CrossMark

Database of synesthetic color associations for Japanese kanji

Daisuke Hamada[1] · Hiroki Yamamoto[1] · Jun Saiki[1]

Abstract Synesthesia is a neurological phenomenon in which certain types of stimuli elicit involuntary perceptions in an unrelated pathway. A common type of synesthesia is grapheme–color synesthesia, in which the visual perception of letters and numbers stimulates the perception of a specific color. Previous studies have often collected relatively small numbers of grapheme–color associations per synesthete, but the accumulation of a large quantity of data has greater promise for uncovering the mechanisms underlying synesthetic association. In this study, we therefore collected large samples of data from a total of eight synesthetes. All told, we obtained over 1000 synesthetic colors associated with Japanese kanji characters from each of two synesthetes, over 100 synesthetic colors form each of three synesthetes, and about 80 synesthetic colors associated with Japanese hiragana, Latin letters, and Arabic numerals from each of three synesthetes. We then compiled the data into a database, called the KANJI-Synesthetic Colors Database (K-SCD), which has a total of 5122 colors for 483, 46, and 46 Japanese kanji, hiragana, and katakana characters, respectively, as well as for 26 Latin letters and ten Arabic numerals. In addition to introducing the K-SCD, this article demonstrates the database's merits by using two examples, in which two new rules for synesthetic association, "shape similarity" and "synesthetic color clustering," were found. The K-SCD is publicly accessible (www.cv.jinkan.kyoto-u.ac.jp/site/uploads/K-SCD.xlsm) and will be a valuable resource for those who wish to conduct statistical analyses using a rich dataset in order to uncover the rules governing synesthetic association and to understand its mechanisms.

Synesthesia is a neurological phenomenon in which ordinary stimuli elicit vivid individual perceptions in unrelated pathways, without a corresponding physical stimulus. For example, grapheme–color synesthetes experience subjective colors upon viewing black alphanumeric characters (Dixon, Smilek, & Merikle, 2004; Ramachandran & Hubbard, 2001). Hearing–color synesthetes similarly "see" colors upon hearing particular sounds (Baron-Cohen, Wyke, & Binnie, 1987). Lexical–gustatory synesthetes "taste" particular flavors and experience particular oral textures upon hearing/reading particular words (Ward & Simner, 2003). These synesthetic perceptions arise without a corresponding sensory input. Therefore, studies on these forms of synesthesia must isolate the processes of consciousness from sensory processing, and in doing so may provide a clue to understanding perceptual consciousness. Indeed, brain-imaging studies conducted among synesthetic participants have provided evidence identifying brain regions important for color perception (Hubbard, Arman, Ramachandran, & Boynton, 2005; Nunn et al., 2002; Sperling, Prvulovic, Linden, Singer, & Stirn, 2006). Synesthesia may also provide insight for understanding various aspects of cognition, including mental imaging (Barnett & Newell, 2008), memory (Luria, 1968; Tammet, 2006), art and creativity (Rich, Bradshaw, & Mattingley, 2005; Ramachandran & Hubbard, 2001), and numeracy (Rich et al., 2005).

Most of the knowledge of synesthesia has come from studies of grapheme–color synesthetes, who experience the most

✉ Hiroki Yamamoto
yamamoto@cv.jinkan.kyoto-u.ac.jp

[1] Graduate School of Human and Environmental Studies, Kyoto University, Yoshida, Nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan

common form of synesthesia. In this form, the visual perception of letters and numbers (graphemes) induces simultaneous perception of a given color (e.g., the letter "F" may be green, and the number "2" may be red; Cytowic & Eagleman, 2009). Two important characteristics of this synesthetic experience should be noted. First, the grapheme–color association is surprisingly strong. Rich et al. (2005) showed that grapheme–color pairs were consistent in individuals, with almost no change since childhood. Asano and Yokosawa (2012) showed that kanji–color pairs also have consistency over time. Although a synesthete occasionally acquires different synesthetic color associations from his or her original ones, these new color associations are temporary and are lost over time (Mills et al., 2002). This consistency of associations has been used as a diagnostic criterion for synesthesia. Using this criterion, the prevalence of grapheme–color synesthesia has been estimated at anywhere from 1 in 2000 (Baron-Cohen, Burt, Smith-Laittan, Harrison, & Bolton, 1996) to 1 in 200 or more (Ramachandran & Hubbard, 2001). The second important characteristic is that the phenomenological aspect of color sensation in grapheme–color synesthesia is very heterogeneous across individuals. Dixon et al. (2004) showed that some grapheme–color synesthetes, termed "projectors," perceived their associated colors visually in external space, characterizing them as existing "out there on the page." Others, termed "associators," perceived their colors in internal space, characterizing them as existing "in my mind's eye" or "in my head." Researchers often use this categorization to classify synesthetes. But currently there are controversies regarding the validity of the "associator" versus "projector" categorization (see Chiou & Rich, 2014).

What determines the associations between a grapheme and a color? There is some evidence of commonality between individuals. English-speaking synesthetes often associate synesthetic colors with the initial letter of common color words, such as "R" being red and "G" being green (Rich et al., 2005; Simner et al., 2005), and with phonological information, such as "I" [a] being white [wa t] (Rich et al., 2005). However, several studies have shown that associations are idiosyncratic between individuals (Laeng, Svartdal, & Oelmann, 2004; Ward, Li, Salih, & Sagiv, 2007). For example, when shown the letter "B," one individual may report blue, another green, and others yellow. Even synesthetic monozygotic twins report different colors stimulated by the same letter (Rich et al., 2005). These aspects of synesthetic associations suggest that direct correspondences between a grapheme and a color (a first-order relation) are too elusive to guide the investigation of possible associative mechanisms.

Given this difficulty with establishing first-order relations, recent studies have begun to explore correspondence between the relationships among graphemes, on the one hand, and the relationships among colors, on the other (a second-order relation; Watson et al., 2012). Brang, Rouw, Ramachandran, and Coulson (2011) showed that similarity between synesthetic colors depended on similarity in the shapes of letters. Eagleman (2010) argued that letters early in their respective alphabets (e.g., A, B, C, D) tended to be associated with colors that are more distinct from each other, whereas letters that come later (e.g., V, W, X, Y) tended to be associated with colors that were quite similar to each other. This ordinal relationship has also been found for Japanese hiragana characters (Asano & Yokosawa, 2013).

Accumulating a drastically larger dataset of first- and second-order relationships between graphemes and colors would be a promising way, if not the only effective way, to understand the mechanisms underlying synesthetic associations. Such a dataset would allow us to more thoroughly analyze synesthetic mechanisms from various angles. There have been forays into this approach; for instance, Rich et al. (2005) examined 192 adult synesthetes and showed that certain letters and numerals are likely to induce specific colors (i.e., R → red, Y → yellow, D → brown). This study also revealed that synesthetic colors could be derived from parts of conventional sequences such as days of the week, letters, or numerals. Eagleman (2010) launched the Web-based Synesthesia Battery, and this project has verified almost 10,000 synesthetes across 15 types of synesthesia and eight languages. In fact, the ordinal relationship described above was found by using this battery. Although these databases have great utility, they have one fundamental limitation: The number of synesthetic colors per synesthete is relatively small. Previous studies have often been limited to color associations with only 26 Latin letters (A–Z) and ten Arabic numerals (0–9), and this relatively small sample per synesthete confines the analysis to the global structure of synesthesia common across synesthetes. If samples per synesthete could be greatly enlarged, we might uncover novel rules governing the fine structure of synesthetic associations.

Therefore, in this study we took a new approach, complementary to existing databases, in which thousands of synesthetic colors were collected from each of two synesthetes, hundreds of synesthetic colors from each of three synesthetes, and about 80 synesthetic colors from each of three synesthetes. This study had two major goals. The first goal was to create a database that includes synesthetic colors for 483 kanji, 46 hiragana, and 46 katakana. Hiragana and katakana are phonetic characters, whereas kanji are Japanese logographic characters, each of which has its own meaning. These characters originated from Chinese and are mainly used to write content words. Kanji have at least two important characteristics, as they relate to synesthetic color associations. The first is that kanji are visually composed of multiple orthographic components, called *radicals*. Most kanji have two or more radicals, and many kanji characters share common radicals. For example, the characters 決, 河, 活, and 流 share the common *sanzui* radical, 氵, which signifies water. Radicals may be linked to colors, and radical commonality may

underlie the generation of synesthetic colors (Asano & Yokosawa, 2012). The second important characteristic is that most kanji have multiple readings, termed the *on-yomi* and *kun-yomi*. *On-yomi* are derived from the original Chinese pronunciation of the character, whereas *kun-yomi* come from the Japanese pronunciation (Tamaoka, 1991). For example, the kanji 青 (meaning "blue") is pronounced /sho/ or /sei/ in its *on-yomi* reading, and /ao/ in its *kun-yomi* reading. By taking advantage of these specific characteristics of kanji, we succeeded in collecting a number of synesthetic colors associated with kanji characters from each of five synesthetes.

The second goal of this study was to propose new research directions using the developed database. The new database should allow more thorough analyses of the structure of the association between a grapheme and a color. We illustrate two promising directions here. The first is the analysis of the first- and second-order relationships between a given kanji character and synesthetic color, taking advantage of kanji's multifaceted properties for the study of regularities. Asano and Yokosawa (2011, 2012) previously reported that synesthetic kanji–color associations depend on both the semantic and phonological information of the character. The larger samples of kanji–color synesthetic pairs in our database will extend these findings regarding the structure of synesthetic associations. The second direction concerns a fundamental yet overlooked issue in the literature: Can any color be a synesthetic color for a synesthete? More explicitly, are synesthetic colors distributed randomly in the color space? Notably, this question cannot be answered without collecting a large amount of data for a single synesthete, which our study provides. We will illustrate techniques of spatial statistics to explore the pattern of distribution of synesthetic colors in the color space.

Method

We collected a large number of synesthetic colors for Japanese kanji, hiragana, katakana, Latin letters, and Arabic numerals from each of eight synesthetes. Five of the synesthetes had colors associated with all kinds of characters (kanji–color synesthetes), and others had synesthetic colors associated only with hiragana, Latin letters, and Arabic numerals (hiragana–color synesthetes). The five kanji–color synesthetes participated in two experiments. The first employed a color selection task to determine the synesthetic colors that they experienced when stimulated with printed graphemes comprising kanji, hiragana, katakana, Latin letters, and Arabic numerals. The second experiment employed a color-matching task to determine their synesthetic color associations for Latin letters and Arabic numerals displayed on a CRT monitor. The three hiragana–color synesthetes participated only in the color-matching task.

Participants

Eight Japanese synesthetes (five females [S.H., A.H., M.H., A.M., and Y.H.] and three males [S.O., H.S., and Y.K.]; age range = 18–21 years) received payment (dependent on the experiment time) for their participation in the experiments. S.H. and S.O. were self-reported projector-type synesthetes, and the others were self-reported associator-type synesthetes. Their scores on the color match consistency test were below 1.0, indicating a consistent and authentic synesthetic experience (Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007).

Color selection experiment

Visual stimuli The test stimuli comprised individual black graphemes printed centrally on a white card, called a "character card" (127 × 89 mm). The character cards, used in elementary education, included 482 kanji, 46 hiragana (あ–ん), 46 katakana (ア–ン), 26 Latin letters (A–Z), and ten Arabic numerals (0–9). Each kanji, hiragana, and katakana character was printed in 42-point HG textbook font (about 15 × 15 mm), and each Latin letter and Arabic numeral was printed in 72-point Arial font (about 25 × 33 mm). Character cards were placed on a desk covered by a black cloth and illuminated by a D65 daylight fluorescent lamp, which has a high color-rendering index of 98 (FL20S D-EDL-D65; Toshiba Inc., Tokyo, Japan). The viewing distance was approximately 40 cm.

Procedure The participant was asked to view a character card and select the color(s) that best matched their perceived synesthetic color by using color chips from the *Munsell Book of Color, Matte Finish Collection* (Munsell Color, 1976). More specifically, in each trial they picked up a character card, placed it on the center of the desk, observed the character, and selected one of the five basic colors (5R, 5G, 5B, 5Y, 5P) that most resembled their synesthetic color. Next, they examined all four sheets for the selected basic color from the *Munsell Book* and selected the color chip that best matched the synesthetic color. Finally, they wrote the color chip's Munsell code in a blank space on the character card. If they perceived multiple synesthetic colors for a character, the procedure was repeated, and the part of the character for which they perceived each synesthetic color was marked by a box. Trials were blocked for kanji, hiragana, and katakana (40 blocks of 12 or 13 cards for kanji; four blocks of 12 or 10 cards for hiragana or katakana, respectively), and the order of the trials was quasi-randomized.

Two of the participants, S.H. and A.H., performed a longer-term experiment (480 kanji characters) than did the others (240 kanji characters). Since the experiment was expected to be very time-consuming, we decided to allow S.H. and A.H. to perform the experiment in their home in their free time, after

practice sessions that were performed in the laboratory. Care was taken to ensure that the experiment was performed in a constant visual environment. The other six synesthetes performed the experiment in our laboratory. All participants were instructed to view the cards under the high color-rendering lamp, to do the experiment regularly, and to keep an experimental diary. The diary confirmed that it took a long time to complete the experiment. Of the synesthetes who had kanji–color associations, S.H. spent 16 h 45 min, 1 h 4 min, and 54 min for the kanji, hiragana, and katakana sets, respectively, over a total of 17 days. A.H. spent 53 h 30 min and 2 h 4 min for the kanji and hiragana sets, respectively, over a total of 23 days. Because of the especially time-consuming nature of the experiment for this participant, the experimenter did not employ the katakana color selection task with A.H. S.O. spent 14 h 41 min, 54 min, and 40 min for the kanji, hiragana, and katakana sets, respectively, over a total of 7 days. M.H. spent 10 h 11 min, 49 min, and 57 min for the kanji, hiragana, and katakana sets, respectively, over a total of 7 days. H.S. spent 9 h 25 min, 1 h 4 min, and 55 min for the kanji, hiragana, and katakana sets, respectively, over a total of 5 days. The experimenters checked for mistakes in recording of the Munsell codes, and found a total of 13 typos of the Munsell codes (two typos for S.H.; six typos for A.H.; one typo for S.O.; two typos for M.H.; and two typos for H.S.) and excluded these selections.

Color matching using a CRT monitor

Visual stimuli For the 26 Latin letters and ten Arabic numerals, the color coordinates (CIE L*a*b*) of each synesthetic color were determined by a color-matching task using a CRT monitor (1280 × 1024 pixels, 100 Hz, NANAO, FlexScan F980). The test character, displayed in Arial font (3° of visual angle), was located on the left side (3° from center) of the screen, and a reference color patch was located on the right side (3° from center). The test character was black, and the background color was gray (CIE $x = .284$, $y = .341$, $Y = 30.0$). The monitor was calibrated with a chromameter (CS-100A; Konica Minolta, Tokyo, Japan) using the Mcalibrator2 software (Ban & Yamamoto, 2013).

Procedure In the laboratory setting, the participant was asked to adjust the color of the displayed reference patch so that it best matched their perceived synesthetic color for the test character. More specifically, after 3 min of dark adaptation followed by 2 min of light adaptation, participants started the color-matching trials. First, the character and the reference patch were displayed on the CRT monitor. The initial color of the reference patch was not shown by cordinates (L*a*b*) determined in the color selection experiment. Based on the cordinates determined in the color selection task, the initial color changed form –5, –5, –10 to +5, +5, +10 randomly for each trials. Next, the participant adjusted the color patch with

a digital input device until he or she felt that it matched the perceived synesthetic color. The input device allowed the participant to alter the L*a*b* color coordinates of the color patch independently in steps of ±0.3, ±1, and ±1. The trial order was randomized.

Analysis

We used the CIE L*a*b* color space to analyze the distribution of synesthetic colors. The CIE L*a*b* coordinates specify colors using values along three axes: light and dark (L*), red and green (a*), and blue and yellow (b*). An L*a*b* color space is a uniform space in which a given spatial distance corresponds to an equivalent perceptual difference; that is, one unit corresponds approximately to one just-noticeable difference. This uniformity would be suitable for interpreting the distribution of synesthetic colors when analyzing them on the basis of a distance metric, as we describe below.

The distribution of synesthetic colors was investigated using spatial statistics analysis, performed with the spatstat (Baddeley, 2010) software, which is an R package for analyzing a spatial point pattern. Such point patterns fall into three general categories: random, clustered, and uniform (Fig. 1 top) (Bailey & Gatrell, 1995). In a *random* pattern, any point is equally likely to occur at any location, and points are not affected by other points. In a *clustered* pattern, many points are concentrated in some regions and are sparse in others. Finally, in a *uniform* pattern, every point is as far as possible from all nearby points. We considered the distribution of synesthetic colors in the a*b* chromaticity plane to be a 2-D point pattern, and then explored which distribution patterns the synesthetic colors exhibited.

More specifically, our exploratory analysis had three steps. First, the synesthetic color's point pattern was visualized in the a*b* chromaticity plane. Second, the density of the point pattern was estimated by kernel estimation using the *density()* function. Kernel estimation calculates the density of points within a specified search radius ($\sigma = 8.46$) around each point. Third, the type of the synesthetic color's point pattern was judged by the L function, computed from Ripley's K function, which is defined as

$$K(r) = E(r)/\lambda. \qquad (1)$$

In the K function, which is a function of search radius r, $E(r)$ is the expected number of points within radius r of an arbitrary point. λ is the density of points (synesthetic colors per unit region), estimated by $\lambda = n/a$, with n points (synesthetic colors) in a given region a. If a point pattern is random, the K function becomes

$$K(r) = \pi r^2. \qquad (2)$$

Fig. 1 Examples of types of point patterns, classified with spatial statistics. (a) When objects (points) comprise a single type, the possible point patterns are random, clustered, or uniform. (b) When objects comprise two types, the possible point patterns are independent, attractive, or repulsive
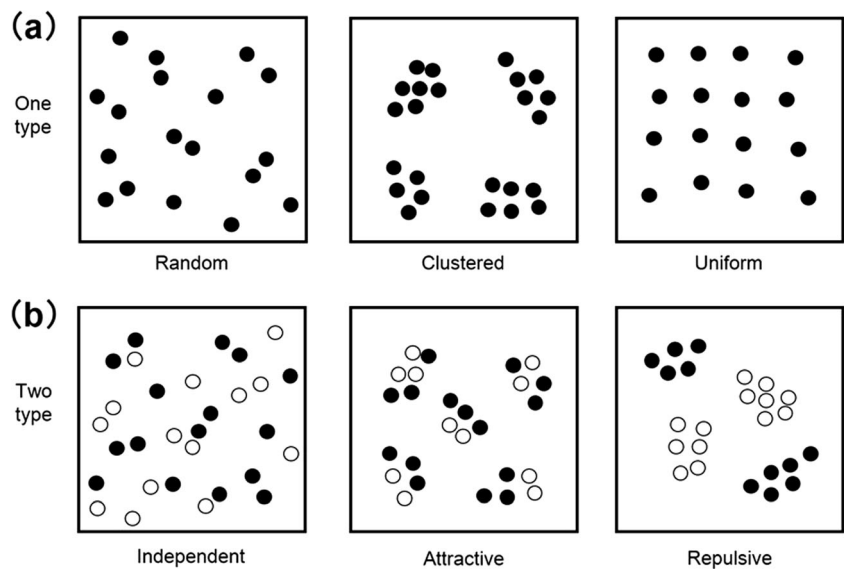
The L function can be computed from the K function as

$$L(r) = \sqrt{K(r)/\pi} - r. \tag{3}$$

Using this L function, we could classify the synesthetic color's point pattern as being one of the three types: random when $L(r) = 0$, clustered when $L(r) > 0$, or uniform when $L(r) < 0$, as is shown in Fig. 2a.

For further analysis, we referred to an extension of the L function, called the *cross-L function*, which can evaluate the dependency of two point patterns. This function is helpful for investigating the interrelation between the synesthetic colors stimulated by different groups of graphemes, such as by grapheme types (e.g., kanji vs. hiragana) or by the learning grade, and therefore complexity, of kanji (e.g., the simple 一 and 二 vs. the more complex 雨 and 雲). The dependency of point pattern pairs was divided into three categories (Bailey & Gatrell, 1995): independent, attractive, or repulsive, shown in Fig. 1b. An *independent* pair would show no dependency between the two patterns, an *attractive* one would have two patterns that bunched together at specific places, and a *repulsive* one would have two mutually exclusive patterns. These types of synesthetic color point patterns were determined by the cross-L function, which is computed from Ripley's cross-K function, defined as

$$K_{AB}(r) = E(r) / \lambda_B. \tag{4}$$

Here, $E(r)$ is the expected number of members of point pattern B within a given radius r of any arbitrary member of the point pattern A. λ_B is the density of the point pattern B, which can be estimated as n_B/a, with a total of n_B members of the point pattern B in a given region a. The cross-L function can be computed from the cross-K function as

$$L_{AB}(r) = \sqrt{K_{AB}(r)/\pi} - r. \tag{5}$$

Using this cross-L function, we could classify the synesthetic color point patterns as one of the three types: independent when $L_{AB}(r) = 0$, attractive when $L_{AB}(r) > 0$, or repulsive when $L_{AB}(r) < 0$, as is shown Fig. 2b.

Results

Synesthetic colors associated with kanji characters

We investigated the synesthetic colors associated with 482 kanji characters for S.H. and A.H., and 240 kanji characters for S.O., M.H., and H.S. We obtained a total of 4874 colors from five synesthetes, as specified using the Munsell color system. Additionally, about 80 synesthetic colors from each of three synesthetes were associated with hiragana characters. Notably, the number of synesthetic colors was larger than the number of kanji characters investigated. On average, a kanji character was associated with 2.26 colors for S.H., 4.61 colors for A.H., 1.92 colors for S.O., 1.55 colors for M.H., and 0.80 colors for H.S. Below, the thousands of synesthetic colors for kanji characters were analyzed.

The first question we asked was "How is a kanji character associated with multiple colors?" There seem to be at least four ways for S.H. The first was that synesthetic colors depended on the component radicals rather than on the character as a whole, and kanji characters are often composed of multiple radicals. For example, the character 決 was perceived in two parts, the aforementioned *sanzui* radical 氵, perceived as yellow, and the remaining 夬 component, perceived as red (Fig. 3a). This radical dependency was also seen in
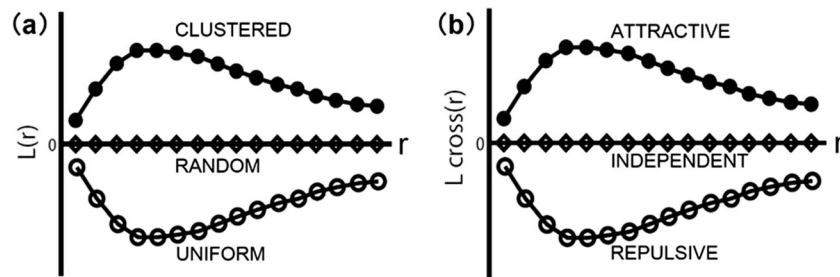
Fig. 2 Example plots of expected *L* functions and cross-*L* functions for each of the three kinds of point patterns in Fig. 1a and b. For the definition of the *L* function, see the main text. (a) Examples when objects comprise one type. If the *L* function is a constant 0, it indicates a random pattern; if it is positive, a clustered pattern; and if it is negative, a uniform pattern. (b)

Examples when objects comprise two types. If the *L* function is a constant 0, it indicates an independent pattern; if it has a positive value, this indicates an attractive pattern; and if it has a negative value, it indicates a repulsive pattern

synesthetes S.O., M.H., and H.S. The second way was that a synesthetic color could depend on pronunciation, and kanji characters often have multiple readings. For example, the character 千 was perceived as yellow by S.H. when read as its *on-yomi* (/sen/), but it was perceived as red when read as its *kun-yomi* (/chi/) (Fig. 3b). The third way was that a kanji character as a whole sometimes had an overall color that differed from the local colors usually associated with its radicals. For example, the character 聞 is composed of two radicals (門 and 耳), which were perceived to have their own local colors (yellow-red and red-purple, respectively), but the character as a whole was perceived to be green (Fig. 3c). The fourth way was that a kanji was sometimes simply perceived to be two different colors. For example, the upper radical (ハ) of the 谷 character was sometimes perceived as purple-blue and sometimes as yellow (Fig. 3d).

For synesthete A.H., most synesthetic colors depended on finer component parts of the kanji characters, beyond the radicals. For example, the character 決 was divided into the radicals 氵 and 夬 by S.H., but A.H. further divided the夬 into the ユ and the remaining 人. The *sanzui* 氵 was perceived as red, the ユ component as a different red, and the 人 component as yellow (Fig. 4a). The character "識" was associated with the largest number of synesthetic colors (12 colors), and these synesthetic colors depended on detailed components of the kanji character (Fig. 4b).

We also investigated 46 hiragana characters, 46 katakana characters, 26 Latin letters, and ten Arabic numerals for S.H., S.O., M.H., and H.S., and all of these characters except the katakana for A.H. Some of these characters were associated with two or more colors for S.H., A.H., and M.H., and thus there were a slightly greater number of synesthetic colors than characters (S.H.: 144 colors for 128 graphemes; A.H.: 192 colors for 82 graphemes; S.O.: 122 colors for 128 graphemes; M.H.: 132 colors for 128 graphemes; H.S.: 83 colors for 128 graphemes). When these were added to the results from the kanji, a total of 1,229 (S.H.), 2,349 (A.H.), 551 (S.O.), 460 (M.H.), and 285 (H.S.) synesthetic colors were collected.

Figure 5 shows all of these colors in the L*a*b* color space for each of the five kanji–color synesthetes. The synesthetic colors seemed to be widely distributed across the three axes. To take a closer look, we projected these distributions in three 2-D planes: a* versus b* (Fig. 5b, f, j, n, and r), L* versus a* (Fig. 5c, g, k, o, and s), and L* versus b* (Fig. 5d, h, l, p, and t). Figure 6 shows all of synesthetic colors for each of the three hiragana–color synesthetes, in a similar fashion to Fig. 5. For both types of synesthetes, the point patterns in the chromatic plane seemed to show some regions where the points were concentrated and others where there were few points (a* vs. b*), suggesting that synesthetic colors were neither randomly nor uniformly distributed in the color space. This type of bias was also present for the luminance dimension. Focusing on the point pattern along the luminance axis for kanji–color synesthetes (L* vs. a* and L* vs. b*), we see a small number of dark colors, with L* less than about 30 (Fig. 5c, d, g, h, k, l, o, p, s, and t). For hiragana–color synesthete A.M., few dark colors can be seen with L* less than about 30 (Fig. 6c and d). In contrast, For Y.K. and Y.H., more dark colors can be seen with L* less than about 30 (Fig. 6g, h, k, and l). All together, it appears that the synesthetic colors are distributed more heavily in some regions in the color space. This tendency is investigated at greater length using spatial statistics in the Example of Applications section below.

We compiled these diverse synesthetic colors and the associations between kanji characters and colors into a database that could be helpful for the study of grapheme–color synesthesia. In the following sections, we describe the contents of the database and introduce new research directions that this database opens up.

Database contents

Figure 7 shows a section of the new database. The database denotes not just the synesthetic colors associated with each kanji character, but also gives additional information for each character, such as the radicals and pronunciations that affected

Fig. 3 Representative results of the color selection experiment from projector-type synesthete S.H. Each figure is a scanned character card. The synesthete wrote the Munsell codes and supplemental comments in the blank spaces on the character cards. (a) An instance in which synesthetic colors varied by radical. The full 決 character was divided into the 冫 radical and the remaining 夬 component, each of which had a distinct color. (b) An instance in which synesthetic colors varied by sound. She described that when the 千 character was read with its *on-yomi* pronunciation, "sen," the color was 7.5Y-8-4, and when it was ready with its *kun-yomi* pronunciation, "chi," the color was 2.5R-5-8. (c) An instance in which the global color differed from the local colors. She added to her notation of the individual colors that "At a glance, I feel also that the whole character is 2.5G-6-12." (d) An instance in which the synesthetic color switched for the same part. She described that the upper part (ハ) of the 谷 was sometimes 5PB-8-4, but sometimes 7.5Y-9-10

Fig. 4 Representative results of the color selection experiment from associator-type synesthete A.H., shown in a fashion similar to Fig. 3. Both panels (a) and (b) show instances in which the synesthetic colors varied in terms of finer components, beyond the radicals. (a) The 決 character was divided into the 冫 radical and the 夬 component, and even the 夬 was divided into the ユ and the remaining 人, each of which had a distinct color. (b) The 識 character was associated with the most synesthetic colors (12), which depended on highly detailed components of the kanji character

the synesthetic colors, as we have described above (see also Table 1). This auxiliary information will potentially be useful for analyzing the first- and second-order relations of the synesthetic associations. The database was implemented as a Microsoft Excel file that comprised multiple sheets, including the kanji, hiragana, katakana, English graphemes (letters and numerals), and search sheets. The details of the contents of each sheet will be presented below.

Kanji sheet The kanji sheet has each kanji character in a row and the various auxiliary information associated with each character in columns, hereafter called "classes." The classes comprise three types: color class, elemental class, and

supplemental class. The color class includes colors, defined using three types of chromaticity coordinates—the Munsell color system, CIE xyY, and CIE L*a*b* color coordinates—in conjunction with a visualization of the color by coloring the cell background.

The elemental class is the category of the various components in the kanji character. This had subclasses of "radical," "constituent" "*on-yomi*," "*kun-yomi*," and "grade." "Radical" describes the common subelements found in different kanji characters. The "constituent" subclass shows the component parts of the kanji character, allowing researchers to investigate the influences of all component shapes, not limited solely to radicals. The "*on-yomi*" and "*kun-yomi*" subclasses list the various readings of the kanji character, as we previously described. "Grade" shows the year in which the given kanji character is taught in Japanese elementary school (first through sixth grade), as stated by the Ministry of Education, Culture, Sports, Science and Technology of Japan's appendix, "List of Kanji by School Year" in the *Course of Study for Elementary School*. Since the school year in which kanji characters are taught is standardized by the Ministry, this item will be potentially useful for research on the longitudinal tendencies of synesthetic colors.

The supplemental class is a category containing complementary information about the kanji characters or synesthetic colors. This is divided into five subclasses: "strokes," "number of colors," "color appearance," "subjective report," and "overall colors." "Strokes" is the number of strokes in the kanji character. "Number of colors" shows the number of synesthetic colors associated with the character. "Color appearance" categorizes the various ways that the synesthete

Fig. 5 Distribution of synesthetic colors collected from kanji–color synesthetes, in CIE L*a*b* uniform color space. The synesthetic colors for kanji, hiragana, and katakana were transformed from the Munsell values to CIE L*a*b* values. (a) Synesthetic colors of observer S.H. in L*a*b* color space. (b) Those colors projected onto the a* versus b*

chromaticity plane. (c) Those projected onto the L* versus a* plane. (d) Those projected onto the L* versus b* plane. Panels (e)–(t) show the color data for observers A.H., S.O., M.H., and H.S. in similar fashion to panels (a), (b), (c), and (d)

perceived the colors, divided into nine categories: (A) colors were divided into each part ("component") of the kanji; (B) colors switched within the same part; (C) colors were divided into each part, but there was a different overall color when the character was seen as a whole; (D) there was an overall color for the character as a whole, but there was also a local color;

(E) the color changed depending on the reading or meaning; (F) colors were divided into each part, but colors also switched in some parts; (G) indicates a synesthete can not show positions of synesthetic colors but can perceived colors somewhere; (H) there was a part with no color; and (I) there were colors in the mind's eye. Finally, "subjective report" contains
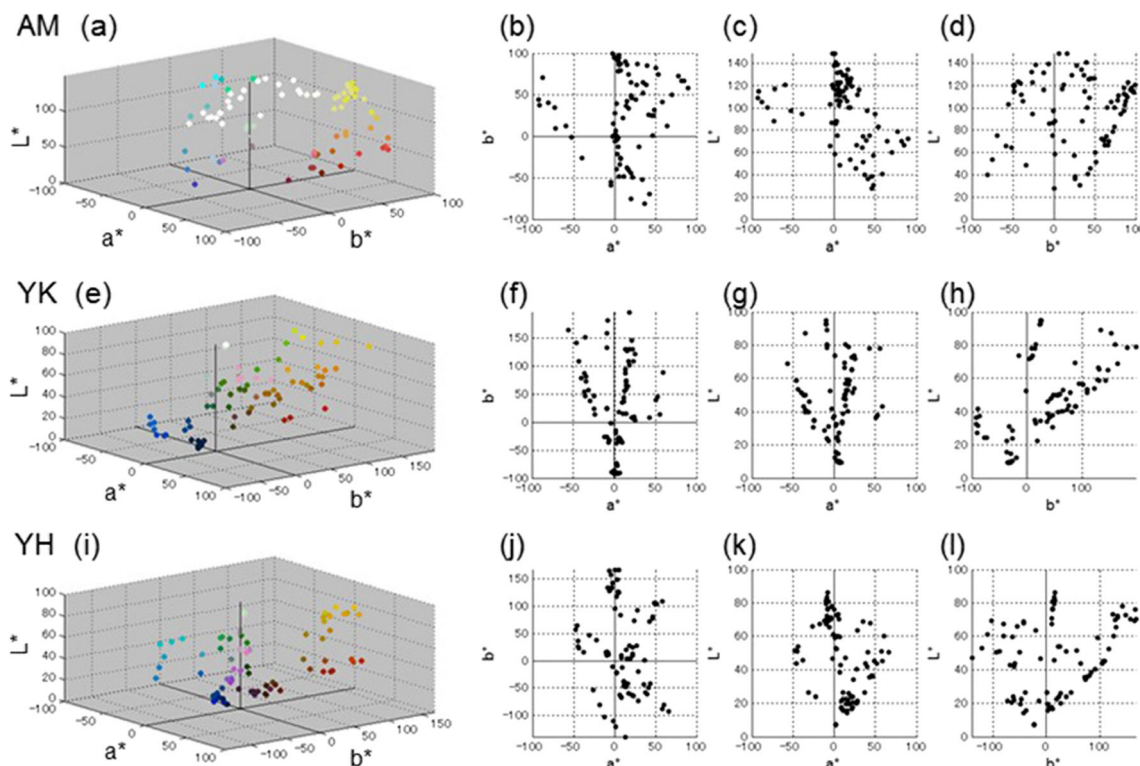
Fig. 6 Distribution of synesthetic colors collected from hiragana–color synesthetes, in CIE L*a*b* uniform color space, shown in similar fashion to Fig. 5. (a) Synesthetic colors of observer A.M. in L*a*b* color space. (b) Those colors projected onto the a* versus b* chromaticity plane. (c) Those projected onto the L* versus a* plane. (d) Those projected onto the L* versus b* plane. Panels (e)–(l) show the color data for observers Y.K. and Y.H. in a similar fashion

additional information on the synesthete's subjective thoughts about the color appearance. "Overall colors" was the overall synesthetic color of a kanji character.

Hiragana, katakana, and English grapheme sheets These sheets had each character in a row and select classes or subclasses associated with the character in the columns. The hiragana and katakana sheets had the color class and two supplemental subclasses (only "number of colors" and "subjective report"). The English graphemes sheet (with letters and numerals) only had the color class, determined as described in the Method section, containing the CIE xyY and L*a*b* coordinates measured in the CRT monitor color-matching task.

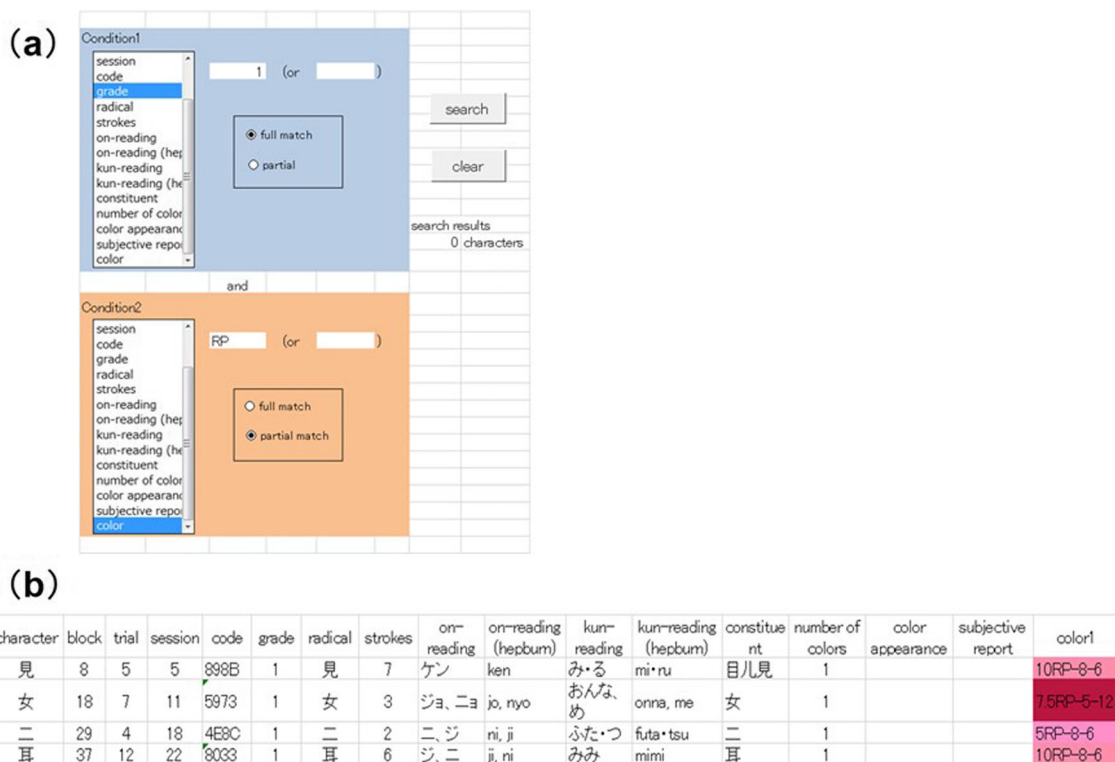| synesthete	character	grade	radical	strokes	on-yomi	on-yomi (hepburn)	kun-yomi	kun-yomi (hepburn)	constituent	number of colors	color appearance	subjective report	overall color	color1	color2	color3
SH	男	1	田	7	ダン・ナン	dan・nan	おとこ	otoko	田力男	2	A			5PB-4-12	5PB-5-12	
SH	青	1	青	8	ショウ、セイ	sho, sei	あお、あお・い	ao, ao・i	月青	2	*D	Overall color is the blue. The part of "月" is 7.5Y-8.5-12.	5PB-4-12	5PB-4-12	7.5Y-8.5-12	
SH	年	1	干	6	ネン	nen	とし、とせ	toshi, tose	ノ一十千年	1				5Y-9-8		
SH	先	1	儿	6	セン	sen	さき	saki	土儿先	1				5Y-8.5-10		
SH	白	1	白	5	ハク、ビャク	haku, byaku	しろ、しら、しろ・い	shiro, shira, shiro・i	白	1				N-9-0.25		
SH	子	1	子	3	シ	shi	こ、ね	ko, ne	子	1				7.5GY-8.5-2		
SH	糸	1	糸	6	シ	shi	いと	ito	幺小糸	2	A			N-8-0.75	N-7-0.25	
SH	字	1	子	6	ジ	ji	あざ	aza	宀子字	1				N-2-0.5		
SH	王	1	王	4	オウ	o	きみ、おおき み	kimi, okimi	王	1				2.5YR-6-14		
SH	三	1	一	3	サン	san	み、みっ・つ	mi, mittsu	一二三	1				2.5PB-8-6		
SH	入	1	入	2	ニュウ、ジュ、ジュウ	nyu, ju, ju	い・れる、はい・る	i・reru, hai・ru	入	1				7.5Y-9-6		
SH	手	1	手	4	シュ	shu	て、た	te, ta	手	1				5PB-5-10		
SH	水	1	水	4	スイ	sui	みず、みな	mizu, mina	水	1				7.5B-8-4		
SH	正	1	止	5	セイ、ショウ	sei, sho	ただ・しい、まさ・に	tada・shii, masa・ni	止正	1				2.5Y-9-2		
SH	犬	1	犬	4	ケン	ken	いぬ	inu	大、犬	2	A			5PB-7-8	7.5Y-9-10	
SH	見	1	見	7	ケン	ken	み・る	mi・ru	目儿見	1				10RP-8-6		
SH	休	1	人	6	キュウ	kyu	やす・む	yasu・mu	イ木休	2	A			2.5GY-8.5-1	7.5GY-6-10	
SH	気	1	气	6	キ、ケ	ki, ke			气メ気	1				10GY-6-10		
SH	六	1	八	4	ロク、リク	roku, riku	む・つ	mu・tsu	一八ハ六	1				2.5YR-6-14		
SH	木	1	木	4	ボク、モク	boku, moku	き	ki	木	1				2.5G-6-10		

Fig. 7 Screenshot of the kanji sheet in the K-SCD

Table 1 Synesthetic color information and auxiliary information for characters in the database

Class	Subclass	Explanation of Subclass
Color		Munsell color system, CIE xyY and CIE $L^*a^*b^*$
Elemental	Grade	The year when a given kanji character is taught in Japanese primary schools
	Radical	Orthographic components of kanji characters
	On-reading	Chinese reading of kanji
	Kun-reading	Japanese reading of kanji
	Constituents	Separable parts of kanji character
Supplemental	Strokes	The number of kanji strokes
	Number of colors	The number of colors in a character
	Color of appearance	Patterns of way of perceiving synesthetic colors
	Subjective report	Additional information on the synesthete's introspection about the "Color appearance"
	Overall color	The overall synesthetic color of a kanji character

The color class includes chromaticity coordinates for Japanese kanji, hiragana, and katakana, Latin letters, and Arabic numerals. The elemental class contains subclasses relating to component values of the kanji characters. The supplemental class has subclasses relating to complementary information about the kanji characters or the perceived synesthetic colors

Search sheet The search sheet can take arbitrary data from the kanji sheet. Figure 8 shows a screenshot of the search sheet, which has two functions for searching the database. The first is an OR search, in which a user can assign a maximum of two factors per subclass. The second is an AND search, for which a user can assign a maximum of two conditions. For these searches, the terms may be fully or partially matched. If a user would, for example, like to search for data on kanji that are first-grade level and were associated with red-purple synesthetic colors, as is shown in Fig. 8a, the user would input "1" for "grade," with "full match" selected, as Condition 1, and input "RP" for color, with "partial match" selected, as Condition 2. Figure 8b shows the result of this example search, containing all returned entries fulfilling the condition.
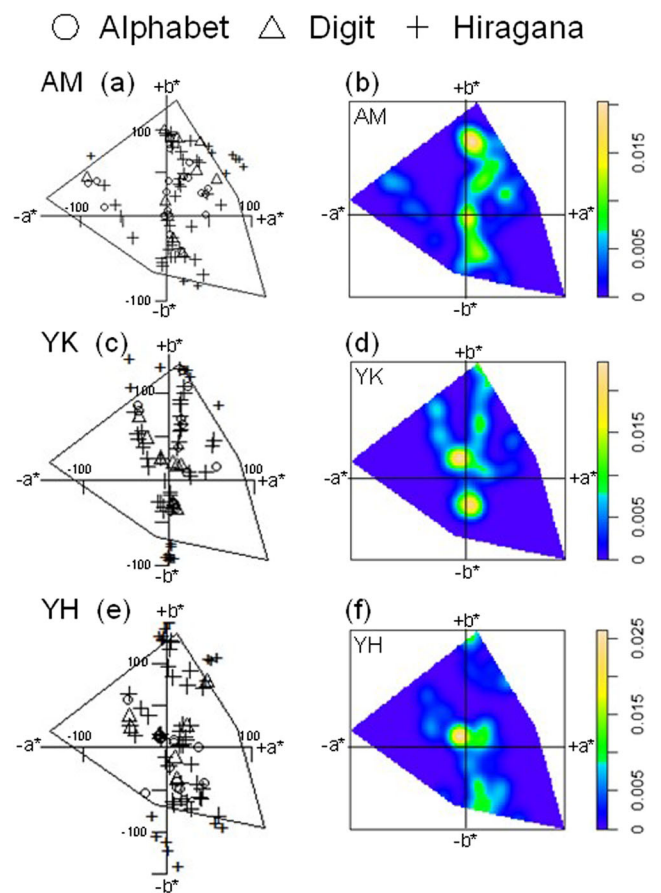

character	block	trial	session	code	grade	radical	strokes	on-reading	on-reading (hepburn)	kun-reading	kun-reading (hepburn)	constituent	number of colors	color appearance	subjective report	color1
見	8	5	5	898B	1	見	7	ケン	ken	み・る	mi・ru	目儿見	1			10RP-8-6
女	18	7	11	5973	1	女	3	ジョ、ニョ	jo, nyo	おんな、め	onna, me	女	1			7.5RP-5-12
二	29	4	18	4E8C	1	二	2	二、ジ	ni, ji	ふた・つ	futa・tsu	二	1			5RP-8-6
耳	37	12	22	8033	1	耳	6	ジ、二	ji, ni	みみ	mimi	耳	1			10RP-8-6

Fig. 8 Screenshots of the search function of the K-SCD, showing (a) a sample of the search fields and (b) the search results

○ Alphabet △ Digit + Hiragana
◇ Katakana × Kanji
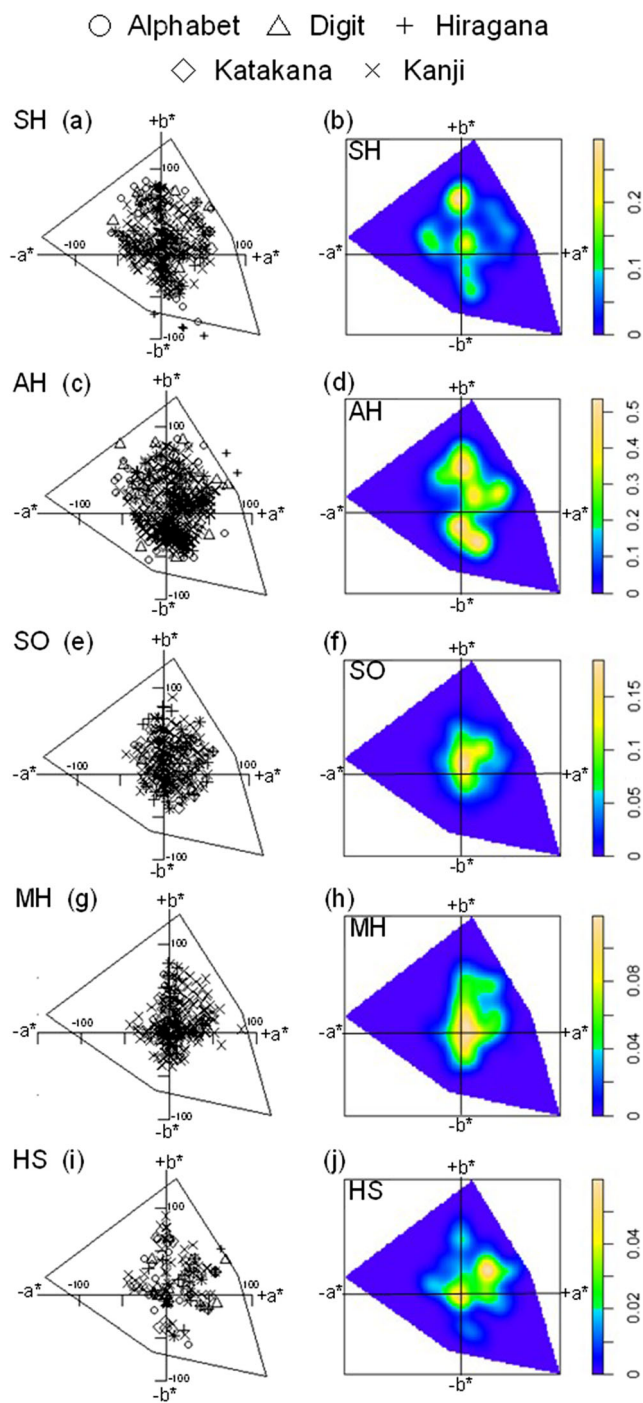
Fig. 9 (a, c, e, g, i) Point patterns of synesthetic colors in the a*b* chromaticity plane for kanji–color synesthetes S.H., A.H., S.O., M.H., and H.S., respectively. (b, d, f, h, j) Density plots of the synesthetic colors, as estimated by kernel estimation for each synesthete. Achromatic synesthetic colors were excluded, and then the density was estimated as before. The color bars denote the density λ of points within a specified search distance ($\sigma = 8.46$) of each point

○ Alphabet △ Digit + Hiragana

Fig. 10 (a, c, e) Point patterns of synesthetic colors in the a*b* chromaticity plane for hiragana–color synesthetes A.M., Y.K., and Y.H. respectively, shown in a similar fashion to Fig. 9

useful for exploring regularity in the associations between kanji characters and synesthetic colors. Here, we illustrate this with an example in which the first- and second-order relationships of the association were examined, focusing on S.H., who had four types of grapheme–color associations.

At first we focused on the first-order relation; namely, which radicals were most frequently associated with what colors? To answer this, we determined the most frequent color among the five basic Munsell colors. In the database search sheet, we assigned each basic color ("R," "Y," "G," "B," and "P") as the "color," with "partial match" selected, as Condition 1. The most frequent color was yellow: 332 of 482 kanji characters had a yellowish color. We then explored the relationship between these yellowish colors and various radicals, to check for common radicals or constituents in the 332 detected characters. We found that yellowish colors were associated with the *sanzui* 氵, *tsukanmuri* ツ, and *ninben* 亻 radicals, in spite of the fact that *sanzui* means water. We then considered a possible second-order relationship: Do characters with shapes similar to these radicals have similar colors? To answer this, we examined three katakana characters with

Example of applications

Exploring the relationship between the multidimensional natures of kanji and synesthetic colors Our database was

Table 2 High-density and low-density regions of synesthetic colors in the a*b* chromaticity

Synesthete	Number of Chromatic Colors	High-Density		Low-Density	
		Region	[a*, b*]	Region	[a*, b*]
Kanji–Color Synesthetes					
S.H.	1018	origin point	[0, 15]	red	[55, 25]
		yellow	[5, −75]		
		blue	[10, −45]		
		green	[−40, 15]		
A.H.	2204	yellow-red	[15, 20]		
		yellow	[0, 65]		
		red	[45, 20]		
		blue	[0, −20]		
		blue-purple	[20, −25]		
S.O.	473	origin point	[0, 10]		
M.H.	399	origin point	[5, 0]		
H.S.	210	origin point	[0, 0]	yellow	[0, 70]
		red	[55, 25]	blue	[5, −50]
Hiragana–Color Synesthetes					
A.M.	80	origin point	[0, 0]		
		yellow	[0, 100]		
		blue	[10, −50]		
Y.K.	84	yellow-green	[−10, 20]		
		blue	[0, −30]		
Y.H.	84	origin point	[−5, 5]		

For hiragana–color synesthetes, only high-density region are described
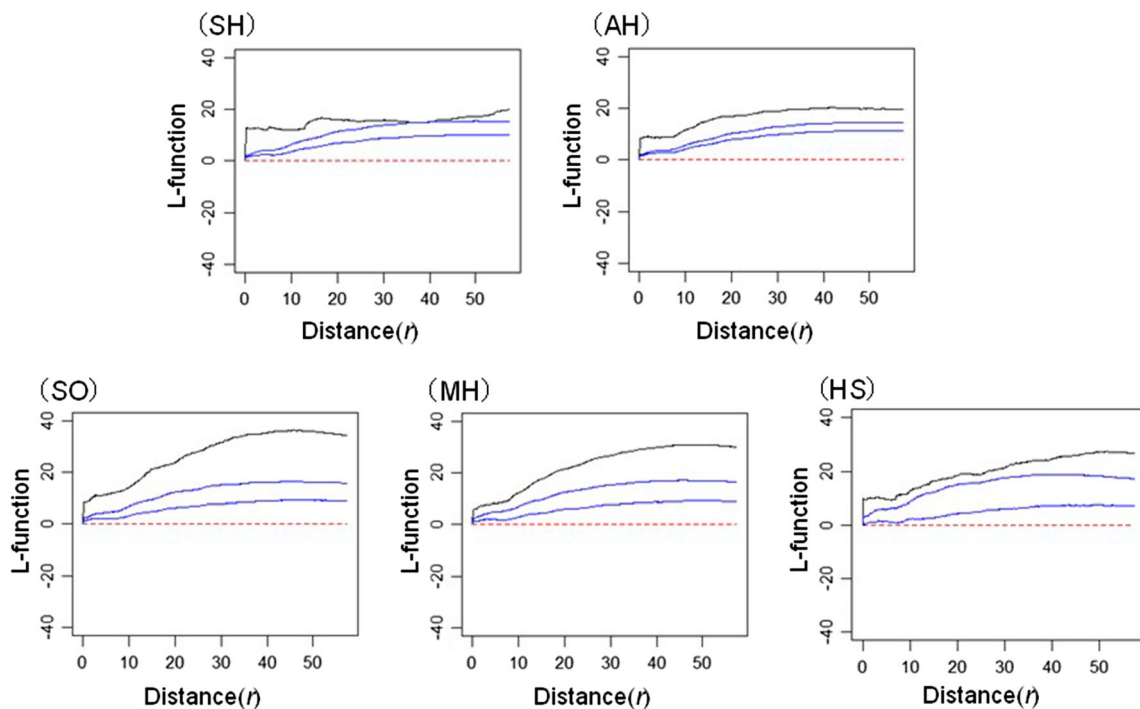
Fig. 11 *L* functions of the point patterns for chromatic synesthetic colors at a distance *r* from a given point for each of the kanji–color synesthetes. The solid black lines indicate the computed *L* function, and the interval between the two blue lines in each panel indicates the 95 % confidence interval for complete spatial randomness (CSR). Specifically, many spatial point patterns (1,000) randomly selected from the Munsell renotation data can be generated under CSR, and Ripley's *K* function can be estimated for each one
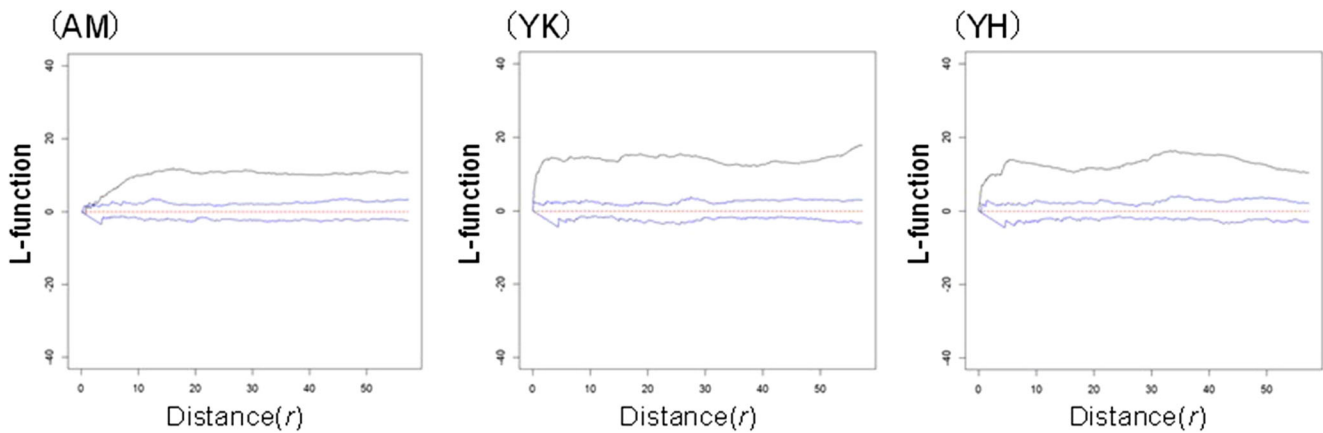
Fig. 12 *L* functions of the point patterns for synesthetic colors at a distance *r* from a given point for hiragana–color synesthetes, shown in similar fashion to Fig. 11. The interval between the two blue lines in each panel indicates the 95 % confidence interval for complete spatial randomness (CSR). Specifically, many spatial point patterns can be generated under CSR, and Ripley's *K* function can be estimated for each one

similar shapes: シ, ツ, and イ. Interestingly, these katakana characters also had a yellowish color. Taken together, for the most frequent color (yellowish), radicals not with similar meanings but similar shapes had similar colors.

Analysis of the global structure of synesthetic colors in the chromaticity space Here we wanted to provide an answer for the question presented in the introduction: Are synesthetic colors distributed randomly in the color space? As we mentioned above, the distribution of synesthetic colors suggested the possibility that the point patterns were not random or uniform, but rather were clustered (see Fig. 5b, f, j, n, and r). To confirm this possibility statistically, we analyzed the point patterns of synesthetic colors using several techniques of spatial statics, which had three steps, noted in the Method section. At first, we projected the colors to the a*b* chromaticity plane, visualizing the point patterns of synesthetic colors (Fig. 9a, c, e, g, and i). Second, we estimated the density of each point pattern by kernel estimation (Fig. 9b, d, f, h, and j). To clarify the distributions of synesthetic colors for kanji–color synesthetes, the density of chromatic colors was also estimated after excluding the achromatic colors (S.H., 211 samples; A.H., 145 samples; S.O., 78 samples; M.H., 61 samples; H.S., 75 samples). As is clearly shown in Fig. 9b, d, f, h, and j, we found clusters of chromatic synesthetic colors for the kanji–color synesthetes. For the hiragana–color synesthetes, as is clearly shown in Fig. 10b, d, and f, there were also many clusters of synesthetic color. The high-density and low-density regions for each synesthete are summarized in Table 2. For two kanji–color synesthetes, S.H. and A.H., over 1,000 synesthetic colors were concentrated in multiple regions in the a*b* chromaticity plane. For the other three kanji–color synesthetes, S.O., M.H., and H.S., over 100 synesthetic colors were concentrated around an origin point.

Finally, to confirm these synesthetic color clusters, we analyzed the types of point patterns using the *L* function (Fig. 11 for kanji–color synesthetes; Fig. 12 for hiragana–color synesthetes). As is shown in Figs. 11 and 12, the *L* functions returned positive values greater than the confidence intervals. This suggests that synesthetic colors are concentrated in multiple regions in the color space; that is, they form "synesthetic color clusters."

Explanations of the script for spatial statistics of synesthetic colors

Since the database was made in the widely used Excel format, it can be easily imported to many kinds of statistical software. For example, the spatial statistics analysis above was done using R, as is shown in Fig. 13. The details of the analysis steps are explained below, to help researchers understand and modify the script for their own use.

> Step 1: Prepare a spatstat package and input the row data into R. Then, the user can input a portion of the database as a text file—for example, the L*a*b* coordinates of synesthetic colors.
> Step 2: Create a point pattern object using a class *ppp* object. The *ppp*(x, y, window = w, marks = mf) function will create the point pattern inside a rectangle. x (a*) and y (b*) are vectors of each of the coordinates of data points. The input *window = w* specifies a range of the polygon. Each coordinate value of the polygon was defined by the most achromatic colors (red: [a*, b*] = [82.28, 22.42]; green: [−139.00, 20.36]; yellow: [11.91, 134.92]; blue: [−14.69, −67.01]; purple: [116.06, −94.45]) from the Munsell renotation data (www.rit.edu/cos/colorscience/rc_munsell_renotation.php). Next, *marks* is defined, which is an additional attribute of each point in the

```
# ScriptSpatStat.R
rm(list=ls()) # remove all the variables from the workspace
print_on <- 1 # print on/off
library(spatstat)
#----- input
cspace1 <- "L" # axis of luminance
cspace2 <- "a" # axis of red and green
cspace3 <- "b" # axis of blue and yellow
subject <- "Data" # name of subject
cspace <- sprintf("%s%s%s", cspace1, cspace2, cspace3) # "Lab"
fn <- sprintf("%s_%s%s", subject, cspace, ".txt") # name of data file
# scolors3D: standard data frame for 3D color coordinates
scolors3D <- read.table(fn,   header=FALSE, sep=",") # read "data_Lab.txt"
names(scolors3D) <- c(cspace1, cspace2, cspace3) # "L" "a" "b"
```

Step1

Load required packages and reading row data into R

```
# SpatStat's marked point pattern objects
# marks: categorical; Kanji, hiragana, katakana alphabet or digit
x <- scolors3D[[2]] # x coordinates of the points of the pattern
y <- scolors3D[[3]] # y coordinates of the points of the pattern
xrange <- c(-100, 100) # vectors of length x
yrange <- c(-100, 100) # vectors of length y

m <- 1:1229; # number of all data
for (i in 1:1085) m[i] <- c("kanji") # kanji
for (i in 1086:1141) m[i] <- c("hiragana") # hiragana
for (i in 1142:1193) m[i] <- c("katakana") # katakana
for (i in 1194:1219) m[i] <- c("alphabet") # alphabet
for (i in 1220:1229) m[i] <- c("digit") # digit
mf=factor(m) # a vector of the same length as x and y
levels(mf) # check that the factor levels are as you intended
# specify the window
w <- owin(poly = list(x = c(-14.69306841, 116.063891, 82.28463313, 11.90915956, -138.9965256),
 y = c(-67.00573254, -94.4521156, 22.42358507, 134.9172171, 20.35588592)))
mpp_Hi <- ppp(x, y, window = w, marks = mf) # create marked point pattern
ppa_Hi <- split(mpp_Hi)$alphabet;
ppd_Hi <- split(mpp_Hi)$digit
pph_Hi <- split(mpp_Hi)$hiragana
ppk_Hi <- split(mpp_Hi)$katakana
ppj_Hi <- split(mpp_Hi)$kanji;
ppad_Hi <- superimpose(pph_Hi, ppk_Hi) # hiragana & katakana
as.data.frame(mpp_Hi)
```

Step2

Defining marks and creating a point pattern object

```
par(mfrow=c(2,2))
plot(ppad_Hi, xlim = xrange, ylim = yrange) # distribution of hiragana & katakana
axis(1, pos = 0, labels = FALSE)
axis(2, pos = 0, labels = FALSE)
plot(mpp_Hi, xlim = xrange, ylim = yrange) # marked point pattern
axis(1, pos = 0, labels = FALSE)
axis(2, pos = 0, labels = FALSE)
if (print_on) { dev.copy2eps(file="Data_RawScatterPlot.eps") }
plot(density(pp_Hi, sigma = 8.46)) # map of kernel density estimation
if (print_on) { dev.copy2eps(file="Data_RawDensityPlot.eps") }
```

Step3

Plot point patterns and The map of kernel density estimation

```
## ----- Analyses #----- For pooled data
Ksc_Hi <- Kest(pp_Hi) # estimate its K function
# plot as the L function defined in shimatani's review # L:= sqrt(K/pi)-r
par(mfrow=c(1,1))
plot(Ksc_Hi, cbind(sqrt(iso/pi) - r, 0) ~ r, ylim = c(-40, 40)) # the plot of the L function
# monte-carlo test; 1% level
# NULL hypo: Poisson null, i.e., CSR (Complete Spatial Randomness)
Esc_Hi <- envelope(pp_Hi, Kest, nsim = 39, rank = 1)
print(Esc_Hi)
plot(Esc_Hi, sqrt(hi/pi)-r ~ r, col = "blue", ylim = c(-40, 40), add = T)
plot(Esc_Hi, sqrt(lo/pi)-r ~ r, col = "blue", ylim = c(-40, 40), add = T)
if (print_on) { dev.copy2eps(file="Data_LfuncWithEnvelope.eps") }
```

Step4

Analyze point patterns based on the distance by means of L function in classification.

```
# cross K function
KCsc_Hi = alltypes(mpp_Hi, "K") #compute the cross-type between all pairs of types
Kkh_Hi=KCsc_Hi[5,3] # katakana vs. hiragana
# monte-carlo test # NULL hypo: , Poisson null, i.e., CSR
Ekh_Hi <- envelope(mpp_Hi, Kcross, nsim = 39, i = "katakana", j = "hiragana")
par(mfrow=c(2,2))
# plot as the cross L function
plot(Kkh_Hi, cbind(sqrt(iso/pi) - r, 0) ~ r, ylim = c(-40, 40))
plot(Ekh_Hi, sqrt(hi/pi)-r ~ r, col = "blue", ylim = c(-40, 40), add = T)
plot(Ekh_Hi, sqrt(lo/pi)-r ~ r, col = "blue", ylim = c(-40, 40), add = T)
if (print_on) { dev.copy2eps(file="HiragaAll_LCrossfuncWithEnvelope1.eps") }
```

Step5

Analyze two types of point patterns by means of crossL function

Fig. 13 Sample R script for our analysis of spatial statistics. Step 5 is included for further analyses of the dependency of two point patterns

pattern (Baddeley, 2010). For example, in addition to the color coordinates of synesthetic colors, a user could also record the kanji, hiragana, katakana, letter, and numeral. This marked point pattern is constructed by the *mf* in the script.

Step 3: Plot the point patterns and their density maps. First, the user can plot the point pattern using the *plot*(*x*, *window* = *w*) function, where *x* is a point pattern. Next, a user can plot a map of the kernel density estimation by a *plot*(*density*(*x*, *σ*)) call, where *x* is a point pattern and *σ* is the standard deviation of the isotropic Gaussian smoothing kernel used for the kernel method.

Step 4: Analyze the point patterns using the *L* function. The user can compute the *K* function for a point pattern *x* using the *Kest* function. Next, the *K* function can be used to plot the computed *L* function using *plot*(*K*, *sqrt*(*iso*/*pi*) ∼ *r*).

Step 5: Analyze the two types of point patterns with the cross-*L* function. The user can compute the cross-type between all pairs of types simultaneously using the command *alltypes*(*x*, *fun* = "*K*"), where *x* is the marked point pattern and *K* is the *K* function. An example of the function for two types of point patterns (katakana vs. hiragana) for synesthete S.H. is shown in Fig. 14.

Discussion

All previous studies on grapheme–color synesthesia have considered dozens of samples per synesthete (often 26 Latin letters and ten Arabic numerals). In this study, we created a new database, which we call the Japanese KANJI-Synesthetic Colors Database (K-SCD). The K-SCD has 1,229 and 2,349 synesthetic colors for each of two synesthetes, associated with 483 kanji, 46 hiragana, 46 katakana, 26 Latin letters, and ten Arabic numerals; 551, 460, and 285 synesthetic colors for each of three other synesthetes; and about 80 synesthetic colors for each of three synesthetes, associated with hiragana, Latin letters, and Arabic numerals. To our knowledge, no study has collected such a large number of synesthetic colors per synesthete. The large samples in the K-SCD should allow investigation of the intrapersonal structure of the grapheme–color synesthetic experience, and possibly lead to promising new ways of studying synesthesia.

We have illustrated two such ways here. Although these are preliminary studies, in the sense that only partial information from the database was utilized, they nevertheless provide important findings. The first example, taking advantage of the multidimensional properties of kanji, analyzed the regularity in the first- and second-order relationships between the radicals of kanji characters and their associated synesthetic colors. This analysis showed that radicals with shapes similar to
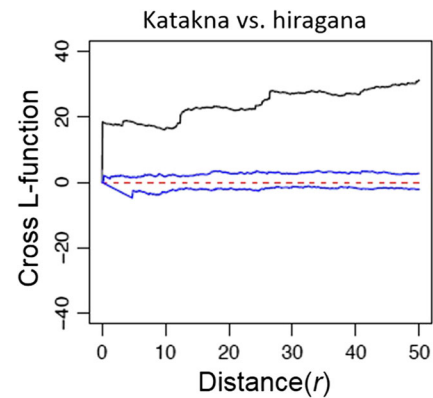
Fig. 14 Cross-*L* function of two types of point patterns for the chromatic synesthetic colors of synesthete S.H. The interval between the two blue lines indicates the 95 % confidence interval for complete spatial randomness

katakana characters have similar colors to those katakana characters. This second-order relationship suggests that the colors associated with certain kanji may originate from those associated with katakana. This possibility is consistent with a previous finding (Asano & Yokosawa, 2012) showing that synesthetic colors for Japanese graphemes acquired at an early age (i.e., numerals, hiragana, and katakana) affect those acquired later in life (i.e., kanji) via their phonological and/or semantic content. The present results suggest the possibility that visual shape also contributes as a key factor to the process of synesthetic learning. Future studies may reveal other factors by analyzing auxiliary information such as the pronunciations, meanings, and grade levels of kanji.

In the second application of K-SCD, we investigated the distribution of synesthetic colors in a*b* chromaticity space, to understand the global structure of synesthesia. The spatial statistical analysis revealed that the synesthetic color distribution was neither random nor uniform. Instead, synesthetic colors formed multiple clusters that were separated from each other in a*b* chromaticity space. This result corroborates a previous similar finding (Yamamoto, 2009) based on hiragana, Latin letters, and Arabic numerals. Why did these synesthetic color clusters emerge? We conjecture that some kinds of first-order relationships determine the core colors of clusters in early synesthetic learning, and then second-order relationships, such as those based on similarity, determine the synesthetic colors of newly learned graphemes. In the second process, a new character may be tinged with a color similar to the core color of the character to which it is most similar, and the color distance from the core may be proportional to the distance in similarity between the two characters in a given grapheme space, leading to clustering. In this scenario, the synesthetic color clusters may reflect both between- and within-category differences in the grapheme similarity space.

The emergence of the synesthetic color clusters may be further analyzed using the cross-*L* function to investigate the

interdependency of two point patterns. Point patterns can be divided into three categories: independent, attractive, or repulsive (see Fig. 1, bottom). Our hypothesis above predicts attractive dependencies between kanji and hiragana characters and between kanji characters across learning grades. This is supported by our preliminary analysis for the interrelation between hiragana characters and Latin letters (Yamamoto, 2009). A preliminary analysis for the interrelation between hiragana characters and katakana characters. (sample codes are shown in Fig. 13) also supports this hypothesis.

In conclusion, the K-SCD provides researchers with a variety of information regarding synesthetic color associations and enables novel research directions. The database, taking advantage of the multidimensionality of kanji characters and the abundance of possible samples, will be beneficial in developing our understanding of synesthesia. The database will be updated continually, as new data from new synesthetes and synesthetic color associations are obtained, or as room for improvement is found.

References

Asano, M., & Yokosawa, K. (2011). Synesthetic colors are elicited by sound quality in Japanese synesthetes. *Consciousness and Cognition, 20,* 1816–1823. doi:10.1016/j.concog.2011.05.012

Asano, M., & Yokosawa, K. (2012). Synesthetic colors for Japanese late acquired graphemes. *Consciousness and Cognition, 21,* 983–993. doi:10.1016/j.concog.2012.02.005

Asano, M., & Yokosawa, K. (2013). Grapheme learning and grapheme–color synesthesia: Toward a comprehensive model of grapheme–color association. *Frontiers in Human Neuroscience, 7,* 757. doi:10.3389/fnhum.2013.00757

Baddeley, A. (2010). Analysing spatial point patterns in R [Workshop notes]. Retrieved from https://research.csiro.au/software/wp-content/uploads/sites/6/2015/02/Rspatialcourse_CMIS_PDF-Standard.pdf

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis.* Harlow, UK: Longman Higher Education.

Ban, H., & Yamamoto, H. (2013). A non-device-specific approach to display characterization based on linear, nonlinear, and hybrid search algorithms. *Journal of Vision, 13*(6), 20. doi:10.1167/13.6.20. **1–26**.

Barnett, K. J., & Newell, F. N. (2008). Synaesthesia is associated with enhanced, self-rated visual imagery. *Consciousness and Cognition, 17,* 1032–1039. doi:10.1016/j.concog.2007.05.011

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: Prevalence and familiarity. *Perception, 25,* 1073–1079.

Baron-Cohen, S., Wyke, M. A., & Binnie, C. (1987). Hearing words and seeing colours: An experimental investigation of a case of synaesthesia. *Perception, 16,* 761–767.

Brang, D., Rouw, R., Ramachandran, V. S., & Coulson, S. (2011). Similarly shaped letters evoke similar colors in grapheme–color synesthesia. *Neuropsychologia, 49,* 1355–1358. doi:10.1016/j.neuropsychologia.2011.01.002

Chiou, R., & Rich, A. N. (2014). The role of conceptual knowledge in understanding synesthesia: Evaluating contemporary findings from a "hub-and-spokes" perspective. *Frontiers in Psychogy, 5,* 105. doi:10.3389/fpsyg.2014.00105

Cytowic, R. E., & Eagleman, D. M. (2009). *Wednesday is indigo blue: Discovering the brain of synesthesia.* Cambridge, MA: MIT Press.

Dixon, M. J., Smilek, D., & Merikle, P. M. (2004). Not all synaesthetes are created equal: Projector versus associator synaesthetes. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 335–343. doi:10.3758/CABN.4.3.335

Eagleman, D. M. (2010). *What has large-scale analysis taught us?* Article presented at the 2010 Meeting of the UK Synaesthesia Association, Brighton, U.K.

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods, 159,* 139–145.

Hubbard, E. M., Arman, A. C., Ramachandran, V. S., & Boynton, G. M. (2005). Individual differences among grapheme–color synesthetes: Brain–behavior correlations. *Neuron, 45,* 975–985. doi:10.1016/j.neuron.2005.02.008

Laeng, B., Svartdal, F., & Oelmann, H. (2004). Does color synesthesia pose a paradox for early-selection theories of attention? *Psychological Science, 15,* 277–281.

Luria, A. (1968). *The mind of a mnemonist.* New York, NY: Basic Books.

Mills, C. B., Viguers, M. L., Edelson, S. K., Thomas, A. T., Simon-Dack, S. L., & Innis, J. A. (2002). The color of two alphabets for a multilingual synesthete. *Perception, 31,* 1371–1394. doi:10.1068/p3429

Munsell Color. (1976). *Munsell book of color: Matte finish collection.* Baltimore, MD: Author.

Nunn, J. A., Gregory, L. J., Brammer, M., Williams, S. C., Parslow, D. M., Morgan, M. J., . . . Gray, J. A. (2002). Functional magnetic resonance imaging of synesthesia: Activation of V4/V8 by spoken words. *Nature Neuroscience, 5,* 371–375. doi:10.1038/nn818

Ramachandran, V. S., & Hubbard, E. M. (2001). Synesthesia: A window into perception, thought and language. *Journal of Consciousness Studies, 8,* 3–34.

Rich, A. N., Bradshaw, J. L., & Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: Implications for the role of early experience in lexical–colour associations. *Cognition, 98,* 53–84. doi:10.1016/j.cognition.2004.11.003

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., & Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology, 22,* 1069–1085. doi:10.1080/02643290500200122

Sperling, J. M., Prvulovic, D., Linden, D. E., Singer, W., & Stirn, A. (2006). Neuronal correlates of colour–graphemic synaesthesia: A fMRI study. *Cortex, 42,* 295–303.

Tamaoka, K. (1991). Psycholinguistic Nature of the Japanese orthography. *Studies in Language and Literature, 11,* 49–82.

Tammet, D. (2006). *Born on a blue day.* London: Hodder & Stoughton.

Ward, J., Li, R., Salih, S., & Sagiv, N. (2007). Varieties of grapheme–colour synaesthesia: A new theory of phenomenological and behavioural differences. *Consciousness and Cognition, 16,* 913–931. doi:10.1016/j.concog.2006.09.012

Ward, J., & Simner, J. (2003). Lexical–gustatory synaesthesia: Linguistic and conceptual factors. *Cognition, 89,* 237–261. doi:10.1016/S0010-0277(03)00122-7

Watson, M. R., Akins, K. A., & Enns, J. T. (2012). Second-order mappings in grapheme–color synesthesia. *Psychonomic Bulletin & Review, 19,* 211–217. doi:10.3758/s13423-011-0208-4

Yamamoto, H. (2009). *An exploratory analysis of the synesthetic color distribution in chromaticity space* [in Japanese]. Paper presented at the 73rd Annual Convention of the Japanese Psychological Association, Japan, Kyoto.