JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 44, NUMBER 6 JUNE, 1954

Chromaticity-Confusion Contours in a Complex Viewing Situation*

RITA M. HALSEY† AND A. CHAPANIS
Psychological Laboratories, The Johns Hopkins University, Baltimore, Maryland
(Received February 4, 1954)

This experiment measured confusion contours for 58 standard colors distributed throughout the CIE constant-luminance diagram. Matches were made to these standards from an assortment of 342 heterogeneous colors arranged in a display which presented 171 of them simultaneously. The viewing conditions approximate those found in certain complex display situations. Twenty subjects were used in the experiment. Each subject was required to indicate those colors on the display board which provided satisfactory matches for each of the standard colors. Contours on the CIE diagram are drawn to show the percentage of times various chromaticities were confused. From these contours we have selected colors which can be discriminated with high levels of accuracy and so are suitable for coding qualitative and quantitative information. In general, the contours follow trends suggested by extrapolation from precise threshold data.

I. INTRODUCTION

A. Background of the Problem

COLORS are used routinely for coding information of various sorts in our everyday life. Not only do they direct our land, sea, and air traffic, but they also do such ordinary jobs as identifying electrical connections, the contents of gas cylinders, and other commercial products. So useful has color been for conveying information that designers of control equipment are considering it for still other and more complex purposes. For example, colors might be used to provide supplementary information on the visual displays in air traffic control centers. In such an application, colors could add a third dimension, altitude, to present two-dimensional radar scopes and plotting boards, or they could be used for coding qualitative information, e.g., type of aircraft.

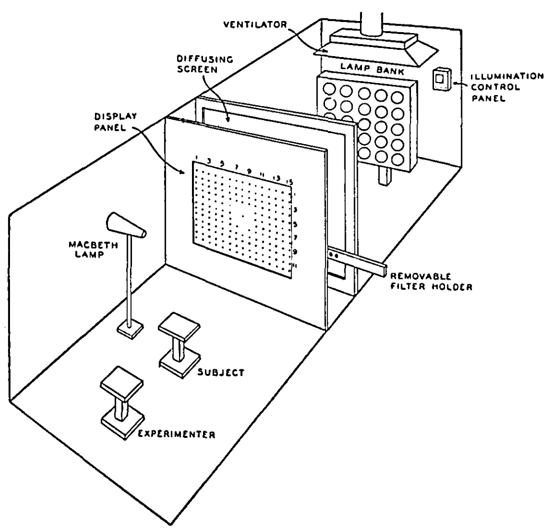

FIG. 1. Highly schematic view of the apparatus and experimental rooms.

Most laboratory studies of color discrimination do not yield data which are directly applicable to such practical situations. In the first place, the typical laboratory study of color discrimination uses ideal viewing conditions, i.e., they provide for maximum discrimination. In addition, the usual criterion for discrimination is the *JND*, or difference which is seen 50 percent of the time. But in practical coding situations it is frequently important to select colors which will never, or rarely, be confused with each other, even when many colors may be present and when the viewing conditions are subject to the contaminating effects of chromatic contrast, after-images, and so on. Considerations such as these make the practical man hesitant about selecting colors for color-coding purposes on the

TABLE I. CIE chromaticity coordinates of the 58 standard colors used in this study.

Filter number	CIE coordinates x	y	Figures containing data	Filter number	CIE coordinates x	y	Figures containing data
6	0.712	0.288	10, 11, 12	184	0.129	0.345	10
20	0.600	0.248	4	186	0.062	0.337	9
29	0.613	0.331	3, 11	193	0.242	0.507	10, 12
43	0.619	0.381	4	201	0.290	0.414	5
63	0.510	0.469	5	206	0.187	0.437	6
85	0.443	0.422	3	214	0.188	0.336	7, 11
95	0.458	0.389	7, 12	216	0.271	0.310	3, 12
100	0.506	0.363	6, 11	220	0.102	0.234	3, 12
101	0.519	0.412	9	225	0.196	0.203	4, 11
104	0.570	0.407	10	230	0.099	0.155	6, 11
110	0.392	0.448	8	233	0.192	0.134	5
113	0.350	0.510	8, 11	236	0.124	0.103	8, 11
117	0.408	0.550	4	240	0.164	0.098	9
124	0.355	0.592	6	244	0.130	0.078	7
132	0.305	0.615	3	251	0.168	0.055	10
152	0.228	0.743	7	256	0.146	0.044	4
160	0.087	0.748	8, 11	268	0.162	0.014	3, 11
167	0.169	0.645	9	276	0.223	0.039	9, 12
172	0.026	0.556	4	283	0.267	0.087	5
178	0.106	0.438	5	285	0.308	0.125	4, 11
288	0.220	0.122	6	317	0.560	0.206	8, 11
290	0.263	0.154	8	320	0.511	0.237	7
291	0.326	0.230	3	321	0.509	0.280	8
294	0.356	0.291	4	324	0.524	0.343	5
300	0.371	0.103	7	332	0.590	0.324	7
304	0.471	0.225	10	335	0.571	0.271	9
306	0.414	0.222	9, 11, 12	337	0.325	0.351	10
310	0.457	0.156	3	340	0.373	0.384	9, 11
313	0.500	0.183	6	342	0.387	0.308	8, 11

* This research was done under Contract N5-ori-166, Task Order I, between the U. S. Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-179, Project Designation No. NR 145-089, under that contract.

† Present address: U. S. Naval Medical Research Laboratory, U. S. Naval Submarine Base, New London, Connecticut.

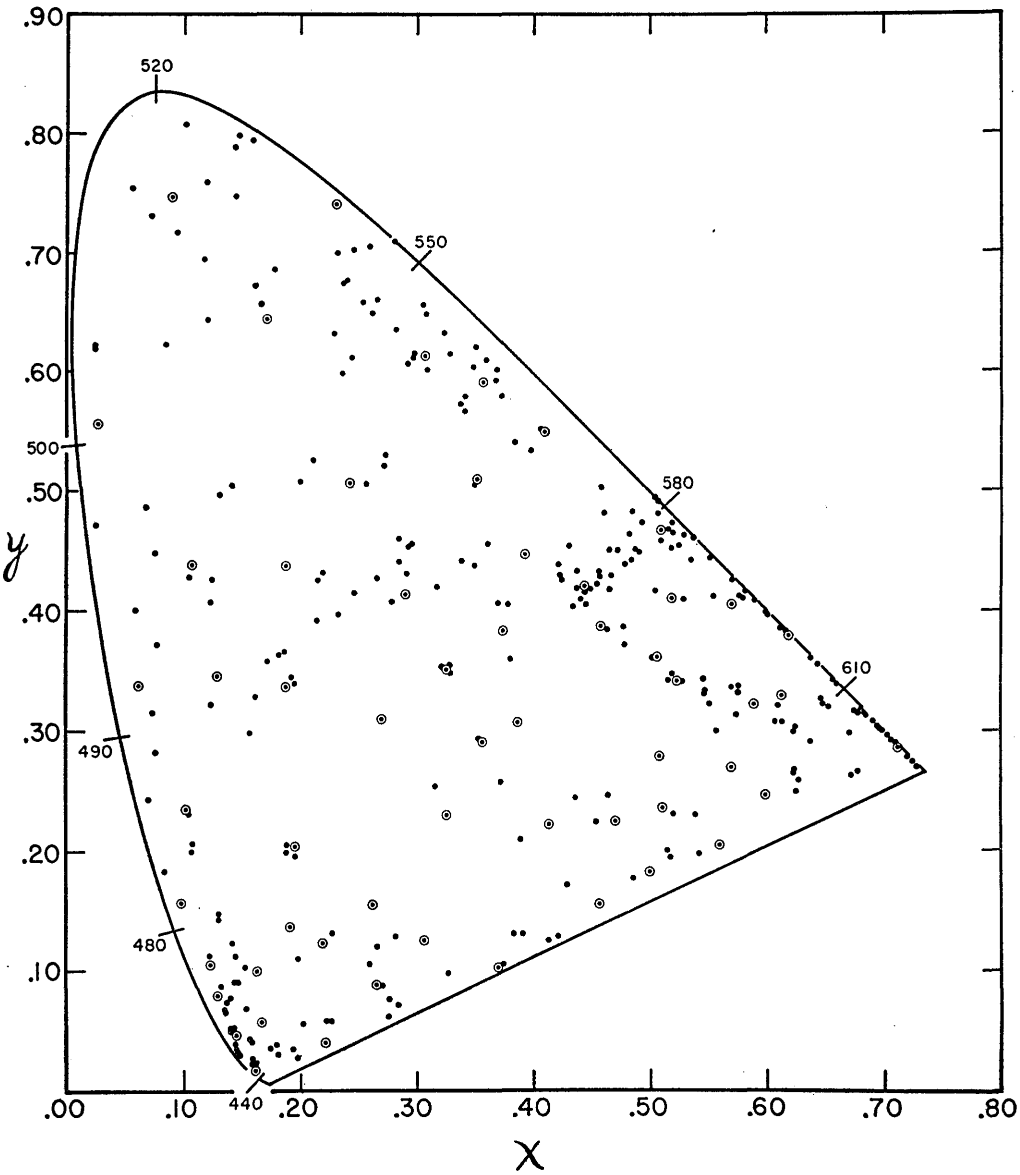

FIG. 2. Locations of the 342 colors used in this study. The 58 standard colors are identified by the circled points. We have not always plotted separate points for colors with nearly identical CIE coordinates.

basis of laboratory data such as those obtained by MacAdam.[1]

The more practical experiments on color discriminability, those by Hill, Holmes, and McNicholas,[2] also have their disadvantages. In the first place, these studies are all concerned with the identification of single, isolated spots of color. In addition, the subjects were usually shown a colored light and were asked to assign it to one of a few color categories, e.g., red, yellow, green, blue, or white. But, as Holmes has pointed out, this forces the subject to group colors which might easily be discriminable. Thus, although a bluish-green and a yellowish-green light might be clearly different, the subject would be forced to call them both "green." In a sense, these earlier studies determined the range of hues which will be accepted under certain labels. For practical coding purposes, however, the colors would be completely divorced from particular names and might, for example, be assigned arbitrary numerals or other sorts of designations.

B. Purpose of this Experiment

The purpose of this experiment was to determine color confusion contours when many heterogeneous stimuli, spatially separated, are in the field of view at the same time. These conditions approximate those

[1] D. L. MacAdam, J. Opt. Soc. Am. 32, 247 (1942).

[2] N. E. G. Hill, Proc. Phys. Soc. (London) 59, 560 (1947); J. G. Holmes, Trans. Illum. Eng. Soc. (London) 6, 71 (1941); J. G. Holmes, in Fischer, Schaeffer, and Sorsby (Editors) *Documenta Ophthalmologica* (Uitgeverij Dr. W. Junk 'S-Gravenhage, 1949), Vol. 3, pp. 240–250; H. J. McNicholas, J. Research Natl. Bur. Standards 17, 955 (1936).

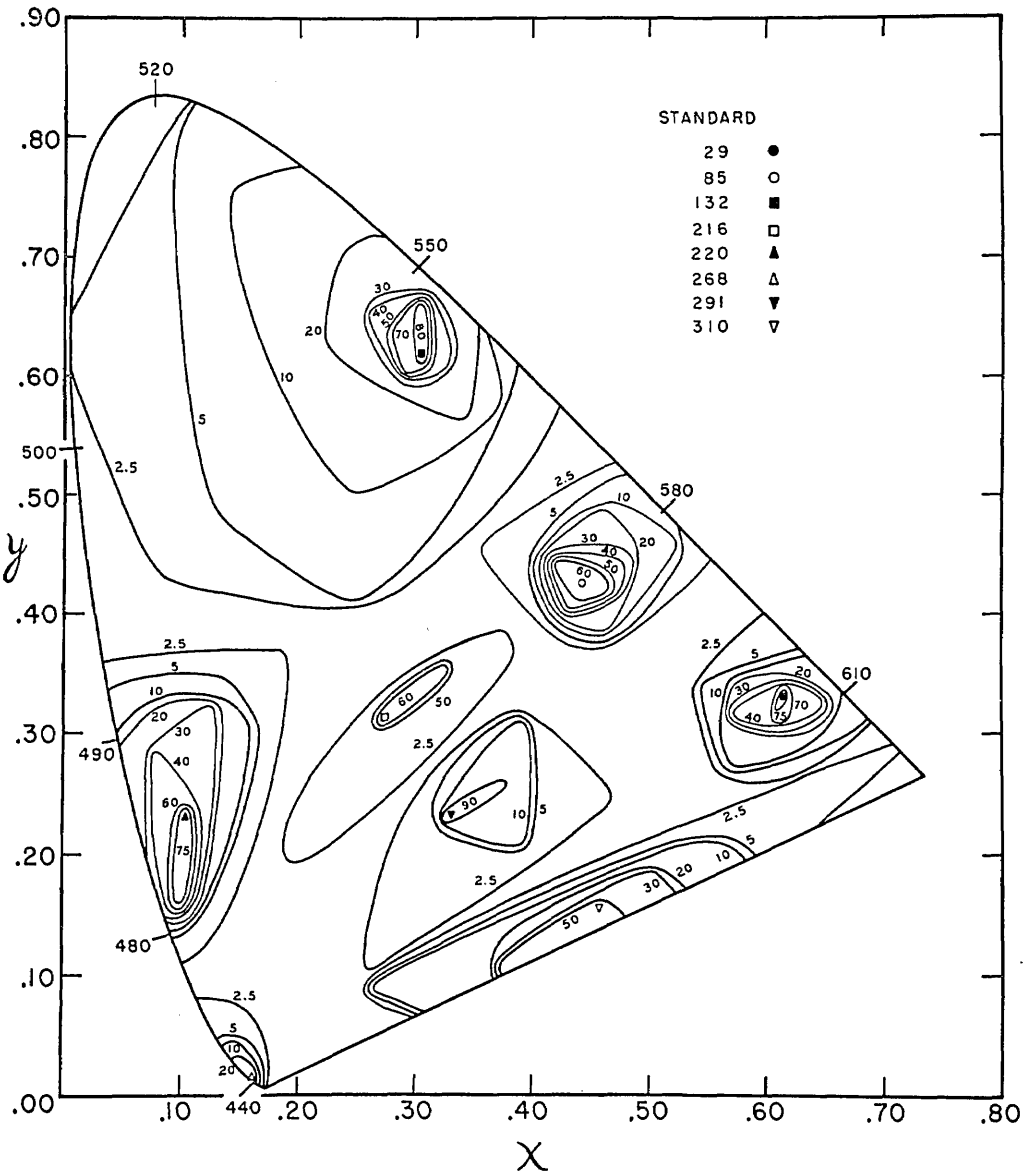

FIG. 3. Confusion contours for eight standards. Each line encloses colors which were confused with the standard a certain percentage of the time. A contour line labeled 40, for example, encloses all those colors which were selected as matches for the standard 40 percent or more of the time.

which may be found in certain complex display situations. This study is the second in a series directed toward the general question: What is the maximum number of colors which can be correctly identified when they are divorced from particular color names? The first study[3] was concerned with the maximum number of spectrum colors which could be easily identified.

II. APPARATUS

A. Over-All Plan of the Apparatus

Figure 1 is a highly schematic view of the apparatus. The entire assembly occupies two rooms with a display panel built into the wall separating the rooms. One room contains a bank of 25 General Electric, 300 watt-105 volt, reflector flood lamps illuminating a large diffusing screen. Light from the diffusing screen passes through a rack of filters behind the display panel and out into the observation room. The apertures producing the spots of color are $\frac{3}{8}$ in. in diameter and there is a separation of 3 in. between the centers of the apertures. The display panel is approximately 45 in. wide by 37 in. high and can accommodate 180 colors at a time.

The lamps used in this experiment were all carefully selected, aged, and calibrated for color temperature.[4]

[3] Rita M. Halsey and A. Chapanis, J. Opt. Soc. Am. 41, 1057 (1951).

[4] A new color temperature meter was designed especially for this purpose [see King, Snider, and Hamburger, J. Opt. Soc. Am. 42, 178 (1952)].

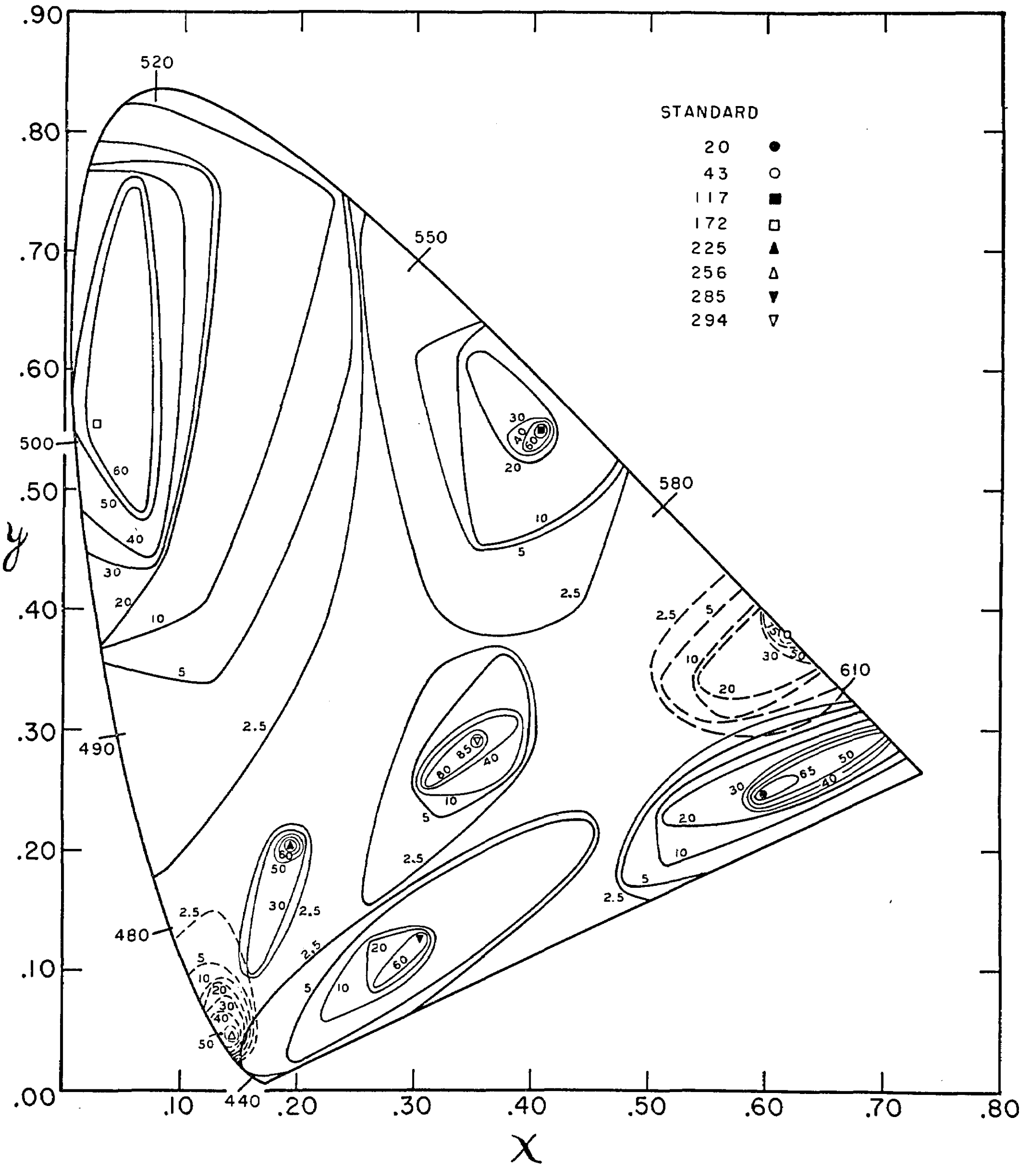

FIG. 4. Confusion contours for eight standards.

To take account of variations in the color temperature of individual lamps, the lamps in any one row of 5 were selected from a much larger lot to yield the same color temperature when operated at approximately the same voltage. Each bank of 5 lights was independently controlled by a continuous voltage transformer and monitored through a control panel and voltmeter. The voltage calibration was designed to produce a color temperature of 2700°K through the diffusing screen. A running-time meter recorded how long the lamps had been in operation, and the lamps were replaced periodically to take account of aging effects on color temperature.

Two Macbeth Daylight lamps, Style ALS-30 (only one of which is shown in the illustration), flooded the test room with a low level of indirect light. The background of the display panel was a neutral grey, roughly equivalent to Munsell N6/. The illumination, provided by the two Macbeth lamps on the background of the display panel, was 0.014 millilambert.

The subject sat 6 feet from the display panel. At this distance each colored spot subtended a visual angle of 18′ and the visual angle of the whole display panel was roughly 35°.

B. Filters

There was a total of 342 colored filters of which 58 were designated standards. The CIE specifications of the filters and standards are shown in Fig. 2, and the CIE specifications of the standards are listed in Table I. We tried to get as uniform coverage of CIE space as was possible with gelatin and glass filters. That we

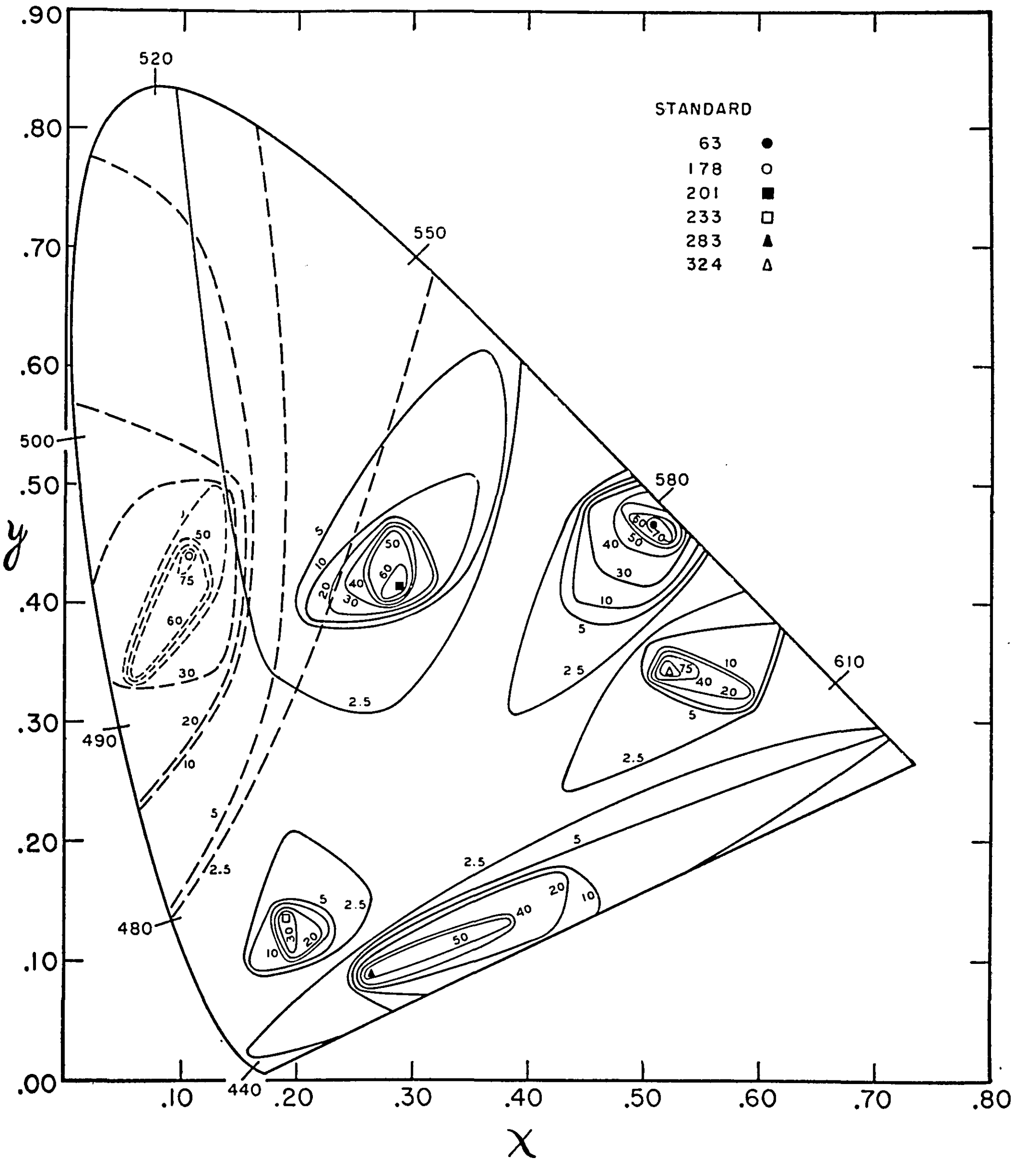

FIG. 5. Confusion contours for six standards.

were not entirely successful is evident from the sparsely populated areas in Fig. 2. Our filters include virtually all of the standard Corning and Wratten series plus combinations of these, theatrical gelatins, and such other filters as we could locate for our purposes.

The brightnesses of all the filters were equated to that of the darkest filter, producing a luminance of 2.7 millilamberts. This equation was done visually and was checked by several persons. Neutral film of various densities was added to each filter until the desired brightness match was obtained. Following the brightness calibration of the filter combinations, all of the filters were measured with the GE recording spectrophotometer.[5] Transmission curves were obtained for all filters and neutral films, in combination whenever possible. However, when the transmission values were too low to be reliable, separate transmission curves were measured for each of the filter components. The CIE coordinates were calculated by the weighted ordinate method, using 31 points from 400 mμ to 700 mμ at 10-mμ intervals. For certain filters which had very low transmission values or very narrow distributions of transmission, however, calculations were made at 5-mμ intervals. All calculations were made independently by two computers.

[5] We are indebted to Mr. L. C. Fitzgerald and Mr. Elmer Bartlett of the Chevrolet-Baltimore Division of the General Motors Corporation and to Dr. George S. Strong of PEMCO, for permission to use their equipment, and to Mr. Michael Davis (also of GM) and Mr. A. E. Snider (of our laboratories) for their painstaking care in running the curves.

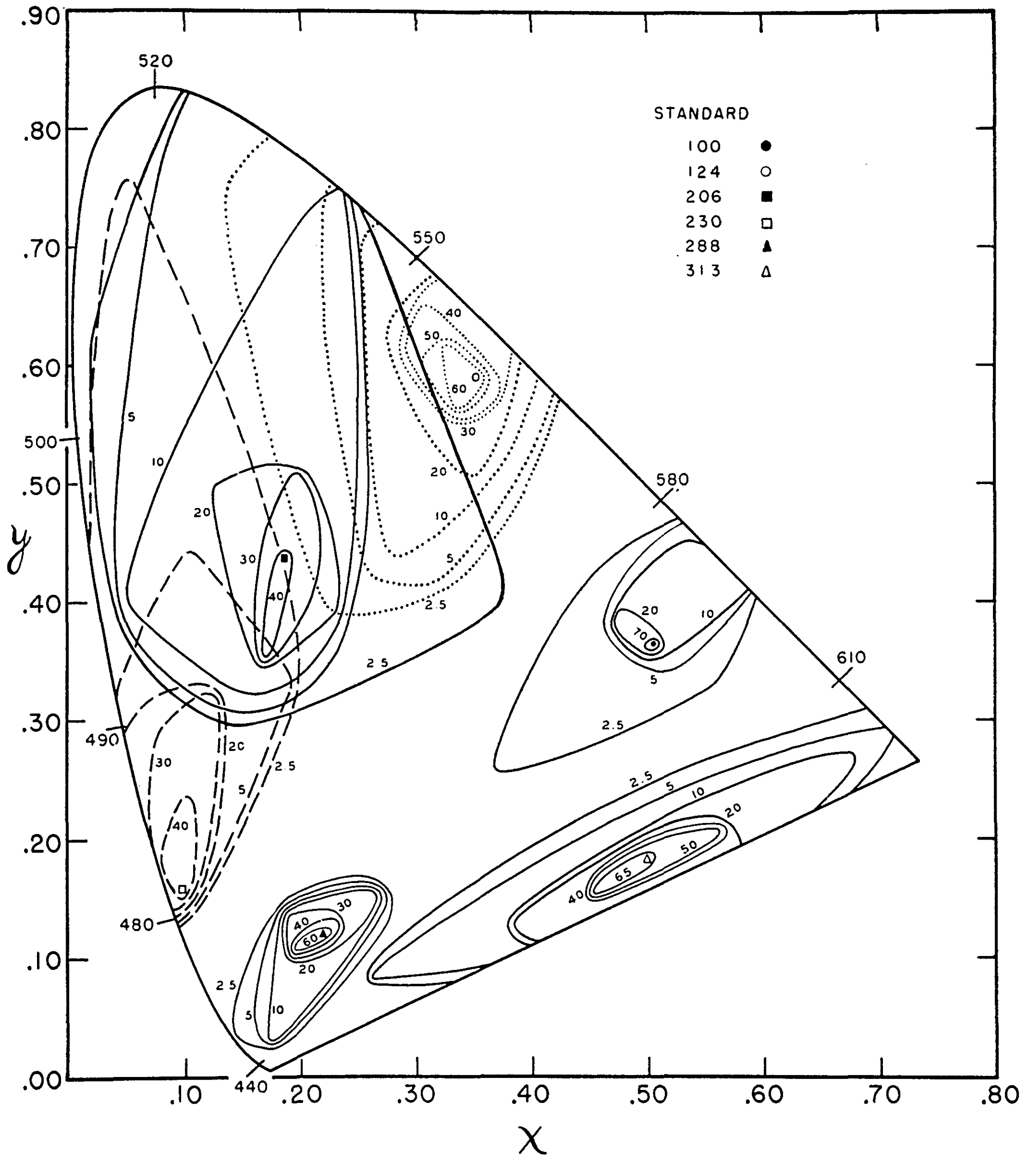

FIG. 6. Confusion contours for six standards.

III. SUBJECTS

Twenty subjects completed the experiment. They were 18 men and 2 women, between the ages of 16 and 35. A large battery of pseudo-isochromatic plates established that all had normal color vision. Each was also given an acuity test with the Bausch and Lomb Ortho-Rater. Fifteen subjects had normal acuity, two wore corrective glasses during the experiment, and three had slightly subnormal visual acuity. The subjects constituted a heterogeneous group as regards their previous experience with color and with psychophysical procedures. One subject (the second author) was familar with the recent work on color discrimination and with the design of the experiment. The majority of the subjects, however, were unsophisticated in this regard. At most, they had "heard about" some color variables from a wide variety of sources, such as courses in physics, psychology, art, and photography. Some of the subjects had some previous experience with psychophysical experiments, although only one was a trained observer.

IV. PROCEDURE

For every subject there were four displays, each consisting of 171 colored lights assigned randomly to the available positions on the display panel. (The eight apertures surrounding the central position were not used; they were masked out to match the background of the panel.) Thus, each color appeared twice for every subject. The standards were presented one at a time in the central position on each display. The subjects' task was to select, from all the colors on the

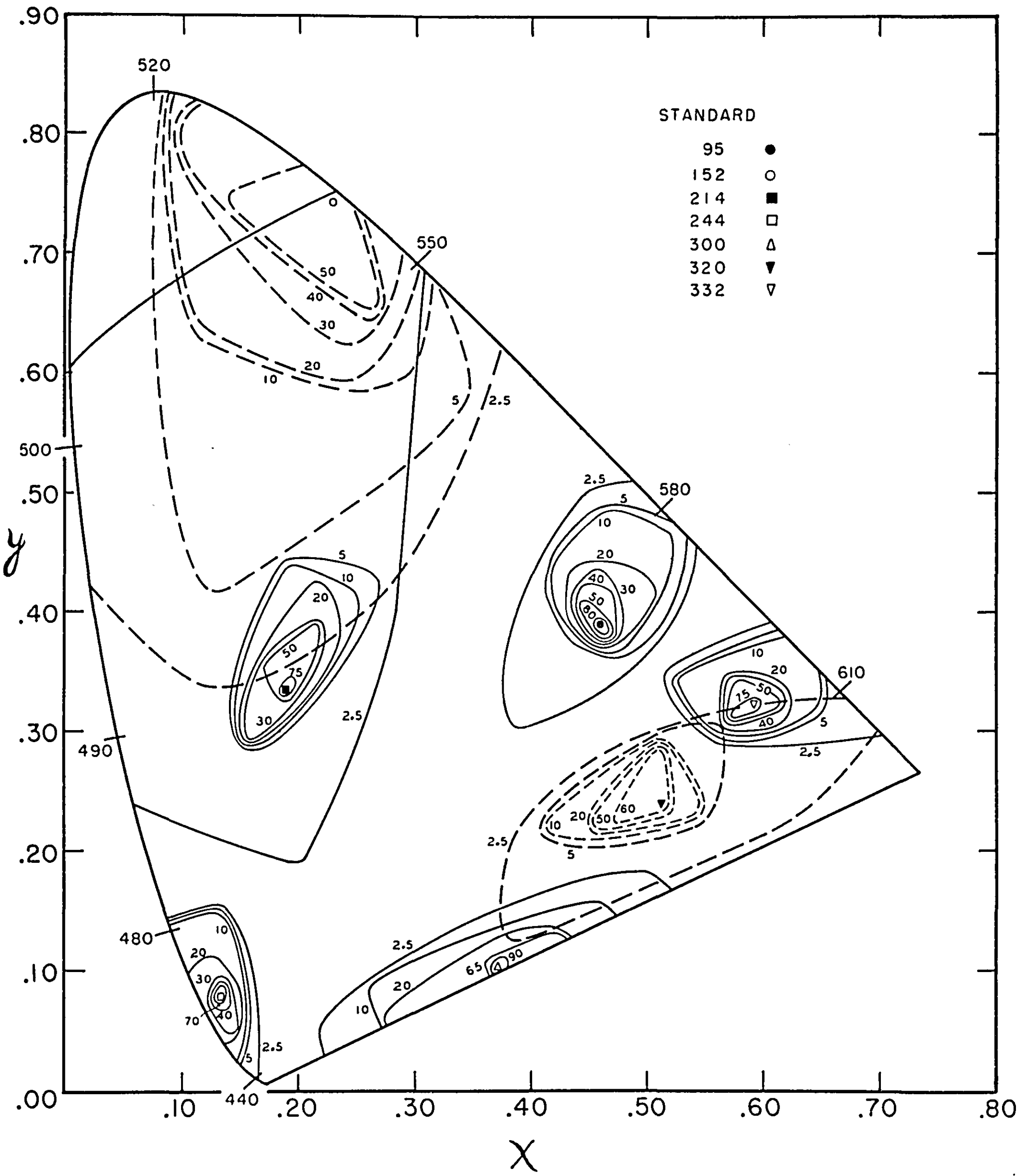

FIG. 7. Confusion contours for seven standards.

display, those which satisfactorily matched each standard. Every standard was compared with every other color twice. Subjects were given unlimited time to make their matches and were carefully instructed to scan each row of lights systematically. Fixation was not controlled. The 80 displays used were all different. A complete set of observations for any subject required from 12 to 24 hours distributed over several days.

V. RESULTS

A. Confusion Contours

The data were analyzed by plotting on the CIE diagram each standard and all of the colors which were matched to it. In addition, the proportion of times that every color was matched to the standard was recorded on the diagram. (The maximum possible number of such matches was 40, that is, 20 subjects each making 2 comparisons between every standard and every other color.) Contours were then drawn to include all those colors which were matched with every standard a certain proportion of the time. The results are shown in Figs. 3 through 10.

The contours shown in Figs. 3 through 10 are empirical contours. They were drawn just outside those colors which were matched with the standard a certain proportion of the time. For example, a contour line labeled 40 encloses all those colors which were selected as matches for the standard 40 percent or more of the time. All such colors lie at least just inside the contour line. Two other arbitrary rules were followed in constructing the contour lines. These are (1) contour lines were never drawn concave; (2) a contour line which

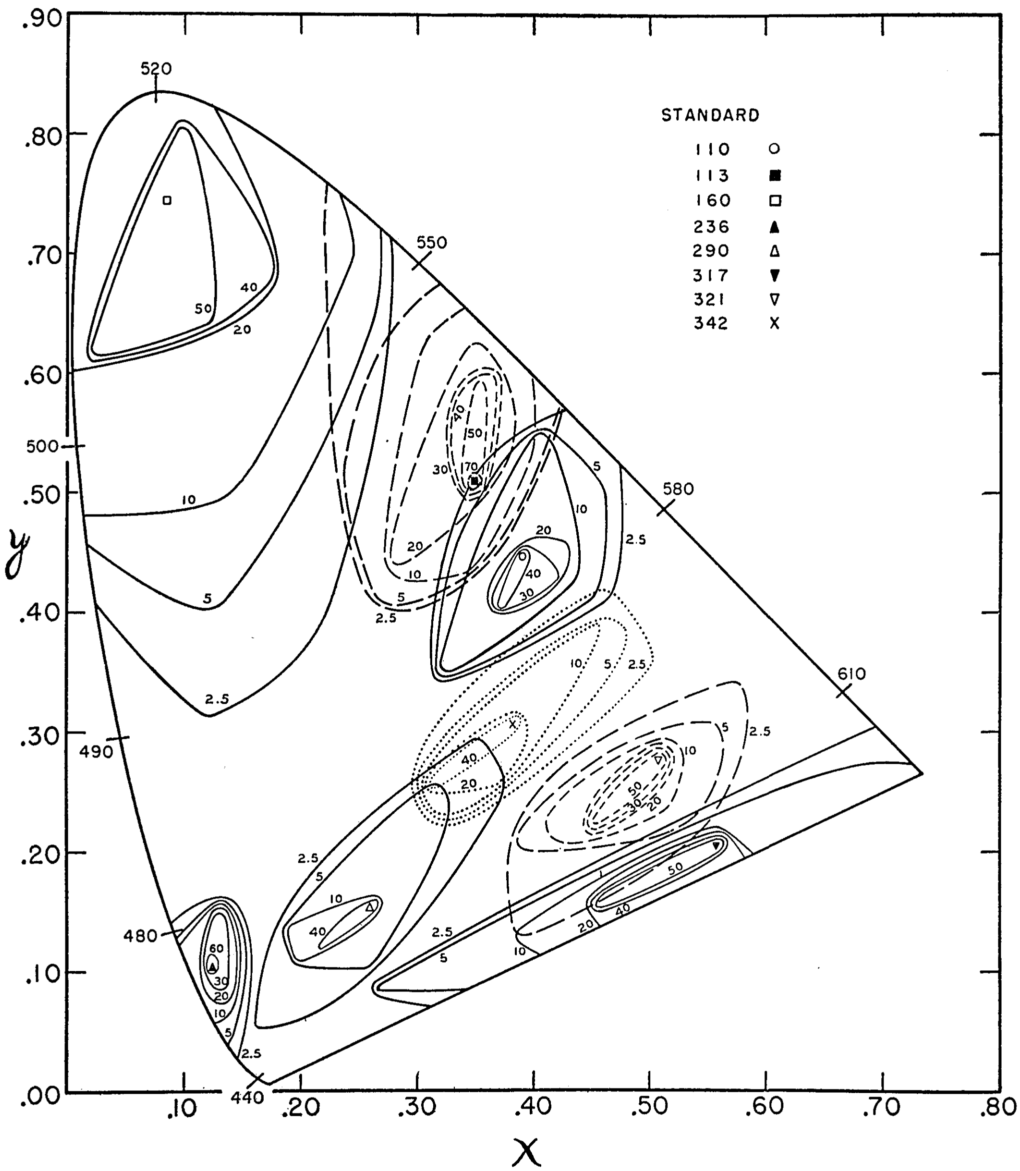

FIG. 8. Confusion contours for eight standards.

indicates a lower frequency of confusion with the standard was always drawn outside those indicating higher frequencies.

All of the original data are represented in Figs. 3–10. Smoothing of these rough contours and interpolation between them can be made by superimposing Fig. 2 upon Figs. 3–10. In many cases, the irregular shapes of our contours are due primarily to the scarcity of colors in certain regions of the CIE diagram where one might expect confusions to occur. For this reason some smoothing and common-sense extrapolations are undoubtedly justified for certain applications.

B. Selection of Coding Colors

Unfortunately there is no sure way of knowing which set of contours is most applicable to specific coding situations. The safest set of colors for coding would be those which show no overlap at all among the confusion contours. Such a set of standards is shown in Fig. 3. No colors were ever confused with two or more of these 8 standards. Other sets of colors can, of course, be selected from Figs. 3 through 10 so that there is no overlap of the 2.5 percent contours. For example, standards 20, 117, 172, 225, 285, 294, and 324 constitute another very highly discriminable set. Still another set is 43, 113, 184, 233, 256, 304, and 342. Eight is the maximum number of standards we have been able to find with no overlap of the 2.5 percent contours. However, if (1) more colors had been used in this experiment, (2) different standards had been used, and (3) interpolation were permitted, we estimate that 10

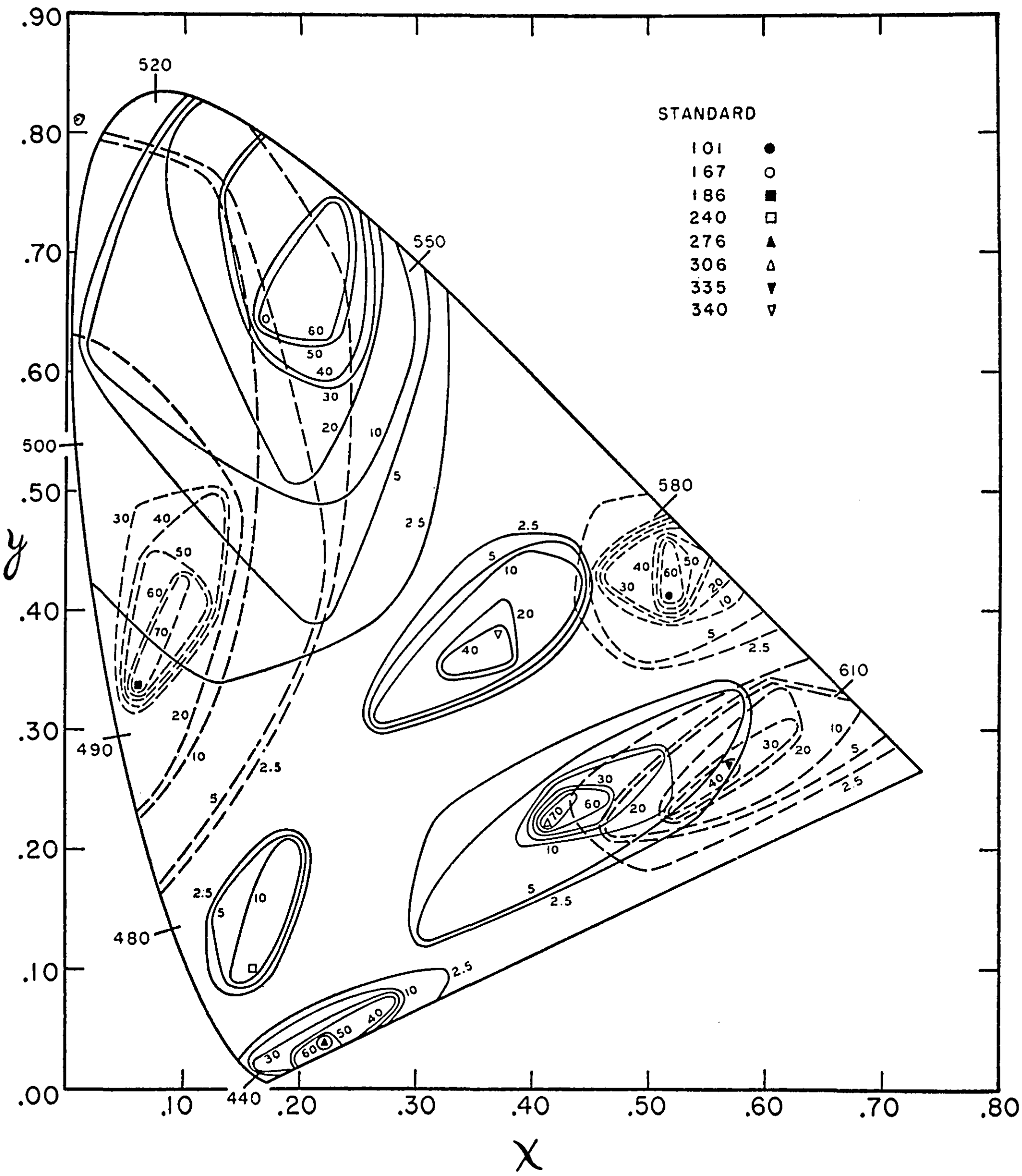

FIG. 9. Confusion contours for eight standards.

colors could be found to yield no confusions at the 2.5 percent level.

There are 11 nonoverlapping 5 percent contours. Two sets of standards discriminable at this level are: 6, 101, 160, 201, 220, 225,[6] 236, 276,[6] 291, 332, 340; and 20, 43, 63, 117, 172, 214, 216,[6] 233, 244, 276,[6] 291. If more complete data were available, we believe that 14 colors could probably be used at the 5 percent level.

We believe that the 10 percent contours are the best and most applicable to general coding situations. These are the most regular and stable of all our contours. They indicate a high level of discriminability and are generally reasonable in the light of such other data as those of MacAdam[1] and Hill.[2] The 2.5 percent contours include those colors which were matched to a standard only once, the 5 percent contours those which were matched twice. Since it may well be that such occasional matches were the result of carelessness, we are inclined to place little confidence in the exact shapes of these low-error contours.

The maximum number of standards found at the 10 percent level of discriminability is 15, although we believe that 18 may be the theoretical maximum. One such set of standards is shown in Fig. 11. Another set of 10 percent contours surround these standards: 6, 63, 110, 124, 172, 214, 230,[7] 240, 256, 285, 290, 317, 321, 332, and 342. Still another set is 6, 43, 63, 117, 178, 193, 216,[6] 233, 236, 268, 285, 294, 306, and 313.

[6] Since there were no contours at the desired level for these standards, we have used 2.5 percent contours in deriving these series.

[7] Twenty percent contours used for these standards.

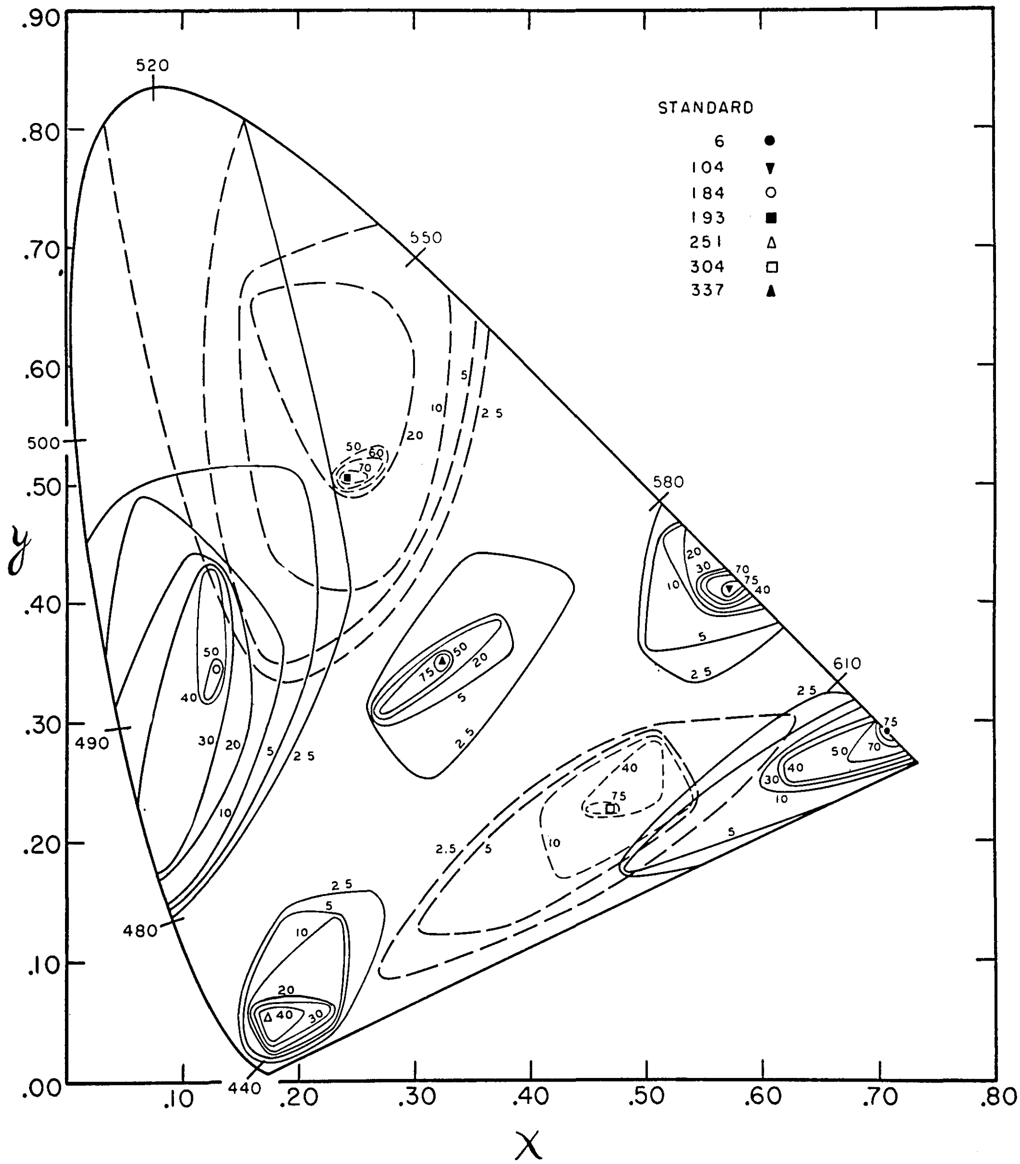

FIG. 10. Confusion contours for seven standards.

Contours for percentages greater than 10 are generally quite small and often irregularly shaped. However, they may be useful for some purposes. Nineteen standards can be selected at the 20 percent level (estimated maximum: 24). One such set is: 6,[8] 43, 100, 101, 110, 117, 152, 184, 201, 240,[8] 251, 256, 288, 294, 300, 306, 317, 335, and 337. At the 30 percent level the maximum number of standards is 28, with an estimated maximum of 35. One such series is: 6, 29, 43, 63, 100, 104, 110, 113, 117, 132, 167, 172, 201, 214, 220, 225, 233, 236, 251, 256, 268,[7] 285,[7] 290,[9] 294,[8] 300,[7] 306, 317,[7] and 324.[7] Contours at higher levels are too unreliable to be useful.

The percentages of confusion represented by these data probably have no absolute meaning for viewing conditions differing from those used here. Our displays presented colors at varying spatial separations—they were sometimes close to each other, sometimes far away. Contrast effects, dazzle, and after-images all undoubtedly contributed heavily to the unreliability of any particular match. At times, grossly different colors looked similar; at other times, identical colors looked different. As a result, no two colors (even those identical in CIE specifications) were ever matched 100 percent of the time. Generally speaking, the best matches for each standard were made only about 90 percent of the time.

[8] Ten percent contours used for these standards.
[9] Forty percent contours used for these standards.

C. On the Use of Colors for Coding when Names Must Be Used

We designed this experiment on the assumption that when colors are to be used for practical coding situa-

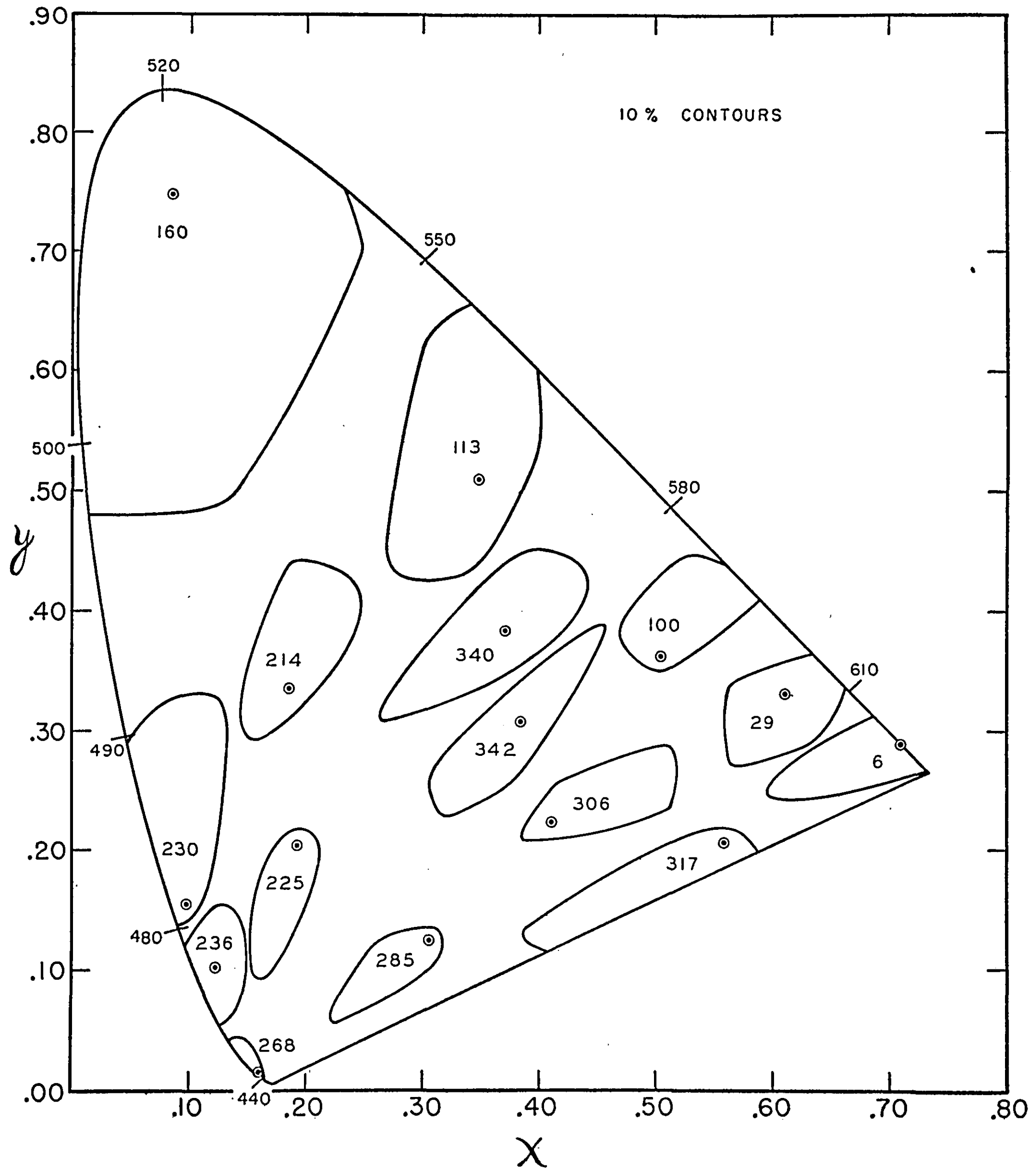

FIG. 11. Ten percent contours for 13 standards. Each contour encloses colors which were confused with the standard 10 percent or more of the time. Since there were no 10 percent contours for filters 225 and 230, those shown here are the 2.5 and 20 percent contours, respectively.

tions, they would be assigned arbitrary names such as numerals or code words. It is possible, of course, that in using the code people would apply their own color names to the colors. For example, the 8 colors in Fig. 3 can easily be assigned names which would be acceptable to most people: orange, yellow, green, white, greenish-blue, violet, pink, and reddish-purple.

It is more difficult to find simple names for the 11 colors at the 5 percent level of discriminability. However, the first set listed could be called: red, yellow-orange, green, whitish-green, greenish-blue, blue, violet, pink, orange, and yellowish-white. In view of the wide individual differences known to exist with regard to color nomenclature, it seems unlikely that there would be much agreement on the names assigned to the 15 or more colors selected from the 10 percent contours.

D. Individual Differences

There were large individual differences among subjects—differences in the size of the confusion contours, in the total number of matches selected, in the consistency with which matches were repeated, and in attitude toward the experiment. None of the differences correlate with differences among subjects in their previous experience with color or with psychophysical procedures. Nor do the perceptual differences correlate with the subject's attitude toward the experiment—some of the most interested and conscientious subjects

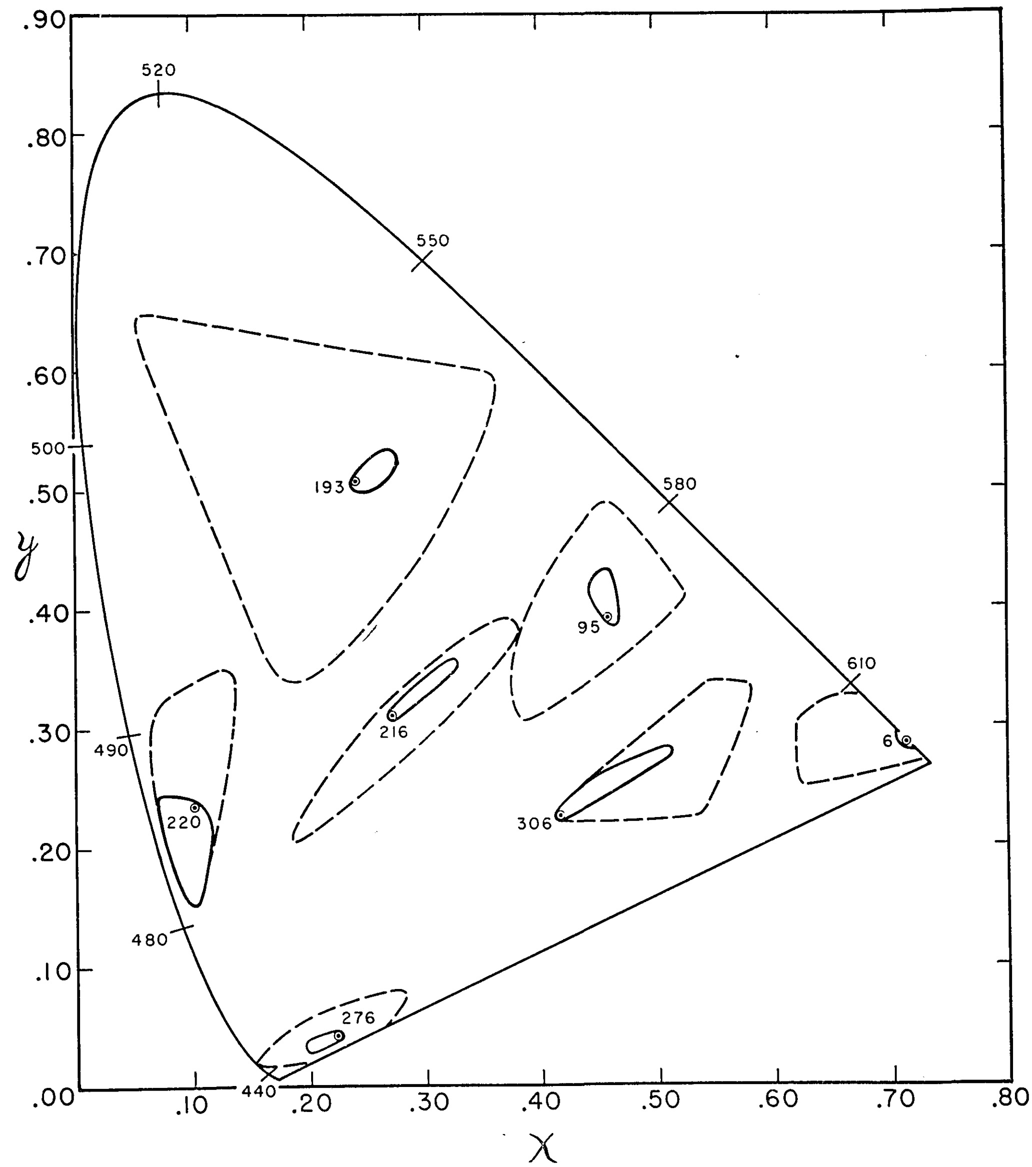

FIG. 12. Confusion contours for two different subjects. The lines enclose all colors which were ever confused with a given standard by that subject. The solid lines are for one of the most discriminating subjects; the dashed line for one of the least discriminating.

yielded the largest discrimination contours. Parenthetically, the junior author gave contours which were generally smaller than that of the median subject; however, he was far from the most discriminating subject.

An example of the differences in the size of contours is shown in Fig. 12. These are for one of the most discriminating and one of the least discriminating subjects. The majority of the subjects gave contours which fall somewhere between the two extremes shown in this figure. The relative sizes of the contours are quite consistent—a subject who produces large contours in one area of the CIE diagram produces large contours in all other areas. The contours for different subjects appear to vary in size only, not in shape, orientation, or location. Thus, we believe that the differences are not indications of different degrees or color deficiency or of physiological sensitivity. Rather, we believe that the differences represent different tolerance levels—the degree of precision the subject was set to undertake.

At the conclusion of the experiment, each subject was carefully interviewed to determine, if possible, how he made his judgments, and what his criteria of a match were. Despite the obvious differences in their behavior and data, all subjects asserted that they accepted a match only if it were perfect: "if the colors were identical," "if there was no difference between them," "if there was perfect agreement," "if there was no disparity in any characteristic," "if they were *exactly* the same."

An analysis of those matches which were made only once during the entire experiment (i.e., those determining the 2.5 percent contours) reveal some interesting facts. Of the 480 such unique matches, 66 percent were made by 3 subjects, the other 17 subjects contributing very few each. (These could have been due to reporting or recording errors, or to fatigue.) The same 3 subjects have the largest number of total matches and the largest confusion contours. They have nothing else in common, being in very different categories with respect to age, experience, and interest.

VI. DISCUSSION

We regard this study as only a start on the general problem of color coding. There are many other variables which might conceivably affect the general sizes of the contours we have found here. For example, variations in the size and luminance of the lights might easily affect the size of the discrimination contours. It is also possible that the use of different room illuminants would interact with the discrimination contours found here. Training the subjects might result in a considerable improvement of discrimination. Finally, reducing the total number of different colors available on the display might well have an effect on the general level of discrimination possible.

There are also other ways of attacking the problem of selecting colors for coding purposes. For example, it would be possible to use a scale of standards arranged in some sort of a series along one side of the display. This would be analogous to the use of the hypsometric colors put alongside geographic maps and charts. We cannot predict from our data how much effect such a series of standards would have on the general level of discrimination.

A more difficult kind of experiment could have been designed by requiring the absolute identification of certain colors rather than allowing comparative judgments as was done in this study. We have some reason to believe that the contours found here are probably valid for situations where the colors must be recognized in isolation, i.e., without reference to a standard. The senior author, who tested most of the subjects in this study, acquired considerable familiarity with the colors during the course of the experiment. Observations by her suggest that contours for absolute judgments, although larger in size, are similar in shape to those reported here. In another experiment, equivalent results were obtained by an absolute judgment method and the color-matching method over a limited color range. There can be no doubt, however, that if absolute judgments were to be used in practical situations, the observers would require extensive preliminary training.

As compared with precise threshold data such as those measured by MacAdam,[1] our confusion contours are very large indeed. This is undoubtedly accounted for by the conditions of our experiment. Ours is perhaps as confusing a visual display as will ever be found in practice. Colors must be compared when they are sometimes widely separated, and they are viewed under conditions which produce dazzle, after-images, successive contrast, and simultaneous contrast. These are, however, precisely the conditions under which colors of this sort would be seen if they were to be used on realistic displays such as are now contemplated.

Despite the obvious differences between our experiment and that of MacAdam, there are several points of similarity between his data and ours. Although precise comparisons are not justified because MacAdam's formulations apply to *small* color differences, the approximate relationship between the two sets of data is of interest. We have constructed ellipses, according to the MacAdam formulas,[10] for some of our standards and find fair agreement between these calculated ellipses and our contours when the latter have been smoothed to take account of the distribution of available colors. In general, our 10 percent contours are roughly 30 times as large as the MacAdam ellipses. In the red, orange, yellow, pink, and blue-violet regions the correspondence is good. Our green contours, however, are considerably wider, and our purple contours more elongated, than the MacAdam predictions. The most serious discrepancies occur in the highly saturated blue-green region, where MacAdam's extrapolations are based on very few data. In this region our contours are larger than MacAdam's and markedly different in the orientation of their axes. In addition, many of our contours are asymmetrical, being wider at the yellow or green end, as would be expected, however, from the progressive changes shown by the MacAdam ellipses for different chromaticities.

A final word of discussion is in order on the tritanopia of the central fovea as it relates to our data. Among others, Willmer and Wright[11] have shown that when it is stimulated with very small test fields—subtending 20′ of arc or less—the eye is unable to discriminate blue-green colors. In this and in other ways the central fovea seems to have the characteristics of a tritanopic form of color blindness. Our spots of light subtended only 18′ at the subject's eye yet our data contain little or no evidence of this central tritanopia. We believe that Willmer and Wright provide the answer to this paradox in their next-to-the-last paragraph, which is entirely concerned with the difficulty of making these observations. Thus, they say ". . . any deviation from direct fixation immediately caused a breakdown in the [tritanopic] matches obtained . . . ." The subjects in our experiment did not fixate any light; instead they were required by the very nature of the task to scan the display almost continually. These facts lead us to believe that the data reported here are not contaminated by central tritanopia.

---

[10] D. L. MacAdam, J. Opt. Soc. Am. **33**, 18 (1943).
[11] E. N. Willmer and W. D. Wright, Nature **156**, 119 (1945).