Beige, Aqua, Fuchsia, etc.: Definitions for Some Non-Basic Surface Colour Names

Paul Centore

Abstract

Many common surface colour names such as "beige" lack technical definitions. This paper uses a data set of about 16,000 reflectance spectra of named physical samples, collected by the Color Association of the United States (CAUS), to infer definitions for 20 common non-basic surface colour names (aqua, beige, coral, fuchsia, gold, lavender, lilac, magenta, mauve, navy, peach, rose, rust, sand, tan, taupe, teal, turquoise, violet, wine), as well as Berlin and Kay's basic names (excepting black). A convex polyhedron in the Munsell tree is constructed for each name. Any colour inside the polyhedron, and only such a colour, is described by that name, with the implicit understanding that the polyhedron's boundaries are inherently fuzzy; the polyhedron's centroid can be taken as the name's focal colour. Accompanying files make the polyhedra publicly available. The current analysis is unique in using a three-dimensional definition, in its large number of surface colours, and in its inclusion of non-basic surface colour names. After presenting the analysis results, comparisons are made to other investigations of naming.

1 Introduction

While colour perception is non-verbal, humans typically assign names to colours. Though commonly understood, colour names lack the objective, technical definitions that are required by many practical applications, especially automated applications. This paper infers such definitions from a data set, collected over several decades by the Color Association of the United States (CAUS), of about 16,000 named and measured physical samples.

A colour name is formulated as a subset of the commonly used Munsell colour system, which classifies object colours by three basic perceptual attributes: hue, value, and chroma. The Munsell system can be drawn as an irregular cylinder, called the *Munsell tree*. The subset defining each name is a convex polyhedron in the Munsell tree. Any colour inside a name's polyhedron is consistent with that name, and the polyhedron's centroid is a single colour that best typifies that name.

In 1969, Berlin and Kay[1] identified 11 *basic* colour names: white, black, red, green, yellow, blue, brown, purple, pink, orange, and gray. With a few exceptions, every language, even the most isolated, develops names for these colours, and the development occurs, with a few variations, in the order listed. Berlin and Kay found the extent of the basic terms experimentally, presenting subjects with a chart, indexed by hue and value, of highly saturated

Munsell chips, and asking them to indicate which colours they considered, for instance, blue. In addition, subjects selected a *focal colour* which they considered the truest exemplar of blue.

Apart from the 11 basic colour names, the English language identifies many non-basic colours, such as *aqua* or *peach*. This paper infers definitions for 20 non-basic names (aqua, beige, coral, fuchsia, gold, lavender, lilac, magenta, mauve, navy, peach, rose, rust, sand, tan, taupe, teal, turquoise, violet, wine) that occur sufficiently often in the CAUS data set. Unlike Berlin and Kay, the entire three-dimensional Munsell system is used rather than just the highest-chroma colours.

For each sample, the CAUS data set provides a name and a reflectance spectrum, from which Munsell coordinates were calculated.[2] To define *beige*, for example, all the samples named *beige* or some variant thereof were plotted in the Munsell tree as a three-dimensional cloud. The cloud, like the other names' clouds, appeared convex, so the convex hull of the beige cloud was found. Since colour names are inevitably fuzzy (e.g., one cannot say exactly where red ends and orange begins), and since a fair number of outliers, which seemed unduly far from the bulk of the cloud, were noticed, the cloud was reduced by discarding all the hull's vertices, and the convex hull was recalculated on the reduced set, producing a polyhedron. This polyhedron defines beige: any Munsell colour inside the polyhedron would be called beige, and no colour outside would. Of course, the polyhedron's boundary is implicitly approximate, because there are no hard dividing lines between beige and non-beige. The centroid of the polyhedron gives a focal colour for beige—a single colour that most typically represents beige.

This paper's main result is such polyhedra for the 20 non-basic colour names, as well as for 10 of Berlin and Kay's 11 basic colour names (*black* was excluded because of insufficient data). Accompanying computer files give the polyhedra's vertex and face structure, and list the names and Munsell coordinates of the samples used in constructing those polyhedra. The paper also presents the polyhedra visually, as projections onto two-dimensional sections of Munsell space. An inclusion test based on a polyhedron can determine whether or not a particular colour would be characterized as, for instance, beige. A table of centroids is also provided.

The paper is organized as follows. First, the Munsell system, which classifies surface colours both rigorously and intuitively, is introduced. Second, the CAUS data set is described, along with some data cleaning. Third, the analysis method is presented, with a step-by-step illustrative calculation of the polyhedron and centroid for beige. Fourth, the polyhedra themselves are presented and discussed, and some applications are suggested. Fifth, the current work is compared to other work on colour naming, and some important issues are highlighted. Finally, a summary is given.

2 The Munsell System

2.1 Perceptual Description

At the start of the 20th century, the American painter and teacher Albert Munsell developed the Munsell colour system, as an educational tool. It has since become common in many visual fields, such as graphics and fashion. It applies only to the colours of physical objects, rather than to coloured light sources. The Munsell system is helpful because it classifies surface colours by three natural perceptual attributes that are basic to art and design: hue, value and chroma.

Hue is universally understood. It says whether a colour is red, yellow, purple, etc. Munsell designates 10 basic hues: R (red), YR (yellow-red, or orange), Y (yellow), GY (green-yellow), G (green), BG (blue-green), B (blue), PB (purple-blue), P (purple), and RP (red-purple). Each basic hue is further subdivided into 4 steps, denoted with a numerical prefix. For example, the four greens are denoted 2.5G, 5G, 7.5G, and 10G. 2.5G is a yellower green, that is closer to GY than it is to BG. 10G is a bluer green, that is closer to BG than it is to GY. A prefix of 10 is sometimes replaced with a prefix of 0 and the next hue. For example, 10G is equivalently written 0BG. In all, then, the Munsell system specifies 40 hues (4 steps for each of the 10 basic hues). These 40 hues are equally spaced perceptually. For example, the hue difference between 2.5G and 5G is the same size as the hue difference between 5G and 7.5G. The 40 hues are discrete stopping points on a continuous hue circle. One could interpolate any desired amount between two adjacent hues. For example, the hue 6GY is a yellowish green that is between 5GY and 7.5GY, but closer to 5GY. White, black, and grays are not considered hues in the Munsell system. An N, for "neutral," is used to designate them.

Munsell value designates how light or dark a colour is. The theoretically darkest black has a value of 0, and is denoted N0. The theoretically lightest white has a value of 10, and is denoted N10. N0 and N10 are theoretical ideals, that actual materials approach, but have so far not reached. Most blacks are about N1, rather than N0. Similarly, common whites are just below N10. Between N0 and N10 are 9 progressively lighter grays, denoted N1, N2, and so on up to N9. The spacing between the grays is perceptually equal. All colours have a Munsell value, not just the neutrals. For example, there are light blues and dark blues. A blue with value 8.5 has the same lightness as N8.5.

Munsell chroma refers to how intense, or saturated, a colour is. For instance, a lemon is an intense yellow, while masking tape is a dull yellow. A dull colour is closer to a neutral gray than an intense colour. The Munsell system denotes chroma numerically. Grays have chroma 0. A colour with a chroma of 10 or greater is generally perceived as saturated, and it is rare to encounter chromas greater than about 16. Colours of low chroma, say 4 or less, are perceived as subdued, with a high gray content. It is often difficult to distinguish the hue of low-chroma colours. For example, one cannot say readily whether masking tape is more yellow or more orange. The hue of high-chroma colours, by contrast, can easily be identified.

Many different colours can have the same hue. Figure 1, for example, shows the "hue leaf" for 10R, a set of colours all of which have hue 10R. The different colours within a hue leaf are specified further by value and chroma. Any colour with a particular hue and value has a maximum attainable chroma, which occurs at the rightmost end of the horizontal bar of colours of that value, in that hue leaf. Figure 1 shows, for instance, that the maximum chroma for hue 10R at value 5 is about 18. The limits in the figure are theoretical limits that practical exemplifications, made with actual pigments or inks, likely cannot reach. The empty boxes indicate colours that are in the Munsell system, but that are beyond the gamut of the process used to produce the figure. Each hue leaf increases in chroma as one moves to the right, and decreases as one moves to the left. The left edge of each hue leaf, in fact,

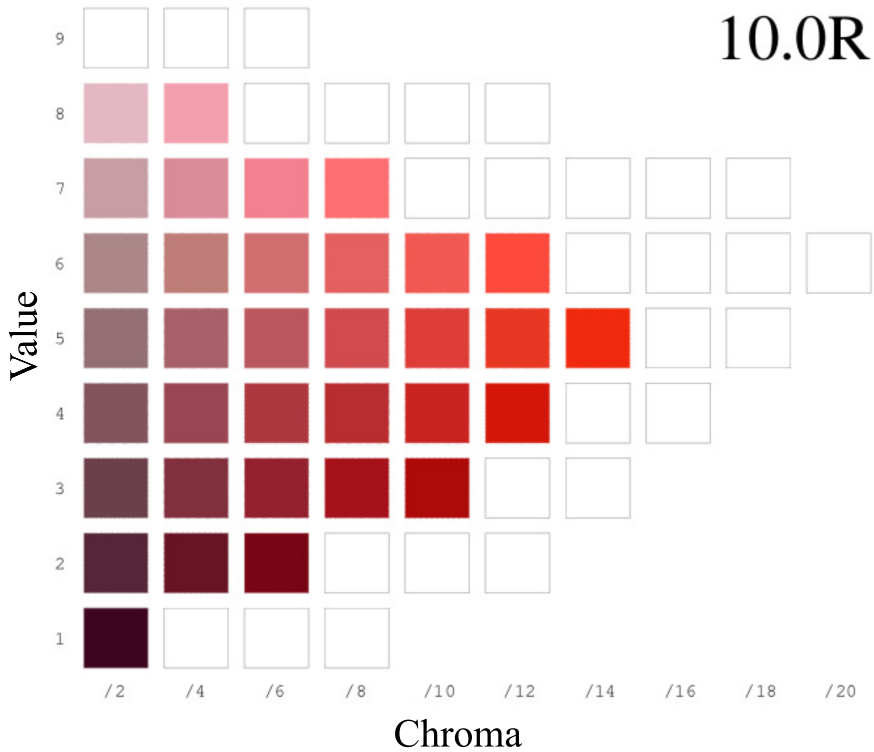

Figure 1: The Hue Leaf for 10R in the Munsell System

transitions smoothly into a vertical neutral axis, consisting solely of grays, whose chroma is 0.

The Munsell notation for a colour takes the form H V/C, where H stands for hue, V stands for value, and C stands for chroma. For instance, the colour 10R 7/6 would be a fairly light (V is 7), moderately intense (C is 6), orangish red (H is 10R). A colour with chroma 0 is a neutral gray, which is denoted NV, where V stands for value. For example, N5 is a gray that is midway between white and black.

A three-dimensional *Munsell tree*, shown in Figure 2, can be constructed by placing the far left, or neutral, edges of all the hue leaves along a common vertical axis. Since the hue leaves have different shapes, the Munsell tree as a whole is irregular, extending out to different distances at different heights, depending on the leaf. The leaves are placed sequentially by hue (red, yellow-red, yellow, green-yellow, etc.), forming a smooth circle. The colours of maximum chroma, for a given hue and value, occur on the tree's outer surface, which consists of the rightmost edges of the hue leaves. Berlin and Kay used these maximum-chroma Munsell chips in their experiments.

Early versions of the Munsell system were collections of hand-painted swatches, which were used as physical standards for judging other colours. A major advance was the 1943 Munsell renotation,[3] which superseded previous versions and is the standard today. The renotation used thousands of visual assessments of paint samples, by 41 human observers, to provide a firm empirical and quantitative basis for the system. Furthermore, the renotation can be inverted[2] to calculate a colour's Munsell coordinates from its reflectance spectrum. A reflectance spectrum is found by measuring a colour with a spectrophotometer; the spectrum

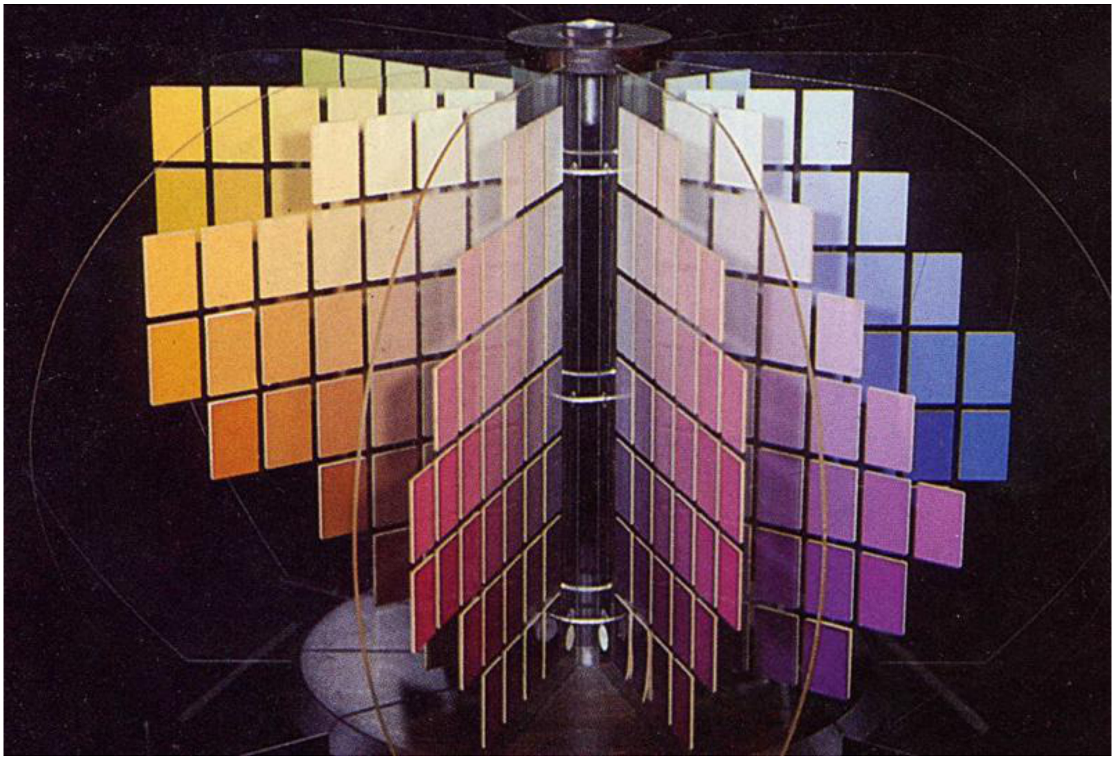

Figure 2: The MunsellTree

gives the percentage of light that the colour reflects at each wavelength.

It should be noted that the Munsell system applies only to surface colours, and not to coloured lights, such as the RGB signals produced by an electronic display. This distinction is important because some recent colour naming studies have used RGB signals as samples, and it is not clear how to convert between RGBs and Munsell coordinates. As a consequence, their data cannot be merged with the CAUS data without some additional analysis.

2.2 Coordinates for the Munsell Tree

The Munsell tree can be conveniently coordinatized with cylindrical coordinates. Let the base of the cylinder be a plane, shown in Figure 3, coordinatized by a hue angle θ and a radial distance r, where r is Munsell chroma. The hue angle starts at 0 on the right-hand horizontal axis. which corresponds to the hue 0R, and increases counterclockwise to 360°, advancing evenly through the hues YR, Y, etc. The vertical axis, denoted z, corresponds to Munsell value. Another convenient coordinatization for the Munsell tree is Cartesian coordinates, in which the vertical coordinate z stays the same, but the horizontal coordinates use the standard polar transformation:

$$x = r\cos\theta \tag{1}$$

$$y = r\sin\theta. \tag{2}$$

The conversions between Munsell, cylindrical, and Cartesian coordinates are straightforward, and this paper will switch between them as needed.

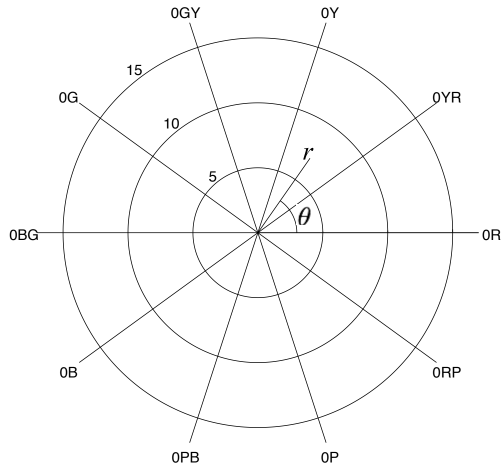

Figure 3: The Hue-Chroma Base Plot for the Munsell Tree

A set of all Munsell specifications that correspond to a particular colour name, such as beige, can be plotted in the Munsell tree. These sets will be found to be convex: if two colours are in the set, then any colour on the line segment that connects them is also in the set. In colour terms, any colour that is between two versions of beige is itself a version of beige. Technically, our set of colours is actually a finite set of discrete points. To make it into a convex set Γ, we take its convex hull, which is the smallest convex set that contains all the colours. The result is a convex polytope,[4] which in the three-dimensional Munsell tree is a polyhedron.

(The concept of a "straight" line segment, which is needed for discussing convexity, actually depends implicitly on the colour space coordinates. If the Munsell tree were spread out so that the hues lay on a straight line rather than on a circle, as was done by Berlin and Kay and many after them, then the straight line between two colours would go through a different sequence of intermediate colours. To avoid a long discussion of mathematical subtleties, however, we will assume that straight lines and convexity are defined in terms of Cartesian coordinates.)

The centroid $C = (C_x, C_y, C_z)$ of a convex set or polyhedron Γ is, roughly speaking, the middle of Γ. Different centroid calculations can be used, depending on the interpretation of the data. One interpretation treats the cloud as a set of N equally weighted point masses at (x_i, y_i, z_i) :

$$C_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{3}$$

$$C_y = \frac{1}{N}\sum_{i=1}^{N} y_i, \tag{4}$$

$$C_z = \frac{1}{N}\sum_{i=1}^{N} z_i. \tag{5}$$

A second interpretation treats Γ as a filled solid, and integrates over it, giving

$$C_x = \frac{\int_{\mathcal{H}} x dx dy dz}{\int_{\mathcal{H}} dx dy dz}, \tag{6}$$

$$C_y = \frac{\int_{\mathcal{H}} y dx dy dz}{\int_{\mathcal{H}} dx dy dz}, \tag{7}$$

$$C_z = \frac{\int_{\mathcal{H}} z dx dy dz}{\int_{\mathcal{H}} dx dy dz}. \tag{8}$$

The two-dimensional example in Figure 4 explains the difference between these two centroid calculations. Suppose we have five data points, four on the corners of a square, and one on the middle of the bottom edge. Their convex hull Γ is the filled unit square. The centroid of the convex hull is just the center of the square, which is marked. The centroid of the data points as point masses, however, is below the center of the square, because Equation (4), which just averages the points' y-coordinates, evaluates to 0.4 rather than 0.5. Adding any number of data points within the square will not change the filled-solid centroid, but can change the point-mass centroid, as we have just seen. In practice, fortunately, the difference between the two centroids is usually minimal, assuming a fair number of data points, distributed without any bias. The point-mass centroid of Equations (3) through (5) is easier to calculate than the filled-solid centroid, and more commonly used (see, e.g. Ref. 8), but the current paper will use the filled-solid form, so that adding interior points will not affect the centroid.

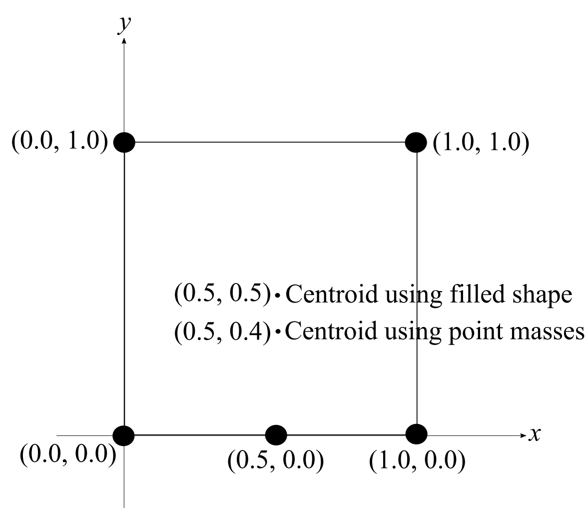

Figure 4: Different Centroids from Point Masses vs. a Filled Solid

3 The CAUS Data Set

The CAUS possesses a collection of fabric samples, taken mainly from the fashion industry. A spreadsheet stores 18,706 lines, one for each sample. Each line gives

1. A year,
2. A season (e.g. fall, winter, etc.),
3. A target market (e.g. men, women, youth, etc.),
4. A name (e.g. Amazon green, Astro orange, etc.), and
5. A reflectance spectrum, presented as a sequence of reflectance fractions, one for every wavelength from 380 to 730 nm in increments of 10 nm.

The data set was augmented by calculating[2] the Munsell specification from the reflectance spectrum for each sample.

Some adjustments were made to the data. A preliminary examination showed that the reflectance percentage, at one wavelength or more, for 2245 samples, exceeded 100%. These samples are believed to be fluorescent. Since the Munsell renotation only applies to non-fluorescent surface colours, these 2245 colours were excluded. In the course of analysis, about 60 highly implausible colour names were found. For instance, a "Futurist blue" sample had Munsell coordinates 4.75R 4.57/15.46, indicating a very saturated red. For lack of an alternative explanation, these samples were dismissed as clerical errors. This decision was sometimes delicate, and ran the risk of introducing unintentional bias. When in doubt, a sample was retained, on the rationale that boundaries are fuzzy, and the polyhedron construction will discard outliers anyway. All the samples that contributed to a polyhedron were checked for a questionable name.

These adjustments left a data set of about 16,400 non-fluorescent samples, with plausible names, and this data set was used to infer definitions for 30 different colour names.

4 Analysis Method

4.1 The Form of a Colour Name Definition

Like most words, colour names arise informally, relying on examples rather than scientific formulations. As a result, a name's definition must be inferred from the way speakers use that word. This task is fairly straightforward for colours, because people can indicate physical samples that exemplify a colour name. In fact, colour-naming experiments typically present a variety of Munsell chips or other samples to a subject, who identifies the ones he would call, for instance, *blue*. Alternatively, a subject might be invited to assign names to samples. This paper infers definitions from the CAUS data set, in which colour names have already been associated with samples, though not as part of a formal experiment.

A colour name, for an object or surface colour, is here defined as the set of all the colours in Munsell space to which that name applies. Geometrically, this set will be constructed as a convex polyhedron in the Munsell tree. Convexity is perceptually reasonable, because one would expect that if two colours share a common name, then any colour directly between them would also share that name. Furthermore, all such sets constructed from the data appear convex, with a fair amount of central clustering, indicating that a convex model is plausible.

The boundaries of colour names seem inherently fuzzy. The same sample, for instance, might be called blue by some observers, green by others, and bluish green by yet others, implying that the dividing line between blue and not-blue is not always clear. The existence of modifiers like *bluish* also imply that observers consciously recognize that a colour is close to blue, but not quite there. To reflect this ambiguity, the polyhedron's boundary, as described in detail in the next section, will exclude some colours near the boundary, even though the samples for those colours were sometimes identified as blue in the CAUS data. Another motivation for this exclusion is the regular appearance of outliers, samples whose Munsell coordinates seem unduly far removed from the bulk of other samples with the same name. A name will apply to a colour if and only if that colour, plotted in the Munsell tree, is within the polyhedron for that name. This definition acknowledges implicitly, however, that different observers might judge that a colour near the polyhedron's boundary, either inside or outside, is, or is not quite, described by that name.

Some previous work on basic colour names also identified a *focal colour*, which is the truest or most typical version of that name. The focal colour was generally near the middle of the set of all colours of that name, which, however, were restricted to the two-dimensional surface of the Munsell tree. The concept transfers easily to the three-dimensional tree, where the centroid of a name's polyhedron is taken as the best exemplar of that name.

4.2 An Example of Constructing a Polyhedron

This section illustrates the calculation of the polyhedron and centroid for the name *beige*. The other names follow this example.

To begin, find all the occurrences of *beige* in the adjusted CAUS data. This step itself requires some care, because the word *beige* sometimes occurs in the data with modifiers such as *parchment beige* or *straw beige*, or in a not-quite-English compound such as *oakbeige*. In addition, the name itself is sometimes a modifier, as in *beige bisque*, or a derivative like *beigette* is used. (All these examples actually occur in the CAUS data.) A simple inclusion test looked for any occurrence of *beige* as a character string within each name; if the string *beige* occurred, the corresponding sample was taken as an instance of the colour *beige*. Exclusion tests were also used, as needed. For instance, when looking for the term *yellow*, any CAUS names with the term *yellowish*, such as *yellowish orange*, were excluded. The list of exclusions was often revised after looking at some unexpected sample names; for instance, the term *yellow green* was not allowed to occur in the list for yellow.

Each appearance of the name was checked for plausibility. In this case, a colour named *beige cloud* was listed, with Munsell coordinates 0.10B 4.27/7.07, a darkish blue that would fit nobody's idea of beige. Colours with such implausible names were eliminated. After these checks and conditions, 277 non-fluorescent CAUS samples were found that were named some form of *beige*.

The Munsell specifications for each of the 277 samples were plotted as a cloud $\mathcal{S}$ of points in the Munsell tree, shown in Figure 5. The same cloud can be plotted in the hue-chroma plane and the chroma-value plane, as shown in Figures 6 and 7. The cloud in the hue-chroma

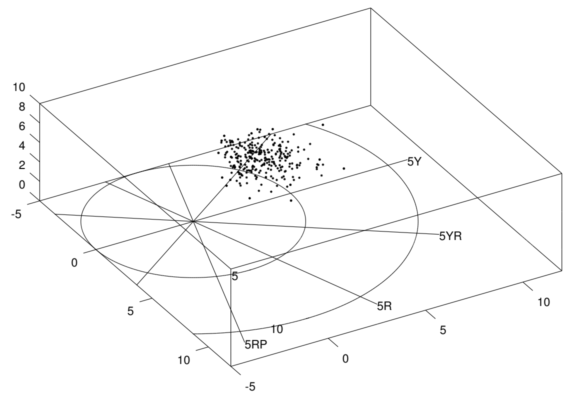

Figure 5: The Colour Name Set for Beige

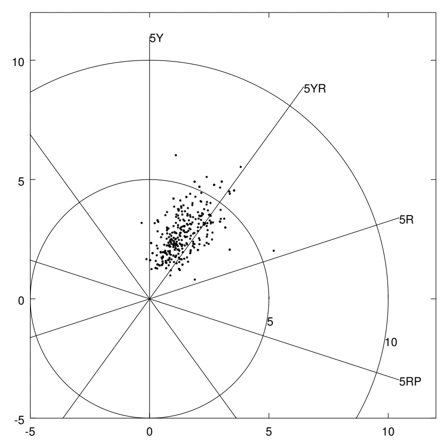

Figure 6: The Colour Name Set for Beige, in the Hue-Chroma Plane

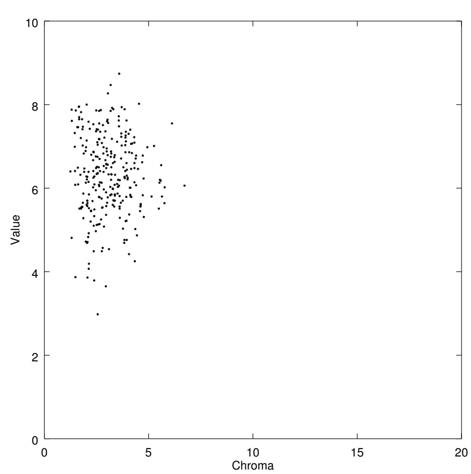

Figure 7: The Colour Name Set for Beige, in the Chroma-Value Plane

plane is the vertical geometrical projection of the three-dimensional cloud. The cloud in the chroma-value plane is a coordinate projection onto z and r, eliminating θ; it is not a geometrical projection, although it is similar to one if the hue sector is narrow. Figure 6 shows that the hues for beige cover the oranges, extending from about 0YR to 10YR, with a few points outside that sector, but all well within the "warm" hues. The chromas, as can be seen from either Figure 6 or 7, extend from about 2 to 6, so the beiges are mostly subdued but not quite neutral. The values in Figure 7 extend from about 4 to 8, so the samples are fairly light, but not as light as white.

Figures 5 through 7 show that a convex polyhedron is a reasonable representation for the beige samples, because they are centrally clustered with no gaps. Some samples, however, are also possible outliers, being somewhat separated from the main mass of the points; a few are particularly visible near the 5R line in Figure 6. On the other hand, it is not incontrovertible that these points are actually outliers. As discussed earlier, furthermore, the boundary of a colour name set is fuzzy, so, rather than make a point-by-point decision about which points to keep and which to discard, the following method was used.

The set $\mathcal{V}$ of vertices of the convex hull $\mathcal{H}$ of $\mathcal{S}$ was found. These vertices are a minimal generating set[4] for $\mathcal{H}$, and all occur on the boundary of $\mathcal{H}$. All other data points are inside the convex hull. Any potential outlier is likely in the set $\mathcal{V}$, so we will discard all points in $\mathcal{V}$, getting the set $\mathcal{S}-\mathcal{V}$. After the discarding, we calculate the convex hull Γ_{beige} of $\mathcal{S}-\mathcal{V}$, which has its own set $\mathcal{V}_{\text{beige}}$ of vertices. This procedure will eliminate genuine outliers; likely, some non-outliers will also be eliminated. The loss of some non-outliers is not a problem, however, because non-outliers are near other points, which will be retained. In fact, the resulting boundary given by $\mathcal{V}_{\text{beige}}$ has the desired fuzziness, because some colours that are outside the boundary but nearby might be considered beige by some observers, though perhaps not by most. We will refer to Γ_{beige} as the *inner convex hull*, and use it as the polyhedron for a colour name.

Figures 8 and 9 show $\mathcal{H}$ and Γ_{beige} for beige, projected into the hue-chroma plane. Figure 8 contains all the points, whether they are outliers or not. In Figure 9, some outliers, as well as some near-outliers, have been eliminated, and can be seen as points outside the inner convex hull. In fact, since this figure is a two-dimensional projection of a three-dimensional solid, there are several other outliers above and below the inner convex hull, that cannot be recognized as outliers. Figures 10 and 11 show further views of Γ_{beige} and the eliminated points. Figure 10 shows Γ_{beige} as a polyhedron in the three-dimensional Munsell tree, and Figure 11 shows Γ_{beige} in the chroma-value plane.

Once the polyhedron Γ_N has been constructed for colour name N, the centroid C_N can be calculated readily using Equations (6) through (8), and then converted back to ordinary Munsell coordinates. The beige centroid works out to 6.7YR 6.1/3.4. This colour is a dull brown with an orangish tint, that is a bit lighter than a middle gray. It corresponds fairly well to an average beige, though perhaps a bit darker than expected.

In summary, we have illustrated the construction of a polyhedron Γ_{beige} and centroid C_{beige} for beige, within the Munsell tree. Colours within the polyhedron would be called beige, and outside colours would not be called beige, with the implicit understanding that the polyhedron's boundary is fuzzy, so colours near the boundary would not be unequivocally beige or non-beige. Colours near the center, on the other hand, could confidently be described as beige, and the centroid is a good choice for the most typical representative for beige.

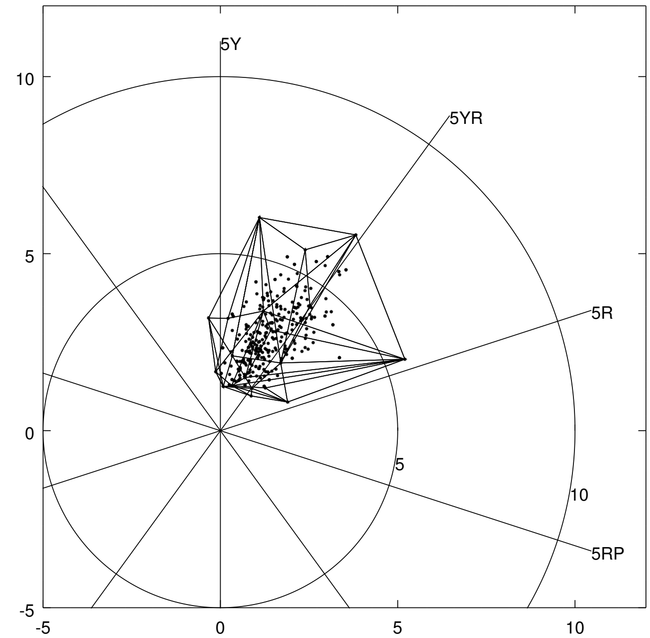

Figure 8: The Outer Convex Hull $\mathcal{H}$ of the Colour Name Set for Beige

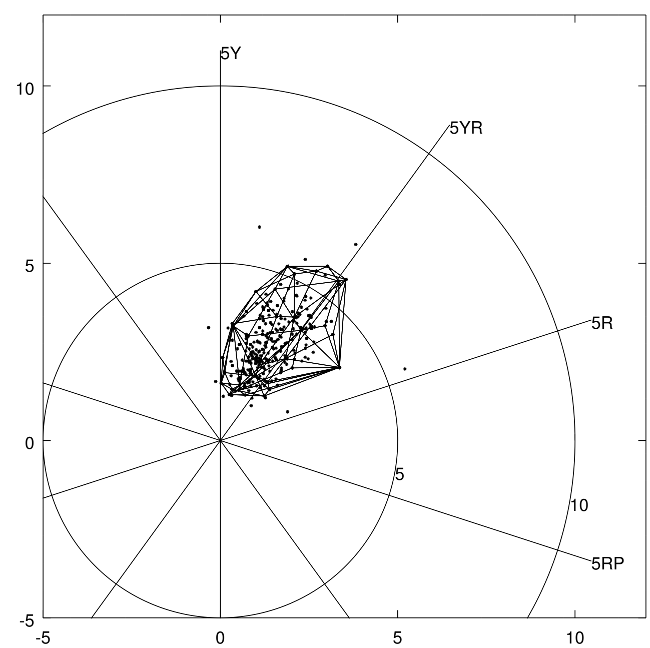

Figure 9: The Inner Convex Hull Γ_{beige} of the Colour Name Set for Beige

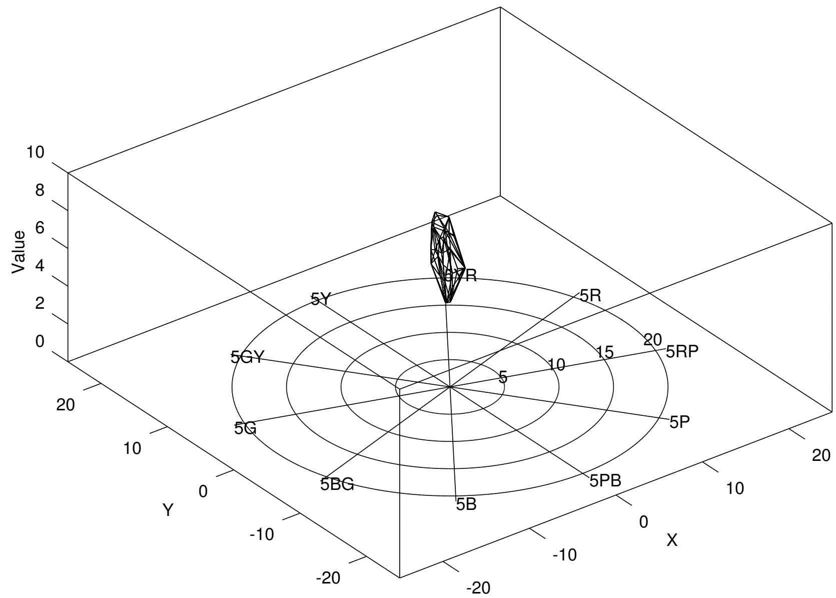

Figure 10: Three-dimensional View of the Inner Convex Hull Γ_{beige} for Beige

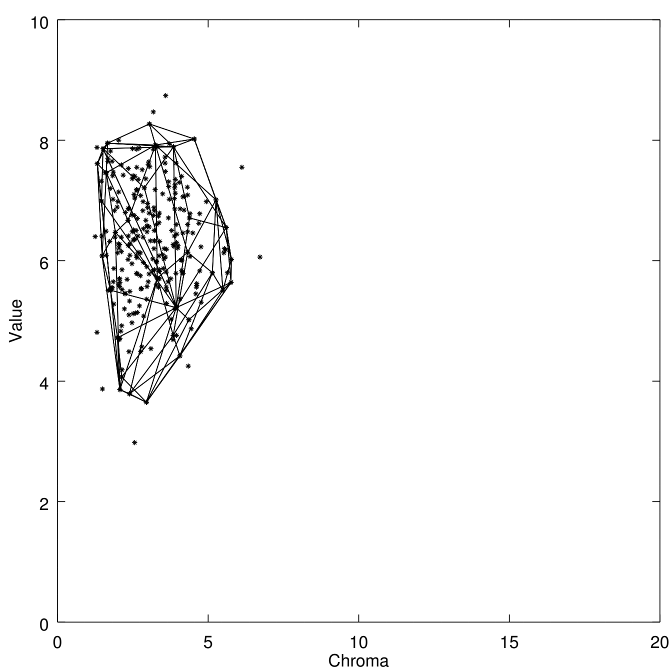

Figure 11: The Inner Convex Hull Γ_{beige} of Beige, in the Chroma-Value Plane

4.3 Scope and Limitations of the Analysis

The constructions illustrated in the previous section were applied to a set of 20 non-basic colour names (aqua, beige, coral, fuchsia, gold, lavender, lilac, magenta, mauve, navy, peach, rose, rust, sand, tan, taupe, teal, turquoise, violet, wine), as well as 10 of Berlin and Kay's 11 basic colour names (black was excluded because of insufficient data), for a total of 30 colour names in all. The names were selected from the words that occurred most frequently in the CAUS name category. No precise cutoff was used, but every name analyzed occurred in at least 25 unique, non-fluorescent samples.

The current analysis has several limitations:

1. As far as is known, the samples were viewed and assigned names under uncontrolled lighting.
2. The data contain many fanciful names such as *autumn glow* or *starlight blue*, suggesting a bias for distinctive names over more typical names.
3. The samples are all from the fashion and clothing industry, and other colour sources (natural colours, artist's pigments, house paints, etc.) do not appear. As a result, some regions of colour space might have been neglected or underrepresented.
4. A related concern is that the fashion industry uses a specialized set of colour terms, or uses standard terms in non-standard ways.
5. Little is known about the "subjects" who assigned the names. Their number is unrecorded and there is no information about their age, gender, educational level, etc. Possibly, committees rather than single individuals assigned some of the names.
6. A sample might have been measured long after its name was assigned, and possibly faded in the interim.
7. The current study is limited to English. While basic colour names have equivalents in all languages, a term such as *navy blue* might have no direct counterpart in other languages.

Despite these limitations, the data reflects how colours are perceived and worked with outside the laboratory, where lighting is usually uncontrolled, and naming or describing colours is informal. An advantage of the CAUS data over planned experiments, in fact, is that the participants assigned names without the self-consciousness, caused by laboratory settings and experimenters, that can distort results. Later, comparison with other studies will show that concerns like those listed above are typical, and probably inevitable when dealing with human perception.

Another important limitation, often overlooked, is that the current analysis deals with surface colours rather than coloured lights. Colour constancy insures that a physical object—but not necessarily a light source viewed directly—is perceived as having the same colour, even when a change in illumination alters the physical composition of the resulting visual stimulus. The Munsell system used in this paper applies only to objects colours, and it is not clear how, or even if, it can be extended to coloured lights. Some recent large-scale naming experiments use electronically produced colours, which are light sources rather than coloured objects; it is uncertain how to transfer our results for the colours of physical items to computer-generated colours, and vice versa.

5 Analysis Results

5.1 Polyhedra

Polyhedra and centroids were found for each of the 30 names. Details for each name are available in a human-readable text file named PolyhedronFiles.zip; other researchers are welcome to use these files for further analysis. Figure 12 shows an example file for *beige*. The file lists the number of samples in the CAUS data, and the centroid for that name, in both Munsell and Cartesian coordinates. The next two sections give all vertices of the convex polyhedron for beige, in both Munsell and Cartesian coordinates. The following section describes each face of the polyhedron as a triangle formed by three vertices. For example, the first triangular face is formed by the 12$^{\text{th}}$, 14$^{\text{th}}$, and 9$^{\text{th}}$ vertices in the list (the choice of Munsell or Cartesian coordinates is irrelevant). Finally, the file lists the samples from the CAUS data; each row gives the sample's assigned name, alongside its Munsell specification.

Tables 1 and 2, referring to non-basic and basic names respectively, list the centroids for the names. Figures 13 through 22 plot the 30 polyhedra geometrically, using projections as described in Section 4.2. For the name *gray* in Figure 15, many points appear to be outside the shaded region, near the neutral axis. In fact, those points are inside the polyhedron, but the polyhedron wraps completely around the circle, so the "projection" of its boundary onto the value-chroma plane produces some artifacts; in fact, the polyhedrons extend all the way to the neutral axis, and include the low-chroma colours that appear outside the shading.

5.2 Analysis of Polyhedra

The 30 polyhedra show that, with the exception of the neutral colours like *gray* and *white*, each colour name has a definite hue component, in that the entire polyhedron is restricted to a limited sector of the hue circle. For basic colour names like *green*, the hue sector can be very wide, covering more than a third of the circle. For other names, like *navy*, the sector is very narrow, being restricted to one or two standard Munsell hues.

Similar results occur for chroma. The basic colour names like *blue* and *green* extend from some very dull samples to some very saturated samples. Neutral colours like *gray* and *white*, of course, only contain dull samples. Many non-basic names, such as *sand* and *fuchsia*, contain a much more limited set of chromas; the chromas for *sand* are uniformly subdued, while the chromas for *fuchsia* are uniformly vibrant—there is no such thing as a dull fuchsia. A few non-basic names, however, such as *gold* and *rose*, manage to span ten or twelve chroma steps, so they can appear in both subdued and intense versions.

Two pairs of names, *pink* and *red*, and *brown* and *orange*, verify some common colour intuitions. First, a light red, or a light red-purple, would likely be denoted pink rather than red. The value-chroma plot for red, shown at the top right of Figure 19, contains very few values greater than 5, while the value-chroma plot for pink, shown at the middle right of Figure 18, contains very few values less than 5, even though the two names span approximately the same set of hues and chromas. While the Munsell hue of pink is red or red-purple, viewers seem to prefer the more descriptive term *pink* for light samples.

The name *brown* seems similarly to describe dark, dull versions of the seven or eight

Colour name: beige
Number of CAUS samples: 277
Centroid in Munsell coordinates: 6.66YR 6.15/3.40
Centroid in Cartesian coordinates: 1.70 2.94 6.15
Polyhedron vertices in Munsell coordinates:
8.77R 5.22/3.94
2.03YR 5.02/4.36
2.58YR 7.21/2.88
...

Polyhedron vertices in Cartesian coordinates:

3.36	2.06	5.22
3.17	2.99	5.02
2.03	2.05	7.21

...

Polyhedron faces (each entry refers to a row in the listing of vertices):

12	14	9
29	18	25
27	16	4

...

Samples, with Munsell coordinates, from CAUS data:

Sandra Pink-Beige	5.89R	6.18/5.58
Highland Beige	6.42R	5.70/2.06
Tenuto Beige	8.77R	5.22/3.94
Winebeige	1.93YR	5.70/3.38
Rose Beige No.2	2.03YR	5.02/4.36
Cocoa Beige	2.10YR	5.81/3.61
Buffalo Beige	2.11YR	7.20/1.75
Pottery Beige	2.22YR	5.45/2.24
Beige	2.29YR	6.83/3.21
Blush Beige	2.36YR	5.56/3.35
Cordbeige	2.58YR	7.21/2.88
Continental Beige	2.63YR	5.51/1.76
Muscade Beige	2.82YR	4.72/1.99
Rose Beige No.2	2.92YR	4.69/3.83
Sandalwood Beige	3.07YR	5.45/4.59
Crouton Beige	3.16YR	6.20/3.54
Townbeige	3.23YR	4.54/3.10
Powder Beige	3.24YR	6.71/4.38

...

Figure 12: Output File for *Beige*

	Colour Name	**Centroid**	**Samples**
1	Aqua	7.4BG 6.2/3.4	119
2	Beige	6.7YR 6.1/3.4	277
3	Coral	6.5R 5.8/8.3	215
4	Fuchsia	4.8RP 4.1/10.3	46
5	Gold	9.8YR 6.4/7.4	362
6	Lavender	5.6P 5.4/4.8	47
7	Lilac	7.8P 5.6/4.8	78
8	Magenta	3.8RP 3.4/9.4	25
9	Mauve	1.2RP 5.1/3.9	181
10	Navy	7.3PB 2.1/3.6	100
11	Peach	2.9YR 7.0/5.9	102
12	Rose	0.5R 5.0/7.7	467
13	Rust	9.4R 3.9/7.4	93
14	Sand	7.6YR 6.3/3.2	123
15	Tan	6.3YR 5.2/4.1	129
16	Taupe	3.2YR 4.7/1.4	76
17	Teal	1.6B 3.3/4.5	43
18	Turquoise	1.6B 5.5/5.9	121
19	Violet	7.0P 3.8/6.2	178
20	Wine	2.7R 3.0/4.9	83

Table 1: Centroids for Non-Basic Colour Names

	Colour Name	**Centroid**	**Samples**
1	Blue	0.6PB 4.7/4.6	1673
2	Brown	2.8YR 4.0/3.7	536
3	Gray	9.6P 5.3/0.5	485
4	Green	2.0G 5.3/4.7	1296
5	Orange	2.5YR 5.7/8.9	378
6	Pink	0.7R 5.8/7.9	594
7	Purple	6.8P 3.5/6.4	226
8	Red	4.3R 3.8/8.8	662
9	White	2.1Y 8.0/2.0	152
10	Yellow	2.6Y 7.4/8.0	394

Table 2: Centroids for Basic Colour Names (except Black)

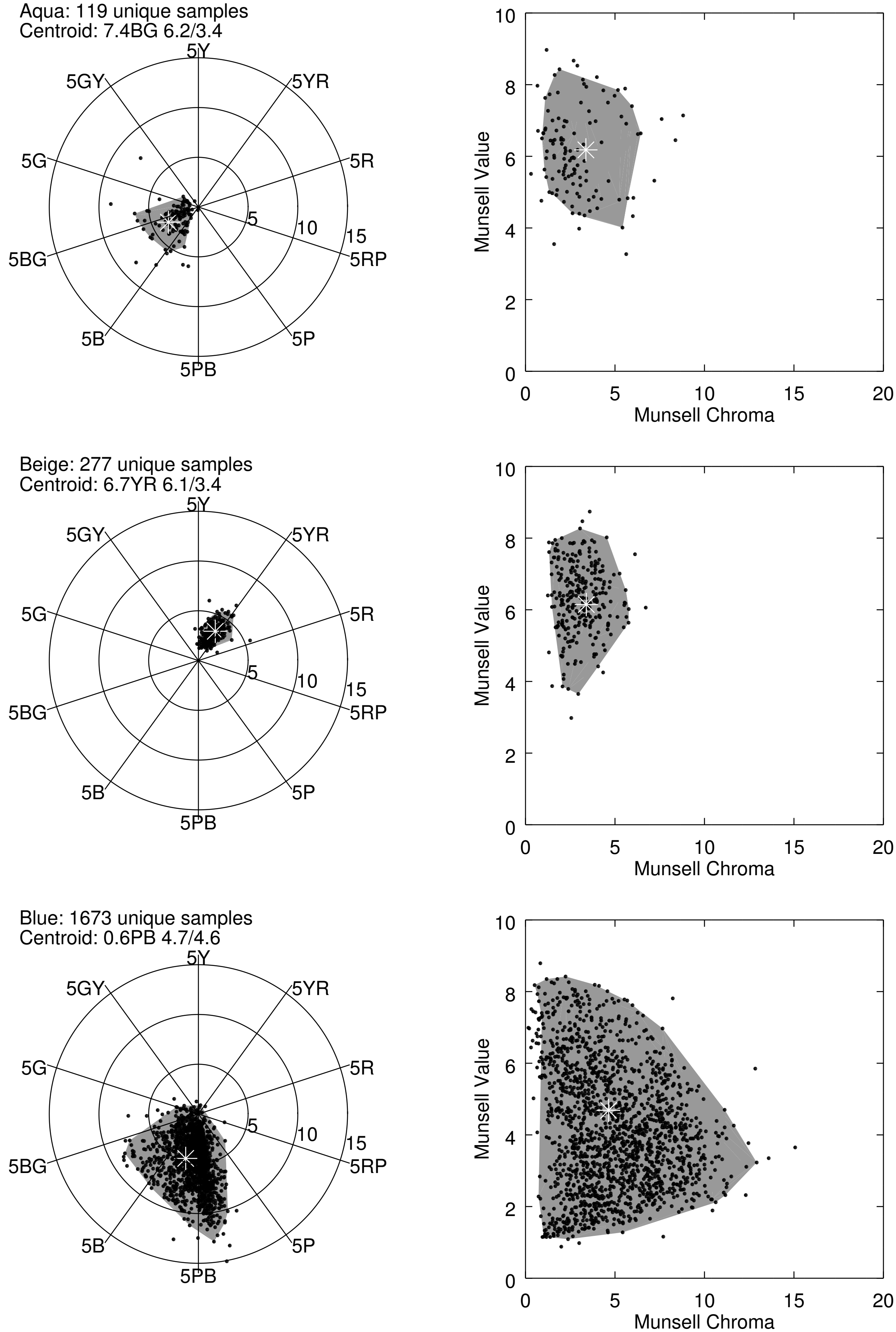

Figure 13: Polyhedra and Centroids for Names (page 1 of 10): Aqua, Beige, Blue

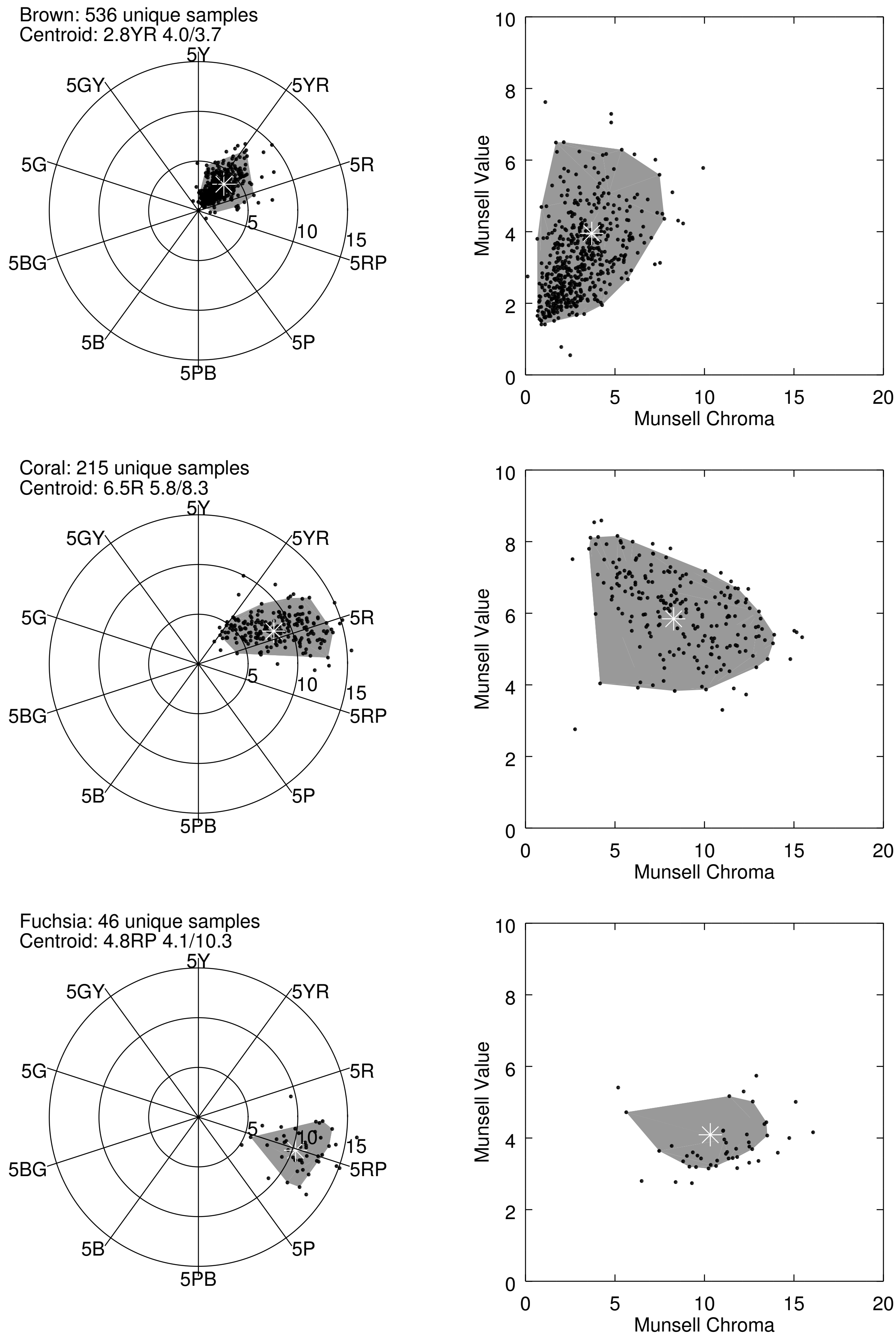

Figure 14: Polyhedra and Centroids for Names (page 2 of 10): Brown, Coral, Fuchsia

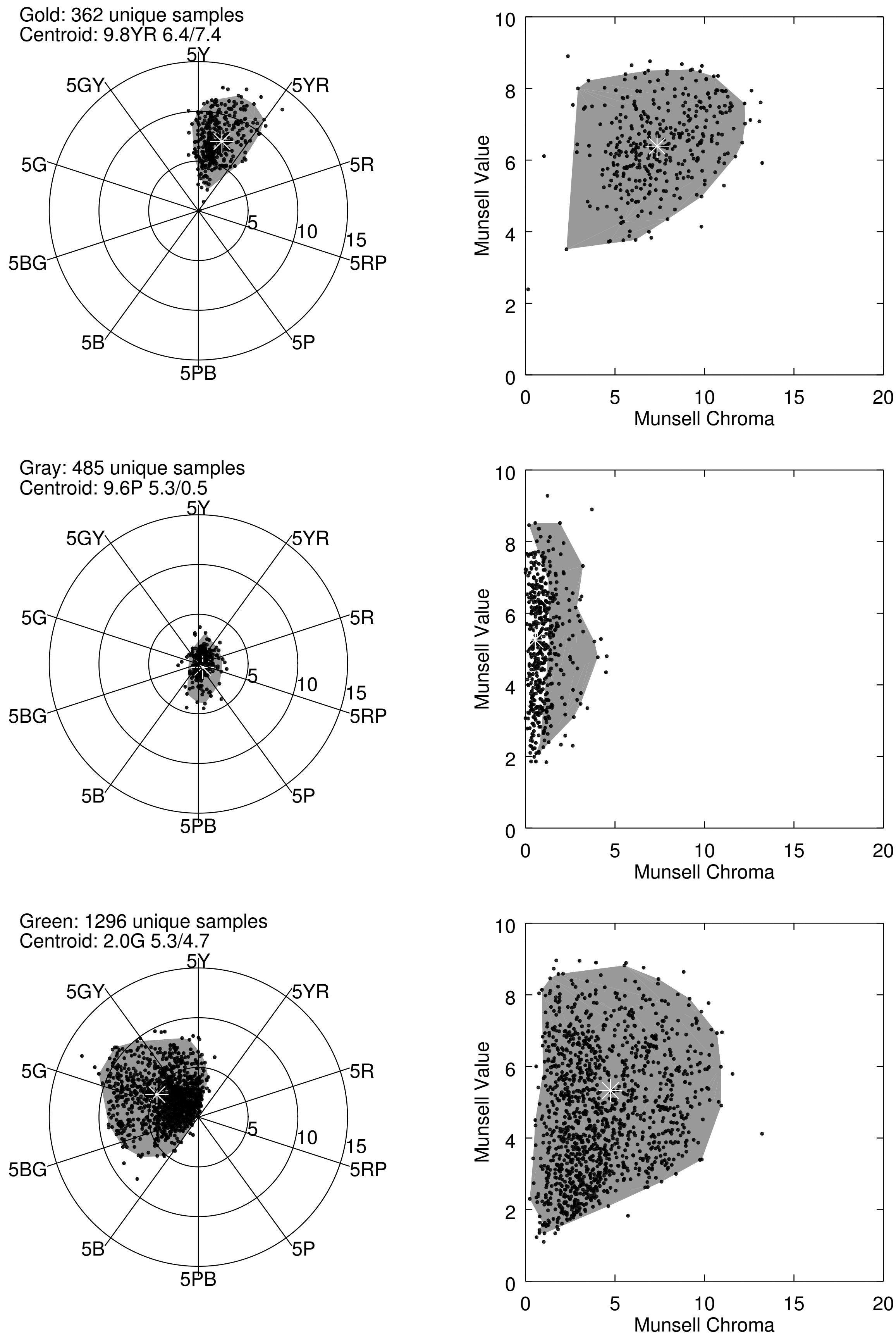

Figure 15: Polyhedra and Centroids for Names (page 3 of 10): Gold, Gray, Green

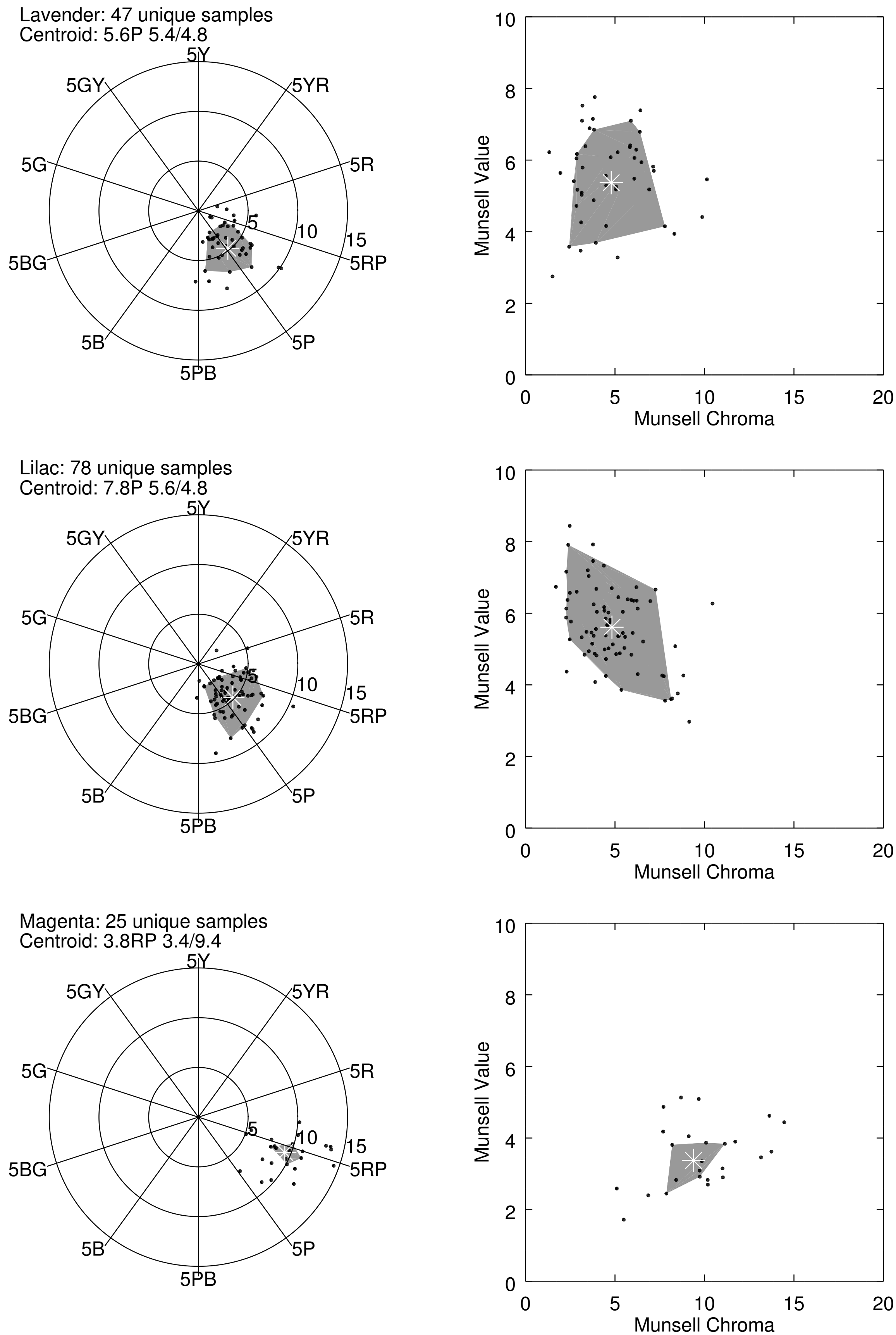

Figure 16: Polyhedra and Centroids for Names (page 4 of 10): Lavender, Lilac, Magenta

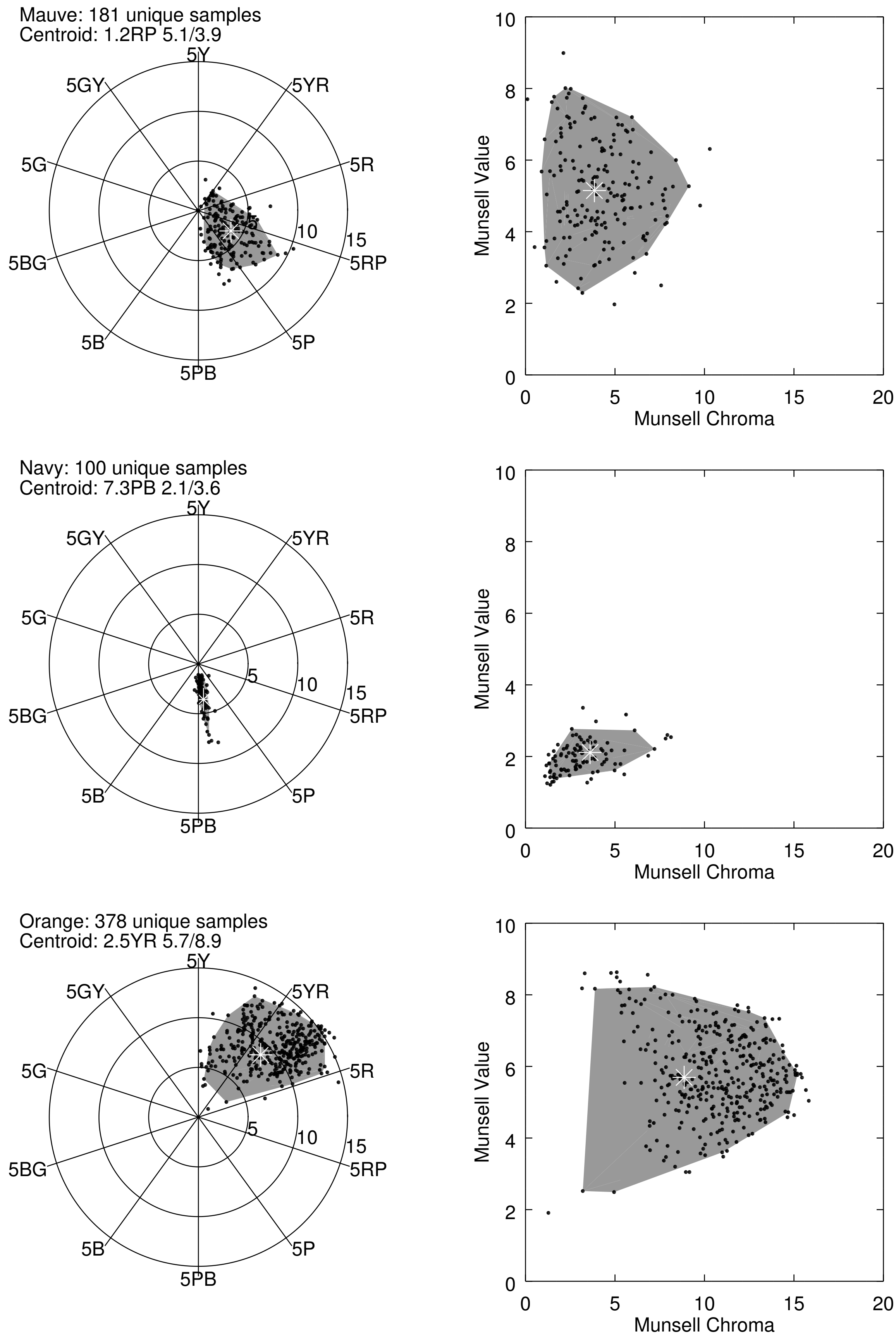

Figure 17: Polyhedra and Centroids for Names (page 5 of 10): Mauve, Navy, Orange

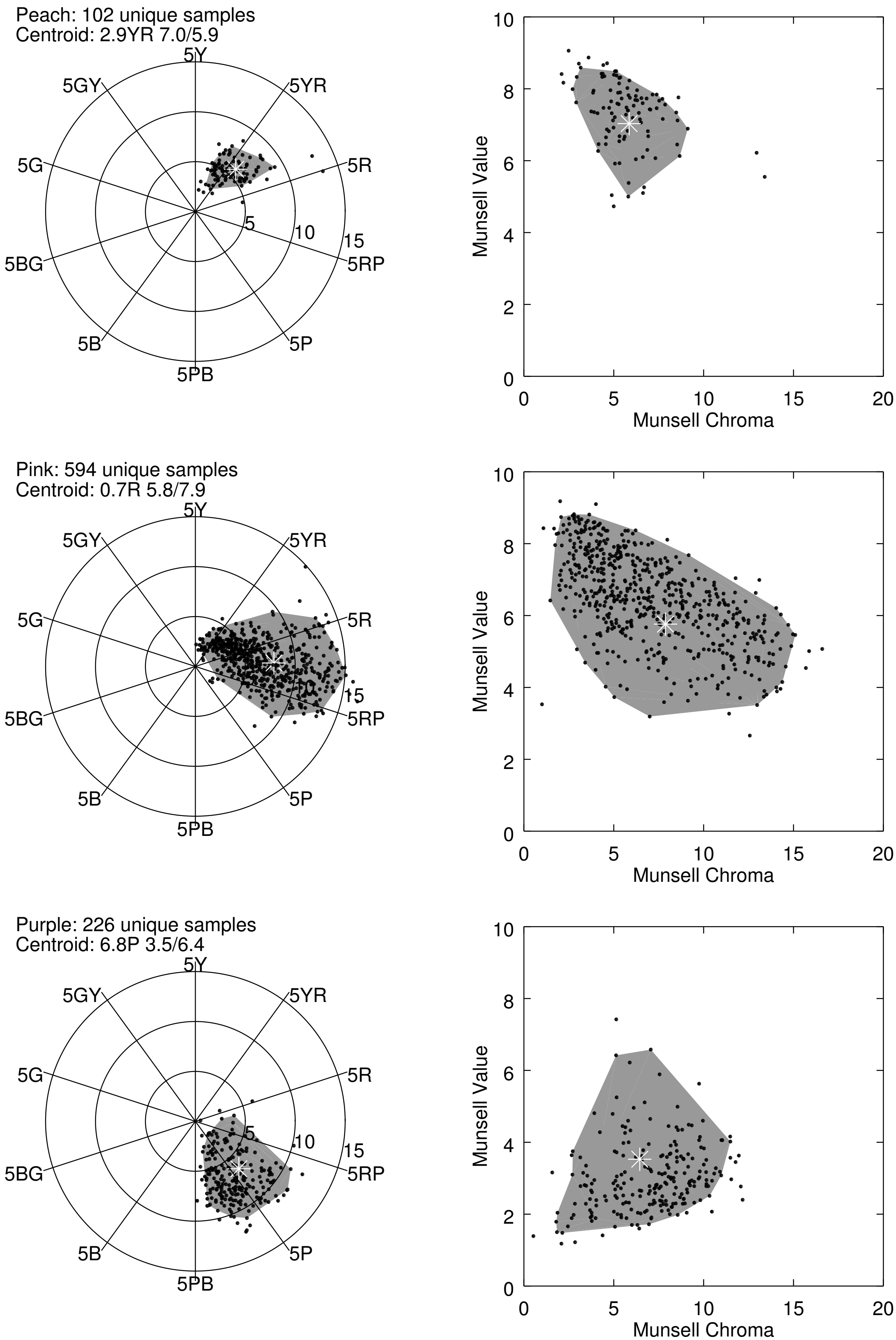

Figure 18: Polyhedra and Centroids for Names (page 6 of 10): Peach, Pink, Purple

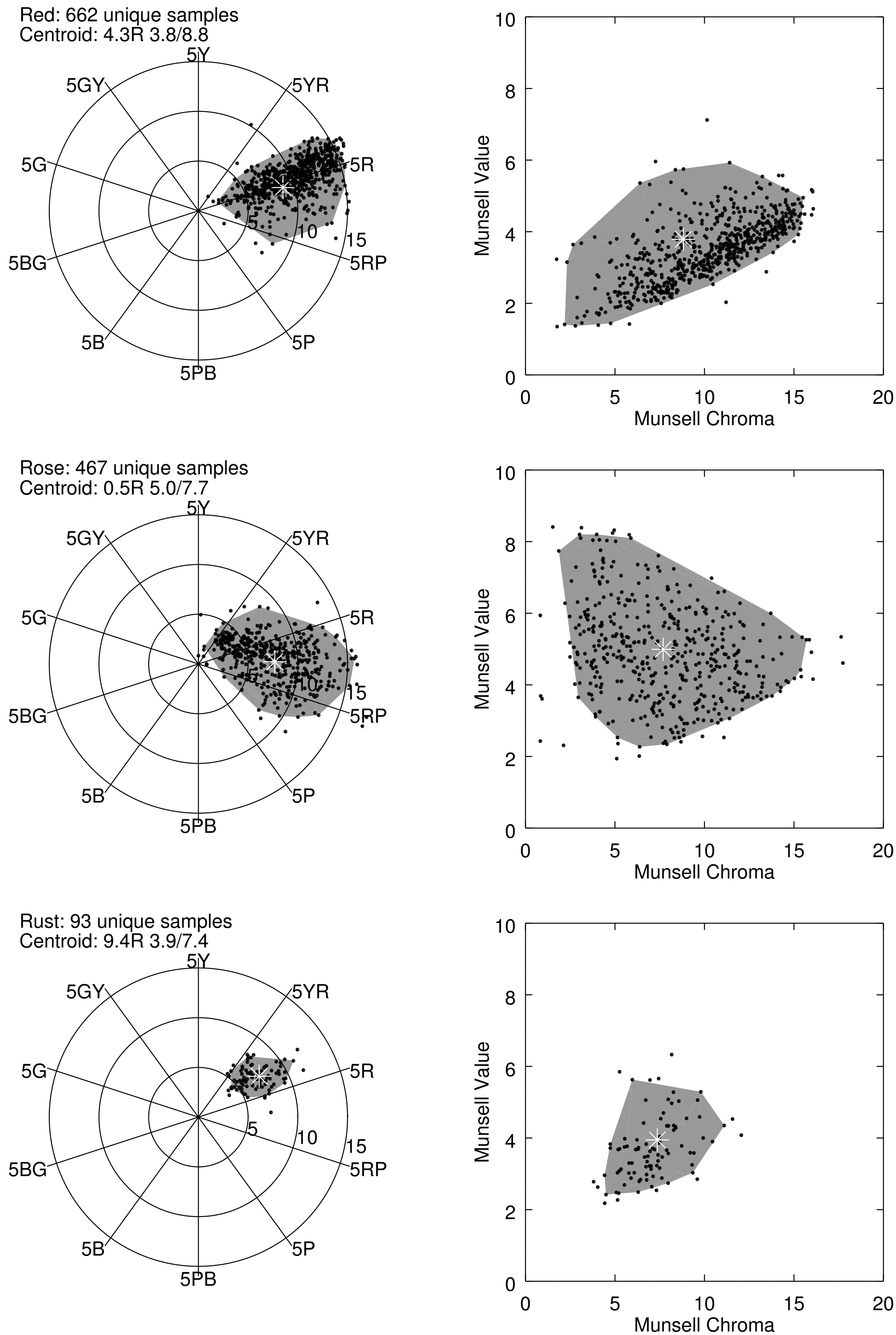

Figure 19: Polyhedra and Centroids for Names (page 7 of 10): Red, Rose, Rust

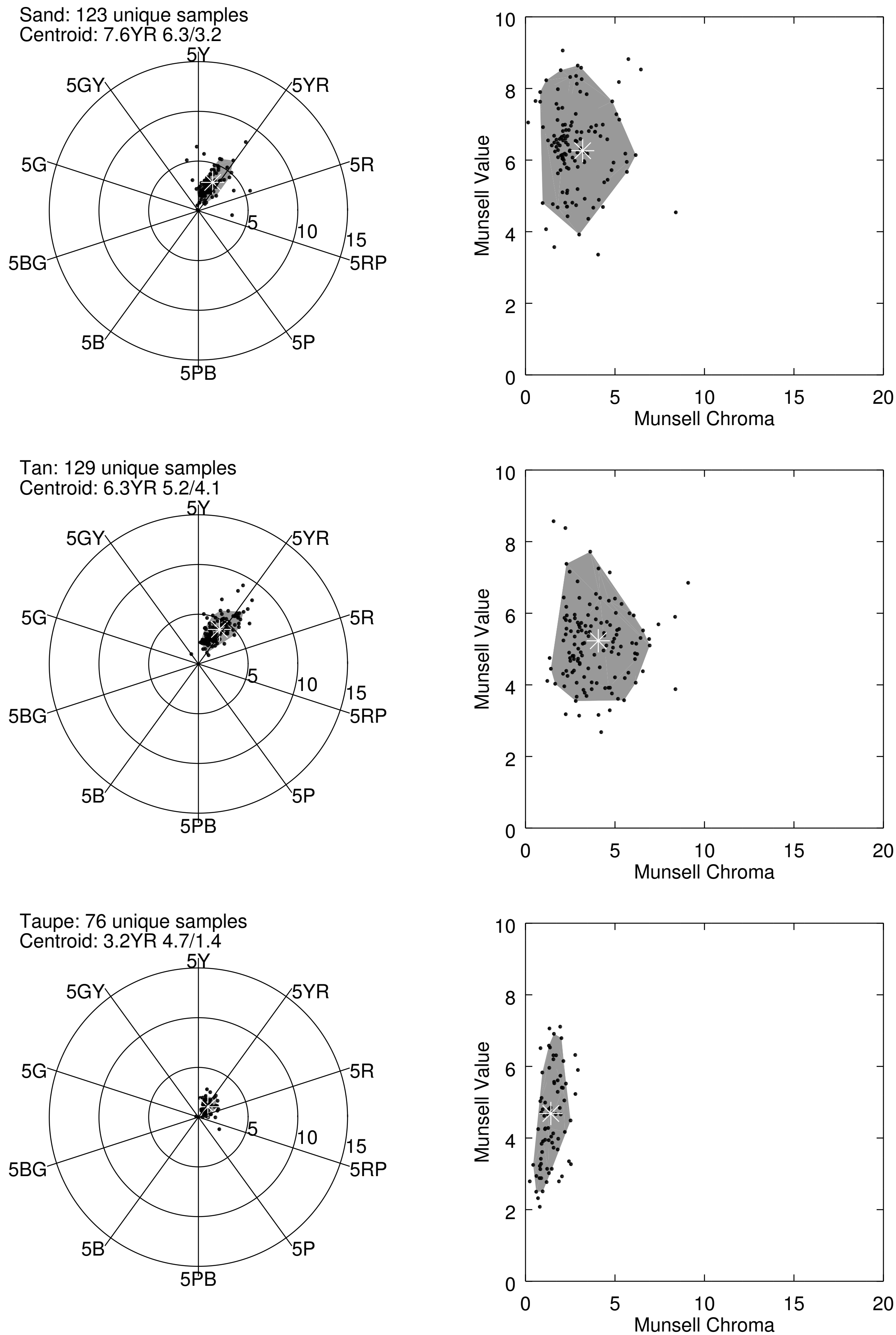

Figure 20: Polyhedra and Centroids for Names (page 8 of 10): Sand, Tan, Taupe

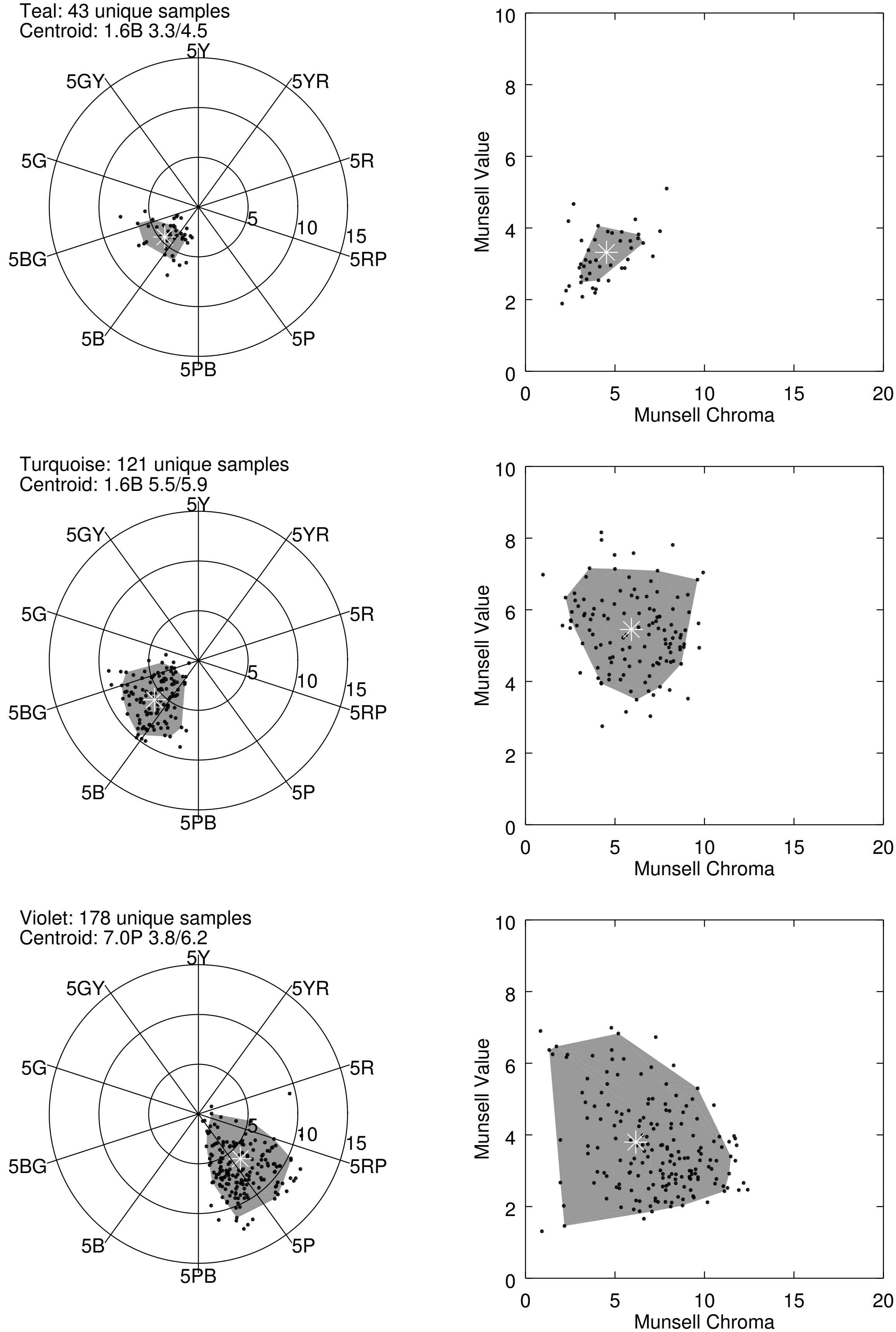

Figure 21: Polyhedra and Centroids for Names (page 9 of 10): Teal, Turquoise, Violet

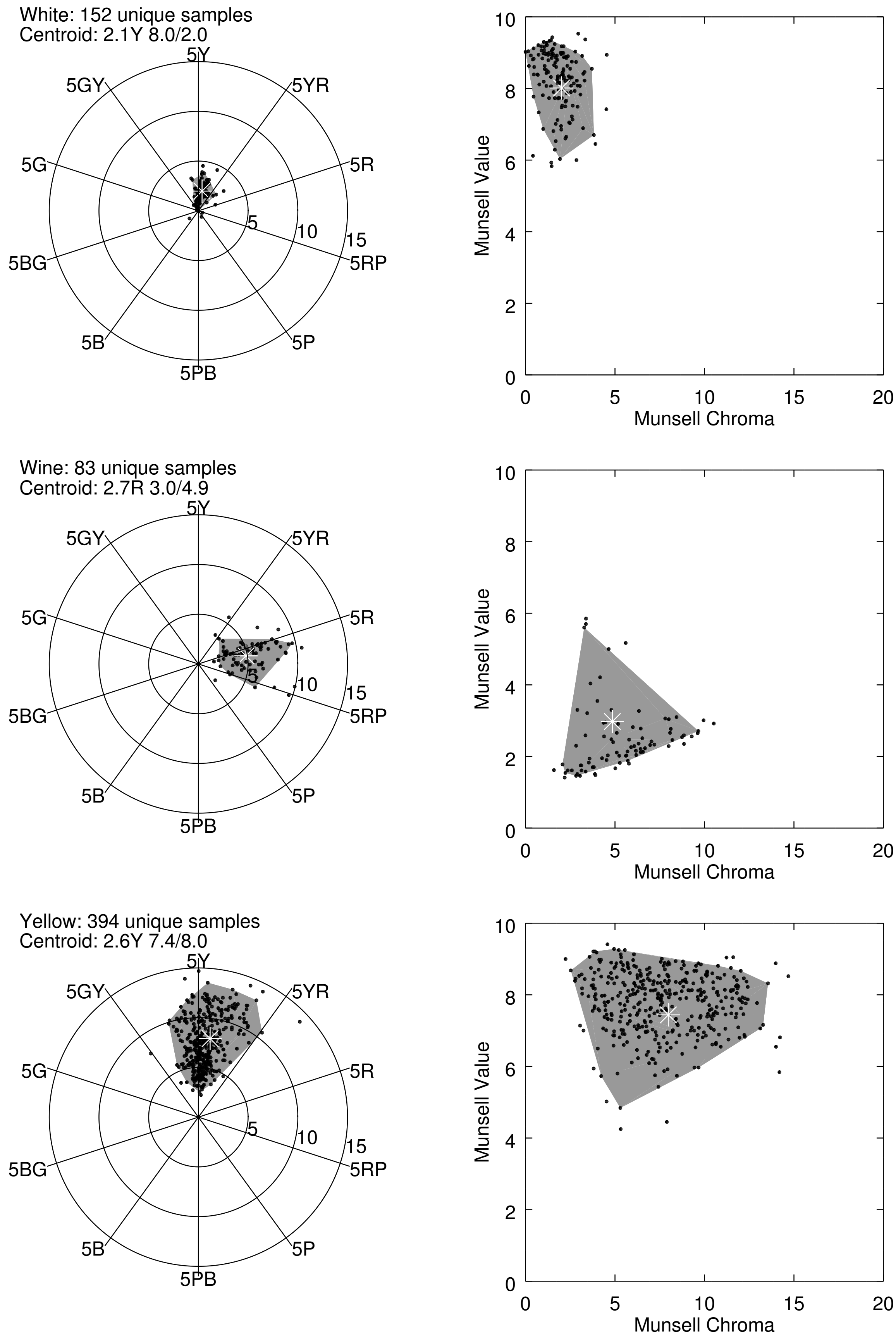

Figure 22: Polyhedra and Centroids for Names (page 10 of 10): White, Wine, Yellow

standard Munsell hues centered around orange. While the hue would be somewhere in the yellow-orange-red, or warm, sector, the term *brown* is preferred to a hue indication. This finding is in accordance with some previous work of Bartleson,[5] which determined that the definition of brown required three dimensions rather than a single hue name. In the current context, the three dimensions of hue (yellow through red), value (dark), and chroma (subdued) are needed to define the polyhedron for brown.

The polyhedron for *white* seems to exhibit a warm bias, in that most of its hues are near yellow and orange, with only a small percentage in the cooler greens, blues, and purples. Theoretically, it would be expected that *white* would apply to any colour of very low chroma and rather high value. With a low enough chroma, the hue is irrelevant, and a uniform sampling of Munsell space would produce very light, very unsaturated colours that evenly fill out a circle in the hue-chroma plane. A possible explanation is sample bias: the CAUS samples might come from a limited number of applications, such as fashion and interior decoration, that tend to prefer warm colours, at least for light neutrals. The polyhedron for *gray*, however, where hue should also be irrelevant, shows colours of all hues, suggesting that the warm/cool distinction is not as important for grays. The warm bias for white also explains the relatively high chroma of 2.0 for the white centroid; had the hues been evenly spaced, cool hues would have averaged out warm hues to produce a chroma near 0.0. The gray centroid, for example, has a chroma of only 0.5, which would have no discernible hue.

5.3 Applications

The calculated polyhedra allow two main applications: determining whether a particular sample can be described by one or more of the 30 colour names, and choosing representative colours for display.

To assign a name (or multiple names) to a physical colour sample, measure it with a spectrophotometer, and calculate its Munsell specification. Convert the Munsell specification to Cartesian coordinates, and use the polyhedron vertex and face data in each of the 30 colour name files to determine whether the sample is inside a polyhedron. If it is, then that polyhedron's colour name can reasonably be applied to the sample. Since many polyhedra overlap, multiple colour names can sometimes be assigned to the same sample. This feature mirrors human usage, in which a colour could plausibly be described as both *taupe* and *beige*, for example, or as simultaneously *pink* and *rose*.

Further distinctions can be made if desired. Since the polyhedra's boundaries are somewhat fuzzy, some applications might prefer to consider a sample near a boundary, whether on the inside or the outside, as only partially satisfying that colour name. If the sample is near the polyhedron's centroid, on the other hand, the name can be assigned with considerably more confidence. A practitioner could require a narrower definition, or accept a broader definition, as a particular application calls for.

The second application is choosing typical representative colours for design or display work. If a client wants a living room wall painted aqua, for example, Munsell specifications can be selected from inside the aqua polyhedron, and the client can make a more refined decision by viewing their exemplifications in a Munsell book. Or a cosmetic designer might be interested in coral lipstick; then the centroid for coral can be taken as a starting point, and the other colours in the coral polyhedron can be used as variations.

6 Comparison with Previous Work

The current analysis shares many common features with previous color naming investigations, but is also unique in some ways:

1. A colour name is defined as a three-dimensional polyhedral subset of Munsell space. Previous work used a single-point definition, or sometimes a two-dimensional subset of the surface of the Munsell tree. (Sivik and Taft[??] worked in three dimensions in the Natural Colour System (NCS), finding “isosemantic” level sets of colours described by a colour name with some degree of fidelity.)
2. The number of named samples is about 16,000, whereas previous surface colour studies typically have only a few hundred named samples.
3. This analysis treats non-basic colour names, while previous surface colour studies usually restricted themselves to basic names.

This section discusses these points in greater detail, in the context of other kinds of comparisons, and suggests some approaches for further work.

6.1 Surface Colours vs. Digital Colours

All the CAUS samples were physical objects, such as fabric swatches, whose colours result from reflection. Digital and electronic devices, by contrast, produce coloured light sources directly. As mentioned earlier, conversions between object and light colours are uncertain, so it is difficult to transfer surface colour names to digital colour names. This fact is unfortunate, because there are two large-scale digital naming studies.

During the early 2000s, Nathan Moroney[6] collected a large digital data set, totaling over 30,000 assigned colour names from over 5000 volunteers. The simple procedure involved a central server that randomly produced RGBs and displayed them over the internet on observers’ own computers. The observers assigned names to the displayed colours, and submitted the names electronically. Data analysis produced a central estimate, or focal colour, expressed as an sRGB triple, for each colour name. Reference 6 is a printed version of the results. Table 3 gives the number of occurrences for the names, using Moroney’s categories of *very common*, *common*, *rare*, and *very rare*. In all, 663 colour names occurred. Moroney’s choice of names was sometimes finer-grained than ours; for instance, *navy* and *navy blue*—whose focal sRGBs are practically identical—are treated as two separate names, while we treated them as one.

Description	Occurrences	Number of Names
very common	> 500	12
common	$50 - 500$	78
rare	$5 - 50$	333
very rare	< 5	240
Total		**663**

Table 3: Statistics for Moroney’s Survey

In 2010, Randall Munroe[7] ran a colour survey similar to Nathan Moroney's, achieving over 5 million colour assessments from 222,500 participants. Munroe's experiment also randomly served RGBs for observers to view on their own computers. Analysis resulted in focal sRGBs for the 954 most common colour names. The names and focal sRGBs have been posted online at www.xkcd.com, and can be freely downloaded as a text file.

6.2 Control vs. Quantity

Colour-naming investigations typically face a tradeoff between the quantity of data collected, and the control of the conditions. Many laboratory experiments[5,8,9,10] present a limited set of carefully measured colour samples to subjects (who have typically been screened for colour-blindness), under the controlled illumination of a light box. This approach reduces variability, but usually only a few hundred colour assessments can be made.

A competing approach, exemplified by Moroney's and Munroe's internet-based surveys, collects a very large quantity of data, at the expense of greater uncertainty. Their internet surveys assumed the sRGB standard[11] held universally, meaning that all monitors were sRGB-compliant, calibrated, and viewed under dim D50 ambient lighting. Likely none of these conditions held exactly, but, for lack of a better alternative, the sRGB model was used, to handle the uncertainty.

Similarly, the CAUS data set features many assessments, but under uncontrolled conditions. The viewing illumination is not known, and could have been different for different samples, or even for the same sample at different times. The participants' motives were also uncertain, and could have more to do with marketing than with accurate description. While the set of 16,000 assignments is large enough to analyze non-basic names, we have to accept the accompanying uncertainty.

6.3 Forced vs. Unforced Choice

Another variable in colour-naming studies is the source of the names that can be assigned. In Berlin and Kay's original work, a subject was given a colour name such as *red*, and shown a grid of Munsell chips of maximum chroma, coordinatized by hue and value. The subject would identify which sections of the grid he considered red, and then choose the chip he thought was most typically red. Many follow-on studies used the same format. Some varied the procedure by asking subjects to choose a name from a small pool of names, This approach might be called *forced choice*.

The current study, by contrast, as well as most computer-based studies, are open-ended, allowing subjects to choose any name they want, resulting in a greater variety of names, and finer definitions. The differing approaches reflect differing aims. Berlin & Kay were more interested in linguistics and cognition than colours, and were looking for perceptual universals that held regardless of language. The CAUS names, on the other hand, aimed, at least on some level, to describe colours.

6.4 Form of Definition

This paper delimits a colour name by a polyhedral subset of three-dimensional Munsell space. Berlin and Kay's original work presented subjects with a two-dimensional grid of high-chroma Munsell chips, consisting of only those colours on the outer surface of the Munsell tree. Subjects indicated the limits of a name such as *red* on this grid. Lower-chroma colours, inside the tree, were therefore not considered. (Some later studies, by Sturges and Whitfield,[9] and Boynton and Olson,[12] did include some interior colours, as did Bartleson's 1976 investigation[5] of brown.) Duller colours like beige and navy blue could therefore not be treated. Furthermore, even most basic colours are incompletely represented, because their lower-chroma versions never occur. This paper's approach, by contrast, is fully three-dimensional, allowing any Munsell colour as a sample.

Other colour definitions are not extended sets at all, but rather a single focal colour, either a Munsell specification or an sRGB, that best represents a given colour name. A single-point definition is helpful, but incomplete, because most observers will readily agree that a name can apply to a range, sometimes a very wide range, of easily distinguishable colours. Moroney's and Munroe's internet studies, where names are assigned to sRGBs rather than Munsell coordinates, could easily perform our polyhedron constructions in the sRGB cube, to map those ranges more informatively. Our polyhedron for *magenta* needed only 25 data points. Table 3 shows that at least 90 of Moroney's names have over 50 data points, and probably many more names have at least 25, so polyhedra could be generated for a large set of names, much larger than the 30 names treated in this paper.

6.5 Possible Further Work

Colour-naming analysis require many human assignments of names to colour samples, and the more comprehensive the analysis, the more assignments are needed. The 16,000 entries of the CAUS data set, for example, allowed previously unanalyzed names like *taupe* and *rust*. Further work would require further large data sets. The data needed has a simple form: a list of physical samples to which a human has assigned a name, and the Munsell specifications (or the reflectance spectra, from which Munsell specifications can be calculated) for those samples.

In addition to the CAUS data set, some well-known earlier data sets could be analyzed:

1. Robert Ridgway's 1912 *Color Standards and Color Nomenclature*,[13] measured in 1949 by D. H. Hamly.[14]
2. Maerz and Paul's 1930 *Dictionary of Color*,[15] measured by various workers,[16] with results used in Kelly and Judd's dictionary of colour names.[17]
3. The Royal Horticultural Society's 1938 *Horticultural Colour Chart*,[18] whose Munsell specifications were determined visually in 1957 by Dorothy Nickerson.[19] This colour chart is currently in its sixth edition, and still available. Measurements have been made for some later editions, too.
4. Gladys and Gustave Plochere's 1948 *Plochere Color System*,[20] whose Munsell specifications were determined visually in 1949 by W. E. Knowles Middleton.[21]

Sect. 9 of Kelly and Judd's *Color Names Dictionary*, contained in Ref. 17, lists references to several other sources.

Studies involving monitor colours use electronic communication to achieve hundreds of thousands of responses with minimal overhead, and minimal time and effort from the participants. Surface colours could be analyzed similarly, though with more overhead. The author, for example, has assembled a shade bank of about 10,000 colours, with measured reflectance spectra, for a particular printer and paper. One could randomly select a dozen RGB colours from the bank, print those colours on a sheet, and mail that sheet to a participant, who could assign names and submit them electronically. This method, of course, would incur costs of printing and distribution, as well as the typical administrative costs.

7 Summary

This paper has used a large data set, of over 16,000 samples, provided by the CAUS, to assign definitions to 30 colour names. 20 of the names (aqua, beige, coral, fuchsia, gold, lavender, lilac, magenta, mauve, navy, peach, rose, rust, sand, tan, taupe, teal, turquoise, violet, wine) are non-basic, and technical definitions for them had not been studied previously. Each name is defined as the set of colours filling a convex polyhedron in the Munsell tree. A Munsell colour inside that polyhedron would be assigned that name, while colours outside wouldn't be; one recognizes implicitly, of course, that the polyhedron is fuzzy, so colours near its boundary cannot be assigned or dis-assigned that name with much confidence. In addition, the centroid of a polyhedron is taken as the most typical representative for a particular name. Illustrations of the 30 polyhedra and centroids are provided, as are text files from which the illustrations can be produced; other researchers are welcome to use these text files for further analysis. After suggesting some applications, this paper's analysis of surface colour names is compared to previous analyses. The current study is novel in that it works directly in three-dimensional colour space, uses a very large data set, and treats non-basic names. The paper concludes with some suggestions for further investigation.

References

1. Brent Berlin & Paul Kay, *Basic Color Terms: Their Universality and Evolution*, CSLI Publications, 1999.
2. Paul Centore, "An Open-Source Inversion Algorithm for the Munsell Renotation," Color Research & Application, Vol. 37, No. 6, pp. 455-464, December 2012.
3. Sidney Newhall, Dorothy Nickerson, & Deane B. Judd. "Final Report of the O. S. A. Subcommittee on the Spacing of the Munsell Colors," JOSA, Vol. 33, Issue 7, pp. 385-418, 1943.
4. S. R. Lay, *Convex Sets and Their Applications*, Dover Publications, 2007.
5. C. James Bartleson, "Brown," Color Research and Application, Vol. 1, No. 4, Winter 1976.
6. Lars Sivik & Charles Taft, "Color naming: A mapping in the NCS of common color terms," Scandinavian Journal of Psychology, Vol. 35, pp. 144-164, 1994.
7. Nathan Moroney, *The Color Thesaurus*, Hewlett-Packard Laboratories, June 2008.

8. Randall Munroe, www.xkcd.com, accessed in June 2017.
9. Robert Benavente, Maria Vanrell, & Ramon Baldrich, "A Data Set for Fuzzy Colour Naming," Color Research & Application, Vol. 31, No. 1, pp. 48-56, February 2006.
10. Julia Sturges & T. W. Allan Whitfield, "Locating Basic Colours in the Munsell Space," Color Research and Application, Vol. 20, No. 6, December 1995.
11. Julia Sturges & T. W. Allan Whitfield, "Salient Features of Munsell Colour Space as a Function of Monolexemic Naming and Response Latencies," Vision Research, Vol. 37, No. 3, pp. 307-313, 1997.
12. Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB, IEC 61966-2-1:1999, International Electrotechnical Commission, 1999.
13. Robert M. Boynton & Conrad X. Olson, "Locating Basic Colors in the OSA Space," Color Research & Application, Vol. 12, No. 2, pp. 94-105, April 1987.
14. Robert Ridgway, *Color Standards and Color Nomenclature*, self-published, 1912.
15. D. H. Hamly, "The Ridgway Color Standards with a Munsell Notation Key," Journal of the Optical Society of America, Vol. 39, Issue 7, pp. 592-599, 1949.
16. A. Maerz & M. R. Paul, *A Dictionary of Color, 1st ed.*, McGraw-Hill, 1930.
17. Dorothy Nickerson, "Interrelation of Color Specifications," Paper Trade Journal, Vol. 125, Issue 19, pp. 153-171, 6 November 1947.
18. Kenneth L. Kelly & Deane B. Judd, "Color: Universal Language and Dictionary of Names," NBS Special Publication 440, 1976. Available online at http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nbsspecialpublication440.pdf or https://ia801701.us.archive.org/9/items/coloruniversalla00kell/coloruniversalla00kell.pdf
19. Royal Horticulture Society, *Horticultural Colour Chart*, Robert F. Wilson, 1938.
20. Dorothy Nickerson, "*Horticultural Colour Chart* Names with Munsell Key," Journal of the Optical Society of America, Vol. 47, No. 7, pp. 619-621, 1957.
21. Gladys & Gustave Plochere, *Plochere Color System: in book form*, self-published, 1948.
22. W. E. Knowles Middleton, "The Plochere Color System: A Descriptive Analysis," Canadian Journal of Research, Vol. 27. Sec. F, No. 1, pp. 1-21, January 1949.

