APPLIED RESEARCH

Asst Technol 2006;18:000–000

Factors That Influence the Performance of Experienced Speech Recognition Users

Heidi Horstmann Koester, PhD

Koester Performance Research

Performance on automatic speech recognition (ASR) systems for users with physical disabilities varies widely between individuals. The goal of this study was to discover some key factors that account for that variation. Using data from 23 experienced ASR users with physical disabilities, the effect of 20 different independent variables on recognition accuracy and text entry rate with ASR was measured using bivariate and multivariate analyses. The results show that use of appropriate correction strategies had the strongest influence on user performance with ASR. The amount of time the user spent on his or her computer, the user's manual typing speed, and the speed with which the ASR system recognized speech were all positively associated with better performance. The amount or perceived adequacy of ASR training did not have a significant impact on performance for this user group.

Key Words: Assistive technology—Communication aids for disabled—Computer access—Outcomes—Physical disability—Speech recognition.

1. INTRODUCTION

Automatic speech recognition (ASR) systems for computer access allow a user to enter text and commands by speaking to the computer. These systems have the potential to improve the productivity and comfort of computer use for a wide variety of users. ASR may be particularly attractive to people with physical disabilities, when nonspeech methods of computer input, such as the keyboard or mouse, may be either too slow or too painful to fully meet their needs.

Although powerful ASR systems providing recognition of thousands of words have been commercially available for years, our understanding of how well they meet the needs of users with physical disabilities is quite limited. Two key performance benchmarks are recognition accuracy, as a percentage of words correctly recognized by the ASR system, and text entry rate, in corrected words per minute (wpm). For users without disabilities, the small amount of available data suggest that after a short initial training, users may achieve 85% to 93% recognition accuracy and a 14-wpm text entry rate (Devine, Gaehde, & Curtis, 2000; Halverson, Horn, Karat, & Karat, 1999; J. Karat, Horn, Halverson, & Karat, 2000). After 20 hours of use in one study, recognition accuracy improved to 94% and the text entry rate ranged from 25 to 30 wpm (Halverson et al., 1999; J. Karat et al., 2000).

To better understand the performance of speech recognition users who have physical disabilities, we recently measured the performance of 23 experienced ASR users. The text entry rate varied widely, ranging from 3 to 32 wpm, as did the recognition accuracy, ranging from 72% to 94% (Koester, 2003b, 2004). The averages (95% confidence intervals) for text entry rate and recognition accuracy were 16.9 wpm (13.5, 20.3) and 85.0% (82.2, 87.9), respectively. We also followed 8 new users with physical disabilities during their first 6 months of ASR use. Performance on a short paragraph after 4 to 6 weeks of use showed a large range, from 60% to 99%, in recognition accuracy and from 1.5 wpm to 72.6 wpm for text entry rate. After 6 months, only 1 of these 8 individuals was still using the ASR system (Koester, 2003a).

Such wide variation defies a simple answer to the question of the performance that users can expect from ASR systems. This study was conducted to begin to address the following questions:

Address correspondence and reprint requests to Heidi Horstmann Koester, PhD, 2408 Antietam, Ann Arbor, MI 48105.

1. Why do some ASR users perform relatively well and others relatively poorly?
2. Are there patterns that might help us understand the range in performance that we see?
3. If so, can we use these patterns to help ASR users achieve better performance?

1.1. Influential Factors in ASR Performance

The many factors that may influence performance with speech recognition have long been a subject of clinical and theoretical discussions. There has been little research to directly assess the relative strength of these factors, but a review of the possibilities is useful as a conceptual basis for the analyses performed in this study. A more detailed discussion of some of these issues can be found in a recent literature review (Koester, 2001). Factors that may influence ASR performance can be loosely organized into five categories, as described below.

1.1.1. Hardware/Software

This refers to the possibility that the specific computing hardware, computing context, and ASR software chosen may influence resulting performance. Because ASR software requires a relatively high amount of computing resources to run optimally, the general assumption has been that a faster hardware configuration, including a faster processor and increased RAM, will lead to better ASR performance (Cantor, 2001; Gardner-Bonneau, 1999; Lenker, 1998). In addition, the details of a particular installation, such as the brand of ASR software selected, the microphone used, and the application software to be used, have also been identified as potentially influential factors (DeRosier, 2002; Gardner-Bonneau, 1999; Jones, Frankish, & Hapeshi, 1992).

1.1.2. ASR Training and Experience

Training refers to ASR training provided by a competent professional and/or computer-based modules specific to speech recognition. A common assumption is that proper use of an ASR system is complex and that significant training is required to become a skilled independent user (Griffith, 1999; Schwartz & Johnson, 1999). Experience relates to how long users have used a particular ASR system and how frequently they use the system that they have. Given that expertise with any skilled task generally improves with practice and experience, it is expected that a greater amount of usage and experience will yield better performance.

1.1.3. ASR Usage Techniques

This category is specific to how the user actually uses his or her speech recognition system. It covers two general areas: dictation style and correction strategies. Dictation style refers to how quickly the user speaks and how many words each utterance contains. Correction strategies refer to the methods employed by the user to fix recognition errors made by the ASR system. One dimension of this is how errors are corrected, in particular, which of the several correction commands are used. Overuse of the "scratch that" command, for example, may actually lead to degraded performance, whereas using the correction dialogue box is generally considered more appropriate because it helps the system learn from its mistakes (Halverson et al., 1999). Halverson et al.'s (1999) study of 12 novice ASR users without physical impairments observed novices' tendency to use scratch that and avoid the more appropriate correction dialogue. Several studies also suggest that those who use the keyboard and mouse to make corrections within the correction dialogue were able to make faster corrections than those who used speech only (C. Karat, Halverson, Horn, & Karat, 1999; Lewis, 1999; Suhm, Myers, & Waibel, 1999).

A second dimension to correction strategy is when errors are corrected. Halverson et al. (1999) distinguished between a proofreading style, in which the user first dictates a stream of text and then in a second step proofreads to detect and fix the recognition errors, and an in-line style, in which recognition errors are detected and corrected utterance by utterance. They suggested that greater use of a proofreading style is a hallmark of ASR expertise and hence may be associated with better ASR performance.

1.1.4. Computer Experience and Usage

Users who have good knowledge and skills in use of their computer in general may be in a better position to become skilled ASR users (Grott & Schwartz, 2001). It is considered difficult for someone who is a brand-new user of both a computer and an ASR system to try to learn both sets of skills simultaneously. A closer match between the tasks a user needs to do and the capabilities of the ASR system has also been proposed as a factor in better ASR performance, although one study's attempt at demonstrating this quantitatively was unsuccessful (Goette, 1998).

1.1.5. User Characteristics

This covers a wide range of demographic, physical, cognitive, and psychological user character-

istics that might possibly affect performance with ASR, including gender, education, employment status, impairment, manual typing speed (if any), speech quality, motivation, and frustration tolerance.

Table 1 lists each of the candidate factors and a hypothesis for how each might affect user performance with speech recognition. There has been a great deal of speculation about the significance of these factors, and these hypotheses generally have good face validity and stem from direct clinical experience with ASR users. However, there is little or no research that has specifically tested the extent to which any one of these factors is truly correlated with ASR performance, as well as the relative strength of these influential factors.

1.2. Research Goals

This study provides more in-depth analysis of the data from our baseline study of ASR performance for experienced users (Koester, 2003b, 2004). Although those data provide useful information regarding expected recognition accuracy and text entry rate for experienced users, further analysis is needed to determine why some users enjoyed much better performance than others did. In particular, our goal is to determine which of the hypothesized factors described above and in Table 1 really mattered to the performance of this user group. To thoroughly address all of these hypotheses would require a much larger data set than ours comprising 23 experienced users of ASR. However, the data set is sufficiently large and rich to allow for the identification of major trends and to begin to understand the relative strength of the numerous factors that influence ASR performance.

2. METHOD

2.1. Overview

Data from experienced ASR users were analyzed to determine the factors that influenced user performance with ASR. Measurements of recognition accuracy and text entry rate with ASR were the dependent variables. Indicators for 20 potential factors were formed from responses to survey questions and other measures with this same group of users. The relationship between these 20 independent variables (representing the possible factors) and the 2 dependent variables (representing actual user performance) was assessed graphically and statistically using scatter plots, bivariate analyses, and multivariate regression modeling.

2.2. Procedures

Twenty-three experienced ASR users participated. All have physical disabilities that affect their ability to use the standard keyboard and mouse, and all had at least 6 months of ASR experience. Table 2 summarizes the characteristics of this participant group. Each participant completed a 53-item survey in a verbal interview with a researcher. Items covered the following topics: ??2

1. general information, including demographics, education, and employment status;
2. background in computer use and information about their computer and computer usage, including applications used and relative time spent on each;
3. type of ASR system, reasons for using ASR, training received, and what they liked and disliked about ASR;
4. other input methods used, reasons for using them, training received, and likes and dislikes; and
5. relative usage of ASR as compared to other input methods for various computer tasks.

Following the survey, task performance with and without the use of ASR was measured using six word-processing and operating system tasks. The text entry rate and recognition accuracy measures used here were obtained from two text entry tasks: transcription of a 75-word paragraph from hard copy and a short composition on a supplied topic. The tasks were identical for each input condition, except that the transcription text and the composition topic were comparable but not the same. The order of input conditions was counterbalanced across participants. Eighteen of the 23 participants could perform the tasks with a nonspeech alternative. Seventeen of these typed directly on the standard keyboard, and 1 used an onscreen keyboard.

Task performance was videotaped, with a closeup of the user's computer screen and an audio track of their speech. Each user action and corresponding system response, and the time of occurrence for each, was transcribed from the videotape into a spreadsheet log. This log formed the raw data for measuring user performance as well as several of the independent variables.

2.3. Dependent Variables

The primary performance variables were the recognition accuracy, or the percentage of words correctly recognized by the ASR system, and the text entry rate, or how quickly correct text was

TABLE 1. Summary of factors that may influence user performance with automatic speech recognition

Category	Specific factor	Hypothesis
Hardware/software	CPU speed RAM	More hardware resources yield faster text entry rate and possibly better recognition accuracy (Cantor, 2001; Gardner-Bonneau, 1999; Lenker, 1998).
	ASR brand	Some ASR software might provide better recognition than others.
	Microphone	Headset microphones generally provide better recognition accuracy than tabletop microphones (DeRosier, 2002; Gardner-Bonneau, 1999; Jones, Frankish, & Hapeshi, 1992).
	Application software	Dictating into the notepad provided by the ASR system provides better recognition accuracy than dictating directly into third-party application software (DeRosier, 2002; Gardner-Bonneau, 1999; Jones et al., 1992).
ASR training and experience	Amount of training	More training yields better performance (Griffith, 1999; Schwartz & Johnson, 1999).
	Perceived adequacy of training	Better training yields better performance (Griffith, 1999; Schwartz & Johnson, 1999).
	ASR usage ASR experience	More experience yields better performance
ASR usage techniques	Correction strategies	Users who employ more appropriate correction strategies will enjoy better performance (Halverson, Horn, Karat, & Karat, 1999).
	Dictation style	Performance may be influenced by dictation factors such as the number of words in an utterance or dictation speed.
Computer experience and usage	Pre-ASR computer experience	Greater computer experience prior to using ASR provides a more solid foundation for learning ASR, leading to better ASR performance (Grott & Schwartz, 2001).
	Computer usage	Higher computer usage indicates a comfort level with the computer, which would be associated with better ASR performance.
	Tasks performed	Users whose tasks involve more text entry, a strength of ASR, may have better ASR performance (Goette, 1998).
User characteristics	Gender	ASR systems are primarily built on male voice models, so males may achieve better performance.
	Education	Higher education and/or employment status may be associated with better ASR performance, possibly through increased vocational need for productivity or better understanding of sophisticated technology.
	Employment status Need computer for job or school	A user who has a clear vocational or education need for computer productivity may be more motivated to learn effective ASR use and may therefore achieve better recognition accuracy and faster text entry rate.
	Manual typing speed	Conflicting hypotheses: 1. Users who have decent function with a non-ASR input method may not spend the time to get really good at ASR, leading to poorer ASR performance (Schwartz & Johnson, 1999). 2. Users who have decent function with a non-ASR input method might use it to effectively complement their use of ASR, leading to better ASR performance (Karat, Halverson, Horn, & Karat 1999).
	Speech quality	Users who speak more like TV newscasters will have better recognition accuracy.
	Psychology	Users with higher motivation, perseverance, and frustration tolerance will achieve better ASR performance (Grott & Schwartz, 2001).

TABLE 2. Characteristics of the 23 study participants

Participant	Sex	Age	Disability	Education	Speech Impaired	Literacy Difficulty	Need computer for work or school
GC1	M	51	SCI, diabetes	BA	No	No	No
JC2	F	54	SCI, C6-7	BA	No	No	No
BC4	F	40	SCI, C6	BS	No	No	No
KC1	F	31	SCI, C5-6	BA	No	No	No
SC1	M	29	SCI, C5-6	HS	No	No	No
DC1	M	48	SCI, C5	MA	No	No	Yes
SC2	M	24	SCI, C5	MS	No	No	Yes
RC1	M	59	SCI, C4-5	HS	No	No	No
EC2	M	24	SCI, C3-4	Some college	No	No	No
SC4	M	25	SCI, C5	BA	No	No	Yes
EC1	M	27	RSI	MS	No	No	Yes
OC1	M	41	RSI	Some college	No	No	No
SC3	F	28	RSI	MA	No	No	Yes
BC3	F	46	R CVA	PhD	No	No	Yes
BC6	F	47	MS	BS	Yes	No	No
DC2	M	58	MS	MA	No	No	No
TC1	M	54	MS	MA	No	No	Yes
BC2	M	22	MD	HS	Yes	No	No
DC3	M	50	MD	MA	No	No	Yes
AC1	F	22	CP	Some college	No	No	Yes
BC5	F	20	CP	Some college	No	No	Yes
JC1	F	15	Arthogryposis	HS	No	No	Yes
MC1	M	47	ALS	BA	Yes	No	Yes

Note: Presence of speech impairment and literacy difficulty was determined from participant self-report. For the 3 participants who cited speech impairments, all characterized them as "mild" and primarily noticeable when fatigued.

generated using speech recognition. The recognition error rate was calculated as the total number of recognition errors during the two text tasks divided by the number of words spoken. This was then converted to recognition accuracy by subtracting it from 1. The text entry rate was calculated as the number of correct characters produced divided by the number of minutes required to produce them. This measure in units of characters per minute was then divided by 5.5 characters per word to yield the text entry rate in wpm. Note that the text entry rate includes any time spent correcting speech recognition errors and user errors.

2.4. Independent Variables

The goal was to operationalize as many of the candidate factors from Table 1 as possible, using the data collected in the survey and during task performance. Table 3 shows the operational definitions and coding schemes for each of the 20 independent variables used in these analyses. We were not able to operationalize four of the factors in Table 1. We did not measure CPU speed or speech quality, beyond asking whether individuals had impairment. Only 3 participants reported any speech impairment, and all 3 characterized these as "mild" and noticeable primarily when fatigued. There was no variation in ASR brand to test that factor. Finally, we did not examine users' psychological characteristics at a level adequate to use in statistical analyses.

2.5. Bivariate Analyses

A series of bivariate analyses was the first step in determining which independent variables had a consistent influence on the dependent variables. The relationship between each independent variable and each dependent variable was graphed for visual inspection and determined statistically by calculating the Pearson correlation. Analysis of variance (ANOVA) tests were performed for independent variables that were coded categorically. Statistical significance for the correlations and ANOVA tests was set at the .05 level.

2.6. Multivariate Analyses

Given the many variables that may play a role in ASR performance, a purely bivariate analysis is

TABLE 3. Independent variables: measurement methods and coding schemes

Category	Indicator	How Measured	Coding
Hardware and software	RAM	System settings on user's computer	Continuous
	ASR delay	Video. Seconds between conclusion of utterance and appearance of ASR system's recognition, divided by number of words in the utterance	Continuous
	Microphone	Survey and video	Dichotomous: 0 = Headset 1 = Tabletop
	Text application	Video	Dichotomous: 0 = ASR notepad 1 = MS Word
ASR training and Experience	Training hours	Survey	Ordinal: 1 = < = 2 hours 2 = 2 – 10 hours 3 = > 10 hours
	Training adequacy	Survey	Ordinal: 1–7 rating
	ASR usage	Survey. Portion of computer time overall that involves ASR use	Ordinal: 1 = < 25% 2 = 25% – 50% 3 = 51% – 75% 4 = > 75%
	ASR text usage	Survey. Portion of text tasks that involves ASR use	Ordinal: 1 = < 25% 2 = 25% – 50% 3 = 51% – 75% 4 = > 75%
	ASR experience	Survey. Number of years or ASR use	Ordinal: 1 = 6 months – 1 year 2 = 1 – 3 years 3 = > 3 years
ASR usage techniques	Scratch that	Video. Percent of correction episodes in which user said "scratch that"	Continuous
	Proofread	Video. Percentage of correction episodes that involved a proofreading strategy, in which correction occurred after completion of dictation	Continuous
	Words per utterance	Video. Average number of words spoken in one utterance	Continuous
	Dictation speed	Video. Average number of words dictated per minute	Continuous
Computer experience and usage	Computer usage	Survey. Number of hours of computer use per week	Semicontinuous Maximum = 20
	Word proc time	Survey. Portion of user's computer time spent on word processing	Ordinal: 0 = < 25% 1 = > 26%
	Pre-ASR computer experience	Survey	Ordinal: 1 – 7 rating
User characteristics	Gender	Survey	Dichotomous: 0 = Male 1 = Female
	Education	Survey	Ordinal: 1 = Less than high school 2 = High school diploma 3 = Bachelor's degree 4 = Graduate degree
	Need computer for job or school	Survey	Dichotomous: 0 = No 1 = Yes
	Typing speed	Video. Words per minute text entry rate for non-ASR input method. Those without a non-ASR input method were coded as 0	Continuous

not sufficient, either on statistical or conceptual grounds. Therefore, multivariate regression analyses were performed to attempt to determine the relative influence of potential factors while taking other factors into account. From a statistical standpoint, 23 cases are not sufficient to exhaustively examine 20 different independent variables but are adequate for the development of a two- or three-factor model, using the rule of thumb of 6 to 10 cases per independent variable (Neter, Wasserman, & Kutner, 1990). Multiple regression models were developed for each of the two dependent variables (recognition accuracy and text entry rate) using the following procedures.

2.6.1. Reduction of Independent Variables

Given the large number of independent variables, the first step in the model-building process was reducing the potential pool to a reasonable size. A combination of theoretical and statistical criteria governed this reduction process. Our primary criterion was to test the independent variables of greatest theoretical interest; specifically, these were variables related to ASR training (training hours and training adequacy), correction strategies (scratch that), and manual typing speed (typing speed). Secondary statistical criteria were based on the bivariate relationships. Multivariate influence was examined for any independent variable that had (a) a visible bivariate relationship on the scatter plot and (b) a statistically significant bivariate correlation with any dependent variable or a bivariate correlation greater than .2 (absolute value) with any dependent variable. These criteria are admittedly somewhat arbitrary but were designed to screen out variables with very little relationship to ASR performance while being conservative in retaining those that might have even a small influence.

We could examine with caution only those variables with a relatively uneven distribution over their range (e.g., only five or six entries in one of the two factor levels); microphone, time spent on word processing tasks (word proc time), and training adequacy fall into that category. We were also cautious with respect to variables that effectively duplicate another independent variable in the list of candidates because of high correlation between two independent variables.

2.6.2. Model Selection and Refinement

The remaining candidate factors after the reduction process were then examined for suitability in a multivariate model. The first step was to find the "best" one-factor model, then determine if any of the remaining factors significantly improved the model enough to warrant a two-factor model. If a two-factor model was found, the remaining factors were again searched for a possible three-factor model. "Best models" at each step were selected based primarily on statistical criteria because the theoretical criteria were already satisfied in reducing the independent variables. A model was judged to be "better" than another if it had a higher adjusted R^2 value, greater statistical significance for each independent variable's model coefficient, stronger partial relationships based on graphic analysis, and more robust satisfaction of regression assumptions. Linear regression model assumptions of linearity, constant variance of residuals, and normally distributed residuals were examined in depth using plots of the residuals (Neter et al., 1990). Cases with excess influence were detected by plotting Cook's distance versus central leveraged value. The effect of these cases was determined by removing them and remodeling; any model highly sensitive to the presence of one or two specific cases was considered to be less robust than other models.

2.6.3. Model Validation

The model selection process yielded one or two candidate models for each dependent variable. Each of these models underwent a more thorough model validation to further examine its theoretical and statistical suitability. Theoretical suitability was assessed by comparing the model structure to theoretical expectations. Statistical suitability was assessed by using a hold-out sample to verify the model structure and the stability of the coefficient values (Neter et al., 1990). The coefficients were remodeled using one half of the original data sample; the cases for the selected sample were selected randomly.

2.6.4. Criteria for Model Interpretation

The main purpose of the multivariate modeling is to identify influential factors and their relative influence on ASR performance. An independent variable was considered to be an influential factor if its standardized Beta coefficient in a multivariate model was significant at the less than .05 level. The relative strength of two or more influential factors in a single model was determined by comparing their standardized Beta coefficients. A secondary purpose was to use the model equations to reason about performance under various conditions. Proper use of regression models for this lat-

TABLE 4. Independent variables and their Pearson correlations with recognition accuracy (Rec Acc) and text entry rate (TER)

		Bivariate correlation	
Category	**Variable**	**Rec Acc**	**TER**
Hardware/software	RAM	0.024	0.158
	ASR delay	0.085	−0.356
	Microphone	−0.152	−0.105
	Text application	−0.081	0.010
ASR training/usage	Training hours	0.001	−0.114
	Training adequacy	0.190	−0.127
	ASR usage	0.090	−0.010
	ASR text usage	0.419*	0.227
	ASR experience	−0.078	−0.004
ASR techniques	Scratch that	−0.681**	−0.598**
	Proofread	0.266	−0.003
	Words per utterance	0.315	0.559**
	Dictation speed	0.132	0.426
Computer experience and usage	Computer usage	0.251	−0.053
	Word proc time	0.413*	0.355
	Pre-ASR experience	−0.311	−0.198
User characteristics	Gender	−0.078	−0.147
	Education	0.338	0.371
	Need computer for job/school	0.397	0.478*
	Typing speed	0.189	0.610**
Other ASR factors	Recognition accuracy	1.0	0.687**

Note: ASR = automatic speech recognition.
*$p < .05$.
**$p < .01$.

ter purpose is possible if the model meets all linear regression assumptions.

3. BIVARIATE RESULTS

A primary purpose of the bivariate analysis was to reduce the pool of independent variables to a more manageable subset that could be analyzed in a multivariate sense. Table 4 shows the Pearson correlations between each of the candidate factors (the independent variables) and the two performance measures (the dependent variables). The following sections describe the extent to which variables in each category met the statistical criteria for further analysis in the multivariate models.

3.1. Recognition Accuracy

3.1.1. Hardware/Software Factors

None of the hardware/software factors met the statistical criteria, so none were retained for the multivariate analyses. The test for Microphone was limited to some extent because only 5 of the 23 participants used tabletop microphones, but at least for these few participants, microphone type made no difference in ASR performance. The text application factor showed a relatively even distribution, and the averages for text entry rate and recognition accuracy were almost identical whether users dictated into the ASR notepad or Microsoft Word.

3.1.2. ASR Training/Usage Factors

Neither of the ASR training factors met the statistical criteria. Figure 1 shows that amount of training did not significantly affect recognition accuracy for this participant group (ANOVA $p = .284$). Because training is such a commonly postulated factor in ASR performance, however, the training indicator with the strongest bivariate relationship to recognition accuracy (training adequacy) was retained for further analysis in the multivariate models. ??3

The amount of ASR usage overall did not show a connection to recognition accuracy, but the amount of ASR usage for text tasks in particular is positively correlated to recognition accuracy. This is primarily because the 4 people who reported using ASR for less than 25% of their text tasks also had relatively low recognition accuracy. A somewhat surprising result is that experience with ASR

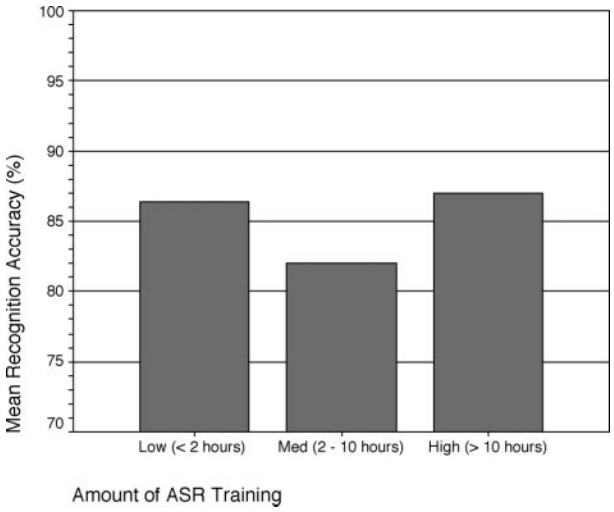

FIG. 1. Recognition accuracy versus amount of automatic speech recognition (ASR) training. Note: For low training, $n = 9$; for medium training, $n = 8$; for high training, $n = 6$.

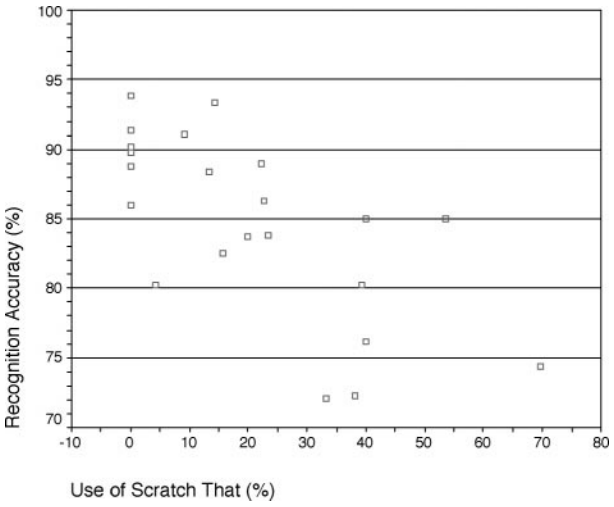

FIG. 2. Scatter plot of recognition accuracy as a function of the use of the "scratch that" correction strategy.

had essentially no relationship to recognition accuracy, with a correlation of −.078. This may be because all of the participants in this study had at least 6 months' experience using their ASR system.

3.1.3. ASR Techniques Factors

Indicators within the category of ASR techniques had a relatively strong relationship to recognition accuracy. More frequent use of scratch that, in particular, was significantly associated with lower recognition accuracy, as shown in Figure 2. The use of a proofreading approach to correcting recognition errors, as well as using a higher number of words per utterance, showed some association with higher recognition accuracy. These correlations were not statistically significant ($p = .266$ and $p = .144$, respectively) but did meet our internal criteria of exceeding .2 and were therefore retained for the multivariate modeling.

3.1.4. Computer Experience and Usage Factors

In the category of general computer usage and experience, greater time spent on word-processing tasks was significantly correlated with higher recognition accuracy. The 17 participants who reported spending less than 25% of their computer time on word processing had an average recognition accuracy of 83.4%. The 6 participants who use word processing 25% of the time or more had an average recognition accuracy of 89.5%, a difference that is statistically significant at $p = .05$. In addition, more hours of computer use overall also showed some association with higher recognition accuracy, although not at the $p = .05$ level. In a counterintuitive result, the amount of computer experience before using ASR was negatively associated with recognition accuracy, although again, the correlation was not statistically significant ($p = .148$). This may be a spurious result because it has no clear theoretical rationale. However, because it meets the statistical criteria, it was retained for further analysis.

3.1.5. User Characteristics Factors

For the category of user characteristics, both level of education and needing a computer for job or school showed similar positive associations with recognition accuracy, although they not were statistically significant ($p = .114$ and $p = .061$, respectively). Figure 3 illustrates that the 12 participants who needed their computer for job or school responsibilities had an average recognition accuracy about 10 percentage points higher than the remaining 11 who did not, but the width of the confidence intervals kept the difference from being statistically significant. Gender showed no real relationship to recognition accuracy. ??4

3.1.6. Summary of Bivariate Analyses

Across all independent variables examined for influence on recognition accuracy, 10 were retained for further analysis in the multivariate models. These are ASR training adequacy, ASR text usage, scratch that, proofread, words per utterance, computer usage, word proc time, pre-ASR experience, education, and need computer for job/school. Only relatively weak bivariate relation-

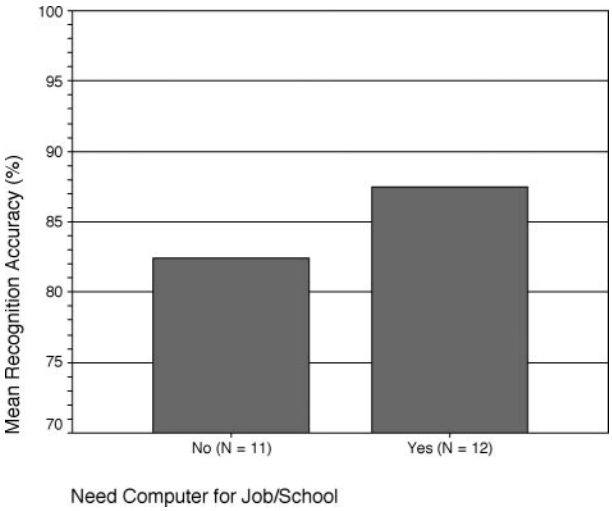

FIG. 3. Recognition accuracy versus a user's need for a computer for job or school requirements. Note: The average recognition accuracy is higher for users who needed their computer for job or school, but the difference is not statistically significant ($p = .061$).

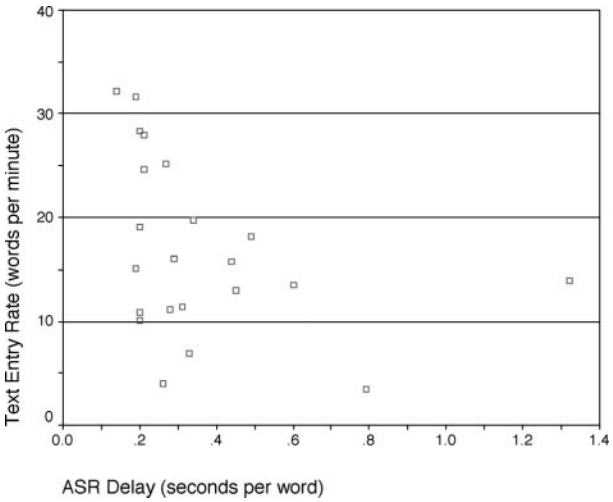

FIG. 4. Scatter plot of text entry rate as a function of the time required for the automatic speech recognition (ASR) system to display its recognitions.

ships were found for factors related to hardware and software, although that may be partly due to limited distribution across all categories. ASR training factors showed surprisingly little connection to recognition accuracy, as did the amount of experience participants had using ASR.

3.2. Text Entry Rate

3.2.1. Hardware / Software Factors

ASR delay had a correlation of $-.356$ ($p = .104$) with text entry rate, which meets the criteria for further consideration in the multivariate analyses. ASR delay is the amount of time required for the system to determine and display its recognition of the user's utterance, measured as the average number of seconds per spoken word. Although there is no requirement for the user to wait until one utterance is recognized before beginning a new one, in practice, many users wait to see if any corrections are necessary before continuing on to the next utterance. Therefore, it is expected that a longer ASR delay would lead to a slower text entry rate, and this is supported by these results (see Fig. 4). However, as Figure 4 also shows, a significant portion of the relationship may be due to the presence of two cases with especially long values for ASR delay. This will be examined further in the multivariate analyses. None of the other hardware or software factors had a notable influence on text entry rate.

3.2.2. ASR Training / Usage Factors

Neither of the ASR training factors was close to meeting the statistical criteria for bivariate influence, so neither was retained for further analysis. The amount of ASR usage overall was unrelated to text entry rate, but the amount of ASR usage for text tasks in particular showed some positive association with text entry rate. This may be simply because ASR text usage is correlated to recognition accuracy, which is in turn correlated with text entry rate. The multivariate analyses will determine if there is any independent influence of this factor.

3.2.3. ASR Techniques Factors

Indicators within the category of ASR techniques had a relatively strong relationship to text entry rate. More frequent use of scratch that was strongly associated with slower text entry rate, with a correlation of $-.598$ (significant at $p < .05$). In addition, faster dictation speeds correlated with faster text entry rate ($r = .426$), as expected. Words per utterance also met criteria. However, this variable was not retained for multivariate analysis for two reasons: (a) it is strongly correlated (at $p < .05$) with two other variables that met the criteria, dictation speed and typing speed, and (b) of the three variables, words per utterance has the weakest theoretical ties to text entry rate. The use of a proofreading approach to correcting recognition errors showed no association with text entry rate.

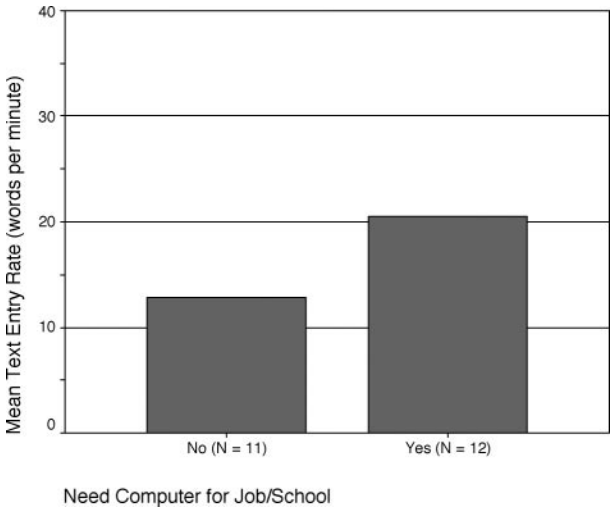

FIG. 5. Text entry rate versus a user's need for a computer for job or school requirements. Note: The average text entry rate is higher for users who needed their computer for job or school (significant at $p = .021$).

3.2.4. General Computer Usage Factors

In the category of general computer usage and experience, greater word proc time was correlated with higher text entry rate, although this may be because word proc time is also correlated with recognition accuracy. Amount of computer use overall showed no association with text entry rate. The amount of computer experience before using ASR was negatively associated with text entry rate, similar to the recognition accuracy results, although this time, the correlation did not meet the statistical criteria for inclusion in multivariate analysis.

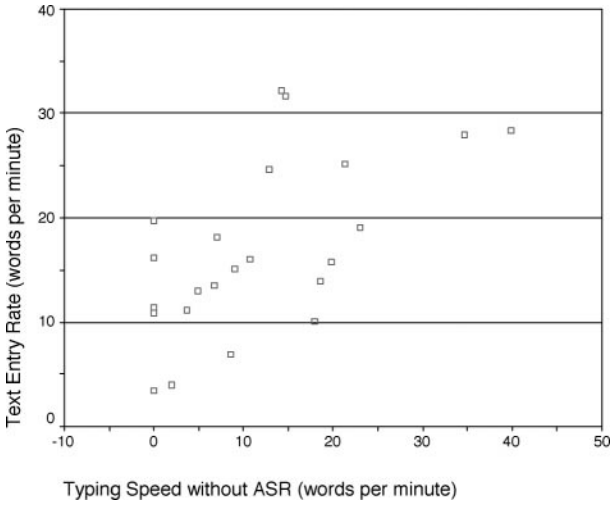

FIG. 6. Scatter plot of text entry rate as a function of manual typing speed. Note: Typing speed was coded as a 0 for users who did not have a non–automatic speech recognition (ASR) method of text entry rate.

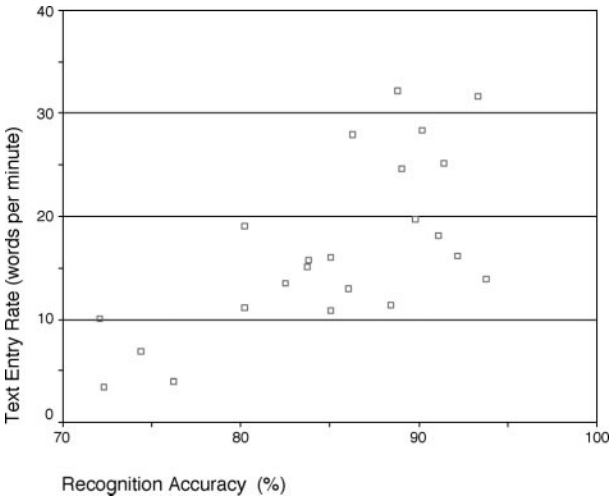

FIG. 7. Scatter plot of text entry rate as a function of recognition accuracy.

3.2.5. User Characteristics Factors

For the category of user characteristics, both level of education and needing a computer for job or school showed positive associations with text entry rate, as they did for recognition accuracy. Figure 5 illustrates that the 12 participants who needed their computer for job or school responsibilities averaged 20.6 wpm with their ASR system, as compared to an average of 12.9 wpm for those who did not need a computer for job or school (significant at $p = .021$). Manual typing speed was also strongly associated with text entry rate (see Fig. 6); that is, users who could type faster without ASR also tended to be faster with ASR. Gender showed no real relationship to text entry rate.

3.2.6. Influence of Recognition Accuracy

We also examined the bivariate relationship between text entry rate and recognition accuracy, primarily because of its theoretical naturalness. Achievement of good recognition accuracy is typically considered to be a prerequisite to fast text entry rate with ASR. As shown in Figure 7, recognition accuracy and text entry rate are indeed highly related for this data set, with a correlation of .687 (significant at $p < .001$). However, recognition accuracy is not the sole determiner of text entry rate. One purpose of the multivariate analyses is to determine what additional factors influence text entry rate, once the effect of recognition accuracy is taken into account.

TABLE 5. Model statistics for bivariate regression model of recognition accuracy as a function of scratch that usage

Factor	Partial Beta (β)	Significance of β	Adjusted R^2
Scratch that	−.681	<.001**	.437

**$p < .01$.

3.2.7. *Summary of Bivariate Analyses*

Across all independent variables examined for influence on text entry rate, nine were retained for further analysis in the multivariate models. These are ASR delay, ASR text usage, scratch that, dictation speed, word proc time, education, need computer for job/school, typing speed, and recognition accuracy. ASR training factors showed relatively little relationship to text entry rate, as did the amount of experience participants had had with ASR.

4. MULTIPLE REGRESSION RESULTS

4.1. Recognition Accuracy

As described above, the goal of multivariate modeling for recognition accuracy is to determine the factors that contribute to high recognition accuracy and the relative strength of their contributions. The process used was to find the "best" models involving one, two, and three independent variables, as described in Section 2.6.2. Based on the bivariate results, the 10 independent variables considered in these multivariate analyses were ASR training adequacy, ASR text usage, scratch that, proofread, words per utterance, computer usage, word proc time, pre-ASR experience, education, and need computer for job/school.

4.1.1. *Model With One Independent Variable*

From the bivariate analyses, the scratch that variable emerged as the strongest single bivariate factor in recognition accuracy. The full regression model is

$$\text{RecAcc} = 89.4 - (0.225)(\text{ScratchThat}). \quad (1)$$

Table 5 shows the key statistics for this model. The coefficients of the model suggest that every decrease of 10 percentage points in the use of scratch that yields 2.25 percentage points of improvement in recognition accuracy. The single-factor model of recognition accuracy as a function of scratch that meets linear regression assumptions, based on plots of the residuals, and has no cases with undue influence.

4.1.2. *Model With Two Independent Variables*

To determine the best two independent variable (2-IV) model, each of the nine remaining independent variables was included in the model with scratch that. Table 6 shows the results for each of the resulting nine models.

The factor of computer usage emerged as the strongest second independent variable, explaining an additional 10% of the variance in recognition accuracy. This was a bit of a surprise, given the relatively low bivariate correlation for computer usage. The standardized model coefficient was the highest of those tested and was the only one significant at $p < .05$. The full model is

$$\text{RecAcc} = 85.456 - (0.247)(\text{ScratchThat}) + (0.480)(\text{CompUsage}). \quad (2)$$

The effect of scratch that in this 2-IV model is relatively unchanged from the single-factor model of

TABLE 6. Statistics for each of nine candidate factors for the second independent variable in a regression model of recognition accuracy

Factor	Partial Beta (β)	Significance of β	Adjusted R^2
ASR training adequacy	.091	.616	.398
ASR text usage	.208	.235	.451
Proofread	.125	.471	.423
Words per utterance	.271	.125	.478
Computer usage	.347	.034*	.535
Word proc time	.123	.500	.421
Pre-ASR experience	−.087	.618	.415
Education	.152	.384	.431
Need computer for job/school	.048	.805	.409

Note: Scratch that was used as the first independent variable in each of the nine models. ASR = automatic speech recognition.

*$p < .05$.

TABLE 7. Statistics for each of eight candidate factors for the third independent variable in a regression model of recognition accuracy

Factor	Partial Beta (β)	Significance of β	Adjusted R^2
ASR training adequacy	.073	.892	.499
ASR text usage	.165	.306	.538
Proofread	.081	.612	.516
Words per utterance	.227	.163	.561
Word proc time	.038	.824	.511
Pre-ASR experience	−1.119	.093	.582
Education	.171	.286	.540
Need computer for job/school	.012	.945	.509

Note: Scratch that and computer usage were the first and second independent variables in each of the eight models. ASR = automatic speech recognition.

Equation 1, which points to good model stability. The coefficient for computer usage suggests that each additional hour of computer use per week may add approximately 0.5 percentage points to recognition accuracy.

This model meets linear regression assumptions for the most part, although the distribution of the residuals deviates from normal to a greater degree than in the one-factor model. No cases with extreme influence were found, but there were two cases with a notably higher centered leverage value than the rest of the data set. The sensitivity of the model to these two cases is explored in more depth in Section 4.1.4.

4.1.3. Model With Three Independent Variables

The final model-building step for recognition accuracy was to determine whether a third influential factor could be found. Each of the eight remaining independent variables was modeled with scratch that and computer usage to see if the model fit would significantly improve. Table 7 shows the results for each of the resulting eight models.

None of the factors tested adds significantly to the model, although the surprising negative association with pre-ASR computer experience reemerged to some extent in the 3-IV model. In addition, the model coefficients for scratch that and computer usage remained significant at the $p = .05$ level in each of the three-factor models tested. This strengthens our confidence in the robustness of the two-factor model.

4.1.4. Model Validation

Two plausible models for recognition accuracy emerged from the model-building process: (a) a single-factor model with scratch that as the predictor (RA Model 1, shown in Equation 1) and (b) a two-factor model with scratch that and computer usage as the predictors (RA Model 2, shown in Equation 2). Although the two-factor model explains more of the variance in recognition accuracy, the single-factor model may still be useful in cases in which not much is known about the Computer Usage factor. Therefore, validation was performed for both models.

The theoretical suitability of RA Model 1 is strong. A significant negative relationship between the use of scratch that and recognition accuracy is to be expected given the way in which speech recognition systems work. Users who repeatedly correct recognition errors using "scratch that" are effectively telling the ASR system to ignore what they just said, thus defeating the ability of the system to learn from its mistakes. If this pattern is continued over time, the lack of learning by the ASR system can degrade recognition accuracy. The theoretical suitability of RA Model 2 is strong as well because it also incorporates scratch that. The rationale for the second predictor, computer usage, is also reasonable, if somewhat less compelling, however. Simply using one's computer more frequently may lead to better performance on computer tasks, and recognition accuracy with ASR is one indicator of that improved performance. However, by similar reasoning, the amount of ASR use in particular would also be expected to have a positive relationship with recognition accuracy, but it does not, for this data set. Therefore, the exact source of the relationship between computer usage and recognition accuracy is not clear, but it does have plausible theoretical grounds.

Statistically, a split-data procedure was used to determine the robustness of each model. Twelve of the 23 cases were randomly selected to form Split-Half 1; the remaining 11 cases formed Split-Half 2. Models for recognition accuracy using scratch that and computer usage as predictors were calculated

TABLE 8. Split-half model validation results for a single-factor regression model of recognition accuracy (RA) as a function of scratch that (ST)

Data set	Model equation	Partial β	Significance of β	Adjusted R^2
Split-half 1	RA = 88.7 − 0.21 (ST)	−.503	.115	.170
Split-half 2	RA = 89.9 − 0.24 (ST)	−.887	<.001**	.763
All Data	RA = 89.4 − 0.23 (ST)	−.681	<.001**	.437

Note: Scratch that explains much more variance in recognition accuracy for Split-half 2 (n = 11) than Split-half 1 (n = 12).
**$p < .01$.

using each subgroup's data. Results are shown in Tables 8 and 9.

These results generally support the full data set conclusion that scratch that is the most influential factor on recognition accuracy. However, it should be noted that scratch that is not statistically significant as a single factor for Split-Half 1 data. For this group, the effect of computer usage was relatively strong, and indeed once this second factor was added, the significance of scratch that closely approached the .05 level. For Split-Half 2, in contrast, scratch that was an important factor in both the one- and two-factor models, and computer usage had very little influence of recognition accuracy for this group. Computer usage appears to be a less consistent influence across all participants than scratch that.

During testing of the regression assumptions for RA Model 2, two data points were observed to have higher values for centered leverage than the rest of the data set. This suggests that these two data points might have more influence on the model than perhaps they should. To test the extent of this influence, RA Model 2 was refit without the two influential data points. When this was done, the values for the model parameter estimates (for the constant and two Beta coefficients) were relatively unchanged, but the coefficient for computer usage was no longer statistically significant (p = .226). This does not necessarily mean that computer usage is not a truly significant factor, but, especially in combination with the split-half modeling results above, it does indicate that its influence is weaker and less robust than scratch that.

On both theoretical and statistical grounds, therefore, the use of scratch that is the most robust and influential predictor of recognition accuracy with ASR of the 20 independent variables tested for this data set.

4.2. Text Entry Rate

The modeling process for text entry rate was very similar to that followed for the recognition accuracy models. To determine the factors that contribute to faster text entry rates and the relative strength of their contributions, we searched for the "best" models involving one, two, and three independent variables. The nine independent variables employed in the multivariate analyses were ASR delay, ASR text usage, scratch that, dictation speed, word proc time, education, need computer for job/school, typing speed, and recognition accuracy.

4.2.1. Model With One Independent Variable

From the bivariate analyses, recognition accuracy emerged as the strongest single bivariate factor in text entry rate (TER). The full regression model is

TABLE 9. Split-half model validation results for a two-factor regression model of recognition accuracy (RA) as a function of scratch that (ST) and computer usage (CU)

		Scratch that		Computer usage		
Data set	Model equation	Partial β	Significance of β	Partial β	Significance of β	Adjusted R^2
Split-half 1	RA = 81.7 − 0.21 (ST) + 0.77 (CU)	−.520	.055	.564	0.041*	0.464
Split-half 2	RA = 90.0 − 0.24 (ST) − 0.01 (CU)	−.883	.001**	.010	0.957	0.733
All data	RA = 85.5 − 0.25 (ST) + 0.48 (CU)	−.748	<.001**	.347	0.034*	0.535

Note: The effect of computer usage is significant only for Split-half 2.
*$p < .05$.
**$p < .01$.

TABLE 10. Statistics for a one-factor regression model of text entry rate as a function of recognition accuracy

Factor	Partial Beta (β)	Significance of β	Adjusted R^2
Recognition Accuracy	0.687	<0.001**	0.446

**$p < .01$.

$$\text{TER} = -56.392 + (0.862)(\text{RecAcc}). \quad (3)$$

Table 10 shows the key statistics for this model. Based on the model, those who achieve a recognition accuracy of 85% (which is approximately the average for this data set) may expect a text entry rate of 15.9 wpm. (The actual observed average was 16.9 wpm.) Improving recognition accuracy to 95% leads to a predicted text entry rate of 24.4 wpm, whereas decreasing it to 75% results in a text entry rate estimate of 7.4 wpm. In Figure 7, a strong bivariate relationship can be seen, but is also clear that recognition accuracy is not the only factor in text entry rate. This single-factor model of text entry rate does meet linear regression assumptions, based on plots of the residuals; the relationship is linear, residuals have approximately constant variance, and their distribution is roughly normal.

4.2.2. Model With Two Independent Variables

Because scratch that and recognition accuracy are so closely correlated ($r = -.681, p < .001$), the presence of both of those factors in a single model is not appropriate. Therefore, there were seven remaining independent variables to consider in the two-factor model. (Models for text entry rate as a function of scratch that are presented in Section 4.2.5.) Table 11 shows the results.

Three independent variables, ASR delay, dictation speed, and typing speed, emerged as significant factors influencing text entry rate, at the $p = .05$ level, in conjunction with recognition accuracy. Typing speed appears to have the strongest influence of the three, based on its stronger Beta significance and higher adjusted R^2. The addition of typing speed explains almost 25% more variance in the text entry rate than recognition accuracy alone. The full model is

$$\text{TER} = -50.729 + (0.744)(\text{RecAcc}) + (0.373)(\text{Typing Speed}). \quad (4)$$

The coefficient for typing speed suggests that a user who can manually type 15 wpm may enjoy a text entry rate with ASR that is approximately 4 wpm faster than that achieved by a user who can manually type at 5 wpm.

Testing the regression assumptions suggests that the model does have some flaws, despite its high R^2 value. The model meets the linearity assumption for linear regression models. However, it exhibits some nonconstant variability in the residual values (heteroskedasticity), such that for higher values of recognition accuracy and for typing speeds between 10 and 20 wpm, the model error is relatively higher. In addition, the residuals show some deviation from normality. This does not affect the significance of the independent variables, but it does reduce the confidence of using the model as an equation to simulate text entry rate under different conditions.

One data point (BC3) was identified as having a moderately high Cook's distance of 0.4. This is below the suggested "high" value of 0.7 for a two-factor model (McDonald, 2002) but nonetheless may exert extra influence on the results. The model was refit without BC3, resulting in minor changes to

TABLE 11. Statistics for each of seven candidate factors for the second independent variable in a regression model of text entry rate

Factor	Partial Beta (β)	Significance of β	Adjusted R^2
ASR delay	−.419	.004**	.648
ASR text usage	−.073	.685	.424
Word proc time	.086	.634	.426
Dictation speed	.345	.043*	.507
Education	.156	.366	.443
Need computer for job/school	.243	.165	.474
Typing speed	.498	.001**	.682

Note: Recognition accuracy was the first independent variable in each of the seven models. ASR = automatic speech recognition

*$p < .05$.
**$p < .01$.

TABLE 12. Statistics for each of six candidate factors for the third independent variable in a regression model of text entry rate

Factor	Partial Beta (β)	Significance of β	Adjusted R^2
ASR delay	−.355	.002**	.808
ASR text usage	.124	.393	.679
Word proc time	−.037	.793	.667
Dictation speed	.106	.535	.623
Education	.092	.705	.674
Need computer for job/school	.086	.548	.672

Note: Recognition accuracy and typing speed were the first and second independent variable in each of the six models. ASR = automatic speech recognition.
**$p < .01$.

the model coefficients and yielding a stronger model overall, with notable improvements in the adjusted R^2 (to .804) and the normality of the residuals. The revised model equation without case BC3 was

$$\text{TER} = -60.76 + (0.866)(\text{RecAcc}) + (0.395)(\text{Typing Speed}). \quad (5)$$

The refitting of the model provides a concrete example of the need for caution in interpreting exact model coefficient values. It also illustrates how a single case can have fairly large influence for relatively small data sets. However, the form of the model and high statistical significance of each factor in it remain unchanged.

4.2.3. Model With Three Independent Variables

The final model-building step for text entry rate was to determine whether a third influential factor could be found. The most likely candidates were ASR delay and dictation speed because both emerged as significant in the 2-IV modeling step. But the other remaining variables were also considered for the sake of completeness. Table 12 shows the results for each of the resulting six models.

Table 12 shows that adding ASR delay to the model adds significant explanatory power, and it is the only added factor that does so. Dictation speed, for example, which was significant in a two-factor model, no longer has significance once the effect of typing speed has been taken into account. The full three-factor model equation is

$$\text{TER} = -52.553 + (0.822)(\text{RecAcc}) + (0.316)(\text{Typing Speed}) - (11.242)(\text{ASR Delay}). \quad (6)$$

Examining the residuals showed that the regression assumptions hold better for this model than the two-factor model does. The problem with heteroskedasticity has been resolved with the addition of the third factor, and the distribution of the residuals is closer to normal. Two data points (BC2 and BC3) were identified as having relatively high Cook's distance and/or centered leverage values. Refitting the model without these two data points yielded a very similar model, however, but with a decrease in the significance of the ASR delay factor, from $p = .002$ to $p = .059$.

4.2.4. Model Validation

Three plausible models for text entry rate emerged from the model-building process: (a) a single-factor model with recognition accuracy as the predictor, (b) a two-factor model with recognition accuracy and manual typing speed the predictors, and (c) a three-factor model with recognition accuracy, typing speed, and ASR delay as the independent variables. The theoretical suitability of these three factors is strong. A significant positive relationship between recognition accuracy and ASR text entry rate is certainly expected. An association between manual typing speed and ASR text entry rate also makes conceptual sense. ASR users with faster manual typing skills may be able to correct recognition errors faster, leading to faster overall performance. In addition, good typing skills provide a better alternative to scratch that when the ASR system garbles a long phrase. Finally, part of the effect of typing speed may simply be that some people are more "speed focused" than others, such that, all other things being equal, the speed-focused individual is likely to outperform his or her peers on any practiced task. The influence of ASR delay is also sensible because many users wait to see what the ASR system outputs are before moving on to the next utterance. The longer they have to wait, the slower their overall performance will be.

TABLE 13. Split-half model validation results for a single-factor regression model of text entry rate (TER) as a function of recognition accuracy (RA)

Data set	Model equation	Partial Beta β	Sig. of β	Adjusted R^2
Split-half 1	TER = −53.6 + 0.838 (RA)	.728	.007**	.484
Split-half 2	TER = −61.6 + 0.907 (RA)	.648	.031*	.356
All Data	TER = −56.4 + 0.862 (RA)	.687	<.001**	.446

Note: Recognition accuracy explains a similar amount of variance in text entry rate for both halves.
*$p < .05$.
**$p < .01$.

Statistically, a split-data procedure was used to help determine the robustness of the one- and two-factor models. (The three-factor model was not analyzed in this way due to an inappropriate number of data points per factor when splitting the data set in half.) Twelve cases were randomly selected to form Split-Half 1. The remaining 11 cases were Split-Half 2. Models for text entry rate using recognition accuracy and manual typing speed as predictors were fit using each subgroup's data. Results are shown in Tables 13 and 14.

The split-half results show fairly consistent model coefficient values and structure. In the two-factor model for Split-Half 2, the strength of each factor is weaker than for Split-Half 1, and the partial coefficient for typing speed in particular is noticeably lower in that group. The weakened fit for Split-Half 2 is partly due to the reduced statistical power in fitting a two-factor model with 11 data points and partly due to the fact that typing speed was not in fact as strongly influential for Split-Half 2 as for Split-Half 1.

4.2.5. Alternative Expression of Text Entry Rate Models

The models for text entry rate developed above used recognition accuracy as a key independent variable. This supports an intuitive understanding of how speech recognition works and ensures that recognition accuracy, as a clinically relevant and readily measurable variable, is explicitly represented in the model. However, given the strong relationship between recognition accuracy and scratch that, it is also possible and at times desirable to represent text entry rate as a function of scratch that. This allows the use of a common set of independent variables when working with the models for recognition accuracy and text entry rate. Refitting the one-, two-, and three-factor models for text entry rate using scratch that instead of recognition accuracy yields the following:

$$\text{TER} = 22.084 - 0.249(\text{Scratch That}) \quad (7)$$

$$\text{TER} = 16.688 - 0.183(\text{Scratch That}) + 0.342(\text{Typing Speed}) \quad (8)$$

$$\text{TER} = 22.410 - 0.218(\text{Scratch That}) + 0.281(\text{Typing Speed}) - 11.873(\text{ASR Delay}). \quad (9)$$

As shown in Table 15, the coefficients for all model parameters are significant at the $p = .05$ level.

5. DISCUSSION

A user's performance with speech recognition software is a complex multivariate construct. Numerous factors may combine to influence performance, most likely in different ways for different individuals. These analyses address the questions of which factors seem to have influence across mul-

TABLE 14. Split-half model validation results for a two-factor regression model of text entry rate (TER) as a function of recognition accuracy (RA) and typing speed (TS)

		Recognition accuracy		Typing speed		
Data set	Model equation	Partial β	Significance of β	Partial β	Significance of β	Adjusted R^2
Split-half 1	TER = −55.5 + 0.80 (RA) + 0.47 (TS)	.691	<.001**	.615	<.001**	.887
Split-half 2	TER = −47.9 + 0.71 (RA) + 0.28 (TS)	.510	.080	.385	.169	.473
All data	TER = −50.7 + 0.74 (RA) + 0.37 (TS)	.593	<.001**	.498	.001**	.682

Note: The model explains more variance in text entry rate for the randomly selected Split-half 1.
**$p < .05$.

TABLE 15. Model statistics for regression models of text entry rate using scratch that instead of recognition accuracy as an independent variable

Model	Factor	Partial Beta (β)	Significance of β	Adjusted R^2
Equation 7	Scratch That	−.598	.003**	.325
Equation 8	Scratch That	−.438	.016*	.492
	Typing Speed	.457	.013*	
Equation 9	Scratch That	−.523	.002**	.625
	Typing Speed	.377	.019*	
	ASR Delay	−.384	.012*	

Note: ASR = automatic speech recognition.
*$p < .05$.
**$p < .01$.

tiple individuals and which ones do not. This discussion summarizes what we have learned about these questions, addresses some limitations of the study, and attempts to draw some clinically relevant conclusions.

5.1. Factors With Significant Influence

Of the 20 independent variables examined, 4 had a significant multivariate effect on recognition accuracy or text entry rate. These were scratch that, computer usage, typing speed, and ASR delay. Only scratch that had a significant influence on both dependent variables, which suggests that it is the strongest factor among those studied for this participant group. The scratch that variable is the percentage of times a user employed the "scratch that" correction strategy when fixing a recognition error. Because this strategy tells the ASR system to disregard what was just said, it is most appropriately used as a means of correcting misspoken words or inadvertent utterances, such as coughs or sneezes. Most users in this study, however, used scratch that much more generally to erase recognition of properly spoken utterances. And those who used scratch that more often tended not to use the more appropriate "correct that" strategy, which allows the speech recognizer to learn from its mistakes. These results corroborate what has long been part of well-designed ASR training: less frequent use of scratch that yields better recognition accuracy and text entry rate. Unfortunately, this principle presented in training does not seem to have been integrated into many users' behavior, as even some of those who received more than 10 hours of training exhibited overuse of "scratch that." (The correlation between scratch that use and ASR training hours was only .003.)

The results also confirm the intuitive suggestion that more computer use overall will yield better recognition accuracy, although there are two caveats that soften the strength of this conclusion. First, the amount of ASR usage was not associated with increased recognition accuracy, with a bivariate correlation of only .090. This detracts from the theoretical strength of concluding that more use leads to better performance. Second, computer usage, by itself, does not improve recognition accuracy. Increased computer usage appears to be helpful only when the proper correction strategy is used. In addition, the computer usage factor, although a significant predictor for the group as a whole, was not a highly robust factor, which detracts from its statistical strength.

The importance of ASR delay in influencing the text entry rate confirms the expected result that faster system recognition leads to a faster text entry rate. Available system RAM and CPU speed are two factors that may enhance recognition speed. This study did not examine the specific influence of CPU speed and found that total system RAM alone was not associated with recognition accuracy or text entry rate. These results support the common clinical recommendation to get the best computer hardware possible when running speech recognition but do not offer specific guidelines about what is most important in that hardware.

Typing speed emerged as a significant influence on text entry rate, although as discussed above, the primary reason for this result is not quite clear. One possibility is that faster typing leads to faster correction of recognition errors, when users type the correction into the correction dialogue box. Looking at the correlation between typing speed and average correction time, there is a generally negative relationship, although it is not quite as strong as might be expected. (The correlation across all participants is −.33, whereas for only those with nonzero typing speed, it is stronger, at −.48.) The mild to moderate correlation suggests that there are additional reasons for typing speed's

influence. Typing speed has a significant bivariate correlation with dictation speed, so to some extent, it is hard to distinguish from the effect of dictation speed, although in multivariate analyses that included both variables, typing speed was definitely the stronger factor. The association with dictation speed supports the idea that some people simply perform faster than others, whether because they place a higher value on high speed or have a higher capacity for it. In other words, part of the reason for typing speed's significance may not be a causal link to text entry rate with ASR; it may simply reflect an association between fast performance on different tasks. Further analysis of individuals' correction behavior is necessary to get a clearer sense of the relationship between typing speed and text entry rate with ASR.

5.2. Factors Without Significant Influence

The remaining 16 independent variables tested did not have a significant multivariate influence on user performance, either on recognition accuracy or text entry rate. Although it can be difficult to draw firm conclusions from negative statistical results, our confidence in these results is higher for variables with relatively even distributions across their range and low bivariate correlations far from statistical significance (e.g., $r < .3$ and $p > .50$). Falling into this category is the three-level variable for ASR training hours. As Figure 1 shows, the data do not exhibit a monotonically positive relationship between the amount of ASR training received and users' recognition accuracy. Results for the second training indicator, training adequacy, also support the conclusion that training was not a key factor in these users' ASR performance, although it suffers a bit from a distribution that is skewed toward high ratings of adequacy.

The relative insignificance of training was an unexpected result. A more detailed look at each of the three training categories provides some insight into this result. In this 23-participant group, there were 9 participants who had less than 2 hours of training, typically only a brief introduction to the system. This low training group included the top 3 fastest ASR users (5 of the top 10) and the top 2 users for recognition accuracy (6 of the top 10). This suggests that for some people, extensive ASR training appears not to be necessary. Another 6 people received more than 10 hours of training (the high training group), and they, too, enjoyed relatively good performance overall, with 4 of the top 10 for text entry rate and 3 of the top 10 for recognition accuracy. The 8 people with training from 2 to 10 hours appear to be at greater risk for mediocre performance. Only 1 user in the top 10 for text entry rate was in this middle training group, and it contained no one in the top 10 for recognition accuracy. This hints at the possibility that if a user is someone who requires more than a short introduction to ASR, they will require at least 10 hours of training for best performance. Unfortunately, 10 hours of training does not seem to guarantee good performance, as 2 of the 6 users with recognition accuracy at 80% or below were in the high training group, with the remaining 4 split evenly between low and midtraining.

Another factor that showed little to no effect on user performance was dictation application. It has been hypothesized that dictating into the ASR system's built-in text application provides better performance than use of third-party word processors such as Microsoft Word (Gardner-Bonneau, 1999), but the results here are inconsistent with this. In this study, with 13 users dictating into Word and 10 dictating into the ASR notepad (self-selected), the means for recognition accuracy and text entry rate for the two applications were almost identical, with very low correlations (less than .10) to match. The ASR notepad group did have a lower between-subjects variation, however, suggesting that perhaps it did provide more consistent performance across participants.

Several factors, including ASR text usage, word proc time, words per utterance, and need computer for job/school, showed considerable bivariate influence on ASR performance but dropped from significance when examined in a multivariate context. Perhaps the most notable of these was need computer for job/school, which was intended to reflect the necessity of productive computer use for an individual. As shown in Figures 3 and 5, this factor had a graphically clear bivariate effect on recognition accuracy and text entry rate, and this effect was statistically significant in the case of text entry rate. However, because need computer for job/school was also highly correlated with the use of scratch that ($r = -.484$) and manual typing speed ($r = 0.355$), its inclusion to multivariate models that include one or both of these variables supplies very little additional explanatory power.

5.3. Limitations of the Study

The primary goal of this study was establishment of baseline performance measures for experienced ASR users, with a secondary goal of assessing influential factors if possible. This primary goal dictated an observational field data approach,

in which user performance was observed and measured in each user's actual setting, with his or her unique technological configuration and life circumstances. This meant that it was not possible to control the assignment of the candidate independent variables, leading to uneven distribution and less rigorous statistical power in some cases. For example, general microphone type (whether headset or desktop) had a very low statistical correlation with user performance, but that is at least partly due to the fact that only 5 of the 23 participants used desktop microphones. Similar caveats apply to the word proc time and training adequacy variables. However, we were fortunate in getting reasonably even distributions across the other independent variables and believe that the advantages in observing a user's performance on their actual system outweighed the limitation of uneven distribution across some independent variables.

A second limitation is that we were not able to assess some potentially important factors at all. In particular, possibly influential user characteristics such as speech quality, frustration tolerance, motivation, and perseverance were not examined. We may be able to operationalize some of these factors in follow-up analyses, but in the meantime, these results offer no new insights with respect to the influence of these variables.

In choosing the modeling methods for this study, a primary consideration was to obtain valid models based on careful theoretical consideration of the relationship between the independent and dependent variables. This is why we considered the possible basis for each factor's influence (Table 1), examined the bivariate relationships graphically, and avoided the use of automated modeling procedures such as stepwise regression. However, even though the models obtained here appear to be valid and reasonably robust, this does not preclude the existence of other good models, particularly for the second and third independent variables.

Finally, a third important issue is the challenge in generalizing these results to particular individuals. The techniques used here reveal factors that were influential across an entire group of users, rather than for particular individuals within the participant group. Although this suggests that these factors may well be significant to many other ASR users, it does not account for the possibly unique combination of factors that may be important to any single individual. For example, although dictation application was not significant across this participant group, a particular user may find that it makes a noticeable difference to her or his performance. Therefore, it is an oversimplification to conclude that the nonsignificant factors here will necessarily be unimportant for every ASR user. However, the fact that a few key factors stood out from the crowd suggests that those factors merit special attention, particularly the use of scratch that.

6. CONCLUSIONS

6.1. Clinical Implications

Although it is difficult to draw strong clinical implications from a study of 23 individuals, the results here suggest the following minimal guidelines:

1. Coach the proper correction strategy. Many clinicians are aware of the desirability of limiting the use of scratch that. These results reinforce that and suggest that it receive primary emphasis.
2. Do not ignore non-ASR input methods such as single-digit typing. These can be used effectively to leverage ASR performance, when used in conjunction with ASR, not solely as a backup method.
3. Get the best hardware possible and configure it appropriately for ASR use. Teach users methods of gauging system performance and monitoring resource use within the operating system.

6.2. Design Implications

Limiting users' tendency to use scratch that may require more than just good coaching by a clinician. This is supported by the fact that many of those who overused scratch that were also those who received more than 10 hours of training from a qualified clinician. Given the strong negative influence of scratch that usage on performance, perhaps the ASR system itself could take some responsibility helping users limit its use. It might not be that difficult for the ASR system to listen to when the user says "scratch that" (it does this already) and keep track of when the context seems appropriate for use of that strategy. For example, the following pattern might suggest inappropriate use of scratch that: A user says a given utterance, followed by "scratch that," and then immediately repeats an utterance whose signal looks very similar to the first utterance (because the user in fact said the exact same thing). If evidence cumulates that this user is overusing scratch that, the system might pop up a dialogue box, in a nonintrusive but noticeable way, describing the problem and making suggestions about how the user can improve.

At a minimum, the potential consequences of scratch that could be featured more prominently in ASR product literature, manuals, help systems, tip-of-the-day dialogue boxes, and other materials.

6.3. Future Work

The results of this study provide a better understanding of the influential factors in user performance with ASR systems, but much more work remains to be done. In particular, the results related to ASR training require further exploration. If these results are replicated, it may be useful to consider identifying those users who may not need extensive ASR training and using the resources saved to provide more in-depth training to those who really need it. The reasons behind the significance of computer usage and typing speed need to be sorted out more clearly as well. If we can better understand the factors that lead to enhanced ASR performance, we will be in a better position to provide ASR users with the conditions they need to achieve their very best performance.

Acknowledgments: This study was funded by U.S. Department of Education Grant #H133E980007. This study was part of the RERC on Ergonomics, conducted from November 2000 to July 2003 while the author was a senior research fellow at the University of Michigan. Thanks to all of the participants in this study for their generous contributions of time, effort, and insights. Thanks also to Ruthvick Divecha for help with transcribing the videotapes and performing statistical analyses.

REFERENCES

Cantor, A. (2001). Speech recognition: An accommodation planning perspective. *Proceedings of CSUN 2001 Conference.* Los Angeles:.

DeRosier, R. (2002). *Speech recognition software as an assistive device: A study of user satisfaction and psychosocial impact.* Unpublished master's thesis, Temple University, Philadelphia.

Devine, E. G., Gaehde, S. A., & Curtis, A. C. (2000). Comparative evaluation of three continuous speech recognition software packages in the generation of medical reports. *Journal of the American Medical Informatics Association, 7,* 462–468.

Gardner-Bonneau, D. (1999). The future of voice interactive applications. In D. Gardner-Bonneau (Ed.), *Human factors and voice interactive systems* (pp.). Norwell, MA: Kluwer Academic.

Goette, T. (1998). Factors leading to the successful use of voice recognition technology. In *Proceedings of the 3rd International ACM Conference on Assistive Technologies; 1998; Marina del Rey, CA* (pp. 189–196). New York: ACM Press.

Griffith, R. (1999). Speech recognition for injury, disability, and prevention. *Proceedings of CSUN 1999 Conference.* Los Angeles:.

Grott, R., & Schwartz, P. (2001, June). Speech recognition from alpha to zulu. Paper presented at the Instructional Course in RESNA 2001 Conference; Reno NV.

Halverson, C. A., Horn, D. B., Karat, C., & Karat, J. (1999). The beauty of errors: Patterns of error correction in desktop speech systems. In M. A. Sasse & C. Johnson (Eds.), *Proceedings of Human-Computer Interaction—INTERACT '99* (pp. 133–140). Amsterdam: IOS Press.

Jones, D., Frankish, C., & Hapeshi, K. (1992). Automatic speech recognition in practice. *Behaviour and Information Technology, 11,* 109–122.

Karat, C., Halverson, C. A., Horn, D. B., & Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the CHI '99 Conference; 1999; Boston, MA* (pp. 568–574). New York: ACM Press.

Karat, J., Horn, D. B., Halverson, C. A., & Karat, C. (2000, April). *Overcoming unusability: Developing efficient strategies in speech recognition systems.* Poster session presented at CHI 2000, ACM Conference on Human Factors in Computer Systems, The Hague, the Netherlands.

Koester, H. H. (2001). User performance with speech recognition: A literature review. *Assistive Technology, 13,* 116–130.

Koester, H. H. (2003a). *Abandonment of speech recognition by new users. 26th annual conference on Rehabilitation Engineering (RESNA); 2003; Atlanta, GA.* Washington, DC: RESNA Press.

Koester, H. H. (2003b). *Performance of experienced speech recognition users. 26th annual conference on Rehabilitation Engineering (RESNA); 2003; Atlanta, GA.* Washington, DC: RESNA Press.

Koester, H. H. (2004). Usage, performance, and satisfaction outcomes for experienced users of automatic speech recognition. *Journal of Rehabilitation Research and Development, 41,* 739–754.

Lenker, J. (1998). Naturally speaking. *Paraplegia News, 52*(6), 37–41.

Lewis, J. R. (1999). Effect of error correction strategy on speech dictation throughput. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 457–461).

McDonald, B. (2002). A teaching note on Cook's distance: A guideline. *Research Letters in the Information and Mathematical Sciences, 3,* 127–128.

Neter, J., Wasserman, W., & Kutner, M. (1990). Applied linear statistical models: Regression, analysis of variance, and experimental designs (3rd ed.). Boston: Richard D. Irwin, Inc.

Schwartz, P., & Johnson, J. (1999). The effectiveness of speech recognition technology. In *Proceedings of RESNA '99 Conference; 1999; Long Beach, CA* (pp. 77–79). Washington, DC: RESNA Press.

Suhm, B., Myers, B., & Waibel, A. (1999). Model-based and empirical evaluation of multimodel interactive error correction. In *Proceedings of the CHI '99 Conference; 1999; Boston, MA* (pp. 584–591). New York: ACM Press.