



HAL
open science

”I have to translate the colors” :description and implications of a genuine case of phoneme color synaesthesia

Lucie Bouvet, Cynthia Magnen, Clara Bled, Julien Tardieu, Nathalie Ehrlé

► To cite this version:

Lucie Bouvet, Cynthia Magnen, Clara Bled, Julien Tardieu, Nathalie Ehrlé. ”I have to translate the colors” :description and implications of a genuine case of phoneme color synaesthesia. 2022. hal-03880364v1

HAL Id: hal-03880364

<https://hal.science/hal-03880364v1>

Preprint submitted on 1 Dec 2022 (v1), last revised 19 Jul 2024 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

“I have to translate the colors”:

description and implications of a genuine case of phoneme color synaesthesia

Bouvet, L.¹⁺, Magnen C.²⁺, Bled. C.¹, Tardieu., J.², Ehrlé, N.³

¹Université de Toulouse, CERPPS, Toulouse, France

²Université de Toulouse, CNRS, Maison des Sciences de l'Homme de Toulouse, France

³Service de neurologie, CHRU Maison Blanche, Reims, France

⁺These authors contributed equally to this work

Corresponding author:

Lucie Bouvet, Université de Toulouse Jean Jaurés, CERPPS, 5 Allées Antonio Machado
31058 Toulouse Cedex, France. Lucie.bouvet@univ-tlse2.fr

Abstract

The aim of the paper was to document further the case of VA who possesses colored speech sound synesthesia.

To do so, different stimuli were presented to VA: isolated vowels, nonsense syllables, and words. VA systematically perceived a color for each phoneme and we observed that only the articulatory and acoustic factors linked with changes in phonemic category influenced the colors perceived. Semi-structured interviews indicated that in the context of conversation, VA had to translate color into phonemes into her constructed linguistic representations which has a cognitive cost. She also reported having a visual cognitive style in which the verbal dimension is almost non-existent.

This is the first time that genuine phoneme color synesthesia has been described. In the discussion, we address several implications of this synesthesia: the cognitive cost of this stable system, the atypical speech perception that it implies, and its relation to a specific cognitive functioning.

Keywords: phoneme color synesthesia, speech perception, visual cognitive style, veridical mapping

1.Introduction

Synesthesia is a subjective phenomenon in which the perception of a specific stimulus automatically triggers an idiosyncratic perception in the same or a different sensory modality. For example, a letter can trigger the perception of a specific color (grapheme-color synesthesia), a word can cause the sensation of flavor (word-taste synesthesia), or a sound can provoke the perception of complex colored photism (colored hearing synesthesia). The idiosyncratic synesthetic association is generally automatic, involuntary and stable over time (Ward, 2013). We previously reported, from a neuropsychological perspective, the case of VA who possesses a complex form of colored hearing synesthesia (Bouvet et al., 2017) in which each sound (music, speech, daily sounds) elicits the perception of a dynamic colored photism. When focusing on speech sounds, we noticed that is the phoneme unit which triggers the colored photism. This form of linguistic synesthesia (i.e., phoneme color synesthesia) has been very little described in the literature. The aim of this paper is to document VA's phoneme color synesthesia from a psycholinguistic perspective by evaluating the variations of her colored perception in different speech contexts (syllables, words, and conversation).

Among the 60 types of synesthesia that have been described, linguistic synesthesias are the most common form (Simner et al., 2006). Among them, grapheme-color synesthesia is the most frequent (1 – 2% of the general population) and the most widely studied one. The study of linguistic synesthesia has led to a better understanding of the implication of non-perceptual processes in this condition. Indeed, ambiguous graphemic stimuli (such as 5 which could be considered as the number five or the letter s) are associated to different colors depending on the context (Dixon et al., 2006). Moreover, in some cases, just thinking about the letter is sufficient to elicit a color perception (Mills et al., 2002). Similarly, visually distinct forms of the same grapheme (written in bold or in italics) can induce the same color (Smilek et al., 2002). Therefore, it is now established that grapheme-color synesthesia is triggered by the language unit per se and not solely by the visual form (Simner, 2012). Although orality is the primary dimension of language, relatively fewer studies have explored speech color synesthesia and questioned the implication of the different psycholinguistic dimensions (i.e., phonemic, lexical) in this form of synesthesia.

At the phonemic level, studies have mainly been carried out on vowels and results generally demonstrated that the acoustic properties of vowel sounds were significant predictors of color choices. In particular, brightness of the synesthetic color is linked to the

density of sounds: high-pitched vowels (/i/, /e/) are linked to brighter colors and low-pitched vowels (/o/, /u/) are linked to darker colors (Marks, 1975). Moreover, Moos et al. (2014) found that lower values of the first formant (F1) (i.e., higher vowels) were greener and yellower, but lower values of the second formant (F2) (i.e., more back vowels) were redder and bluer. In a recent study, Cuskley et al. (2019) replicated these earlier findings with synesthetes and nonsynesthetes, evidencing a cross-modal association between vowel and color in the general population. They also investigated whether vowel category was a better predictor of color choices than acoustic factors, as changes in acoustic structure are correlated with changes in vowel category. They showed the important role of categorical perception over and above acoustic cues for both populations (Cuskley et al., 2019). Their study demonstrates that acoustic and phonemic levels are linked and influence each other during speech perception.

At the lexical level, very few studies have been carried out on the question of the underlying processes in this form of synesthesia. Among them, Baron-Cohen et al. (1987) reported the case of a chromatic lexical synesthesia, EP, for whom hearing words triggered the perception of color. Her synesthesia was investigated by presenting different types of words (words of different semantic categories and weekdays) orally as well as letters of the alphabet. For each word, she provided a fine description of the color perceived. A potential link between color and semantic or phonological level was tested by looking at the colors associated to words that were semantically close, i.e. words that belonged to the same semantic field (e.g. *man*, *masculine*, *male*) or words that were phonologically close, i.e. words with homophones that vary only by one phonemic unit (e.g. [mœn], [mu:n], [mœʊn]). EP did not associate the same color or similar colors to groups of words that were related either on the phonological or the semantic dimension. The authors concluded that there is no semantic or phonological link between words and colors even though a direct link between phonemic or lexical units and synesthetic colors was not tested. Moreover, they observed that for words, EP provided a single color whereas for nonwords she provided several colors that could be linked to the colors triggered by the letters of the alphabet. One can note that this conclusion concerns only the graphemic level whereas the association between color and phonemes was not tested. In another study, Baron-Cohen et al. (1993) tested if the synesthetic perception of 9 subjects with colored hearing synesthesia was mediated by a phonemic or rather by a graphemic representation. To do so, they presented a list of words orally that included homophones ([sʌn] ‘sun/son’; [beə] ‘bear/bare’; [weə] ‘where/wear’), words with shared initial phonemes but with different initial graphemes (e.g., [nɒk] / [naɪs] ‘knock/nice’), and

nonwords. Participants had to give a verbal description of the color that was triggered. For all subjects, the synesthetic color was triggered by the initial letter of the word, not by the phoneme. The authors concluded that the participants possessed grapheme-color rather than phoneme-color synesthesia.

These results are emblematic of most studies that have focused on linguistic synesthesia, which have demonstrated that it is mainly induced at the graphemic level (Simner, 2007). When examining the experimental settings of these studies, however, one notes a recurrent confusion due to the methodological difficulty of working with the oral dimension of language. In many cases, when participants are asked to associate a color to a linguistic sound, the experimenter uses the graphemic representation as in the following example: “*what color is associated to the letter “b” in the word ‘balance’?*” In fact, it is impossible to produce a consonant sound without the support of a vowel. From an experimental point of view, this implies that these consonants are presented either in a syllable, or in a word or graphemically. The disadvantage of such methods is that they summon several types of representations which are not directly those that are targeted. Recent studies on synesthesia started to highlight the relations between different linguistic representations (acoustic, phonemic and graphemic) and synesthetic mappings (Cuskley et al., 2019). For instance, Kang et al. (2017) showed that graphemes sharing a phonetic feature tended to induce similar synesthetic colors across languages. Thus, to evaluate as accurately as possible the oral dimension of linguistic synesthesia, it seems necessary to investigate the different linguistic representational levels separately and, at the same time, to consider their mutual influences.

In our previous paper, we documented a case of phoneme color synesthesia (Bouvet et al., 2017). We demonstrated the consistency of VA’s synesthetic associations between phonemes and colors by evaluating the phonemic level in the context of a word. Speech sounds were pronounced by the experimenter in a sentence form “the sound X as in the word Y” where X is the oral pronunciation of a grapheme for consonants and oral vowels or a syllable which refers to the nasal vowels (e.g., “an” [ã]). The objective of the present paper was to investigate whether VA’s synesthesia varied as a function of stimuli types and different tasks. To this end, we developed three tasks which involve different types of linguistic units: isolated vowels, nonsense consonant-vowel (CV) syllables, and words. In all tasks, we used natural voices in order to approximate “real world” conditions in our experimental tasks. The

first task aimed to observe whether a natural variation of speech (voice pitch) modulated VA's synesthetic perception for isolated vowels. Moreover, in this task, we also manipulated syllables to assess whether VA continued to provide a specific color for each phoneme or whether she provided a color for the syllable as a unit. Nonsense CV syllables were used in order to assess the influence of coarticulation. The aim of the second task was to observe whether the semantic context modulated VA's synesthetic perception. The experimental settings enabled us to determine whether the different graphemic representations of the same phoneme influenced her synesthetic perception. Lastly, we interviewed VA on how her synesthesia manifested itself in a conversational context.

2. Case description

VA, a young woman, was 20 years old at the start of the study. When we first met her, she was addressed for a neuropsychological assessment with the diagnosis of social phobia and schizophrenia. After conducting a neuropsychological assessment and cognitive evaluation, it appeared that VA possessed a complex colored hearing synesthesia which at first was considered as schizophrenia, in which each sound (music, language, daily sounds) elicited the perception of a dynamic photism. Moreover, she also possessed savant abilities for astrophysics as well as higher-than-average cognitive abilities (for more details see Bouvet et al. 2017). After this first study, she agreed to continue the research with the aim of better understanding her synesthesia and her cognitive functioning.

3. Testing phoneme color synesthesia for vowels and for nonsense CV syllables

3.1. Material and methods

The linguistic material consisted of audio recordings of French vowel and consonant phonemes. The recordings were performed in the sound perception lab of the CCU platform¹ in an audiometric booth using a Sennheiser MD46 omnidirectional microphone, a TASCAM DM-3200 mixing console and a MacPro computer equipped with the Reaper software. Four adult native speakers of French were recruited to record speech stimuli: two men and two women with two different high and low pitches. The vowel phonemes [i, e, ε, a, u, o, ɔ, y, ø, œ, ê, ã, ð] were pronounced in the carrier sentence « *Je dis /V/ comme dans ...* » [*I pronounce /V/ as in ...*]. The target vowels were written down with their IPA transcription as well as the

¹ <https://mshs.univ-toulouse.fr/plateformes/les-plateformes-technologiques/ccu-cognition-comportements-et-usages/petra/>

example word, which appeared at the end of the sentence orthographically. The speakers were checked for their understanding of the symbols before recording. They were asked to read the carrier sentences one by one, twice in a row, for a total of 52 stimuli (13 vowels* 4 speakers). Target vowels were then extracted with Audacity® from the audio recordings. The consonant phonemes [p, t, k, b, d, g, m, n, f, s, ʃ, v, z, ʒ, ʁ, l] were produced three times in a CV context where the vowels were [i, a, u], for a total of 192 stimuli (16 consonants* 3 contexts* 4 speakers). The stimuli were listened to by a phonetician (CM) and selected according to their perceptual typicality. For each phoneme, we selected one exemplar produced by a male and one by a female speaker. Out of the 244 stimuli created, half were selected. In total, each vowel exemplar was pronounced twice (2 speakers) and each consonant exemplar was pronounced 6 times (2 speakers * 3 contexts).

The 122 stimuli were orally presented randomly using the Psychopy software (Peirce, 2007). After hearing each sound with the headphone (Sennheiser HD 380 pro), VA had to indicate the corresponding color (or colors) closest to her perception with the color chart (MACal 9800 pro© which contains 102 colors). The experimenter noted down the color reference. The sound was played only once. VA was told that she could take as much time as she needed to give her answer.

3.2. Analysis

Color dimension. For the analyses, each color was converted to CIELab space (the correspondence was given directly by MACal ©). It expresses color using three different values. *L* is the lightness that goes from 0 (darkest black) to 100 (brightest white). The color channels are *a* and *b*. The *a* axis represents the green/red colors with green at negative values and red at positive values. The *b* axis represents blue/yellow colors with blue at negative values and yellow at positive values. To observe a potential influence of speaker gender on color, T-student statistics were performed between the speaker's gender and the Lab values for vowels solely.

Formants analysis. For isolated vowels, formants were extracted and analyzed using Praat© software (see Figure 1). Formants are independent of voice pitch and are more characteristic of the type of vowels. Formant values vary with tongue position in vowel production. F1 values are lower when the tongue is higher in the mouth (i.e., high vowels such as /i/) and higher when the tongue is lower in the mouth (i.e., low vowels such as /a/). F2 values vary

according to tongue frontness. To evaluate a potential link between formants (F1 & F2) and colors, Pearson correlations were performed between formants and Lab values.

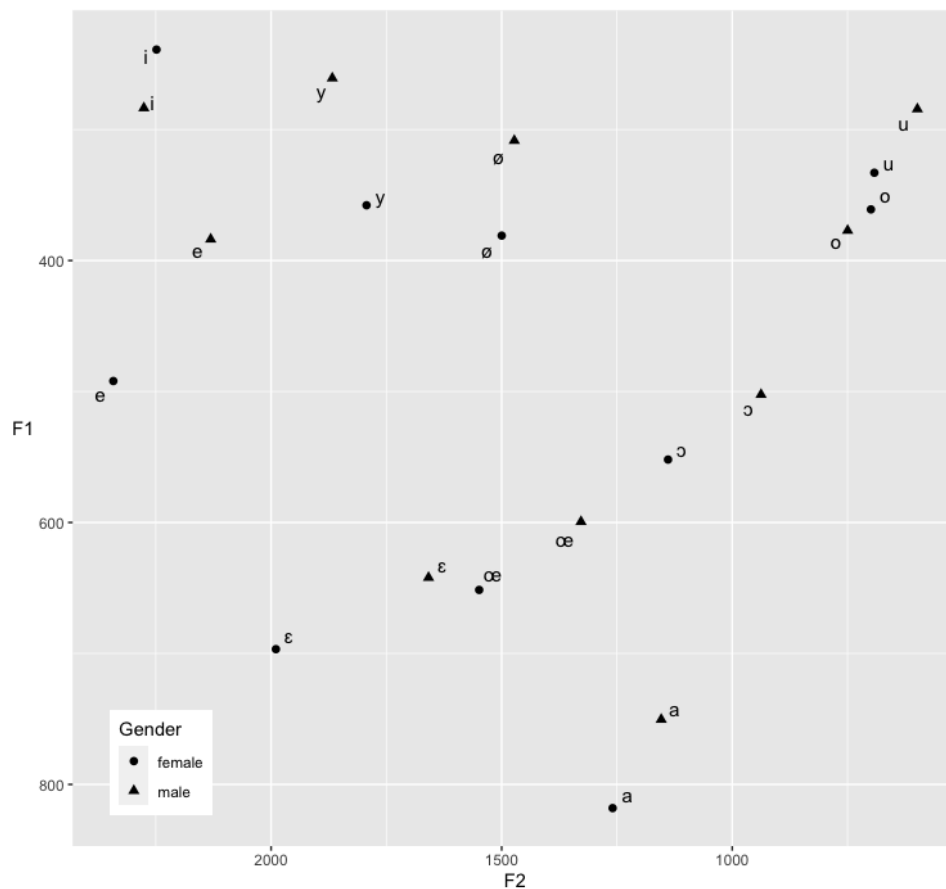


Figure 1. Representation of F1 and F2 values of vowels. Triangles represent vowels produced by male speakers and points represent vowels produced by female speakers.

Consonants. Consonants can be defined by their articulatory mode, either occlusive [p, b, k, g, t, d, m, n] or constrictive [ʃ, f, ʒ, l, ʋ, v, s, z]), and the place of their articulation (bilabial [p, b, m], labiodental [f, v], dental [n, d, t, s, l, z], dorso palatal [ʒ, ʃ] and velar [g, k, ʋ]). The articulatory mode defines how the sound is produced by blocking the airstream. For occlusives, the airstream is momentarily blocked whereas for constrictives, the air passage is narrowed by constriction of the vocal tract. The place of articulation refers to the place along the vocal tract where the articulators (tongue, teeth, lips or glottis) make contact.

To test a possible link between the articulatory mode (occlusive vs constrictive) and the chosen colors, as Lab values do not follow the assumption of normality (Shapiro-Wilk p values < .001), we performed a non parametric test (Mann-Whitney U test). To test if there

was a link between the place of articulation of the consonants and the chosen colors, as the place of articulation is an ordinal variable, we performed Spearman's rank correlation.

3.3. Results

As previously reported (Bouvet et al., 2017), VA provided a specific color for each phoneme. Consonants and vowels were analyzed separately.

Isolated vowels. For 8 of the 13 vowel categories (see Figure 2), VA choose exactly the same color for the same vowel pronounced by the two different speakers [i, e, a, o, ɔ, ø, ã, õ] (61.5% of the vowels). For the 3 vowel categories /y/, /œ/ and /ẽ/, the colors attributed to the different exemplars were in the same hue. For the two vowel categories /ɛ/ and /u/, the colors attributed to the different exemplars were of different hues. No influence of the speaker gender was observed on the Lab values (all $p > .6$). Regarding the influence of formants, we observed a significant negative correlation between F1 and a values ($r = -0.692, p < .001$), which indicates that the lower the F1 values are, the redder the colors. A significant negative correlation between F1 and b values ($r = -6.48, p < .005$) was also observed, indicating that the lower the values on F1, the yellower the colors. A significant correlation between F2 and L values ($r = .723, p < .001$) was also observed, with vowels with higher F2 values being associated to lighter colors.

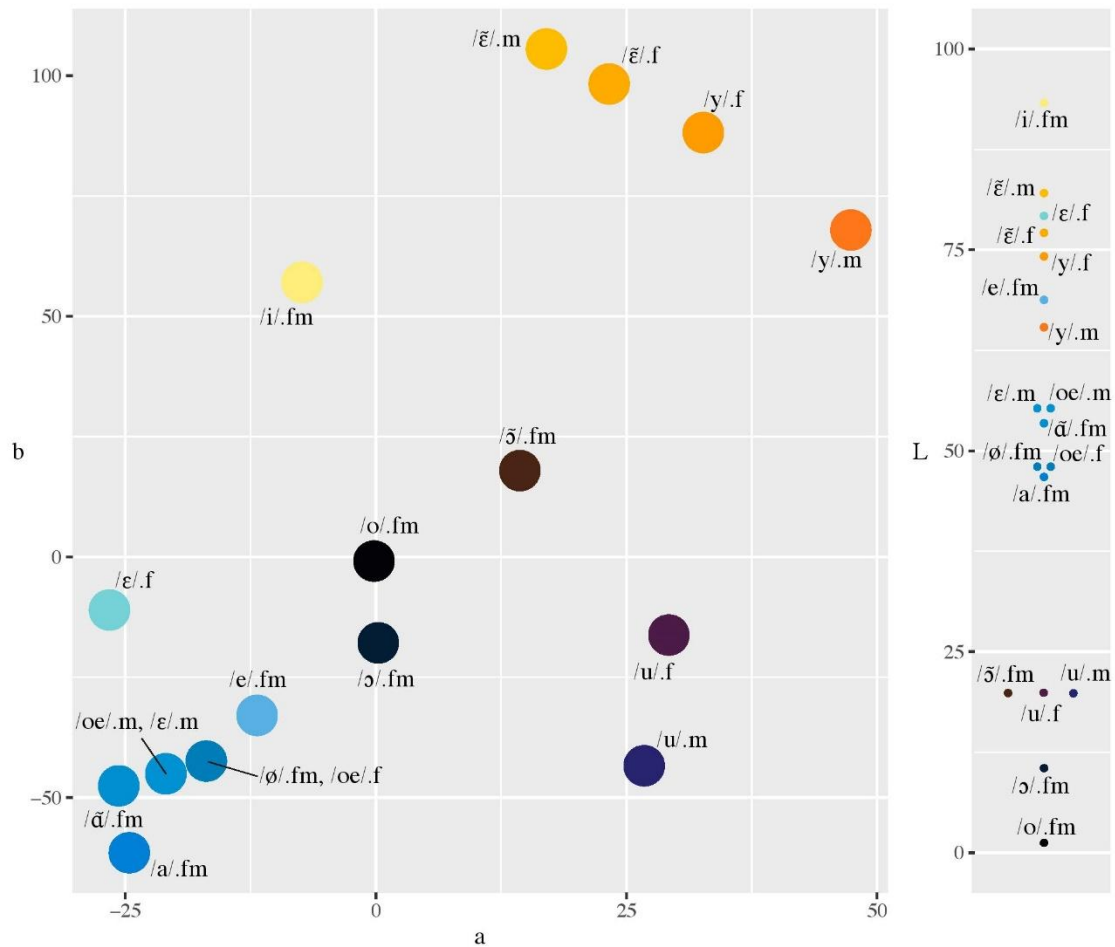


Figure 2. Representation of colors chosen by VA in the Lab color system for vowels (2 trials). *f* indicates vowels pronounced by a female speaker and *m* indicates vowels pronounced by a male speaker.

Consonants and vowels in syllabic context.

For each CV syllable, VA provided two different colors. Thus, vowels and consonants were analysed separately.

Vowels in syllabic context. We analyzed colors from the 3 different consonant contexts [a, y, u] (see Figure 3). There were 32 exemplars for each phoneme consonant context. For [a] three different colors were given: luminous cyan² (20 times), sky blue (11 times), azur blue (1 times), luminous cyan being the color given for [a] presented alone. The speaker gender does not explain this difference of color ($p > .1$). For [i], the same color was given for the 32 exemplars: light yellow. It is the same color as for [i] alone. For [u], the same color (eggplant) was given for 31 exemplars, and the color dark violet was given for the other exemplar. These

² The names of the colors are given by the MACal color chart.

two colors were given for the two exemplars [u] presented alone. Thus, no variation in color with the consonant context was observed.

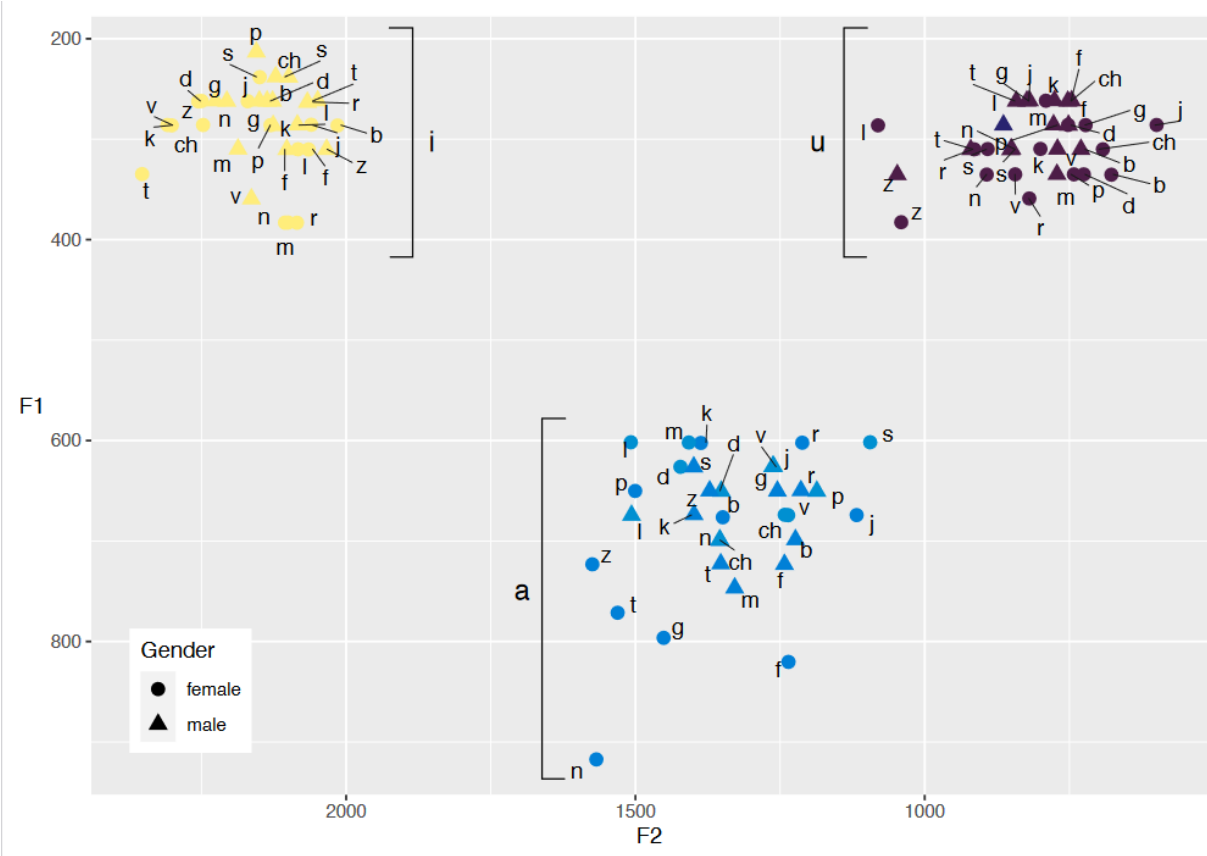


Figure 3. Representation of colors chosen by VA for [a], [i], [u] in different consonant contexts according to the F1 and F2 values.

Consonants. For 12 consonants ([p, k, b, d, g, n, f, s, ʃ, v, ʒ, ʁ], i.e., 75% of consonants), VA chose exactly the same color for the six exemplars. For the four remaining consonants [m, l, t, z], a few differences in colors were observed especially for the *b* and the lightness dimensions within the consonant category. As can be seen in Figure 4, these differences in color are minimal. Consequently, no effect of vocalic contexts and the speaker's gender was observed on the color chosen for consonant categories.

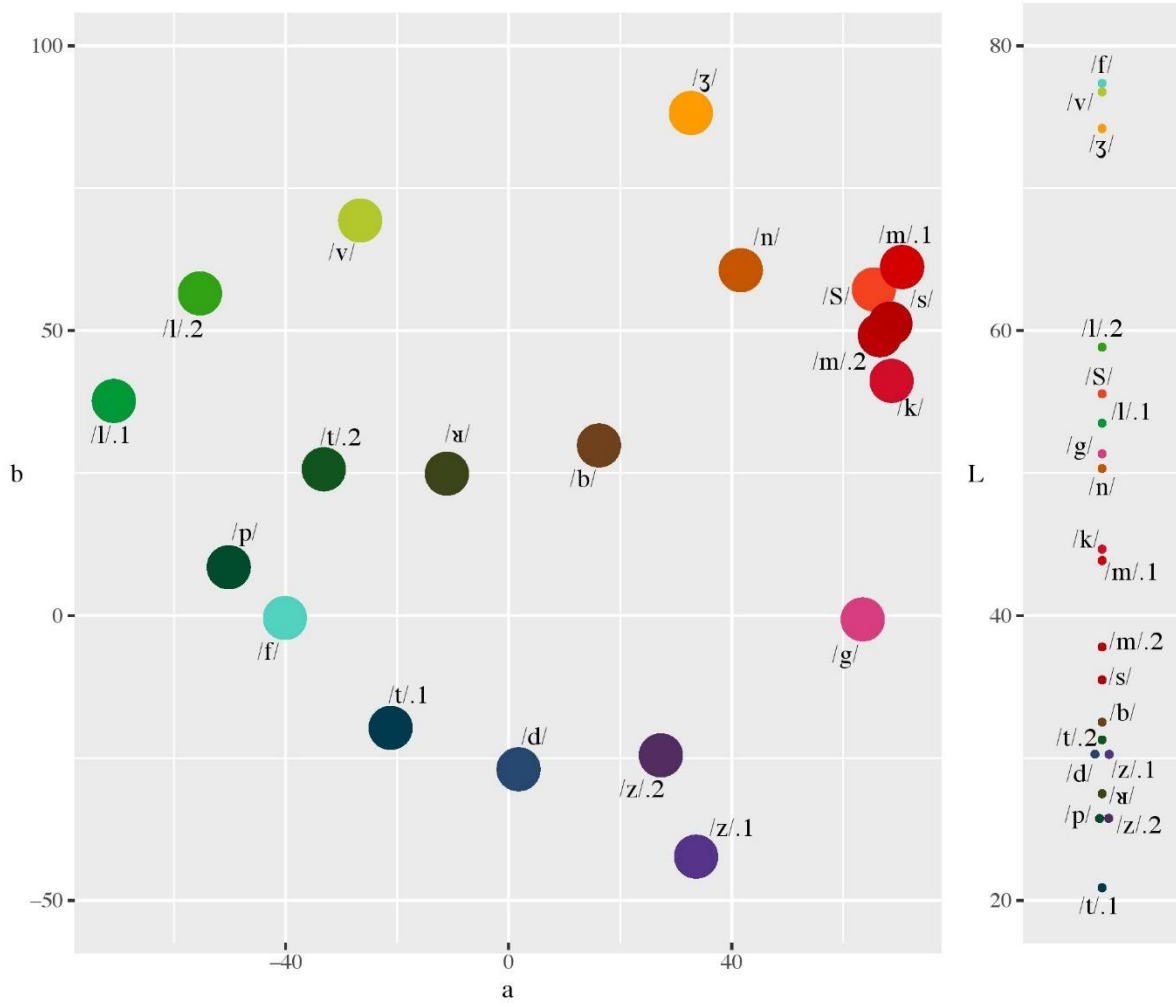


Figure 4. Representation of colors chosen by VA for consonants (6 trials) as a function of the Lab color system. Numbers indicate the two different colors chosen for the six versions of the consonants.

Regarding the link between colors and articulatory mode, colors associated to constrictive consonants were lighter ($U = 321, p < .001$) and yellower ($U = 854, p < .05$) than colors associated to occlusive consonants. No difference on the a value was observed ($U = 906, p = .07$) (see Figure 5).

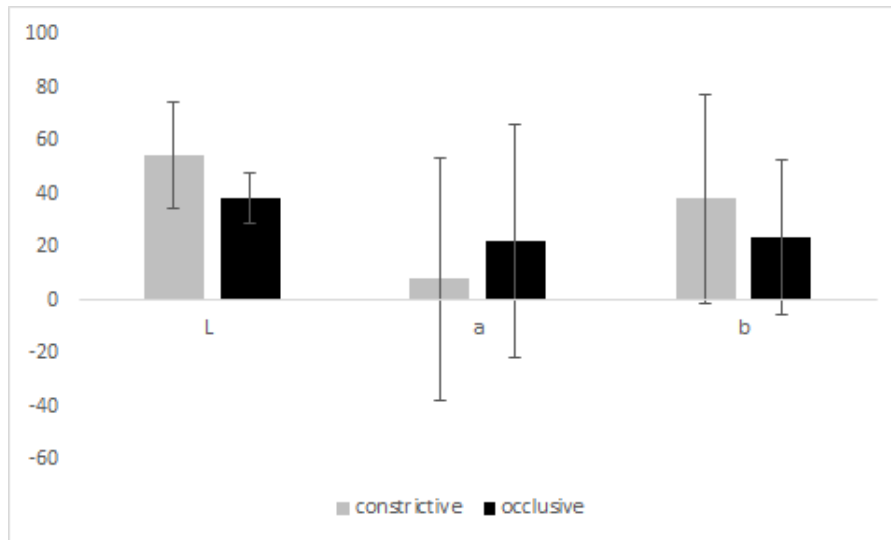


Figure 5. Mean Lab values and standard deviation for constrictive and occlusive consonants.

Regarding the link between colors and the place of articulation, we observed a significant correlation with *a* value. The more posterior the place of articulation of the consonants is, the redder they are ($r = .033, p < .001$) (see Figure 6). Correlations with other Lab values (*L* and *b*) were not significant ($p > .5$). Thus, there is a link between synesthetic colors and the articulatory mode as well as the place of articulation of consonant phonemes.

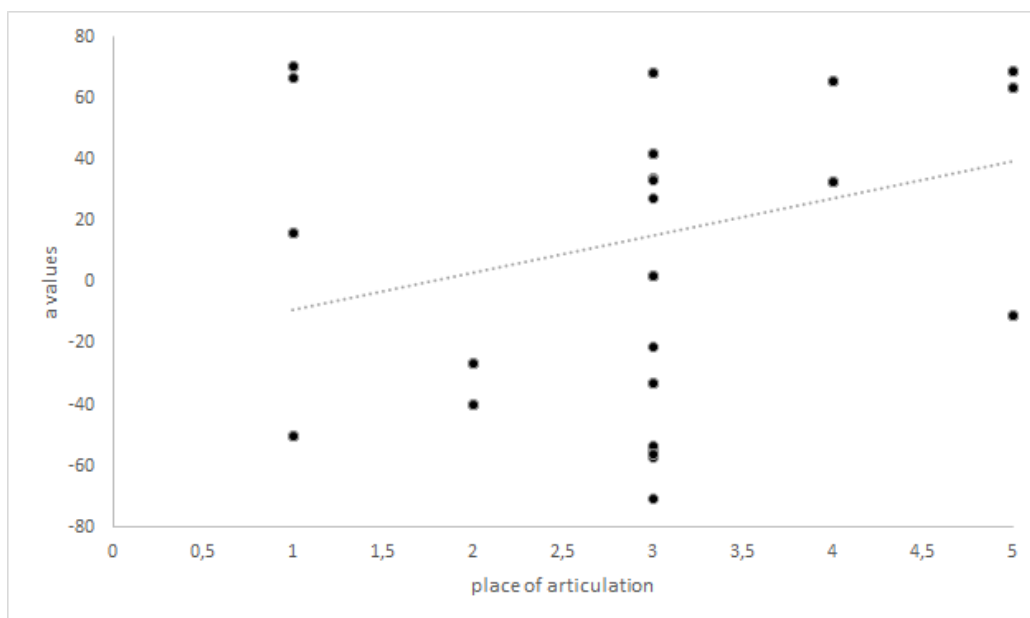


Figure 6. Correlation between place of articulation and *a* value

4. Testing phoneme color synesthesia in the context of words

4.1. Material and methods

We selected 60 monosyllabic words composed of phonemes that elicit different associations of the 6 color categories: green and blue, green and yellow, red and blue, green and orange, orange and blue, and with 3 colors. Each color category comprised 10 words. The two-color categories comprised 5 bi-phonemic words with the CV form and 5 tri-phonemic words with the Consonant/Vowel/Consonant (CVC) form in order to balance the length of the words in each group. The three-color categories of words were exclusively composed of 10 CVC words (see supplementary information for more details). Among the selected words, there were different graphemic representations for 5 vowel phonemes and for 3 consonant phonemes (see Table A.1 in Appendix).

The words' frequencies of use were verified in different databases (MANULEX and LEXIQUE). Each word was recorded by a female and a male speaker (for a total of 120 words, 60 monosyllabic words* 2 speakers) and was produced with a definite article "le/la" (*the*). The recordings were performed in the same conditions as previously.

The words were presented only once via headphones (Sennheiser HD380 pro) using the PsychoPy® software. The order of presentation was randomized. For each word, VA had to indicate the corresponding color (or colors) closest to her perception with the color chart (MACal 9800 pro® which contains 102 colors). VA was instructed to ignore the definite article for her color choices. The corresponding codes were noted down by the experimenter.

4.2. Analysis

Color dimension. As for isolated vowels and CV syllables, each color was converted to CIELab space (the correspondence was given directly by MACal®) in order to perform analysis. It expresses color using three different values. *L* is the lightness that goes from 0 (darkest black) to 100 (brightest white). The color channels are *a* and *b*. The *a* axis represents the green/red colors, with green at negative values and red at positive values. The *b* axis represents blue/yellow colors, with blue at negative values and yellow at positive values. To observe a potential influence of speaker gender on color, T-student statistics were performed between the speaker's gender and the Lab values

Euclidean distance. To observe a potential influence of speaker gender on color, T-student statistics were performed between the speaker gender and Lab colors. To evaluate the role of the semantics on VA's color choices, we compared the colors given for phonemes presented alone (vowels and nonsense CV syllables, previous experiment) and the colors given for the phonemes in words. To do so, we calculated the Euclidean distance (d) between the mean of the Lab values for each main phoneme category across the two different tasks according to the equation:

$$d = \sqrt{(L1 - L2)^2 + (a1 - a2)^2 + (b1 - b2)^2}$$

The score reflects how far away the two colors are in the color space. The lower the score, the closer the colors are in this color space. According to Rothen et al. (2013), a score below 109 is an indicator of consistency.

4.3. Results

Influence of semantic representation

As in the previous task, VA attributed a specific color to each phoneme within the words. For the bi-phonemic words, she gave 2 colors and for the tri-phonemic words she gave 3 colors. Otherwise, the colors of the phonemes did not influence each other since, despite the different phonemic contexts, the triggered colors remained the same within 12 out of 28 phoneme categories (42%). When the colors within the same phonemic category were different, the colors selected were in the same hue. These color differences were not explained by the speaker gender as no difference between the female and male versions was found in the Lab values. No influence of the speaker gender was observed on the Lab values (all $p > .1$).

In order to explore a possible influence of the semantic dimension on colors, we compared the colors provided for isolated phonemes and the colors provided for phonemes in words by computing the Euclidean distance (d). For 9 phoneme categories, the d value was 0, indicating that VA chose exactly the same color for the phoneme in the two contexts (meaningless syllables and words). For 18 phoneme categories, the d value was below 109. For 1 phoneme category (/ã/, $d = 153$), the color was completely different between the two tasks (see Figure 7). Consequently, no influence of the semantics was observed on VA's phoneme color synesthesia.

























































	Isolated phoneme color	Phoneme in word color	<i>d</i> value		Isolated phoneme color	Phoneme in word color	<i>d</i> value
/a/			0	/k/			9.4
/ɔ/			0	/l/			20.8
/b/			29.8	/m/			8.1
/ʒ/			20.1	/n/			0
/d/			17.5	/o/			3.1
/e/			0	/u/			47.4
/ɜ/			16.8	/p/			0
/ø/			14	/ʁ/			0
/œ/			41.8	/s/			5.3
/f/			0	/t/			8
/g/			56.9	/ʎ/			5.9
/i/			0	/v/			0
/ʒ/			17.4	/ã/			153.2
/ẽ/			5.4	/õ/			6.63

Figure 7. Mean colors for phonemes when presented alone (experiment 1) or inside words (experiment 2). The *d* value indicates the Euclidean distance between the mean Lab values between the two tasks. The lower the value, the closer the two colors are.

Influence of graphemic representation

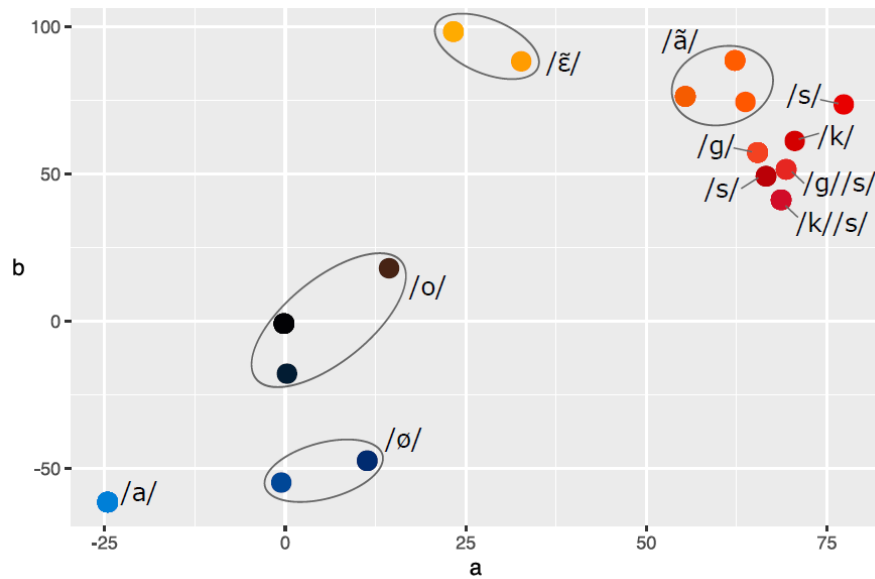


Figure 8. Representations of colors selected for phonemes with different graphemic representations on the a and b values of the Lab system.

It can be seen that for phonemes for which there are different graphemic representations (with two or three letters), VA provided a single color. Moreover, for each phoneme, there were very slight color variations which are not linked to the graphemic variations (see Figure 8). Consequently, no effect of the graphemic representation was observed on VA's synesthesia.

5. Phoneme color synesthesia in the context of conversation

5.1. Method

Two semi-directed interviews were conducted with VA by the author CB. The purpose of these interviews was to better understand the functioning of VA's synesthesia and how her synesthesia manifested itself in daily situations (such as conversation). Part of the interview consisted in having VA listen to daily sounds (we asked her to describe what she was feeling). The interviews were conducted at the hospital of Reims in a quiet room. Each interview lasted for approximately 30-45 minutes. The interviews were audio-recorded and the verbatim of the interview was transcribed. From these interviews, we selected the parts that addressed the question of her synesthesia in the context of conversation.

In order to perform a thematic analysis, the transcribed interview data were analysed by the author LB. The transcripts were read and reread by LB and an initial set of themes were

produced. The themes were reviewed by CB and CM, evaluating how well they captured the coded data. Here, we present a free translation of her statements.

5.2.Results

A complex form of colored hearing synesthesia.

VA described her synesthesia as a more complex phenomenon than the simple perception of colors. Colors could have different shapes depending on the acoustic parameters: “*Color has a shape.*” “*In the visual field one feels the vibrations, it makes like mini wavelets like very closely-spaced vertical lines.*” “*The frequency is flatter here, it changes less so one has the impression that the sound is flatter, it has less shape.*” “*The sound [...] is rather something reddish, one has the impression of a squarish shape.*” As sound is a physical phenomenon that evolves in time and space, so does her synesthesia: “*The way sounds evolve corresponds to movements, the colors move in time and in space.*” “*The sound moves and the color follows the sound.*” “*It is likely to change according to the change of frequency, volume, timbre and also the time that it lasts.*” VA also has a 360-degree synesthetic space: “*There is space around me 360 degrees and I can see all around me in synesthesia, I can see things behind me*”. Finally, VA explained that her synesthesia can also imply the kinesthetic modality: “*The higher the volume is, the more the colors will invade the visual field and the closer they come to me... in some cases they can touch me.*”

Processes of translation into speech perception and production

To access the meaning of what is being said, VA described two distinct processes. First of all, the correspondence between sound and color is automatic and these two dimensions (sound and color) are closely linked. For VA, colors are meaningful only in relation with sounds. “*Colors only have meaning if they are associated with sounds*” “*Colors, for me, have no meaning at all, they are not associated with emotions.*” The automaticity of this association is unidirectional: “*I can possibly remember the sound associated with the color but from memory that's all... it's not spontaneous at all! The association only goes one way.*”

VA described a multisensory way of thinking, mostly dominated by the visual modality: “*If I think of a cake, I won't think of the word cake: I will think of all the sensations I have when I see, touch and eat the cake. All the senses are suddenly activated.*” “*I think in video constantly, in pictures.*” VA explained that she does not think in words: “*I don't think in*

words.” “I do not remember the words, I see the meaning of what I hear: it is more than an image because there are all five senses.” Consequently, when she has to produce speech, i.e. when she speaks, she has to translate this multisensory way of thinking into words. “[when I think] all the senses are suddenly activated. So, I have to translate them back into words when I speak”. When she is the receiver, i.e., when she has to perceive and to understand speech, another type of translation process occurs. To access the meaning, she has to translate the perceived colors. Contrary to the automatic correspondence between sound and color, the translation process between color and language is not straightforward: “The meaning of words is a work of translation in my head and this translation work requires a much greater effort than being able to look at the colors. I have to translate the colors.” “If I want to be able to understand what people are saying when they speak, I also have to focus on synesthesia because otherwise I don't understand the words.” These processes of translation made it difficult for VA to interact with others: “It's hard to understand what I hear and to express myself.”

Sensory overwhelm and communication difficulties

Because of her particular way of processing speech (processes of translation, hypersensitivity), holding a conversation requires a great deal of effort. She feels overwhelmed by all the information and that causes her to lose track of the conversation, especially when the interaction lasts a long time and the speech flow is fast. “For example, when [someone] explained to me how to fill out my papers, and then she went from one subject to another, and she made a long long speech and then I couldn't hold on to the end, because there were too many colors” “After a while, I switch off because I end up focusing on colors and voice rather than focusing on the meaning of words”. “I have a lot of trouble holding a conversation for a long time, if it lasts it's a terrible effort.” “When the perceptions are really strong, for example [...], when there is someone who speaks quickly and then for a very long time, it is completely exhausting.”

A visual cognitive style

VA explained that since she processes information in a sensory/visual way, she has difficulties apprehending and evolving in a world that she described as mostly verbal: “For most people, the world is partly verbal, for me it's a foreign language.” “Teaching methods are extremely verbal and I am not. It requires a huge verbal memory and I would like to learn

more visually.” “I don't have the same perception of things as most people so that suddenly in terms of studies and work, it doesn't work.”

6. Discussion

In the present study, we evaluated the consistency of VA's synesthesia with different types of stimuli (isolated phonemes, syllables, and words). The aim was to investigate whether acoustic or semantic factors influenced VA's synesthesia. First of all, across these different types of stimuli, we systematically observed that VA provided one specific color for each phoneme. We also observed that synesthetic colors were determined by some factors linked to phonemic categories. For vowels, we observed a correlation between formants (F1 & F2) and color values. For VA, higher vowels are redder and yellower, while lower vowels are greener and bluer. Moreover, front vowels are lighter. These results are partially in line with those observed in synesthetes (Cuskley et al., 2019; Moos et al., 2014) but also in the general population (Cuskley et al., 2019). This indicates that VA's synesthesia for vowels fits into the framework of general cross-modal mapping. For consonants, we observed a correlation between colors and the articulatory mode and the place of articulation. Constrictive consonants are associated to lighter and yellower colors than occlusive consonants and back consonants are associated to a redder color. Thus, synesthetic colors are linked to articulatory and acoustic factors that are correlated with changes in phonemic category. This interaction between acoustic and phonemic levels during speech perception has already been described (Cuskley et al., 2019). On the other hand, the perceived colors were not influenced by voice pitch, nor by phenomena of intrasyllabic coarticulation, which are acoustic factors that are unrelated to phonemic categories.

In the context of words, contrary to what might have been expected, we did not observe any effect of semantic knowledge on her synesthetic perception. She continued to provide one color for each phoneme that composed the word and no influence of the graphemic representations was observed. Moreover, the colors triggered for isolated phonemes were similar to the colors triggered for phonemes within words. Taken together these results indicate that her synesthesia manifests itself at the level of the phoneme categories. To our knowledge, it is the first time that a genuine phoneme color synesthesia has been evidenced. Previous studies on this topic have always observed that the synesthesia was influenced either by semantic factors (Baron-Cohen et al., 1993) or acoustic factors inherent to the speaker (Fernay et al., 2012).

In the light of this atypical synesthesia and consequently this atypical way of processing speech, we conducted a thematic analysis of the interviews in order to apprehend the manifestation of VA's synesthesia in the context of everyday conversation. First of all, VA described her synesthesia as a dynamic spatialized colored phenomenon. The dynamic aspect of her synesthesia is linked to acoustic variations of sound. Second, in the context of conversation, she automatically perceives colors that she has to translate into her constructed linguistic representations. Paradoxically, this process of translation does not seem automatic and has a high cognitive cost. Thus, holding a conversation is a difficult exercise for her as it generates a multitude of synesthetic stimuli. These overflowing sensations are intrusive and make her lose the thread of the speech. Moreover, her semantic representation appears to be predominantly visual. This causes her difficulties apprehending and evolving in a world that is mostly verbal.

Considered altogether, these data highlight two contrasted aspects of her synesthesia. The colors associated to the phonemic representations constitute a very stable system, whereas the acoustic parameters of sounds influence different aspects of the synesthesia that are linked to the color such as shapes, localization in space, or kinesthetic sensations. The experience of localized geometric objects in auditory-visual synesthesia has been previously documented (Chiou et al., 2013) but this is the first time that it has been observed in the context of speech and daily communication. The fact that acoustic factors activate dynamic visual representations provides access to a rich speech perception in which all factors are perceived at the same level. Usually, listeners select among speech variations the most reliable information for the recognition of phonological sequences and meaning. This selection is an automatic process without conscious awareness (Johnson & Sjerps, 2021; Strange, 2011). For VA, this automatic process seems to be challenged by the profusion of information to which she has access. This could explain the invasive dimension of her synesthesia.

The nature of her segmental speech perception to access word meaning raises several questions. During the 80s, this word recognition process was the cornerstone of lexicon access models (see for example the TRACE model, Elman & McClelland, 1984; McClelland & Elman, 1986). More recently, these models have been judged to have too high a cognitive cost (Wauquier-Gravelines, 1999) because they presuppose remembering each phoneme in the time course of the speech in order to access the meaning. They were also deemed to be too close to an orthographic vision of speech perception (Grosjean & Gee, 1987) and not to account for the dynamic nature of speech processing. Nowadays, listeners are considered as

linguistic experts who can use different-sized units (phonemes, syllables, words) to match the speech signal to units of meaning in order to understand conversational interactions (for a review see Nguyen et al., 2009). These results highlight that VA's synesthesia is unique in that it involves a high cognitive cost and belongs to particular type of cognitive functioning.

One question that emerges is why VA developed this segmental processing of speech. We can hypothesize that VA developed a strategy to organize the excess of visual information on the sole basis of the phonemic representation system. A similar extremely atypical strategy to organize information has been described in autism and in the acquisition of some specific abilities (synesthesia, hyperlexia, absolute pitch) via the mechanism of veridical mapping (Mottron et al., 2013). This mechanism allow the enhanced detection and memorization of isomorphisms between perceptual and non-perceptual structures (for a concrete example see Bouvet et al., 2014). It is interesting to note that VA learnt to read alone and quite early (at 3 years old; see Bouvet et al., 2017). Her segmental method of processing speech is reminiscent of the learning process of reading when children decipher words grapheme by grapheme, unit by unit. Thus, her early reading skill might have influenced this segmental speech processing and might have helped her to detect regularities between oral linguistic units and visual representations.

It has been recently pointed out that synesthesia is one manifestation of a particular neurodiverse phenotype (Ward, 2021) and that the exploration of the mechanism underlying synesthesia can reveal a particular cognitive profile. A cognitive style can be defined as the heuristic used by an individual to process information about his or her environment (Kozhevnikov, 2007). In the interviews, VA described a visual cognitive style when she explained that she thought solely in images. In the previous study on VA (Bouvet et al., 2017), we hypothesized that her vivid imagery was linked to the development of her synesthesia. With this new study, we can go one step further and hypothesize that a visual cognitive style can be linked to the possession of synesthesia. The presence of visual cognitive style has been described in certain conditions, notably in the autism spectrum (Bled et al., 2021; Kunda & Goel, 2011). Interestingly, a higher prevalence of synesthesia has also been described in this condition (Baron-Cohen et al., 2013; Neufeld et al., 2013). Thus, it is possible that some psychopathological conditions can be related to a particular cognitive functioning and that synesthesia is a manifestation of this cognitive functioning. Further research might explore this issue.

To conclude, using different methodologies we have evidenced for the first time a genuine phoneme color synesthesia. We have also discussed the fact that this synesthesia

might be unique as it reflects a particular way (i.e., segmental) of processing speech and might emerge from a sensory/visual way of processing information. Further research could consider the interest of combining different types of methodologies intra-individually in order to better apprehend some complex forms of perception.

Acknowledgements

We are greatly indebted to VA, who consented to participate to this research.

Funding

This work was supported by the MSHS-T (Maison des sciences de l'homme et de la société de Toulouse)

References

- Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured speech perception : Is synaesthesia what happens when modularity breaks down? *Perception*, 22(4), 419- 426. <https://doi.org/10.1068/p220419>
- Baron-Cohen, S., Johnson, D., Asher, J., Wheelwright, S., Fisher, S., Gregersen, P., & Allison, C. (2013). Is synaesthesia more common in autism? *Molecular Autism*, 4(1), 1- 6. <https://doi.org/10.1186/2040-2392-4-40>
- Baron-Cohen, S., Wyke, M. A., & Binnie, C. (1987). Hearing words and seeing colours : An experimental investigation of a case of synaesthesia. *Perception*, 16(6), 761- 767. <https://doi.org/10.1068/p160761>
- Bled, C., Guillon, Q., Soulières, I., & Bouvet, L. (2021). Thinking in pictures in everyday life situations among autistic adults. *PLOS ONE*, 16(7), e0255039. <https://doi.org/10.1371/journal.pone.0255039>
- Bouvet, L., Barbier, J.-E., Cason, N., Bakchine, S., & Ehrlé, N. (2017). When synesthesia and savant abilities are mistaken for hallucinations and delusions : Contribution of a cognitive approach for their differential diagnosis. *The Clinical Neuropsychologist*, 31(8), 1459- 1473. <https://doi.org/10.1080/13854046.2017.1288269>

- Bouvet, L., Donnadieu, S., Valdois, S., Caron, C., Dawson, M., & Mottron, L. (2014). Veridical mapping in savant abilities, absolute pitch, and synesthesia : An autism case study. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00106>
- Chiou, R., Stelter, M., & Rich, A. N. (2013). Beyond colour perception : Auditory-visual synaesthesia induces experiences of geometric objects in specific locations. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 49(6), 1750- 1763. <https://doi.org/10.1016/j.cortex.2012.04.006>
- Cuskley, C., Dingemans, M., Kirby, S., & van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia : Categorical perception and structure in vowel-color mappings in a large online sample. *Behavior Research Methods*, 51(4), 1651- 1675. <https://doi.org/10.3758/s13428-019-01203-7>
- Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P., & Merikle, P. M. (2006). The Role of Meaning in Grapheme-Colour Synaesthesia. *Cortex*, 42(2), 243- 252. [https://doi.org/10.1016/S0010-9452\(08\)70349-6](https://doi.org/10.1016/S0010-9452(08)70349-6)
- Elman, J. L., & McClelland, J. L. (1984). Speech Perception as a Cognitive Process : The Interactive Activation Model. In N. J. Lass (Éd.), *Speech and Language* (Vol. 10, p. 337- 374). Elsevier. <https://doi.org/10.1016/B978-0-12-608610-2.50013-3>
- Fernay, L., Reby, D., & Ward, J. (2012). Visualized voices : A case study of audio-visual synesthesia. *Neurocase*, 18(1), 50- 56. <https://doi.org/10.1080/13554794.2010.547863>
- Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25(1), 135- 155. [https://doi.org/10.1016/0010-0277\(87\)90007-2](https://doi.org/10.1016/0010-0277(87)90007-2)
- Johnson, K., & Sjerps, M. J. (2021). Speaker Normalization in Speech Perception. In *The Handbook of Speech Perception* (p. 145- 176). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119184096.ch6>

- Kang, M.-J., Kim, Y., Shin, J.-Y., & Kim, C.-Y. (2017). Graphemes Sharing Phonetic Features Tend to Induce Similar Synesthetic Colors. *Frontiers in Psychology*, 8. <https://www.frontiersin.org/article/10.3389/fpsyg.2017.00337>
- Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology : Toward an integrated framework of cognitive style. *Psychological bulletin*, 133(3), 464- 481. <https://doi.org/10.1037/0033-2909.133.3.464>
- Kunda, M., & Goel, A. K. (2011). Thinking in Pictures as a cognitive account of autism. *Journal of Autism and Developmental Disorders*, 41(9), 1157- 1177. <https://doi.org/10.1007/s10803-010-1137-1>
- Marks, L. E. (1975). On colored-hearing synesthesia : Cross-modal translations of sensory dimensions. *Psychological Bulletin*, 82(3), 303- 331.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1- 86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- 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(11), 1371- 1394. <https://doi.org/10.1068/p3429>
- Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-Modal Associations in Synaesthesia : Vowel Colours in the Ear of the Beholder. *I-Perception*, 5(2), 132- 142. <https://doi.org/10.1068/i0626>
- Mottron, L., Bouvet, L., Bonnel, A., Samson, F., Burack, J., Dawson, M., & Heaton, P. (2013). Veridical mapping in the development of exceptional autistic abilities. *Neuroscience & Biobehavioral Reviews*, 37(2), 209- 228. <https://doi.org/10.1016/j.neubiorev.2012.11.016>

- Neufeld, J., Roy, M., Zapf, A., Sinke, C., Emrich, H. M., Prox-Vagedes, V., Dillo, W., & Zedler, M. (2013). Is synaesthesia more common in patients with Asperger syndrome? *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00847>
- Nguyen, N., Wauquier, S., & Tuller, B. (2009). The dynamical approach to speech perception : From fine phonetic detail to abstract phonological categories. In *Approaches to Phonological Complexity* (p. 191- 218). De Gruyter Mouton. <https://doi.org/10.1515/9783110223958.191>
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8- 13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Rothen, N., Seth, A. K., Witzel, C., & Ward, J. (2013). Diagnosing synaesthesia with online colour pickers : Maximising sensitivity and specificity. *Journal of Neuroscience Methods*, 215(1), 156- 160. <https://doi.org/10.1016/j.jneumeth.2013.02.009>
- Simner, J. (2007). Beyond perception : Synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Sciences*, 11(1), 23- 29. <https://doi.org/10.1016/j.tics.2006.10.010>
- Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology*, 103(1), 1- 15. <https://doi.org/10.1348/000712610x528305>
- Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K., & Ward, J. (2006). Synaesthesia : The prevalence of atypical cross-modal experiences. *Perception*, 35(8), 1024- 1033.
- Smilek, D., Dixon, M. J., Cudahy, C., & Merikle, P. M. (2002). Concept driven color experiences in digit-color synesthesia. *Brain and Cognition*, 48(2- 3), 570- 573.
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech : A working model. *Journal of Phonetics*, 39(4), 456- 466. <https://doi.org/10.1016/j.wocn.2010.09.001>

Ward, J. (2013). Synesthesia. *Annual review of psychology*, 64(1), 49- 75.

<https://doi.org/10.1146/annurev-psych-113011-143840>

Ward, J. (2021). Synaesthesia as a model system for understanding variation in the human

mind and brain. *Cognitive Neuropsychology*, 38(4), 259- 278.

<https://doi.org/10.1080/02643294.2021.1950133>

Wauquier-Gravelines, S. (1999). Segmentation lexicale de la parole continue : La linéarité en

question. *Recherches linguistiques de Vincennes*, 28, Art. 28.

<https://doi.org/10.4000/rlv.1217>

Table A.1. Different graphemic versions of the same phoneme as well as a word example used in the task "Testing synesthesia in the context of words"

phoneme	graphemic version	word example
/a/	1. 'a'	/sak/ 'sac'
	2. 'at'	/ʃa/ 'chat'
	3. 'as'	/pa/ 'pas'
	4. 'e'	/fam/ 'femme'
/ɛ̃/	1. 'ain'	/pɛ̃/ 'pain'
	2. 'in'	/vɛ̃/ 'vin'
/ã/	1. 'an'	/bãd/ 'bande'
	2. 'anc'	/bã/ 'banc'
	3. 'ant'	/gã/ 'gant'
	4. 'ent'	/vã/ 'vent'
	5. 'ang'	/sã/ 'sang'
/o/	1. 'au'	/fon/ 'faune'
	2. 'eau'	/po/ 'peau'
	3. 'ot'	/so/ 'sot'
/ø/	1. 'oeu'	/nø/ 'nœud'
	2. 'eue'	/kø/ 'queue'
/g/	1. 'g'	/goʃ/ 'gauche'
	2. 'gu'	/gɛp/ 'guêpe'
/k/	1. 'c'	/kod/ 'code'
	2. 'qu'	/kø/ 'queue'
/s/	1. 's'	/sup/ 'soupe'
	2. 'c'	/ʃãs/ 'chance'

