Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n3p588

Classes of Colors and Timbres: A Clustering Approach

By Mannone, Distefano, Santini

Published: November 20, 2022

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 03, November 2022, 588-605
DOI: 10.1285/i20705948v15n3p588

Classes of Colors and Timbres: A Clustering Approach

Maria Mannone* [a,b], Veronica Distefano[b], and Giovanni Santini[c]

[a] *University of Palermo, Engineering Department, Italy*
[b] *Ca' Foscari University of Venice, DAIS and ECLT, Italy*
[c] *Xi'an Jiaotong-Liverpool, Suzhou, China*

Published: November 20, 2022

Similarities between different sensory dimensions can be addressed considering common "movements" as causes, and emotional responses as effects. An imaginary movement toward the "dark" produces "dark sounds" and "dark colors," or, toward the "bright," "brighter colors" and "brighter sounds." Following this line of research, we draw upon the confluence of mathematics and cognition, extending to colors and timbres the gestural similarity conjecture, a development of the mathematical theory of musical gestures. Visual "gestures" are seen here as paths in the space of colors, compared with paths in the space of orchestral timbres. We present an approach based on clustering algorithm to evaluate the association between color bands and orchestral timbres. The analysis is based on 8 indicators which represent and describe participants' background and associations to be tested. The indicators include socio-demographic information and color class options from the color space, to associate with each given timbre class. We clustered responders into homogeneous groups where the within-group-object dissimilarity is minimized and the between-group-object dissimilarity is maximized. The partitions are obtained with k-modes. While participants' background does have an influence in their answers, the overall behaviors confirms the existence of different space regions for different timbres, supporting our hypothesis of perceived similarities similarities between color and timbre classes. In fact, the cluster analysis confirms identifiable blocks. Our pioneering study on a small dataset may open the way toward a future and deeper comprehension of complex color-timbre perceived connections.

*Corresponding author: maria.mannone@unive.it; mariacaterina.mannone@unipa.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

keywords: crossmodal perception, similarity, timbre, color, clusters

1 Introduction

Nature or nurture? Individual or collective? Particular or universal? Several questions arise while investigating art in light of science, especially while dealing with colors and sounds. Inspiration and irrationality may lead to special artwork that can be later rationally investigated. Vice versa, scientific concepts can lead to the creation of new artworks and experiments. Seeking similarity also means researching the uniqueness within the variety. Associations between colors and sounds can be contextualized in the framework of synesthesia and crossmodal correspondences (Corno, 2011; Marks, 1987). In psychology, "synesthesia" denotes automatic and involuntary multisensory perception—one hears a sound and sees a precise color, consistently associated with that sound. Crossmodal correspondences are "systematic associations found across seemingly unrelated features from different sensory modalities" (Parise, 2016). Every synesthetic subject is consistent in their associations even after long periods. Conversely, crossmodal correspondences, the audiovisual-object theory (Kubovy and Schutz, 2010), and the hypothesis of a supramodal brain (Rosenblum et al., 2016) concern multisensory association shared among different people. Synesthesia aroused the interest of artists and scientists: Messiaen (Bernard, 1986), Scelsi (Castanet and Cisternino, 2001), Kandinsky (Cooper, 2013; Cytowic, 1995; Ione and Tyler, 2003; Kandinsky et al., 1994; Robert, 1975), Ligeti (Ligeti et al., 2004; Peacock, 1985), Scriabin, Baudelaire, Rimbaud, Nabokov, Eisenstein, Hockney, Feynman, Locke (Locke, 1689), Leibniz, Darwin, Sachs (Bernard, 1986; Ione and Tyler, 2003; Jewanski et al., 2009; Peacock, 1985), Ciurlionis (Crnjanski and Tomač, 2019), Berlioz (Barbiere et al., 2007).

Here, we deal with color synesthesia (chromesthesia). It has been shown that arts can improve the quality of people lives (Michalos, 2005); synesthesia and crossmodal correspondences help understand connections between arts, and thus they can further advance the positive action of multimodal artworks.

Former experiments. Several studies agree regarding musical instruments and colors associations (Jones, 1973). Other works focus on the strength of color-saturation and loudness (Panek and Stevens, 1966), as well as color-saturation and pitch (Huang, 1998a). Correspondences between certain characteristics of vision and hearing are investigated (Marentakis, 2020; Marks, 1974); including higher pitch / louder sound, color luminosity / sound loudness, saturation / timbre / duration associations (Caivano, 1994), color luminosity / space location, timbre/visual shape correspondences (Adeli et al., 2014). Some experiments investigate similarities and differences between synesthetic and non-synesthetic people, also involving eye-movements, attention, and emotions (Hagtvedt and Brasel, 2016; Bresin, 2005; Barbiere et al., 2007). Musical-spatial analogies might also be asymmetrical: e.g., musical changes in one direction evoke significantly stronger spatial analogies than their opposites (Eitan and Granot, 2005). Other studies (Crnjanski and Tomač, 2019) investigate the association between color brightness, lines, and timbres, and the relationship between sound pitch and color saturation for product choices (Huang et al., 2019).

Mathematical and computational approaches. Colors and timbres admit a common mathematical description: a superposition of simple elements with opportune coefficients—Fourier series for sound, color superposition for light, a Fourier-like series and transform for colors (Guan et al., 2014). Some studies focus on the physics of color, seeking for analogies with the physics of complex sounds (Isaac, 2018). Machine learning is used to automatically associate colors and images (Lee et al., 2019), using interactive interfaces (Peacock, 1985), and Virtual Reality color-scanners (VR), such as the Synesthesizer, a synthesizer based on cross-synthesized physical models regulated by machine learning (Santini, 2019). In the pioneering study by Grey (Grey, 1977), a space of timbres is defined, where the more similar the timbres, the closer the points in the space. Grey proposed a hierarchical clustering approach to build up clusters or closely-related timbres. This computational approach is very close to the orchestral music practice, where instruments are clustered in "sections" according to their main characteristics. Reuter et al. (Reuter et al., 2018) considered a palette of given colors, and asked 40 participants to choose orchestral sounds (instruments and ranges, from a given set) that, in their opinion, were best matching each color.

Our approach. We aim to verify the presence of any timbre-bands/color-bands associations through a clustering approach. In fact, we focus on classes comparisons, rather focusing on one-to-one correspondences, which may be highly subjective. Our study is thus inspired by loose associations, to be described with fuzzy logic (Isaac, 2018), category theory (Mac Lane, 1971), and gestural similarity (Mannone, 2018). We hypothesize that color-timbre associations are mediated by a *common perceptive reaction*. In this sense, our research is connected with Palmer's studies (Palmer et al., 2013), where emotion is seen as a connection between sound and color. Because we focus on classes of colors and timbres, and, consequently, classes of participants' answers and thus classes of participants grouped according to their answers, we choose to consider all questionnaire answers as a categorical/nominal variables (Ammar E. Z., 2012).

While (Reuter et al., 2018) considers one-to-one associations, in our experiment we consider instead classes of colors as color bands, with orchestral timbres already grouped into characteristic orchestral features. We also gave participants the opportunity to choose, through a color picker, the precise tonality they judged to be the best match.
Our findings confirm some of the results of (Reuter et al., 2018), regarding the correspondence between low pitches (and close harmonic, as for the low-register piano cluster of note) and dark colors, and "brilliant" sound with more luminous colors, light colors and high-register winds, yellow-ish and reddish colors with trumpets and trombones (evident in associations 2, 7, and 8 of our experiment).
The main characteristics of our study are the chance of choosing the precise color in the space of colors; the theoretical hypothesis of classes correspondence rather than one-to-one correspondence, and thus the use of categorical information; the underlying hypothesis of perceptive similarity.

In this context, amongst clustering methods, the k-modes method is widely exploited while dealing with categorical data, for its clarity of use and its suitability for big data analysis. We apply the k-modes method to a small dataset.

The paper is structured as follows. In Section 2, we summarize the theoretical background and the research framework of our study. In Section 3, we describe the adopted clustering algorithm with the motivation of our choices. In Section 4, we present and discuss our results. Finally, Section 5, ends the paper, envisaging further and future lines of research.

2 Theoretical background

While music and mathematics are apparently seen as distant disciplines, from Pythagoras their intersection are numerous, and they inspire today's scientific research. A branch of mathematical music theory focuses on musical gestures (Mazzola and Andreatta, 2007; Arias, 2018; Clark, 2020), described through the formalism of category theory (Mac Lane, 1971). Category theory is an abstract branch of mathematics, whose basic components are points and arrows are points and arrows (morphisms) between them, to model abstract transformations. Categories are nowadays applied in different fields, including physics, chemistry (Spivak, 2014), and music (Mazzola and Andreatta, 2007).

In the framework of gesture theory, it has been developed a study on gestural similarity (Mannone, 2018). In a nutshell, it is a mathematical and perceptive condition where specific visual sketches and musical sequences appear as being produced by the same creator gesture, and thus perceived as "similar." The "gestural similarity conjecture" has been verified in some preliminary experiments (Mannone and Papageorgiou, 2020), see the upper side of Figure 1.

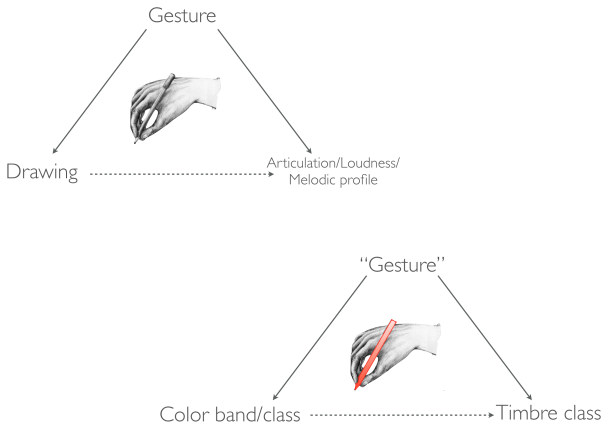


Figure 1: Triangular diagrams to synthesize the conjecture of gestural similarity (up) and of chromo-gestural similarity (down).

We focus here on relationships between "points" and "gestures" in the space of timbres and in the space of colors. Gestures in the space of colors are transitions from a color to another one. Gestures in the space of timbres are transitions from a timbre to another one. Their formalism and mutual interactions have also been described through categories, and in particular, through bicategories and bigroupoids, because each gesture can be inverted (Mannone et al., 2022). Thus, 'colors' and 'timbres' can be compared through the common notion of 'gesture' (Figure 1, bottom).

In this article, we present an experimental study to verify the amount of perceived associations between portions of color space and portions of timbre space. In this research, meant to explore the conditions of *chromo-gestural similarity*, we do not consider any one-to-one mapping or any direct correspondence between sound and color, but an 'indirect' comparison based on perceptive similarity. We can depict the perception of a synesthetic person as an arrow picking up some points from the space of colors, and mapping them onto some points of the space of timbres—always the same target, as in a function. However, the arrows for the non-synesthetic person do not reach the same target all the time, but a range of targets. Such a range of targets for one person, and for another person as well, lies within an equivalence class.

3 Methods

3.1 Clustering method

Clustering techniques are an example of unsupervised machine learning. We use clustering techniques to divide a data population into a certain number of groups, where observations within a cluster are more similar to each other than to the observations in other clusters. Clustering methods, developed in statistics and machine learning, are used for data processing in several fields of application, that include image processing, medicine, economics, and medicine (McLachlan and Basford, 1988; Melnykov and Maitra, 2010; Ramey, 1985), including Covid-19 pandemic data (Keser and Kocakoç, 2021). We use clustering techniques to divide a data population into a certain number of groups, where observations within a cluster are more similar to each other than to the observations in other clusters. Cluster analysis is an unsupervised learning technology to identify hidden patterns across data. Input data are not labeled, and the division into classes maximizes intra-class similarity and minimizes inter-class similarity.

Each clustering technique is characterized by a unique and precise notion of similarity and a specific metric (Hartigan and Wong, 1979). Clustering algorithms can be classified into two classes: hierarchical algorithms and partitioning algorithms.

In hierarchical clustering, clusters are formed by iteratively dividing the patterns through a divisive or an aggregative approach (Nielsen, 2016). Once defined a measure of similarity between two groups of points, such as a metric and a linkage, one can perform splitting and merging procedures (Hartigan and Wong, 1979).

The metric allows one to evaluate the distance between each pair of points within the clusters. The linkage, a function of the metric, permits to measure the similarity between two clusters.

Partitioning algorithms assign data to a certain number of clusters by iteratively optimizing some objective function (Reynolds et al., 2009). In this study, we choose to focus on a partition algorithm known as k-modes type clustering for categorical data, presented by Huang (Huang, 1998a). The k-modes algorithm is an adaptation of the k-means algorithm. It has been designed to cluster categorical data sets (Bai et al., 2012; Kim and Hyunchul, 2008). According to Huang (Huang, 1998a), the k-modes is in general faster than the k-means thanks to the fewer number of iterations needed to converge (Wilde et al., 2020; Dorman and Maitra, 2020). The k-modes introduces the following modifications: replaces cluster means with cluster modes; uses a simple matching dissimilarity measure for categorical objects, and utilizes a frequency-based method for updating clustering modes, minimizing the cost function.

With respect to the k-means, the k-modes introduces the following modifications: it "uses a simple matching dissimilarity measure for categorical objects, it replaces means of clusters with the modes, and it uses a frequency-based method to update the modes" (Huang, 1998a).

Aiming to subgroup the population and identify patterns across data, we define the number K of clusters and we select the k initial modes. Then, we allocate every object to the nearest mode. After a small number the iterations, the algorithm converges. Similarly as the k-means, the k-modes algorithm provides locally optimal solutions, that are heavily dependent on the initial mode or on the objects' order; this fact constitutes the main disadvantage of the considered method.

The considered distance between categorical objects in the k-mode algorithm is the simple match dissimilarity measure. According to (Huang, 1998a, 289), the algorithm can be described as follows:

> Let $\mathbf{X}$ be a set of categorical objects described by m categorical attributes, $\{A_1, A_2, ..., A_m\}$. The mode of X is defined to be a vector $Q = [q_1, q_2, ..., q_d]$ such that the function is minimized, where the distance is defined as:
>
> $$D(\mathbf{X}, Q) = \sum_{i=1}^{n} d_1(X_i, Q) \tag{1}$$

The simple-matching dissimilarity measure (Kaufman and Rousseeuw, 1990; Huang, 1998a,b) is a well-known measure used for categorical data. According to (Çilingtürk and Ergüt, 1977; Kaufman and Rousseeuw, 1990), the simple matching distance is defined as follows. Let X and Y be two categorical objects, characterized by m attributes. The simple matching distance (Kaufman and Rousseeuw, 1990) between X and Y is given by

$$d(X, Y) = \sum_{k=1}^{m} \delta(x_k, y_k), \quad \delta(x_k, y_k) = \begin{cases} 0 & (x_k = y_k) \\ 1 & (x_k \neq y_k) \end{cases}$$

In our research, we can consider the versions with and without weights, 30 as the maximum number of iterations.

Categorical data clustering plays a major role in social, economic or medical fields, to find similar patterns in a defined population, identify decisions and targets for healthcare, measure attitude and options in social sciences. For the process of data clustering, the k-modes algorithm has been widely used, mainly due to its easiness of implementation and capacity to handle big data amount. Variations of the k-modes algorithm include mixture models for categorical data, fuzzy models, an iterative initial points refinement algorithm for categorical data, a genetic clustering algorithm (Gk-mode). Many algorithms have been proposed for the clustering of categorical data (Saha and Das, 2015; Liu et al., 2020). Here, we focus on the classic k-modes algorithm, because it is the most suitable tool for the analysis of a categorical dataset, obtaining subgroups based on elements' dissimilarities. For our analysis, we also consider centroids and voxel representations.

In our study, the number of clusters for k-modes clustering is set to $k = 4$. We choose 4 clusters because we are considering groups of 4 colors for each task. In fact, we evaluated people behavior accordingly to their answers to the color picking task and the band association task.

We calculate distances through the simple-matching method (Hamming distance). The steps to perform the k-modes clustering are summarized in Algorithm 1. All analysis have been performing using the R statistical software.

Algorithm 1 k-modes pseudocode

```
Choose a value of k
Select initial centres (modes)
Calculate the distances between objects to the cluster modes
Update the modes value and repeat the step to calculate the distances
while Changing cluster membership do
    Repeat the process
end while
```

3.2 Questionnaire

We prepared an online anonymous form, a questionnaire "Sound and Color Associations," with multiple-answer questions, open questions, sound examples, and color pickers. A few general questions were followed by a 2-part experiment with 4+4 associations to be made. The form is available in the folder at `https://tinyurl.com/uyrzr79s`. The first questions regarded artistic activity, age range, and synesthetic experiences, respectively. We had the following demographic questions in the first part of our questionnaire:

- Q1. Are you a musician or a visual artist?
    1. Musician
    2. Visual Artist
    3. Both musician and visual artist
    4. None of the two
- Q2. What's your age range?
    1. 8-17
    2. 18-29
    3. 30-44

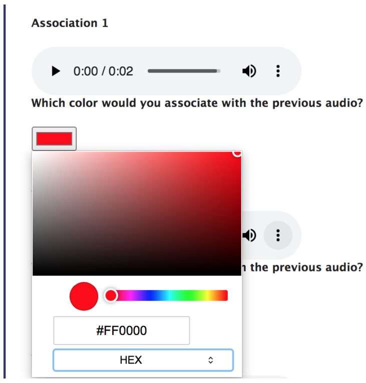
   4. 45-60
   5. 60+

- Q3. Are you a synesthetic subject/have had synesthetic experiences?
   1. Yes
   2. No
   3. I don't know

The second part of the questionnaire (containing eight questions) regarded an association task. Participants were asked to associate given timbre combinations and colors, without any explanations. During this association task, participants listened to a single chord orchestrated differently. The different orchestrations were simulated with a sound library.

The first part of the association task (four questions, Associations 1-4) regarded sound-color free associations; the second part (other four questions, Associations 5-8) regarded color-band and sound associations.

The first four associations allowed the use of a color picker (Figure 2), outputting hexadecimal color values. Those values were then translated into RGB (Red Green Blue) coordinates.



Note. structure of the Associations 1-4

Figure 2: Structure of questionnaire for associations 1-4. Clicking on the color box, participants were able to use the color picker.

Each answer to each one of these questions is a point in the space of colors. To highlight color classes, we divided the color space into cubic boxes, grouping close answers (in fact, space closeness corresponds to tone/luminosity/hue closeness). The color of the so-obtained *voxels* depends upon the centroid of the colors it contains. Opacity depends on the concentration of these answers: the more the choices of 'black,' the more black and the less transparent the corresponding voxel will be. In this way, we converted a continuous variable, the coordinates of each color, into 125 voxels, which we consider as a categorical variable because of our choice to focus on classes. Voxels have been obtained through a Python code.

The remaining four associations presented, each of them, four gradients of colors (as a collection of four colored squares) to choose from (Figure 3). In this case, the output was a number between 1 and 4, according to the chosen set of colored squares.

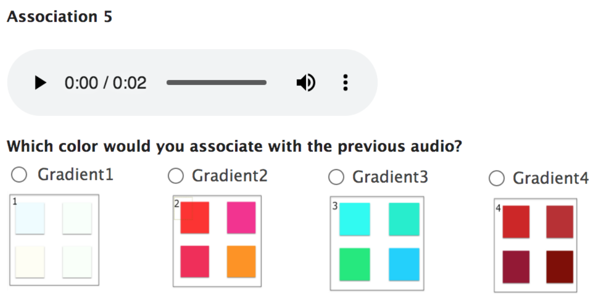


Figure 3: Structure of questionnaire for associations 5-8

Regarding sound samples, we chose a C-Major chord for all musical samples. The orchestration of the sound samples, and the indicated score loudness (which influences on timbres), is the following:

1. Sound for Association 1: 2 flutes, 2 oboes, 1 clarinet, Violin I, Violin II, ***f***
2. Sound for Association 2: Full orchestra, ***ff***
3. Sound for Association 3: 3 Clarinets, ***p***
4. Sound for Association 4: Piano, ***f*** (low-pitch chord)
5. Sound for Association 5: 1 flute, 1 oboe, 1 clarinet, Violin I, Violin II, ***pp***
6. Sound for Association 6: 3 flutes, ***p*** (relatively high-pitch chord)
7. Sound for Association 7: 3 trumpets, ***f***
8. Sound for Association 8: 2 trombones, 3 celli, ***f***.

The expected results were indicatively warm/brilliant colors for Associations 1 and 3, colder/darker colors for Associations 2 and 4, and, for the gradient association-task (Associations 5-8), the sequence 3-1-2-4. Fluctuations are also given to different color-screen characteristics.

4 Results

We had 52 participants to our experiment, recruited over emails and social networks during the first decade of June 2020.

Population. Concerning the population, the 62.2% is constituted by musicians, while the 25.5% was not a musician nor a visual artist. The age range is comprised, for the 41.2%, between 18-29 years, while the 37.3% declared to belong to the range 30-44; the

11.8% is between 45 and 60 years old. With a reference to synesthetic experiences (Q3), it is interesting to note that the most part of participants (45.1%) declared of not having had any of such experience, while the 31.4% declared of not knowing.
Amongst participants who declared of being a musician, the 43.8% replied "No" regarding the Q3 question. The 53.8% of non-musicians and non-visual artists replied "No" to Q3, while the 38.5% of them replied "I don't know."

The expected associations for Associations 4-5-7-8 were the options 3-1-2-4, respectively. Concerning these questions, musicians provided the following answers: the 50% selected the option 3 for Association 5; for Association 6, the 40.6% selected the option 1; for Association 7, the 56.3% chose option 2, and for Association 8, almost all musicians chose the option 4. Thus, we find that the most part of musicians chose the expected associations.

Cluster population is divided in the following way: cluster 1 includes the 17.6% of participants; cluster 2 the 23.51%; cluster 3 the 49%, and cluster the 9.8%.

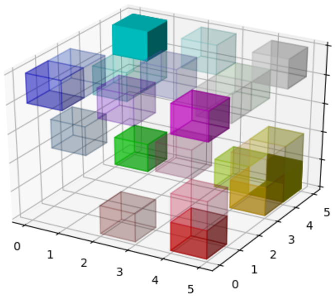
Observing the cluster population with respect to question Q1, clusters 1 and 2 are constituted by musicians for the 88.9% and 75%, respectively. In cluster 3, there is an equidistribution between musicians, visual artists, and both of them. In cluster 4, the 80% of people are neither musicians not visual artists.

Each picture in Figure 4 shows the colors associated to the first four sounds (Associations 1–4), as distribution of associations in the RGB (Red Green Blue) space. The RGB space can be described as a portion of 3D space with coordinates between 0 and 255 along the x (red), y (green), and z (blue) axes. These coordinates express the amount of the three primary colors included in a given color. For example, black has RGB coordinates (0, 0, 0), while white (255, 255, 255). As the RGB space provides more than 16 million combinations, we found that a quantization of answers could be more expressive. In fact, respondents would unlikely be able to distinguish between a yellow (255, 255, 0) and a yellow (254, 254, 0). For this reason, we decided to use voxels, i.e., volumetric pixels. Each voxel is a cube that samples the RGB space at a fixed step. The details of participants' subgroups for each association task can be found in the folder "clusters" of the repository at `https://tinyurl.com/uyrzr79s`. As an example, in Figure 5 we show participants' answers to the Association task 4, clustered according to their answers to the other questions.

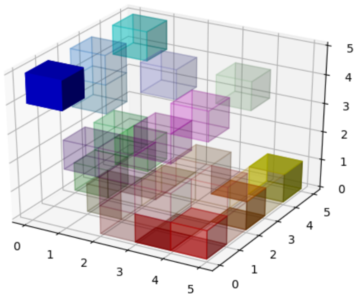
Algorithm 2 voxel partition pseudocode

Divide each axis in 5 equal parts
Create a *voxelArray* of 125 voxels based on the quantized coordinates
Import the *.csv* file with color associations
for each *voxel* in *voxelsArray* **do**
    Calculate number n of color choices that fall inside the cube volume
    Assign alpha value according to n
end for
Render the colored voxel with transparency values

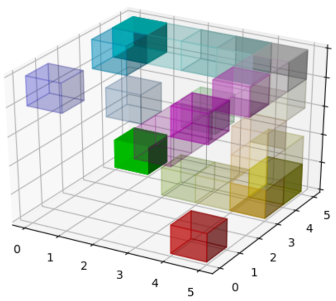
In order to construct the voxels, we divided each axis in 5 equal segments, by finding 6



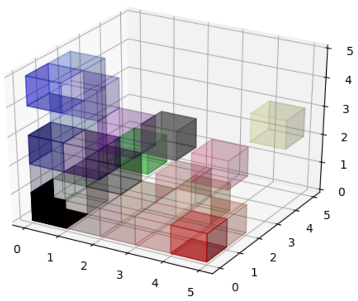
(a): Colors associated with 2 flutes, 2 oboes,



(b): Colors associated with a full-orchestra



(c): Colors associated with a 3-clarinet chord, *p*.

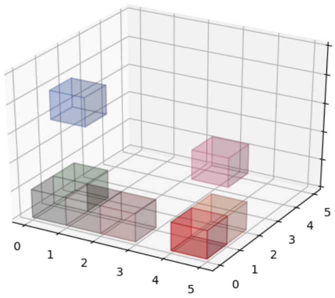


(d): Colors associated with a low-pitch piano chord, *f*.

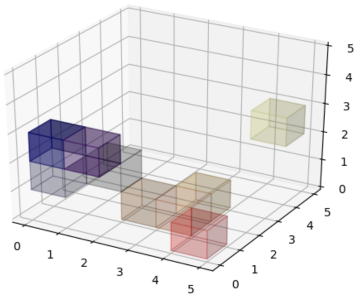
Figure 4: Color associations for each timbre (part I of the test, Associations 1–4), corresponding to a different orchestration of a C-major chord. The more intense the color of a cube, the more frequent the answers in that portion of space. Colors in (c) are similar to (a) but more "bright." Space distributions of (c) and (d) are almost complementary.

equally distant integers between 0-255 (extremes included). Thus, we sampled the RGB space in 125 voxels. Each voxel was associated with the color given by the coordinates of the center of each cube. We then used a python script to process a *.csv* file containing the answers for each association. From the online form described in 3.2, we gathered color information in hexadecimal format. Therefore, we had first to convert hexadecimal values into RGB coordinates. After that, for each association, we counted how many responses were falling inside each voxel.

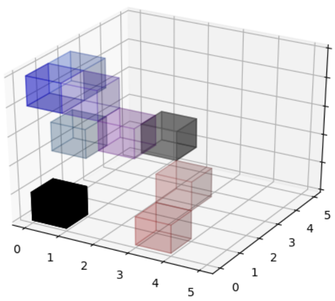
The number of answers would provide a value for the transparency of the cube (0 answers = invisible). Algorithm 2 reports a pseudo-code for the creation of voxels. In



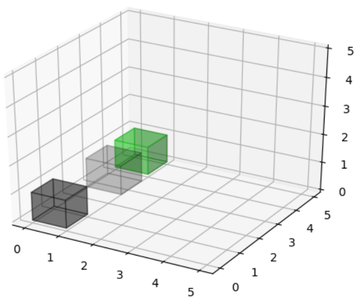
(a)



(b)



(c): Black-ish colors: the most expected answers.



(d): This partition contains green, a less expected association for the proposed timbre.

Figure 5: Partition of color selections displayed in Figure 4 (d), that is, Association 8, according to participants' replies to the second part of the test. Despite fluctuations, almost none of the participants selected "brilliant" or very luminous colors for this question, confirming our expectation of a "dark" color association.

order to represent the cluster distribution described in 3.1, we added one column in the *.csv* file. Numbers from 1 to 4 were used to indicate the inclusion of each row (i.e., respondent) in one of the four clusters. For each voxel's transparency value, we only counted respondents assigned to one cluster at a time.

To visually summarize the overall behavior of participants (Figure 4), for each one of the first four associations we also computed the centroid of color choices, shown in Figure 6. The centroid is meant to provide the "baricenter" of the selections, indicating the point of minimal distance from all choices. The computation was realized by calculating the geometric centroid (mean of coordinates) of color choices in the RGB space, where each choice is a point in that space. As expected, we find that centroids for Association 1 and

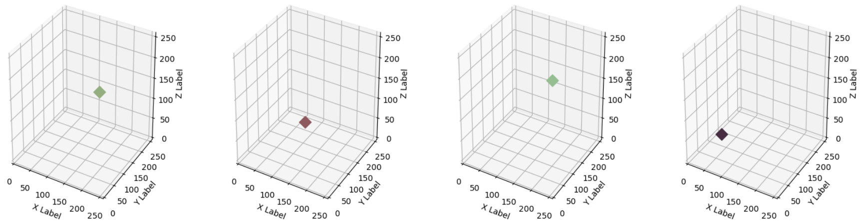


Figure 6: Centroids of color choices for associations 1-4.

Association 3 are close (with a little brighter color for Association 3), while the centroid for Association 2 belongs to a warm color region, and the centroid for Association 4 is a very dark color.

Participants were also asked to associate some sounds (the same chords played by different orchestral combinations) with a color band amongst four color bands (Table 1). The expected answers (as the best match) to each question were: 3, 1, 2, 4. The majority of participants chose the third answer (third color-band choice) to the first question (first sound), the first answer to the second question, the second answer to the third question, and the fourth answer to the fourth question. In particular, the third and fourth sounds were more distinctly associated with the expected questions.

Table 1: For each timbre, participants selected a color band. The expected color-band selection was 3-1-2-4 for timbres 1-2-3-4, respectively. The number of participants that selected these options are highlighted in bold.

color bands	timbre 1	timbre 2	timbre 3	timbre 4
1	16	**25**	6	3
2	12	7	**31**	7
3	**24**	17	9	2
4	0	3	6	**39**

5 Discussion and Conclusions

We faced the problem of associations between colors and sounds, investigating the correspondence between orchestral-section chords and blurred color bands. We started from the gestural similarity conjecture applied to colors. Here, we consider as gestural generators unseen paths in the space of parameters, moving toward "warm and brilliant" and then toward "dark and cold," or again "dark and warm." The parameter spaces are the conceptual space of colors and timbres, with "gestures" as paths, to be compared according to common perceptive/emotional reactions to visual and auditory stimuli. We designed an experiment comparing points in the space of sounds and points in the space of colors. Participants filled an online questionnaire with some kind of color associations. Our results confirmed the expectations, showing a variability circumscribed within recognizable regions of the color space. Color-distribution analogies reflected timbral analogies. Our results strongly suggest the emergence of multi-sensory properties.

Further research will focus on chromatic/timbric transitions. From a computational point of view, the next objective will be to replicate the analysis while increasing the number of participants. With more people and more parameters, we might catch more nuances of timbre-color perceived connections, as an overall greater diversity with respect to the considered small dataset. Working with an increased observed population will allow us to explore eventual macro-connections of classes of colors and classes of timbres. Future studies may uncover less-expected patterns of behavior which may emerge from a wider sample. From a computational point of view, we can combine our analysis with graphic techniques for centroid-based categorical data clustering (Manisera, 2011). We also aim to compare different clustering techniques. Our study can lead to a future predictive model, to design a sonification technique based on shared timbre-color similarities. For instance, a neural-network model can be fed with the results of our study, to refine the choice of color and timbre features and automatically predict possible well-working associations. The statistical approach may for example be borrowed from (Athanasiadis and Ioannides, 2021).

The sense of the present research ultimately is aiming to know more on the arts and on our own perception of complex phenomena such as timbres and colors. Are our senses and our 'artistic perception' giving us information about nature and physics? Our study can lead to new signs of progress in multisensory data exploration, sound/visual design, and ultimately, human-perception knowledge.

Declarations

The authors received no funds for this research. No sensitive data were collected during the experiment. Ethical approval was not applicable in this research. The consent to participate was given by accepting the invitation via email. The research did not involve minor subjects. Data and materials are available at the folder `https://tinyurl.com/uyrzr79s`. The authors contributed equally. The authors declare no conflict of interest.

References

Adeli, M., Rouat, J., and Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Frontiers in Human Neurosciences*, 30.

Ammar E. Z., L. P. (2012). K-modes clustering using possibilistic membership. In *IPMU 2012, Part III, CCIS 299*, pages 596–605.

Arias, J. S. (2018). Spaces of Gestures are Function Spaces. *Journal of Mathematics and Music*, 12(12):89–105.

Athanasiadis, I. and Ioannides, D. (2021). Selection of features and prediction of wine quality using artificial neural networks. *Electronic Journal of Applied Statistical Analysis*, 14(2):22657.

Bai, L., Liang, L., Daang, C., and Fuyuan, C. (2012). A cluster centers initialization method for clustering categorical data. *Expert Systems with Applications*, 39(9):8022–8029.

Barbiere, J. M., Vidal, A., and Zellner, D. (2007). The Color of Music: Correspondence through Emotion. *Empirical Studies of the Arts*, 25(2):193–208.

Bernard, J. W. (1986). Messiaen's Synaesthesia: The Correspondence between Color and Sound Structure in His Music. *Music Perception*, 4(1):41–68.

Bresin, R. (2005). What is the color of that musical performance? In *Proceedings of the International Computer Music Conference - ICMC 2005*, pages 367–370.

Caivano, J. L. (1994). Color and Sound: Physical and Psychophysical Relations. *Color: Research and Applications*, 19(2):126–133.

Castanet, P. A. and Cisternino, N. (2001). *Viaggio al centro del suono*. Luna Editore, La Spezia.

Çilingtürk, A. M. and Ergüt, O. (1977). Hierarchical Clustering with Simple Matching and Joint Entropy Dissimilarity Measure. *Journal of Modern Applied Statistical Methods*, 13(1):329–338.

Clark, T. L. (2020). On the topological characterization of gestures in a convenient category of spaces. *Journal of Mathematics and Music*.

Cooper, J. (2013). Kandinsky as a Synesthete. Game Anim. Retrieved from `https://www.gameanim.com/2013/07/30/kandinsky-as-synesthete/`.

Corno, D. (2011). Sinestesia. Enciclopedia Treccani, `http://www.treccani.it/enciclopedia/sinestesia_%28Enciclopedia-dell%27Italiano%29/`.

Crnjanski, N. and Tomač, D. (2019). Musical Perception and Visualization. Music and Spatiality. Paper read at Music and Spatiality. 13th Biennale International Conference on Music Theory and Analysis.

Cytowic, R. E. (1995). Synesthesia: Phenomenology and neuropsychology: A review of current knowledge. *Psyche: An Interdisciplinary Journal of Research on Consciousness*, 2(10).

Dorman, K. S. and Maitra, R. (2020). An Efficient k-modes Algorithm for Clustering Categorical Datasets. arXiv, `arXiv:2006.03936`.

Eitan, Z. and Granot, R. (2005). How music moves: Musical parameters and listeners' images of motion, Perceptual characterization of motion evoked by sounds for synthesis control purposes. *Music Perception*, 23(3):221–247.

Grey, J. (1977). Multidimensional Perceptual Scaling of Musical Timbres. *The Journal of the Acoustical Society of America*, 61:1270–1277.

Guan, T., Zhou, D., Xu, C., and *et al.* (2014). A novel RGB Fourier transform-based color space for optical microscopic image processing. *Robot. Biomim.*, 1(16).

Hagtvedt, H. and Brasel, S. A. (2016). Cross-Modal Communication: Sound Frequency Influences Consumer Responses to Color Lightness. *Journal of Marketing Research*, 53(4):551–562.

Hartigan, J. A. and Wong, M. A. (1979). A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108.

Huang, L., Qian, D., and Li, O. (2019). The cross-modal interaction between sound frequency and color saturation on consumer's product size perception, preference, and purchase. *Psychology and Marketing*, 37(7):876–899.

Huang, Z. (1998a). A fast clustering algorithm to cluster very large categorical data sets in data mining. *Research issues on data mining and knowledge discovery*, 2:283–304.

Huang, Z. (1998b). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2:283–304.

Ione, A. and Tyler, C. (2003). Neurohistory and the Arts: Was Kandinsky a Synesthete? *Journal of the History of Neurosciences*, 12(2):223–226.

Isaac, A. M. C. (2018). Prospects for timbre physicalism. *Philosophical studies*, 175:503–529.

Jewanski, J., Day, S., and Ward, J. (2009). A Colorful Albino: The First Documented Case of Synaesthesia, by Georg Tobias Ludwig Sachs in 1812. *Journal of the History of the Neurosciences*, 18(3):293–303.

Jones, T. D. (1973). *Art of Light and Colour.* Van Nostrand Reinhold, New York.

Kandinsky, W., Lindsay, K. C., and Vergo, P. e. (1994). *Complete Writings on Art.* Da Capo Press, New York.

Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley & Sons, New York.

Keser, I. K. and Kocakoç, I. D. (2021). Functional Cluster and Canonical Correlation Analysis of EU Countries by Number of Daily Deaths and Stringency Index During Covid-19 Pandemic. *Electronic Journal of Applied Statistical Analysis*, 14(2):22891.

Kim, K. and Hyunchul, A. (2008). A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications*, 34(2):1200–1209.

Kubovy, M. and Schutz, M. (2010). Audio-Visual Objects. *Review of Philosophy and Psychology*, 1(1):41–61.

Lee, J. Y., Moon, C. B., and Kim, B. M. (2019). The Color-Music Relationship Direct Modeling using Machine Learning. In *Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 629–631, Zagreb, Croatia.

Ligeti, G., Roelcke, E., and Peroni, A. (2004). *Lei sogna a colori?* Alet, Padova.

Liu, C., Wang, X., Huang, Y., Liu, Y., Li, R., Li, Y., and Liu, J. (2020). A Moving Shape-based Robust Fuzzy k-modes Clustering Algorithm for Electricity Profiles. *Electric Power Systems Research*, 187:106425.

Locke, J. (1689). *An Essay Concerning Human Understanding.* Clarendon Edition, London.

Mac Lane, S. (1971). *Categories for the Working Mathematician.* Springer, New York.

Manisera, M. (2011). A graphical tool to compare groups of subjects on categorical variables. *Electronic Journal of Applied Statistical Analysis*, 4(1):1–22.

Mannone, M. (2018). Introduction to Gestural Similarity in Music. An Application of Category Theory to the Orchestra. *Journal of Mathematics and Music*, 12(2):63–87.

Mannone, M. and Papageorgiou, D. (2020). Gestural Similarity, Mathematics, Psychology. Hints from a First Experiment and some Applications between Pedagogy and Research. In *Collected Papers of the Academy of Arts of Novi Sad*, volume 8, pages 137–153. University of Novi Sad, Novi Sad, Serbia.

Mannone, M., Santini, G., Adedoyin, E., and Cella, C. E. (2022). Color and Timbre Gestures: An Approach with Bicategories and Bigroupoids. *Mathematics*, 10(4).

Marentakis, G. (2020). Pictorize: Transforming Image Region Luminosity to Sound Brightness for Monitoring Location. In *SMC 2020*, pages 268–275, Torino.

Marks, L. E. (1974). On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness. *The American Journal of Psychology*, 87(1/2):173–188.

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3):384–394.

Mazzola, G. and Andreatta, M. (2007). Diagrams, Gestures and Formulae in Music. *Journal of Mathematics and Music*, 1(1):23–46.

McLachlan, J. and Basford, K. E. (1988). *Mixture models: inference and applications to clustering*. Marcel Dekker, New York.

Melnykov, V. and Maitra, R. (2010). Finite mixture models and model-based clustering. *Survival Analysis*, 4:80–116.

Michalos, A. X. (2005). Arts and the quality of life: an Exploratory study. *Social Indicators Research*, 71.

Nielsen, F. (2016). Hierarchical Clustering. In *Introduction to HPC with MPI for Data Science*, Part of the Undergraduate Topics in Computer Science book series (UTICS), pages 195–211. Springer, Cham.

Palmer, S. E., Schloss, K. B., Xu, Z., and Prado-Leon, L. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, 110:8836–8841.

Panek, D. W. and Stevens, S. S. (1966). Saturation of red: a prothetic continuum. *Perception & Psychophysics*, 1:59–66.

Parise, C. V. (2016). Crossmodal Correspondences: Standing Issues and Experimental Guidelines. *Multisensory Research*, 29(1-3):7–28.

Peacock, K. (1985). Synesthetic Perception: Alexander Scriabin's Color Hearing. *Music Perception*, 2(4):483–505.

Ramey, D. B. (1985). Nonparametric clustering techniques. In Kotz, S., Read, C. B., Balakrishnan, N., and Vidakovic, B., editors, *Encyclopedia of statistical science*, volume 6, pages 318–319. Wiley, New York.

Reuter, C., Jewanski, J., Saitis, C., Czedik-Eysenberg, I., Siddiq, S., Kruchten, S., and Oehler, M. (2018). Colors and Timbres: Consistency and Tendencies of Color-Timbre Mappings in non-synesthetic Individuals. Poster, Giessen University Annual Conference.

Reynolds, A. P., Richards, G., de la Iglesia, B., and Rayward-Smith, V. J. (2009). Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms*, 5:475–504.

Robert, E. (1975). *Sound Structure in Music.* University of California Press, Berkeley and Los Angeles.

Rosenblum, L. D., Dias, J. W., and Dorsi, J. (2016). The Supramodal Brain: Implications for Auditory Perception. *Journal of Cognitive Psychology*, 29(1):65–87.

Saha, A. and Das, S. (2015). Categorical fuzzy k-modes clustering with automated feature weight learning. *Neurocomputing*, 166:422–435.

Santini, G. (2019). Synesthesizer: Physical Modelling and Machine Learning for a Color-Based Synthesizer in Virtual Reality. In Montiel, M., Gomez-Martin, F., and Agustín-Aquino, O., editors, *Mathematics and Computation in Music, MCM 2019*, volume 11502 of *Lecture Notes in Computer Science*, Cham. Springer.

Spivak, D. (2014). *Category Theory for the Sciences.* MIT Press, Cambridge, Massachusetts.

Wilde, H., Knight, V., and Gillard, J. (2020). A novel initialisation based on hospital-resident assignment for the k-modes algorithm. arXiv, `arXiv:2002.02701`.