Examining Mobile Phone Text Legibility while Walking

Terhi Mustonen, Maria Olkkonen, and Jukka Häkkinen
Nokia Research Center
Helsinki, Finland
{terhi.mustonen/ext-maria.olkkonen/jukka.hakkinen}@nokia.com

Abstract
In this study, alternative methods for studying legibility of text while walking with a mobile phone were examined. Normal reading and pseudo-text search were used as visual tasks in four walking conditions. Visual performance and subjective evaluation of task difficulty were used as measures of text legibility. According to the results, visual performance suffers from increasing walking speed, and the effects are greater on reading velocity for pseudo-text search. Subjects also use more homogenous strategies when reading compared to pseudo-text search, and therefore it is concluded that reading is a more useful measure of legibility. Subjective measures are found to be more sensitive to small variations in legibility than objective measures, and give additional information about task demands. Hence, without both objective and subjective measurements important information about legibility in different conditions and with different tasks will be lost.

Categories & Subject Descriptors: H.5.2. [**User Interfaces**]: Ergonomics, Evaluation/ methodology.

General Terms: Human Factors.

Keywords: Legibility, mobile phones, walking, pseudo-text.

INTRODUCTION
In addition to technological and user-related attributes, text legibility in mobile phones is influenced by environmental factors such as surround vibration. Vibration caused by vibrating environments (e.g. cars, trains, or escalators) can interfere with an individual's ability to observe quickly and accurately, and can alter both visual performance and subjective visual comfort [1]. One such condition is that of reading from a mobile display while walking, which demands that the gaze be stabilized during body motion. Every heel strike with the ground sends a shockwave through the body to the head, causing transient vibration, which, if visual acuity is to be maintained, must be countered [2]. Vision is impaired by vibration when adjacent details become blurred or confused, and fine details, i.e. high spatial frequencies, are usually the most degraded. Furthermore, the effect of vibration on legibility is dependent on the requirements placed upon the individual while performing the task [1].

CHI 2004, April 24–29, 2004, Vienna, Austria.
ACM 1-58113-703-6/04/0004.

When visual information is processed in poor quality conditions, subjects are assumed to develop a strategy of slowing down to prevent errors [3]. Thus, visual performance, e.g. speed and accuracy of reading, is assumed to reflect the ability to extract information from the display. Traditional reading tasks, in which coherent texts are used, are ecologically valid means to study the visual quality of displayed text. However, they have some disadvantageous features that limit their use. In connected text the meaning of the sentence is dominant and not every word demands equal attention [4]. Moreover, the varying difficulty of texts makes standardization difficult, which prevents the repetition of identical stimuli. Finally, since a coherent text can only be used once, generation of test material is very time consuming [3]. The pseudo-text task, in which meaningful sentences are replaced with letters and numerals that are arranged in rows like normal text, has been developed to avoid these problems and has been included as a part of the ergonomic ISO 9241-3 standard [3, 5]. Eye-movements in pseudo-text are thought to resemble eye-movements during normal reading [3, 6].

Visual performance measured as reading or search speed does not necessarily reflect visual comfort. It is known that in high quality conditions, scaled comfort may increase with improving quality while objective performance measures do not [6]. Moreover, subjective estimates have been shown to be sensitive indicators of the disruption caused by vibration. For example, a greater subjective impairment may be reported across a wider frequency range and for all vibration magnitudes than that which is measured objectively, and reading speed may be maintained even though reading difficulty is increased [7].

The focus of the present study was to compare alternative methods in studying mobile phone text legibility while walking. Reading and pseudo-text were used as visual tasks in four different walking conditions, and both subjective and objective performance measures were used.

METHODS

Test subjects
Six subjects with normal or corrected-to-normal vision participated in the test. Their ages ranged from 26 to 36 years of age (mean 29, sd 3.5). All subjects had normal visual acuity when measured from a 40 cm viewing distance with a standard eye chart, and normal near distance contrast sensitivity for five spatial frequencies (1.5-12 CPD).

Test conditions

Mobile phone text legibility was tested in four walking conditions as follows: 1) Walking in a corridor, which enabled clear headway with a subjectively natural speed *("Walk")*. Individual walking speeds varied from 3.4 km/h-4.5 km/h (mean 3.7, sd 0.4). 2) Walking on a treadmill at a natural speed that was measured during the performance in condition 1 *("Own speed")*. 3) Walking on a treadmill at the fixed speed of 1.5 km/h *("1.5 km/h")*. 4) Walking on a treadmill at the fixed speed of 3 km/h *("3 km/h")*.

In addition to walking, the task was performed in a stationary position as a control *("Control")*. The lighting of the test room was kept constant at 400 lx during the tests.

Visual stimuli and tasks

The mobile phone used in the test was a Nokia 7650 (display resolution 176*208, pixel pitch 0.198 mm). Two types of visual stimuli, real text and pseudo-text, were applied (Figure 1). The smallest character size in the application was chosen to present the visual stimuli since small details are most sensitive to the detrimental effects of vibration. Consequently, the height of uppercase "H" (9 pixels) was 2 mm, which corresponds to 0.29 degrees of visual angle from a viewing distance of 40 cm.

Real text

Natural texts *("Real text")* were taken from Finnish newspaper articles with neutral text content. A single text was comprised of complete sentences on ten text lines, and contained approximately 200 characters depending on word length. A test sequence was composed of ten consecutively presented texts that together formed a meaningful text section. In the real text task, subjects were instructed to read texts through carefully as fast as possible.

Pseudo-text

Pseudo-text stimuli *("Pseudo-text")* were random strings of uppercase and lowercase letters and spaces. Each stimulus consisted of 10 lines with 20 characters each, including embedded spaces. The number of characters was equal to that used in real text stimuli. Pseudo-texts contained 4 to 7 target characters ("H"), which were randomly placed with the restriction that the lines did not start or end with them.

In the pseudo-text task, subjects were asked to search for target characters in blocks of pseudo-text by scanning the text line-by-line from the top left to the bottom right as fast as possible. A test sequence consisted of ten consecutively presented pseudo-texts.

Objective and subjective measures

In the real text task, visual performance was measured as reading velocity (*"VR",* characters/s). In the pseudo-text task, search velocity (*"VS",* characters/s) and errors (number of missed target characters) were used as objective measures of text legibility.

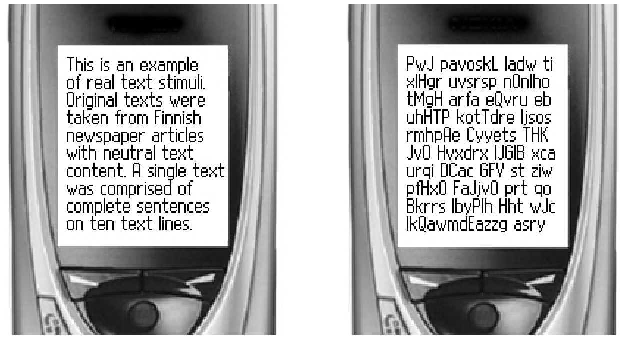

Figure 1. Real text (left) and pseudo-text (right) stimuli.

Subjective experience of the task demand was assessed with the NASA Task Load Index (TLX). This multidimensional rating procedure provides an overall workload score based on a weighted average of ratings on six subscales (Table 1) [8]. Ratings are given in a 100-point scale that ranges from low to high (good-poor for Performance).

Table 1. NASA TLX dimensions. Subjective ratings of the contributions of each dimension to task load are used as weights, with which the task load ratings for each dimension are multiplied. The average of these products is taken as the task load index.

Dimension	Description
Mental Demand	How much mental or perceptual activity is required
Physical Demand	How much physical activity is required
Temporal Demand	How much time pressure there is due to the rate of the task
Effort	How hard one has to work in order to accomplish the goals of the task
Performance	How successful one is in accomplishing the goals
Frustration level	How insecure, stressed etc. one feels during the task

Procedure

Test sequences containing ten real text or pseudo-text stimuli were presented in all test conditions twice on two separate days. The order of stimulus presentation and the texts associated with different conditions were counterbalanced among the test subjects.

Each visual task was started by pressing the control button below the display on the right side. Pressing the same key indicated the end of the reading/search. Furthermore, targets detected in the pseudo-text were recorded by pressing the joystick-button of the phone. In the real text task subjects were asked two questions about the content of the text after each sequence to confirm that texts were read through properly. Weights for the TLX dimensions were defined for the pseudo-text and the real text tasks at the beginning of the test. The task load ratings on six TLX dimensions were collected after each test sequence. Results were analyzed with a repeated measures analysis of variance.

RESULTS

Visual performance while walking

Visual performance deteriorated with increasing walking speed. The main effect of the walking condition [F=(4, 55) = 7.77, $p < .001$] as well as of the type of visual task [F(1, 58) = 16.63, $p < .01$] on speed of visual processing (characters/s) was significant. The processing speed was faster in the real text reading than in the pseudo-text search in all walking conditions (Figure 2). Moreover, the walking condition affected processing speed significantly in the real text task [F(4, 25) = 6.70, $p < .001$] but not in the pseudo-text task ($p > .05$). The error rate was affected more by increasing the walking speed than the search velocity in the pseudo-text task, but this result did not reach significance ($p < .07$).

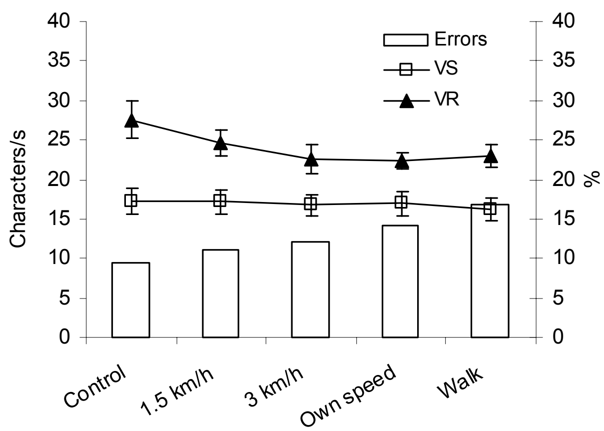

Figure 2. Reading velocity (VR) and search velocity (VS) are on the left ordinate and the error percent is on the right ordinate. The subject's own speed was always faster than 3 km/h (3.4 km/h- 4.5 km/h). Error bars are standard errors of the mean.

Walking and task load in the visual tasks

The task load increased with increasing walking speed (Figure 3), although large interindividual differences in the amount of increase were obtained. Both the walking condition [F(4, 55) = 7.34, $p < .001$] and the type of visual task [F(1, 58) = 6.18, $p < .05$] were found to influence the subjective evaluation of the task demand. There was a clear difference in the task load caused by real text reading and the pseudo-text task, as the latter was more demanding in the majority of cases. Furthermore, there was more variance in the assessment of the real text task compared to the pseudo-text task. The effect of the walking condition on task load was significant both in real text reading [F(4, 25) = 2.91, $p < .05$) and in the pseudo-text task [F(4, 25) = 10.41, $p < .001$].

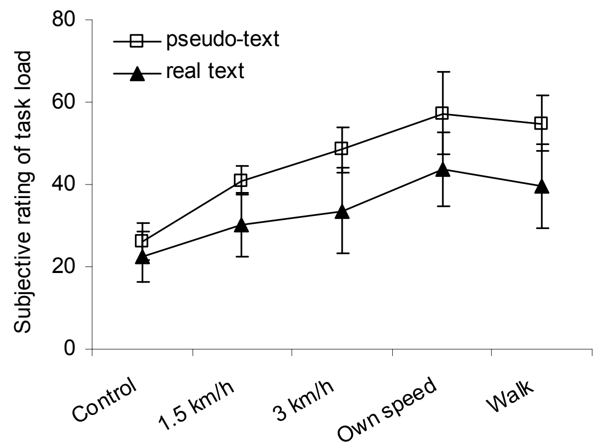

Figure 3. NASA TLX for reading task and pseudo-text task. Error bars are standard errors of the mean.

When the six task load dimensions were considered separately, the walking condition affected all dimensions ($p < .05$) except *Effort.* The effect of the visual task was significant in *Performance* [F(1, 58) = 6.93, $p < .05$] (Figure 4) and *Frustration* [F(1, 58) = 5.23, $p < .01$]. In these two dimensions, the task load caused by the pseudo-text was higher (Figure 4). Moreover, task load dimensions that were the most influenced by walking were not the same in real text reading and in the pseudo-text task (Table 2).

Table 2. The effect of the walking condition on task load assessment for six TLX dimensions in the two visual tasks. The values are F(4, 25) statistics from ANOVA with repeated measures. The level of significance is indicated by * ($p < .05$), ** ($p < .01$), *($p < .001$), and ns (non significant).**

	Real text	Pseudo-text
Mental demand	2.85*	5.10**
Physical demand	8.44***	2.22 ns
Temporal demand	3.29*	3.65*
Effort	0.79 ns	2.84*
Performance	1.28 ns	13.61**
Frustration	1.96 ns	5.17**

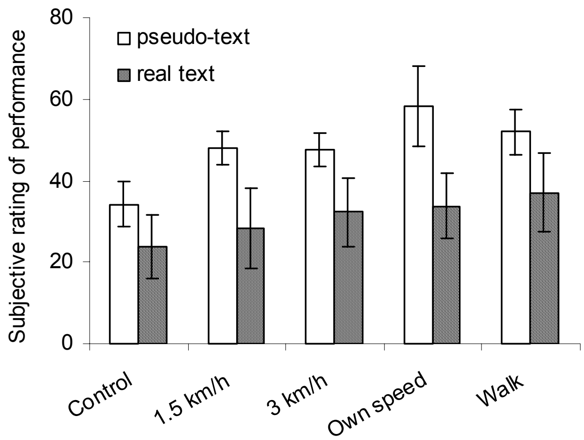

Figure 4. Evaluation of own performance in the two visual tasks as an example of the NASA TLX dimensions. Error bars are standard errors of the mean.

DISCUSSION

Methods of studying mobile phone legibility in different walking conditions were examined, and the results clearly show that no one method alone is sufficient to test legibility reliably and validly. Visual performance deteriorates with increased walking speed both in the reading of real text and in the pseudo-text search, but the effects of walking seem to be different in the two tasks. Normal reading is generally faster than pseudo-text search, and reading velocity is clearly affected by walking. The pseudo-text task, in which the error rate was a somewhat better indicator of task difficulty than search velocity, is less sensitive to changes in the walking condition. This may result partly from the small number of characters that can be presented at once on a small display, which might not allow for small variations in performance to be apparent. Interestingly, two alternative strategies for performing the pseudo-text task were found. Some subjects made more errors as the walking speed increased while keeping the search velocity constant, while other subjects kept the error rate constant and slowed down the speed of their search. According to this finding, pseudo-text does not set similar constraints to visual processing as does reading connected text, which indicates the artificial nature of pseudo-text. When testing legibility on small displays, velocity of reading is a more sensitive and useful measure of legibility than the pseudo-text performance.

Walking affects subjective evaluations in the pseudo-text task substantially and the task load in the pseudo-text search is higher than that in the reading of connected text. Together with results from objective measures, this indicates that subjects are able to compensate for task demands while performing the pseudo-text search. In particular, pseudo-text and reading differ in the level of frustration and evaluation of one's own performance. Even though the task load in these two subscales increase in both tasks, the pseudo-text is assessed as more frustrating, and one's own performance is evaluated to be poorer irrespective of the test condition. Frustration caused by the pseudo-text task as such may confuse results and interfere with legibility testing.

Despite differences in performance and task load, the general results of objective and subjective measures are parallel; as the subjective task load increases, performance declines. In subjective measures the differences between conditions are more obvious, however, and hence they clearly are more sensitive to quality variations.

Because it ultimately depends on the subject whether the increasing task difficulty most affects the pseudo-text search, the reading or the subjective evaluations, the small number of test subjects in this study limits the generalization of the results. Legibility in small displays should be studied further and with more subjects and character sizes. A more detailed study of the functioning of pseudo-text and reading in legibility research would also be pivotal. Meanwhile, the results of the present study strongly suggest that several methods need to be used in order to measure legibility in different conditions. Most importantly, the use of subjective measures in addition to visual performance should be included in legibility testing.

REFERENCES

1. Griffin, M. J. Handbook of human vibration. Academic Press, London (1990).
2. Hillman, E. J., Bloomberg, J.J., McDonald, P.V. & Cohen, H.S. Dynamic visual acuity while walking in normals and labyrinthine-deficient patients. *Journal of Vestibular Research*, 9 (1999), 49-57.
3. Roufs, J.A.J. and Boschmann, M.C. Text quality metrics for visual display units: I. Methodological aspects. *Displays*, 18 (1997), 37-43.
4. Bouma, H. Visual reading processes and the quality of text displays. In: E. Grandjean and E. Vigliani (eds.), Ergonomic aspects of visual display terminals, 101-114. Proc. Int. Workshop, Milan (1980).
5. ISO 9241-3:1992 Ergonomic requirements for office work with display terminals (VDTs), Part 3: Visual display requirements, Amendment 1: Annex C (normative): Visual performance and comfort test.
6. Boschmann, M.C. and Roufs, J.A.J. Text quality metrics for visual display units: II. An experimental survey. *Displays*, 18 (1997), 45-64.
7. Griffin M.J. & Hayward, R.A. Effects of horizontal whole-body vibration on reading. *Applied Ergonomics*, 25 (1994), 165-169.
8. NASA TLX subjective workload assessment scale: http://iac.dtic.mil/hsiac/docs/TLX-UserManual.pdf