This is a preprint of an article published in COLOR research and application, 31(1):48-56, 2006.
http://www3.interscience.wiley.com/journal/35037/home

A Data Set for Fuzzy Colour Naming

Robert Benavente, Maria Vanrell, Ramon Baldrich

Computer Vision Center / Dept. Informàtica (UAB), Building O – CampusUAB
08193 Bellaterra (Barcelona), Spain

Abstract: Although colour naming has been studied from different points of view, the automation of this visual task has not been an active topic until recently. In computer vision, colour naming has been posed as a fuzzy-set problem where each colour category is modelled by a function that assigns a membership value to any given sample. However, the success in the automation of this process relies on having an appropriate psycho-physical data set for this purpose. Up to the present, such a dataset has not been available.
In this paper we present a data set obtained from a colour-naming experiment. In this experiment, we have used a scoring method to collect a set of judgements adequate for the fuzzy modelling of the colour-naming task. The data set is composed of 387 colour reflectances, their CIELab and Munsell values, and the corresponding judgements provided by the subjects in the experiment. These judgements are the membership values to the eleven basic colour categories proposed by Berlin and Kay[1]. All these data have been made available online (http://www.cvc.uab.es/color_naming) and, in this paper we provide a wide analysis of it. To prove the suitability of the proposed scoring methodology, we have computed a set of common statistics in colour-naming experiments, such as consensus and consistency, on our dataset. The results make us to conclude the coherence of our data with previous experiments and, thus, its usefulness for the fuzzy modelling of colour naming.

Key words: Colour categorization; Basic colour terms; computational models; fuzzy sets

INTRODUCTION

The anthropological study of Berlin and Kay about colour naming, published in 1969, initiated an interesting research about this topic. In their work, Berlin and Kay stated the existence of a universal and limited number of basic colour categories in any language. In the most evolved ones, the set of basic colour terms includes eleven categories (white, black, red, green, yellow, blue, brown, purple, pink, orange, and grey). After their work, some experiments have confirmed their results.[2-6]

The work we present in this paper is not framed in anthropological issues neither in colorimetry goals, but it is framed in the computer vision field. In computer vision, the goal on colour naming research is to find a computational process that, after a learning step, is able to assign names to the colours contained in a digital image. In this field, there have been some attempts to obtain such a computational model for colour naming.[7-10] The computational automation of the colour-naming task has been posed as the problem of finding a statistical model able to assign a membership value to the different colour categories to any visual stimulus. These approaches follow the idea proposed by Kay and McDaniel[11] that considers colour naming as a fuzzy process.

An automatic colour-naming system should assign colour names in the same way as a real human observer. Thus, the success on the automation of the colour-naming task mainly relies on an appropriate learning step. In computer vision, a learning process refers to the procedure of fitting a mathematical model to known data, in our case psycho-physical data, to simulate a specific task. For colour naming, the psycho-physical judgements to be fitted must be obtained from a colour-naming experiment. Hence, considering colour naming as a fuzzy process implies defining a model able to assign to a given colour sample, in our case a RGB triplet from a digital image, a membership value to all colour categories according to the psycho-physical data.

Data from most of the previous colour-naming experiments[3,4,12,13] are not adequate to model the colour-naming space as a fuzzy process. These experiments are usually done by asking subjects to assign a unique colour name to a given stimulus without any other judgement. However, some variations of the monolexemic method have been proposed and some experiments have used ratings to evaluate colour samples[5,14]. Troost and De Weert[15] introduced the idea of describing colour samples with a naming vector where each dimension corresponds to a colour term. This representation was also adopted by Speigle and Brainard[16] by using an 11-dimensional vector where each dimension corresponds to one of the basic colour terms. In that experiment, the ratings ranged from 0 to 9 and were provided by the subjects who where asked to "rate how good an X the stimulus is", where X was each of the eleven basic colour names. Unfortunately, these experiments were not thought to be used for fuzzy colour-naming modelling but were part of colour constancy studies and the corresponding data sets are directed to that purpose.

To provide a sufficient data set of colour judgements adequate for the computational modelling of the colour-naming task, we have performed a colour-naming experiment using a methodology similar to the used by Speigle and Brainard. In our experiment, subjects were asked to assign a membership value to all the colour categories considered. The experiment is restricted to the eleven basic colour terms proposed by Berlin and Kay because previous experiments[3-5] showed that basic names are used more consistently and with more consensus than non-basic names. This fact is essential for computational approaches due to the problems caused by outliers, that is, low consistent and low consensus samples. In addition, non-basic colour terms correspond to small regions on colour space and this is difficult to model from a computer vision point of view.

Therefore, the main goal of this paper is not to replicate any of the previous experiments. Our final goal is twofold. First, to make available to the scientific community a complete dataset derived from a fuzzy colour naming experiment. Such dataset can be the basis of a future mathematical model of the colour naming task by means of a fuzzy model. Second, we have computed common statistics used to evaluate colour naming experiments to demonstrate the validity of these data. We have shown that data obtained from a scoring methodology present similar statistics to data obtained in previous monolexemic experiments. After our analysis we can assure that our data is valid in a perceptual sense, that is, data is not affected by the method used which a priori could introduce more subjectivity to the judgements obtained. Hence, any computational fuzzy modelling of colour naming that uses these data as learning set will have a solid perceptual basis.

The rest of the paper is organized as follows. First, we introduce the fuzzy framework for colour naming in which our work is framed. Second, we explain the methodology followed in our experiment. Then, we present the analysis of the method used in the experiment and, finally, the last section is devoted to the discussion of some conclusions from this experiment.

FUZZY SETS FOR COLOUR NAMING

Previous works[9-10] based on the idea proposed by Kay and McDaniel of considering the colour-naming task as a fuzzy decision have shown the possibilities of this approach to automate colour naming. From this point of view, the best way to mathematically model these decision functions is by considering the basis of the fuzzy set theory.[17]

A fuzzy set is described by its membership function. In colour naming, we can consider that any colour category, C_k, is a fuzzy set with a membership function, $f_{C_k}(\underline{x})$, which assigns, to any colour sample $\underline{x}$, a membership value $f_{C_k}(\underline{x})$ within the [0,1] interval. This value represents the certainty we have about stimulus $\underline{x}$, has to be named with the linguistic term t_k, corresponding to category C_k. From this point of view, the first step of any colour-naming modelling process will be the definition of the membership functions for each colour category. Once these functions are defined, it will be possible to compute a colour descriptor such as:

$$CD(\underline{x}) = (f_{C_1}(\underline{x}),..., f_{C_n}(\underline{x})) = (m_1,...,m_n) \quad \text{where} \quad m_k \in [0,1] \ \ \forall k = 1,...,n \quad \text{and} \quad \sum_k m_k = 1 \qquad (1)$$

$CD(\underline{x})$ describes the membership relation of $\underline{x}$ to each colour category, m_k is the certainty value associated to $\underline{x}$ by f_{C_k} and n is the number of categories considered. In our case n=11 and the categories considered in the model are the corresponding to the eleven basic colour terms proposed by Berlin and Kay, that is $t_k \in$ {*'white'*, *'black'*, *'red'*, *'green'*, *'yellow'*, *'blue'*, *'brown'*, *'purple'*, *'pink'*, *'orange'*, *'grey'*} with k = 1,...,11. Therefore the colour descriptor $CD(\underline{x})$ defined above is a vector of 11 components and the information contained in such descriptor can be used by a decision function, $N(\underline{x})$, which decides the colour name of a given stimulus $\underline{x}$. The most easy decision rule is to choose the maximum from $CD(\underline{x})$ (equation 2).

$$N(\underline{x}) = t_k \quad | \quad m_k = \max_{i=1..11}\{f_{C_i}(\underline{x})\} \qquad (2)$$

Once we have given the essential definition, to automate colour naming we need to model the membership functions, $f_{C_k}(\underline{x})$. Several expressions have been proposed to this end[7,10]. Whatever the selected function is, it will imply to estimate a set of parameters based on psycho-physical data. In the next section we present the experiment that will provide the data set to fit the $f_{C_k}(\underline{x})$ functions.

METHOD

Subjects

The subjects that took part in the experiment were 10 researchers (5 male and 5 female) from our Lab. They were between 24 and 30 years old and all of them were volunteers. All the subjects were tested with the Ishihara and Farnsworth D-15 tests to guarantee they had normal colour vision. All the subjects were bilingual Catalan and Spanish speakers with an advanced level of English since the experiment was developed using the English terms for the eleven basic colour categories.

Stimuli

The stimuli used were a total of 387 samples which included 36 achromatic and 351 chromatic samples. Each sample had a size of 24×16 cm to assure a wide visual angle. All the samples were printed by a HP DesignJet 2500CP Plotter, and their reflectance functions were measured with a PhotoResearch PR-650 spectro-radiometer.

The selection of samples was done according to the requirements of our computer vision application. Because we were using a specific sensor (SONY DXC-930), we were interested in taking a set of reflectances according to three criteria: to cover as much as possible the gamut of our sensor, to avoid overlapping between samples and to have a reasonable number of samples. To this end, first we sampled the Munsell colour space at each hue, at each value unit (from 3 to 9) and at the highest chroma value available. The selected samples were printed with the plotter. Obviously, the printed samples suffered a deviation from the real Munsell due to the use of the plotter. Therefore, the reflectances of the printed samples were measured with the spectro-radiometer and the RGB values of our samples for our sensor were obtained by the following equation:

$$x_i = \int E(\lambda)S(\lambda)R_i(\lambda)d\lambda \tag{3}$$

where x_i represents the *i*th component of vector x=(R,G,B), that is, the RGB response to a sample with $S(\lambda)$ as surface reflectance function, $E(\lambda)$ as the spectral distribution of the illuminant, and $R_i(\lambda)$ as the spectral sensitivity of the *i*th sensor of the device.

The RGB values of all these samples were represented on the rg plane of the sensor. To cover some gaps that appeared in this space we printed some additional surfaces to have the final set of 351 samples that accomplished the three criteria mentioned above.

Since the spectra of the selected samples are known, the coordinates of the samples set at any colour space can be computed. Thus, we provide the CIELab values computed according to standard equations[18] using the CIE Illuminant D65 and the Two Degree Standard Observer. The corresponding Munsell values have been computed with the 'Munsell conversion – Version 4.01' software from Gretagmacbeth (http://www.gretagmacbeth.com). The set of spectra as well as the corresponding CIELab and Munsell values of the samples used in the experiment have been made available online to the scientific community (http://www.cvc.uab.es/color_naming).

Apparatus

The experiment was developed in a dark room. The subjects were sat in an adjustable chair in front of a booth where the samples were presented. They were at a viewing distance of 50 cm. from the sample that was presented on a support painted with a neutral grey corresponding to Munsell N7. Because the final goal of our work is to define a model that assigns colour names in the same way as a human observer, we must consider the effects of the colour constancy mechanisms of the human visual system. Thus, the inside of the booth was white to assure that psycho-physical data would be acquired considering the colour constancy mechanisms of the human subjects doing the experiment. The samples were illuminated from the top of the booth by an illuminant with a correlated colour temperature (CCT) of 5955K and a Luminance of 150 cd/m^2. Figure 1 shows a diagram of the experimental conditions.

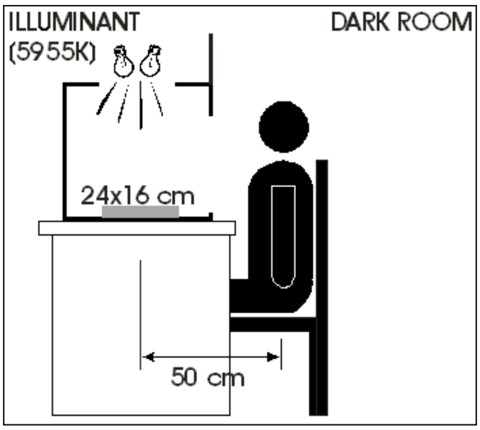

FIG. 1. Scheme of the experiment conditions. The experiment was developed under controlled conditions in a dark room to assure that samples were only illuminated from the top of the booth.

Procedure

The procedure followed was very similar to the one followed by Speigle and Brainard. Subjects were instructed to use only the eleven basic colour terms. These are white, black, red, blue, green, yellow, purple, pink, orange, brown and grey. Any other colour name was not allowed.

For each one of the samples presented, the observer was asked to distribute a total score of 10 points among the 11 possible colour names according to the certainty they had about the sample belonging to the different categories. Thus, if the subject was absolutely sure about the colour name of a sample, then the 10 points had to be assigned to the category corresponding to that name. Otherwise, if there was a doubt between two or more names, the 10 points had to be distributed between the categories corresponding to those names (e.g. blue - 4 and green - 6). Hence, the result of the naming for each sample is a colour descriptor of 11 components, one for each basic colour term. No time limitations were set to give a response.

The 387 samples were presented one at a time, twice each, to the ten subjects. This means a total number of 7740 observations. The samples were first presented in random order, and the reverse was used in the second trial. For each sample, the scores from the 10 subjects were normalized to the [0,1] interval and averaged to obtain the mean colour descriptor of the sample. The set of colour judgements obtained is available at: http://www.cvc.uab.es/color_naming/.

RESULTS

As we have said before, the goal of this work is to present the colour-naming experiment made to obtain a set of fuzzy colour-naming judgements for the eleven basic colour categories. That is, we do not study the existence of the eleven basic colour categories nor their location in the colour space as some previous studies have done. Our goal is to obtain the membership values of a wide set of samples and for each colour category. Moreover, and in order to prove the feasibility of the methodology followed, in this section we show that usual statistics computed on monolexemic colour-naming experiments agree with the ones we can derive from our data. The statistics analysed are consistency, consensus, position of focal colours and centroids, and the confusion matrix in the use of colour terms. The analysis of consistency and consensus was done over the whole set of samples while the analysis of the other three statistics only considered the chromatic categories due to the reduced number of samples that were considered achromatic by the subjects. After this analysis, we will be able to state that the methodology followed in the experiment is a generalization that comprises the results of monolexemic experiments.

Consistency

Consistency is a measure of the degree of coincidence in the two evaluations that each subject makes for each sample. In previous works, consistency has been calculated by counting the number of times that a colour sample has been given the same name by the same subject on the two observations of the sample. Due to the differences between these works and our experiment, some variations of the consistency measure are proposed. In our case, we do not have a colour name for each sample, but a colour descriptor of membership values with 11 components, one for each of the 11 basic colour names. The proposed measures are the following:

Consistency based on identical membership values: Judgements for a sample are consistent when the same subject gives exactly the same values to all the categories in the two observations of the sample.

Consistency based on the highest membership value: Judgements for a sample are consistent when the same subject gives the highest membership value to the same category in the two observations of the sample, no matter how the values are distributed among the 11 categories.

Consistency based on the city-block metric with a threshold (τ): Judgements for a sample are consistent when the city-block distance (equation 4) between the two colour descriptors given by the same subject is lower than the value of the threshold τ.

$$d = \sum_{k=1}^{11} \left| CD_k(\underline{x}_i)^{j,1} - CD_k(\underline{x}_i)^{j,2} \right| \tag{4}$$

where $CD_k(\underline{x}_i)^{j,1}$ is the k-th component of the colour descriptor, that is the score given to the colour category C_k, of the i-th sample for the j-th subject in the first observation, $CD_k(\underline{x}_i)^{j,2}$ is the same component for the second observation.

Global Consistency: The city-block distance between the two colour descriptors given by the same subject is normalized to one and the global consistency is calculated according to equation 5:

$$Consistency = \frac{\displaystyle\sum_{i=1}^{i=ns} \sum_{j=1}^{j=np} 1 - \frac{\sum_{k=1}^{11} \left| CD_k(\underline{x}_i)^{j,1} - CD_k(\underline{x}_i)^{j,2} \right|}{2}}{ns \cdot np} \tag{5}$$

where $CD_k(\underline{x}_i)^{j,1}$ is the k-th component of the colour descriptor of the i-th sample for the j-th subject in the first observation, $CD_k(\underline{x}_i)^{j,2}$ is the same component for the second observation, *ns* is the number of samples, and *np* is the number of subjects in the experiment. Notice that the sum of the differences is divided by the highest possible difference which is 2. This measure has been designed to take into account the nature of the fuzzy methodology of our experiment.

The consistency values were calculated for the four measures proposed. Obviously, the results for the less restrictive measures (global consistency and highest membership value) are the best, but the values obtained by the other measures are also good. According to the measure based on identical membership values, subjects scored the samples consistently in 2501 of the 3870 evaluations. This means that 64.63% of the samples were scored consistently. If we consider a threshold of 0.4 in the calculation of the consistency (i.e. if the city-block distance between the two colour descriptors is lower than 0.4, we consider the naming of the sample is consistent), the number of samples scored consistently increases up to 2762, which means that consistency was reached 71.37% of the time. If we only consider the colour name with the highest membership value to study consistency, the number of samples scored consistently is 3327 (85.97%). Finally, the application of the global consistency measure provides a result of 85.43% of consistent use of the colour names in the experiment. These results are summarized in table I.

TABLE I. Consistency results obtained according to the different proposed measures.

Consistency method	Coincidences	Percentage
Highest membership value	3327	85.97%
Global consistency	3306.2 (*)	85.43%
City-block consistency with τ=0.4	2762	71.37%
Identical membership values	2501	64.63%

(*) Result obtained from equation 5

This is a preprint of an article published in COLOR research and application, 31(1):48-56, 2006.
http://www3.interscience.wiley.com/journal/35037/home

If we assume that in a monolexemic colour-naming experiment the name assigned to any sample would have been the one corresponding to the category with the highest membership value in our experiment, we can compare the consistency results from Boynton & Olson[3] and Sturges & Whitfield[4] experiments with our measure based on the highest membership value. In those works the consistency values for basic colour terms were 75% and 84.4% respectively. As can be seen, our consistency value (85.97%) is very similar to the results obtained in these previous experiments. Moreover, if we consider the consistency measure based on identical membership values, the value obtained (64.63%) is not bad if we take into account that an exact coincidence of the membership values is much harder to obtain than a coincidence of only the highest value.

Consensus

Consensus is a measure of the agreement in the judgements between subjects. In the previous experiments, the consensus in a particular sample was computed by counting the number of times that the most used name for that sample had been given by all subjects in all the trials. Hence, perfect consensus for a sample was achieved when all subjects used the same colour name on both trials of the sample.

In our experiment, we have 20 11-dimensional colour descriptors for each sample, obtained from the two trials of the 10 subjects. From these 20 colour descriptors, we have computed three different consensus measures.

Consensus based on the membership values: Consensus is reached when all subjects have assigned the same values to the eleven categories in all the evaluations of the sample. Hence, in our case, a sample will have consensus if it has 20 identical colour descriptors.

Consensus based on the highest membership value: Consensus is reached when all subjects have assigned the highest membership value to the same category in all the evaluations of the sample.

Consensus based on the highest membership value with a threshold (τ): Consensus is reached when the highest membership value has been assigned to a certain category in at least τ evaluations of the sample. For example, if τ is set to 15, consensus will be reached if the category corresponding to the highest membership value is the same for, at least, 15 of the 20 colour descriptors.

Consensus was computed for the three measures defined above. The first measure gave as a result that consensus was obtained for 43 of the 387 samples; this is, for 11.11% of the samples. As happened with consistency, the second measure can be considered the equivalent to the measure computed in previous experiments. In the present experiment, total consensus based on the highest membership value was obtained for 136 of the 387 samples. This means that 35.14% of the samples were given the highest value to the same category by all the subjects. If we compare these results to the consensus obtained in previous works, we can see that perfect consensus in the experiments of Boynton & Olson and Sturges & Whitfield was obtained on 30% and 23% of the samples respectively. In our case, we obtain a slighter higher consensus probably due to the fact that we do not allow non-basic colour names which are used with lower consensus than basic colours.

Using the third statistic, the percentage of consensus increases gradually as the threshold is reduced. Thus, when only 15 coincidences on the category which has been assigned the highest membership value are required, consensus is of 75.97%, and this value increases up to 98.19% when the threshold is set to 11 coincidences, which is the minimum possible majority. In figure 2, the evolution of consensus value in terms of the threshold (number of coincidences required) is presented.

This is a preprint of an article published in COLOR research and application, 31(1):48-56, 2006.
http://www3.interscience.wiley.com/journal/35037/home

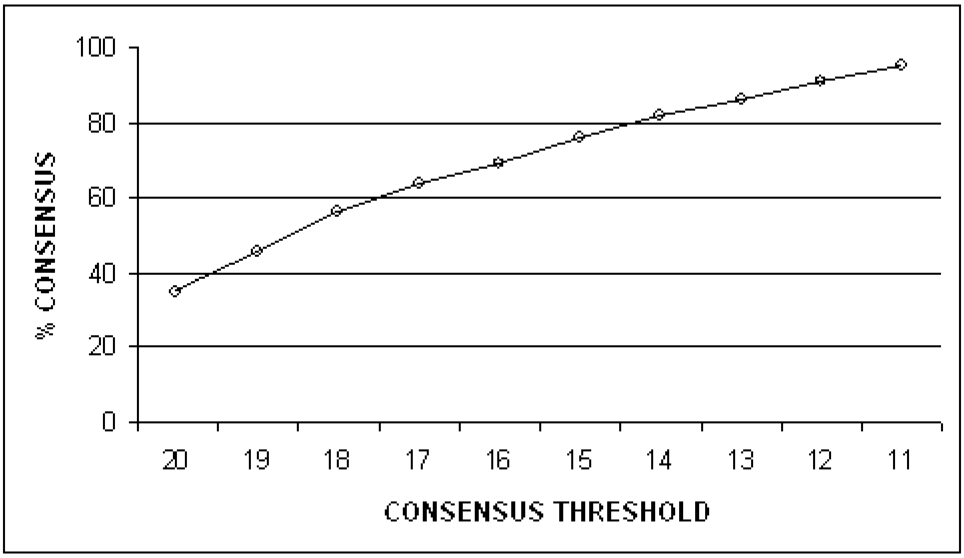

FIG. 2. Percentage of consensus when the criterion for defining consensus is relaxed from 20 coincident responses (of 20 possible) to the smallest possible majority of 11 coincidences.

Focal colours

Focal colours are defined as the best examples for each colour category. In previous experiments, focal colours have also been defined as the fasted named samples with consensus for each colour category. In our experiment, response time was not considered. Hence, we have selected as focal candidates all the samples which have been assigned 10 points to the same colour category by all the subjects in the two evaluations of the sample. For two categories, orange and yellow, there is no sample with total consensus and the samples with the highest mean membership value in those categories have been selected as focals. In figure 3, our focal candidates and the focals from Berlin & Kay and Sturges & Whitfield experiments are shown in terms of hue and value on Munsell colour space.

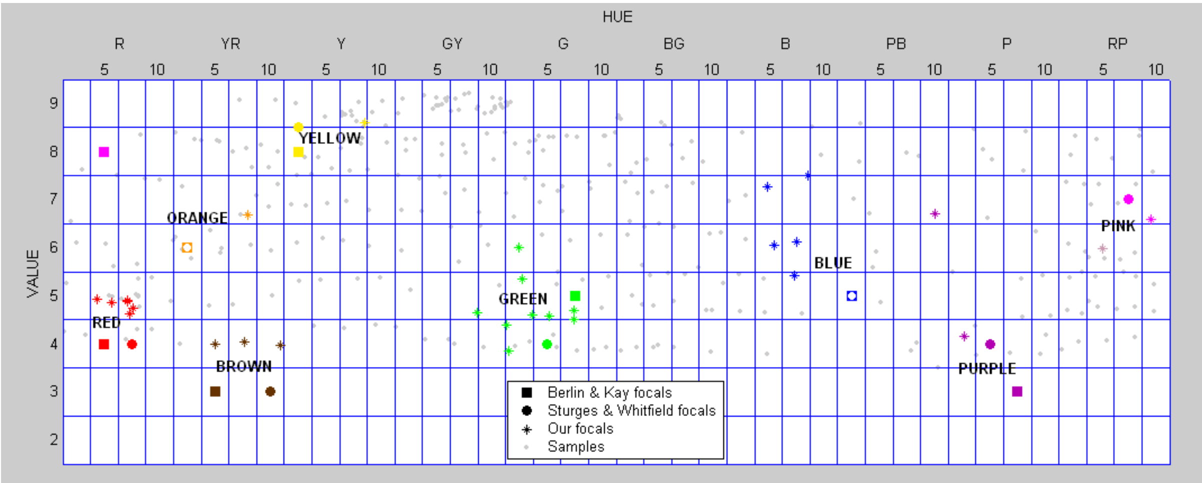

FIG. 3. Location of candidate focals obtained in our experiment and focals from two previous works (Berlin & Kay and Sturges & Whitfield). Samples used in our experiment are shown as grey dots.

As can be seen in the figure, the consensus samples from our experiment lie, in general, near the focal colours found by Berlin & Kay and Sturges & Whitfield. The focals from Sturges and Whitfield have, in most of the cases, a candidate focal of our experiment at a distance lower than one Munsell chip

(2.5 hue units or 1 value unit). This distance is higher for three categories: orange, yellow and blue. In the case of orange, the shift of the focal could be due to cultural reasons, because it had been previously detected in a preliminary experiment[10] that Catalan and Spanish speakers located the orange region at a different area from English. In the case of blue, the problem could be due to the different criteria of selection of samples in the experiments, since the focal locations are highly dependent on the stimulus set.

Centroids

Centroids are a measure of the central tendency of the location of the categories in the colour space. In previous experiments the centroids of each category have been computed by averaging the values of all the samples named with the corresponding term, and weighted according to whether the term was used once or twice. In our case, we have computed the centroids as the average of the values of the samples, but weighted according to the membership values provided by the subjects for each colour category (equation 6).

$$Centroid_i = \frac{\sum_{j=1}^{n} m_i^j \underline{x}_j}{\sum_{j=1}^{n} m_i^j} \qquad (6)$$

where $\underline{x}_j$ is the j-th sample and m_i^j is the membership value of the j-th sample for category C_i.

Table II shows the centroids obtained in our experiment and in two previous monolexemic works.[12,13] Table III allows a comparison to the previous results by giving the CIELab differences between experiments. In figure 4, the centroids for the chromatic categories of the three experiments are presented in terms of Munsell Hue and Value.

TABLE II. Centroids obtained in our experiment and in two previous studies.

	Boynton & Olson			Sturges & Whitfield			Our experiment		
	Hue	Value	Chroma	Hue	Value	Chroma	Hue	Value	Chroma
Red	3.5R	3.97	9.19	6.01R	3.91	11.72	6.55R	5.02	10.82
Orange	1.73YR	5.91	10.1	3.88YR	6.21	11.45	7.42YR	6.77	7.9
Brown	3.85YR	4.41	4.52	7.69YR	4.32	5.78	6.88YR	6.01	4.72
Yellow	2.51Y	7.56	8.28	3.35Y	7.86	10.44	7.89Y	7.96	6.82
Green	8.21GY	6.07	5.07	1.73G	5.35	7.52	1.21G	6.38	4.16
Blue	8.44B	5.42	5.12	8.23B	4.87	7.33	2.67B	6.46	4.45
Purple	6.37P	4.53	5.22	4.75P	4.18	7.86	8.69P	5.65	5.82
Pink	2.04R	6.04	5.68	2.15R	6.14	8.59	8.38RP	6.72	6.73

TABLE III. Comparison of centroids obtained in our experiment (BVB) and in previous studies (Boynton & Olson (BO) and Sturges and Whitfield (SW)), in terms of the CIELab differences.

	BO vs. SW	BO vs. BVB	SW vs. BVB
Red	15.75	12.56	17.69
Orange	14.05	20.83	19.81
Brown	12.17	18.81	19.02
Yellow	16.03	24.55	14.61
Green	17.19	21.27	10.47

This is a preprint of an article published in COLOR research and application, 31(1):48-56, 2006.
http://www3.interscience.wiley.com/journal/35037/home

Blue	10.93	22.89	14.38
Purple	12.78	20.03	12.08
Pink	12.45	12.86	8.99
Mean diff.	13.92	19.23	14.63

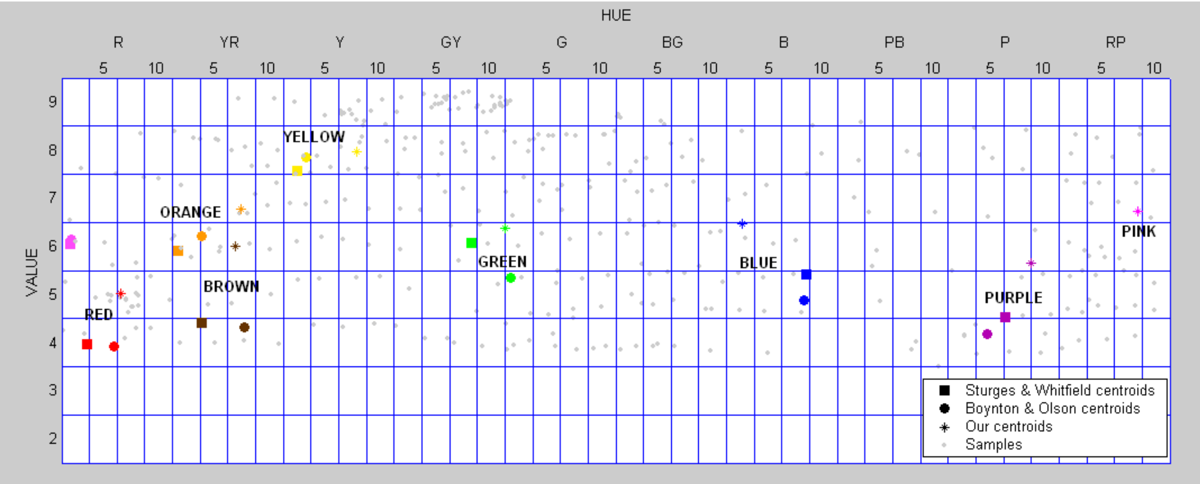

FIG. 4. Location of centroids obtained in our experiment and in two previous works (Boynton & Olson and Sturges & Whitfield). Samples used in our experiment are shown as grey dots.

As can be seen in table II and in figure 4, the centroids found in our work have, in general, a difference of one value unit to the centroids of previous works. Again, this is due to the different criteria in the samples selection which has been previously explained. The fact that the samples were selected in order to cover the sensor space has brought a lack of samples in the low value areas of the Munsell space. The value difference is considerably smaller for those categories that are located in the areas of high value in the Munsell system, such as pink and yellow. The bigger differences are found for the categories located in the low value areas (brown, purple and red) because they are more affected by the lack of low value samples in the experiment. The mean difference to Boynton and Olson study is 19.23 CIELab units and the difference to Sturges and Whitfield is 14.63, while the mean difference between the centroids found by Boynton and Sturges was of 13.92. Another possible cause of these differences in the results is the different illumination conditions in the experiments. Boynton and Olson used an illuminant with a temperature of 3200K and Sturges and Whitfield used an illuminant of 6500K while our experiment was done under a 5955K illuminant. The mean difference between centroids is smaller compared to Sturges and Whitfield who used an illuminant similar to ours. However, the difference between Sturges and Boynton is also small, but the illuminants are quite different. Hence, the influence of illumination conditions in the results is not clear and needs further study. Anyway, we can see that the differences between our centroids and the ones from previous studies are reasonable. Moreover, the higher difference found in our case seems to be due to the high dependence of the centroids locations on the set of stimuli selected as was concluded by Speigle and Brainard.

Linked Colours

In previous studies the confusion matrix in the use of basic colour terms was analysed to study the way in which basic colour terms were related. Boynton and Olson defined two colours as linked when a majority of subjects applied the same different basic colour terms to describe a particular colour sample on the two observations of the sample. According to Sturges and Whitfield this inconsistency indicated that the sample was perceived as a mixture of colours that contained elements of the two

This is a preprint of an article published in COLOR research and application, 31(1):48-56, 2006.
http://www3.interscience.wiley.com/journal/35037/home

terms used. With the information from the confusion matrix, it is possible to build a three-dimensional model of the relationships between the basic colours.

In our case, the experiment has provided more information about these relationships than previous experiments and we can use the whole set of membership values to build the confusion matrix. Hence, our confusion matrix has been built by counting the number of times that any subject has scored two particular colour names for the same sample. This definition of the confusion matrix agrees with the idea of Sturges and Whitfield that when a sample is named with two different basic names on the two observations of the sample it is perceived as a mixture of the two colours. In our experiment, we assume that if subjects give values to two or more categories for the same sample, it is because none of the basic colour terms describes exactly the sample and they doubt about which name must be given to the sample. Table IV shows the confusion matrix for our experiment (white and black are omitted due to lack of sufficient data). To avoid some outliers detected on data, values less than 2 points have not been considered.

TABLE IV. Confusion matrix computed from the results of our experiment. Each cell of the matrix shows the number of times that a sample has been given values to the categories intersecting on the cell. The number of subjects who scored the two categories for at least one sample is given in parentheses.

	Red	Green	Yellow	Blue	Orange	Purple	Grey	Pink	Brown
Red	----				38 (10)	14 (6)		39 (9)	38 (9)
Green		----	90 (9)	274 (10)	1 (1)		55 (9)		44 (9)
Yellow		90 (9)	----		32 (8)		5 (4)	1 (1)	59 (10)
Blue		274 (10)		----		29 (7)	24 (7)		
Orange	38 (10)	1 (1)	32 (8)		----			25 (6)	70 (10)
Purple	14 (6)			29 (7)		----	13 (5)	57 (9)	5 (3)
Grey		55 (9)	5 (4)	24 (7)		13 (5)	----	6 (4)	30 (5)
Pink	39 (9)		1 (1)		25 (6)	57 (9)	6 (4)	----	19 (3)
Brown	38 (9)	44 (9)	59 (10)		70 (10)	5 (3)	30 (5)	19 (3)	----

From the confusion matrix a three-dimensional model of the relationships between the basic colour categories has been built. Figure 5 shows a schematic representation of this model. The criteria to decide whether two colours are linked are similar to the used by Sturges and Whitfield. As in their work, we represent all the relationships of the confusion matrix (excluding confusions from a unique subject) and the width of the line linking two colours indicates the confusion between the two colours.

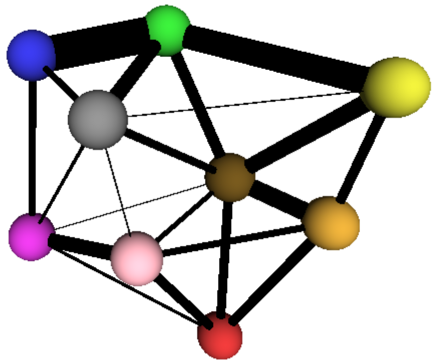

FIG. 5. Schematic representation of the linkages between basic colour categories (excluding black and white). The spheres represent the position of colours on the chromatic plane of the CIELab space. The model is viewed in perspective from L=160.

The analysis of the linkages in Figure 5 shows that the colours that are linked and unlinked are the same as in previous works, except for yellow-grey and pink-grey which were unlinked in both previous studies. However, the linkages observed in these two cases are very weak and confusions between those colours also appear in the confusion matrices of the previous studies.

Hence, these results mean that although the method employed seems to include more subjectivity, the rating of samples is done in a perceptual sense. Thus, for example, nobody has given values to red and green or blue and yellow at the same time, which would be contradictory to the opponent theory of colour perception. Hence, the comparison of the results in terms of linked colours also supports the validity of the method and the equivalence to previous studies about colour naming.

CONCLUSIONS

In this paper, we have presented the colour-naming experiment developed to obtain a set of colour judgements useful to be used as the basis for the fuzzy modelling of the colour-naming task. The set of fuzzy judgements, the reflectances and, the CIELAB and Munsell values of the samples used in the experiment are available online (http://www.cvc.uab.es/color_naming). Making psycho-physical data available to the research community is an important step, at least in the computer vision field, because it can contribute to achieve faster progress in this field. Moreover, it will allow testing and comparing different models over the same data sets. This fact is also very important in our field due to the complexity of obtaining that kind of data sets for testing.

The methodology used in this experiment was very similar to the one used by Speigle and Brainard. The results of our experiment have been compared to the ones from previous monolexemic experiments to show that although the experimental methods are considerably different, the results in terms of some usual measures in colour naming are similar. This fact allows converting the results of the present experiment to a monolexemic experiment by just assigning to each sample the colour name with the highest membership value in the colour descriptor provided by the subjects from the experiment.

The analysis of the results also supports the statement of Speigle and Brainard about the heavy influence of the stimuli gamut in the location of focals and centroids. Hence, some deficiencies in the selection of the samples used in the experiment have caused that our focals and centroids are not exactly in the same locations as in some previous works. Although these differences are not important for our purpose, it would be interesting to find a way to analyse colour-naming results that is not so dependent on the set of stimuli used as the computation of focal colours and centroids.

Once we have the set of fuzzy colour-naming judgements the next step is to find a mathematical model that fits the data from our experiment. Using such data as learning set will allow us to obtain a model of colour naming that provides the same colour-naming classifications as a real human observer. The results obtained in previous attempts[9,10] to automate the colour-naming task are encouraging. The present data set will help to improve the results obtained in these previous works.

Another important topic in the automation of colour-naming task is to take into account the colour constancy problem.[19] The present experiment was done under a near daylight illuminant. However, the variation of the illumination conditions could make the colour-naming judgements to change. If it was the case, it would explain some of the differences found in the results of different colour naming studies that did not use the same illuminants. Hence it would be interesting to replicate the present experiment under different illuminants in order to obtain a set of judgement which allows studying and modelling the effects of the illumination changes over the human colour naming.

ACKNOWLEDGEMENTS

This work has been partially funded by projects TIC2000-0382 of Spanish Ministerio de Ciencia y Tecnología and TIN2004-02970 of Spanish Ministerio de Educación y Ciencia.

1. Berlin B, Kay P. Basic Color Terms: Their Universality and Evolution. Berkeley: University of California; 1969.
2. Uchikawa K, Boynton RM. Categorical color perception of Japanese observers: Comparison with that of Americans. Vision Research 1987;27:1825-1833.
3. Boynton RM, Olson CX. Salience of chromatic basic color terms confirmed by three measures. Vision Research 1990;30:1311-1317.
4. Sturges J, Whitfield TWA. Salient features of Munsell color space as a function of monolexemic naming and response latencies. Vision Research 1997;37:307-313.
5. Guest S, Van Laar D. The structure of colour naming space. Vision Research 2000;40:723-734.
6. Lin H, Luo MR, MacDonald LW, Tarrant AWS. A cross-cultural colour-naming study. Part I: Using an unconstrained method. Color Research and Application 2001;26:40-60.
7. Lammens JM. A Computational Model of Color Perception and Color Naming [PhD thesis]. Buffalo [USA]; University of New York; 1994.
8. Yendrikhovskij SN. A Computational Model of Colour Categorization. Color Research and Application 2001;26(S1):S235-S238.
9. Benavente R, Tous F, Baldrich R, Vanrell M. Statistical Modelling of a Colour Naming Space. Proceedings of the 1st European Conference on Color in Graphics, Imaging and Vision. Poitiers (France): 2002. p 406-411.
10. Benavente R, Vanrell M, Baldrich R. Estimation of Fuzzy Sets for Computational Colour Categorization. Color Research and Application 2004; 29:342-353.
11. Kay P, McDaniel CK. The linguistic significance of the meaning of basic color terms. Language 1978;54: 610-646.
12. Boynton RM, Olson CX. Locating basic colors in the OSA space. Color Research and Application 1987;12:94-105.
13. Sturges J, Whitfield TWA. Locating basic colors in the Munsell space. Color Research and Application 1995;20:364-376.
14. Boynton RM, Fargo L, Collins BL. Categorical color rendering of four common light sources. Color Research and Application 1990;15:222-230.
15. Troost JM and De Weert CM. Naming versus matching in color constancy. Perception & Psychophysics 1991;50:591-602.
16. Speigle JM, Brainard DH. Is Color Constancy Task Independent? Recent Progress in Color Science. Springfield: The Society for Imaging Science and Technology, Karen Braun and Reiner Eschbach Eds; 1997.
17. Klir GJ, Yuan B. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Pearson Education POD. Upper Saddle River: Prentice Hall, 1995.
18. Wyszecki G, Stiles WS. Color Science: Concepts and Methods, Quantitative Data and Formulae. New York: John Wiley & Sons, 1982.
19. Finlayson GD, Hordley SD, Hubel PM. Color by Correlation: A Simple, Unifying Framework for Color Constancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001;23:1209-1221.