Task Performance Metrics on Liquid Crystal Displays

Dave Bockus and Jenny Guay Brock University Dept. of Computer Science St. Catharines, Ontario, Canada L2S 3A1 <u>bockusd@brocku.ca</u> Jg03zr@brocku.ca

Abstract

Twenty participants performed a selection and targeting task where performance indices were recorded. User performance varied as the sizes and resolutions of the screen changed during task performance. Performance was calculated using formula's represented in the Fitts' and Hicks' Models. Results show that the largest screen with the highest resolution had good performance times but also a smaller screen with a lower resolution yielded good performance times. The results indicate that bigger is not always better when working on LCD screens and that the user can become optically challenged when trying to work with a high resolution on a smaller screen. Results show that there is a relationship between the perceived font size on the screen and the users' ability to perform comprehension and recognition while performing certain tasks.

Key Words

Task performance Metrics, Fitts' Model, Hicks' Model, LCD Displays, Resolution vs. Size, Recognition Tasks

Introduction

Liquid Crystal Display (LCD) screens have replaced the now obsolete CRT Monitors. The introduction of this new display technology brought about clearer screens with purer colours and faster response times. People interact with LCD screens on a daily basis at home and in the workplace. Each user has different settings on the screens they use, such as different resolutions and sizes. As well they may sit various distances away from their screens. In their daily usage of computers these users will complete various navigational and selectional tasks, each affected by the view presented by their individual screen settings. Depending on the willingness of the individual user, the resolution can be varied to increase screen real estate, thus providing more space to perform tasks. This is a known phenomenon once a user obtains a larger screen. But does this increased screen size and higher resolution truly improve task efficiency? Using variations on known models such as Fitts' Law [12] and Hicks' Law (Heim 2008), a measure of task performance can be correlated to the resolution and screen size. This will provide a metric in which to set resolution with respect to screen size to obtain optimal task performance.

To test for a correlation between the display size, resolution and the task performance, experimentation using selection and navigational tasks is done. It is hoped that as resolution and screen size are varied, the user will gravitate to those size and resolution settings which maximise task performance. Thus by providing tasks which can be accurately measured with known metrics, small discriminations in task performance can be measured.

Experimentation will be controlled to ensure results are measured with some confidence, and that individual nuances and biases do not skew the results. Participants fill out a short questionnaire to determine eligibility thus screening out subjects with preconditions which can potentially corrupt the study by affecting independent variables. The tasks that will be completed are modeled loosely after Fitts' and Hicks' Law. A Fitts' model is used to measure simple motor skills when confronted with varying screen size and resolution. The task will be a simple targeting program, measuring the time it takes to select as the size and resolution change. A Hicks' model is used to help determine if screen size and resolution affect the ability to perceive targets for selectional tasks. The Hick's task measures readability and selection, as the scale is increased. After the data has been collected, task performance time is plotted against the two independent variables, screen size and resolution.

Back Ground

There have been many studies done regarding task performance on a single monitor compared to multiple; however few have been done comparing monitors of slightly different sizes. A study more applicable to the home market. Recent research has focused on task performance in virtual environments and using various reading comprehension tests. There is a correlation between small monitors and significantly larger monitors, with the larger improving task performance in general. However there are no studies that focus on using models, such as Hicks' and Fitts' Law to determine which factors affect task performance. Hicks' Law regards decision time in the context of selection, while Fitts' law is the movement time in navigational tasks [9]. When testing Hicks' Law, a measure of reaction time should be considered against the number of items that are present. The task should involve selecting a specific object from a list, with the list size increasing, after each iteration. This being the most pure form of the model. For Fitts' law, the movement time is in a ratio with the distance that was moved. An appropriate task would be dragging and dropping an icon onto a folder icon [11]. In this way, the two models have a physical measure that can be compared [7]. There are studies that relate physical fields of view to software fields of view, but the tasks that the experiments involved are very similar in that the physical field of view and software field of view were kept at a 1:1 ratio [2]. In order to decide how to relate the display size, resolution and angle, human visual perception must be considered. Since the tasks in this experiment were programmed, good design practices were followed. Each individual's perception is different regarding experience with brightness, contrast, flicker, motion and colour, these variables will be kept constant [15].

Progression toward LCD screens opened the door for new research opportunities in Human Computer Interaction, regarding usability and perception. There is recent research that relates a wider physical field of view to increased eye hand coordination and task performance [3]. The tasks involved in the research were searching and locomotion tasks, which effect the completion time because of the complexity and cognitive activity involved. The physical field of view is easy to control, but the monitor settings must be taken into account as a difference in brightness, contrast, colour, and font type could contaminate the results [4]. As well, outside factors such as visual and mental fatigue can degrade task performance [16]. Readability tests were inconclusive if the monitor size and resolution had an impact on the reading [6]. However, a study testing search speed had better results, with the speed being faster on high resolution rather then low resolution [16]. The search speed is more closely related to Fitts' law, and indicates that for a

navigation task involving searching, the bigger monitor with a higher resolution will have better results. An ideal way of representing these results is to plot the performance data against the shortest navigation distance. This becomes valid, because as the display size and resolution increases, different amounts of navigation will be required in each task [4]. When designing a navigation task, the different types of navigation must be taken into account. The lower resolution monitors require more virtual navigation such as zooming, and the high resolution can exceed human perceptual and require more physical limitations navigation [2]. The physical navigation can be limited by determining the optimum viewing distance from the user to the monitor. What must be taken into consideration is that the number of pixels on a flat screen that can be resolved by head and eye movement only occurs when the user is standing very close to the monitor and looks left and right. As the distance from the monitor increases, fewer pixels can be resolved [1]. In order to maintain a constant viewing angle, the comparison between viewing a projected display and a smaller LCD monitor can be made. The distance from the projector can be scaled down accordingly when viewing a monitor. In Figure 1, the scaling down of the visual angle is represented with respect to the distance to the screen. This type of setting will be used for the monitor settings versus task performance experiment, with small monitors only. Never the less, this ensures a constant viewing angle, eliminating contamination from this factor.

In order to enforce this constant viewing angle for each participant, one study had two posts on either side of the projector/monitor and then fishing line was strung across at eye level. When the user sat down the distance was adjusted so the fishing line was across where the eyes should be centered [10]. Other

Fig. 1-The scalable distance to be used from user to the monitor

research has suggested reading is best done when the eyes are looking downward, this can be considered the optimum viewing angle, and should be kept constant throughout the experiment [15]. However this is not consistent with how LCD screens are used. Empirical evidence shows that monitor placement is haphazard, and usually a function of the available furniture rather then good ergonomic practices.

There has been research done regarding the different fonts designed specifically for screen display versus ones designed for print. The formatting of the fonts for print are different then the screen display fonts. Screen display fonts have larger x-heights, meaning when compared with other fonts, they appear larger at the same font size, to improve readability and recognition. A suggested optimum print font size that has been tested is 9-10 point, anything above or below this has caused reading performance to degrade [13]. An optimum font size for reading off of computer screens is generated by using the screen display fonts scaled to the print size, which correlates to a normal print font. The type of font used has been found to affect legibility, and should be taken into consideration when conducting the task performance experiment. A comparison of anti-aliased fonts to bitmap fonts was researched but the results were not significant enough to conclude that either font was better for readability [5]. Another factor when choosing fonts is the case the words will be used in most often [13]. Fonts designed specifically for screen display have vast differences between letters, such as upper case

'I' that can often look similar to a lower case 'I' [5]. Having the letters as lowercase makes it easier to read and identify letters and words [13]. Since the experiment to be conducted is meant to be practical, any words used will be mixed case.

Deviating from a ratio of comparative 'fields of view' such as geometric fields of view, and display fields of view; can drastically affect what the person sees. It causes miniaturization or magnification and should be held constant so as to not skew the results. Studies have identified that there is a difference in visual navigation techniques between men and women, but using certain viewing angles and display sizes can eliminate this bias. Women navigate by landmark so a wider visual angle allows easier navigation by this technique [8]. This is only relevant in a virtual environment, and has not been proven to effect regular LCD screen navigation.

If the complexity of a document the user is trying to view is greater then what can be displayed, extra navigation is required [3]. The reduced amount of information on computer screens is a critical factor in performing tests related to the Hicks' and Fitts' model [13]. The key is to determine at what point the display size is not large enough to display an optimum amount of information. Current technology is available on larger LCD monitors that have resolutions so high that two pages of information can be displayed side by side. However since these monitors are not in the home consumer market in significant masses. the most commonly available monitors are used in this experimentation [15]. There have been many studies testing readability but there are few that test legibility, which is usually measured by identification and selection tasks. When testing readability or legibility, the colour of the font and background used must be considered. By using the most common colour scheme, dark text on a light background and requiring participants to have normal-tocorrected vision, biases and inaccurate results can be avoided. Even though studies have shown that using positive contrast, light text on dark, is easier to read because the flicker rate is less apparent, the current LCD's have flicker rates that are hardly noticeable, thus negating any previous concerns [13]. An interesting approach to finding the optimum settings for a monitor was to have participants in a study adjust these settings themselves. However this requires that participants have some computer literacy and makes the results hard to measure and record [13]. Furthermore, these settings are highly personal preferenced.

Adjusting the visual angle alone is not guaranteed to have an impact on reading performance. This goes for any other single factor, a combination of factors must be adjusted to yield any sort of difference. Although studies have shown that reading from paper is generally faster then reading from a computer screen, using high resolution displays can yield almost equal reading speeds for both mediums [13]. After the size and resolution correlations have been determined, the next step is to consider the design of the tasks. The interface for the tasks should make use of common design principles, including visibility. Visibility is obtained by using high quality displays and making information easily visible to the user. By following this guideline, it will reinforce any differences in performance that occur when the display settings have been adjusted, thus not attributing to bad design [14].

Experiment

There have been many studies concerning readability and task performance on CRT monitors, but there have been few that have used LCD monitors. This study will test the difference in task performance when paired with different screen resolutions and display sizes on LCD monitors. The objective is to show there is better task performance using a large monitor on higher resolution. This will be done using navigational and selectional tasks to simulate Fitts' and Hicks' law, which are validated metrics.

Pre-Experiment:

The first stage was selecting volunteers, and having them fill out a pre-experiment questionnaire to determine their eligibility. The questionnaire determined the participant's computer experience, with qualified ones having used a computer at least once a week. By using the computer once a week, it was surmised that the participant had basic mouse and keyboard skills, ensuring competent eyehand coordination. Information regarding the age of the participants was collected, having them choose an age range between 18 and 45. In a study done by Statistics Canada, the age of the highest computer users was less than 45 [2]. For the lower part of the age range, it is expected the lowest age is 18, since the study was conducted in a university environment. Vision was taken into account, asking participants to have normal or normal to corrected vision. Gender was also required, to ensure equal numbers of both.

The Equipment

The tasks were simple executable files written in visual studio. A Pentium 3 computer with windows XP is an acceptable minimum for running the files. The computers were equipped with a QWERTY keyboard and an optical mouse. Three different sizes of monitors, 15, 17 and 19 inches, with standard diagonal measurement were used. The chair the participants sat on was adjustable. Height was kept constant throughout the experiment once adjusted for the individual. This assured each participant was looking downward at the screen with their feet touching the ground. The chair distance emulated a safe viewing distance of approximately 25 inches in accordance with Figure 1.

The Variables

The independent variables are the screen size and screen resolution. The three resolutions tested were, 800 x600, 1024x768, and 1280x1024. The three screen sizes used, 15, 17 and 19 inch displays and were tested with the three screen resolutions and two tasks. The 15 inch monitor was only tested with two resolutions because of hardware limitations at the higher resolution. The dependent variables are the task performance time, and error rate. navigational task. For the the task performance is calculated using the index of difficulty of the task over the movement time. For the selectional task, the task performance is calculated as strictly the selection time. The number of items to choose from was kept constant. Constants in this experiment were determined by reviewing previous studies and seeing how factors skewed results. Since the most readable combination of colours for font and background is black on white, it was used throughout the experiment [1]. LCD screens have different types of colour quality; medium colour (16 bits), high colour (32 bits), and true colour. The colour quality was kept at 16 bits and the refresh rate of the monitors was set to 75HZ. Lack of noticeable flicker on LCD monitors meant that the refresh rate could be The brightness and kept constant [16]. contrast was set to 50/50 and kept constant throughout the experiment. The font type is set to Verdana, a font specifically designed for screen display, with a larger x component making it easier to read. A font size of 12 point at 96 dpi was used (default setting) because it is the default size used in computer applications. Since the experiment is being conducted in a university environment, 12 point was optimal because it is the size most

used when writing assignments and tests. The monitor height was adjusted so the participants when looking straight ahead were staring at the top edge of the monitor. Thus participant needed to look slightly down at the screen contents. (Ball 2006) had shown that optimal viewing angle requires the user's eyes looking downward at the screen.

The Tasks

The first task the participants performed loosely simulated Hicks' law, concerning decision time in the selection from a list of items. To obtain a performance index the item list was held constant. In the middle of the screen there was a word. Around this word, 7 buttons with various words on them (Fig 2). With each click of the middle word, the 7 buttons changed to simulate a selection task. The user clicked the matching word on one of the 7 buttons. Once this selection was made the 7 buttons and the middle word are reset. The user must then click on the middle word to start the process over. The time it takes from clicking the middle word to finding the corresponding selection is measured. As well, the number of words in the selection is measured, ranging from one to three. The number of words displayed in the centre and on the 7 buttons is synergetic, so that the user must read and comprehend during the selection process. This prevents eliminating choices based on the number of words. An error rate is also recorded measuring incorrect word selection.



The second task simulated Fitts' law, focussing on movement time in navigational tasks. A button is placed on the left center of the screen (Fig 3). After clicking on the leftmost button, a second button appears to the right at a random distance. Clicking on the right button finishes the iteration; the user must click the left button for the cycle to repeat. The position of the right button is a random distance to the left of the first button. The time it takes to navigate to and click the right button is recorded. The distance between the buttons is recorded in pixels.

Each task was performed under combinations of the different resolutions and sizes. Each task was performed using the 3 different monitor sizes, and three resolutions. The target size was held at 32x75 pixels. Thus as resolution and monitor size varied, so did the relative target size and distance to the target, in a proportioned setting. When analyzing the Fitts' formulas, a constant (K) which represents proportional change in target size and distance between targets as monitor size and resolution vary is represented as:

$$ID = log2(2KA/KW)$$
(1)

This formula represents the index of difficulty, and illustrates how proportional change in distance and target size become non

factors. The variable A is the amplitude and measures the distance between the two buttons during each iteration. The W is the width of the cross-section of the right button which was constant throughout the experiment. The constant K in formula 1 is eliminated allowing the ID to be unaffected by the scaling induced by screen size and resolution. Thus, the movement time MT now becomes the only independent variable. In order to determine an average index of performance (IP), as the task progresses, the average movement time, served as a measure of task performance. This index of performance is given as:



Formula 2 then gives a standard forum in which to compare the effects of changing monitor size and resolution (Mackenzie 1992).

Post Experiment

The participants filled out an exit questionnaire to determine which resolution and size they preferred and how they felt the slight differences changed their task performance, as well as any complications they ran into. The results were calculated according to the appropriate ratios for each task, and then were displayed in graphs and spreadsheets for easy comparison. The data was compared between the two tasks as well as within each task. If the difference in results between the tasks are greater then the difference within the tasks, the results between the tasks are considered significant. If the case is the opposite, the difference was not considered significant between the two tasks.

The Results

The first task focused on selection and readability, which measured selection from a list of items. The task measured an index of performance represented simply as the movement time MT in seconds (Fig 4). In the task the number of words in each iteration ranged from one to three. Comparing the times show that for iterations with only one word, the time was lower then for searching for two or three words, as can be expected. There is no noticeable pattern as the resolutions and sizes change (Fig 5). A couple values are distinct; the worst performance was



on the 15 inch, 800x600 for all three lengths. As well one of the best performers for all three was the 17 inch, 1024x768. It does appear that the gradient curve gets much rougher as the number of words increase. Hence more difficult to comprehend. This indicates that there are some resolutions and monitor size combinations which definitely

improve or hinder readability for selections of 3 words.

The time that was recorded was in milliseconds and started when the user clicked the middle area until they clicked the word they were searching for. The results showed that the best performers were the 19 inch at 1280x1024, the 19 inch at 800x600 and yet again the 17 inch at 1024x768. There was a decrease in performance at the 17 inch 1280x1024 combination, attributed to ability to optically differentiate selections. overloading the user. The 19 inch, 1280x1024 also suffered a decreased performance. This could be attributed to the 19 inch, 1280x1024 being the last combination the user completed



during the experiment. Fatigue could be attributed to the loss of performance. The poorest performer was the 15 inch at 800x600, with its small surface area and pixelly resolution, the readability was greatly decreased. As well each user started off on this iteration, suggesting a small learning curve could also distort the results.

In the second task, the targeting one, the index of performance that was calculated was a relative measure, because a recognized bit rate was not used. (Mackenzie 1992).This was due to the varying of the target position. The index of performance was based on the time it took to move between the targets and was averaged over the population of data (Fig 6). In essence



Fig-6 Task 2 IP

the IP calculated is an average over 1200 samples per subject. The index was calculated as the index of difficulty over movement time. The index of difficulty ID, which allows scaling based on the resolution and monitor size was calculated using formula (2). The target in the first task was kept at a constant size throughout and had a cross-section of 32 x 125 pixels. This allowed it to scale proportionately along with the distance.

The results show that the 17 inch monitor at 1024x768 vielded the best index of performance. Not surprisingly the 17 inch at 1280x1024 performed poorly. This occurred once again because the optically resolution and monitor size overloaded the user, and prevented them from completing the task as effectively as other resolutions. For the 19 inch monitor, as the resolution went up, the index of performance remained relatively constant. The 19 inch combined with the 1280x1024 resolution performed the best out of the 19 inch, also beating out the 17 inch 1280x1024 because of the increased physical size. This illustrates the importance physical size has on the users ability to perform. The 19 inch 1280x1024 did not overload the user because the increased surface area of the monitor was able to effectively display a clear

resolution. The poorest results were from the 15 inch monitor, which was not entirely unexpected. The 15 inch monitor at the resolution of 800x600 had the worst index of performance. The performance of the 15 inch at 800x600 was so different from the IP of the other combinations that it could be considered an outlier. A reason for this could be that there is a small learning curve involved, and since every user started at 15, 800x600, their results would be biased negatively. The learning curve was short so after the first round, the learning would have been accomplished.

Conclusions

Task 1

A font size in the range of 4.2 mm on screen gave the best results with performance degrading as the size moved away from this metric. As the font size went up, the performance slowly started to decrease, and it decreased dramatically when the font size fell below 4.2 (See Fig 7 and 8). The height of 12 point Verdana font on paper was measured to be ~3.40 mm (Fig-7) as a comparative measure. The closest measurement to this was on the 15 inch screen at a resolution of 1024x768. This indicates that while 12 point Verdana may be easy to read on paper, on the screen it causes performance degradation. Resolutions with a font size above the print size did the best. As soon as the font size approached or dipped below the print size, performance went down.

Many factors contribute to this, for instance blurriness due to high resolution on a surface which can not support it, or graininess as a result of low resolution. It is apparent that many factors contribute to ideal reading conditions. The study did show that a balance



of screen size and resolution must be obtained to facilitate prime conditions. This also reinforces the concept that people read differently on screen then they do on paper since a larger average font size around 4.2 mm tends give the best reading and comprehension performance. As the font size goes up, the error rate in Task 1 noticeably decreases. The resolutions with the lowest error rate were the 19 inch at 1024x768 and 1280x1024 which also had the highest and 3rd highest font sizes, respectively. Once again when the font size gets near 4.2 mm, the error rate increases



dramatically(Fig-9). From the error data and the Task 1 IP, 4.2 mm was the magic number that caused the results to change dramatically as it was approached. Both (Fig-8) and (Fig-9)



show that once the font size goes below 4.2 mm the index of performance decreases and the error rates increase. The number of errors for the 15 inch at 800x600 is an anomaly because it performed badly in all aspects. This task dealt primarily with fine motor skills and it appears that if the resolution gets too small, those skills are affected, as shown on the 17 inch at 1280x1024 results.

Task 2

Analysis of the Data relating to the targeting task shows that the Index of Performance stabilizes when the distance is greater then 300 pixels (Fig-10). On shorter distances, the IP does not follow a clear trend but varies widely. The reasoning for this: clicking and cognitive aspects dominate at short distances, but become less of a factor across larger distances, since movement time tends to take over as a dominant factor. The IP scales with the distance after 300 pixels between targets, however, the task only involves targeting and not comprehension. Thus, only small gains are observed across the varying resolution and monitor size combinations.



Overall, the best performers over both tasks were the 17 inch at 1024x768 and the 19 inch at 1024x768 (Fig-11). The 17 inch at 1280x1024 as well as the 15 inch at 800x600 was the poorest performers for both tasks. The user's optical abilities were overloaded with the 17"- 1280x1024 combination due to the small font size; the 15"- 800x600 had too little





surface area and a grainy resolution that facilitated a blurry appearance. The relative size of the object also has a big effect, as the objective size decreases, so does performance between both tasks. As seen in the graph (Fig-11), the indices of performance follow each other for both tasks, yielding similar results. Analysis reveals that the size and resolution the participants were most comfortable with was the 17 inch at 1024x768. This shows that a large screen at a high resolution is only best if the user wants to display a lot of information on the screen, however they should expect trade-off of lower а performance.

Future Work

The selection and targeting task could be refined in order to measure an artificial performance index. This would be used to describe and measure a global usability with respect to monitor size and resolution. The results could have a broader scale and practical applicability.

From the results there appears to be a relationship between monitor resolution and size, at certain values. These values are critical because they represent and facilitate a decrease in task performance when venturing outside of these centre lines. Further study and observations would shed light on the nature of these critical factors, identifying those inflection points which separated useful combinations from those which clearly fall into the undesirable category.

References

- 1. Ball, R., Booker, J., North, C., Shupp, L., Yost, B. Evaluations of Viewport size and curvatures of Large, High-resolution displays. *Graphics Interface* 123-130.
- 2. Ball, R., North, C. Effects of tiled high-resolution display on basic visualizations and navigation tasks. *CHI* /*Late Breaking Result: Posters.* 1196-1199.
- 3. Baudisch, P., Bellotti, V., Good, N., Schraedley, P. Keeping things in context: A comparative evaluation of focus plus context screens, overviews and zooming. *CHI* 4(1). 259-266.
- 4. Bowman, A.D., Chen, J., Ni, T. Increased Display Size and Resolution Improve Task Performance in Information-Rich Virtual Environments. *Graphics Interface*. 139 -146.
- 5. Boyarski, D., Fortizzi, J., Neuwirth, C., Regii, S. H. A study of fonts designed for screen display. *CHI*. 87-94.
- 6. Bridgeman, B., Lennon, L.M., Jackenthal, A. Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Perfomance. *Applied Measurement in Education.*, *16* (3). 191-205.
- 7. Card, S.K., Moran, T. P., Newell, A. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associate, Publishers, London, 1983.
- 8. Czerwinki, M., Tan, D. S., Robertson, G. G. Women take a wider view. *CHI* 195-202.
- 9. Ed: Gilbert, N., Monk, A. F. *Perspectives on HCI: Diverse Approaches.Handley*. Academic Press, California, 1995.
- 10. Gergle, D., Pausch, R., Scupelli, P. G., Tan, D. S. With similar visual angles, larger displays mprove spatial performance. *CHI* (5(1)). 217-224.
- 11. Heim, S. *The Resonant Interface: HCI foundations for interaction design*. Pearson, New York, 2008.
- 12. MacKenzie, I.S. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction*, 7 (1). 49.
- 13. Mills, C.B., Weldon, L. J. Reading text from computer screens. ACM, 19 (4). 329-358.
- 14. Norman, D. *The design of everyday things*. Basic Books, New York, 1988.

- 15. Plaisant, C.S., Shneiderman, B. *Designing the user interface: Strategies for effective Human-computer interaction*. Pearson, Maryland, 2005.
- 16. Ziefle, M. Effects of display resolution on visual performance. *Human Factors*, 40 (4).