105
81

EFFECTS OF INSTRUMENT PANEL LUMINANCE AND HUE ON SIMULATOR DRIVING PERFORMANCE AND DRIVER PREFERENCES

by

Daniel Imbeau

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial Engineering and Operations Research

APPROVED:

Walter W. Wierwille, Chairman

John G. Casali

Robert D. Dryden

Paul T. Kemmerling

Harry L. Snyder

December, 1987

Blacksburg, Virginia

EFFECTS OF INSTRUMENT PANEL LUMINANCE AND HUE ON SIMULATOR DRIVING PERFORMANCE AND DRIVER PREFERENCES

by

Daniel Imbeau

(ABSTRACT)

Twenty-four subjects, each having a valid Virginia driver's license (males and females of ages from 20 to 73), read aloud words presented on two displays while driving a simulated vehicle in night-time conditions. The words, emulating printed legends found on automobile instrument panels, were presented in different hues (eight levels), brightnesses (two levels), character sizes (four levels), and word complexities (two levels).

The brightness levels had been subjectively determined in a preliminary experiment by subjects representative of the older and younger segments of the driver population. Each of two groups of drivers determined one brightness level that was subjectively equal among the eight hues. For each word presented, six reading and driving performance measures were taken. Also, subjective attractiveness, subjective comfort, and subjective ease of readability of each hue by brightness treatment combination, were measured.

Globally, the results tend to indicate that color of illumination per se had a reliable effect on subjective preferences but a negligible effect on reading and driving performance. Brightness had an impact on performance only with the smaller character sizes. For the larger sizes, brightness level as selected by the subjects (in the preliminary experiment) had a negligible effect on performance. Character size had marked effects on both performance and subjective preferences. The two smaller character sizes tested yielded significant performance decrements for older drivers while the two larger sizes yielded best performance and were better accepted by all subjects. Word complexity did show a significant effect on glance time at the displays with all character sizes.

The results of the experiment were transformed into a set of guidelines for use in design of automobile instrument panels. A number of recommendations for future research are also included.

ACKNOWLEDGMENTS

I wish to thank Dr. Walter W. Wierwille for his guidance throughout this project. Without his patience and direction, this dissertation would never have materialized.

I also wish to thank the following personnel at Virginia Polytechnic Institute and State University:

- Dr. Harry L. Snyder, for very helpful technical suggestions in planning the experiment,
- Brian C. Hayes, for his extensive aid with data gathering,
- Dr. Jonathan F. Antin, for his aid with the initial literature search, and
- Willard W. Farley, for his aid in calibrating the color measurement equipment.

I also wish to thank Dr. John G. Casali, Dr. Robert D. Dryden, Prof. Paul T. Kemmerling, and Dr. Eric P. Smith for their valuable suggestions during this project.

I would also like to thank the following personnel in the C-P-C Human Factors Group at GM:

- Laurie Wolf and Gail Chun, for their constructive technical suggestions and for coordination with GM personnel, and
- Dr. Brian Peacock for administrative support.

Finally, I want to thank specially the Institut de Recherche en Santé et Sécurité du Travail du Québec for its financial support during the whole span of my doctoral studies.

To my wife

Renée

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The increase in new sources of information being presented to the driver of the modern automobile has stimulated much human factors research. However, one design feature of standard automobile instrument panel has yet to be definitively investigated and resolved, namely, which color should be used to illuminate the instrument panel for night driving. Various colors have been used, including white, green, blue-green, orange, and red. Since different automobiles use different colors, there is a clear need to determine if there is any significant performance difference associated with the use of one or another color.

The problem of automobile instrument panel illumination color is on the surface relatively simple. However, informed human factors researchers who understand the physiological and psychophysiological aspects of human vision recognize that the question of color of illumination for automotive instrument panel is complex. Indeed, a research team could easily spend many years of effort directed at obtaining a full answer to the question of color. Hue, saturation, luminance, glare, dark adaptation, visual impairment, age effects, character size, font, fatigue, and subjective preferences are variables of interest, each having many possible settings. As a consequence, the number of factorial settings in the parameter space is in the thousands.

The technical literature does not directly answer the main questions associated with automobile instrument panel illumination color. As might be expected, the references generally found deal with age related impairments, dark adaptation and red illumination, luminance specifications, visibility, glare, and general rules for visual presentation. While an informed reader could make intelligent guesses as to what colors and levels might be acceptable, there would be no assurance at all that an optimum selection had been made. It appears that experimentation must be used to obtain answers to the important questions. This document describes an experiment that was directed at obtaining such answers.

LITERATURE REVIEW

Design Issues

When dealing with instrument panel illumination design, the following visual functions are important to consider:

- accommodation
- light transmissivity through the eye as a function of age
- color vision
- dark adaptation
- field of vision

The human eye undergoes certain changes with normal aging. These changes generally cause predictable decrements in the visual functions. Hence, it is expected that the limits in the design of an automobile instrument panel for use by the general public will be imposed by older drivers, as Mourant and Langolf (1976) suggest.

Other design issues of importance for instrument panel illumination design are fatigue, subjective preferences, and readability of material presented in the form of legends. Issues related to symbology will not be addressed here since printed information specifically is of prime interest.

Accommodation

> Accommodation is the process responsible for maintaining a clear retinal image in the presence of varying vergence of light entering the eye. Accordingly, it allows clear vision over a range of distances and additionally can enable the hyperope to compensate his or her error of refraction to obtain improved vision. In humans, accommodation involves

> altering the ocular refracting power through change in crystalline lens curvature (Carter, 1982, p. 125).

The stiffening of the lens with age renders accommodation difficult, and explains why, in general, people gradually become farsighted with aging (i.e., presbyopia). For instance, the nearest focusing distance for the 16 year old may be 80 mm, 250 mm at age 45, and 1000 mm at age 60 (Fowkes, 1984). This is especially true at low background luminances or low illuminances (Briggs, 1986; Hughes and Neer, 1981; Vander, Sherman, and Luciano, 1977).

Acuity. Visual acuity is the ability of the eye to discriminate very small objects. It depends largely on accommodation, which determines the sharpness of the image projected on the fovea. Nearsightedness (myopia) and farsightedness (hyperopia and presbyopia) are consequences of troubles in accommodation. Such problems are usually easy to correct using corrective lenses. However, the bifocal wearer poses a problem because maximal focus is either at normal reading distance or at several feet. Adequate viewing of the legends on an instrument panel then becomes a problem because they may be out of focus; legends fall at a distance which is between the two best focus distances. Hence, hyperopia and presbyopia have important implications for instrument panel design.

Acuity depends on luminance contrast, that is, the relative luminance difference between an object to be detected (i.e., target) and its background. As adapting luminance decreases, more contrast is needed to detect the target.

The highest resolution is achieved in photopic illumination when the target image is projected onto the fovea. When the image is displaced toward the periphery of the eye, acuity decreases rapidly. At scotopic levels of illumination, acuity is highest when the target is offset about eighteen degrees from the primary line of sight (Olzak and Thomas, 1986).

Legibility. Legibility of alphanumeric characters is dependent on their size and their luminance contrast; large and/or high contrast characters are more legible and, generally, as character size decreases more contrast is needed to maintain legibility (Fowkes, 1984; Snyder and Maddox, 1980). This is true for stroke as well as for dot matrix characters. Fowkes (1984) recommends that the minimum size that a character should subtend at the eye is 20 minutes of arc (or about 5 mm viewed at a distance of 800 mm) and that the contrast ratio should be at least 8:1 in automobile instrument panel design. These recommendations are in general agreement with the Human Factors Society (1986) and the Society of Automotive Engineers (SAE, 1984) which further suggests that "... large, dimmer emitters are more legible than small, bright emitters" (SAE, 1984, p. 14). Since higher luminance contrast is associated with higher perceptibility, no maximum contrast ratio is specified. It is also suggested in this draft standard that between-character spacing should be in the order of one stroke width.

Another variable related to alphanumeric characters that is important to consider in instrument panel design is the font used. McCormick and Sanders (1982) note that white characters on a black background should have a thinner stroke than black characters on a white background, because of the irradiation effect in which white features appear to spread into black areas. Dark adaptation and highly illuminated displays tend to accentuate the effect. White letters on a black background are best suited for work at low illumination levels according to the authors and, in such conditions, the optimum stroke width-to-height ratio is between 10 to 12.5%. The SAE (1984) recommends somewhat similar ratios for instrument panel design: 10 to 15%. A lower limit of 8% is suggested by the Human Factors Society (1986). Character width-to-height ratios between 70 and 90% are generally recommended for letters and 60% for numerals (Human Factors Society, 1986; McCormick and Sanders, 1982; Sanders and McCormick, 1987).

The font of alphanumeric characters should be sans serif for good legibility; the Leroy, Lincoln/Mitre and military standard MIL-M-18012B fonts are generally recommended for their high legibility in adverse viewing conditions (McCormick and Sanders, 1982; SAE, 1984).

Mourant and Langolf (1976) performed a study on display luminance, letter-to-background luminance contrast, and letter size required on automobile instrument panels to meet a 95% recognition performance criterion for older drivers under mesopic viewing conditions. The letters were white on a darker background. They used letters with a stroke width-to-height ratio of 17% and kept the background illumination constant in all conditions so that the illuminance at the driver's eyes was 0.053 lux, which is typical of night driving according to the authors. The character sizes used were 2.3, 4.3, 6.4, and 8.4 mm viewed at 81.3 cm, that is, about 10, 18, 27, and 36 minutes of arc (arcmin) subtended at the eye, respectively. The letter-to-background contrast ratios tested were 1.25:1, 2:1, and 25:1.

The general findings were that for older subjects, no adequate luminance level could be found for the smaller character size (i.e., 10 arcmin) at any of the contrast ratios as well as for the 18 arcmin characters at the 1.25:1 contrast ratio; the subjects did not have sufficient acuity to read such small characters regardless of luminance level. Concerning the other letter sizes, luminance thresholds for legibility were found for older people, however, they were generally much higher than those found for younger subjects (Mourant and Langolf, 1976).

The authors then compared the actual letter sizes and luminances found in three automobiles with their research results. They concluded that only younger subjects could read with ease the legends in these cars. They determined that for older subjects to be able to read these legends for which the character size was equal or smaller than 4.1 mm (17 arcmin when viewed at 81.3 cm) and the letter-to-background luminance contrast ratio greater than 17:1, the luminance levels should be increased to 22.27 cd/m^2, which is much higher than the actual 0.51 to 3.43 cd/m^2 range measured in the automobiles. This theoretical level of luminance (i.e., 22.27 cd/m^2) would likely be distracting for the driver and possibly present potential for glare. The authors thus recommend increasing letter height to 6.4 mm (27 arcmin when viewed at 81.3 cm) or more to allow the majority of older people to read the legends under the actual luminance conditions found in automobiles (Mourant and Langolf, 1976).

These authors also noted that luminance contrast ratio (over a 1.25:1 to 25:1 range) had little effect on response time for letter heights greater than 6.4 mm, whereas for small letters, the higher the contrast the shorter was the response time. These results are in agreement with those of Williams (1967).

Acuity and color. Reynolds (1971) measured the luminance required to accurately read letters of various sizes transilluminated with green, yellow, and white electroluminescent light and aviation red incandescent light, in the dark (spectral distribution for the green, white, and yellow lights provided by the author). The performance criterion was 100% accuracy. He found that as the character size was decreased, more letter luminance was required, which is compatible with previous findings. Also, he found no differences in the luminance required for reading characters of various colors subtending less than 24.17 arcmin at the eye. For greater character sizes, however, (statistically speaking) significantly more luminance was required for red than for yellow or green, and also more luminance was required for white than for green. The ranking of the colors for this larger size was the following (from the highest luminance required to the lowest): red, white, yellow and green. Although significant, the luminance differences were very small; in the order of 0.0017 cd/m^2 between the extremes, red and green. These results are similar to those of Carr (1967) who did not find differences in acuity under low red illumination versus low white illumination for small letters (illumination level of 0.22 lux, no spectral nor colorimetric data provided). Note that the studies of Reynolds (1971) and Carr (1967) do not provide any data on the characteristics of the alphanumeric characters used (i.e., font, width-to-height ratio, and stroke width-to-height ratio).

Reynolds (1971) notes that his findings are in general agreement with the literature, that is, color of illumination specifically has little effect on acuity or legibility of emmetropic subjects (i.e., subjects with good corrected or uncorrected vision) at low levels of ambient illumination. Furthermore, Merrifield and Silverstein (1986) note that if luminance is sufficient, color has little effect on visual acuity, except for short wavelengths. Corresponding data specifically concerning subjects with reduced vision such as older people are not available in the literature other than the study of Mourant and Langolf (1976), which used achromatic light. Their study clearly demonstrates that data for individuals with good vision cannot be generalized easily to the whole population of drivers.

When dealing with colored illumination, one must remember that a chromatic stimulus is often perceived to be brighter than an achromatic one of the same luminance. In other words, as saturation of a chromatic stimulus increases, the ratio of its perceived brightness to that of a white stimulus of the same luminance (the so called B/L or B/Y ratio) is greater than one and increases up to values as high as nine. This phenomenon results principally from the failure of the luminance measure to describe (or correlate with) perceived brightness of somewhat large chromatic stimuli above detection threshold. This ratio remains fairly constant for yellow (Booker, 1981; Howett, 1986; Wyszecki, 1986).

Ware and Cowan (1983) have developed a polynomial correction factor and a brightness formula that permit the determination of relative brightness differences between chromatic stimuli. The correction factor for each chromatic stimulus is computed as follows:

$$Cs = 0.256 - 0.184\, y_s - 2.527\, x_s\, y_s + 4.656\, x_s^3\, y_s + 4.657\, x_s\, y_s^4$$

where x and y equal the CIE 1931 chromaticity coordinates of the stimulus.

The brightness estimate of each stimulus, is calculated as follows:

$$\log (Bs) = \log (Ls) + Cs$$

where B is an estimate of the brightness and L is the measured luminance of each stimulus. Results such as those obtained by Mourant and Langolf (1976) with achromatic characters on an achromatic background should probably be interpreted with care when dealing with chromatic characters and backgrounds.

Research has shown that both 1976 CIELUV and 1976 CIELAB spaces, which are based on surface colors and thus have a convergent geometry, are clearly nonuniform when one tries to use them to predict equivalent achromatic contrast from color contrast. Satisfactory transformations can, however, be derived (Post, Costanza, and Lippert, 1982). Furthermore, research has shown that in these convergent spaces, the relationship between ΔE distances (i.e., color difference) and luminance contrasts between colors is nonmonotonic; in nonconvergent spaces, color difference always varies monotonically with luminance difference. Hence, the convergent spaces as defined constitute inappropriate bases for prediction of visual performance measures which are monotonically related to luminance contrast (Lippert, 1986; Lippert and Snyder, 1986). Lippert and Snyder (1986) have shown, however, that rescaling the axes of the CIELUV space yields a model which performs comparably to nonconvergent spaces. For instance, expansion of the L* axis reduces the rate of convergence and makes the model "behave" more like a nonconvergent one. Use of such a rescaled model requires, on the other hand, more computations due to the transformations involved. The Human Factors Society (1986) recommends that a nonconvergent space, namely the Yu'v', be used to assess legibility of information on a CRT display (i.e., ΔE Yu'v') and that the CIELUV be used to assess discriminability of colors used for color coding (i.e., color differences: ΔE CIELUV). The latter metric should not be used for colors having a small luminance difference due to its inherent nonlinearities.

These authors suggest that for adequate legibility, colored symbols should differ from their colored backgrounds (i.e., chromatic contrast) by a minimum of 100 ΔE Yu'v' distances.

They also note that highly saturated blue should not be used for the presentation of fine detail against a dark background, because the center part of the fovea (i.e., where acuity is best) is relatively insensitive to that color. Donohoo and Snyder (1985) suggest that high purity blue backgrounds cause a disruption in accommodation and hence should not be used. Also, as reported by Reynolds (1971), red light renders accommodation difficult for presbyopic subjects and thus should most probably be avoided for instrument panel illumination.

Galer and Simmonds (1985) performed a study to determine the optimal color for automotive instrument panel illumination. They tested five broad spectrum colors (blue-green, red, green, orange, yellow) for illuminating a cluster of analog displays equipped with orange pointers (no spectral data or chromaticity coordinates were provided for any of the colors). The driving task was simulated using a fixed-base driving simulator. The authors do not specify the exact illumination conditions in the experiment; they note that the conditions were those found in normally lit streets, that is, up to 100 lux. Eighty subjects participated in the study. Their task was to drive along the computer generated road presented on a VDU and to state the speed of the vehicle (10 times per instrument panel color) and determine whether that speed was within the speed limits shown on the computer VDU (nine readings per driver per instrument panel) when prompted by the computer.

Their results showed no differences between the display colors on accuracy of reading the speed, in speed of reading as measured by response time, and on deciding whether the speed was within speed limits (Galer, 1984, 1986; Galer and Simmonds, 1985). It must be noted, however, that the use of an orange pointer optimized for blue-green lighting in the displays may have biased the experiment. Indeed, under some illumination color conditions, the chromatic contrast of the pointer and its background would be expected to be poor. Hence, it can be speculated that the results would have been different if the color of the pointer had been varied to optimize its chromatic contrast with the background for each illumination color condition tested. Consequently, the results are not applicable to the general problem of illumination and hue selection for instrument panels.

At any rate, the illumination conditions used in their experiment were more in the photopic vision range (Hood and Finkelstein, 1986) and hence this study is difficult to compare with those presented previously in which substantially lower illumination conditions were used. However, it can be said that the reading accuracy results as well as the response time results tend to disagree with the literature if reading a speedometer and reading strings of characters are comparable tasks. Snyder (1980) notes that the literature generally shows differences between the colors concerning legibility under similar ambient illumination conditions. For instance, Palmai, Schanda, and Heine (1980) had 51 subjects read flashing sequences of three digits on LED displays of three different colors (red, green, and yellow; color spectral composition provided by the authors) all having the same 40 cd/m^2 luminance, under daylight illumination conditions of 300 lux. They found that the yellow display yielded a lower rate of legibility errors and that performance with green was worse than red (Palmai et al., 1980). No information was provided on the characteristics of the alphanumeric characters used.

Lippert (1984) measured the reading speed (i.e., inverse of time required to read accurately 98% of the time) for achromatic, yellow-green, and red numerals presented in strings of variable length (i.e., 3, 4, and 5 numerals) against eight uniform background chromaticities (achromatic, red, orange, yellow, green, blue, violet, and purple) and at seven luminance modulations (background luminance varying from 24 to 46 cd/m^2), on a head-up display. The numerals subtended approximately 33 arcmin at the eye of the subjects (positive presentation). He found faster reading speeds for red numerals regardless of background hue, even without luminance contrast. The author cannot tell whether this result was due to the slightly higher luminance (i.e., 2 cd/m^2) of the red numerals, the purity of the red hue used, or the presence of a dominant wavelength near the peak sensitivity of one of the three retinal receptors (i.e., red at 610 ± 10 nm). Also, he found no important differences between performance with achromatic and yellow-green numerals. The author concludes that, at luminance contrasts less than 1.6:1, legibility is improved when either characters or background of red and purple hues are used (Lippert, 1984; Lippert and Snyder, 1986).

All of these studies are somewhat difficult to compare because each one uses a different performance criterion for comparing the different colors or color and background combinations. For instance, Reynolds (1971) determined a minimum luminance level for accurate reading (100% accuracy criterion), Palmai et al. (1980) determined a percentage of reading errors for a constant level of illumination, and Lippert (1984) used reading speed as a criterion. Also, the results obtained with high illumination levels are apparently different from those obtained at low levels. At low levels of illumination color appears to have little effect on acuity and legibility for emmetropic subjects, which is not the case at high illumination levels.

Light Transmissivity through the Eye as a Function of Age

Light transmission is adversely affected in the aging eye due to opacities in the lens, reduced clarity of the vitreous liquid, lens thickening, and irregularities of the lens which cause scatter and veiling glare. Such changes, together with degradation in the pupillary mechanism, result in more than a 50% reduction in light reaching the retina at age 50 and a 66% reduction at age 60. The reduction in transmission is greater in the 300-450 nm (blue) range because it is compounded by the selective optical density due to the yellow pigment found in the lens. The density of the lens decreases rapidly above 450 nm, and the lens transmits over 90% of the incident light for wavelengths longer than 580 nm in a normal eye. However, the pigmentation increases with age so that large differences can be found between young and old observers (Carter, 1982; Ordy, Brizzee, and Johnson, 1982).

The consequences of the reduced light transmission are glare sensitivity and poor visibility through contrast reduction at the retina. The latter is reflected by a reduced ability to notice detail and poor night vision; mesopia is approached at significantly higher illumination levels in the aging eye (Briggs, 1986; Faye, 1986; Fowkes, 1984; Hood and Finkelstein, 1986; Hughes and Neer, 1981; Olzak and Thomas, 1986).

Considering such impairments, Briggs (1986) notes that the workstation design conditions in which the functioning of older persons will improve consist mainly of those which increase the amount of light reaching the retina, those which minimize an individual's need to adapt to large changes in level of illumination prior to performing a visual task, and those that shorten the time required to make decisions on the basis of changing visual information. Faye (1986) suggests the use of large dark print and of lighting sources that avoid glare as possible management strategies to cope with these problems. Pitts (1982) suggests that if illuminance is to be increased in an attempt to aid older people, the increase should be in the longer wavelengths, since the short ones are more scattered and absorbed in the aging eye. However, increases in illumination present potential for glare, which will be discussed shortly.

The fact that in the aging eye mesopia is approached at significantly higher luminance, compounded with the reduction in the number of cones, results in a degraded color sensitivity especially under low luminance conditions (Fowkes, 1984; Pokorny and Smith, 1986).

Glare. Glare refers to the effect produced by luminance within the visual field that is sufficiently greater than the luminance to which the eyes are adapted to cause annoyance, discomfort, or loss of visual performance and visibility (Cushman and Crist, 1986). Glare can be categorized according to its effects: visual discomfort (discomfort glare) or visual performance reduction (disability glare). Glare can be specular (i.e., from a reflected source of light) or direct.

The effect of glare on vision has been described as "... an illuminance that has the effect of increasing the luminance of a visual target and its background, so that the target/background contrast in the retinal image is reduced" (Pulling, Wolf, Sturgis, Vaillancourt, and Dolliver, 1980, p. 108). Glare is amplified by light scatter in the eye which, as mentioned before, reduces image contrast on the retina. Hence, sensitivity to glare increases with age; research data show an abrupt increase in sensitivity to glare at about age 40 and resistance to glare has been found to diminish in later years with accelerating rapidity (Carter, 1982; Pulling et al., 1980).

There are various external sources of direct and specular glare in the night driving environment, such as headlamps from oncoming vehicles with or without high beams activated as well as reflections in the rear-view mirrors of headlamps from following automobiles (Finlay and Wilkinson, 1984; Olson and Sivak, 1984). Olson and Sivak (1984) found that glare as low as 1 lux from reflections in the rear view mirrors resulted in decrements in visibility distance. They report that with a glare onset of 75 lux, the dark adapted threshold was temporarily elevated 8.5 times for the younger subjects (age 25) and 11.5 times for the older subjects (65+). The times required to reach threshold stability, assuming steadiness of the glare source, were about 45 seconds for the young and 70 seconds for the old, and the final level reached was about 4.5 to 6 times the dark adapted threshold for young and old subjects, respectively (Olson and Sivak, 1984).

Mourant and Langolf (1976) report that another possible source of glare which is of more interest is the instrument panel illumination; if it is too bright, glare can result. These authors estimated that the veiling luminances from instrument panel lights set to maximum in three cars were less than one-tenth of the veiling glare disability threshold for subjects of age over 30. This means that the maximum intensity level of the instrument panel illumination found in the cars they tested could be increased tenfold before panel lighting would become a source of disability glare (Mourant and Langolf, 1976). However, illuminations well below disability glare level could be annoying and distracting to certain drivers, especially to the younger ones for whom visual capabilities are better. The authors suggest that to increase legibility of instrument panels, especially for older drivers, larger letters should be used in an effort to avoid resorting to high display luminances. This way, lower luminance levels would be required and the majority of the population could read the display accurately (Mourant and Langolf, 1976).

The literature does not address the issue of glare caused by colored light. In virtually all studies, achromatic light has been used. However, it can be inferred that high instrument panel luminances should be avoided. As far as external sources of glare are concerned (e.g., from headlamps) remedies such as polarized headlights, glare screens on roads, good road delineation, high-mast highway lighting, and driving restrictions have been proposed, while tinted windshields and glasses have been discouraged (Pulling et al., 1980). It seems, however, that little control over glare associated with external sources can be exercised through instrument panel illumination design.

Color Vision

Effects of age. Due to the progressive increase in pigment absorption of the crystalline lens with age, blue colored objects appear darkened and there is a general bias of colors away from the blue end and toward colors of longer wavelengths; it is like looking through a yellow filter. Hence, discrimination along the yellow-blue continuum is more affected than along the red-green continuum (Carter, 1982; Human Factors Society, 1986). Discrimination between hues of closely related dominant wavelengths is also decreased and this is especially true at low photopic or upper mesopic light levels. This loss in discrimination is mainly due to the reduced light transmission of the ocular media and restriction imposed by the aging pupillary mechanism on the quantity of light entering the eye.

Color blindness. A secondary category of color vision defects includes congenital deficiencies, collectively referred to as color blindness. The most common form of color blindness is partial color blindness or dichromatism (i.e., protanomaly and deuteranomaly) in which an individual can match any mixture of light with only two primaries instead of three as in normal vision; the person will confuse the red and green hues. This form of color vision defect occurs in about 2.1 percent of the population (Human Factors Society, 1986).

A less serious form of red-green defect is anomalous trichromatism in which there is a weakness rather than a failure of either the red or green component of the color system. People with such a defect will confuse red and green hues and will require more of either red or green, depending on the weakness, in a mixture of red-green for a match to a yellow. This more common color vision defect occurs in 6.4% of the population. Color vision defects affect many more men than women (Merrifield and Silverstein, 1986).

Color vision tests include the Nagel anomaloscope, pseudo-isochromatic charts (or plates), the Farnsworth-Munsell hundred hue test, the hue circle, and other objective tests. Pseudo-isochromatic charts are suitable for a general screening of protan and deutan defects. For this purpose, they are probably the most commonly used color vision test due to their low cost, simplicity, and ease of administration (Padgham and Saunders, 1975).

The age and color vision characteristics of the end users are important considerations in automobile instrument panel design. When potentially confusing color pairs are chosen (i.e., red-green and yellow-blue), luminance coding should be used where possible, since chromatic coding cannot help some individuals in discriminating the different messages (Society of Automotive Engineers, 1984). To accommodate people who have troubles distinguishing colors that differ only in the amount of red or green, The Human Factors Society (1986) recommends that colors should differ in the amount of blue as well as differ in the amount of red and green in visual displays.

Merrifield and Silverstein (1986) recommend that:

> For situations where older and/or unscreened operators are anticipated, only redundant forms of information coding should be employed and the number of displayed colors should be restricted to three or four. If color coding is used to code critical information and such individuals will be expected to use the display, the selection of a color set that can accommodate red/green color defects should be considered (Merrifield and Silverstein, 1986, pp. 58-59).

Dark Adaptation

Adaptation of the eye to different levels of luminance is controlled by two mechanisms. First, the pupil changes in size depending on the amount of light incident on the eye; its size increases as adapting luminance decreases. However, the maximum change in intensity of light entering the eye due to the pupil alone is in the order of one log unit. If we consider that the eye is sensitive to some 13 log units of intensity of light, clearly the pupil accounts for a small part of the process that controls sensitivity during dark adaptation (Hood and Finkelstein, 1986).

The second mechanism becomes dominant in low ambient illumination conditions, when the pupil can not dilate any further to admit more light into the eye. In such conditions, the cones lose much of their sensitivity and the rhodopsin which was deactivated in high ambient illumination becomes regenerated, resulting in an increase in rod sensitivity (Vander et al., 1977). Hence, the photoreceptor contribution is gradually shifted from the cones to the rods as illumination decreases. The dark adaptation process takes about 30 minutes to be complete, whereas adaptation to light takes place within seconds (one or two minutes at most) (McCormick and Sanders, 1982). Mesopic levels of luminance are considered to be comprised of those between 10^{-3} and 3 cd/m^2.

Mourant and Langolf (1976) report that pavement luminance at night may be as high as 2.05 cd/m^2 on expressways with artificial lighting or as low as 0.034 to 0.171 cd/m^2 on rural highways from headlamps alone (Fowkes, 1984). These luminances are clearly in the range of mesopic vision. Hence, dark adaptation is important in night driving because the driver's ability to detect features and targets outside of the vehicle in such low ambient illumination conditions must be preserved and restored as quickly as possible when disturbed. Instrument panel illumination should thus be designed accordingly.

Presenting a dark adapted eye with a light source reduces dark adaptation to a degree that depends on the intensity of the light source, on the size of the light source, and on the duration of exposure. Cones dark adapt much faster than rods, but the latter are much more sensitive (Hood and Finkelstein, 1986).

The effect of duration of presentation of the light source merely results in a longer dark adaptation time that is proportional to the duration of exposure. The effect of increasing the size of the adapting light produces an effect equivalent to increasing duration of exposure (Hood and Finkelstein, 1986).

Wavelength also seems to affect dark adaptation but to a very small extent. In a study of the influence of colored lighting for aircraft instrumentation on scotopic vision (instrument panel luminance level constant at 0.17 cd/m^2), Reynolds (1971) noted that aviation red incandescent light increased dark adapted thresholds 42% compared to 94% for white electroluminescent light and 127% for green electroluminescent light (spectral composition of white and green light provided by the author). However, the absolute differences in threshold were very small: less than 7.6×10^{-6} cd/m^2. Such small differences in threshold levels may explain the results of a previous study performed by Carr (1967) in which he found no differences in the influence of red incandescent versus white incandescent lighting at the same illumination levels (i.e., 0.22 lux) on the dark adaptation thresholds. The finding that red light interferes less with dark adaptation is consistent with the general belief that red light is helpful for those who must be prepared to work under conditions of low ambient illumination (Grether and Baker, 1972).

However, Reynolds (1971) concludes that the absolute advantages of red illumination are minimal and only apply if all extraneous sources of light can be controlled. He notes that "... pilot exposure to bright ground lights, bomb bursts, or flares will destroy any advantage in dark adaptation gained by the use of red instrument lighting" (Reynolds, 1971, p. 38). In night driving, which is similar to piloting at night in many aspects, uncontrolled extraneous sources such as oncoming headlights, streetlights, and other sources of light in the driver's field of view null the advantage of red illumination for dark adaptation. Hence, preserving dark adaptation should not be retained as a motivating factor for the choice of red lighting for instrument panel illumination.

Field of Vision

Merrifield and Silverstein (1986) report that "Displayed information that requires a high degree of visual resolution, such as small alphanumeric, graphic, and sensor images, must be foveally fixated to visually extract that information from the display" (Merrifield and Silverstein, 1986, p. 58). Critical information can still be displayed up to 10° to 15° in the visual periphery. In these circumstances, yellow-blue colored light should be preferred to red-green.

Fatigue

"Fatigue effects upon the driver are shown, in increasing order of importance; in inattention to individual components of the driving task, erratic visual search behaviour, and finally, short periods of unconsciousness or 'microsleep'" (Fowkes, 1984, p.219).

Mourant and Langolf (1976) note that high levels of instrument panel lighting, even if not a source of disability glare, can be annoying and distracting especially to the younger driver (i.e., discomfort glare) and can reasonably be considered as a potential source of fatigue. Thus, high display luminance should be avoided or adequate luminance adjustment should be provided. The authors suggest that the luminance levels actually found in the automobile instrument panels they tested (i.e., 0.51 to 3.43 cd/m^2) appeared adequate.

Pulling et al. (1980) note that prolonged glare can be presumed to cause muscular fatigue and attitudinal tenseness, which degrade driving skill especially for the elderly. This probably explains why older people usually do not like to drive at night. Thus, even if external sources of glare cannot be controlled, internal glare sources should be minimized as much as possible.

Concerning colored illumination, it was mentioned earlier that red lighting makes accommodation difficult for presbyopic persons and thus should probably be avoided for instrument panel lighting.

Helander (1986) notes that chromostereopsis is a depth effect caused by the different refractive power of the eye at short and long wavelengths in which highly saturated red and blue (and to a lesser extent red and green, and blue and green) on a dark background are perceived at different distances. Because it is difficult for the eye to focus on both colors at the same time, the accommodation mechanism has to drive the focus back and forth. Such conditions present potential for fatigue through discomfort. However, it is unlikely that visual fatigue of a driver could result specifically from using this combination of colors on an instrument panel since the amount of time spent looking at it is always fairly small. Rather, use of such a combination of colors could result in annoyance.

Osaka (1985) measured the visual fatigue of subjects solving arithmetic problems presented on a VDU associated with the following colors: blue, red, white, yellow, and green (P22 phosphor, chromaticity coordinates provided by the author). The subjects were dark adapted prior to the experiment. He found that yellow and green were the less fatiguing colors as measured by the critical flicker frequency (CFF) method both in the fovea and at the periphery. Red and blue were the most fatiguing colors (Osaka, 1985). However, whether CFF is an accurate measure of fatigue seems questionable. The author does not specify any luminance/brightness contrast in his experiment. Also, no information is provided on the characteristics of the alphanumeric characters used.

Subjective Preferences

The literature on subjective preferences is rather sparse, and only a few references were found. Christ (1975) noted in a review of the color coding literature that people generally prefer chromatic to achromatic displays.

Simmonds, Galer, and Baines (1981) compared five display formats (electro-mechanical, revised electro-mechanical, curvi-linear, electronic dial, and digital) each including a speedometer and a tachometer. Each format was tested in three conditions: first it was presented on slides, then in a fixed-base driving simulator, and finally in actual roadway driving. Each condition included day and night driving. Four hundred subjects participated in the experiment. Overall, all the electronic displays were liked but the digital display was the most preferred of all formats. This preference was strongest among older drivers (age 50+), reportedly because the numbers displaying speed were large and were easy to compare with speed limits. The authors also noted that the digital display was read more accurately than the other types of display; it resulted in nearly 100% accuracy. There were virtually no reading errors in any conditions of presentation, whether during the day or night (Simmonds et al., 1981). The authors do not provide any information on the characteristics of the alphanumeric characters used in the displays nor on the illumination color and levels of the displays tested.

Galer and Simmonds (1985), in a study on optimal color lighting for automotive displays, had drivers evaluate five different broad band colors along the following dimensions: ease of reading, ease of deciding whether the speed was within a speed limit, distraction while driving, attractiveness, choice for own car, and general preference. The authors note that "The colors were matched luminance, selected for consumer appeal. The luminance was matched using normal company [The Ford Motor Company Limited] procedures" (Galer and Simmonds, 1985 p. 2). The authors provide no details on these company procedures.

They found that overall, people preferred most the blue-green display, with the yellow following closely behind. The red display was the least preferred but it was considered the least distracting. It is interesting to note that people over the age of 50 found yellow easiest to read, most attractive, and least distracting, whereas people below the age of 30 found the exact same qualities for green. On the other hand, people below the age of 30 found yellow most distracting whereas people over the age of 50 found green most difficult to read. Most drivers who wore bifocal lenses generally preferred the yellow display. It is interesting to note that the red display was found soothing and restful while the orange was found reassuring by some subjects.

While these results are interesting, they must be interpreted with great care. As mentioned earlier, the experimenters used an analog display with orange pointers in their study; this pointer color has been optimized for blue-green lighting. Thus, it is clear that under some illumination colors, the chromatic contrast of the pointer and its background would be expected to be poor, which is just what has been noted judging by the subjects' comments pertaining to the ease of reading under red and orange illuminations. The authors even note that the red filters masked the pointer. Moreover, luminance of the display does not appear to have been well controlled in this experiment as subjects found, for instance, the green display glaring and lacking of contrast while the red one was too dense and also lacking of contrast. Obviously, this experiment lacked control over variables that were very likely to play an important role on the subjects' preferences, namely chromatic contrast and luminance contrast.

Clearly, such results show that in using analog displays to evaluate the best instrument panel illumination color, several precautions have to be taken. For instance, if chromatic contrast between the pointer and the background cannot be optimized for detection, luminance contrast should, so that at least one basis of comparison exists between the different colored displays. As mentioned earlier, no differences were found between the different colors in the response times, and it can be expected that if each color is optimized at least in terms of pointer-to-background contrast, then differences in these variables could likely be observed as the literature tends to suggest (Snyder, 1980). The results of this study can probably not be expected to be widely generalizable to instrument panel illumination nor representative of people's subjective evaluation obtained under well controlled conditions.

The SAE (1984) suggests that older people prefer saturated colors whereas younger people prefer desaturated colors. The sources of this finding are not given, however.

Readability

McCormick and Sanders (1982) define readability as: "A quality that makes possible the recognition of the information content of material when represented by alphanumeric characters in meaningful groupings, such as words, sentences, or continuous text" (McCormick and Sanders, 1982 p. 89). While legibility is concerned with the characteristics of individual characters, readability is concerned with the recognition of groupings of characters (words and sentences).

The Human Factors Society (1986), suggests that character height should not exceed 24 arcmin since larger characters "... may inhibit the reading process [i.e., readability] by reducing the number of character positions that may be foveally viewed per fixation" (Human Factors Society, 1986, p. 29).

Readability formulae use counts of language variables in a piece of writing (such as number of syllables per hundred words or number of words per sentence) to provide an index of probable difficulty for readers. It is a predictive tool in the sense that no reader participation is needed. General acceptance of readability formulae can be observed from the development and widespread use of over 30 such formulae, plus at least 10 variations, up to 1960 (Klare, 1963, 1974-75). One of the most often used modern readability formulae is the Flesch's "Reading Ease" yardstick (Flesch, 1948).

Klare (1963) notes that "... the most important unit in the study of readability is the word. It is a natural unit of analysis in written material; it is the most often used of all factors in readability prediction; it accounts for the greatest amount of variance in available factor analyses of readability elements" (Klare, 1963 pp. 164-165).

Thorndike and Lorge (1944) have shown by their *Teacher's Work Book of 30,000 Words*, that words occurring frequently in the English language are easier to read. The authors provide a yardstick to assess the readability level (or reading difficulty) of words based on their frequency of occurrence in the language. For instance, they note that words occurring more often than three times per million words can be read easily by 11th grade students whereas only words occurring more often that 50 times per million words can be easily read by second grade students. The complete yardstick is the following:

grade	*occurrences/million words*
1-2	100+
3	50+
4	20+
5-6	10+
7-8	6+
9	5+
10	4+
11	3+
12	2+

King-Ellison and Jenkins (1954) have further shown that reading speed is increased for words that are more frequent in the English language: they found a -0.99 correlation between word frequency and recognition time. Another variable that also plays an important role in recognition time is word length: the shorter the word, the more rapid the recognition (Klare, 1963). Klare (1963) notes that "... more frequent words are recognized faster than less frequent both because the reader apparently needs to see less of them and because they tend to be shorter" (Klare, 1963 p. 167). The author further notes that reading speed and readability are closely related and that the characteristic of words most often measured in readability studies is, directly or indirectly, that of frequency (Klare, 1963).

Summary

The literature does not provide a direct answer to the instrument panel illumination color question, but it does show that the design constraints are imposed mainly by the limited visual capacity of older drivers.

For these people, illumination in the longer wavelengths (except red), combined with high luminance contrast and large letters presented on a digital display, appears to be advantageous. However, it is not certain that such a design would suit the younger drivers' preferences.

Dark adaptation should not be considered as a key factor in the selection of an instrument panel illumination color. Glare and fatigue aspects are important to consider and seem related as far as instrument panel design is concerned. Also, the data on subjective preferences associated with instrument panel illumination color appear unreliable. Moreover, it must be noted that most of the studies were not performed in realistic driving situations; that is, displays and drivers are always in completely static positions (i.e., no vibration or slight movements allowed) so that reading conditions are optimal, which is certainly not the case in real driving situations (i.e., experiment validity considerations). When dealing with alphanumeric characters, the studies seldom provide data on the associated characteristics (i.e., font, size, etc.).

Finally, studies dealing with color have compared dominant wavelengths, but never are the effects of varying the saturation within one hue studied. It appears that at low luminances, color has little effect on acuity and legibility for normal subjects. On the other hand, at high luminances, red and yellow appear to yield the best legibility performance for these subjects. For multicolored displays, minimum color differences are prescribed for adequate discriminability.

RESEARCH OBJECTIVES AND EXPECTED RESULTS

The goal of this research was to determine which automobile instrument panel illumination colors are most suitable for night driving based on performance and subjective measures. More specifically, the impacts of illumination color, luminance level, character size, legend complexity, age, and gender on performance and preferences as well as interactions between these variables were investigated. Specific concerns of this research were:

- How is word recognition performance (e.g., response time and accuracy) in a night driving situation affected by variations in hue, brightness, character size, and word complexity?
- How is visual demand of an instrument panel legend reading task affected by variations in hue, brightness, character size, and word complexity, while driving at night?
- How is driving performance affected by variations in hue, brightness, character size, and word complexity in an intruding legend reading task?
- What are the subjective preferences of drivers concerning hue, brightness, and character sizes of legends displayed on an instrument panel?
- What are the corresponding age and gender differences?

Based on the literature review, the following results were expected:

- The performance measures (i.e., response time, visual demand, accuracy, and driving behavior) should each show differences as a function of age. It is unlikely that gender would show any major effects with the exception that older men and women might differ.

- As character size and/or brightness are decreased, concurrent decreases in performance should be observed especially for older drivers.
- The differences in performance between hues should be small when sufficient brightness is provided except for short wavelengths.
- Word complexity should have a significant effect on performance.

METHOD

Experimental Design

The design used was a mixed factors design including four within-subjects and two between-subjects factors (Figure 1).

Within-subjects factors. A 2 x 2 x 4 x 8 full factorial design including the following within-subjects factors: brightness, word complexity, character size, and color, was used.

Color and brightness. Eight different hues (light blue, blue-green, green, amber, orange, reddish orange, red, and white) were used in the experiment. Each was presented at two brightness levels (low and high) as determined in a preliminary experiment (described later). Figure 2 shows the colors on the CIE 1931 xy chromaticity diagram and Figure 3 shows the colors on the CIE 1976 u'v' diagram. The colors and the brightness levels are described in more detail in the preliminary experiment section of the results.

Characters. The four character heights used in the presentation of stimuli were 1.5, 2.5, 3.7, and 5.5 mm, respectively. When viewed at a distance of about 750 mm these heights subtend angles of 7, 11, 17, and 25 minutes of arc (arcmin), respectively, at the eye. By including the 7 arcmin character size, additional information could be obtained on performance effects under extreme reading conditions.

Eleven arcmin is about the smallest character size expected to be used in automotive instrumentation, whereas 25 arcmin is about the largest.

The font chosen closely matched the MIL-M-18012B (a capitalized type) with a width-to-height ratio of 78% and a stroke width-to-height ratio of 11%. This stroke width was chosen because the reading task involved illuminated letters on a black background with very high letter-to-background luminance contrast ratios (provided later), which

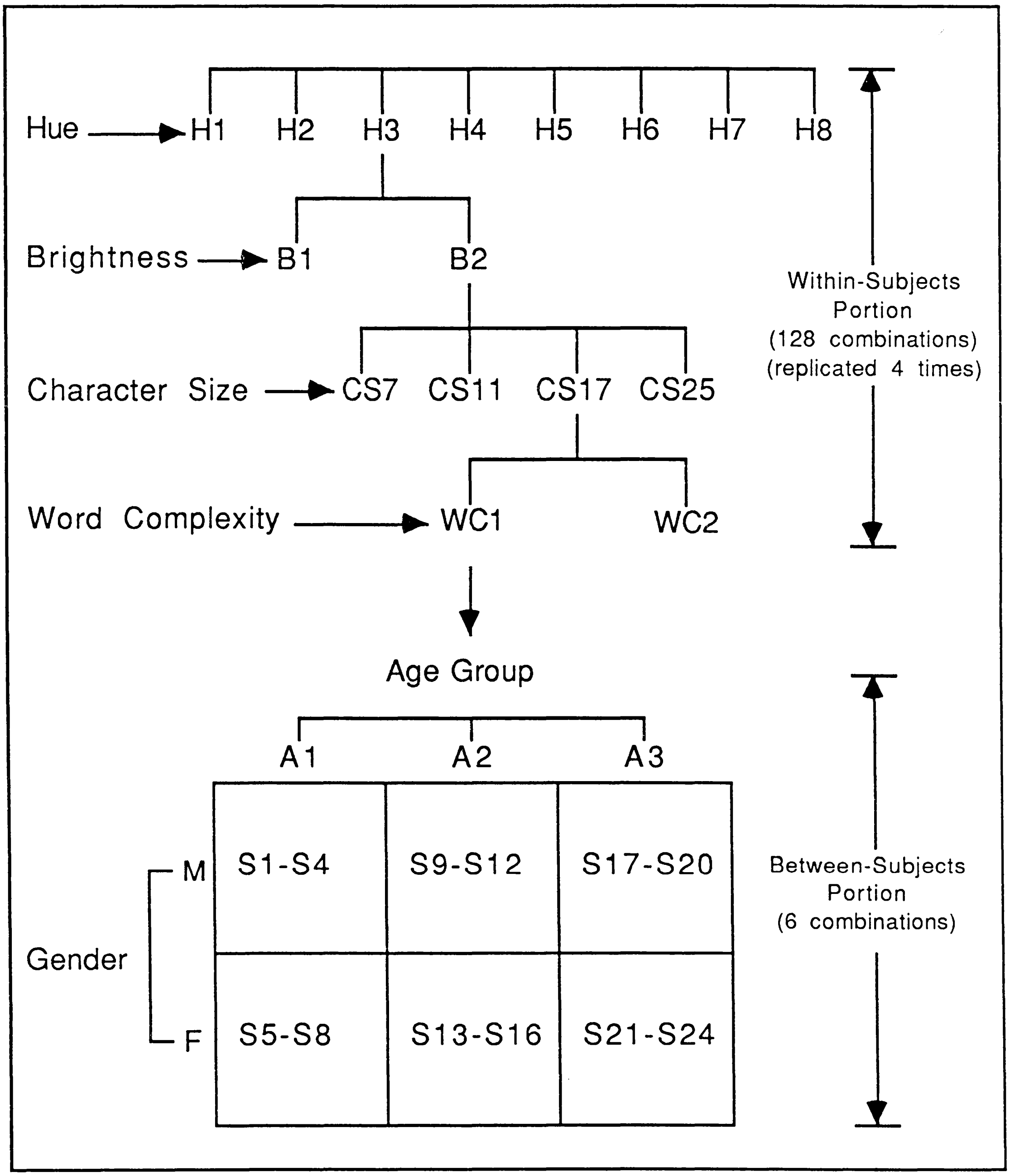


Figure 1. Experimental design.

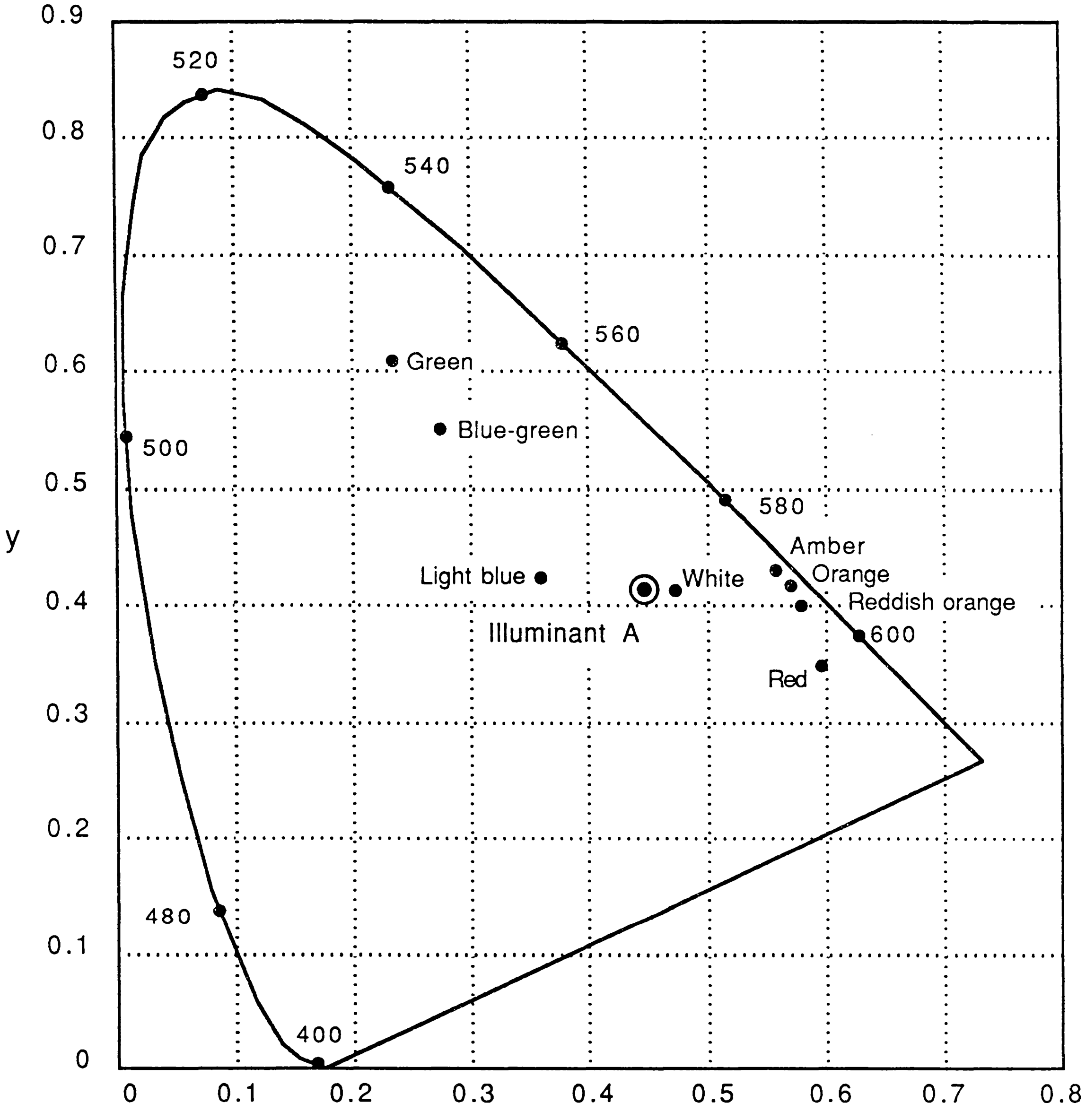


Figure 2. Colors on the CIE 1931 xy chromaticity diagram.

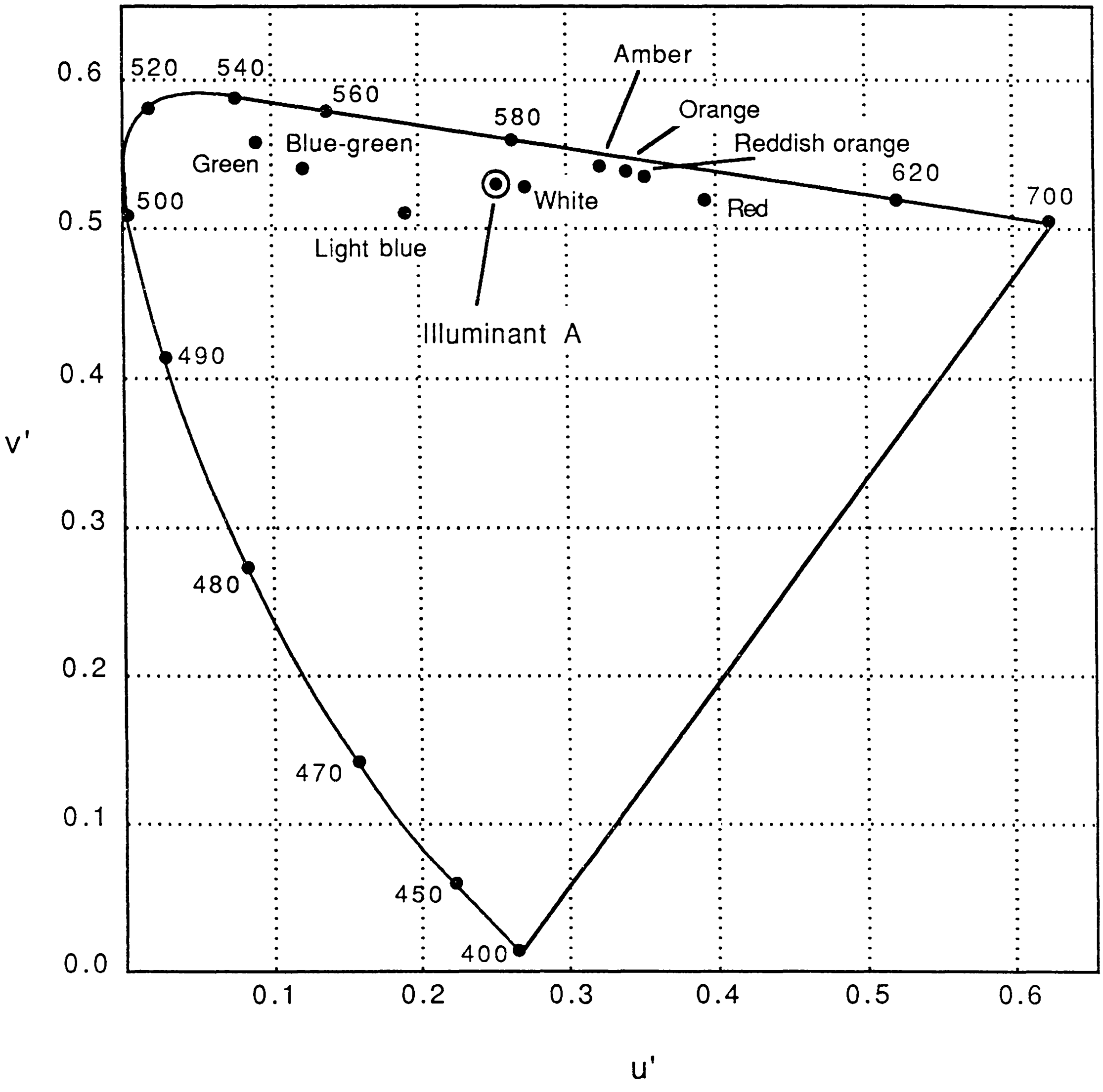


Figure 3. Colors on the CIE 1976 u'v' chromaticity diagram.

results in stroke widths that appear wider than they actually are (i.e., irradiation)(Sanders and McCormick, 1987).

Word complexity. Stimuli presented to the subjects on the displays consisted of words categorized into two levels of complexity (high and low). The level of complexity was determined by combining two commonly used word readability variables: length and frequency of occurrence in the English language.

The low word complexity level included one-syllable words occurring 20 or more times per million words read in the English language. According to Thorndike and Lorge's criteria (Thorndike and Lorge, 1944) these words should normally be read easily by children in school grades one through four. The high word complexity level consisted of two-syllable words occurring between 6 and 20 times per million words read in the English language. These words can normally be read easily by children in grades five through eight. These words have been selected using the *Word Frequency Book* by Carroll, Davies, and Richman (1971). In total, 576 words were selected; the complete list is presented in Appendix A.

Three words of randomly selected complexity were presented in the same character size at once on an individual slide. These words were arranged in a vertical column and were centered on the vertical axis of the slide. The vertical spacing of the words on a typical slide is given in Appendix B. By using three words at once, the subject was exposed to a situation similar to that found in an automobile instrument panel, that is, several labels may be grouped together. Also, exposure to a given color was more prominent using this scheme of presentation.

Note that even though the subject was presented with three words simultaneously, he or she was concerned with only one of these at a time (either the top, middle, or bottom one), as will be described later in the procedure section.

Between-subjects factors. Gender and age group were the two between-subjects factors. The age groups were as follows: young ($19 \le age \le 30$), middle ($31 \le age \le 50$), and older ($51 \le age \le 73$). Details on the subjects who participated in the study are given in the Subjects section, later.

Treatment presentation. Each subject was presented with 128 different treatment combinations (8 colors x 2 brightnesses x 4 character sizes x 2 word complexities), each replicated 4 times. Four replications were selected in an effort to keep the total time spent by a subject in the laboratory within reasonable limits. Hence, each subject experienced a total of 512 trials.

With the 576 words selected, 18 series of 32 randomly selected words were created. However, only 16 (8 colors x 2 brightness levels) out of these possible 18 series were required for each subject in the experiment, and hence the 16 were selected randomly from among the 18.

Because there were eight subjects per age group (i.e., four males and four females), color presentation order was completely counterbalanced using an 8 by 8 Latin square. The Latin square used in each group was different and orthogonal to those used in the other groups.

The two brightness levels were presented consecutively for each color to minimize adaptation between colors and also to maximize the time a subject experienced a given color. It was believed that subjective judgments (described later in the Measures section) would be more reliable as a result. The Low-High and High-Low brightness sequences were presented an equal number of times for each color, in each gender by age group. Figure 4 shows the color and brightness presentation scheme for the younger subjects.

Each 32 word series included four replications of each of eight word complexity by character size treatment combinations (2 word complexities x 4 character sizes).

Younger Drivers

Order of Presentation ↓	Subject S1 F	 S2 M	 S3 F	 S4 M	 S5 F	 S6 M	 S F	 S M
1{1 2	H1{Lo Hi	H2{Lo Hi	H3{Lo Hi	H4{Lo Hi	H5{Lo Hi	H6{Lo Hi	H7{Lo Hi	H8{Lo Hi
2{1 2	H2{Hi Lo	H3{Hi Lo	H4{Hi Lo	H5{Hi Lo	H6{Hi Lo	H7{Hi Lo	H8{Hi Lo	H1{Hi Lo
3{1 2	H8{Lo Hi	H1{Lo Hi	H2{Lo Hi	H3{Lo Hi	H4{Lo Hi	H5{Lo Hi	H6{Lo Hi	H7{Lo Hi
4{1 2	H3{Hi Lo	H4{Hi Lo	H5{Hi Lo	H6{Hi Lo	H7{Hi Lo	H8{Hi Lo	H1{Hi Lo	H2{Hi Lo
5{1 2	H7{Lo Hi	H8{Lo Hi	H1{Lo Hi	H2{Lo Hi	H3{Lo Hi	H4{Lo Hi	H5{Lo Hi	H6{Lo Hi
6{1 2	H4{Hi Lo	H5{Hi Lo	H6{Hi Lo	H7{Hi Lo	H8{Hi Lo	H1{Hi Lo	H2{Hi Lo	H3{Hi Lo
7{1 2	H6{Lo Hi	H7{Lo Hi	H8{Lo Hi	H1{Lo Hi	H2{Lo Hi	H3{Lo Hi	H4{Lo Hi	H5{Lo Hi
8{1 2	H5{Hi Lo	H6{Hi Lo	H7{Hi Lo	H8{Hi Lo	H1{Hi Lo	H2{Hi Lo	H3{Hi Lo	H4{Hi Lo

H1: Blue-Green
H2: Green
H3: Orange
H4: Light blue
H5: Red
H6: Reddish orange
H7: White
H8: Amber

Figure 4. Hue and brightness presentation scheme for the younger subjects.

Presentation order of the replicated treatment combinations was random. Each word series was presented in one color by brightness combination throughout.

Thus to summarize:

- each subject experienced a total of 512 stimuli
- these were divided and presented into 16 series of 32 words (i.e., 8 colors by 2 brightness levels)
- all of the 32 words in a series was presented in one color and one brightness
- four replications of each character size by word complexity combination were presented randomly in each series (i.e., 2 word complexities x 4 character sizes x 4 replications = 32 words in one color by brightness combination)

Subjects

A total of 40 subjects participated in the complete study, 16 in the preliminary experiment and 24 in the main experiment. All had a valid Virginia driver's license. Those who participated in the preliminary experiment were different from those who participated in the main experiment.

Only subjects in the younger and older age groups described previously participated in the preliminary experiment. Subjects from the three age groups participated in the main experiment. In both experiments, each group included 4 females and 4 males. In the main experiment, the ages ranged from 20 to 73 years, with means and standard deviations of 23.6 and 2.7 for the younger age group; 41.9 and 6.4 for the middle age group; and 58.4 and 7.2 for the older age group. The distance driven per year per subject ranged from 3200 to 56000 km, with a mean of approximately 19700 km. All subjects were paid volunteers.

Measures

In the main experiment, upon presentation of three words (stimulus), the subject's task was to read aloud only one of them. The following dependent variables were measured for each stimulus presented:

- response time from stimulus presentation to a correct answer
- response time regardless of correctness
- glance time on the screen regardless of correctness
- percentage of correct answers
- percentage of correct and incorrect answers
- lane deviation variance regardless of correctness

Response time was defined as the time elapsed from presentation of the stimulus until the moment at which the subject vocalized a response. The projector advance mechanism provided an adequate auditory cue to the subject upon presentation of a new stimulus.

Glance time was defined as the total time the subject had to look at the stimulus prior to a vocal answer.

Lane deviation was recorded on-line and the variance was computed (in m^2) for each stimulus. This measure is sensitive to lateral movements within the lane but is not affected by absolute position in the lane. That is, if the driver maintains the simulated vehicle at a constant position within the lane, the variance will be zero no matter what this position is (lane center, side, etc.).

For each of these dependent measures, 128 final values were computed from the original 512 stimuli by averaging over the four replications. This was accomplished for each subject.

The other dependent measures in the main experiment were:

- subjective attractiveness
- subjective comfort
- subjective ease of readability

The subject was asked to assess verbally on a seven-point Likert scale each of these subjective measures. The rating scales are presented in Appendix C.

While only one measure of attractiveness and only one measure of comfort were given by the subject for each color by brightness combination, four measures of ease of readability were taken, that is, one for each of the four character sizes.

Apparatus

Driving task. The driving task was simulated using the driving simulator located in Virginia Tech's Vehicle Analysis and Simulation Laboratory. This simulator is described in Appendix D. In brief, it mimics a midsize, rearwheel-drive American sedan. It has been validated by Leonard and Wierwille (1975) against a comparable instrumented automobile. The simulator provides a safe, realistic, accurate, and well controlled test environment in which to perform the experimentation.

Several dry pavement luminance measurements were performed using a standard American sedan in the Blacksburg area at night (no moon). Based on the values collected, the roadway scene display luminance was set to a value of about 4.5 cd/m^2, in the center of each of the displayed lanes.

A constant level of random lateral wind gusts was simulated to provide a light task loading to the subject. The cross-wind gusts were generated by a low frequency random noise generator followed by a low pass filter. The wind gusts had constant spectral characteristics and density functions.

Stimulus generation. An Apple Macintosh microcomputer and a font editor program were used to design and generate a font with the required stroke width and width-to-height ratio. The words were then printed on paper in columns of three, all centered with respect to a central vertical axis. These outputs were then scaled, photographed, and the negatives mounted in slide mounts. In total, 192 slides each containing three words were generated.

Displays. The slides were presented using two carousel type slide projectors (Kodak Medalist I). One projector was installed on each side of the driving simulator central instrument panel to approximate information presentation in actual automotive displays. The projectors were positioned so that the words appeared at viewing distances of about 680 mm to 770 mm for the shortest to the tallest drivers used in the study. Seat travel was constrained to limit the range on viewing distances between subjects. The words appeared at an angle of about 20° horizontal from the straight ahead line of sight for an average subject, and at about the same height as that of the speedometer center, that is, at an angle of about 15° below horizontal from the straight ahead line of sight. The words thus appeared at a composite angle of about 25° from the straight ahead line of sight for an average driver. Each carousel stored 96 slides.

The projectors were driven by a Tandy TRS-80 Model III microcomputer. This was accomplished through the use of a general purpose converter interface as well as a projector driving interface. This computer was programmed to make the slides appear at constant 8.75-s intervals. A slide could actually be seen during 8 s, since it required 0.75 s to appear when the projector's advance mechanism was activated. The computer was also programmed to alternate between the left and right projectors, which produced an equal number of stimuli on each side.

Twelve volt automotive lamps (#921) were used to replace the standard projection lamps, normally found in the projectors, to obtain a lighting that is representative of what is found in automobile instrument panels. These bulbs offered the additional advantage of generating less heat, thus preventing overheating of the color filters and slides. A constant-current regulated power supply was the source of power for the bulbs. It was used to maintain constant illumination regardless of power line fluctuations.

Color. The opaque color filters used were mounted on the front of the bulb drawer unit, just behind the slide slot inside the projectors. Figure 5 depicts the placement of the color filters within the right projector. Each change of color filters involved withdrawal of the drawer unit. The heat absorbing lens was removed from the projector since it was no longer required and also because it was slightly tinted, thus altering the spectral characteristics of the color filters. A light shield had to be added to the unit to minimize light scatter within the projector.

To reduce luminance, neutral density filters could be added behind the color filters as well as directly in front of the bulb. Luminance could also be varied by changing the current in the bulb filament through adjustment at the power supply. However, such a procedure changed the spectral characteristics of the light emitted by the bulb. Hence, transparent (gel-type) color filters were stacked behind the original color filters and neutral density filters (i.e., on the front of the drawer unit) to minimize the shift in light chromaticity due to a current change; light yellow filters were used for high currents and light blue for low currents. The specifications of all color filters are described in the Color Matching section of the results.

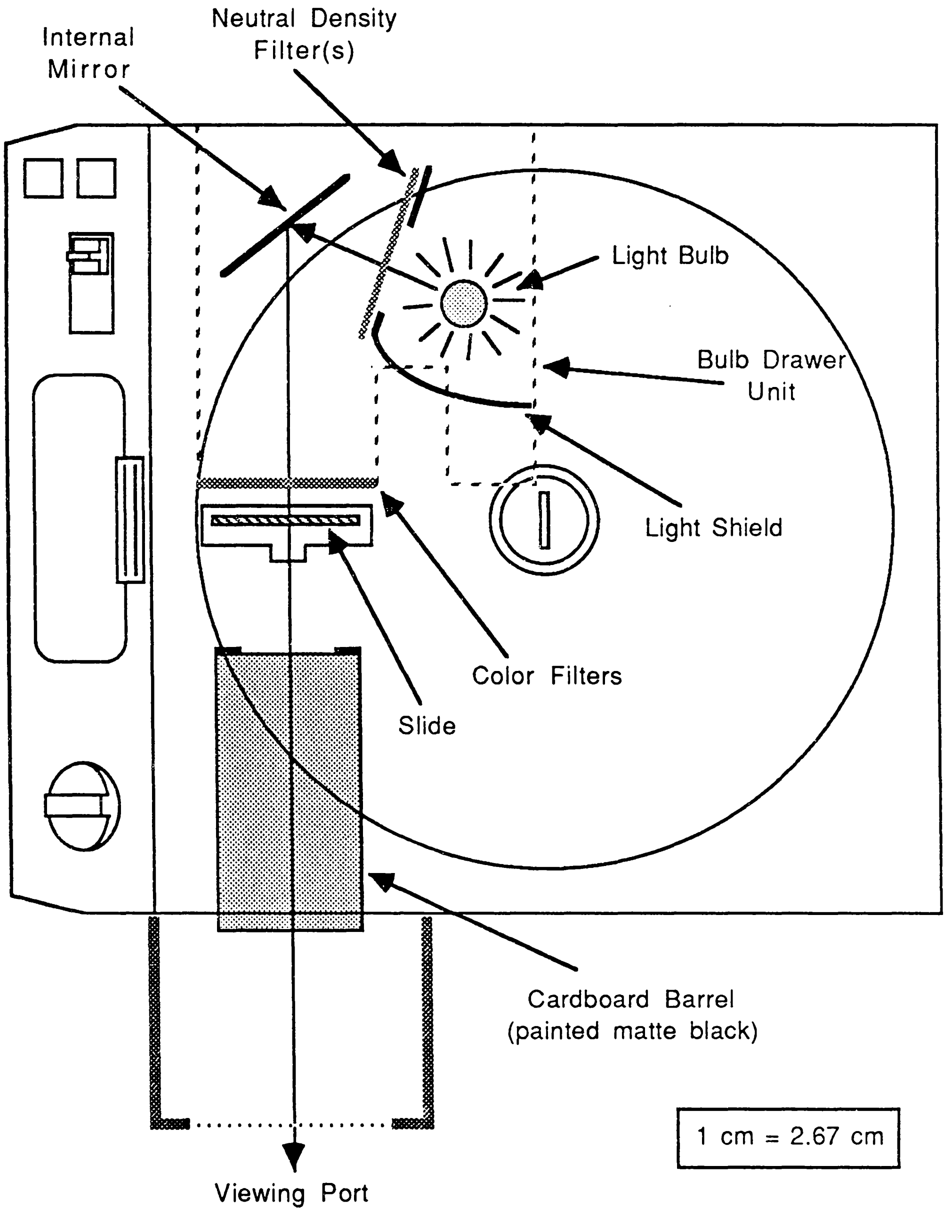


Figure 5. Top view of the right hand projector.

Due to the opacity and diffusion quality, of the color filters provided, the previewing screens found on the projectors could not be used. Rather, the projectors were modified and positioned so that the subject looked directly at the slide inside the projector. This was accomplished by removing the focusing lens from the projection opening. Since the slide is normally 13.3 cm deep inside the projector, a cardboard barrel painted with matte black paint was inserted in that opening to absorb and stop any light diffused within the projector. Hence the subject could read a slide by looking directly at it through the projection opening of the projector. The diameter of the cardboard barrel was large enough to permit head movements without loosing sight of parts of the slide. The equivalent distance from the bulb to the image seen by the subject was 17.8 cm in both projectors.

Owing to the size of the projector and location of the projection opening on the projector (i.e., on the left hand side when facing the projector), the slide on the left hand side had to be viewed by reflection in a mirror to achieve the 25° composite angle between the image and the straight ahead line of sight. Hence, the subject was looking at a virtual image on the left hand side projector. However, the mirror frame was hidden behind a cardboard panel so that it was impossible for the subject to know whether the image seen was real or virtual when the lights in the room were turned off. Placement of the reflection mirror is shown in Figure 6.

The slide mounts were made of white plastic. They were barely visible in this otherwise dark tunnel. This offered the advantage of providing a good reference for the eye of the subject to focus and thus eliminated the undesired effect of words appearing to "float in the air."

The simulator central instrument panel illumination color and brightness were subjectively matched to the displays by using transparent (gel type) color filters and by

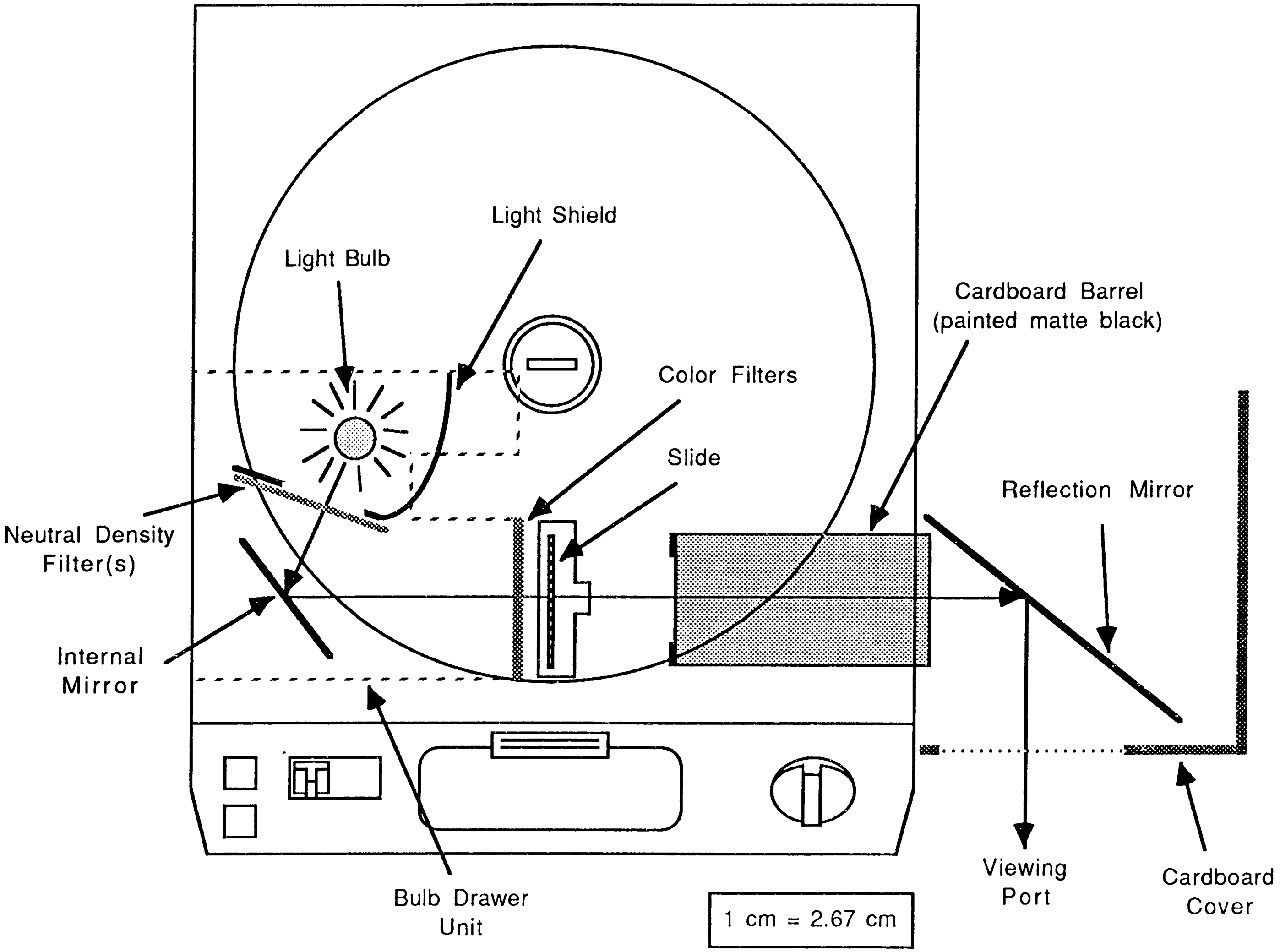


Figure 6. Top view of the left hand projector.

varying the instrument panel bulb current so as to present the subject with one uniform lighting across the displays and the central instrument panel of the simulator.

Chromaticity and luminance measurement. A Minolta CS-100 chroma meter equipped with a close-up lens (#122) and a data processor DP-101 were used to measure the chromaticities of the colors generated on the displays. The measuring distance from the meter reference point to the color target was between 323 and 368 mm when using close up lens #122 as specified by the manufacturer.

This equipment was also used to measure the luminance and contrast ratio of the stimuli presented, as well as the simulated roadway scene luminance (see preliminary experiment section).

Measuring apparatus. Lane deviation was measured and recorded by a second microcomputer TRS-80 Model III interfaced with the simulator. This computer was also used simultaneously to generate the displayed roadway curvature.

Response time was measured using a voice operated logic signal (vox) interfaced with the TRS-80 Model III used to drive the projectors. The vox included a microphone and an interface. The computer was programmed to start calculating the response time 0.75 s after the command to advance the slide (i.e., time for the slide to be visible to the subject). The response time interval ended with the first vox logic change. Although calibrated to filter the noise generated by the simulator, any first sound uttered by the subject after the presentation of the slide, such as a cough, would trigger the vox. Thus, careful checking of the computed response times was necessary.

A low light level video camera was mounted on the side of the simulator and was directed toward the subject's face. One of the experimenters observed a video monitor (connected to the camera) during the experiment and depressed a hand-held pushbutton whenever the subject glanced at either projector. The experimenter held the button in the depressed position for the duration of the glance and released it when the subject's glance returned to driving. The pushbutton signal was sent to the TRS-80 Model III microcomputer, which then computed glance time to the projector. A videocassette tape recorder was used to record the video camera signal for later review and correction of glance time where necessary. Vocal responses of the driver were recorded for backup purposes on the audio track using a second microphone.

Procedure

Preliminary experiment. The goal of the preliminary experiment was to determine the two brightness levels for each of the eight colors that were to be used in the main experiment.

This was accomplished in the following way: for each subject in each group (younger and older), one subjectively comfortable brightness based on the subject's preferences was established for each of the eight colors using a 15 minute of arc character size, a size that is in the middle of all sizes used in the main experiment. Based on the literature review, the assumption was made that the older drivers would select higher luminances than the younger drivers for corresponding colors. Hence, the luminance of the median current value in the bulb obtained in the younger age group for a given color would constitute the low brightness condition to be used in the main experiment for that color, and the luminance of the median current value in the bulb found in the older age group for that same color would constitute the high brightness condition in the main experiment.

Procedure for the preliminary experiment. Upon volunteering to participate in this part of the study, each subject was given a static visual acuity test and a color vision test using a Titmus II vision tester. All subjects were required to have a minimum visual acuity of 20/40 (corrected or uncorrected) as well as normal color vision. This device uses pseudo-isochromatic plates as a color vision test.

As an initial setup, blue-green filters were placed in the projectors and on the central instrument panel of the simulator. Blue-green was used as the reference color throughout because it has been used extensively in automobiles and, hence, it was likely to be familiar to subjects. A typical three word-slide was inserted into both projectors (i.e., two identical slides). The character size on the slides was 15 arcmin. One projector, selected randomly for each subject, was then turned on. The simulator roadway scene display was also turned on, and the simulator central instrument panel lighting was kept off.

The subject was then asked to sit in the simulator. The lights in the room were turned off. After four minutes in the dark, the subject was asked to determine the brightness that was most comfortable for night driving. Following the subject's instructions, the experimenter performed the luminance adjustment by varying the current in the projector bulb. When a subjectively comfortable level had been reached, the second projector was turned on and its luminance was adjusted by the experimenter to match that of the first one. This was accomplished by setting the current to the same value as that of the first projector (the projectors had been checked earlier to insure that they yielded equal luminances for equal currents).

The subject was then asked to examine both projector displays and tell if the brightness was appropriate according to his judgement. If not, further adjustments were made until a comfortable level was reached. Once this goal was attained, the central instrument panel lighting was turned on and the subject was instructed to adjust its brightness to match that of the projector displays to obtain a uniform brightness across the central instrument panel and the displays. Once this step was completed, the subject was again asked if further brightness adjustments either on the projector displays or on the simulator central instrument panel were needed. When the final levels were obtained, the current in the bulbs as well as the potentiometer setting for the central instrument panel lighting were recorded. These values constituted the reference against which the other colors would be matched.

To determine the preferred brightness for any color (other than blue-green), a color filter was inserted into one of the projectors, again selected at random. A blue-green filter was then installed in the other projector and its brightness was adjusted to the reference value. The simulator central instrument panel lighting was turned off. The subject was then instructed to match the brightness of the color to that of the blue-green. Again, following the subject's instructions, the experimenter performed the current adjustments. When a match was obtained, the blue-green filter was replaced by a filter of the color being matched and the current was adjusted to obtain the same brightness as in the other projector. A color filter was then placed on the simulator central instrument panel and the subject was asked to match its brightness to that of the projector displays to obtain uniform brightness across the central instrument panel and the displays. Once the subject was satisfied with the adjustments, the current value in the bulbs as well as the central instrument panel potentiometer setting were recorded. This procedure was then repeated for all six remaining colors.

Subjects always used their original blue-green reference setting to set the brightness of the new color. A subject was always given ample time to perform the adjustments. The order in which the seven colors (excluding blue-green) were presented was randomized for each subject.

Once all brightness levels were obtained for all eight colors, the subject was debriefed, paid, and dismissed. (The subjects did not drive the simulator in the preliminary experiment.)

It should be noted that one layer of neutral density filter was put just behind each color filter on the front of the bulb drawer unit, to avoid hot spots in the image.

Color matching. The goal of the color matching part was to obtain a best match to the chromaticity coordinates specified using CIE Standard Illuminant A (i.e., 2856°K) for the color filters provided.

For purposes of color matching, the Minolta model CS100 chroma meter and model DP101 processor were used. This processor offers the possibility of storing the chromaticity coordinates of up to four colors to be used as measurement references. When the chromaticity coordinates of a new color are measured, measurement error is minimized when the closest reference color is used, that is, when the distance (on the chromaticity diagram) between the color to be measured and the reference color (stored in the processor) is the smallest. For instance, the red color was measured using a red reference. Appendix E provides the the chromaticity coordinates of the four reference colors that were used for all chromaticity measurements.

The L*a*b* chromaticity coordinates provided with the color filters to be used in the projectors were then transformed to CIE 1931 chromaticity coordinates. These chromaticity coordinates, which will be called the "theoretical" chromaticity coordinates, are given in Appendix F.

The CIE 1931 chromaticity coordinates as well as the luminance of one of the projector displays were then measured for each color filter using the corresponding median current values obtained in the preliminary experiment (8 colors x 2 luminance levels). A slide with a transparent square (10 x 10 mm) was used to fill the measuring area of the chromaticity meter with colored light. Also, one layer of neutral density filter was put behind the color filter to avoid any hot spot in the area being measured. These coordinates were called "practical" chromaticity coordinates. The group median current value obtained in the preliminary experiment for each color was used for the measurements instead of the group median luminance value because adjustments had been done on the current and not on the luminance in the preliminary experiment and also, because the relationship between current and luminance was not linear.

Since in the preliminary experiment, brightness was varied by changes in the current only, it was expected that the spectral characteristics of the light would be different than that of a CIE Standard Illuminant A. The chromaticity coordinates of the colors experienced by the subjects in the preliminary experiment (practical chromaticity coordinates) would then be slightly different than those provided with the color filters and specified using CIE Standard Illuminant A (theoretical chromaticity coordinates). Hence, the luminance values obtained were corrected so that they could be used with the theoretical chromaticity coordinates. This correction was accomplished by using the Ware and Cowan (1983) equations described earlier in the literature review.

With such a correction, the brightness of a display having the practical chromaticity coordinates and the luminance value (i.e., median luminance value) obtained in the preliminary experiment was expected to be the same as the brightness of a display having the theoretical chromaticity coordinates and the corrected luminance value, for a given color filter. Practical and theoretical chromaticity coordinates as well as obtained and corrected luminance values were different, whereas brightness was expected to be equivalent.

Given the fact that the bulbs in the projectors closely approximated CIE Standard Illuminant A when an opaque light blue filter was put over the mirror, and the bulb driven at a current of 1.227 amps, color filters, neutral density filters, and transparent color filters used for correction of the bulb spectral characteristics as described earlier (Appendix G) were all combined until a best match to the theoretical chromaticity coordinates and the corrected luminance values was reached. The specific combinations of various filters as well as the current values to a best match for each of the 16 color by brightness treatment combinations were recorded and later used in the main experiment.

In summary, on completion of the preliminary experiment, corrected subjective brightness levels and accurate chromaticity coordinates had been developed for each of the 16 color by brightness combinations. These treatment combinations were then used in the main experiment. Figure 7 presents a summary of the procedure for the preliminary experiment.

Main experiment. Upon volunteering to participate in the study and verification of a valid driver's license, each subject was asked to read the participant's informed consent form which included a general description and instructions about the experiment (Appendix H). The experimenter then attempted to clarify any questions the volunteer had. If the individual desired to participate in the experiment, he or she was asked to sign the consent form. After written consent had been obtained from the subject, he or she was asked to carefully read an instruction sheet describing the subjective measures and the corresponding scales (Appendix C). The experimenter then attempted to clarify any questions the subject had.

Next, each subject was given a static visual acuity test as well as a color vision test using the same testing apparatus as in the preliminary experiment. All subjects were required to have a minimum visual acuity of 20/40 (corrected or uncorrected) as well as normal color vision. The experimenters then made sure that the subject could hear clearly the clicking sound emitted by the projector during a slide change.

The subject was then asked to sit in the simulator. The seat belt was buckled and the two microphones placed on the subject's collar. The simulator was then activated and the subject drove the simulator for a few minutes to "get the feel" of it. During this practice

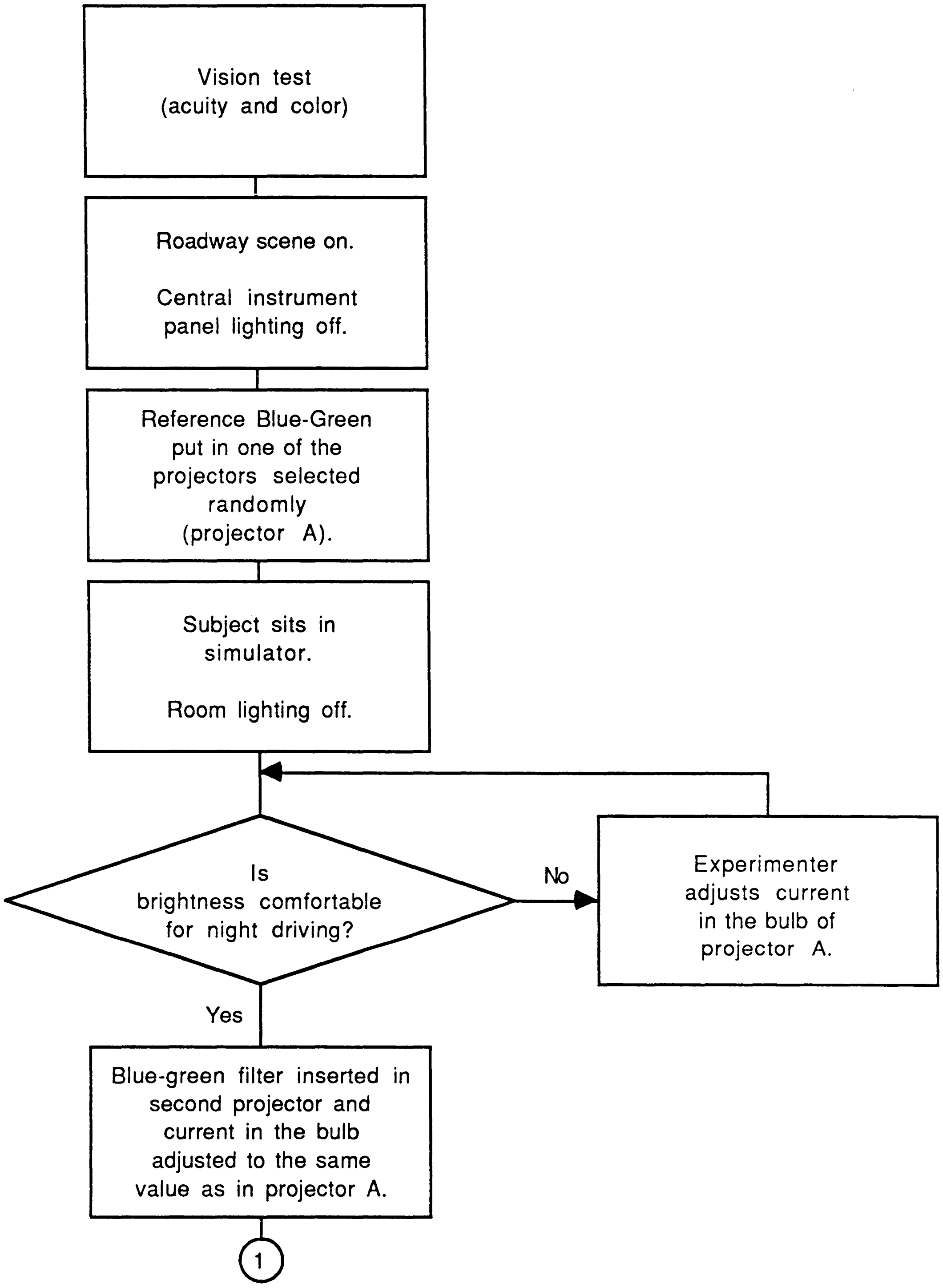


Figure 7. Procedure for the preliminary experiment.

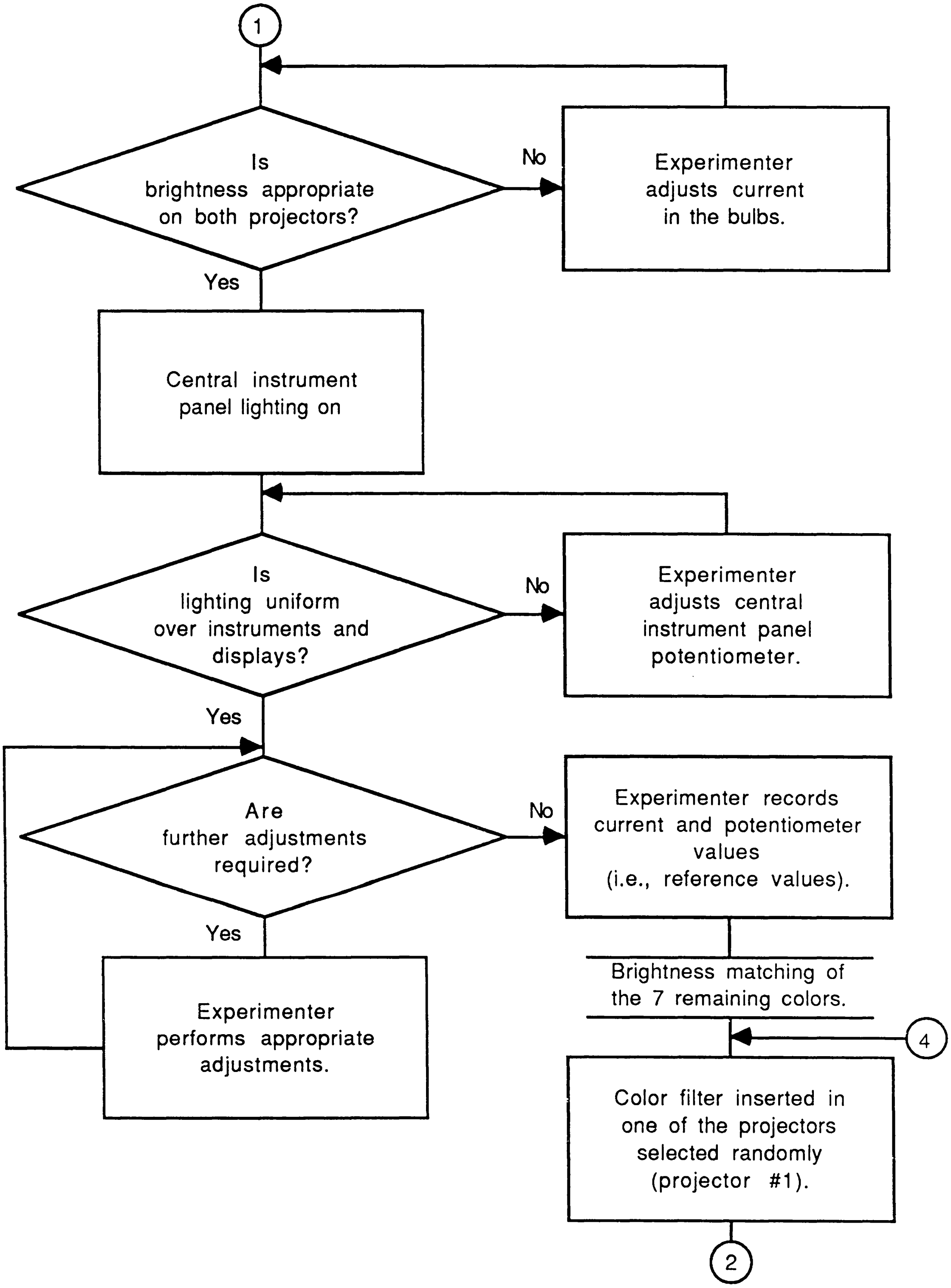


Figure 7. Procedure for the preliminary experiment (cont'd).

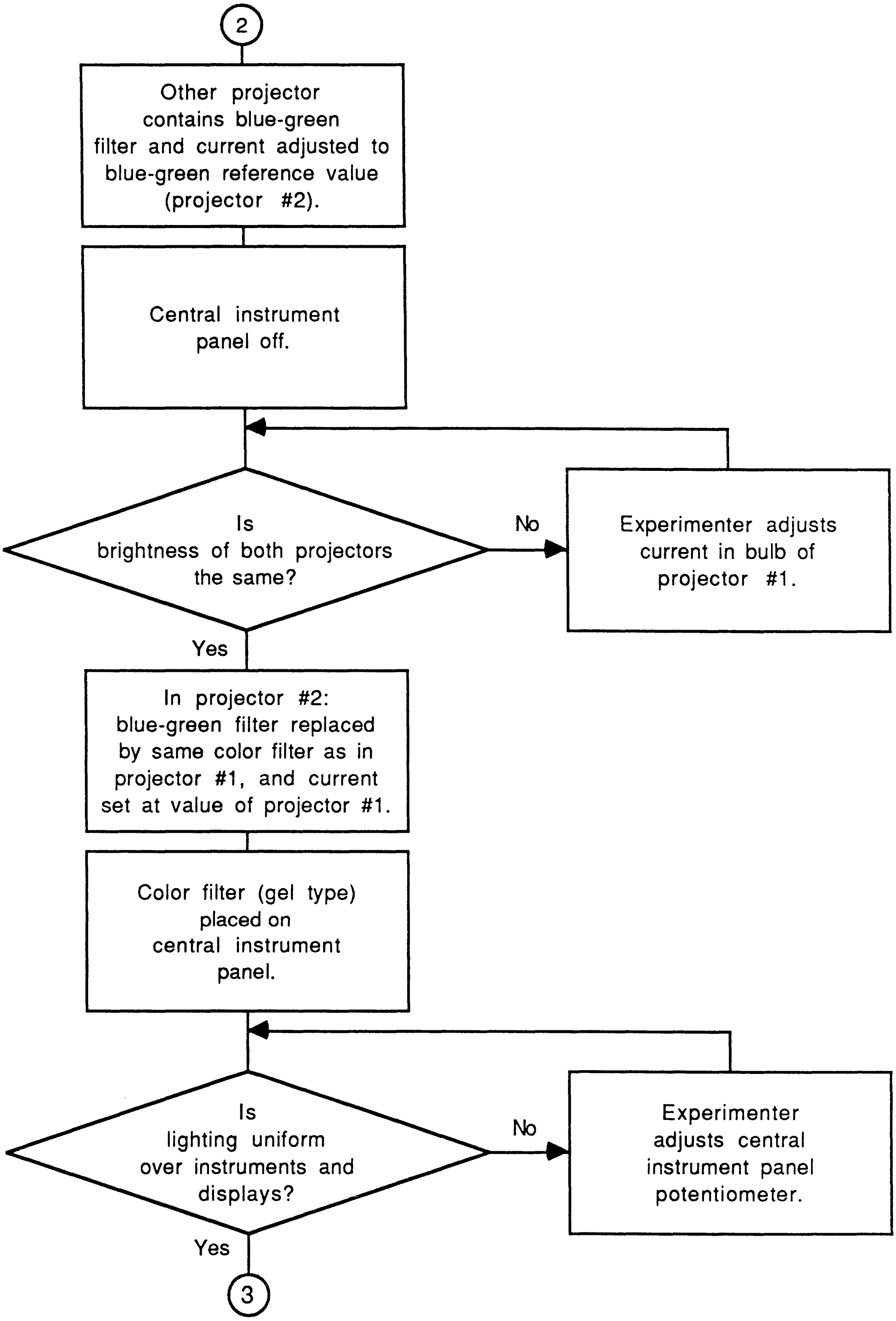


Figure 7. Procedure for the preliminary experiment (cont'd).

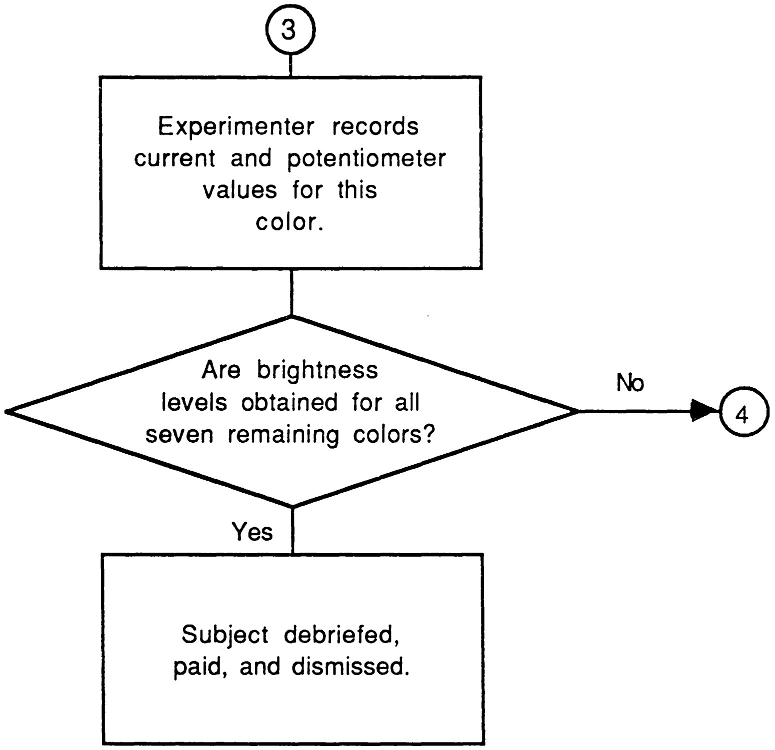


Figure 7. Procedure for the preliminary experiment (cont'd).

run, the lights were turned off to allow the subject to dark adapt. The lights were then be kept off until the 10-min break about halfway through the experiment.

Once the practice run was completed, the subject was asked to stop the vehicle and was then given further verbal instructions on the driving and word reading tasks which include the following:

- While being shown one of the slides, the subject was instructed to read only one of the words (top, middle, or bottom) on the slide as quickly as possible upon hearing the click of the projector during the run. He or she would be reminded at the beginning of each 32 word run which of the three words to read throughout the run.
- The subject was instructed to look at the projector displays only to read a word and to otherwise look on the road, since driving was the main and most important task. This instruction was given to avoid glances to the projector before the slide appeared.
- The subject was instructed to drive at a speed between 55 and 60 mph (the speedometer was in mph) in the right hand lane at all times.
- The subject was instructed to speak loudly when reading the words so that his or her answer would be registered by the vox. The subject was also reminded not to make any sound other than answers if possible.
- The subject was informed that wind gusts were simulated and thus he or she could not assume that the vehicle would remain in its lane position if no steering correction were made.

After these instructions, the simulator was activated and the subject asked to accelerate to the proper speed. Once the speed was reached, the subject was told which of the three words to read on each slide throughout the following run. The experimenters then activated the computer programs and the words started to appear. One experimenter viewed the video monitor and actuated the glance pushbutton while the other checked the words read by the subject for correctness.

The preprogrammed driving scenario alternated between curved and straight stretches of road. During the run, the subject was reminded to check the speed whenever it was not within range.

After a run was completed, the subject was asked to stop the simulated vehicle and the subjective measures were taken by one of the experimenters. While making judgments verbally, the subject could view slides showing a grid (Appendix B) so that he or she could still see the color. Verbal answers were used because they did not interfere with the subject's dark adaptation (written answers would have required some lighting due to the darkness in the room). Once the subjective measures were taken, the experimenters changed color filters, adjusted the current in the bulbs, set the simulator central instrument panel potentiometer, and set the carousels for the next run. The subject was then instructed to bring the vehicle up to speed once again. This process was repeated for seven consecutive runs.

A 10-min break was then given to the subject. During this break, the room lights were turned on and the subject left the simulator to stretch, walk about, and refresh if necessary. After this break, the subject returned to the simulator, the seat belt was buckled, and the microphones were attached. After extinguishing the room lighting, the subject drove freely for a few minutes to allow for readaptation. The experiment then continued for the remaining nine runs.

Upon completion of the 16 runs, the subject was debriefed, paid, and dismissed. The entire experiment lasted about two and one-half hours per subject.

RESULTS AND DISCUSSION

For the sake of simplicity, the following shortened color names will be used in the figures and tables showing the results of the study:

LIBLU = Light blue
BLUGR = Blue-green
GREEN = Green
AMBER = Amber
ORANG = Orange
REDOR = Reddish orange
RED = Red
WHITE = White

Preliminary Experiment and Color Matching

The median current values obtained for each color in the preliminary experiment as well as the corresponding luminances and CIE 1931 chromaticity coordinates for the two groups of subjects are given in Appendix I. The subjects provided relatively consistent current values as shown in Figures 8 and 9.

Appendix J presents the corrected (or theoretical) luminance values of each color and brightness level computed with the Ware and Cowan (1983) equations, from the luminance values corresponding to the median currents obtained in the preliminary experiment (in Appendix I). Note that the maximum change was less than 5% of the original value.

Appendix K presents the luminance values and CIE 1931 chromaticity coordinates of the best matches to the theoretical luminances (Appendix J) and theoretical CIE 1931 chromaticity coordinates (Appendix F). As shown in Table 1, the difference in luminance between the value obtained in the matching operation and the theoretical value is below 1%

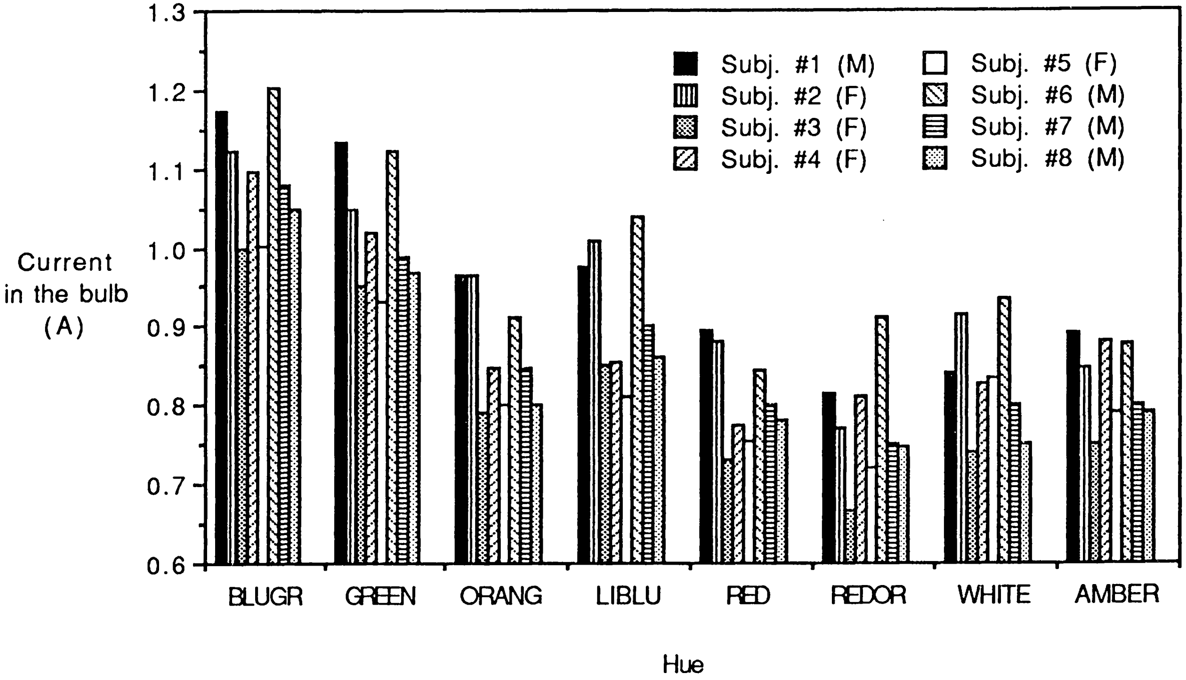


Figure 8. Current values for the younger subjects in the preliminary experiment (F = female, M = male).

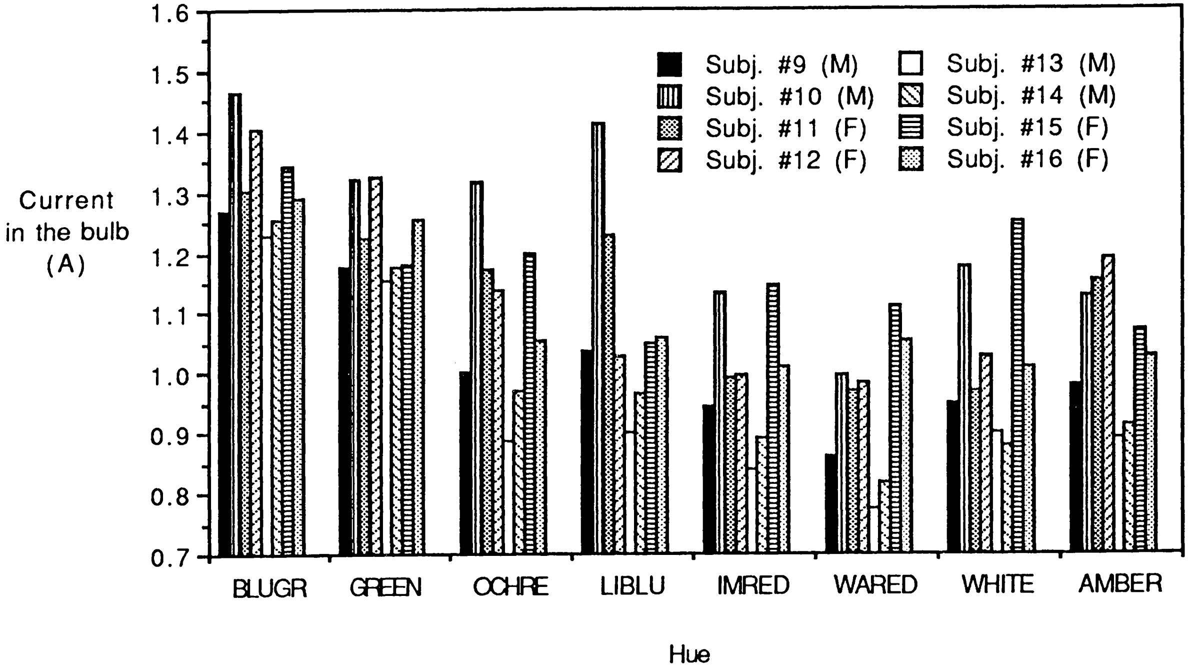


Figure 9. Current values for the older subjects in the preliminary experiment

(F = female, M = male).

TABLE 1

Luminance and Chromaticity Differences Between the Matched Colors and the Theoretical Colors

Color	*Brightness*	ΔY % [1]
BLUGR	Low	0.00
	High	-0.12
GREEN	Low	0.61
	High	0.13
ORANG	Low	0.44
	High	0.43
LIBLU	Low	0.00
	High	0.38
RED	Low	0.54
	High	-0.90
REDOR	Low	0.00
	High	0.14
WHITE	Low	0.00
	High	0.00
AMBER	Low	-0.30
	High	-0.20

---

[1] Difference between the theoretical luminance values (i.e., Corrected Y in Appendix J) and the luminance values obtained in the color matching operation (i.e., Matched Y in Appendix K).

for all colors. Appendix L gives the transformation of the matched CIE 1931 chromaticity coordinates to both CIE 1976 $L^*a^*b^*$ and CIE 1976 $L^*u^*v^*$ spaces. The excitation purity and dominant wavelengths of the matched colors are given in Appendix M.

As expected, the older subjects selected higher luminance values than their younger counterparts in the preliminary experiment ($t\,(14) = 3.496$, $p < 0.005$)(Figure 10). This is especially true for reddish orange. Both curves in Figure 10 appear to show a similar overall pattern. Since the difference between the luminance values obtained in the preliminary experiment and the values obtained in the color matching operation is very small (i.e., less than 5%), the curves associated with each set of values, when plotted as a function of dominant wavelength, will be virtually identical (i.e., Figure 10). One could then wonder why older people selected a very high luminance value for reddish orange. A close examination of Figure 10 reveals that the hues in the longer wavelengths —reddish orange, orange, and amber— all had fairly high luminance values except red. Also, Appendix M shows that these colors also had the highest excitation purity (i.e., saturation). It can be speculated that older people, being presbyopic, do not have the accommodation power to focus the image exactly on the retina at more saturated longer wavelengths. Hence, the image being focused slightly behind the retina appears blurred, and consequently dimmer. At a constant saturation, as the dominant wavelength moves toward the red end, the effect is worsened and to compensate, subjects increase the luminance. Red was an exception because it was much less saturated than the other colors in the longer wavelengths.

The finding that red exacerbates presbyopia tends to support this concept (Reynolds, 1971). It makes even more sense for the bifocal wearers (which were a majority among older people) since the distance at which the words appeared is just between best focus distances of their eye glasses. Some subjects wearing bifocals have

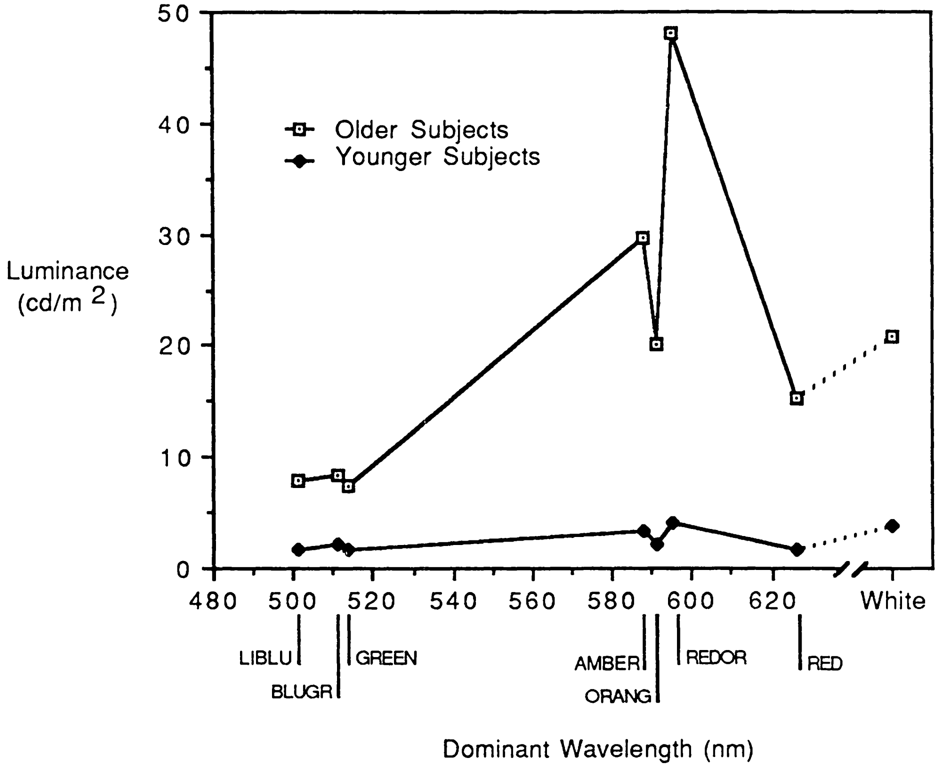


Figure 10. Luminance values obtained in the preliminary experiment as a function of hue and age group.

reported that with reddish orange, the two smaller character sizes appeared doubled and thus, were more difficult to read. Since this information has not been collected systematically, it cannot be generalized to all subjects wearing bifocals in the experiment.

The results for white and amber do not agree with this concept, however. People in the older age group may simply have preferred brighter white. For amber, an alternate explanation will be given later when performance measures are analyzed.

The filter combinations and the current values in the bulbs used to obtain the matched colors are described in Appendix N.

As expected with the equipment used, the contrast ratios (i.e., Lum_{high}/Lum_{low}) on the displays were very high; they ranged from 330:1 to 9860:1 (Appendix O). A display-to-scene contrast ratio was computed; the values ranged from 0.37:1 to 10.96:1 (Appendix O). Appendix O also provides ΔE distances between the illumination color of the letters and their background computed for the standard Yu'v' space as well as for a Yu'v' space rescaled to better describe reading performance (see Lippert, 1986 or Lippert and Snyder, 1986).

Main Experiment

To keep the figures and tables simple, the following convention concerning character size will be used: *CS7* = 7, *CS11* = 11, *CS17* = 17, and *CS25* = 25 minutes of arc (arcmin).

Note that the objective measures which are the following:

- response time from stimulus presentation to a correct response (or time to a correct response),
- response time regardless of correctness (or time to any response),
- glance time at the display for a correct response,

- percentage of correct responses,
- percentage of correct and incorrect responses,
- variance of lane deviation regardless of response

and the subjective measures:

- subjective attractiveness
- subjective comfort
- subjective ease of readability

will be the object of separate statistical analyses, since they do not involve the same set of independent variables as described previously; for instance, the subjective measures do not involve Word Complexity.

Main Experiment: Objective Measures

Missing data. The data reduction process showed that some of the subjects, particularly those in the older age group, did not provide a response at any of the four replications, within the allowed eight-s interval, for many of the treatment combinations. Consequently, some cells in the experimental design did not include response times for all four subjects; the design was thus unbalanced. Statistical analysis of such a design was attempted using the Statistical Analysis System (SAS) package release 5.16, available at Virginia Tech's computing center. However, the memory required for the model (i.e., 1,609 megabytes) used in this study was beyond the system's capabilities. Even a trimmed version of the model including only the main effects and two-way interactions still necessitated too much memory.

To cope with this problem, eight seconds was used as a replacement for all missing values of response time in the design. The assumption was that, in these no response cases the subject would have eventually provided a response. However, it would have taken longer than the eight seconds allowed, and thus, eight seconds was the best estimate of that response time. Such substitutions were done only once in the younger age group, six times in the middle age group and 118 times in the older age group. The proportion of data points thus modified in each age group was 0.1% in the younger age group, 0.6% in the middle age group, and 11.5% in the older age group. Clearly, the data from the younger and middle age groups remained unmodified.

Objective measures: MANOVA. A multivariate analysis of variance (MANOVA) was performed on the six dependent variables for all effects in the model (Table 2); conversion of Wilk's Lambda to an *F* ratio was used as the decision criterion.

All main effects except gender were significant (i.e., age *(A)*, character size *(CS)*, hue *(H)*, brightness *(B)*, and word complexity *(WC)*). In addition the following interactions were significant: *CS x A, B x A, H x CS, B x CS, B x H, WC x CS, B x CS x A, CS x H x B, B x CS x A x G,* and *CS x WC x B x A x G.* In subsequent analyses, the *B x CS x A x G* and *CS x WC x B x A x G* interactions will be neglected because of the usual difficulties in interpreting high order interactions.

As expected gender was not significant as either a main effect or a lower order interaction.

Objective measures: ANOVAs and post hoc tests. Each effect was then tested in separate analyses of variance; one for each dependent variable. The names of the dependent variables have been shortened as follows in the tables and figures:

- response time to a correct response or a no response = *RTCNO*
- response time regardless of response = *RTANY*

TABLE 2

Multivariate Analysis of Variance: Objective Measures

Effect	*df*	*F ratio*	*p*	
Between-Subjects				
A (Age)	2	2.23	0.0419	*
G (Gender)	1	2.21	0.1087	
A x G	2	0.39	0.9552	
Within-Subjects				
CS (Character Size)	3	9.45	0.0001	*
CS x A	6	2.62	0.0001	*
CS x G	3	0.56	0.9247	
CS x A x G	6	0.76	0.8417	
CS x S/AG	54			
H (Hue)	7	1.82	0.0016	*
H x A	14	0.85	0.8295	
H x G	7	1.01	0.4629	
H x A x G	14	0.75	0.9489	
H x S/AG	126			
B (Brightness)	1	43.24	0.0001	*
B x A	2	3.71	0.0025	*
B x G	1	2.88	0.0521	
B x A x G	2	2.04	0.0627	
B x S/AG	18			
WC(Word Complexity)	1	12.84	0.0001	*
WC x A	2	0.96	0.5107	
WC x G	1	0.37	0.8877	
WC x A x G	2	0.67	0.7638	
WC x S/AG	18			

TABLE 2 cont'd

Effect	*df*	*F ratio*	*p*	
H x CS	21	1.45	0.0010	*
H x CS x A	42	0.93	0.7533	
H x CS x G	21	0.92	0.7361	
H x CS x A x G	42	0.73	0.9994	
H x CS x S/AG	37			
B x CS	3	6.21	0.0001	*
B x CS x A	6	2.99	0.0001	*
B x CS x G	3	1.34	0.1710	
B x CS x A x G	6	1.60	0.0229	*
B x CS x S/AG	54			
B x H	7	1.60	0.0113	*
B x H x A	14	1.05	0.3562	
B x H x G	7	0.79	0.8236	
B x H x A x G	14	1.06	0.3382	
B x H x S/AG	126			
WC x CS	3	3.28	0.0001	*
WC x CS x A	6	1.29	0.1373	
WC x CS x G	3	0.47	0.9669	
WC x CS x A x G	6	0.91	0.6168	
WC x CS x S/AG	54			
WC x H	7	1.04	0.3978	
WC x H x A	14	0.93	0.6434	
WC x H x G	7	0.99	0.4930	
WC x H x A x G	14	1.14	0.1919	
WC x H x S/AG	126			
WC x B	1	1.65	0.2105	
WC x B x A	2	1.68	0.1311	
WC x B x G	1	1.19	0.3690	
WC x B x A x G	2	1.01	0.4688	
WC x B x S/AG	18			

TABLE 2 cont'd

Effect	*df*	*F ratio*	*p*	
CS x WC x H	21	1.04	0.3705	
CS x WC x H x A	42	0.77	0.9955	
CS x WC x H x G	21	1.02	0.4148	
CS x WC x H x A x G	42	1.01	0.4597	
CS x WC x H x S/AG	378			
CS x WC x B	3	1.26	0.2246	
CS x WC x B x A	6	1.21	0.2056	
CS x WC x B x G	3	1.30	0.1988	
CS x WC x B x A x G	6	1.56	0.0294	*
CS x WC x B x S/AG	54			
CS x H x B	21	1.35	0.0066	*
CS x H x B x A	42	1.06	0.2679	
CS x H x B x G	21	0.98	0.5557	
CS x H x B x A x G	42	1.01	0.4434	
CS x H x B x S/AG	378			
H x WC x B	7	1.13	0.2645	
H x WC x B x A	14	0.76	0.9443	
H x WC x B x G	7	0.88	0.6855	
H x WC x B x A x G	14	0.87	0.7932	
H x WC x B x S/AG	126			
CS x H x WC x B	21	1.22	0.0535	
CS x H x WC x B x A	42	0.84	0.9590	
CS x H x WC x B x G	21	1.04	0.3542	
CS x H x WC x B x A x G	42	0.92	0.8069	
CS x H x WC x B x S/AG	378			

- glance time for a correct response or a no response = *GTCNO*
- percentage of correct responses = *%CORR*
- percentage of correct and incorrect responses = *%CINC*
- variance of lane deviation regardless of response = *LDANY*

The main results of the ANOVAs are presented in Tables 3 to 15 in the same order as in Table 2.

Post hoc tests were then performed on all significant main effects and interactions. For interactions, simple-effects tests were used. The Ryan-Einot-Gabriel-Welsh multiple-range method (SAS, 1985) was chosen for all means comparisons because it offers power and good control over the mean experimentwise error rate. The results of the comparisons for a given effect are shown in the corresponding figure; any two means having at least one letter in common are not significantly different at the 0.05 alpha level. For the interactions, two sets of letters were required. The first set includes letters not in parentheses while the second set includes letters in parentheses. For each level of the factor on the x axis of the figure, the letters not in parentheses describe comparisons among the different levels of the other factor(s) involved in the interaction. Hence, these letters must be read vertically on the figure since they compare points corresponding to a single level of the factor on the x axis. The set of letters in parentheses describes horizontal means comparisons, that is, comparisons among the different levels of the factor on the x axis and for a single level of the other factor(s) involved in the interaction.

Each effect (i.e., interactions and main effects) is examined separately in the remainder of this section, in the following order: *CS x H x B*, *H x CS*, *B x H*, *H*, *B*, *CS*, *B x CS x A*, *CS x A*, *B x A*, *B x CS*, *A*, *WC x CS*, and *WC*.

TABLE 3

Summary of the Effect of Age *(A)* on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	430.506	7.06	0.0054	*
RTANY	412.308	6.78	0.0064	*
GTCNO	23.773	8.16	0.0030	*
%CORR	6.616	6.63	0.0069	*
%CINC	6.650	6.67	0.0068	*
LDANY	22.193	4.20	0.0319	*

[1] The Age main effect has 2 *df*, the error *(S/AG)* has 18 *df*.

TABLE 4

Summary of the Effect of Character Size *(CS)* on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*
RTCNO	803.679	48.72	0.0001 *
RTANY	788.752	50.19	0.0001 *
GTCNO	67.635	53.51	0.0001 *
%CORR	11.074	23.59	0.0001 *
%CINC	5.963	13.65	0.0001 *
LDANY	6.911	23.59	0.0001 *

[1] The Character size main effect has 3 *df*, the error *(CS x S/AG)* has 54 *df*.

TABLE 5

Summary of the Effect of the Character Size by Age *(CS x A)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	150.692	9.13	0.0001	*
RTANY	140.861	8.96	0.0001	*
GTCNO	2.140	1.69	0.1404	
%CORR	3.397	7.24	0.0001	*
%CINC	3.571	8.18	0.0001	*
LDANY	0.576	1.97	0.0868	

[1] The *CS x A* interaction has 6 *df*, the error *(CS x S/AG)* has 54 *df*.

TABLE 6

Summary of the Effect of Hue *(H)* on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	4.946	2.80	0.0096	*
RTANY	5.184	3.01	0.0058	*
GTCNO	1.123	6.19	0.0001	*
%CORR	0.084	2.77	0.0104	*
%CINC	0.044	1.46	0.1868	
LDANY	0.097	0.42	0.8904	

[1] The Hue effect has 7 *df*, the error *(H x S/AG)* has 126 *df*.

TABLE 7

Summary of the Effect of Brightness *(B)* on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	116.365	118.98	0.0001	*
RTANY	113.822	135.28	0.0001	*
GTCNO	7.503	65.71	0.0001	*
%CORR	1.699	100.43	0.0001	*
%CINC	1.058	39.68	0.0001	*
LDANY	0.727	7.93	0.0114	*

[1] The Brightness main effect has 1 *df*, the error *(B x S/AG)* has 18 *df*.

TABLE 8

Summary of the Effect of the Brightness by Age *(B x A)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	17.694	18.09	0.0001	*
RTANY	16.063	19.09	0.0001	*
GTCNO	0.623	5.46	0.0140	*
%CORR	0.353	20.89	0.0001	*
%CINC	0.367	13.77	0.0002	*
LDANY	0.125	1.37	0.2796	

[1] The *B x A* interaction has 2 *df*, the error *(B x A x S/AG)* has 18 *df*.

TABLE 9

Summary of the Effect of Word Complexity *(WC)* on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	5.819	16.72	0.0007	*
RTANY	7.582	21.15	0.0002	*
GTCNO	2.121	19.18	0.0004	*
%CORR	0.003	0.73	0.4043	
%CINC	0.014	3.92	0.0634	
LDANY	0.257	2.28	0.1485	

[1] The Word Complexity main effect has 1 *df*, the error *(WC x S/AG)* has 18 *df*.

TABLE 10

Summary of the Effect of the Hue by Character Size *(H x CS)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	1.936	2.82	0.0001	*
RTANY	2.093	3.24	0.0001	*
GTCNO	0.251	2.83	0.0001	*
%CORR	0.055	2.47	0.0004	*
%CINC	0.028	1.33	0.1516	
LDANY	0.105	0.81	0.7085	

[1] The *H x CS* interaction has 21 *df*, the error *(H x CS x S/AG)* has 378 *df*.

TABLE 11

Summary of the Effect of the Brightness by Character Size *(B x CS)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	50.770	37.05	0.0001	*
RTANY	48.144	37.81	0.0001	*
GTCNO	2.553	17.73	0.0001	*
%CORR	1.132	33.86	0.0001	*
%CINC	0.737	18.82	0.0001	*
LDANY	0.367	4.01	0.0120	*

[1] The *B x CS* interaction has 3 *df*, the error *(B x CS x S/AG)* has 54 *df*.

TABLE 12

Summary of the Effect of the Brightness by Character Size by Age *(B x CS x A)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	5.703	37.05	0.0017	*
RTANY	4.830	3.79	0.0032	*
GTCNO	1.375	9.55	0.0001	*
%CORR	0.153	4.60	0.0008	*
%CINC	0.224	5.74	0.0001	*
LDANY	0.340	3.71	0.0037	*

[1] The *B x CS x A* interaction has 6 *df*, the error *(B x CS x S/AG)* has 54 *df*.

TABLE 13

Summary of the Effect of Brightness by Hue *(B x H)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	1.677	3.58	0.0015	*
RTANY	1.634	3.67	0.0012	*
GTCNO	0.190	1.72	0.1094	
%CORR	0.051	3.33	0.0028	*
%CINC	0.029	2.84	0.0089	*
LDANY	0.094	0.60	0.7547	

[1] The *B x H* interaction has 7 *df*, the error *(B x H x S/AG)* has 126 *df*.

TABLE 14

Summary of the Effect of the Word Complexity by Character Size *(WC x CS)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	0.373	1.72	0.1730	
RTANY	1.094	4.83	0.0048	*
GTCNO	0.232	7.69	0.0002	*
%CORR	0.001	0.44	0.7249	
%CINC	0.020	7.40	0.0003	*
LDANY	0.075	0.71	0.5495	

[1] The *WC x CS* interaction has 3 *df*, the error *(WC x CS x S/AG)* has 54 *df*.

TABLE 15

Summary of the Effect of the Character Size by Hue by Brightness *(CS x H x B)* Interaction on the Dependent Measures

Dependent Variable	*MS*	*F ratio* [1]	*p*	
RTCNO	1.149	2.63	0.0001	*
RTANY	1.060	2.63	0.0001	*
GTCNO	0.108	1.54	0.0624	
%CORR	0.041	2.68	0.0001	*
%CINC	0.024	2.09	0.0035	*
LDANY	0.085	0.80	0.7243	

[1] The *CS x H x B* interaction has 21 *df*, the error *(CS x H x B x S/AG)* has 378 *df*.

Character size by hue by brightness interaction (CS x H x B). The character size by hue by brightness interaction was significant in four of the dependent measures: both response time measures and both accuracy measures.

Figure 11 shows the percentage of correct responses as a function of dominant wavelength, character size, and brightness. Accuracy is lowest with the 7 arcmin character size, especially at low brightness. In that case, accuracy varies by as much as 28% depending on the hue used; that is, from 52% of correct responses for green to 80% for white. The variations are much less important and not statistically different among the hues at high brightness for this character size (6.8%). For the other character size by brightness treatment combinations, there is no statistically significant difference in accuracy among the hues.

Figure 12 shows the time to a correct response as a function of dominant wavelength, character size, and brightness. This figure is almost an exact inverse of the previous figure. The 7 arcmin character size at low brightness yielded variations of up to 1.42 s in response time depending on the hue used; that is, from 3.45 s for white to 4.86 s for green. At high brightness, the corresponding maximum variation is on the order of 0.70 s and significant. The 11 arcmin character size yielded a maximum variation of 0.3 s in response time among the hues which is also significant. Finally, there is no significant difference in response time among the hues for the 11 arcmin at high brightness as well as for both larger character sizes at any brightness level.

Hue had a significant effect on accuracy with the 7 arcmin character size at low brightness only. However, the effect on response time was significant with the 7 arcmin at both brightness levels as well as with the 11 arcmin at low brightness. In these specific conditions no hue can be singled out as the worst; only somewhat overlapping groupings of colors can be observed. On the lower end are green and light blue; in the middle,

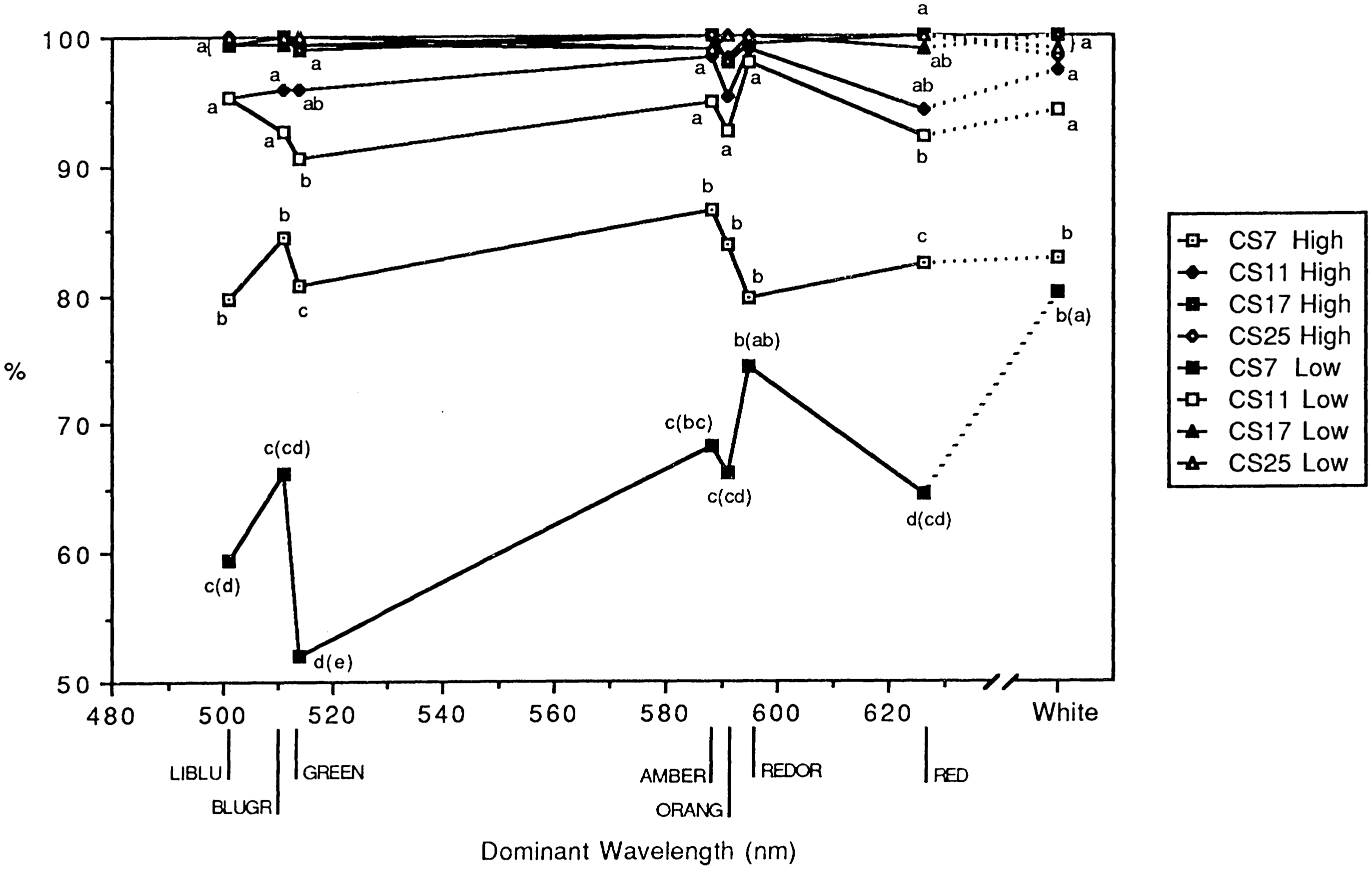


Figure 11. Percentage of correct responses (%CORR) as a function of dominant wavelength, character size, and brightness.

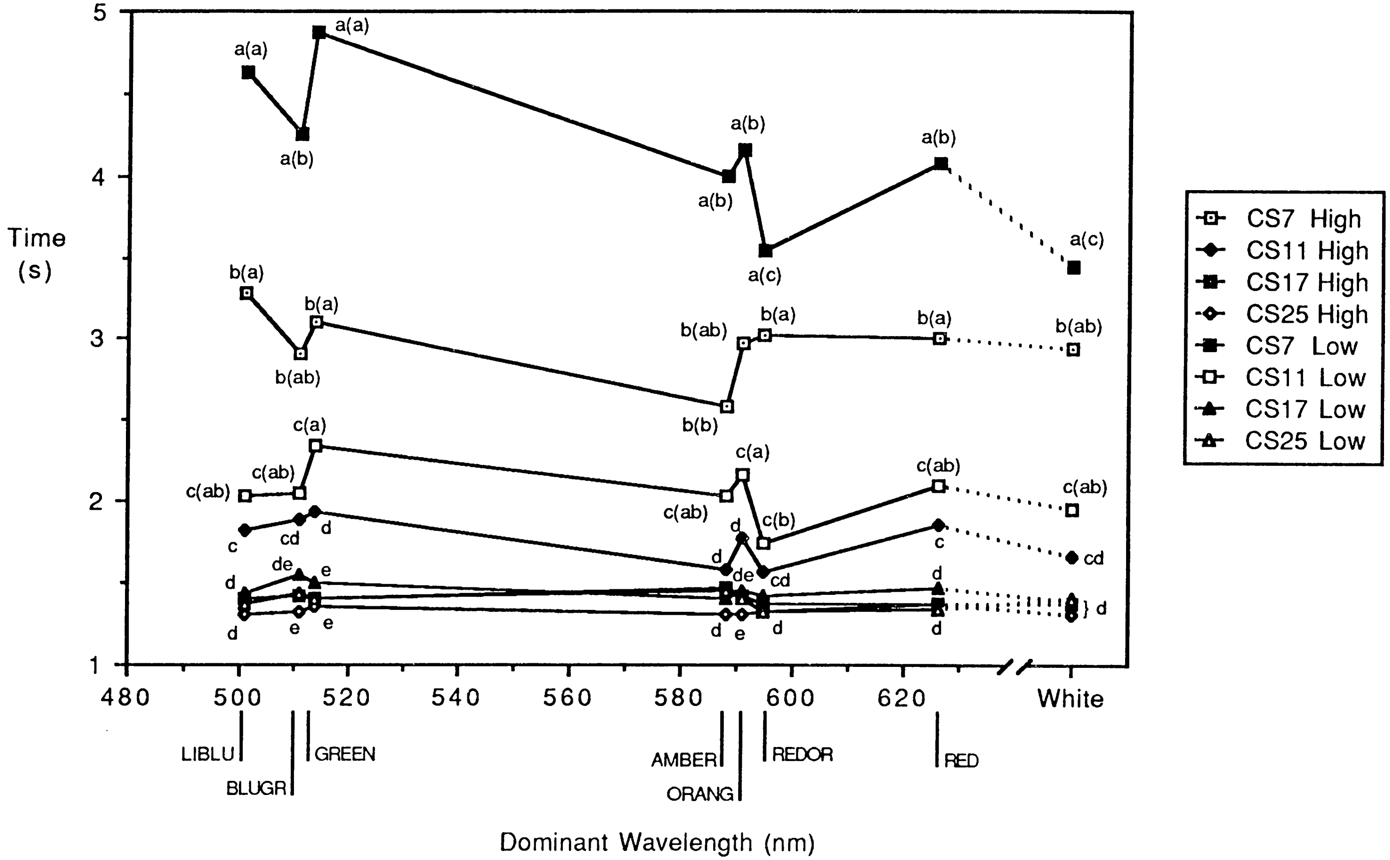


Figure 12. Time to a correct response or no response (RTCNO) as a function of dominant wavelength, character size, and brightness.

blue-green, amber, orange, and red; and at the other end, reddish orange and white (the "best" hues).

The fact that no difference in performance was observed among the hues with the 17 arcmin at both brightness levels supports the conclusion that the subjects selected luminance values that yielded fairly equivalent brightness among the hues, in the preliminary experiment. (The 17 arcmin size is very close to the 15 arcmin character size used in the preliminary experiment.)

When more luminance was provided to the 7 arcmin character size, the improvement in response time was statistically significant for all hues. Concerning accuracy, the improvement was significant for all hues but white and reddish orange. With the 11 arcmin, the improvement in performance associated with an increase in luminance consisted of a significant reduction in response time for green, amber, and orange only. With the two larger character sizes, no significant improvement in performance was observed following an increase in luminance.

The two other measures (i.e., time to any response and percentage of correct and incorrect responses) were not shown because they duplicate the results observed with the measures just analyzed, as will be the case with the *H x CS* and *B x H* interactions.

Hue by character size interaction (H x CS). The hue by character size interaction was significant for four measures: both response time measures, glance time at the displays, and percentage of correct responses.

Figure 13 shows the percentage of correct responses as a function of dominant wavelength and character size. Figure 14 shows the time to a correct response or no response for the same independent variables. As noted before, hue had an effect on performance only with the smallest character size.

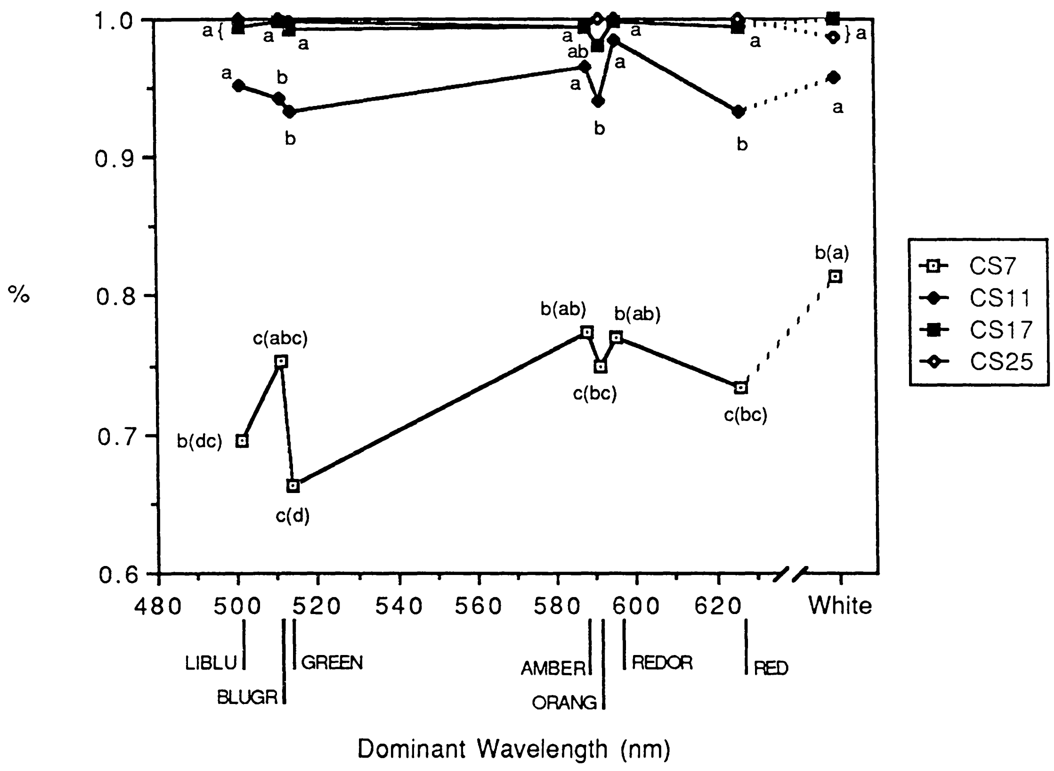


Figure 13. Percentage of correct responses (%CORR) as a function of dominant wavelength and character size.

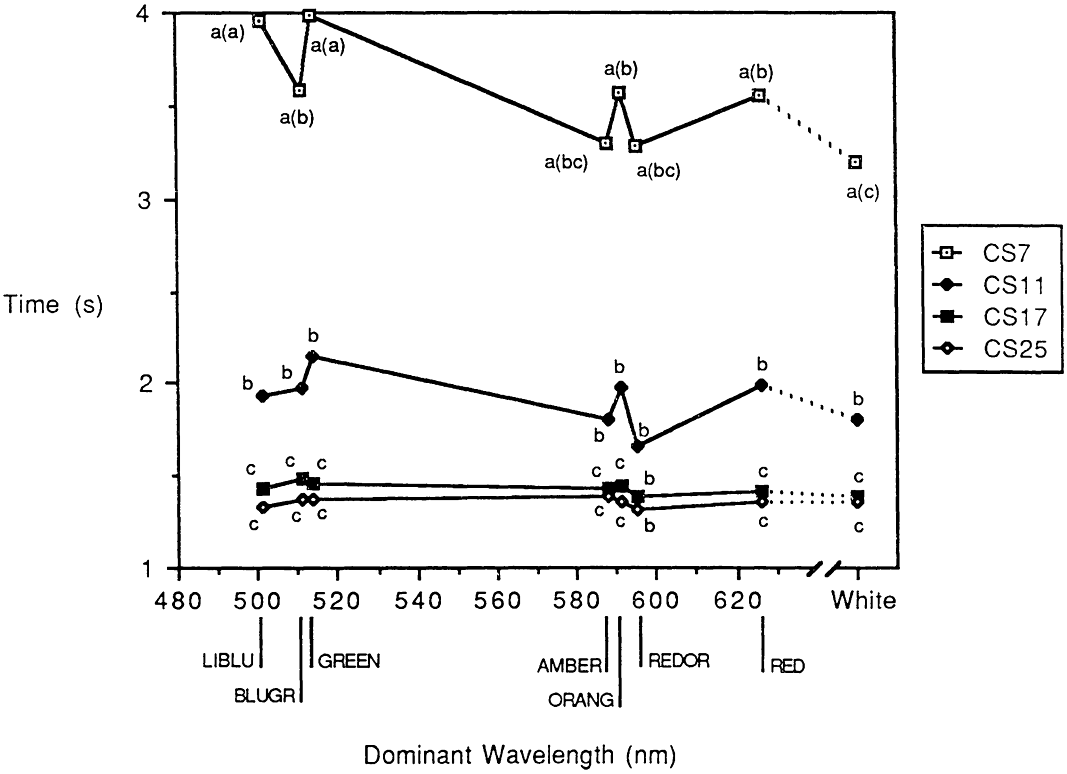


Figure 14. Time to a correct response or no response (RTCNO) as a function of dominant wavelength and character size.

As character size was increased from 7 arcmin to 11 arcmin, the improvements in accuracy and response time were significant for all hues. When character size was increased from 11 arcmin size to 17 arcmin, the improvement observed in response time was statistically significant for all hues but reddish orange. For accuracy, a corresponding improvement was observed only with blue-green, green, and red. The two largest character sizes yielded equivalent performance with any hue.

The other measures (i.e., time to any response and glance time for a correct response or a no response) were not shown because they yielded very similar results.

Brightness by hue interaction (B x H). The brightness by hue interaction was significant for four measures: both response time measures and both accuracy measures.

Figure 15 shows the percentage of correct responses as a function of dominant wavelength for each brightness level. Figure 16 shows the time to a correct response or no response for the same independent variables. Hue had an effect on accuracy only at low brightness whereas for the response time measure, hue had an effect at both brightness levels. There is little doubt that the differences noted among the hues are due to the 7 arcmin character size. High brightness yielded a significantly shorter response time for any hue. The same was observed for accuracy except for reddish orange and white.

One could wonder whether performance for a given hue was related to its luminance value (cd/m^2). Table 16 presents the Pearson correlation coefficients between luminance and two performance measures (accuracy and response time) for the low brightness condition. Correlations between these measures and ΔE distances in the Yu'v' space as well as in a Yu'v' space rescaled to better describe reading performance (Lippert, 1986; Lippert and Snyder, 1986) are also given. The results show that overall performance (i.e., all character sizes) correlates very well with luminance. The correlation coefficients are identical when ΔE distances from the standard Yu'v' space are considered instead of

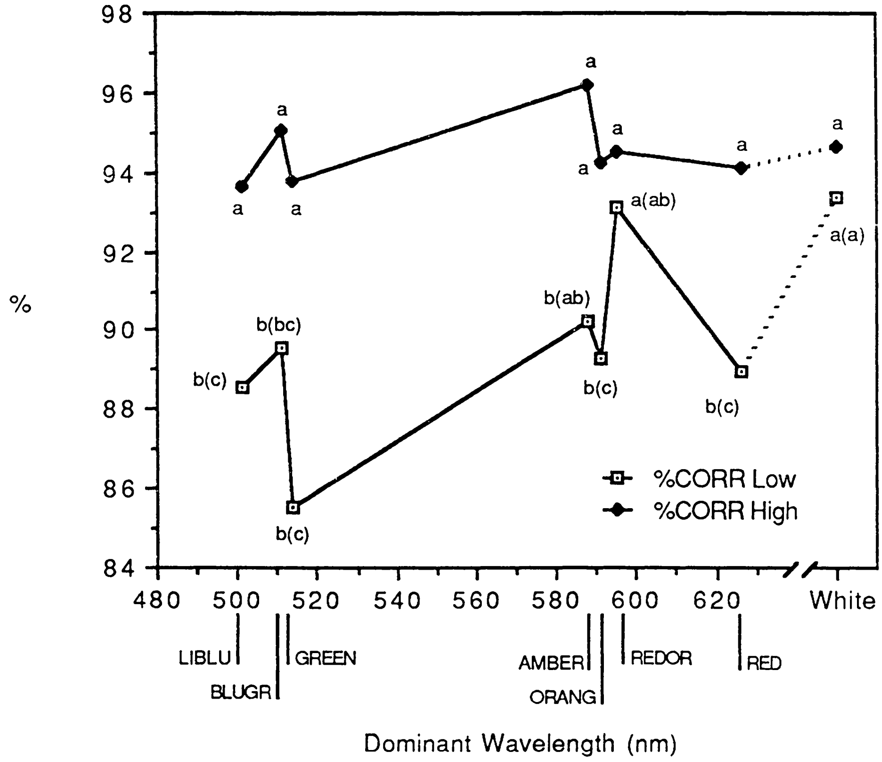


Figure 15. Percentage of correct responses (%CORR) as a function of dominant wavelength and brightness.

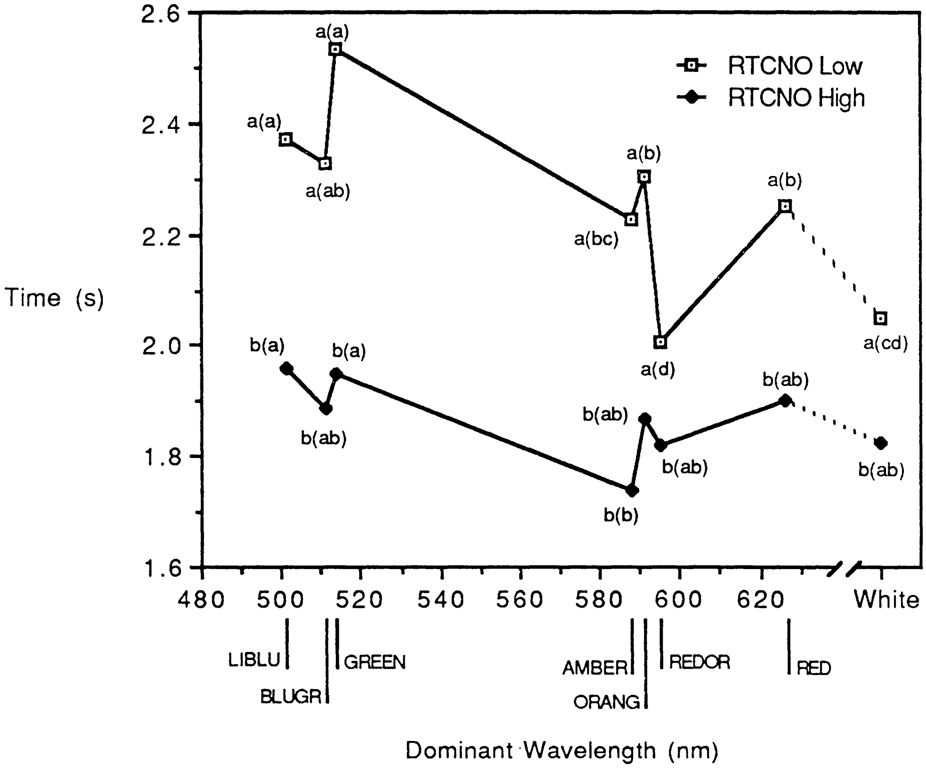


Figure 16. Time to a correct response or no response (RTCNO) as a function of dominant wavelength and brightness.

TABLE 16

Pearson Correlations between Luminance and Performance for the Low Brightness Condition

Measure	*Luminance* *Y* (cd/m^2)	ΔE *Yu'v'*	ΔE *rescaled* *Yu'v'*
%CORR (all character sizes)	0.904	0.904	-0.078
%CORR (7 and 11 arcmin)	0.907	0.907	-0.077
%CORR (17 and 25 arcmin)	-0.030	-0.030	-0.018
RTCNO (all character sizes)	-0.900	-0.901	-0.117
RTCNO (7 and 11 arcmin)	-0.904	-0.904	0.015
RTCNO (17 and 25 arcmin)	-0.533	-0.534	-0.223

Note: A correlation coefficient of at least ±0.705 is required for $p < 0.05$ and a correlation coefficient of at least ±0.818 is required for $p < 0.01$.

luminance. However, overall performance does not appear to correlate with ΔE distances from the rescaled Yu'v' space.

Performance associated with the smaller characters only (7 and 11 arcmin), which represents extreme performance, appears to correlate slightly better with luminance (or ΔE Yu'v') than overall performance. On the other hand, performance associated with the larger characters only (17 and 25 arcmin), which represents the other extreme in performance, does not correlate as well with luminance (or ΔE Yu'v'). Again, performance associated with either grouping of character sizes does not appear to correlate with ΔE distances from the rescaled Yu'v' space.

Table 17 presents the corresponding Pearson correlation coefficients for the high brightness condition. This time, only overall response time correlates with luminance (and ΔE Yu'v'). The correlation between performance for the smaller character sizes is relatively high but not significant. Performance associated with the larger character sizes does not correlate with luminance (or ΔE Yu'v'). Here again, performance with either grouping of character sizes does not appear to correlate with ΔE distances from the rescaled Yu'v' space. It is interesting to note that, although not all significant, the correlation coefficients are consistently smaller in the high brightness condition.

In the low brightness condition, performance associated with the two smaller character sizes correlates better with luminance (or ΔE Yu'v') than performance associated with the larger character sizes. Hence, it can be speculated that the differences in performance noted among the hues with the smaller characters, could be related to the difference in luminance among the hues rather than to a difference in their chromatic characteristics. This conclusion would be in agreement with the literature, that is, illumination color specifically has little effect on acuity or legibility of subjects with good corrected or uncorrected vision.

TABLE 17

Pearson Correlations between Luminance and Performance for the High Brightness Condition

Measure	*Luminance* *Y (cd/m^2)*	*ΔE* *Yu'v'*	*ΔE rescaled* *Yu'v'*
%CORR (all character sizes)	0.384	0.384	-0.105
%CORR (7 and 11 arcmin)	0.415	0.415	-0.147
%CORR (17 and 25 arcmin)	0.094	0.094	0.200
RTCNO (all character sizes)	-0.715	-0.715	0.000
RTCNO (7 and 11 arcmin)	-0.671	-0.671	-0.027
RTCNO (17 and 25 arcmin)	-0.326	-0.326	0.244

Note: A correlation coefficient of at least ±0.705 is required for $p < 0.05$ and a correlation coefficient of at least ±0.818 is required for $p < 0.01$.

Normally, as character size is decreased, more luminance is required to maintain legibility. However, if character size is decreased while a relatively low luminance level is kept constant (as in the low brightness condition of the experiment), the amount of light reaching the eye is reduced and the reading conditions gradually approach the threshold at which luminance contrast is important. Given the fact that, in the preliminary experiment a comfortable brightness level for each color was determined using a slide with a 15 arcmin character size, it is very likely that the viewing conditions were difficult with the smaller 7 and 11 arcmin character sizes in the low brightness condition of the experiment. Therefore, in such circumstances a slight increase in luminance should lead to a significant improvement in performance. In the experiment, the colors all had different luminances in the low brightness condition, and thus the luminance contrast ratios were different. The difference in performance among the colors observed for the smaller character sizes at low brightness in Figure 12 could then be explained by their different luminance contrast ratios which probably were in critical ranges, as the correlation coefficients tend to suggest.

For the larger character sizes, however, ample luminance was present in the low brightness condition since the luminance level had been selected for a smaller character size (15 arcmin) in the preliminary experiment. Owing to the large size of these letters, the amount of light reaching the eye was greater than with the 7 and 11 arcmin sizes. Hence, reading conditions were far above the threshold for the large characters and in such circumstances luminance is less important. Correlation between performance and luminance could thus be expected to be smaller as observed.

Although these speculations make sense, it must be kept in mind that they originate from correlations only, and they are impossible to prove from these data since luminance was not a systematically manipulated variable in this experiment.

It is interesting to note that a very high luminance value for reddish orange in the high brightness condition (Figure 10) did not yield a correspondingly higher mean accuracy and a shorter mean response time (Figures 15 and 16). The explanation provided earlier could explain this result which further supports the conclusion that the subjects, in the older age group at least, selected a fairly equivalent brightness among the hues in the preliminary experiment.

For amber, even if the difference in accuracy with any other color is not statistically significant in the high brightness condition, a slightly higher accuracy can still be observed in Figure 15. Also, the shortest response time in the high brightness condition was obtained with amber. The better overall performance obtained with amber could be explained by its higher luminance (reddish orange excluded). Therefore, the older subjects probably selected a slightly brighter level for amber compared to the other hues in the brightness matching operation of the preliminary experiment.

Hue main effect (H). The hue main effect was significant for four measures: both response time measures, glance time to a correct response, and percentage of correct responses.

Figure 17 shows both response time and glance time for a correct response or no response as a function of dominant wavelength. Figure 18 shows the percentage of correct responses as a function of dominant wavelength. There is a statistically significant difference in accuracy as well as in response time between green and both reddish orange and white. The maximum difference —between green and white— is 0.33 s for response time and 4.4% for percentage of correct responses. A similar result is observed for glance time: the maximum difference is 0.16 s and also statistically significant. There is little doubt that these differences are due to the 7 and 11 arcmin character sizes. Besides these

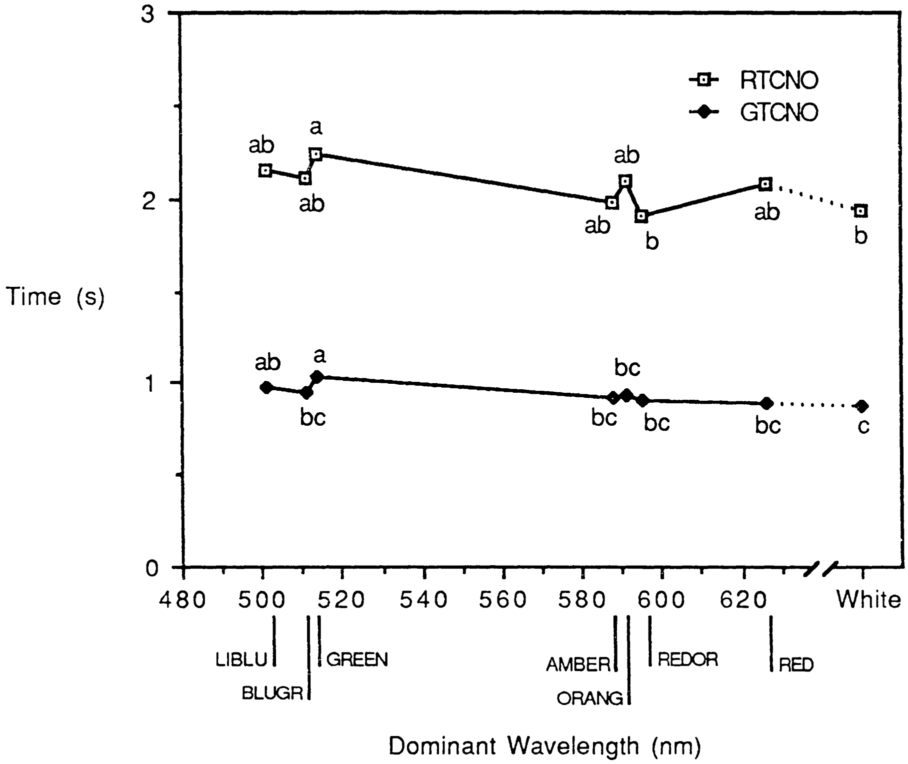


Figure 17. Time to a correct response or no response (RTCNO) and glance time for a correct response or a no response (GTCNO) as a function of dominant wavelength.

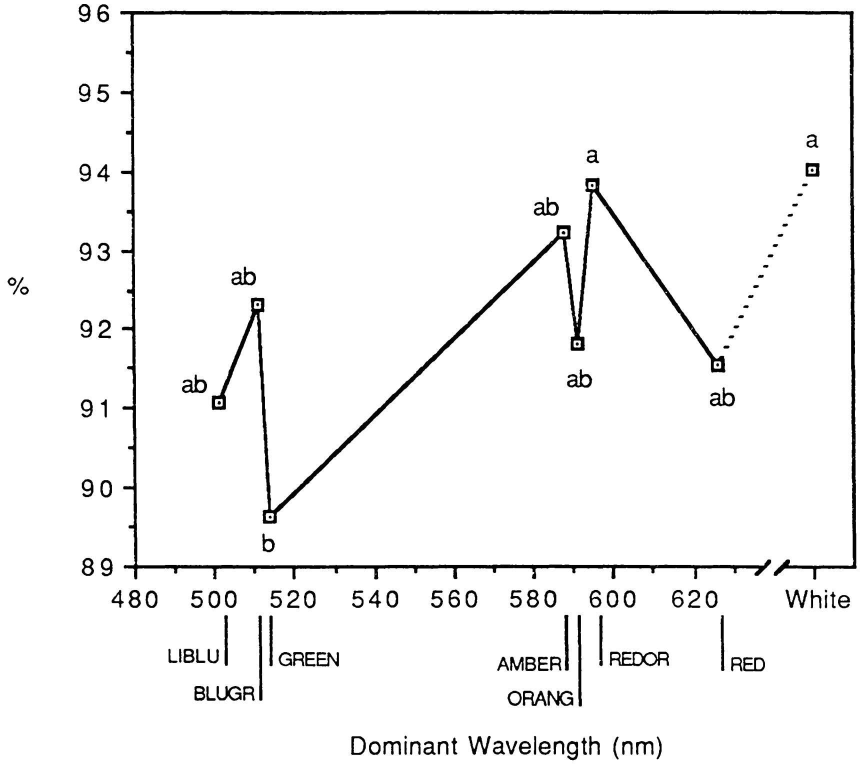


Figure 18. Percentage of correct responses (%CORR) as a function of dominant wavelength.

"extremes" —green and white— the differences are not significant among the other colors with respect to response time, glance time, and accuracy.

Concerning percentage of correct and incorrect responses, no significant difference was found among the hues. The fact that the *F* test was significant in the ANOVA but that the multiple-range test failed to find a difference among the means is unusual but possible. Indeed, the *F* test is testing whether any possible combination of the means differs from zero whereas the multiple-range test is testing whether very specific combinations of the means differ from zero.

It is worth noting that none of the interactions involving age and hue simultaneously were significant. This indicates that no significant difference in performance among the age groups was found with any color.

Brightness main effect (B). The brightness main effect was significant for all six measures. Lower brightness globally involved a longer mean glance time (i.e., 0.99 s vs 0.89 s) which translated into a longer mean response time (i.e., 2.26 s vs 1.87 s) and lower mean accuracy (i.e., 89.8% correct vs 94.5%, and 92.4% correct or incorrect vs 96.1%). Lower brightness also intruded more on the driving task as shown by the lane deviation measure (i.e., 0.091 m^2 vs 0.087 m^2).

Character size main effect (CS). The character size main effect was significant for all six measures. Figure 19 shows both measures of accuracy as a function of character size. The 7 arcmin character size yielded a significantly lower accuracy and a significantly higher no response rate. No difference was found among the three other character sizes.

Figure 20 shows both response time and glance time for a correct response or no response as a function of character size. The two smaller character sizes definitely yielded significantly degraded performance.

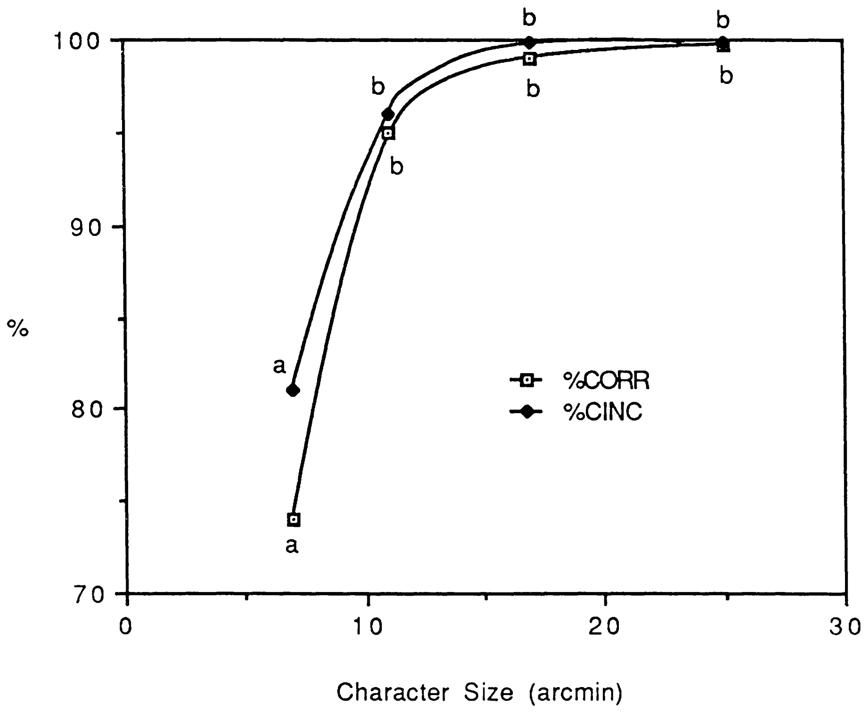


Figure 19. Percentage of correct responses (%CORR) and percentage of correct and incorrect responses (%CINC) as a function of character size.

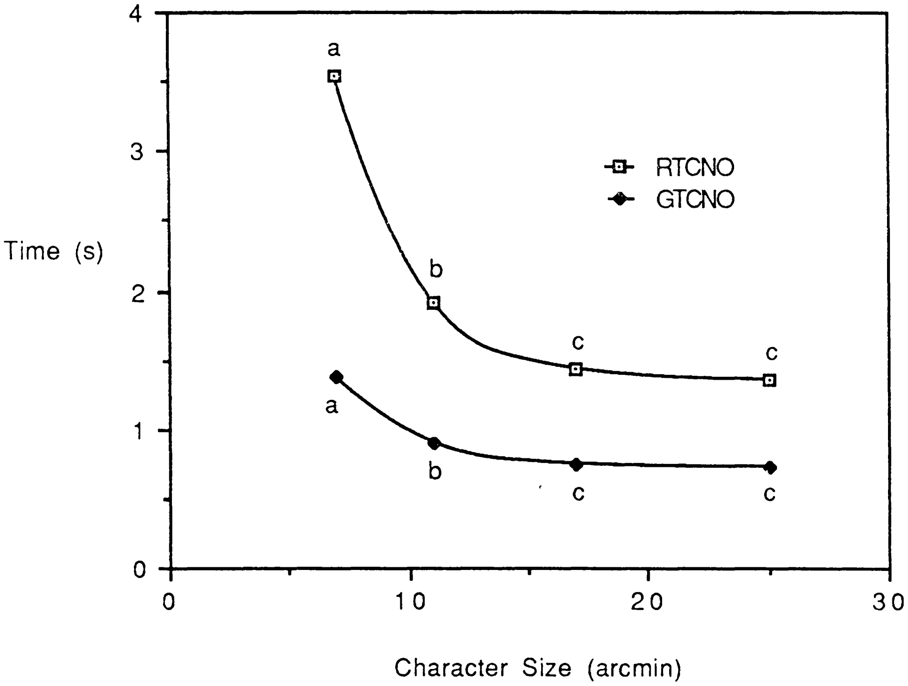


Figure 20. Time to a correct response or no response (RTCNO) and glance time for a correct response or a no response (GTCNO) as a function of character size.

Figure 21 shows the lane deviation measure as a function of character size. This figure clearly shows that the 7 arcmin character size had a significant overall degrading effect on driving behavior.

These figures show that the 11 arcmin character size did require more time to be read and yielded longer mean response time than the larger character sizes but the consequences on accuracy and driving behavior although observable were not significant.

Brightness by character size by age interaction (B x CS x A). The brightness by character size by age interaction was significant for all six measures. Figure 22 shows the percentage of correct responses per age, character size, and brightness. Accuracy was heavily dependent on character size and brightness for the older subjects; it varied from 35% for the smallest letters at low brightness to 99% for the two larger character sizes at either brightness level. The middle age drivers showed degraded performance only with the smallest letters at both brightness levels. Their performance was comparable to that of the younger age group in any character size by brightness condition except with the 7 arcmin at low brightness. Comparing with the 17 and 25 arcmin sizes for which best performance was achieved, the younger drivers showed significantly lower performance only with the 7 arcmin characters at low brightness.

When compared to the younger and middle age groups, the older drivers showed significantly degraded performance with the two smaller character sizes at any brightness level whereas with the two larger sizes, their performance was equivalent.

Figure 23 shows the percentage of correct and incorrect responses for the same independent variables. The curves show the same general patterns observed previously. The proportion of no responses varied significantly as a function of character size and brightness for the older drivers with values ranging from about 60% with the 7 arcmin character size at low brightness to 0% for the larger characters. The middle age group had a

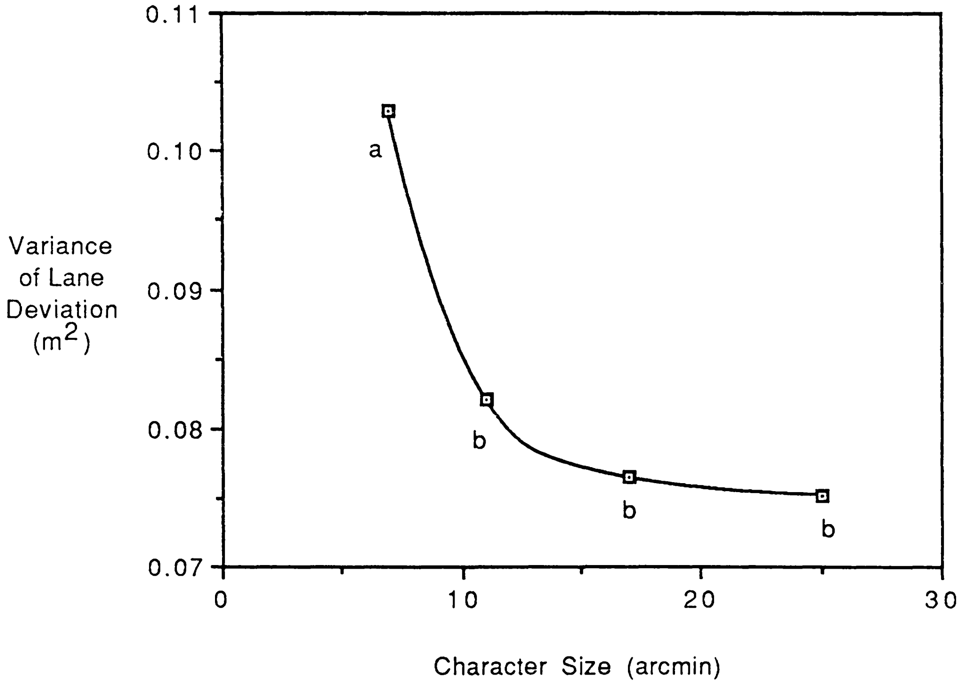


Figure 21. Variance of lane deviation regardless of response (LDANY) as a function of character size.

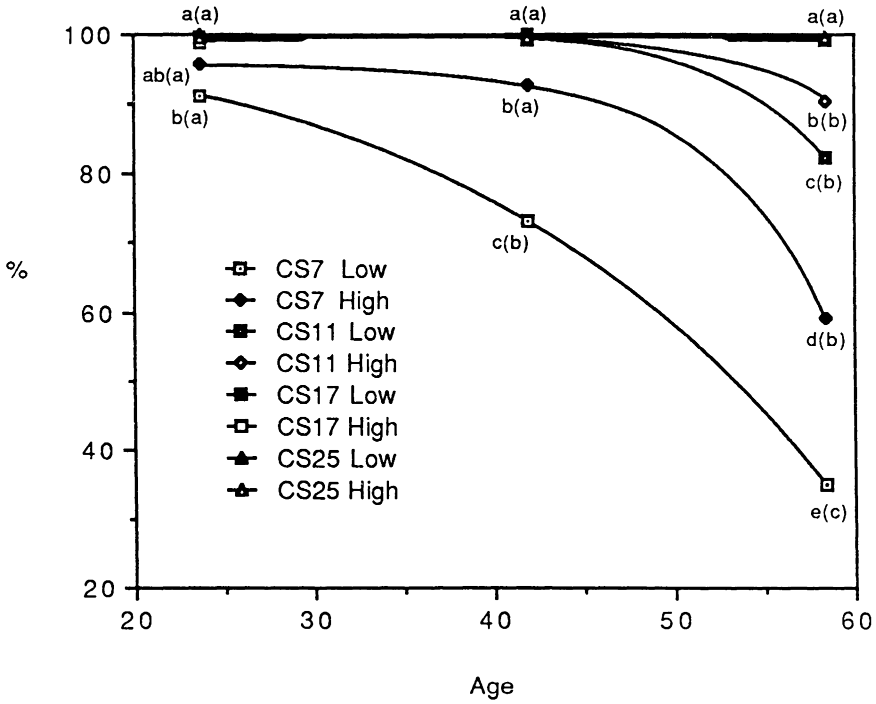


Figure 22. Percentage of correct responses (%CORR) per brightness, character size, and age.

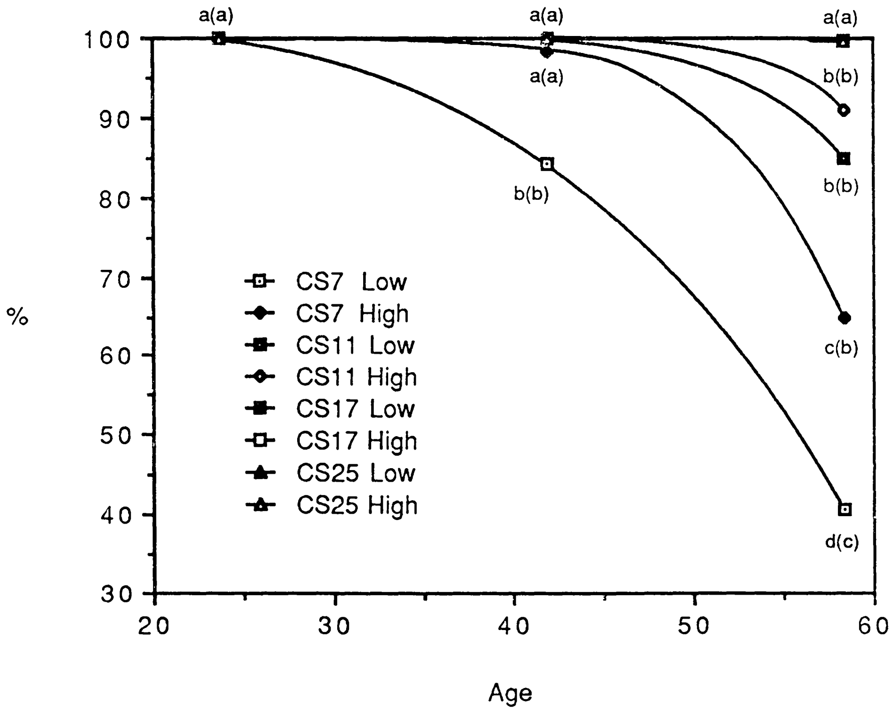


Figure 23. Percentage of correct and incorrect responses (%CINC) per brightness, character size, and age.

proportion of no responses significantly different from zero only with the 7 arcmin character size at low brightness (i.e., about 15%).

The proportion of incorrect responses (i.e., difference between both accuracy measures) varied from about 6% in the worst conditions (i.e., small letters at low brightness for older drivers) to 0% in the best conditions (i.e., larger letters for the younger and middle age groups) with an overall average of about 2.2% for all subjects in all conditions.

Figure 24, shows the average time to a correct response or no response *(RTCNO)* as a function of age, character size, and brightness. The results for the mean time to a correct, incorrect, or no response *(RTANY)* will not be shown because they almost exactly duplicate those observed for *RTCNO*. The correlation coefficient between these two measures is better than 0.9913 in every age group; both measures give virtually the same information as a result of the very small proportion of incorrect responses.

Figure 24 is almost an exact inverse of Figure 22. The difference observed in response time between the older and middle age groups is larger than between the middle and younger age groups with both the 7 and 11 arcmin character sizes at any brightness. Time to a correct response was dependent on brightness for the 7 arcmin character size; the improvement in performance observed is significant for every age group at high brightness. A similar improvement was present with the 11 arcmin size only for the older drivers.

Middle age and younger drivers showed fairly equivalent response times with the 11, 17, and 25 arcmin character sizes at any brightness. The older drivers, however, still showed a longer mean response time with the 17 arcmin character size at low brightness when compared with the middle age group. The 17 arcmin character size at high brightness as well as the 25 arcmin at both brightness levels yielded best and equivalent performance in all age groups.

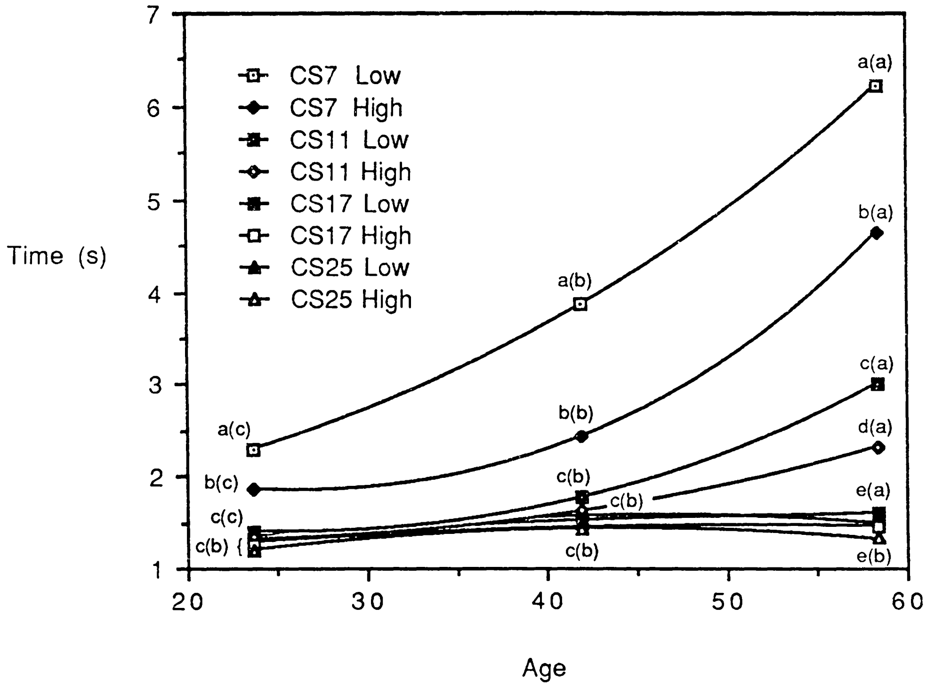


Figure 24. Time to a correct response or no response (RTCNO) per brightness, character size, and age.

Figure 25 shows the glance time for the same independent variables. The general curve patterns resemble those observed in the previous figure except that with glance time, the difference among the age groups is significant with any character size. Also, it is surprising to note that glance time for the most difficult condition (i.e., 7 arcmin letters at low brightness) was significantly shorter for the older drivers than for their middle age counterparts.

If a subject were to take more time before giving a response because the word displayed is more difficult to read, he or she should glance longer at the display. This consequence should have a detrimental effect on the driving behavior; that is, variance of lane deviation should increase proportionally to glance time. Hence, response time, glance time, and variance of lane deviation should all show similar patterns. However, the result observed in Figure 25 for the 7 arcmin character size at low brightness is not consistent with this assumption; the older subjects who had a higher proportion of no responses and a longer average time to a correct response than the other groups in this treatment combination did show an average glance time that was significantly longer than that of the younger age group but shorter than that of the middle age group. Furthermore, with the 7 arcmin character size at high brightness, older drivers had a significantly longer mean glance time than middle age drivers, but the difference was not as large as could have been expected from examination of time to correct response curves (i.e., Figure 24).

This unexpected result can be explained by a behavior that was typical of some older subjects during the experiment; several of these subjects "gave up" when the character size was too small, especially at low brightness. The consequences of such a behavior are:

- no response recorded at all and hence, a response time of eight seconds was substituted for the missing data,

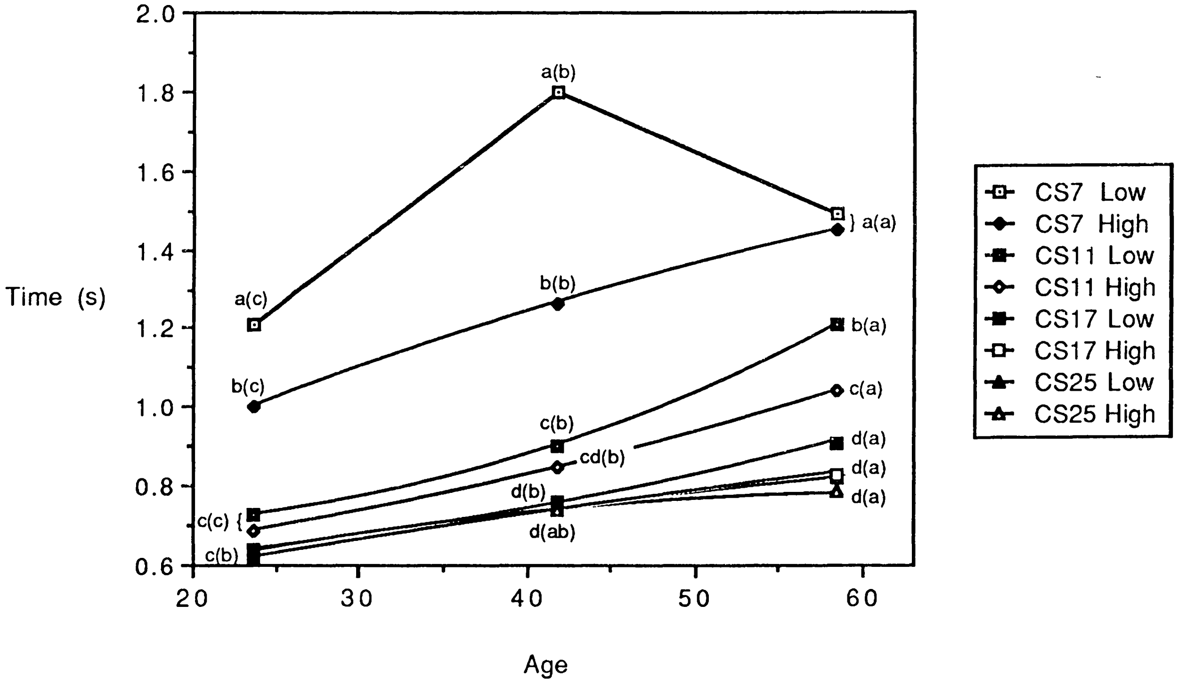


Figure 25. Glance time for a correct response or a no response (GTCNO) per brightness, character size, and age.

- very short glance time since the subject glanced at the display and determined that the task was too difficult to perform, and
- small variance of lane deviation since the subject's attention was redirected to the driving task earlier than usual.

Figure 26 shows the variance of lane deviation for the same independent variables. Again, the 7 arcmin character size yielded the lowest performance in every age group. A surprising result is that the older drivers consistently showed smaller variance of lane deviation than their middle age counterparts with every character size by brightness treatment combination except with the 11 arcmin character size at low brightness. Also, as in Figure 25 the variance of lane deviation observed for the older drivers with the 7 arcmin characters size at both brightness levels was lower than expected particularly at low brightness. This result is consistent with the particular behavior of some of the older drivers. Finally, the middle age group showed an unexpectedly high variance of lane deviation for the 25 arcmin character size at high brightness.

The fact that older drivers often ignored the task when it appeared too difficult is a consciously adopted strategy that cannot be ignored and which is manifested in the actual results. Hence, this fact must be kept in mind when interpreting the performance observed for the older drivers specifically with the 7 arcmin character size at both brightness levels as well as with the 11 arcmin character size at low brightness.

Character size by age interaction (CS x A). The character size by age interaction was significant for both measures of response time and both measures of accuracy. Figure 27 shows the percentage of correct responses as a function of age and character size. Figure 28 shows the percentage of correct and incorrect responses for the same independent variables. The older drivers definitely had a lower reading accuracy as well as

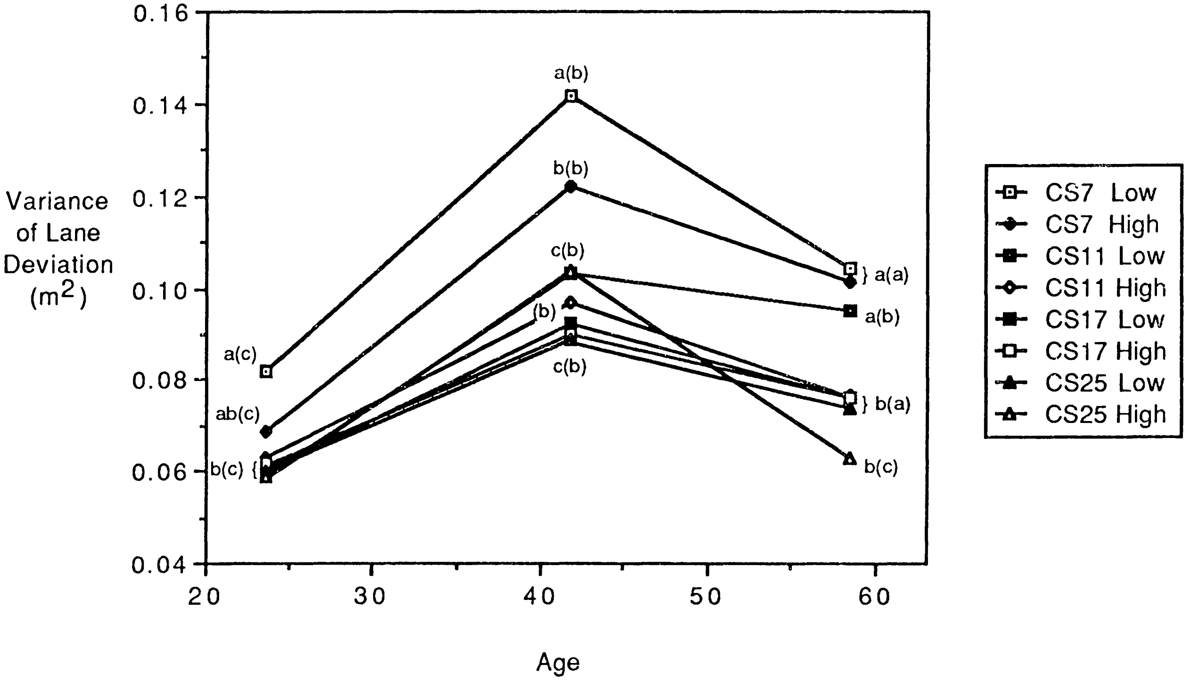


Figure 26. Variance of lane deviation regardless of response (LDANY) per brightness, character, and age.

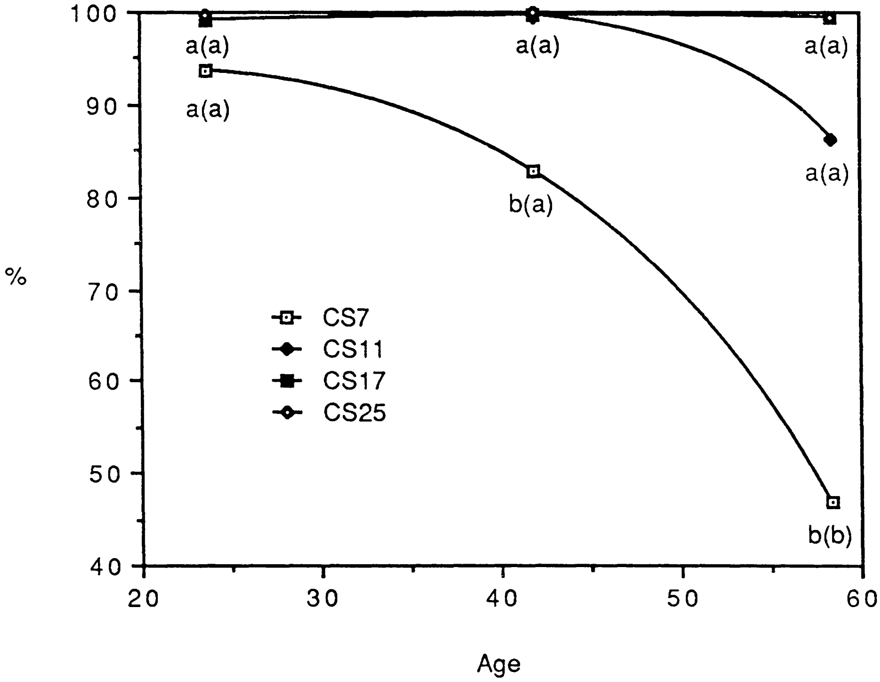


Figure 27. Percentage of correct responses (%CORR) per character size and age.

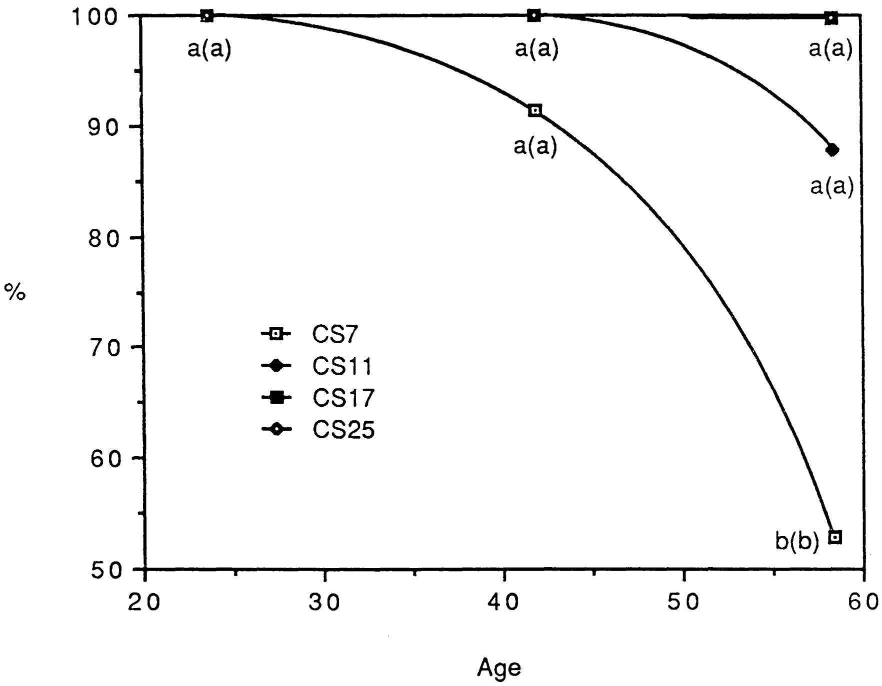


Figure 28. Percentage of correct and incorrect responses (%CINC) per character size and age.

a higher proportion of no responses than the other age groups with the 7 arcmin character size. No other difference in accuracy was observed among the character sizes and the age groups.

Figure 29 shows the time to a correct response as a function of character size and age. Again, the same patterns as before can be observed: the 7 arcmin character size generated a significantly longer mean response time for the older and middle age group drivers when comparing with the larger sizes. Mean response time for the three larger character sizes was not statistically different except for the older drivers who still showed degraded performance with the 11 arcmin character size.

The results concerning character size found thus far are in general agreement with the literature on visual displays. For instance, the Human Factors Society (1986) notes that threshold character height for rapid and accurate legibility of single characters is 11 to 12 arcmin depending on resolution and character definition. In this experiment, the results associated with the 7 arcmin character size clearly demonstrate that it is below that threshold if accurate legibility can be defined as 95% or better accuracy. On the other hand, the 11 arcmin character size yielded performance, at any brightness level, that was equivalent to the 17 and 25 arcmin sizes (which both yielded best performance) for both the younger and middle age group drivers. For the older drivers, however, degradation in performance could be observed with the 11 arcmin character size especially at low brightness (Figures 22, 23, 24, 25, and 29). Hence, considering the virtually ideal resolution and character definition for all character sizes in this experiment (contrast ratios of 330:1 to 9860:1) in conjunction with the performance of older drivers, the 11 arcmin character size appears to be below the threshold height for rapid and accurate legibility for this category of drivers. This conclusion can be largely explained by the fact that most of them did not have the accommodation power necessary to read such small letters and thus had to rely on their

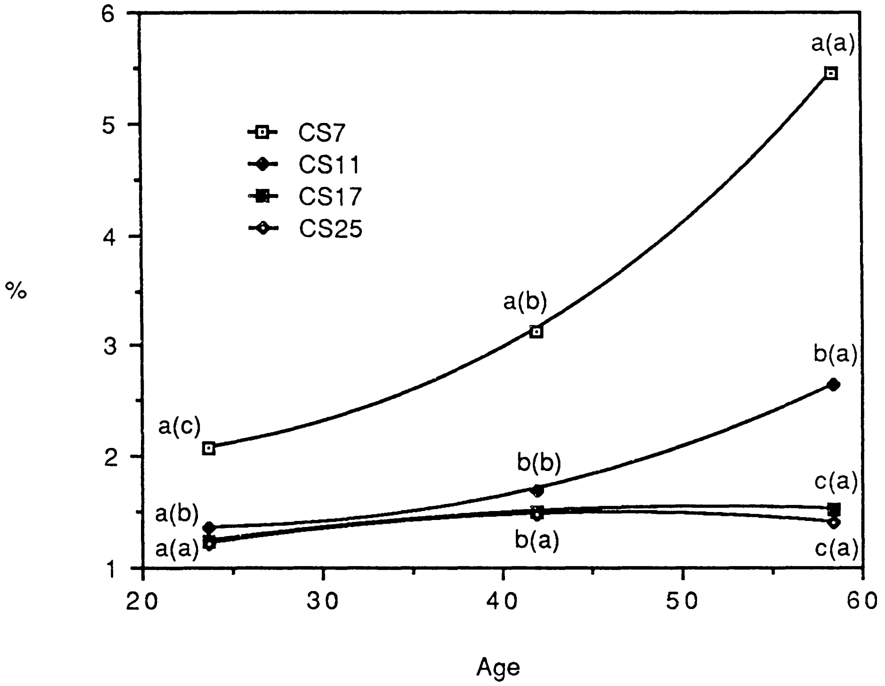


Figure 29. Time to a correct response or no response (RTCNO) per character size and age.

bifocals, which all of them wore. However, the viewing distance for the words, which was representative of legend viewing distances found in actual automobiles, was between their best focus distances so that they had to lean to read the words presented with the 11 arcmin character size (and with the 7 arcmin for that matter).

The poorer performance observed for the older drivers with the 11 arcmin character size is consistent with the findings of Mourant and Langolf (1976). Using a 95% recognition performance criterion, they found that a character size subtending 10 arcmin at the eye was too small for older drivers at any luminance level and contrast ratio they tested. Thus, it is very likely that even more important decrements in performance would have been observed for the older drivers with the 11 arcmin character size at lower luminance contrast ratios.

The Human Factors Society (1986) notes that given good resolution and character definition, the threshold character height for comfortable reading is between 16 and 18 arcmin. The results obtained for the younger and middle age group drivers with the 17 arcmin character size agree very well with this criterion since best performance was observed in these age group with this size (as well as with the 25 arcmin size). However, given the significantly longer mean glance and response times of older drivers with the 17 arcmin at low brightness when compared with the middle age group drivers (Figure 24, and 25), 17 arcmin is probably close to the threshold character size for rapid and accurate but not yet comfortable reading of high resolution and high character definition automobile legends for older drivers.

The Human Factors Society (1986) notes that the preferred character size for both readability and legibility is between 20 and 22 arcmin. Mourant and Langolf (1976) recommend, however, that 27 arcmin should be used for automobile instrument panel legends because it was found to be easier to read than 18 arcmin for older people given luminance contrast ratios lower or equal to 25:1. In general, legibility increases as character size increases (Snyder and Maddox, 1980; Easterby and Zwaga, 1984). Hence, in light of these results and given an equivalent mean glance time between older and middle age drivers with the 25 arcmin character size, 25 arcmin is probably near the threshold character size for comfortable reading of automobile legends for older drivers.

Brightness by age (B x A) and brightness by character size (B x CS) interactions. The brightness by age interaction was significant for all measures but variance of lane deviation, while the brightness by character size interaction was significant for all six measures. These interactions were discussed implicitly in the earlier section on the brightness by character size by age interaction, and since they do not include any new information, they will not be discussed further.

Age main effect (A). The age main effect was significant for all six measures. Figure 30 shows both measures of accuracy as a function of age. Both measures of accuracy display identical patterns. Older drivers had significantly lower reading accuracy and a higher no response rate than other drivers.

Figure 31 shows the mean time to a correct response as well as the mean glance time for correct responses as a function of age. The curves are consistent with the previous observations. Figure 32 shows the lane deviation measure as a function of age. Since these dependent measures now represent averages over all four character sizes, caution must be exercised in their interpretation.

In Figure 32, older drivers could normally have been expected to show a larger mean variance of lane deviation value than their middle age counterparts instead of the observed lower value since the difference between these two age groups can probably not be all accounted for only by the inclusion of the data associated with the no responses

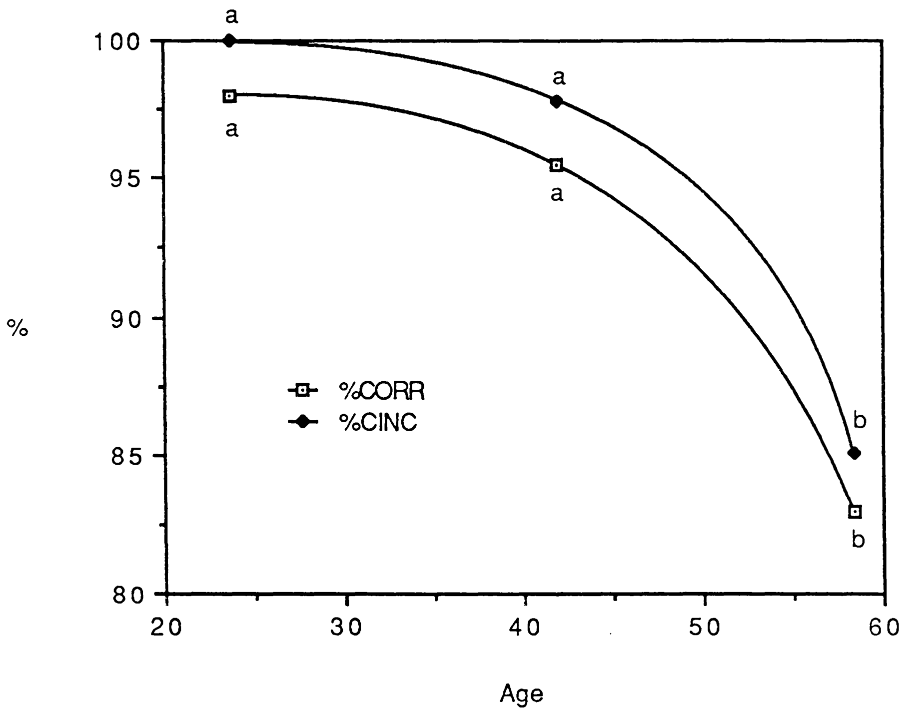


Figure 30. Percentage of correct responses (%CORR) and percentage of correct and incorrect responses (%CINC) as a function of age.

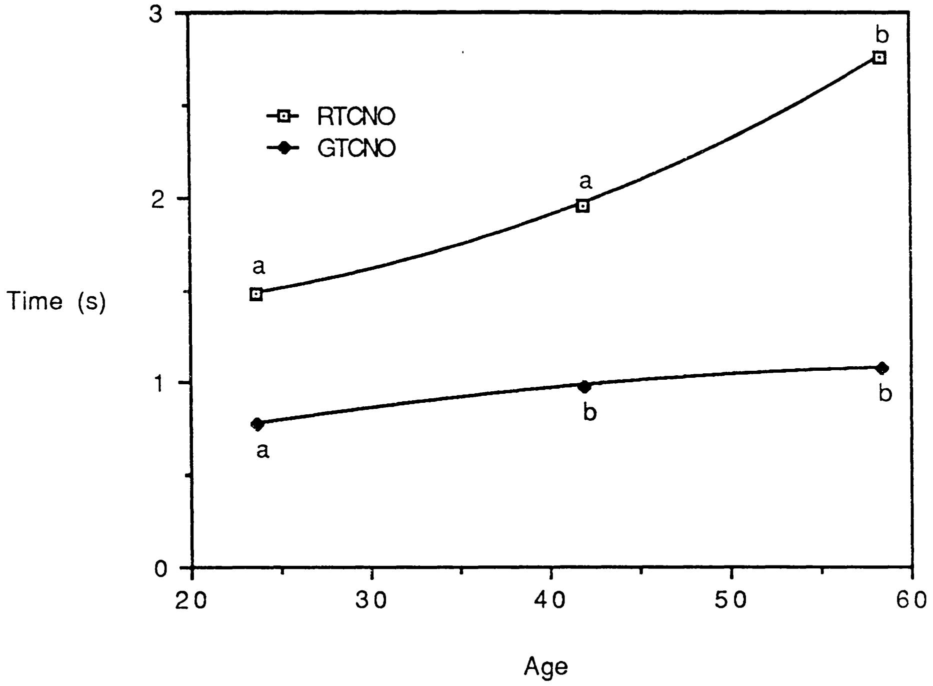


Figure 31. Time to a correct response or no response (RTCNO) and glance time for a correct response or a no response (GTCNO) as a function of age.

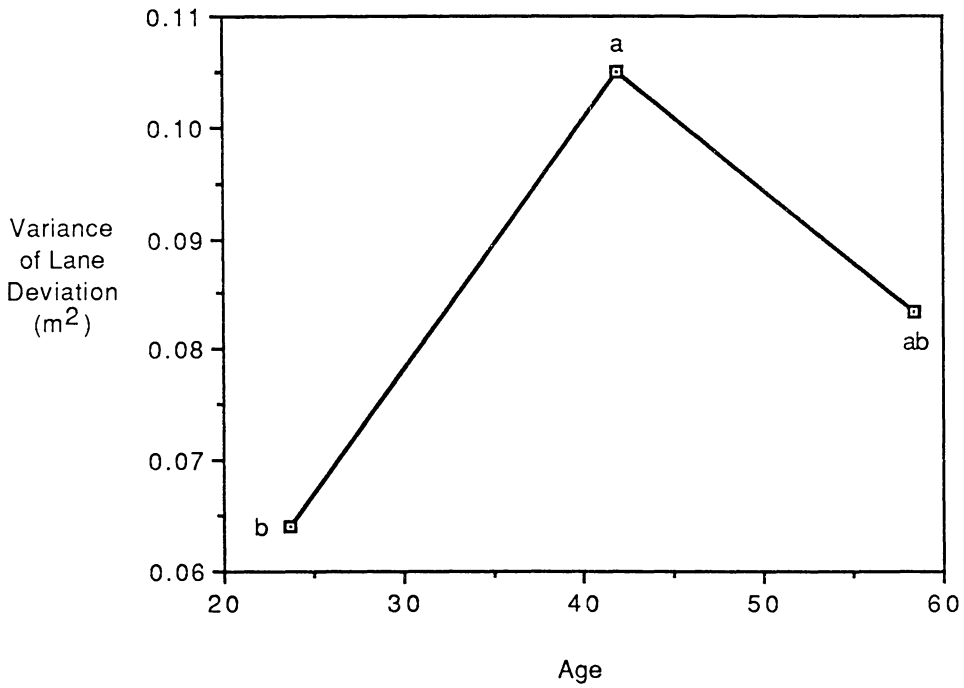


Figure 32. Variance of lane deviation regardless of response (LDANY) as a function of age.

(11.5% of the data for this group). Thus, the older drivers most probably had a more stable overall driving behavior than the middle age group in the experiment.

This more stable driving behavior can be explained by the fact that some of the older subjects were incapable of keeping a constant speed between 55 and 60 mph as instructed, and actually drove a little slower (i.e., 50 to 55 mph). Since the effect of the simulated crosswinds on the vehicle was proportional to the speed of the vehicle, subjects driving slower experienced a reduced crosswind effect compared to subjects who drove at the normal speed. A reduced crosswind effect results in a less disturbed driving behavior and hence, a smaller mean value of variance of lane deviation.

The relatively small variance of lane deviation value observed for the older age group when compared to the middle age group most probably reflects the two following behavioral strategies: they ignored the small character sizes and drove slower.

Word complexity by character size interaction (WC x CS). The word complexity by character size interaction was significant for three measures only: time to any response, glance time to a correct response, and percentage of correct and incorrect responses. Figure 33 shows the percentage of correct and incorrect responses as a function of character size and word complexity. Word complexity appears to have played a role only with the 7 arcmin character size; that is, high complexity words (or "difficult" words) yielded a significantly higher proportion of no responses than low complexity words (or "easy" words).

Figure 34 shows the mean time to any response as well as the mean glance time to a correct response or a no response as a function of character size. High complexity words required a significantly longer mean glance time than low complexity words with every character size. The maximum difference is on the order of 0.1 s. For response time, a

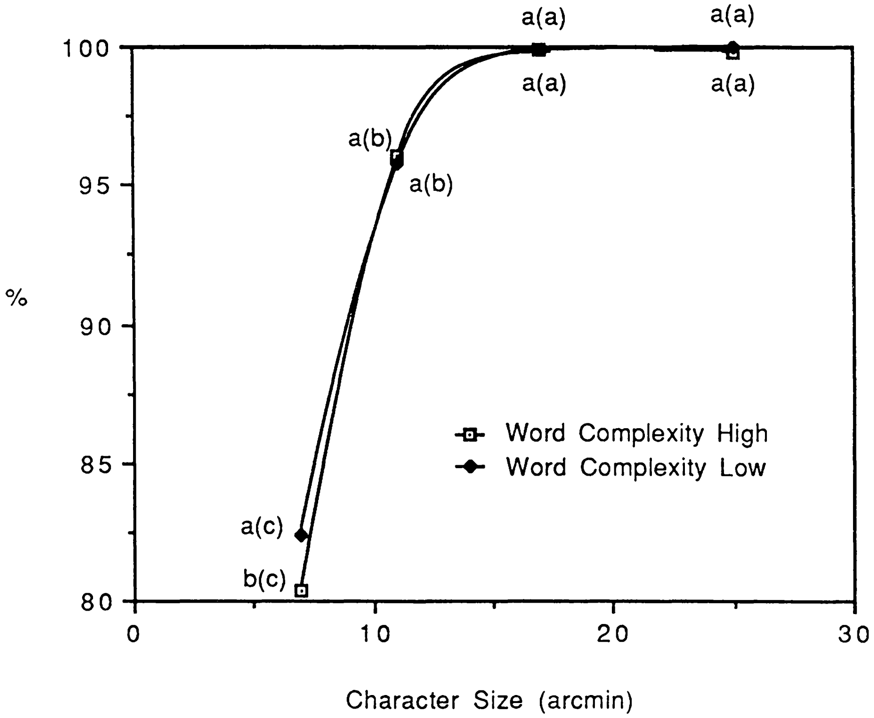


Figure 33. Percentage of correct and incorrect responses (%CINC) as a function of character size and word complexity.

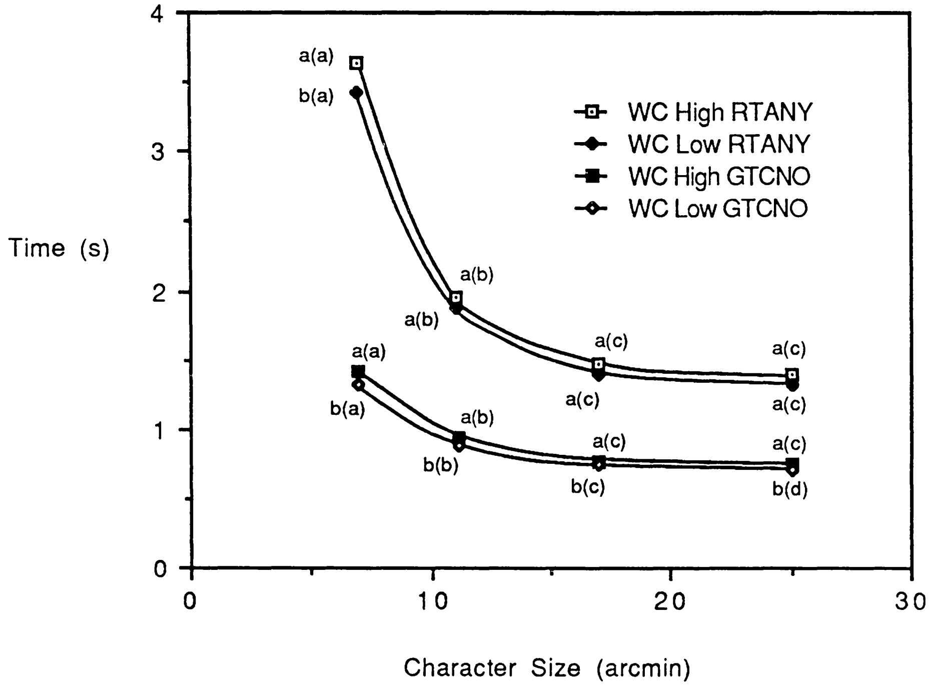


Figure 34. Response time regardless of response (RTANY) and glance time for a correct response or a no response (GTCNO) as a function of character size and word complexity (WC).

significant difference between high and low complexity words was observed for the 7 arcmin character size only. Such results were rather expected.

Word complexity main effect (WC). The word complexity main effect was significant for three of the measures: time to a correct response (or no response), time to any response, and glance time at the displays. Higher complexity involved a longer mean glance time (i.e., 0.96 s vs 0.91 s), which translated into a longer mean response time (i.e., 2.11 s vs 2.02 s).

One reason why word complexity did not yield greater differences in the measures is probably because both categories of words were too similar to each other; i.e., the range of frequency in the English language of the words selected were probably not different enough. It is likely that if very rare words had been used in the high word complexity category, greater differences in the dependent measures between word complexity categories would have been observed.

Main Experiment: Subjective Measures

The three subjective measures (attractiveness, comfort, and ease of readability) did not involve the same set of independent variables. Attractiveness and comfort were measured for all hue by brightness treatment combinations only while ease of readability was measured for every hue by brightness by character size treatment combination. Therefore, two sets of analyses were performed.

The subjective measures, which were judgments made by subjects on a seven-point Likert scale, were assumed to consist of data measured on an interval scale in the analysis. It is often reasonable to consider Likert scales as interval scales and thus the more powerful parametric methods can be used for statistical analysis.

Attractiveness and comfort: MANOVA. A multivariate analysis of variance (MANOVA) was performed on both attractiveness and comfort (Table 18). Conversion of Wilk's Lambda to an F ratio was used as the decision criterion. Only the brightness and hue main effects were significant for the these measures. Again, the gender effect did not result in significance.

Attractiveness and comfort: ANOVAs and post hoc tests. Both main effects were then tested in analyses of variance of each dependent variable. The results are given in Table 19.

Comparisons of means were then performed on the hue main effect. The Ryan-Einot-Gabriel-Welsh multiple-range method (SAS, 1985) was used again. Figure 35 shows the average attractiveness rating as a function of hue. Light blue was judged least attractive while orange, reddish orange, white, and amber were judged most attractive. Red, blue-green, and green fell somewhere in between. In regard to comfort, the multiple-range test failed to find any significant difference among the colors.

High brightness rated higher than low brightness in both measures: 1.36 versus 0.79 for attractiveness, and 1.51 versus 0.71 for comfort.

None of the interactions involving age was significant which indicates that all subjects had similar preferences in terms of attractiveness and comfort. Such a result was unexpected since according to the literature the younger subjects should have been expected to find lower brightness more comfortable than high brightness (Mourant and Langolf, 1976). Indeed, the younger drivers selected subjectively comfortable luminance levels for night driving that were significantly lower than those selected by the older drivers in the preliminary experiment. Hence, the fact that younger subjects found the higher luminance levels more comfortable is rather surprising.

TABLE 18

Multivariate Analysis of Variance for Attractiveness and Comfort

Effect	*df*	*F ratio*	*p*	
Between-Subjects				
A (Age)	2	0.4	0.7943	
G (Gender)	1	0.09	0.9111	
A x G	2	0.75	0.5667	
S/AG	18			
Within-Subjects				
H (Hue)	7	2.14	0.0105	*
H x A	14	1.09	0.3484	
H x G	7	1.11	0.3472	
H x A x G	14	1.49	0.0578	
H x S/AG	126			
B (Brightness)	1	17.77	0.0001	*
B x A	2	0.13	0.9692	
B x G	1	1.01	0.3860	
B x A x G	2	0.15	0.9639	
B x S/AG	18			
B x H	7	1.44	0.1333	
B x H x A	14	0.93	0.5686	
B x H x G	7	1.18	0.2944	
B x H x A x G	14	1.10	0.3328	
B x H x S/AG	126			

TABLE 19

Summary of the Effect of Hue *(H)* and Brightness *(B)* on the Dependent Measures

Effect: *Dependent Variable*	*MS*	*F ratio*	*p*	
Hue[1]:				
Attractiveness	5.970	2.69	0.0124	*
Comfort	5.004	2.32	0.0296	*
Brightness[2]:				
Attractiveness	31.510	28.45	0.0001	*
Comfort	61.760	37.49	0.0001	*

[1] The Hue main effect has 7 *df*, the error *(H x S/AG)* has 126 *df*.

[2] The Brightness main effect has 1 *df*, the error *(B x S/AG)* has 18 *df*.

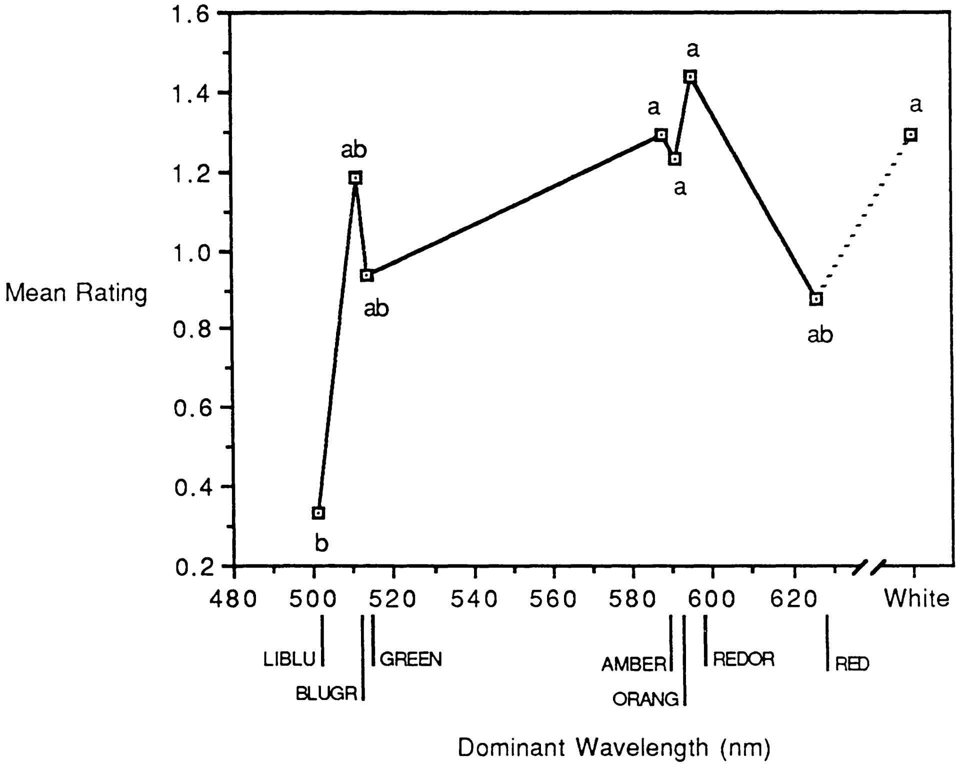


Figure 35. Mean attractiveness rating as a function of dominant wavelength.

Ease of readability: ANOVA and post hoc tests. The results of an analysis of variance on ease of readability are shown in Table 20. The three main effects —character size, hue, and brightness— were significant as well as the following interactions: hue by character size *(H x CS)*, brightness by character size *(B x CS)*, and character size by hue by brightness by age *(CS x H x B x A)*. The latter interaction will be disregarded due to the usual difficulties in interpreting high order interactions.

Hue by character size interaction (H x CS). Figure 36 shows the ease of readability rating as a function of hue and character size. As expected, the two larger character sizes were rated significantly higher than the smaller ones. The difference is significant among all sizes at any hue except between the 17 and 25 arcmin for both reddish orange and white. A difference among the hues was observed for all character size but the largest.

Table 21 presents the correlations between luminance and ease of readability ratings obtained for each hue by character size treatment combination. The correlations are highest for the 7 and 11 arcmin character sizes at low brightness. This could suggest that for these character sizes, luminance may have had an effect on the subjective ease of readability that was stronger than hue; that is, the subjects could have based their judgment (or been influenced) more on the reading conditions associated with the smaller character sizes, rather than uniformly among all character sizes.

Hue main effect (H). Figure 37 shows the ease of readability rating as a function of hue. A statistically significant difference could be observed among the hues; however, one must consider that the difference between the extremes, white (the highest) and light blue (the lowest), was on the order of one half of one rating unit.

TABLE 20

Analysis of Variance for Ease of Readability

Effect	*df*	*MS*	*F ratio*	*p*	
Between-Subjects:					
A (Age)	2	13.82	0.75	0.4862	
G (Gender)	1	11.69	0.63	0.4359	
A x G	2	22.21	1.21	0.3223	
S/AG	18				
Within-Subjects:					
CS (Character Size)	3	1813.09	270.34	0.0001	*
CS x A	6	10.12	1.51	0.1927	
CS x G	3	5.56	0.83	0.4833	
CS x A x G	6	11.33	1.69	0.1412	
CS x S/AG	54				
H (Hue)	7	7.37	4.85	0.0001	*
H x A	14	1.85	1.22	0.2710	
H x G	7	2.69	1.77	0.0990	
H x A x G	14	1.25	0.83	0.6399	
H x S/AG	126				
B (Brightness)	1	78.84	31.61	0.0001	*
B x A	2	1.81	0.72	0.4983	
B x G	1	4.38	1.76	0.2018	
B x A x G	2	0.20	0.08	0.9248	
B x S/AG	18				

TABLE 20 cont'd

Effect	*df*	*MS*	*F ratio*	*p*	
B x CS	3	4.74	6.02	0.0013	*
B x CS x A	6	0.41	0.35	0.9074	
B x CS x G	3	0.22	0.28	0.8417	
B x CS x A x G	6	0.71	0.90	0.4993	
B x CS x S/AG	54				
B x H	7	1.28	1.86	0.0821	
B x H x A	14	0.96	1.40	0.1613	
B x H x G	7	0.57	0.83	0.5652	
B x H x A x G	14	1.06	1.53	0.1099	
B x H x S/AG	126				
CS x H x B	21	0.28	0.95	0.5291	
CS x H x B x A	42	0.44	1.48	0.0328	*
CS x H x B x G	21	0.24	0.81	0.7040	
CS x H x B x A x G	42	0.29	0.97	0.5210	
CS x H x B x S/AG	378				
H x CS	21	1.14	2.40	0.0006	*
H x CS x A	42	0.43	0.90	0.6442	
H x CS x G	21	0.56	1.16	0.2834	
H x CS x A x G	42	0.29	0.60	0.9767	
H x CS x S/AG	378				

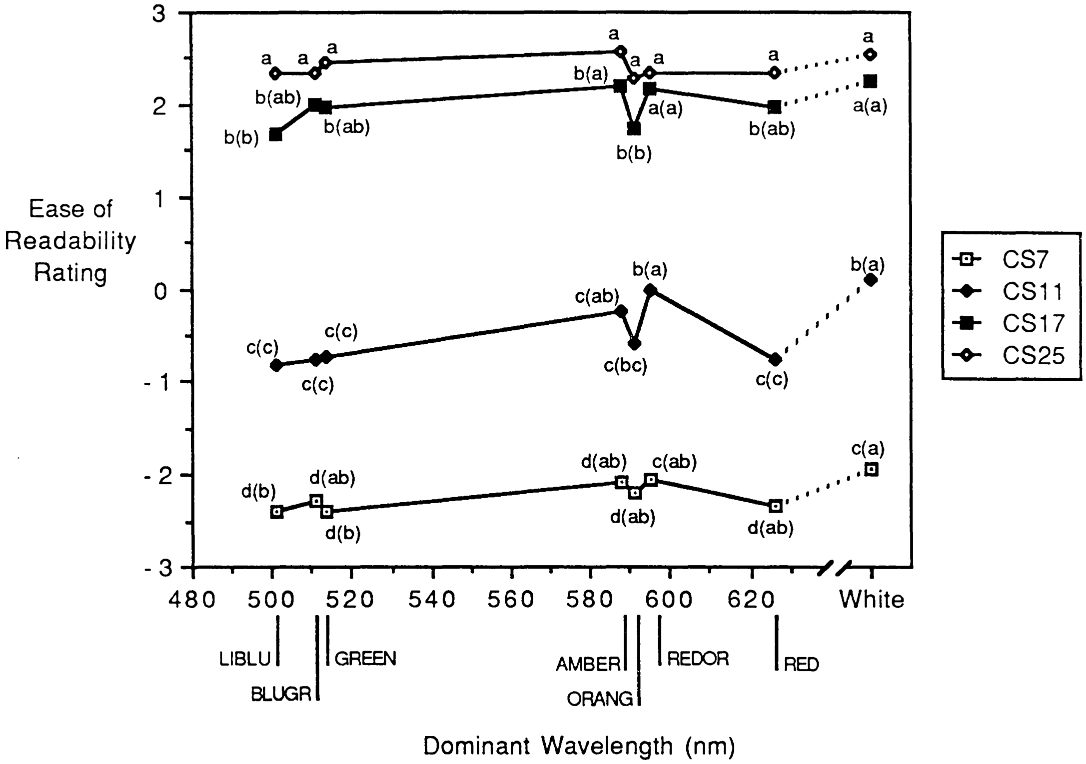


Figure 36. Mean ease of readability rating as a function of dominant wavelength and character size.

TABLE 21

Pearson Correlations Between Ease of Readability Ratings and Luminance Values per Character Size and Brightness Level

Brightness	*Character Size (arcmin)*			
	7	*11*	*17*	*25*
Low	0.960	0.952	0.696	0.452
High	0.765	0.778	0.490	-0.039

Note: a correlation coefficient of at least ±0.705 is required for $p < 0.05$ and a correlation coefficient of at least ±0.818 is required for $p < 0.01$.

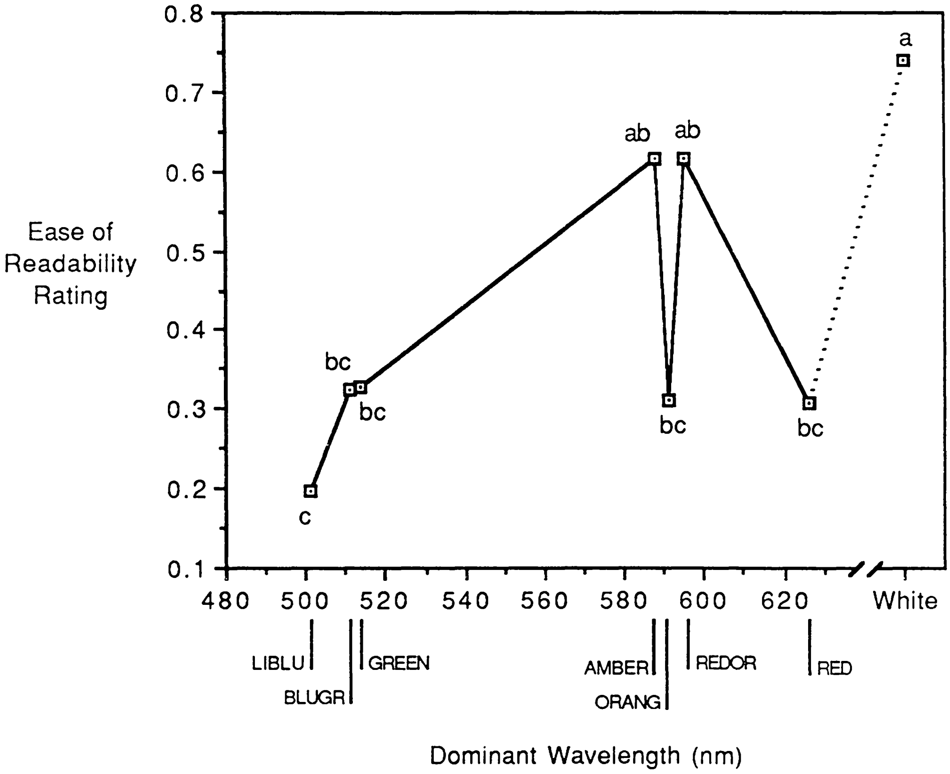


Figure 37. Mean ease of readability rating as a function of dominant wavelength.

Character size main effect (CS). Figure 38 shows the mean ease of readability rating as a function of character size. As expected, the two larger sizes were judged much easier to read than the smaller ones. The difference was significant among all sizes except between the two largest. The greatest difference was between the 11 and 17 arcmin character sizes and is on the order of 2.5 rating units.

Brightness by character size interaction (B x CS). Figure 39 shows the ease of readability rating as a function of character size and brightness. The difference between the two brightness levels was significant with the 7, 11, and 17 arcmin sizes and fairly constant; it was equivalent to about one half of one rating unit.

Correlations

Table 22 presents the Pearson correlation matrix for all of the variables studied in the experiment. The coefficients were calculated using the values obtained for each hue by brightness treatment combination. The correlations among the measures (objective and subjective) are all relatively large (i.e., 0.697 is the lowest) which indicates good agreement in their predictive ability. The correlation between the two response times is almost perfect as noted earlier.

The correlations between the measures and luminance do not show any interesting results. However, as described earlier, correlation between luminance and ease of readability ratings as well as between luminance and performance (i.e., accuracy and response time) did show better colinearity for the conditions associated with the smaller character sizes at low brightness.

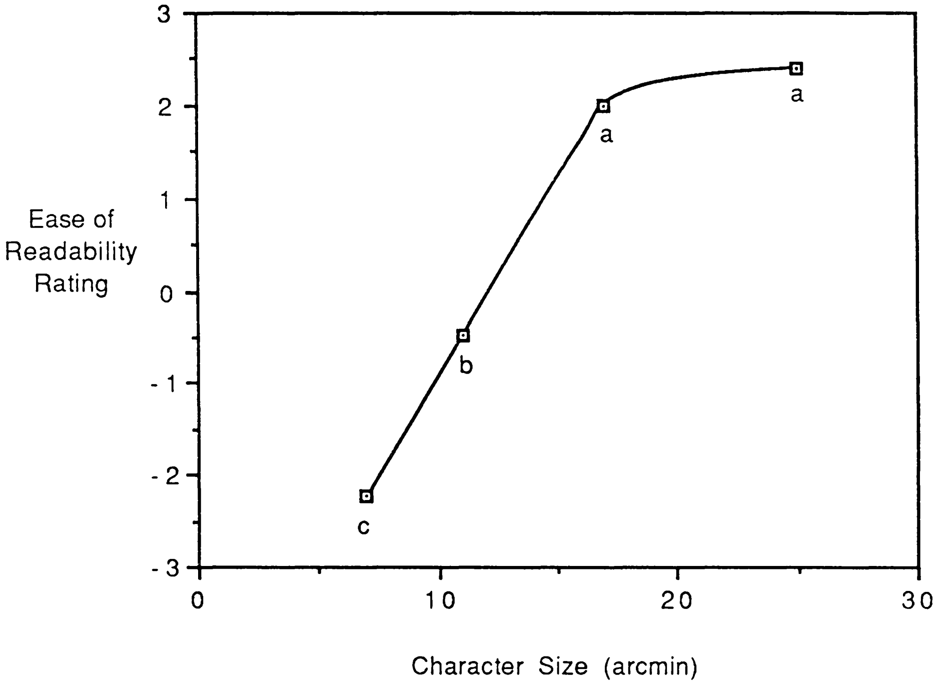


Figure 38. Mean ease of readability rating as a function of character size.

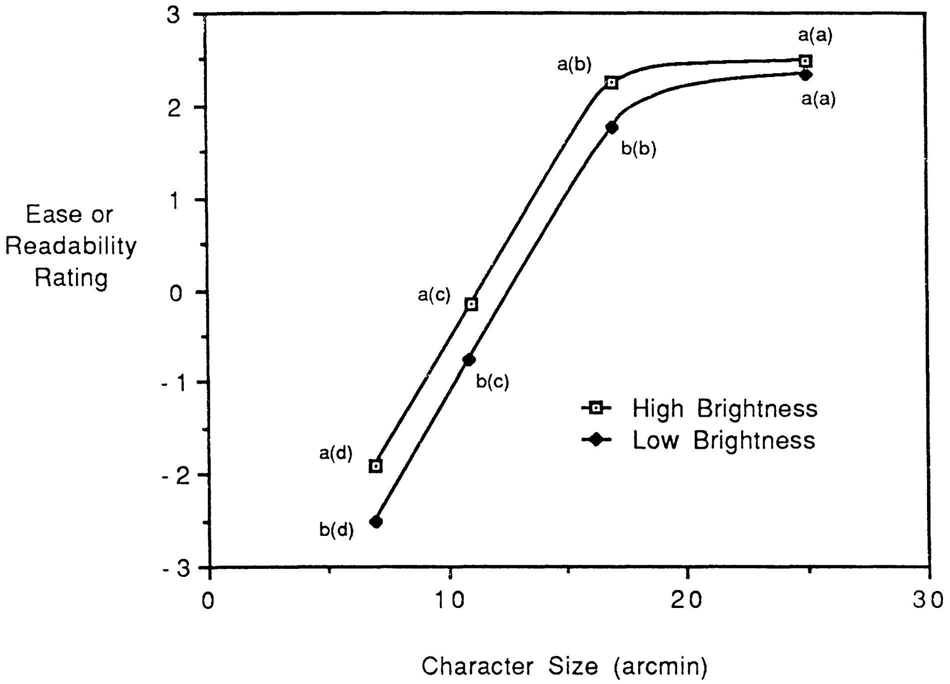


Figure 39. Mean ease of readability rating as a function of character size and brightness.

TABLE 22

Pearson Correlations Between all Dependent Measures and Luminance

Variable	RTCNO	RTANY	GTCNO	LDANY	%CORR	%CINC	ATTRA	COMFT	EASRD	LUMIN
RTCNO	1									
RTANY	0.998	1								
GTCNO	0.898	0.901	1							
LDANY	0.768	0.783	0.818	1						
%CORR	-0.988	-0.982	-0.877	-0.745	1					
%CINC	-0.981	-0.980	-0.848	-0.697	0.989	1				
ATTRA	-0.790	-0.804	-0.739	-0.788	0.783	0.761	1			
COMFT	-0.900	-0.911	-0.832	-0.783	0.891	0.868	0.900	1		
EASRD	-0.925	-0.927	-0.791	-0.802	0.897	0.878	0.839	0.925	1	
LUMIN	-0.696	-0.706	-0.567	-0.591	0.628	0.667	0.568	0.549	0.695	1

Note: the correlation coefficients were calculated from the values of the dependent variables for each hue by brightness combination. A correlation coefficient of at least ±0.50 is required for $p < 0.05$ and a correlation coefficient of at least ±0.57 is required for $p < 0.01$.

RTCNO: time to a correct response or no response
RTANY: time to any response
GTCNO: glance time for a correct response or a no response
LDANY: variance of lane deviation to any response
%CORR: percentage of correct responses
%CINC: percentage of correct and incorrect responses
ATTRA: subjective attractiveness
COMFT: subjective comfort
EASRD: subjective ease of readability
LUMIN: luminance value

CONCLUSIONS

When reading the conclusions which follow, the specific conditions of the experiment, particularly the very high luminance contrast ratios used, must be kept in mind.

- Character size appears to be the most important of the variables studied here and its limits are imposed by older drivers.
  - The 7 arcmin size should not be used since it is particularly difficult to read for older drivers and thus leads to significant decrements in driving and reading performance. Some older drivers systematically ignored words presented with this character size.
  - The 11 arcmin size yielded very good performance for the younger and middle age drivers. For the older drivers, however, a substantial decrement in performance was observed with this size. There is little doubt that an even more important degradation in performance would have been observed with lower luminance contrast ratios. Some older drivers also ignored words presented with this character size. Therefore, the 11 arcmin size should be avoided whenever possible.
  - The two larger character sizes (i.e., 17 and 25 minutes of arc) both yielded best performance. The two larger sizes were also found significantly easier to read than any of the other sizes. Given the limited visual capabilities of older drivers, their slightly lower performance with the 17 arcmin size when compared to the middle age drivers in this experiment, and the fact that the very high resolution and character definition achieved in this experiment can be difficult to reproduce in actual automobile instrument panels, preference

should most probably be given to the 25 arcmin size as suggested in the literature.

- Hue does not appear to have an impact on driving and reading performance when the two larger character sizes are used at any brightness level tested in this experiment. This is also true for the 11 arcmin character size at high brightness. Hence, no hue needs to be disqualified for these conditions with respect to performance.

- Brightness as set in this experiment has an impact on performance only for the smaller character sizes. For the larger sizes, the effect of brightness (luminance) is negligible as long as it varies from and above a minimum brightness (luminance) level set by the driver. Higher brightness was rated more attractive than low brightness by all subjects. The same result was observed for comfort which was rather unexpected. The driver should thus be allowed to set the brightness (luminance) of the instrument panel illumination at will, especially for the reddish orange color.

- Word complexity as defined in this experiment, that is, by a combination of frequency of occurrence in the English language and number of syllables, had a significant impact on glance time at the displays with all character sizes. Legends should thus include simple, commonly used, and appropriate words.

- Based on the assumption that perceived brightness of each color was the same within the given brightness conditions used in the experiment, the following conclusions on hue are true for all age groups:

- Reddish orange, white, orange, and amber appear to be the most attractive colors while light blue appears to be the least attractive. In the middle are blue-green, green, and red.
- No significant difference was found among the colors with respect to comfort.
- White was found to be the easiest color to read and light blue the worst. In the middle are all other colors.

If the assumption is not true, then subjective ease of readability may be more influenced by luminance of the most difficult characters to read than by hue itself.

- Some of the drivers will consciously ignore a legend if they cannot read it; a legend that will not be read by a certain category of drivers because of a very small character size and/or an inappropriately low luminance should most probably not be found on an automobile instrument panel.

DESIGN GUIDELINES

The experiment described in this report and the conclusions drawn can be converted into a set of design guidelines. These guidelines, if followed, should result in instrument panels having high readability. Furthermore, the guidelines should result in a high level of acceptability by drivers.

By and large the guidelines can be based on performance and preference bounds, that is, those values at which performance begins to degrade or preference begins to recede. The domain inside the bounds, that is, on the side in which there are no performance or preference decrements, can be considered as appropriate for guidelines and for actual use.

Guideline 1 (Performance Related)

For the eight hues used in the experiment, the luminance should be adjustable under control of the driver. The recommended range of adjustment is a function of the hue. Figure 10 (as well as Appendix O) provides the data for this recommendation.

The figure shows that reddish orange should have a range of approximately 3 to 49 cd/m^2, whereas green should have a range of 2 to 7 cd/m^2. The other hues fall between these limits, as shown in the figure.

This guideline is based on the idea that younger drivers will tend to use the lower end of the luminance adjustment range while older drivers will tend to use the upper end.

Again, it must be stressed that selection of a given hue dictates the use of a specific luminance range. Of course it is possible to offer the driver a larger range than that specified for each hue. However, the range should not be so much larger that the control has too high a gain over the specified range. In other words, the range specified in Figure 10 should take up a substantial portion of the control movement range, thereby making fine adjustment possible.

Guideline 2 (Performance Related)

The larger character sizes should be used wherever possible. In this experiment four character sizes were used corresponding to 7, 11, 17, and 25 minutes of arc (i.e., 1.5, 2.5, 3.7, and 5.5 mm of height at a 750 mm viewing distance). The two larger sizes (17 and 25 arcmin) show no degradation in performance at very high contrast ratios, regardless of hue or age group, and therefore can be recommended. However, taking into account the results of other legibility research as well as the limited visual capabilities of older drivers, preference should be given to the 25 arcmin character size.

The 11 minutes of arc character size does yield significant performance degradation for older drivers especially at lower luminance levels. Therefore, it is recommended that it should be avoided whenever possible.

The smallest character size used in the experiment, 7 minutes of arc, is not currently in use in automobiles, and it should not be used. Very serious performance decrements occur for older drivers with this character size. Middle age and younger drivers also incur performance decrements. Thus, going to smaller character size is not a feasible method of increasing the number of functions on the automobile instrument panel.

In summary then, the 17 arcmin and particularly the 25 arcmin character sizes are recommended. The 11 arcmin should be avoided whenever possible. The 7 arcmin, which is not currently used, should never be used.

Guideline 3 (Performance Related)

Word complexity in the experiment exhibited significant effects on glance time at the displays with every character size. Although the difference observed between low and high word complexity were small, they suggest that when the driver is placed in a stressful situation, performance decrements will occur with the complex words. Thus, the recommendation is that the most appropriate, simple, and common words be used for labeling (i.e., good human factors engineering should be applied).

Guideline 4 (Preference Related)

In regard to subjective preference, there is difference in attractiveness of the various hues. These preferences, fortunately, were independent of brightness. Figure 35 shows that light blue is considered definitely less attractive than orange, reddish orange, white, and amber. Strictly speaking, the remaining colors —blue-green, green, and red— do not differ from light blue (or the other colors for that matter). However, their mean values for attractiveness are substantially larger. Stretching things a bit then, attractiveness would be ordered as follows:

- Most attractive:
  orange, reddish orange, white, amber
- Next most attractive:
  blue-green, green, red
- Less attractive:
  light blue

In general the results suggest that seven of the eight colors are considered attractive, while one color, light blue, is considered close to neutral on an attractiveness/unattractiveness scale. It could be argued that light blue is still considered acceptable since on the average it was not rated as unattractive.

Guideline 5 (Preference Related)

Figure 32 shows that subjectively rated comfort did not differ among the eight hues. Again, comfort did not interact with brightness. Therefore all eight colors can be considered as equally comfortable.

Guideline 6 (Preference Related)

Figure 36 contains the information on subjective readability as a function of hue and character size. Since the 7 arcmin has been eliminated by Guideline 2, it will not be considered further in this guideline. Starting with the 11 arcmin, it is clear that subjects rated the readability of the hues reliably differently. Allowing some license in the interpretation of the post-hoc tests, subjective readability could be divided into two groups as follows:

- Highest readability:
  reddish orange, white, amber
- Next Highest:
  blue-green, green, orange, light blue, red

It should be noted, however, that at this character size (11 arcmin) the range of the mean ratings is from -.08 to +0.1, in round numbers. Thus, on the average, drivers did not find this character size to be very readable.

For the 17 arcmin character size there is also a statistically reliable difference among the hues as shown in Figure 36. However, the practical difference in the ratings is relatively small. In any case, interpreting the post-hoc tests quite liberally, readability might be placed in three overlapping groups as follows:

- Most readable for the 17 arcmin:
  reddish orange, white, amber
- Next most readable for the 17 arcmin:
  blue-green, green, red
- Lowest readability ratings for the 17 arcmin:
  light blue and orange

However, once again it is emphasized that the difference among the three groups is really quite small and that with the 17 arcmin size, all hues are considered quite readable.

With the 25 arcmin character size, there is no reliable difference in subjective readability as a function of hue. Thus, any hue appears acceptable.

To summarize this guideline, it is suggested that subjective readability difference is quite small as a function of hue. However, slight preference might be given to reddish orange, white, and amber over the other hues. This preference should be considered to be of secondary importance because of its small magnitude.

Guideline 7 (Performance and Preference Related)

A rather expected result of this research is that subjects performed better at the high brightness level for the various hues regardless of age. Figures 15, and 16 demonstrate these improvements. Furthermore, they rated high brightness as more attractive, more comfortable, and more readable (Table 19 and Figure 39). Therefore, the recommendation associated with this finding is that drivers should adjust the brightness control to the highest level that does not produce intolerable glare for them.

Ranking of Guidelines

The seven guidelines presented in this section are ordered according to type, that is, performance, preference and the combination of performance and preference. These guidelines are not presented in what would be considered a descending order of importance.

While the ranking of importance could be argued, it is felt that one should be given, based mainly on the relative magnitude of the differences that were found. The ranking from most important to least important would be as follows, each including a brief review statement:

Guideline 1.

Each hue should have a luminance adjustment range by the driver as specified in Figure 10 (and Appendix O).

Guideline 2.

The 17 and 25 arcmin character sizes should be used wherever possible with preference given to the 25 arcmin size. The 11 arcmin size should be avoided whenever possible. The 7 arcmin size should never be used.

Guideline 7.

Drivers should be able to select the highest comfortable luminance level.

Guidelines 4, 5, and 6.

In terms of subjective attractiveness and readability, ordering is roughly the same:

- reddish orange, white, and amber received the highest rating
- blue-green, green, orange, and red received acceptably high ratings
- light blue received acceptable ratings, but was considered neutral in terms of attractiveness

All hues were found equally comfortable.

Guideline 3.

Words used for labels and legends should be simple and commonly used while also being appropriate.

RECOMMENDATIONS FOR FUTURE RESEARCH

It should be recognized that the conclusions drawn and the recommendations made in this report are largely the result of the single experiment that was conducted. While it is believed that the experiment was as relevant as it could be made, not all questions could be answered. Every time researchers deal with color, they are faced with many possible independent variables, each with several important parameter settings. Only a small subset can be addressed in any given experiment.

The recommendations for future work are based largely on investigation of other important independent variables that could not be studied in the experiment described in this document. If some of this research were carried out, it would aid in determining any changes needed in the guidelines as well as possibly producing additional new ones.

The most important remaining independent variable that should be studied is driving scene luminance and glare. It is well recognized that scene luminance and glare may vary over a large range in night driving. At the one end of the scale is the lonely blacktop country road damp from rain, while at the other end is the well-lit multilane highway with dense traffic. Obviously, these extreme conditions differ markedly both in legend-to-surround luminance contrast ratio and in legend-to-driving scene luminance ratio. In the present experiment only one setting of driving scene luminance was used, a value that was considered to represent the middle of the range of variation.

Another important independent variable is instrument panel illumination method. In the present experiment, drivers read words using backlit displays. It would be important to determine any difference in performance based on other illumination technologies, including LEDs, vacuum fluorescents, conventional front lit, and possibly CRTs and transilluminated LCDs. Each of these technologies may possess certain idiosyncrasies that could cause differences in results and corresponding guidelines.

Yet another variable that should be studied is instrument panel color effects. Because different instrument panel colors could change the contrast ratio, instrument panel color could be quite important. The recommendation to study the effect of instrument panel color should allow for effects of variations in scene luminance which has already been suggested.

It would be worthwhile to examine stroke width-to-height ratio as an independent variable and to determine the effects on both performance and subjective preference.

It should be recognized that the present study made use of a specific cross-comparison technique for setting luminance levels. In particular, blue-green was always used as a reference. After setting the "best" level for blue-green, subjects subsequently set the other hues by comparing to blue-green. This approach produced what is essentially an equal-brightness type of experiment. Guideline 1 in the guideline section thus recommends luminance adjustment ranges based on an equal-brightness preliminary experiment. There is, however, another type of approach that could be taken. Instead of cross-comparing to a common hue (blue-green) it would be possible to allow subjects to select a luminance setting for each hue based on "best" setting for the driving scene. This procedure would not necessarily produce equal brightness, but instead would produce the "best" subjective setting, possibly the setting they would most likely use in their own automobile. While it is possible that differences between this "best" setting approach and the cross-comparison approach previously used may be relatively small, it would be worthwhile to experimentally determine the differences as a check on the earlier results. Should there be any major differences, they should be examined in detail.

Finally, a topic worthy of additional attention is the fact that reddish orange did not yield significantly better performance than other colors, despite its very high luminance value in the high brightness condition. A follow-up study could be set up in which performance would be measured for various saturation and luminance levels of the color. Other hues of equivalent saturation and luminance such as amber or orange could be used as controls.

REFERENCES

Booker, R.L. (1981). Luminance-brightness comparisons of separated circular stimuli. *Journal of the Optical Society of America, 71*, 139-144.

Briggs, R.P. (1986). Visual changes among older workers: implications for workstation design. In *Proceedings of the Human Factors Society Annual 30th Meeting* (pp. 801-803). Santa Monica, CA: Human Factors Society.

Carr, R.M. (1967). The effect of color coding indicator displays on dark adaptation. *Human Factors, 9*, 175-180.

Carrol, J.B., Davies, P., and Richman, B. (1971). *Word frequency book.* New York: American Heritage.

Carter, J.H. (1982). The effects of aging upon selected visual functions: Color vision, glare sensitivity, field of vision, and accommodation. In R. Sekuler, D. Kline, and K. Dismukes (Eds.), *Aging and human visual function*, (pp. 117-120). New York: Alan R. Liss, Inc.

Christ, R.E. (1975). Review and analysis of color coding research for visual displays. *Human Factors, 17*, 512-570.

Cushman, W.H. and Crist, B. (1986). Illumination. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 670-695). New York: Wiley.

Donohoo, D.T. and Snyder, H.L. (1985). Accommodation during color contrast. *Digest of the Society for Information Display*, 200-203.

Easterby, R. and Zwaga, H. (1984). *Information design.* New York: Wiley.

Faye, E.E. (1986). Coping with impaired vision and aging. In *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 795-797). Santa Monica, CA: Human Factors Society.

Finlay, D. and Wilkinson, J. (1984). The effects of glare on the contrast sensitivity function. *Human Factors*, *26*, 283-287.

Fowkes, M. (1984). Presenting information to the driver. *Display Technology*, *5*, 215-223.

Galer, M.D. (1984). The application of ergonomics in the design of automotive displays. *Displays*, (October), 224-228.

Galer, M.D. (1986). Human factors in the design and assessment of in-vehicle information systems. *International Journal of Vehicle Design*, (Special Issue on Vehicle Safety), 338-343.

Galer, M.D. and Simmonds, G.R.W. (1985). *The lighting of car instrument panels-drivers' responses to five colours.* (SAE-850328). Society of Automotive Engineers, Inc.

Grether, W.F. and Baker, C.A. (1972). Visual presentation of information. In H.P. VanCott and R.G. Kinkade (Ed.), *Human engineering guide to equipment design* (Rev. Ed.) (pp. 41-121). Washington, D.C.: U.S. Government.

Helander, M.G. (1986). Design of visual displays. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 507-548). New York: Wiley.

Hood, D.C. and Finkelstein, M.A. (1986). Sensitivity to light. In K.R. Boff, L. Kaufman, and J.P. Thomas (Ed.), *Handbook of perception and human performance, Volume 1: Sensory processes and perception* (Chapter 5). New York: Wiley.

Howett, G.L. (1986). The coming redefinition of photometry. *Journal of the Illuminating Engineering Society*, (Summer), 5-18.

Hughes, P.C. and Neer, R.M. (1981). Lighting for the elderly: a psychobiological approach to lighting. *Human Factors*, *23*, 65-85.

Human Factors Society (1986). *American national standard for human factors engineering of visual display terminal workstations.* Santa Monica, CA: Human Factors Society.

King-Ellison, P., and Jenkins, J.J. (1954). The durational threshold of visual recognition as a function of word-frequency. *American Journal of Psychology, 67,* 700-703.

Klare, G.R. (1963). *The measurement of readability.* Ames, Iowa: Iowa State University.

Klare, G.R. (1974-75). Assessing readability. *Reading Research Quarterly, 1,* 62-102.

Leonard, J. and Wierwille, W.W. (1975). Human performance validation of simulators: Theory and experimental verification. In *Proceedings of the 19th Annual Meeting of the Human Factors Society* (pp. 446-456). Santa Monica, CA: Human Factors Society.

Lippert, T.M. (1984). *Color contrast effects for a simulated CRT headup display.* Unpublished Master's Thesis, Virginia Polytechnic Institute and State University.

Lippert, T.M. (1986). Color-difference prediction of legibility performance for CRT raster imagery. *Digest of the Society for Information Display,* 86-89. Playa del Rey, CA: Society for Information Display.

Lippert, T.M. and Snyder, H.L. (1986). *Unitary suprathreshold color-difference metrics of legibility for CRT raster imagery.* Virginia Polytechnic Institute and State University Technical Report HFL/ONR 86-3.

McCormick, E.J. and Sanders, M.S. (1982). *Human factors in engineering and design* (5th Ed.). New York: McGraw-Hill.

Merrifield, R.M. and Silverstein, L.D. (1986). *The development and evaluation of color systems for airborne applications: fundamental visual, perceptual, and display systems considerations.* (Tech. Report no. NADC-86011-60). Seattle: Boeing Commercial Airplane Company, Crew Systems Technology.

Mourant, R.R. and Langolf, G.D. (1976). Luminance specifications for automobile instrument panels. *Human Factors, 18,* 71-84.

Olson, P.L. and Sivak, M. (1984). Glare from rear-view mirrors. *Human Factors, 26,* 269-282.

Olzak, L.A. and Thomas, J.P. (1986). Seeing spatial patterns. In K.R. Boff, L. Kaufman, and J.P. Thomas (Ed.), *Handbook of perception and human performance, Volume 1: Sensory processes and perception* (Chapter 7). New York: Wiley.

Ordy, J.M., Brizzee, K.R., and Johnson, H.A. (1982). Cellular alterations in visual pathways and the limbic system: implications for vision and short-term memory. In R. Sekuler, D. Kline, and K. Dismukes (Ed.), *Aging and human visual function,* (pp. 79-114). New York: Alan R. Liss, Inc.

Osaka, N. (1985). The effect of VDU colour on visual fatigue in the fovea and periphery of the visual field. *Displays,* (July), pp. 138-140.

Padgham, C.A. and Saunders, J.E. (1975). *The perception of light and colour.* New York: Academic Press.

Palmai, E.W., Schanda, J., and Heine, G. (1980). Visibility of different coloured led displays. *Display Technology, 2,* 123-127.

Pitts, D.G. (1982). The effects of aging on selected visual functions: dark adaptation, visual acuity, stereopsis, and brightness contrast. In R. Sekuler, D. Kline, and K. Dismukes (Ed.), *Aging and human visual function*, (pp. 131-159). New York: Alan R. Liss, Inc.

Pokorny, J. and Smith, V.C. (1986). Colorimetry and color discrimination. In K.R. Boff, L. Kaufman, and J.P. Thomas (Ed.), *Handbook of perception and human performance, Volume 1: Sensory processes and perception* (Chapter 8). New York: Wiley.

Post, D.L., Constanza, E.B., and Lippert, T.M. (1982). Expressions of color contrast as equivalent achromatic contrast. In *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 581-585). Santa Monica, CA: Human Factors Society.

Pulling, N.H., Wolf, E., Sturgis, S.P., Vaillancourt, D.R., and Dolliver, J.J. (1980). Headlight glare resistance and driver age. *Human Factors, 22*, 103-112.

Reynolds, H.N. (1971). The visual effects of exposure to electroluminescent lighting. *Human Factors, 13*, 29-40.

SAS Institute Inc. (1985). *SAS User's Guide: Statistics Version 5 Edition.* Cary, NC: SAS Institute Inc.

Sanders, M.S. and McCormick, E.J. (1987). *Human factors in engineering and design* (6th Ed.). New York: McGraw-Hill.

Simmonds, G.R.W., Galer, M.D., and Baines, A. (1981). *Ergonomics of electronic displays.* (SAE-810826). Society of Automotive Engineers, Inc.

Snyder, H.L. (1980). *Human visual performance and flat panel display image quality.* Human Factors Laboratory, Blacksburg, VA. Report prepared for Office of Naval Research, Code 455. Arlington, VA.

Snyder, H.L. and Maddox, M.E. (1980). On the image quality of dot-matrix displays. In *Proceedings of the S.I.D., 21* (1), pp. 3-7.

Society of Automotive Engineers (1984). *Ergonomic aspects of electronic instrumentation: A guide for designers*. Warrendale, PA.

Thorndike, E.L. and Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University.

Vander, A.J., Sherman, J.H., and Luciano, D.S. (1977). *Physiologie humaine*. Montréal, Canada: McGraw-Hill.

Williams, C.M. (1967). Legibility of numbers as a function of contrast and illumination. *Human Factors, 9*, 455-460.

Ware, C. and Cowan, W.B. (1983). *Luminance to brightness conversion: A two degree factor based on a large population sample.* Technical paper submitted to the CIE for consideration as a provisional recommendation, 1983. (This paper may be obtained by writting to the authors at: Division of Physics, National Research Council of Canada, Ottawa, Ontario, Canada, K1A OR6).

Wyszecki, G. (1986). Color appearance. In K.R. Boff, L. Kaufman, and J.P. Thomas (Ed.), *Handbook of perception and human performance, Volume 1: Sensory processes and perception* (Chapter 9). New York: Wiley.

Wyszecki, G. and Stiles, W.S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.). New York, NY: Wiley.

APPENDIXES

Appendix A: Word Lists

This Appendix presents the words that were as stimuli on the slides. There were two levels of readability; high and low.

Readability = high

Criteria:

- One syllable words
- Equivalent school grades 1, 2, 3 and, 4

(i.e., occurrences ≥ 20 times per million words read)

BURST, CHECK, PAID, GUN, MILES, MAP, CAMP, DOWN, MARK, WAY, SCHOOL, SHOT, CLEAR, LONG, BAG, LOST, FRUIT, WORDS, FORM, NECK, NEAT, STREAM, SOUTH, EAST, BACK, SPEED, DRINK, BEST, FREE, HARD, NORTH, COAL, GOOD, GRASS, RATE, NEW, STREET, DAY, HEAT, DUST, RIGHT, STEP, WORTH, WORK, CHILD, BUS, USE, TRIP, PLACE, OIL, GUIDE, THREE, STEAM, REACH, SEAT, PART, CASE, WEST, LOT, LEG, WELL, EIGHT, HELP, SOFT, AGE, FENCE, YEARS, PLAIN, OFF, WALL, OLD, LAKE, FALL, GREAT, DOG, MASS, FLAT, SET, STORM, LAST, TEST, WRONG, FLIGHT, TRACK, LEFT, HEIGHT, NEXT, SOUND, TAIL, SAND, CLOCK, FLOW, COAT, SIGN, DUST, BORN, SHOW, GUESS, HIT, BELL, NIGHT, TONE, HOUSE, HUGE, WAVE, SWEET, HIGH, RIDE, DEER, FOOD, LUNCH, LAND, MUD, CAT, SIDE, RACE, WHEAT, PLATE, FOUR, DEAD, SCORE, HEAD, THIN, FAR, STONE, BEACH, LIGHT, SENSE, SUN, RULE, GRAIN, BAR, ROOM, PULL, SHOP, POINT, END, TRAIL, FIVE, HOME, CLIMB, THICK, PUSH, CAR, BREAK,

ROUGH, TRAIN, ROLL, TURN, FRESH, MOVE, DRIVE, ROOF, TREES, PARK, DOOR, GLAD, COLD, CUT, SIT, HALF, HALL, CORN, MAY, SHORT, BLOCK, BOARD, LIFT, FRONT, WASH, RING, SMILE, CLOSE, PRICE, WIND, STICK, CAVE, NOSE, REST, COAST, POST, SIX, CROSS, FIRE, COOL, CUP, FULL, PASS, TOWN, CLEAN, SICK, LOG, GAS, HOT, FATE, BRANCH, FAST, SAFE, RAYS, COST, POND, TEN, CLUB, BOX, WEAR, ROPE, PIE, ARM, START, ROW, PIECE, DECK, STOP, YARD, ROD, BANK, FINE, DARK, CAGE, ROAD, SHORE, POLE, WARM, GAME, FUEL, CARD, PATH, DRY, NINE, CAP, ICE, STEEL, JET, SALT, FLOOR, KEY, SHADE, LOAD, SHIP, FORTH, THIRD, TRUCK, SPOT, RANCH, SEAL, LEAF, SNOW, RAIN, WHEEL, HILL, TEACH, SIZE, STRIP, PRESS, TRADE, SKY, CLOTH, DANCE, LAMP, MAIN, TRUNK, BLOW, STATE, FIELD, BOAR, QUICK, CHAIN, LOW, WOLF, WIDE, LOUD, DUCK, FOOT, SHED, VERB, SLOW, SPIN, GLASS, NOISE, HOUR, FAIR, CRACK, ROCK, TUBE, REAR, WALK, BROKE, ROUTE

Readability = low

Criteria:

- Two syllable words
- Equivalent school grades 5, 6, 7 and, 8

(i.e., 6 < occurrences < 20 times per million words read)

LAZY, RIDERS, SODIUM, TIMBER, MANKIND, LADIES, HUNTED, CAMEL, PLATFORM, CIRCUIT, SMOKING, PICNIC, LIMIT, PILOTS, MANY, ENTRY, OBVIOUS, PROJECTS, PRINTING, BEAVER, JUNIOR, TERRY, WARMER, EXPERT, BRITAIN, PUPPY, STRONGLY, PRETEND, SAILOR, BRILLIANT, CONTROLS, FATHERS, CABBAGE, SHINY, EXPAND, JEWELS,

RULERS, AFFAIRS, RESPONSE, TRADING, AFFORD, REPAIR, ENTERS, SUNSET, NICKEL, CLOVER, SPINNING, MANNERS, CONTENT, REPLY, RABBIT, CARGO, BLANKETS, LIGHTER, TENNIS, HEALTHY, PERMIT, ACCENT, CELLAR, APPLY, WILLIE, WINNER, EXCUSE, COOLER, FAILURE, SANDY, CRYSTAL, SAVING, DELIGHT, ANXIOUS, LINING, PARTIES, ORGANS, CANYON, FORMAL, CONSTRUCT, MOONLIGHT, TRULY, COUNTY, MENTION, CHIMNEY, STRIKING, BACON, WORKER, KANSAS, PORTION, SLEEPY, CREDIT, EXTEND, RETURNS, HUMOR, PROMISE, APPROACH, LATEST, CHARLIE, CANNON, POISON, SEASONS, HANDLING, FAVOR, COLLAR, CRAZY, COMFORT, USELESS, SARAH, ADDRESS, RIBBON, PILLOW, DESTROY, ERRORS, DRAMA, QUARTERS, BUNDLE, POLAR, DARING, LEVER, CACTUS, BEDROOM, MUDDY, HOBBY, UNION, IMPACT, PREFER, RAINBOW, SQUIRREL, CAMPING, AIRPORT, TURKEY, ANNUAL, SPLENDID, MOTORS, MUSCLE, CONTACT, DANGERS, BILLION, AIRCRAFT, PITCHER, CONTRACT, NETWORK, MINOR, MAYOR, GENIUS, FASTEN, OUTPUT, ARTHUR, STATUE, DRINKER, HAROLD, BELIEF, RAINY, TENDER, LIQUIDS, GARBAGE, MEDIUM, ACCOUNTS, SLENDER, FOUNTAIN, HEATING, TREASURE, NATIONS, BAMBOO, BARREL, SIDEWALK, SAUCER, TROUBLES, SOVIET, TARGET, ABROAD, RAPID, SPIRITS, VOWELS, RACCOON, BUCKET, METER, MERRY, SERVICE, TICKET, PRISON, BUTTON, CHARGES, PATIENT, CASTLE, FRICTION, NOVEL, DAYLIGHT, GALLON, PUMPKIN, HONEY, ROGER, CONCERT, WEAPON, LEVELS, DOCTOR, YANKEE, CHILDREN, DAIRY, ANGER, TIGER, PARROT, TOKYO, VICTIM, STABLE, SICKNESS, PRINCESS, NAVY, TOURISTS, HIGHWAY, DINING, STUPID, WARRIOR, FOOTSTEPS, SHOVEL, FEATHER, AWFUL, ROYAL, TEMPER, DIAMOND, COLDER, BULLET,

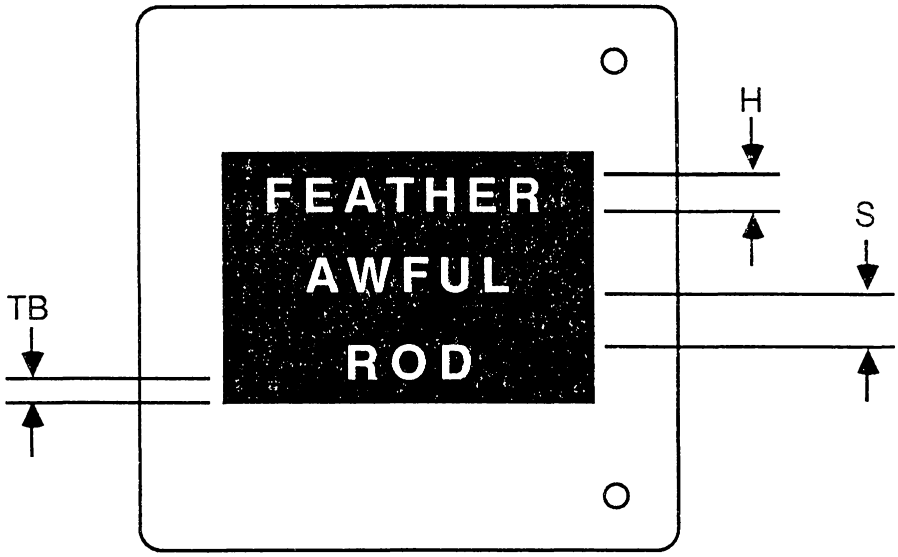
MONDAY, SPIDERS, MASTER, DOZENS, BUREAU, ARTISTS, AGENT, RUSSIANS, CONCEPT, MONSTER, CHANCES, ACTORS, BUBBLE, ROBOT, HUNGER, CARPET, TUNNEL, RADAR, POLLY, BRAZIL, RADIOS, NUMBER, DOWNWARD, TALLER, PARENT, FUNCTIONS, INDEX, CATSUP, SLOWER, PARTNER, PUPIL, REWARD, SPEAKER, VISION, SCARLET, SWEATER, HOTEL, DISPLAY, ADULTS, BALLOONS, FRACTIONS, HOUSTON, CEMENT, FREQUENT, PAINTING, SESSIONS, OWNERS, VESSEL, ALARM, PRAYER, LINDA, CONFLICT, PENCILS, WINDY, PENNIES, ANCHOR, WIDOW, NANCY, ROBIN, CHAMPION, SAFER, BORROW, JUSTICE, KATHY, CUSTOM, COUSIN, CHERRY, OUTFIT, STURDY

Summary:

	<u>*Readability*</u>	
	high	*low*
Number of words:	288	288
Average word length (letters):	4	6
Longest word length (letters):	6	9

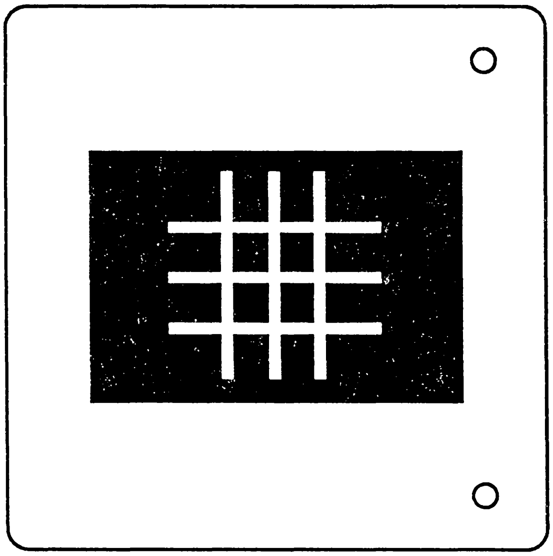
Appendix B: Slide Specifications

This appendix shows a typical slide as well as some specifications for all character sizes.



Character size (arcmin)	*H (mm)*	*S (mm)*	*TB (mm)*
7	1.5	6.6	2.5
11	2.5	6.0	2.2
17	3.7	4.5	1.5
25	5.5	1.5	1.5

This is the grid which the subject viewed while making a subjective evaluation of the hue by brightness combination:



Appendix C: Instructions for Subjective Preferences

During the study, you will be asked to assess the attractiveness and distractibility of the different combinations of color and illumination level on the screens on which you will read the words presented. Also, you will be asked to assess the readability of each character sizes in which the words will be presented.

At the end of each segment of the experiment, you will be prompted by one of the experimenters to provide your answers while the other experimenter changes the color and illumination level in preparation for the next segment. There will be 16 segments in each session and thus, 16 sets of 6 answers to provide. You will have to base your judgments on the scales that follow, so please read them carefully:

Attractiveness of the color (or aesthetic):

- 3	- 2	- 1	0	+1	+2	+3
Extremely Unattractive	Unattractive	Somewhat Unattractive	Neutral	Somewhat Attractive	Attractive	Extremely Attractive

Comfort of the color (or annoyance):

- 3	- 2	- 1	0	+1	+2	+3
Extremely Un-comfortable	Un-comfortable	Somewhat Un-comfortable	Neutral	Somewhat Comfortable	Comfortable	Extremely Comfortable

Ease of Readability:

- 3	- 2	- 1	0	+1	+2	+3
Extremely Difficult	Difficult	Somewhat Difficult	Neutral	Somewhat Easy	Easy	Extremely Easy

All of your answers have to be numbers; those found on the scales. For the Ease of Readability scale, you will have to provide an answer for each of the 4 character sizes. For instance, if you think that the smaller characters were difficult to read you should answer "-2" when the experimenter prompts you for an answer for this character size.

Remember that all three scales have the same anchor points:

-3 = Extremely (on the negative side)
-2 =
-1 = Somewhat (on the negative side)
0 = Neutral
+1 = Somewhat (on the positive side)
+2 =
+3 = Extremely (on the positive side)

You can "fit" your assessment in the proper place.

For instance, the experimenter would ask you the following question: What attractiveness rating would you give to the color you just experienced? Suppose you think it was no more but no less than attractive, you would then answer "+2", and so on with distractibility and ease of reading. Do you have any question?

Appendix D: The Driving Simulator

The driving simulator at Virginia Polytechnic Institute and State University is a versatile, closed-loop, moving-base device. It has been used for numerous research studies involving effects of various parameters on driver control, effects of fatigue, effects of physical motion, driver workload, and evaluation of new techniques in automotive technology.

The simulator provides the driver with a closed-loop roadway display in coordination with the motions of yaw, roll, and lateral and longitudinal translation. Four channels of sound along with vibration are also provided. Brief descriptions of the simulator's subsystems follow.

Visual System. Generation of the simulated roadway image is accomplished by a special purpose computing system. The generated signals are initially displayed on the face of a cathode ray tube. A CCTV camera scan converts this image and transmits it by cable to a CCTV monitor mounted above and behind the dash on the simulator. A Fresnel lens with an effective focal length of 50.8 cm (20.0 in), located between the monitor and the human operator, decreases the apparent roadway image proximity to the driver and enhances the illusion of distance [effective distance 10.1 m (33 ft)]. Additional realism is provided by a plexiglas windshield and a sheet metal mock-up representing a hood and fenders immediately in front of the dash. The field of view provided by the CCTV monitor and lens subtends 39° vertically and 48° horizontally. During simulation, all room lights are turned off so that only the roadway display and the illuminated speedometer are visible to the driver.

Motion System. The simulator is composed of an upper and lower platform, three main struts, and four motion servos. The upper platform consists of a standard automotive configuration including bucket seat, dashboard (with speedometer), steering wheel, brake and accelerator pedal, and the visual display equipment described above. The steering system is active and can be adjusted to match the characteristics of most production cars. The upper platform is pivoted at each end which permits roll motion about an axis 33.0 cm (13 in) from the upper platform floor. The roll motion is accomplished by the roll servo which is attached between the upper and lower platforms. The lower platform, while providing support for the upper platform, is supported by nine precision rubber-wheeled casters which enable the platform to rotate in yaw and to translate longitudinally and laterally on a large lucite sheet. The platform motions are generated by the combined action of three servo-operated struts which have one end pivoted about a floor-anchored support. Two of the servos are aligned with the longitudinal axis of the simulator and provide longitudinal translation and rotation in yaw. The third servo is oriented laterally and provides the lateral translation. Each motion servo is monitored by a feedback potentiometer and is controlled by its own electrohydraulic valve which receives signal inputs from the analog computer. A hydraulic pump, regulated at 10.4 MPa (1500 psi), provides fluid power to the motion system. Acoustical insulation of the pump unit controls the noise level in the simulator room. A 2.5 gallon accumulator and associated valving are used to maintain constant fluid pressure at any required flow-rate. Fluid temperature is controlled by an integral refrigeration-type heat exchanger.

Vehicle Simulation. Simulation of the vehicle dynamics is performed by an EIA TR-48 computer. Driver inputs to the steering wheel, accelerator, and brake pedals are sensed by potentiometers and converted to electrical signals. These signals are the inputs to the vehicle model simulated on the computer. Outputs of the model are analog voltages of vehicle velocity, roll, yaw, lateral position, and longitudinal position, which form the signals applied to the motion servos. These signals are also applied to the driver's speedometer and to the image generation circuitry which continuously adjusts the visual display characteristics (position, perspective, velocity, etc.) to correspond to the simulated vehicle state.

The vehicle model for the simulation allows for rotations in roll and yaw as well as lateral and longitudinal translation. Four separate inputs are provided; namely, steering wheel displacement, accelerator/brake displacement, aerodynamic force (wind gust), and road curvature. The model consists of a set of simultaneous time-varying differential equations relating the inputs to the vehicle motion components. This approach approximates the dynamic response of the vehicle and permits matching of the model responses of either measured full-scale responses or to responses generated by digital simulation models.

Stimulus Generation and Measures Computation. The simulator is supported by three microcomputers with multipurpose interfaces. These systems allow a wide variety of stimuli to be generated and used in the simulation. They also allow the precise on-line measurement of many kinds of performance parameters. Measures are stored on disk and then later transmitted to the University's central computing facility for statistical analysis.

Both an eight-channel chart recorder (Sanborn 350) and a 14-channel instrumentation tape recorder (Honeywell 5600E) are available on a dedicated basis for the simulation. Furthermore, a variety of special purpose equipment is available including video cameras, video recorders, audio cassette recorders, physiological monitoring equipment, and a breathalyzer. This equipment can be configured in numerous ways to fit the needs of a given experiment.

Appendix E: Reference Colors

This appendix provides the chromaticity coordinates of the references colors used for all chromaticity measurements (i.e., the processor was calibrated to these references).

Four references were used: a reference white, a red, a green, and a blue. The reference white was a CIE Standard Illuminant A (2856°K) and the other colors were obtained by using standardized color filters in conjunction with the CIE Standard Illuminant.

Reference	*Filter transmittance*	*Luminance (Y) (cd/m^2)*	*x*	*y*
White	--	340.0	0.4512	0.4144
Red	0.2471	84.2	0.6641	0.3291
Green	0.1836	62.5	0.1971	0.6548
Blue	0.0466	15.9	0.2066	0.2056

Appendix F: $L^*a^*b^*$ to Yxy Transformation

This appendix gives the L*a*b* coordinates that were provided with the eight color filters used in the experiment as well as the corresponding CIE 1931 chromaticity coordinates.

Color	L^*	a^*	b^*	*Y* (*cd/m²*)	*x*	*y*
Blue-green	16.02	-31.42	1.78	2.10	0.2768	0.5469
Green	20.35	-46.85	6.74	3.08	0.2319	0.6130
Orange	10.90	-29.01	-0.70	1.25	0.2523	0.5449
Light Blue	36.01	-18.78	-11.26	9.01	0.3595	0.4249
Red	51.42	44.72	35.26	19.63	0.6010	0.3561
Reddish orange	75.91	39.34	94.74	49.74	0.5878	0.4028
White	62.66	4.93	12.48	31.18	0.4752	0.4144
Amber	58.86	21.71	84.52	26.88	0.5683	0.4255

The coordinates provided for Orange did not correspond to the color of the filter received; the coordinates indicated that the color should have been in the green region of the chromaticity diagram whereas the corresponding filter was clearly in the yellow region. Hence, the chromaticity of the Orange filter was measured using a CIE Standard Illuminant A closely approximated by putting a light blue filter on the mirror inside the projector and driving the bulb to 1.227 amps. The results were as follows:

Color	*Y* *(cd/m²)*	*x*	*y*
Approximate Illuminant A	666.0	0.4503	0.4168
Orange	13.6	0.5836	0.4113

The values obtained for Orange were thus used in the Color Matching part of the study.

Appendix G: Characteristics of the Correcting Color Filters

As mentioned in the text, transparent (gel type) color filters were used to correct the spectral characteristics of the light emitted by the bulb at various currents. At low currents, the color temperature of the bulb was in the yellow portion of the CIE 1931 chromaticity diagram; hence, a number of blue filters depending on the current value were used. At high currents, the color temperature of the bulb shifted toward the blue region, hence, a number of yellow filters again depending on the current value were used.

It is important to note that these filters helped to approximate a CIE Standard Illuminant A at various current values and even though this approximation can be considered as fairly good given the time and the equipment that were available, at no point in this experiment was a CIE Standard Illuminant A replicated.

CIE 1931 chromaticity coordinates of the transparent filters were measured using a CIE Standard Illuminant A.

Color	*Y* (cd/m^2)	x	y
White	340.0	0.4515	0.4152
Light Yellow	305.0	0.4602	0.4197
Yellow	293.0	0.4813	0.4303
Light Blue	200.0	0.4100	0.4087

Appendix H: Participant's Informed Consent

Brief Description of the Experiment. You are being asked to participate in an experiment in which dashboard lighting will be changed in illumination level and in color. The objective is to determine which colors and levels are best from the standpoint of your reading performance and your preference. The experiment will be performed in a driving simulator in the Vehicle Analysis an Simulation Laboratory. This simulator has been used for many previous experiments.

Your part as a participant in the study is to drive the simulator at a constant 55 mph just as you would on an interstate highway when observing the speed limit. While driving, you will read aloud the words presented one at a time on each of two screen as quickly and as accurately as possible. The words will appear at constant time intervals. The words will be displayed in different colors and sizes on the screens. At several points in the experiment, you will be asked to assess several characteristics of the displays.

You must remember that the driving task is the most important, and hence, you must keep the vehicle in a proper position on the road and perform as best as you can the secondary word reading task.

To participate in this experiment, you must have a valid driver's license. Your visual acuity as well as your color vision will be tested prior to the experiment. These tests are very simple and are not harmful in any way. You can expect to be withdrawn from the experiment if your vision does not meet the minimum criteria set for this study. Also, your hearing must be adequate, and your reading comprehension of English must be adequate. Additionally, the experimenters will ask you a few questions about your driving, such as the miles you drive per year, amount of night driving you do, etc.

If you decide to participate, you will be paid $5.00 per hour for your participation (at the end of your participation). The entire experiment is expected to last approximately three hours total. The research team consists of two graduate students, Brian Hayes and Daniel Imbeau, under the direction of Dr. Walter W. Wierwille, the principal investigator and Faculty Member in the IEOR Department.

Further instructions will follow your reading and signing this consent form. As the form indicates, you have the right to discontinue participation at any point in the experiment. This includes the right to withdraw after you have read the instructions, or any time thereafter.

A member of the experimental team will answer any questions you may have. However, in cases that may affect the outcome of the experiment, the team member may delay a detailed answer until you have completed your sessions.

Please do not discuss the experiment with other persons who may become subjects. We expect all data to be taken by August 10, 1987. Following this date, feel free to discuss the experiment with anyone you wish.

Finally we want to point out that it is possible that at times you may feel frustrated or stressed, and the task may seem difficult. Your performance on the task reflects the difficulty of the task, not your personal abilities and talents. Furthermore, after you have completed the experimental sessions, your data will be treated with anonymity.

Informed Consent:

1- You have the right to stop the experiment in which you are participating at any time if you feel that it is not agreeable to you. If you should decide to withdraw while a simulator run is in progress, you must inform the experimenter so that the simulator motion can be stopped before you leave the simulator. Should you terminate the experiment, you will receive pay only for the time you actually participated.

2- If you have any problems with or questions about the research itself, you may contact Dr. Walter W. Wierwille at the phone number and address given at the end of this consent form. If you have any questions about your rights as a participant, you may contact Mr. Charles D. Waring, Chairman of the Institutional Review Board at Virginia Tech (703) 961-5283. If you wish to receive a summary of the results of the research, please include your address (six months hence) with your signature below. If more detailed information is desired after receiving the results summary, please contact the Vehicle Analysis and Simulation Laboratory, and a full report will be made available to you.

3- The risks involved in this experiment include possible interference in your activities following the experiment because of fatigue and the possibility of injury if you attempt to leave the driving simulator while it is in motion.

4- The discomforts involved in this experiment include possible fatigue due to the length of the simulator driving session, which will probably last about two and a half hours.

The faculty and graduate students involved in this study greatly appreciate your help as a participant. We hope that this experiment will be an interesting experience for you. Your data along with that of other participants should aid in the eventual improvement of automobile dashboard lighting design.

Your signature below indicates that you have read this document in its entirety, that your questions have been answered, and that you consent to participate in the study described. If you include your printed name and address below, a summary of the experimental results will be sent to you.

______________________	______________________
Signature	print name
	______________________
	street
	______________________
	city, state, zip

Vehicle Analysis and Simulation Laboratory
IEOR Department
Virginia Tech
Blacksburg, Virginia 24061
951-7962

Appendix I: Median Current Values, Luminances (Y), and CIE 1931 Chromaticity Coordinates from the Preliminary Experiment

Older Age Group (high brightness):

Color	*Median Current (A)*	*Y (cd/m^2)*	*x*	*y*
BLUGR	1.296	8.34	0.2811	0.5253
GREEN	1.202	7.40	0.2135	0.5849
ORANG	1.098	20.30	0.6021	0.3960
LIBLU	1.042	7.96	0.4117	0.4312
RED	0.993	15.30	0.6274	0.3505
REDOR	0.978	48.00	0.6108	0.3860
WHITE	0.992	20.80	0.5280	0.4148
AMBER	1.050	29.60	0.5871	0.4091

Younger Age Group (low brightness):

Color	*Median Current (A)*	*Y (cd/m^2)*	*x*	*y*
BLUGR	1.088	2.17	0.2942	0.5386
GREEN	1.004	1.65	0.2332	0.5821
ORANG	0.848	2.21	0.6111	0.3801
LIBLU	0.880	1.67	0.4450	0.4372
RED	0.790	1.76	0.6338	0.3420
REDOR	0.762	4.09	0.6227	0.3705
WHITE	0.831	3.89	0.5519	0.4065
AMBER	0.824	3.23	0.6015	0.3924

These values corresponded to a filter setup that included one layer of neutral density filter inserted behind the color filter tested. The two filters were mounted on the front of the bulb drawer unit. No other filter of any kind was included.

Appendix J: Theoretical Luminance Values

Color	*Brightness*	*Obtained Y* [1] *(cd/m^2)*	*Corrected Y (cd/m^2)*	*ΔY %.*
BLUGR	Low	2.17	2.16	0.5
	High	8.34	8.32	0.2
GREEN	Low	1.65	1.64	0.6
	High	7.40	7.43	0.4
ORANG	Low	2.21	2.28	3.2
	High	20.30	20.71	2.0
LIBLU	Low	1.67	1.67	0.0
	High	7.96	7.93	0.4
RED	Low	1.76	1.83	3.9
	High	15.30	15.74	0.9
REDOR	Low	4.09	4.25	3.9
	High	48.00	49.23	2.6
WHITE	Low	3.89	4.08	4.9
	High	20.80	21.40	2.9
AMBER	Low	3.23	3.34	3.4
	High	29.60	30.16	1.2

[1] Luminance values obtained in the preliminary experiment.

Appendix K: Luminances and CIE 1931 Chromaticity Coordinates of the Best Matches to the Theoretical Values

Color	*Brightness*	*Matched Y* *(cd/m^2)*	*ΔY %*	*Matched x*	*Matched y*
BLUGR	Low	2.16	0.00	0.2759	0.5490
	High	8.31	-0.12	0.2736	0.5450
GREEN	Low	1.65	0.61	0.2245	0.6086
	High	7.44	0.13	0.2284	0.6204
ORANG	Low	2.29	0.44	0.5765	0.4151
	High	20.80	0.43	0.5812	0.4160
LIBLU	Low	1.67	0.00	0.3699	0.4328
	High	7.96	0.38	0.3482	0.4358
RED	Low	1.84	0.54	0.6051	0.3345
	High	15.60	-0.90	0.6052	0.3472
REDOR	Low	4.25	0.00	0.5905	0.3928
	High	49.30	0.14	0.5824	0.4051
WHITE	Low	4.08	0.00	0.469	0.4311
	High	21.40	0.00	0.4727	0.4223
AMBER	Low	3.33	-0.30	0.5604	0.4306
	High	30.10	-0.20	0.5614	0.4325

Appendix L: Transformation of the CIE 1931 Chromaticity Coordinates to CIE 1976 $L^*a^*b^*$ and $L^*u^*v^*$ Coordinates

Color	*Brighness*	*Y*	*x*	*y*	*L**	*a**	*b**	*u**	*v**
BLUGR	Low	2.16	0.2759	0.5490	16.31	-31.95	1.99	-28.37	4.77
	High	8.31	0.2736	0.5450	34.62	-50.12	1.92	-60.43	9.52
GREEN	Low	1.65	0.2245	0.6086	13.53	-38.81	4.23	-29.00	5.55
	High	7.44	0.2284	0.6204	32.79	-64.22	9.97	-70.12	14.81
ORANG	Low	2.29	0.5765	0.4151	16.94	11.55	34.95	18.00	5.03
	High	20.80	0.5812	0.4160	52.73	24.73	86.93	57.88	16.39
LIBLU	Low	1.67	0.3699	0.4328	13.65	-10.26	-4.40	-10.20	-0.30
	High	7.96	0.3482	0.4358	33.90	-21.65	-10.05	-31.32	-1.60
RED	Low	1.84	0.6051	0.3345	14.62	23.88	10.69	30.62	-1.06
	High	15.60	0.6052	0.3472	46.45	44.79	29.32	90.86	0.21
REDOR	Low	4.25	0.5905	0.3928	24.48	19.24	35.42	33.61	5.37
	High	49.30	0.5824	0.4051	75.64	37.07	88.06	90.39	19.83
WHITE	Low	4.08	0.4690	0.4311	23.93	-0.55	9.17	1.03	3.72
	High	21.40	0.4727	0.4223	53.38	1.88	13.47	6.60	6.49
AMBER	Low	3.33	0.5604	0.4306	21.32	9.35	39.33	17.23	7.12
	High	30.10	0.5614	0.4325	61.74	19.17	88.33	49.59	21.26

Appendix M: Excitation Purity and Dominant Wavelength of the Experimental Colors

Color	*Excitation*[1] *Purity (%) Low Brightness*	*Dominant Wavelength (nm)*	*Excitation*[1] *Purity (%) High Brightness*	*Dominant Wavelength (nm)*
BLUGR	40	511	40	510
GREEN	53	513	54	515
ORANG	97	591	99	591
LIBLU	18	501	23	500
RED	59	637	68	615
REDOR	89	596	93	593
WHITE	30	581	28	584
AMBER	93	589	99	588

[1] A CIE Standard Illuminant A was used as the achromatic stimulus (Wyszecki and Stiles, 1982).

Appendix N: Filter Combinations and Currents in the Bulbs

Color (1 layer)	*Bright-ness*	*Filters behind Color Filter*				*Light Blue on Mirror*	*Filters in Front of Bulb*	*Current (A)*
		Light Yellow	*Yellow*	*Blue*	*Neutral*			
BLUGR	Low	-	1	-	-	1	-	1.231
	High	1	-	-	1	-	-	1.332
GREEN	Low	1	1	-	-	-	-	1.060
	High	-	2	-	-	-	-	1.296
ORANG	Low	1	-	1	1	1	2	1.155
	High	-	-	-	-	1	-	1.294
LIBLU	Low	-	-	1	1	-	2	1.046
	High	1	-	-	1	1	-	1.210
RED	Low	-	1	2	-	-	-	0.896
	High	1	2	2	-	-	-	1.192
REDOR	Low	-	-	2	1	1	-	1.022
	High	1	-	2	-	1	-	1.256
WHITE	Low	2	-	1	1	1	-	0.865
	High	-	-	-	1	1	-	0.947
AMBER	Low	-	1	1	1	1	-	1.046
	High	1	-	-	1	1	-	1.264

---

[1] The White color filters have not been used for the White color. Only the other filters mentioned in the table were used. The neutral density filter was put in the front of the transparent filters in the case of White Low.

Appendix O: Contrast Ratios

Color	*Brightness*	*Y* *(cd/m^2)*	*Contrast Ratio* [1] *(L_{max}/L_{min})*	*Display/Scene Contrast Ratio* [2]	*ΔE* [3] *Standard Yu'v'*	*ΔE* [4] *Rescaled Yu'v'*
BLUGR	Low	2.16	432	0.48	2.158	158.474
	High	8.31	1662	1.85	8.306	158.748
GREEN	Low	1.65	330	0.37	1.651	161.201
	High	7.44	1488	1.65	7.436	161.564
ORANG	Low	2.29	458	0.51	2.290	162.031
	High	20.80	4160	4.62	20.796	162.618
LIBLU	Low	1.67	334	0.37	1.666	154.815
	High	7.96	1592	1.77	7.955	155.386
RED	Low	1.84	368	0.41	1.847	172.341
	High	15.60	3120	3.47	15.596	171.050
REDOR	Low	4.25	850	0.94	4.248	164.843
	High	49.30	9860	10.96	49.295	163.419
WHITE	Low	4.08	816	0.91	4.076	156.198
	High	21.40	4280	4.76	21.395	156.598
AMBER	Low	3.33	666	0.74	3.328	160.245
	High	30.10	6020	6.69	30.095	160.422

---

[1] $L_{min} \approx 0.005\ cd/m^2$

[2] Luminance of roadway scene = 4.5 cd/m^2

[3] Standard: $\Delta E = ((\Delta Y)^2 + (\Delta u')^2 + (\Delta v')^2)^{0.5}$, Background was black (i.e., x = 0.3333, y = 0.3333)

[4] Rescaled: $\Delta E = (((155/Y_{max})\ \Delta Y)^2 + (367\ \Delta u')^2 + (167\ \Delta v')^2)^{0.5}$, (see Lippert, 1986)