



**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. *Consciousness and Cognition*, 2023, 111, 10.1016/j.concog.2023.103509 . hal-03880364v3

**HAL Id: hal-03880364**

**<https://hal.science/hal-03880364v3>**

Submitted on 11 Apr 2023 (v3), 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 synesthesia

Bouvet, L.<sup>1+</sup>, Magnen C.<sup>2+</sup>, Bled. C.<sup>1</sup>, Tardieu., J.<sup>2</sup>, Ehrlé, N.<sup>3,4</sup>

<sup>1</sup>Université de Toulouse, CERPPS, Toulouse, France

<sup>2</sup>Université de Toulouse, CNRS, Maison des Sciences de l'Homme de Toulouse, France

<sup>3</sup>Service de neurologie, CHRU Maison Blanche, Reims, France

<sup>4</sup>MC<sup>2</sup>Lab, UR 7536, Université Paris Cité, France

<sup>+</sup>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 this paper was to assess the genuineness of a case of phoneme color synesthesia by evaluating the influence of several psycholinguistic levels with different stimuli (isolated vowels, nonsense syllables, and words). Results demonstrate the robustness of the synesthesia no matter what the type of stimuli. To explore how this form of synesthesia manifested itself in everyday conversation, interviews were also conducted. VA reported that, in the context of conversation, phonemes still evoked colors that she had to translate in order to access the meaning. She also reported that her mental representations were multisensorial and that the verbal dimension was almost non-existent. We address several implications of this phoneme color synesthesia: the atypical speech perception that it implies, the cognitive cost of this stable system, and its relation to a specific cognitive functioning.

**Keywords:** phoneme color synesthesia, speech perception, multisensory representations, veridical mapping, segmental speech processing, dual coding theory,

## 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 (e.g., a letter can trigger the perception of a specific color or a sound can provoke the perception of complex colored photism). Several criteria are used in order to evaluate the genuineness of the synesthesia: the idiosyncratic synesthetic association is generally automatic, involuntary and stable over time (Ward, 2013). 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. Although orality is the primary dimension of language, practically no studies have evidenced phoneme color synesthesia. We previously reported, from a neuropsychological perspective, the case of VA who possesses phoneme color synesthesia. The aim of this paper was to further document VA's phoneme color synesthesia from a psycholinguistic perspective by evaluating the variations of her colored perception as a function of several linguistic levels (acoustic, semantic and graphemic) and its manifestation in a conversational context.

Several studies have explored the implication of the different psycholinguistic levels (i.e., acoustic, phonemic, lexical and graphemic) in speech color 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 pitch<sup>1</sup> of vowels: high-pitched vowels (/i/, /e/) are linked to brighter colors and low-pitched vowels (/o/, /u/) are linked to darker colors (Marks, 1975). On similar lines, Fernay et al. (2012) evidenced that voice pitch is linked to the luminance (higher pitch being lighter). Moreover, Moos et al. (2014) explored the link between acoustic formants<sup>2</sup> and colors. They found that lower values of the first

---

<sup>1</sup> Pitch refers to our perception of the fundamental frequency. Fundamental frequency, often called F0, is the frequency at which the vocal cords vibrate in vowels and voiced consonants. It generates prosodic variations, i.e. of melody and intonation, which contribute to the identification of the sex, the age and the identity of the speaker, as well as to the meaning of the spoken messages.

<sup>2</sup> 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.

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. They furthermore evidenced the predominant role of categorical phoneme perception over and above acoustic cues (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. The authors observed that for words, EP provided a single color whereas for nonwords she provided several colors that could be linked to those triggered by the letters of the alphabet. However, this conclusion concerned only the graphemic level as the association between color and phonemes was not tested. In another study, Baron-Cohen et al. (1993) tested whether the synesthetic perception of 9 subjects with colored hearing synesthesia was mediated by a phonemic or rather by a graphemic representation. 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). However, one can note a recurrent confusion due to the methodological difficulty of working with the oral dimension of language. Indeed, except for vowels, speech sounds are often presented in a syllable, a word or by referring to their graphemic correspondence (i.e. with letters). The problem is that this summons several types of representations which are not directly those that are targeted. Recent studies on synesthesia started to highlight the association 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 among grapheme color synesthetes. 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.

The goal of the present study was to evaluate the influence of the different linguistic variables (acoustic, semantic and graphemic) on a previously reported case (VA) of phoneme color synesthesia (Bouvet et al., 2017). The underlying objective was to evaluate the genuineness of this phoneme color synesthesia, as a “pure” case of this type of synesthesia has not been previously described. We systematically observed whether VA’s synesthesia varied with the perception of different types of linguistic stimuli (from phonemes to words) that combined different types of representations and that varied gradually. In the first task, we used isolated vowels and consonant-vowel (CV) syllables. We investigated the influence of acoustic parameters on VA’s synesthetic perception. For isolated vowels, we tested two acoustic parameters: vocalic formants and voice pitch (F0). If VA’s synesthesia relies on acoustic parameters we could expect, according to the previous literature, that: 1) higher vowels (F1) would be lighter, greener and yellower whereas more back vowels (F2) would be darker, redder and bluer, and 2) higher voice pitch would be associated to brighter colors. For CV syllables, if VA continued to perceive a synesthetic color for each unit, we also investigated the factors that influenced the color attribution for the consonant-color associations. On this point, this work is exploratory as there is currently no study on this type of association. We also assessed whether the constituent elements of the syllable had an influence on the synesthetic colors. In the second task, we used words to investigate the influence of semantic and graphemic factors on VA’s synesthetic perception. Lastly, in a third task, we interviewed VA on how her synesthesia manifested itself in a conversational context. The purpose was to determine if the same underlying mechanisms of speech processing evidenced with experimental protocols occurred in her daily life.

## **2. Case description**

As full details of VA’s developmental history were reported in the previous paper (Bouvet et al., 2017), we will give only the main lines here. The acquisition of oral language was atypical, as she started to speak in complete sentences at 16 months but did not pronounce a single word before then. She learnt to read by herself quite early (3 years old). At this age she was also able to perform mental calculations. She showed a particular interest in dinosaurs and space. Despite her intellectual precocity, VA experienced motor and coordination difficulties. She had difficulties developing social relationships, and was socially excluded at school. She experienced difficulties in formal learning during her studies as she struggled to understand ambiguous instructions. She performed poorly in mathematics. She reported seeing spatialized dynamical photisms when hearing sounds and also claimed to have

developed an astrophysical theory. She consulted psychiatrists during infancy and as an adult. None of them mentioned a diagnosis of autism but social phobia and schizophrenia diagnoses were made. The aim of the previous paper was to highlight how neurological and psychiatric conditions can be confused. We evidenced, via a systematic neuropsychological evaluation, that VA possessed higher than average cognitive abilities. We also evidenced in VA the possession of a complex sound-color (speech and non-verbal sounds) synesthesia. For speech, it was found that phonemes in particular triggered the perception of color, while for non-verbal sounds, the perception of a spatialized dynamical colored photism was described. We also observed that VA did indeed possess special abilities in astrophysics. After this first study, she agreed to continue the research with the aim of better understanding her phoneme color synesthesia and her cognitive functioning.

### **3. Testing the influence of acoustic parameters on VA's synesthesia**

#### **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 platform<sup>3</sup> 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 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. 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. Out of the 244 stimuli created, the phonetician (CM) selected the best exemplar of each phoneme produced by a male and by a female speaker on the basis of its perceptual typicality. In total, there were 122 stimuli: 26 vowels (13 vowels\* 2 speakers) and 96 CV syllables (16 consonants\* 3 contexts\* 2 speakers).

---

<sup>3 3</sup> <https://mshs.univ-toulouse.fr/plateformes/les-plateformes-technologiques/ccu-cognition-comportements-et-usages/petra/>

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 long as she needed to give her answer.

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.

## 3.2. Analysis and results

### 3.2.1. Isolated vowels

For 8 of the 13 vowel categories (see Figure 1), 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.

*Analysis.* Voice pitch (F0) and formants (F1 & F2) were extracted and analyzed using Praat© software (see supplementary data). To evaluate a potential link between the voice pitch, formants and colors, as certain values did not follow the assumption of normality (Shapiro-Wilk  $p$  values < .001), Spearman correlations were performed.

*Results.* Regarding voice pitch, no significant correlation was observed on the different Lab values with F0 [  $L$  :  $rs(24) = -.27, p = .17$  ;  $a$  :  $rs(24) = .36, p = .07$ ,  $b$  :  $rs(24) = .34, p = .09$ ]. Regarding the influence of formants, we observed a significant negative correlation between F1 and  $a$  values, ( $rs(24) = -.59, p < .001$ ) which indicates that the lower the F1 values are, the redder the colors. A significant negative correlation between F1 and  $b$  values was also observed ( $rs(24) = -.48, p = .01$ ), indicating that the lower the values on F1, the yellower the



colors. A significant correlation between F2 and  $L$  values was also observed ( $r_s(18) = .83, p < .001$ ), with vowels with higher F2 values being associated to lighter colors.

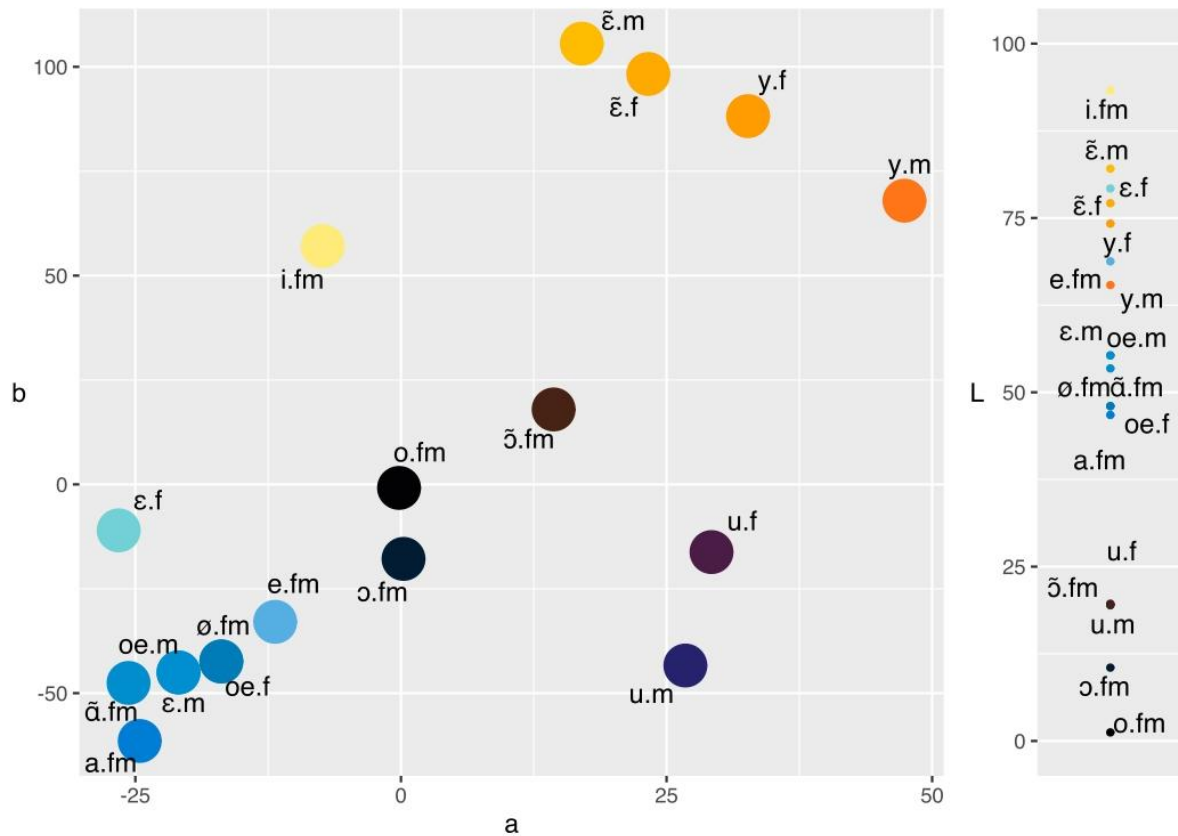


Figure 1. Representation of colors chosen by VA for vowels (2 trials) in the Lab color system. Axis labels indicate the color dimension ( $a$  green/red colors and  $b$  blue/yellow). On the right panel, the colors of the vowels are represented along the lightness dimension ( $L$ ).  $f$  indicates vowels pronounced by a female speaker and  $m$  indicates vowels pronounced by a male speaker.

### 3.2.2. CV syllables

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

*Consonants in a CV syllable context.* For 12 consonants ([p, k, b, d, g, n, f, s, ʃ, v, ʒ, ʁ], i.e., 75% of the 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 2, these differences in color are minimal.

Analysis. Consonants can be defined by their articulatory mode (either occlusive [p, b, k, g, t, d, m, n] or constrictive [ʃ, f, ʒ, l, ʁ, 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, ʁ]). 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. Analyses were conducted on these articulatory parameters to evaluate their influence on synesthetic color. 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.

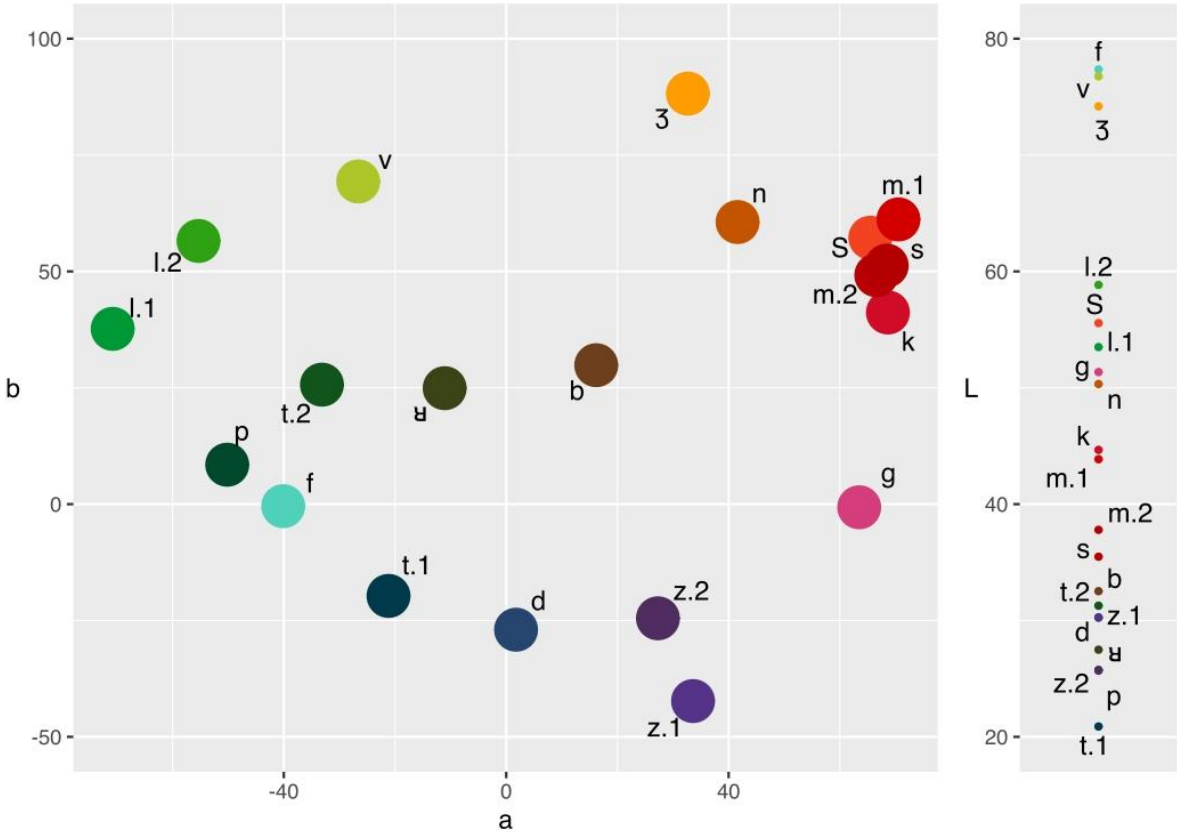


Figure 2. Representation of colors chosen by VA for consonants (6 trials) in the Lab color system. Axis labels indicate the color dimension (*a* green/red colors and *b* blue/yellow). On the right panel, the colors of the consonants are represented along the lightness dimension (*L*). Numbers indicate the two different colors chosen for the consonant.

Results. Regarding the link between colors and articulatory mode, the colors associated to constrictive consonants ( $M = 54.3$ ) were lighter ( $U = 621, p < .001$ ) and yellower ( $M = 38.1$ ) ( $U = 854, p = .07$ ) than the colors associated to occlusive consonants ( $L : M = 38.3; b : M = 23.5$ ). No significant difference on the  $a$  value was observed ( $U = 906, p = .07$ ; constrictive  $M = 7.93$ ; occlusive  $M = 22.4$ ).

Regarding the link between colors and the place of articulation, we observed a significant correlation with the  $a$  value. The more posterior the place of articulation of the consonants is, the redder they are ( $rs(94) = .033, p < .001$ ). Correlations with other Lab values ( $L$  and  $b$ ) were not significant ( $L: rs(94) = .08, p = .44; b: rs(94) = -.03, p = .73$ ).

*Vowels [i, a, u] in CV syllables.* The colors attributed to each vowel within different syllabic contexts are represented in Figure 3. For [a] three different colors were given: luminous cyan<sup>4</sup> (20 times), sky blue (11 times), and azur blue (1 times), luminous cyan being the color given for [a] when presented in isolation. For [i], the same color was given for all the 32 exemplars, namely light yellow. It is the same color as for [i] when presented in isolation. For [u], the same color (eggplant) was given for the majority of 31 exemplars, and the color dark violet was given for one exemplar. These two colors were given for the two exemplars of [u] when presented in isolation. Thus, no variation in color within the syllabic context was observed. Consequently, no further analyses on the influence of acoustic parameters on color were performed.

---

<sup>4</sup> The names of the colors are given by the MACal color chart.

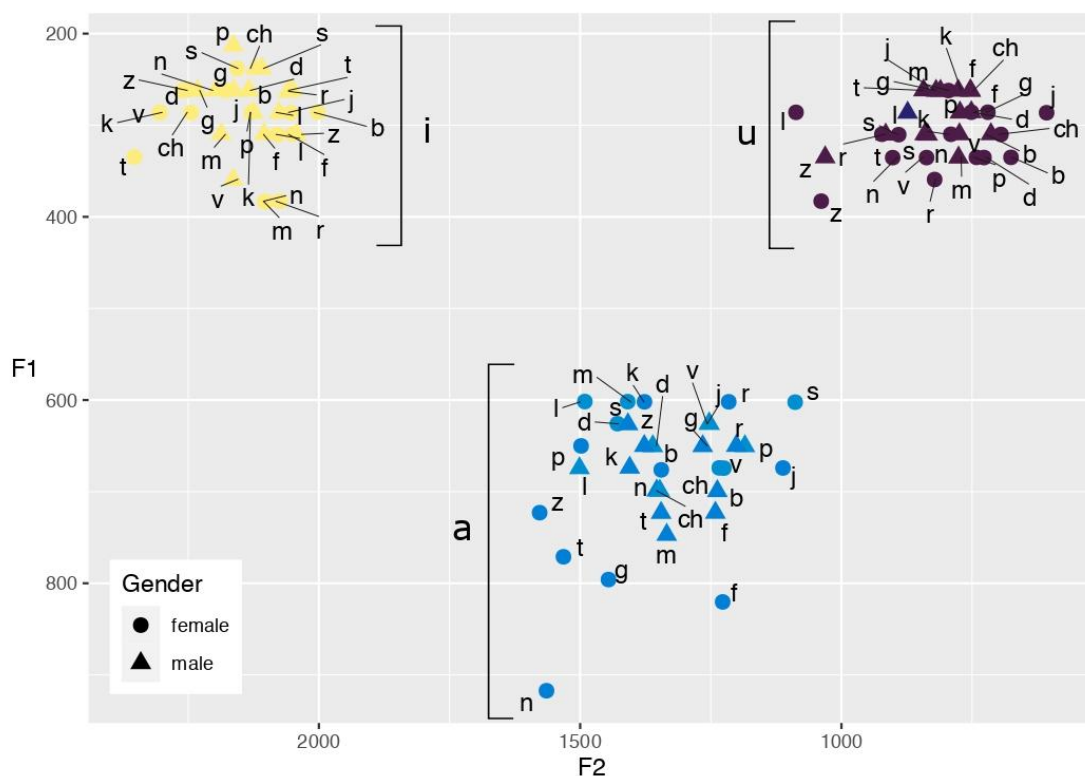


Figure 3. Representation of the colors chosen by VA for the 32 exemplars of each vowel /a/, /i/, /u/ in different consonant contexts according to the F1 and F2 values. Axis labels indicate vowel quality (vowel height on the F1 axis and front/back vowels on the F2 axis). Note that for [a], the difference between colors is minimal and difficult to discern.

### 3.3. Intermediate discussion

In this first task, we used isolated vowels and consonant-vowel (CV) syllables to investigate the influence of acoustic parameters on VA's synesthetic perception. First, we observed that changing the linguistic unit had almost no influence. For CV syllables, VA systematically provided a color for each unit. Interestingly, the phenomenon of coarticulation had no impact on the synesthetic color, as no significant variation in color within a consonant/vowel category was observed, no matter what the association of vowels or consonants. Moreover, the colors associated to vowels in a syllabic context were globally similar to those associated to the same vowels when presented in isolation. Furthermore, we observed that the synesthetic colors were determined by some factors linked to phonemic categories. For isolated 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), indicating that VA's synesthesia fits into the framework of

general cross-modal mapping to some extent. However, no significant relation between voice pitch and color was observed as previously evidenced (Fernay et al., 2012). For consonants, we observed a correlation between colors and the articulatory mode and the place of articulation. Constrictive consonants were associated to lighter and yellower colors than occlusive consonants and back consonants were associated to a redder color. Taken together, these results indicate that, in the case of VA, synesthetic colors are linked to articulatory and acoustic factors that are correlated with changes in phonemic category. 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 following task, we used words to investigate the influence of semantic and graphemic parameters on VA's synesthetic perception.

#### **4. Testing the influence of semantic and graphemic parameters on VA's synesthesia**

To test the influence of the semantics on VA's synesthesia we used words composed of phonemes associated to different synesthetic colors. Under the assumption that VA's synesthesia was influenced by the meaning, we wanted to be able to evaluate whether the position of phonemes within the words defined the synesthetic color. Furthermore, in the corpus, some phonemes had different graphemic representations. This allowed us to evaluate their possible influence on the synesthetic colors.

##### **4.1. Material and methods**

We selected 60 monosyllabic words, all nouns, 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.

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. As in the previous task, each color was converted to CIELab space in order to perform analysis.

**4.2. Analysis and results**

*Semantic influence.* For the bi-phonemic words, VA gave 2 colors and for the tri-phonemic words she gave 3 colors. Despite the different phonemic contexts, the colors of the phonemes did not seem to influence each other since the triggered colors remained identical 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 (see Figure 4). We observed that the meaning of the word has no significant influence on VA’s color attribution mechanism.

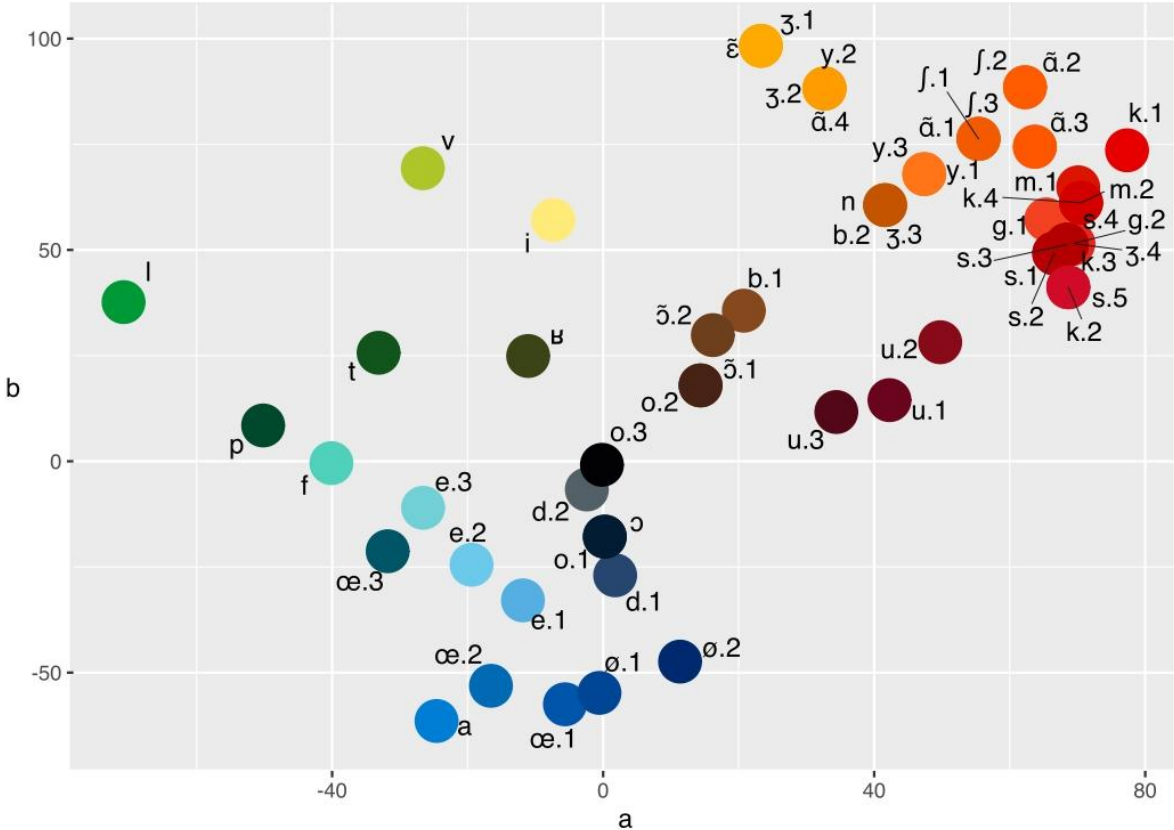


Figure 4. Representation of colors chosen by VA for phonemes in a semantic context in the Lab color system. Axis labels indicate the color dimension (a green/red colors and b

blue/yellow). Numbers indicate the different colors chosen for the different versions of the phonemes.

*Graphemic influence.* Among the selected words, there were different graphemic representations for 5 vowel phonemes and for 3 consonant phonemes (see Table 1). It can be seen in Figure 5 that for phonemes for which there are some graphemic representations made of two or three letters, VA provided a single color. Moreover, for each phoneme, there were very slight color variations which cannot be linked to the graphemic variations.

Phoneme	Graphemic version	word example
/a/	'a'	/sak/ 'sac'
	'at'	/ʃa/ 'chat'
	'as'	/pa/ 'pas'
	'e'	/fam/ 'femme'
/ɛ/	'ain'	/pɛ/ 'pain'
	'in'	/vɛ/ 'vin'
/ɑ/	'an'	/bɑd/ 'bande'
	'anc'	/bɑ/ 'banc'
	'ant'	/gɑ/ 'gant'
	'ent'	/vɑ/ 'vent'
	'ang'	/sɑ/ 'sang'
/o/	'au'	/fon/ 'faune'
	'eau'	/po/ 'peau'
	'ot'	/so/ 'sot'
/ø/	'oeu'	/nø/ 'nœud'
	'eue'	/kø/ 'queue'
/ɛ/	'g'	/go/ 'gauche'
	'gu'	/gɛp/ 'guêpe'
/k/	'c'	/kod/ 'code'
	'qu'	/kø/ 'queue'
/s/	's'	/sup/ 'soupe'
	'c'	/ʃɑs/ 'chance'

Table 1. The different graphemic versions of 8 phonemes as well as a word example

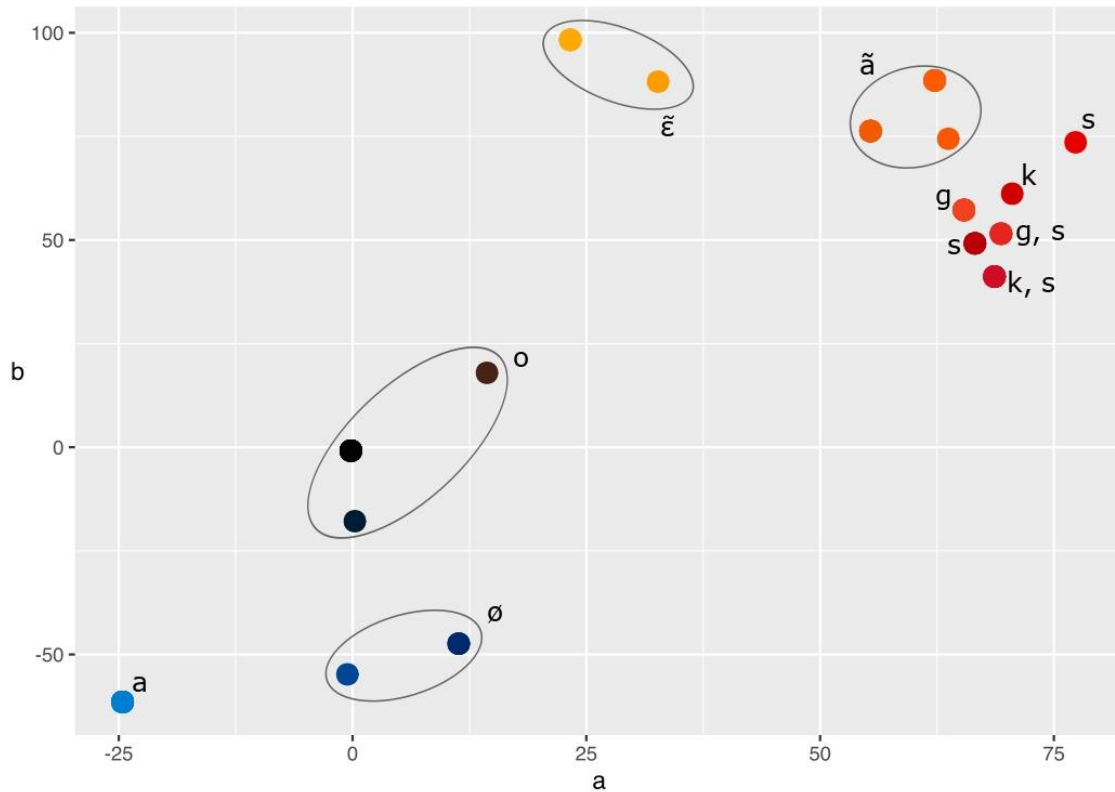


Figure 5. Representations of colors selected for phonemes with different graphemic representations on the  $a$  (green-red dimension) and  $b$  (blue-yellow dimension) axes of the Lab color system.

*Testing the robustness of VA's phoneme color synesthesia.* Since no semantic or graphemic influences on VA's synesthesia were found, we tested the robustness of the phoneme color synesthesia. We compared the colors given for phonemes presented outside a semantic context (isolated vowels and CV syllables of the previous experiment) and the colors given for the phonemes in a semantic context (phonemes that composed the words in the present experiment). 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.

For 9 phoneme categories, the  $d$  value was 0, indicating that VA chose exactly the same color for the phoneme in the two contexts (outside and within a semantic context). For 18 phoneme



categories, the  $d$  value was well below 109. For one phoneme category (/ǎ/,  $d = 153$ ), the color was completely different between the two tasks (see Figure 6).

























































	Phoneme outside a semantic context	Phoneme within a semantic context	$d$ value		Phoneme outside a semantic context	Phoneme within a semantic context	$d$ value
/a/			0	/k/			9.4
/ɔ/			0	/l/			17
/b/			29.8	/m/			8.1
/ʃ/			20.1	/n/			0
/d/			17.5	/o/			3.1
/e/			0	/u/			47.4
/ɜ/			16.8	/p/			0
/ø/			38	/ʌ/			0
/œ/			17.4	/s/			5.3
/f/			0	/t/			7,6
/g/			56.9	/ɣ/			5.9
/i/			0	/v/			0
/ɜ/			17.4	/ǎ/			153.2
/ɛ̃/			5.4	/š/			6.63

Figure 6. Mean colors for phonemes when presented outside (experiment 1) or within a (experiment 2) semantic context. 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.

### 4.3. Intermediate discussion

When testing VA's synesthesia with words, we observed that changing the linguistic unit had, as previously observed with syllables, practically no influence. Contrary to what might have been expected, we did not observe any effect of semantic knowledge on VA's synesthetic perception unlike the study of Baron-Cohen et al. (1993). VA continued to provide one color for each phoneme that composed the word. What is striking is the similarity

of the colors triggered for isolated phonemes and CV syllables and the colors triggered for phonemes within the context of a word. These two different tasks were conducted several years apart. This demonstrates the robustness of the synesthesia. Moreover, no significant influence of the graphemic representations was observed, in contrast to what is found in the majority of linguistic synesthesias (Simner, 2007). Taken together, these results indicate that her synesthesia manifests itself at the level of phoneme categories. Note that these results are based on experimental studies in which VA was asked to provide colors for controlled linguistic units. Words were presented outside their pragmatic context and the result obtained may not reflect the process VA uses in everyday conversation. In order to determine if this process also occurs in a conversational context, we interviewed VA.

## **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). 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 was 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**

#### ***Manifestation of her synesthesia during speech perception***

To access the meaning of what is being said in a conversational context, VA described two distinct processes. First of all, the association between sound and color is automatic and these two dimensions 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 may 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.”*

When she has to perceive and to understand speech, a 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.”*

### **A multisensory predominantly visual cognitive style**

This process of translation gives her access to a multisensory representation of the meaning: *“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.”* These multisensory representations seem to constitute her main way of thinking which is 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.”* *“I don't think in words.”* 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”.*

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.”*

### **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. The processes of translation make it difficult for VA to interact with others: *“It's hard to understand what I hear and to express myself.”* She also 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.”* These double processes of translation made it difficult for VA to interact with others: *“It's hard to understand what I hear and to express myself.”*

### **5.3. Intermediate discussion**

Several conclusions can be drawn from this thematic analysis. First of all, for VA, sounds and colors are automatically associated and in the context of conversation the perception of a phoneme automatically triggers the perception of the associated color. What is quite surprising is that the colors are translated and substituted by a multisensory representation of the meaning. These multisensory representations seem to constitute her main way of thinking and cause her difficulties apprehending and evolving in a world that is mostly verbal. Furthermore, when she has to produce speech, she has to translate her multisensory representations into speech. Paradoxically, these processes of translation (speech/colors to multisensorial representations and multisensorial representations to speech) are not automatic and imply a high cognitive cost. Globally, holding a conversation is a difficult exercise for her.

## **6. General 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, semantic or graphemic factors influenced VA's synesthesia. Across these different types of stimuli, we systematically observed that VA provided one specific color for each phoneme. A very stable system emerged from the association between colors and phonemic representations. This observation suggests that VA processes speech segmentally either in experimental conditions when she has to associate colors with different types of linguistic units or in daily conversation. This observation raises several questions that we address in the following.

First of all, her segmental speech perception to access word meaning is atypical and questions current models of the field. During the 80s, the word recognition process via phonemes 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). The case of VA highlights that access to the meaning uniquely via phonemes is possible but that this rigidity of functioning has a high cognitive cost. We can also wonder why VA developed this atypical speech processing. We can hypothesize that VA used a specific mechanism to represent language with a stable but costly system in order to organize the excess of visual information. This speech process is reminiscent of the learning process of reading when children decipher words grapheme by grapheme, unit by unit. It is interesting to note that VA learnt to read alone and at quite an early age (at 3 years old; see Bouvet et al., 2017). 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. A similar atypical mechanism to organize information has been described in the acquisition of some specific abilities (synesthesia, hyperlexia, absolute pitch) via the mechanism of veridical mapping in autism (Bouvet et al., 2014; Mottron et al., 2013).

Thematic analysis also evidenced a way of thinking dominated by multisensory – mainly visual – representations in which verbal representations are practically absent. In order to communicate, she described a process of translation from verbal to non-verbal representations. This description raises several questions. Most psycholinguistic models consider that our knowledge about words is stored in a mental lexicon. This mental lexicon contains semantic, syntactic, morphological, phonological and orthographic properties of words. In this view, all information (both verbal and non-verbal) is processed via this mental lexicon (Lahiri & Marslen-Wilson, 1991; Segui, 2015). However, the dual coding theory (Paivio, 2010) stipulates that verbal and non-verbal information are processed via two distinct but interrelated systems. The latter view accounts better for VA's difficulties in translating verbal into non-verbal information and vice versa. The cognitive cost of shifting between the

two types of representation seems to underline her communication difficulties. We can wonder to what extent communication difficulties observed in certain psychopathological disorders such as autism may be related to this difficulty of translation between verbal and non-verbal representations, as autistic individuals are often described as sensory/visual thinkers (Bled et al., 2021; Kunda & Goel, 2011).

Finally, the question arises as to how this atypical speech processing and non-verbal thinking style are related. It was recently pointed out that synesthesia is one manifestation of a particular neurodiverse phenotype (Ward, 2021). In the case of VA we can hypothesize that her segmental speech processing (of which the phoneme color synesthesia is a manifestation) can be related to her particular way of thinking. As discussed earlier, using a segmental way to process speech would be a costly and inefficient process for most people. For VA, however, the use of phonemes might be the most efficient way to process speech as she does not use linguistic representations to think. Phonemes are the smallest unit of sound and meaning in speech and make it possible to decipher speech without access to the mental lexicon.

There are several limitations to this study. First of all, as a case study, results cannot be generalised to other groups in the population. However, case studies have the power to shed light on a particular (mainly compensatory) mechanism and can contribute to an understanding of the general functioning (Shallice, 1979). The uniqueness of VA's synesthesia is related to her atypical/segmental way of processing speech and her case could contribute to reflection on theories of speech perception. Moreover, her non-verbal way of thinking and the cognitive cost of translation into verbal representations might also contribute to a better understanding of the organization of our mental representations. Another limitation is that there is no control group included in this study. This methodological decision was motivated by the goal of the study which was to demonstrate the genuineness of the phoneme color synesthesia. All the tasks were designed around VA's synesthesia; the participation of a control group was therefore not relevant. Lastly, the use of interviews to explore VA's functioning can also be questioned as it provides subjective data. Thus, They may not reflect reality. However, this methodology allowed us to have access to some cognitive processes that would otherwise have remained inaccessible.

To conclude on this case study, using different methodologies we have evidenced a case of genuine phoneme color synesthesia. 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 taught us so much by participating in this research.

### **Funding**

This work was supported by the MSHS-T (Maison des sciences de l'homme et de la société de Toulouse) and by the French National Research Agency (VISUAL ANR-19-CE28-0012)

### **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., 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., 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>
- Cuskley, C., Dingemanse, 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>

- 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>
- 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)
- 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>
- 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>
- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form : A phonological approach to the recognition lexicon. *Cognition*, 38(3), 245- 294. [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R)
- 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)
- 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>
- 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>
- Paivio, A. (2010). Dual coding theory and the mental lexicon. *The Mental Lexicon*, 5(2), 205- 230. <https://doi.org/10.1075/ml.5.2.04pai>
- 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>
- Segui, J. (2015). Évolution du concept de lexique mental. *Revue de neuropsychologie*, 7(1), 21- 26.
- Shallice, T. (1979). Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology*, 1(3), 183- 211. <https://doi.org/10.1080/01688637908414450>
- 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., 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.
- 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>

