Comparison Of Menu Displays For Ordered Lists

Saul Greenberg and Ian Witten

Man-Machine Systems Laboratory
Department of Computer Science
The University of Calgary
2500 University Drive NW
Calgary, Canada T2N 1N4

Abstract — Menu selection is a popular way of accessing databases and command sets. When the database can be categorized on a semantic basis, each menu entry usually reflects a "field of knowledge". However, some information — such as the ordered sequential lists found in telephone directories — cannot be easily categorized, and must use a presentation scheme which divides entries into ranges using lexical ordering.

This paper reports a human factors investigation of six possible menu displays for alphabetically-ordered lists. It offers clear guidance as to which menu format provides the best human performance in terms of scanning speed and error rate. Novice and expert computer users are also compared, with results indicating that novices have slower scanning speeds and greater sensitivity to menu displays than experts. Differences between human performance in scanning root menus and menus buried deep in hierarchies are also studied. The results suggest that the traversal of alphabetic menu hierarchies should avoid, as much as possible, descending deep into the tree, for user efficiency deteriorates with depth.

Keywords: Menu displays, directory access, human factors experiments, user-computer interaction.

Intended audience: Designers of man-machine interfaces; all those interested in human interaction with computers and public information systems.

Introduction

Menu schemes provide a simple way for naive and casual users to interact with a system, using only numeric keypads (eg Cuff, 1980; Martin, 1973). Videotext and other textual databases systems (such as Zog; Robertson, *et al*, 1981) are normally built as hierarchical tree structures, where each menu item leads to subdivisions of "fields of knowledge" (Tompa, 1982). Several authors have undertaken investigations of different aspects of category menu representation. Latremouille and Lee (1981) ran experiments testing the effect of adding descriptors to index items in Videotext, while Kiger's (1984) experiments studied the depth/breadth trade-off in a menu hierarchy. Engel *et al* (1983) proposed a novel method for redesigning category menus into a *map* which allows random access into a videotext database.

However, some information cannot easily be categorized on a semantic basis. Many systems must display large, ordered, sequential lists such as those found in a telephone directory or dictionary. Although directories may sometimes be accessed semantically — through yellow pages — the alphabetic listing of names found in the white pages must be accessed via other means. In the absence of a natural *a priori* taxonomy, the only reasonable presentation scheme is to divide the name space into ranges using normal lexical ordering.

One study (Tombaugh and McEwen, 1982) compared taxonomic and alphabetic retrieval methods on Videotext. Their findings revealed little difference in mean search time and number of menu pages accessed between the two. However, a preference towards alphabetic displays was shown indirectly; users switched twice as often towards an alphabetic display when given a choice. But in general, little is known about the display of sequential ranges of menu items. Even the effect of menu length is unknown. Human factors experiments investigating the depth/breadth tradeoff in menu hierarchies have concluded that eight items is about the best length for taxonomic menus (eg Kiger, 1984). If the user is searching for an optimal match between his mental construct and a menu item, this can be interpreted in terms of the well-known phenomenon of "chunking" in short-term memory (Miller, 1956); for the best match cannot be determined until all items have been examined. But this reasoning does not apply to menus whose items represent ordered ranges of a list.

A separate concern from the length of the menu is the manner in which the ranges are presented. It is this which we examine here.

Alternative menu displays

Dictionaries contain very long alphabetically-ordered lists. But people do not normally search these lists sequentially when searching for a word definition. First, the book is usually opened to the approximate position of the item in the text. Further search is then facilitated through reading the keyword at the top of the page, which describe the precise location of the current page in the list. The keyword on the left page supplies a lower range delimiter for that page, for it indicates the first, and alphabetically lowest, word in the list. Similarly, an upper range delimiter on the right page indicates the last alphabetically uppermost word in that page. Telephone directories, on the other hand, supply the full range delimiter, as illustrated by "Restaurants — Roofing" appearing on a single page. Any of the above methods can be used in a computerized menu system to indicate what range of the list is covered by a menu item. The characteristics of displays using these delimiters are the subject of this study.

A *range* is defined as a segment of an ordered list bounded by a lower delimiter and an upper delimiter. When divided into ranges, numbered for menu selection, a list has the form

. . .
n. lower delimiter –> upper delimiter
n+1. lower delimiter –> upper delimiter
n+2. lower delimiter –> upper delimiter
. . .

The ranges are disjoint, and ordered so that the upper delimiter of one precedes the lower delimiter of the next.

A user's selection process is based on his ability to identify the range indicated by a menu item. One method of menu presentation simply uses the above display (see also Table 1a). However, it is evident that this contains at least two separate kinds of redundancy. In the first place, (almost) half the delimiters in the display are implicitly supplied by the neighboring ranges. In the second place, many of the individual delimiters could likely be truncated without sacrificing information relevant to the selection process.

The first kind of redundancy follows from the fact that the upper delimiter of item n provides no information relevant to the selection process that is not contained in item n+1's lower delimiter. However, it may be that this redundant information helps the user to make his selection, certainly in the case where the name sought happens to be item n's upper delimiter. One alternative for menu display is therefore to eliminate the upper-delimiter column entirely, except for the upper delimiter of the very last entry. This last entry is necessary for error recovery in the event that the desired selection is outside the current range. An alternative is to eliminate the lower-delimiter column instead, except for the first entry. This defines three types of menu displays:

- *full-range delimiter* — both upper and lower delimiters are shown for each menu entry (Table 1)
- *lower-range delimiter* — only the lower delimiter is shown on a menu entry, the upper one being given implicitly by the next menu entry (Table 2)
- *upper-range delimiter* — only the upper delimiter is shown on a menu entry, the lower one being implied by the previous menu entry (Table 3).

The second kind of redundancy is that truncation is possible without losing information because only enough characters of each delimiter need be displayed to distinguish it from the preceding or following range. The truncation operation is not entirely trivial; indeed, dictionaries and telephone directories do not attempt it. When the two range-delimiters have a common prefix, that prefix plus one further character will suffice. However, complications arise when one delimiter is itself a prefix of the other. With the occasional addition of a string-terminating character (eg "."), these problems can be solved to give an unambiguous truncation operation. Using this, a delimiter can have one of two ***truncation levels:***

- *complete delimiter* — the complete word is used as a delimiter (Tables 1a, 2a, 3a)
- *root delimiter* — the word is truncated as much as possible without sacrificing ordering information (Tables 1b, 2b, 3b).

These two factors, taken together, give a total of six different display methods. The differences are certainly significant in terms of the total number of characters displayed. The most verbose display will show both delimiters in full, for each menu item; while the tersest will show one truncated delimiter instead. In an example menu with ordinary surnames, the number of characters displayed can differ by a factor of around four. It is by no means clear which method makes it easiest for the user to scan the menu quickly and accurately.

(1a) Complete Delimiter			(1b) Root Delimiter		
1) Arbor	–>	Barney	1) A	–>	Barn
2) Barrymore	–>	Dacker	2) Barr	–>	Dac
3) Danby	–>	Estovitch	3) Dan	–>	E
4) Farquar	–>	Kalmer	4) F	–>	Kalmer.
5) Kalmerson	–>	Moreen	5) Kalmers	–>	More
6) Moriarty	–>	Praleen	6) Mori	–>	Pra
7) Proctor	–>	Sageen	7) Pro	–>	Sage
8) Sagin	–>	Ulston	8) Sagi	–>	Ul
9) Unger	–>	Zlotsky	9) Un	–>	Z

Table 1a & b: Full Range Delimiter

(2a) Complete Delimiter	(2b) Root Delimiter
1) Arbor	1) A
2) Barrymore	2) Barr
3) Danby	3) Dan
4) Farquar	4) F
5) Kalmerson	5) Kalmers
6) Moriarty	6) Mori
7) Proctor	7) Pro
8) Sagin	8) Sagi
9) Unger	9) Un
— (Zlotsky)	— (Z)

Table 2a & b: Lower Range Delimiter

(3a) Complete Delimiter	(3b) Root Delimiter
— (Arbor)	— (A)
1) Barney	1) Barn
2) Dacker	2) Dac
3) Estovitch	3) E
4) Kalmer	4) Kalmer.
5) Moreen	5) More
6) Praleen	6) Pra
7) Sageen	7) Sage
8) Ulston	8) Ul
9) Zlotsky	9) Z

Table 3a & b: Upper Range Delimiter

Indeed, the issue is further compounded when menu *span* is considered. The *span* is defined as the alphabetical segment displayed by the complete menu page. The root page usually covers a *wide span*; such as the complete alphabet (Table 4a). Progressive selections through the menu hierarchy leads to correspondingly *narrow spans*, where only small alphabetic intervals are covered (Table 4b). In wide spans, discrimination of ranges can usually be done by the first or second letter. Narrow spans have a high number of delimiters with common roots, implying that a large number of letters must be used to allow discrimination (Table 4). The greater complexity of narrow span menus suggests that user performance may deteriorate markedly upon progression through deeper layers of the tree hierarchy.

(4a) Wide Span			(4b) Narrow Span		
1) Arbor	–>	Dacker	1) Mayland	–>	Mazury
2) Danby	–>	Kalmer	2) McAdam	–>	McAlden
3) Kalmerson	–>	Praleen	3) McAlpin	–>	McArter
4) Proctor	–>	Ulston	4) McBean	–>	McBeath
5) Unger	–>	Zlotsky	5) McBeave	–>	McCabe

Table 4a & b: Wide and Narrow Menu Spans

Pilot experiments have been designed and run to attempt both to quantify any differences between the six displays in terms of scanning speed and error rate, and to test qualitative differences of user preference. Computer experience of the user, and menu spans, are also taken into account. The tested null hypotheses are

- Six menu display systems based on combinations of truncation and delimiter methods do not differ significantly from each other in terms of user scanning speed and error rate.
- Menu span has no significant effect on user scanning speed and error rate.
- User experience has no significant effect on user scanning speed and error rate.

Method

Subjects. The subjects are forty-eight paid volunteers (university students). Half of them are solicited from senior computer science courses, and are labeled as "computer experts". The remainder are obtained from a junior computer science course intended for students majoring in other areas, and are labeled "computer novices".

Subject use. The experiment is a 3-level (range delimiter) by 2-level (truncation) by 2-level (computer experience) by 2-level (menu span) mixed factorial design (Figure 1). Each subject is assigned to both levels of truncation and width, using counterbalanced ordering to overcome transfer effects. Half of the subjects are computer experts and half are novices, giving a total of 8 subjects per cell.

		Truncated		Not Truncated	
		Narrow	Wide	Narrow	Wide
Full	Novice	S1–8	S1–8	S1–8	S1–8
	Expert	S9–16	S9–16	S9–16	S9–16
Upper	Novice	S17–24	S17–24	S17–24	S17–24
	Expert	S25–32	S25–32	S25–32	S25–32
Lower	Novice	S33–40	S33–40	S33–40	S33–40
	Expert	S40–48	S40–48	S40–48	S40–48

Figure 1: Mixed Factor Anova Design

Apparatus. A Corvus Concept microcomputer connected to a VAX-11/780 is used to display all material for the experiment on a high resolution bit-mapped screen. Keystrokes are timed locally on the Corvus so that precise measurements can be made (to within 50 msec). Instructions to subjects are given first verbally and then on-line.

Design. Subjects are randomly assigned to one of the three *range delimiter* groups. Each subject of each group is exposed to two sets of menu displays employing different truncation methods. Each truncation level alternates wide and narrow span menus.

Procedure. Each set presents instructions on how to use that particular menu, followed by a practice session of twenty-five trials and a test session of eighty trials. Each trial comprises three parts.

- A name is presented to the user, who reads it and presses <return>. Timing begins.
- A menu display appears below the name and is scanned by the user to identify the range which contains the name. The user types the number of the appropriate menu item followed by <return>. Timing ends.

Although the menus are originally generated randomly, they are the same across all subjects.

Measures. The dependent measures are scanning time and error rate. The independent measures are the truncation and range-delimiter methods, menu span and experience of subject. Scanning time is the time in seconds between the appearance of the menu and the user's selection. Any selection which did not contain the indicated name is considered to be an error. Both scanning time and error rate are collected as averages after all the trials. In addition, for qualitative comparison with the quantitative results, a forced-choice decision about preferred truncation style is solicited.

Motivation. It is desirable to keep the subject's motivation level constant to avoid uncontrolled variation within a subject. A cash prize is therefore awarded to the subject with the best performance, where performance is a combination of error rate and scanning time.

Results

Scanning speed and error rate were analyzed independently through use of the analysis of variance statistical package (P4V) supplied by BMDP statistical software (Dixon *et al*, 1981). All statistically significant results are significant at or beyond the 0.05 level. A qualitative questionnaire on truncation revealed no overall subject preference for either method.

Scanning Speed. A 3×2×2×2 analysis of variance (Kirk, 1968) on scanning speed revealed a significant main effect for range delimiters, experience, and menu span, with an interaction between truncation and menu span[1].

Post-hoc comparisons among cells of the interaction matrix, using tests of simple main effects (Kirk, 1968), showed significant differences between menu span at both levels of truncation, whereas truncation was not significant at either level of menu span. Figure 2 illustrates the magnitude of the menu span/truncation interaction. The nature of this interaction – the small effect of truncation contrasted to the large effect of span – allows us, for all practical purposes, to ignore it.

[1] Range delimiters: $F(2,42) = 3.23, p < 0.05$
Experience: $F(1,42) = 5.52, p < 0.05$
Menu span: $F(1,42) = 216.04, p < 0.01$
Truncation × Menu span: $F(1,42) = 14.78, p < 0.01$

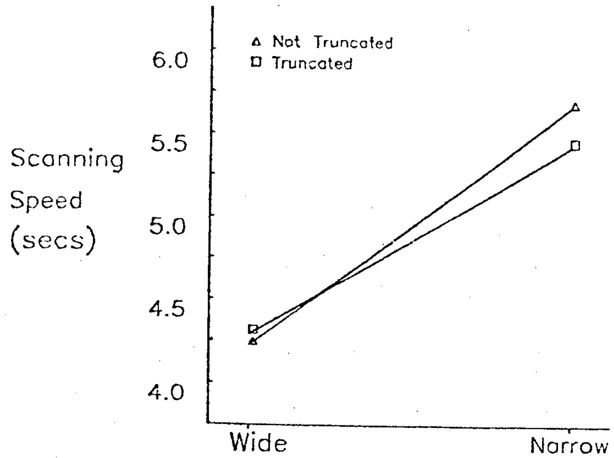

Figure 2: Truncation/Span Interaction (Scanning Speed vs. Span)

Contrast comparisons (Kirk, 1968) between the different levels of range delimiters showed significant differences between full and lower delimiters, full and upper delimiters, but none between lower and upper. These results are illustrated in Table 5, which supplies the differences among cell means of the range delimiter levels.

Individual cell means of the significant main effects are supplied in Table 6. Of note here is that all significant results on main effects have a difference between levels of at least one second.

(means)	Full	Lower	Upper
Full = 5.72	—	1.15*	1.31*
Lower = 4.57		—	0.16
Upper = 4.41			—

* $p < .05$

Table 5: Differences of Range Delimiter Means (Seconds)

Factor	Level	Mean (Secs)	Error (%)
Range Delimiter	Full	5.72	8.125
	Upper	4.41	6.055
	Lower	4.57	10.660
Experience	Novice	5.44	9.922
	Expert	4.36	6.641
Span	Wide	4.25	4.219
	Narrow	5.54	12.340
Overall		4.9	8.281

Table 6: Cell Means of Significant Main Effects and Their Levels

Error Rate. A 3x2x2x2 analysis of variance (Kirk, 1968) on error rate revealed a significant main effect for range delimiters, experience, and menu span, and interactions between range with experience, and range with span[2].

Post-hoc comparisons among cells of the range by experience interaction matrix , using tests of simple main effects (Kirk, 1968), showed significant differences for experience only at the lower range delimiters. Figure 3 graphically illustrates the experience/range delimiter interaction, in which novices have a much greater error rate than experts at the lower range delimiter menu only.

Similarly, post-hoc comparisons among cells of the range by span interaction matrix , using tests of simple main effects (Kirk, 1968), showed span significant at all levels of range delimiters, but range delimiters significant only at the narrow span level. Figure 4 graphically illustrates the range delimiter/span interaction, in which error rate is sensitive to delimiter display type at narrow span menus only, with lower range delimiters having the greatest error rate.

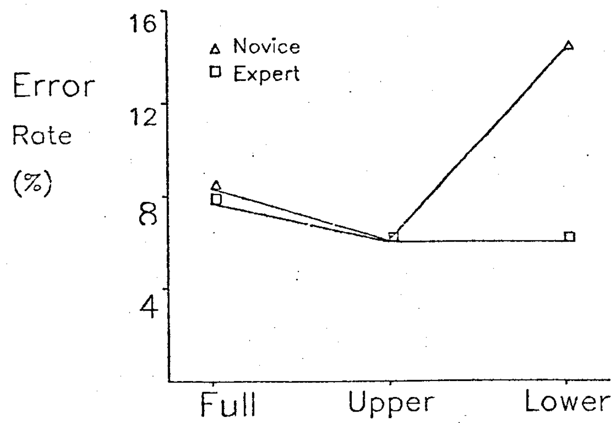

Figure 3: Range/Experience Interaction (Error Rate vs. Range Delimiters)

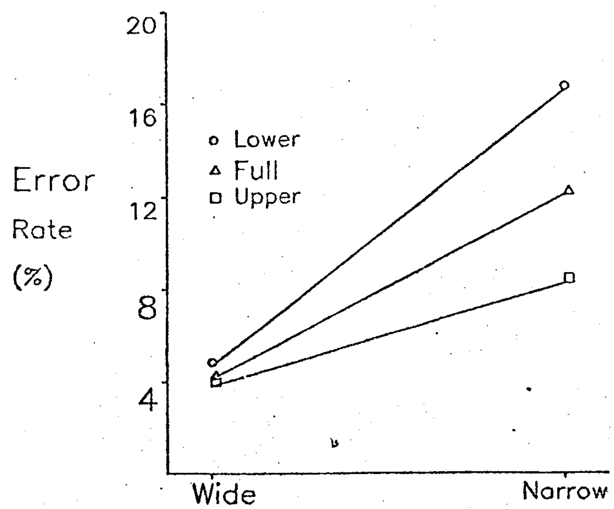

Figure 4: Range/Span Interaction (Error Rate vs. Span)

[2] Range Delimiters: $F(2,42)= 3.68$, $p< 0.05$
Experience: $F(1,42)= 5.57$, $p< 0.05$
Menu Span: $F(1,42)=77.86$, $p< 0.01$
Range x Experience: $F(2,42)= 4.69$, $p< 0.05$
Range x Span: $F(2,42)= 5.54$, $p< 0.05$

Discussion

This study examines six menu displays — created by a combination of truncation and delimiter techniques — at different levels of user experience and menu span. The null hypothesis presented earlier states that display type, user experience and menu span have no effect on any of the measured variables. The test for the null hypothesis for truncation — that truncation has no effect on scanning speed or error rate — provides no evidence to support its rejection. However, all other main effects and some two-way interactions support rejection of the null hypotheses. The results suggest the following points:

1. Full range delimiters are slower in subject scanning speed than either lower or upper delimiters (Table 6).
2. Lower range delimiters have more errors per subject than either full or upper delimiters in a narrow span menu (Figure 4).
3. Novices have more errors at the lower range delimiter than at full or upper delimiters (Figure 3).
4. Narrow span menus are slower for subjects to scan than wide span menus (Table 6).
5. Novices have slower scanning speeds than experts (Table 6).

Conclusions

The results of this experiment supply clear guidelines. The upper range delimiter is preferable to any other kind, for it is faster than the full delimiter and less error prone than the lower delimiter. This is not surprising. Full range delimiters suffer from an over-redundancy of information which inhibits search efficiency. Lower range delimiters promote a bad search strategy, because an extra backup step exists. The correct choice precedes the menu entry which is alphabetically greater than the target. Neither the backup step nor the redundancy is present in an upper range delimiter display.

Truncation techniques do not affect speed or errors. This mildly surprising fact is probably due to the user scanning only the first few letters of each word, rather than the complete chunk. In addition, the equal split between subjects over truncation preference support the conclusion that truncation has little relevance. Thus the selection of truncation method is at the discretion of the designer.

As expected, experienced users are more efficient in almost all measures than novices. But this experiment used general computer experience to differentiate novices and experts. No user had prior experience with any of the menu displays. Perhaps expert computer users are more confident with computer interactions and more adaptable to different systems. Or perhaps they are merely better typists. Whatever the cause, it would obviously be a mistake to disregard computer experience in choice of menu display.

Narrow span menus are poorer in all measures than wide span ones. Menu span simulates the difference between root pages (wide span) and pages deep down the menu hierarchy (narrow span). Therefore we conclude that human performance degrades with tree depth in an alphabetic menu hierarchy. This conclusion is indirectly supported by Greenberg (1984), who examined a University telephone directory accessed via a dynamically-structured menu hierarchy. A personalization scheme minimized the user's descent into the hierarchy by restructuring the tree to reflect the "popularity" of entries (Witten, Greenberg and Cleary, 1983). The results indicate that the degree of improvement in user performance was greater than the reduction in menus scanned, probably due to the user's less-frequent need of descending deep into the hierarchy (Greenberg, 1984).

A significant difference between means, whether it be error rate or scanning speed, is of no consequence if the magnitude of difference is small in relation to the task at hand. For example, scanning speed differences in menu types only slightly exceeds one second, a seemingly small amount. However, this represents a difference of approximately 25% in the overall selection time! The findings also indicate that proper selection of menu styles and strategies can reduce the incidence of errors by half. In real applications, an error in menu choice is associated with a high time and frustration penalty, for the user must backtrack to the previous menu. Clearly, the magnitude of difference in both scanning speed and error rate is important.

The final conclusion is simple. Upper delimiter menus are recommended with truncation at the discretion of the designer. Designers should expect novices to have slower scanning speeds and greater sensitivity to menu displays than experts. And finally, traversal of menu hierarchies should avoid, as much as possible, descending deep into the tree; otherwise user efficiency deteriorates.

Acknowledgement. This research is supported by the Natural Sciences and Engineering Research Council of Canada.

Bibliography

Cuff, R.N. (1980) "On casual users" *Int J Man-Machine Studies, 12*, 163-187.

Dixon, W.J., Brown, M.B., Engelman, L., Frane J.W., Hill, M.A., Jennrich, R.I., and Toporek, J.D. (1981) *BMDP statistical software*. University of California Press, Berkeley.

Engel F.L., Andriessen J.J., and Schmitz, H.J.R. (1983) "What, where and whence: means for improving electronic data access" *Int J Man-Machine Studies, 18*, 145-160.

Greenberg, Saul (1984) "User Modeling in Interactive Computer Systems" MSc Thesis, Department of Computer Science, University of Calgary.

Kiger, J. (1984) "The depth/breadth trade-off in the design of menu-driven user interfaces" *Int J Man-Machine Studies, 20*, In press.

Kirk, R. (1968) *Experimental design: Procedures for the behavioural sciences*. Wadsworth Publishing Company, Belmont, California.

Latremouille, S. and Lee, E. (1981) "The design of videotex tree indexes: the use of descriptors and the enhancement of single index pages" *Telidon Behavioural Research, 2*, Department of Communications, May.

Martin, T. (1980) "Information retrieval" in *Human interaction with computers*, edited by Smith and Green, pp 161-175. Academic Press, London.

Miller, G.A (1956) "The magical number seven plus or minus two: some limits on our capacity for processing information" *Psychology Revue, 56*, 81-97.