Cross-modal association between vowels and colours: A cross-linguistic perspective

Peggy P. K. Mok,[1] Guo Li,[1] Joanne Jingwen Li,[1] Hezul T. Y. Ng,[2] and Him Cheung[3,a)]
[1]*Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, Hong Kong*
[2]*Department of Psychology, The Chinese University of Hong Kong, Shatin, Hong Kong*
[3]*Department of Psychology, The Education University of Hong Kong, Tai Po, Hong Kong*

(Received 31 May 2018; revised 10 December 2018; accepted 12 March 2019; published online 23 April 2019)

Previous studies showed similar mappings between sounds and colours for synaesthetes and non-synaesthetes alike, and proposed that common mechanisms underlie such cross-modal association. The findings between vowels and colours, and between pitch and lightness, were investigated separately, and it was also unknown how language background would influence such association. The present study investigated the cross-modal association between sounds (vowels and pitch) and colours in a tone language using three groups of non-synaesthetes: Cantonese (native), Mandarin (foreign, tonal), and English (foreign, non-tonal). Strong associations were found between /a/ and red, /i/ with light colours, and /u/ with dark colours, and a robust pitch effect with a high tone eliciting lighter colours than a low tone in general. The pitch effect is stronger than the vowel associations. Significant differences among the three language groups in colour choices of other vowels and the strength of association were found, which demonstrate the language-specificity of these associations. The findings support the notion that synaesthesia is a general phenomenon, which can be influenced by linguistic factors. © *2019 Acoustical Society of America*.
https://doi.org/10.1121/1.5096632

[BVT] Pages: 2265–2276

I. INTRODUCTION

The study of synaesthesia, the involuntary cross-activation of senses in which stimuli in one sensory modality (e.g., sounds) automatically trigger experiences in a different modality (e.g., colours), has received growing attention in recent years (e.g., Simner and Hubbard, 2013; Lockwood and Dingemanse, 2015). Studies have demonstrated that synaesthetic-like cross-modal mappings between sounds and colours can also be found in normal populations, but with synaesthetes showing much higher consistency in their responses than non-synaesthetes do (Simner *et al.*, 2005; Moos *et al.*, 2014). They suggested that common psycholinguistic mechanisms underlie such cross-modal association in synaesthetes and non-synaesthetes alike (Simner, 2007; Ward *et al.*, 2006). Nonetheless, few studies have examined the effects of linguistic background on such association. We investigated the vowel and colour mappings in a tone language with three groups of non-synaesthetes: Cantonese (native), Mandarin (foreign, tonal), and English (foreign, non-tonal). The tone dimension allows us to examine the interaction between pitch and vowel quality on colour association. Our findings demonstrate the language-specific synaesthetic mappings with a much-needed cross-linguistic perspective for the growing research into cross-modal associations.

A. Common mechanisms for cross-modal associations

Synaesthesia is a rare condition with various forms. The overall prevalence of synaesthesia is 4.4% (Simner *et al.*, 2006). The most common types of synaesthesia involve linguistic units, e.g., words and graphemes, as triggers (inducers) that generate concurrents in the visual domain, e.g., colours and shapes (Baron-Cohen *et al.*, 1993; Simner, 2007). Simner (2007) discussed synaesthesia as a psycholinguistic rather than a mere low-level perceptual phenomenon as synaesthetic experiences can be triggered by a range of linguistic units, which shows that the inducer-concurrent mappings can be conceptually mediated. For instance, symbols that are visually distinct but belong to the same linguistic (conceptual) category, such as 6, six, and *6*, may trigger the same colour experience. Many studies have shown that similar patterns are found for both synaesthetes and non-synaesthetes, with synaesthetes showing much higher consistency in their responses (e.g., Marks, 1974; Ward *et al.*, 2006; Moos *et al.*, 2014). Martino and Marks (2001) and Marks (2013) distinguished strong and weak synaesthesia. Strong synaesthesia describes the unusual experiences of individual synaesthetes, while weak synaesthesia refers to the milder forms of cross-modal connection revealed through language and perception. Weak synaesthesia is most clearly evident in cross-modal metaphorical language, e.g., warm colour and sweet smell. They suggested that common neural processes underlie both forms of synaesthesia.

Similarly, Ward *et al.* (2006) pointed out that sound-colour synaesthesia can be best explained as an exaggeration of cross-modal mechanisms common to us all rather than a privileged pathway present only in synaesthetes. They speculated that the same cognitive mechanism is used by both synaesthetes and non-synaesthetes, but the mechanism

a)Electronic mail: cheunghim@eduhk.hk

differs in terms of precision and automaticity between the two groups. Marks (1975) and Simner (2012) argued that synaesthetes and non-synaesthetes appear to lie on a continuum. The "gold standard" of consistency in defining true synaesthesia (i.e., test-retest consistency over time) may have excluded those who do experience synaesthetic associations, but their association patterns may not be so consistent (Simner, 2012). Simner (2012) proposed a working definition of synaesthesia as a neurological hyper-association that aims to be more inclusive of its variants, consistent or not. Ward and Mattingley (2006) suggested that the ultimate value of research into synaesthesia is its ability to inform theories of normal or typical cognition. All these suggestions can help to demystify synaesthetic experiences and highlight the value of investigating cross-modal associations in normal populations, the target participants of the current study.

B. Associations between vowels and colours

Among various linguistically triggered synaesthesia types, phoneme (speech sounds)—colour synaesthesia is not prevalent (Baron-Cohen *et al.*, 1993; Day, 2005; Simner, 2007), consisting of only around 10% of all forms of synaesthesia (Ward and Cytowic, 2010), but very consistent mappings were reported. Strong mappings were found in synaesthetes: the vowel /a/ with red, front vowels /i/ and /e/ with brighter colours (e.g., white/yellow/green), and /u/ and /o/ with darker colours (e.g., black/brown/blue).

Not many studies have examined sound-colour mappings. Jakobson (1962) first pointed out the close connections between certain vowels and colours in coloured hearing: the vowel /a/ with red, /e/ and /i/ with brighter colours (e.g., white and yellow), and /o/ and /u/ with darker colours (e.g., black or blue). He suggested that two dimensions of colour, chromatic-achromatic and light-dark, are related to vocalic compactness and tonality, respectively. Compact vowels have the first two formants ($F1$ and $F2$) close together resulting in a concentration of acoustic energy in a narrow region of the spectrum, like /a/ with a small $F2$-to-$F1$ ratio, and diffuse vowels have the $F1$ and $F2$ being far apart from each other, like /i/ with a large $F2$-to-$F1$ ratio. Vocalic compactness relates to the vertical dimension of the vowel space. The vowel /a/ is maximally chromatic (pure red), and diffused vowels like /i/ have reduced chromaticity (yellow). The light-dark dimension of colours mainly relates to vocalic tonality, which refers to the front-back dimension of vowels as determined by the frequency of the second formant $F2$. Front vowels like [i] were proposed to be lighter than back vowels like [u]. As chromaticity decreases (i.e., for high vowels), the importance of the light-dark contrast will be increased.

Ryalls (1986) developed Jakobson's ideas further. He proposed that there was an organizing principle in the vowel-colour mapping: vowel "primes" ([i,a,u]) were associated with the "visual primes" (i.e., the primary colours yellow, red, and blue because all colours can be derived from these three colours), and he used such associations to derive hypothetical associations for other vowels. For example, the vowel [e] would be orange because it is between [i] (yellow) and [a] (red), and [o] would be purple because it is between [u] (blue) and [a] (red). In addition, he suggested that the acute-grave (i.e., high front vs back) and diffuse-compact features of vowels parallel the chromatic (yellow-blue and green-red) and light-dark (white-black) contrasts of colours: acute-grave with yellow-blue; acute-grave with white-black; diffuse-compact with red-green. Ryalls (1986) proposed a perceptual link at the neurological level between the sense modalities of vision and hearing.

Marks's (1975) meta-analysis of many case reports of coloured hearing revealed similar agreements across synaesthetes as also noted by Jakobson (1962) above: the vowel /a/ with red, /e/ and /i/ with yellow and white, /o/ with red and black, and /u/ with blue, brown, or black. While the empirical basis of Jakobson's finding is not clear, Marks converted the number of responses in the many synaesthesia reports he collected into scores on each of the three bipolar dimensions: yellow-blue, red-green, and white-black. He found that the most notable feature of the colours of vowels is how their brightness varies: /i/ and /e/, sounding relatively high in vowel "pitch," are brightest; and /o/ and /u/, sounding relatively low in vowel pitch, are darkest. Vowel pitch here should not be confused with vocal pitch ($F0$). Mark proposed that it is the second formant ($F2$) which is most closely related to the intrinsic vowel pitch, and the vowel pitch predicts the whiteness or blackness of colour association. He also found that, except for /i/, there is a positive relation between greenness-redness and the ratio of $F2/F1$. As the ratio increases (front vowels), greenness increases and redness decreases. He admitted that the yellow-blue dimension was more difficult to interpret. In any case, he found that at least two visual dimensions (white-black and red-green) showed clear correlation with vowel formant frequencies.

Marks (1975) also related vowel pitch ($F2$) and the $F2/F1$ ratio to the distinctive features of gravity and compactness, respectively. Vowel brightness correlates with gravity: acute (high-"pitched," higher $F2$, /i e/) vowels are brighter than grave (low-pitched, lower $F2$, e.g., /o u/) vowels. Compactness correlates with the red-green contrast: compact vowels (e.g., /a/) yield red colours, while diffuse vowels (e.g., /i/) yield green colours.

Some recent studies with non-synaesthetes also support Jakobson's and Marks's findings. Wrembel (2007) tested Polish participants' mapping of 6 Polish vowels with 11 basic colours (Berlin and Kay, 1969). Wrembel (2009) also tested the mapping of 12 English vowels and 11 basic colours by 2 groups of Polish students. In both studies, bright colours (yellow, green) were associated with high front vowels, whereas dark colours (brown, blue, black) were attributed to back vowels, while open sounds tended to be perceived as red and central vowels are mapped onto achromatic grey. Wrembel, however, did not measure any vowel formants and did not define the colour quantitatively.

The few previous studies on vowel-colour relationships above did not specify how they defined various colours. It is reasonable to assume that they adopted prototypical colours (e.g., Wrembel, 2009). This can explain why green is considered a light colour [with red-green-blue (RGB) values of 0, 255, 0], even though the same colour can have different shades (e.g., light green vs green vs dark green). Moos *et al.* (2014) improved on this by analysing vowel-colour relationships quantitatively. They used 16 vowels (8 primary cardinal vowels /i e ɛ a ɑ ɔ o u/ and 8 vowels synthesized to be intermediate between the cardinal vowels) and 16 colours, including shades of some colours (11 focal colours: white, black, red, green, yellow, blue, brown, grey, orange, pink, and purple, and five further colours: dark green, light green, pale pink, cyan, and dark blue to fill the gaps in the colour space). They also used a chromameter to measure the Commission Internationale de l'Eclairage (CIE) coordinates of the 16 colours on the monitor used for the experiment. The CIE coordinates were converted into CIELUV coordinates resulting in L* (luminance), u* (red-green), and v* (yellow-blue). Doing so enabled them to disentangle luminance and hue (e.g., Hamilton-Fletcher *et al.*, 2017). They found significant influence of $F1$ and $F2$ on colour association: a higher u* (redder) correlates with decreasing $F2$ and particularly strongly with increasing $F1$ (i.e., /a/). A rising $F2$ (e.g., for front vowels) correlates with greenish colours. No significant results were found for v*, which represents the blue-yellow axis. The influences of $F1$ and $F2$ on u* held true across participant groups, but with a stronger $F1$ influence for synaesthetes.

One striking finding among the above studies is the consistent mappings between several vowels and colours in spite of the differences in methodology and participant groups. The low vowel /a/ is typically red, front-unrounded vowels /i/ and /e/ are associated with bright colours (white/yellow/green), and back-rounded vowels /u/ and /o/ are associated with dark colours (black/blue/brown). These studies also demonstrated that formant frequencies, especially $F2$, are influencing such mappings. Nevertheless, there is one neglected aspect. The comparison between front-unrounded vowels and back-rounded vowels involves two independent vocalic dimensions, which jointly affect $F2$: front-back and unrounded-rounded. Front vowels (e.g., /i y/) have higher F2 than back vowels (e.g., /u ɔ/), and unrounded vowels (e.g., /i ɛ/) also have higher $F2$ than rounded vowels (e.g., /y œ/). Their $F2$ values form a continuum: /i/ > /ɛ/, /y/ > /oe/ > /ɔ/ > /u/. Even for the study by Moos *et al.* (2014) involving 16 vowels, the two dimensions were still not separated because there was no front-rounded vowel (e.g., /y/ and /œ/) or back-unrounded vowel (e.g., /ɯ/ and /ɤ/), which are much less commonly found than front-unrounded and back-rounded vowels in the world's languages (Maddieson, 1984). So far, no study has explored whether frontness or rounding is more influential in vowel-colour mapping.

In addition, most previous studies on synaesthesia were based in English, although synaesthesia cases in other languages like French, German, Polish, and Chinese were also reported (Marks, 1975; Wrembel, 2009; Simner *et al.*, 2011). These studies do suggest that cross-modal association is likely to be a general phenomenon. Nevertheless, no study has compared how different language backgrounds would affect the vowel-colour associations, especially for non-synaesthetes. Some studies demonstrated that the grapheme-colour mappings can be transferred from $L1$ to $L2$ in bilingual synaesthetes as well as in laboratory settings (see review in Mroczko-Wąsowicz and Nikolić, 2013). As the same persons were involved in these studies, this naturally explains the finding of cross-linguistic equivalents evoking the same colours in the $L1$-to-$L2$ synaesthetic transfer. Thus, the influence of language background on sound-colour association is still largely unknown.

C. Association between pitch and lightness

Compared to the few studies focusing on vowel-colour mappings, there are many more studies exploring the cross-modal association between pitch and lightness (e.g., Marks, 1974, 1975, 1989; Melara, 1989; Hubbard, 1996; Marks *et al.*, 2003; de Thornley Head, 2006; Ward *et al.*, 2006). They all found a salient and consistent mapping for synaesthetes and non-synaesthetes alike; that is, high pitch sounds are associated with light colours or high luminance. Higher tones are judged "whiter" than lower ones. This pitch-lightness correspondence is evident even in the perceptual matches of children as young as four (Marks *et al.*, 1987).

Most of the previous studies used only black/grey/white contrasts to examine the pitch-lightness mapping (e.g., Marks, 1974; Melara, 1989; Hubbard, 1996), although the effects of pitch can be seen in colour (hue) as well. From his literature review, Marks (1974) noticed that yellowness is more associated with a high pitch and blueness with a low pitch. Ward *et al.* (2006) observed that both synaesthetes and non-synaesthetes had a monotonic increase in lightness (defined by the Munsell lightness values of their colour choices) with pitch.

In addition, most previous studies used non-speech sounds to investigate the pitch-lightness association, e.g., musical notes (de Thornley Head, 2006; Ward *et al.*, 2006) or pure tones (Marks, 1974, 1989; Hubbard, 1996). The difference in fundamental frequency ($F0$) of the stimuli is also quite large, e.g., 100–10 000 Hz in Marks (1974), either 174.6 Hz or 1046.5 Hz in Melara (1989), and 200–3417 Hz in Hubbard (1996), although de Thornley Head (2006) demonstrated that the pitch-lightness association could also be found in musical notes with a much narrower pitch range (262–524 Hz). This suggests that the pitch-lightness association is mapped relatively rather than absolutely (Brunetti *et al.*, 2018).

Given the relative mapping, it is of interest to see whether and how the pitch-lightness association could be demonstrated in linguistic vowel sounds with an even narrower pitch range. Moos *et al.* (2013) investigated the synaesthetic colour and visual texture perceptions in response to different types of male voice qualities (e.g., modal, falsetto, whisper). They found that a higher $F0$ (as in falsetto, mean 232 Hz) led to lighter colour choices across participant groups, whereas a low $F0$ (as in creak, mean 92 Hz) resulted in darker colour associations. The materials in the study by Moos *et al.* (2013) were two short spoken passages from a story, which clearly demonstrated that the pitch-lightness effect can be found in speech sounds as well, although not necessarily individual vowel sounds. Nevertheless, there are other spectral differences between various voice qualities besides pitch, so the observed pitch effects are likely modulated by other spectral properties.

Fernay *et al.* (2012) used four vowels (/i ɛ ɑ u/) spoken with both male (60, 90, 120, 150 Hz) and female (180, 210, 240, 270 Hz) voices. They found that both the synaesthete and the ten control subjects tended to associate high pitch with lighter colours (luminance of the chosen colours), although the association seems to be more linear for the synaesthete than the controls who showed a more categorical distinction (male vs female voices). Nonetheless, the pitch difference in the study by Fernay *et al.* (2012) was confounded with the sex of the speakers producing the stimuli. Male and female voices differ in many other aspects in addition to pitch. Whether the pitch-lightness effect can be seen across vowels produced with the same voice still awaits investigation. Also, as their data were collapsed across vowels, it is not known if the pitch effect would show up in individual vowels as well, especially those vowels with robust colour mappings as discussed above, i.e., /i/ (light colours) and /u/ (dark colours).

D. The present study

Previous studies have found robust vowel-colour and pitch-lightness mappings in both synaesthetes and non-synaesthetes. However, the two effects were investigated separately. It is unclear which association is more dominant and how they may interact, especially for vowels with strong colour associations, e.g., /a/ with red. Our study answered these questions by investigating the vowel-colour association in a tone language, Cantonese. Tone languages are ideal for our purpose because linguistic pitch difference is intrinsically coded with vowel qualities, allowing exploration of colour mapping with both vowels and pitch simultaneously. The linguistic pitch difference in lexical tones is much narrower (around 100 Hz only) than in musical notes or pure tones used in previous studies. If the pitch-lightness effect can also be found in such a small pitch range, it will clearly demonstrate the robustness of this effect.

In addition to being a tone language, another reason why we chose Cantonese is that there are four rounded vowels contrasting in frontness: two front rounded vowels /y/ and /œ/ and two back rounded vowels /u/ and /ɔ/. There are also two pairs of front vowels contrasting in rounding: /i/ vs /y/ and /ɛ/ vs /œ/. Their F2 values form a continuum: /i/ > /ɛ/, /y/ > /œ/ > /ɔ/ > /u/. They allow us to examine whether frontness or rounding is more influential in vowel-colour mapping, and whether the colour mapping is determined by a continuous factor *F*2, as suggested by previous studies, or whether the mapping is affected by categorical vowel properties of frontness or rounding.

The choice of Cantonese is also motivated by the fact that Chinese is written logographically. The Chinese writing system bears little correspondence to actual pronunciation. Previous studies have shown that many of the coloured hearing synaesthetes were actually grapheme-colour synaesthetes (Baron-Cohen *et al.*, 1993; Ward *et al.*, 2006). Although Chinese characters can also induce coloured sensation (Simner *et al.*, 2011; Hung *et al.*, 2014), the low grapheme-phoneme correspondence in Chinese characters can ensure that any mappings found in our study can be attributed to actual speech sounds instead of the writing system. Unlike in mainland China where all students learn the official *pinyin* system for Mandarin, Cantonese speakers in Hong Kong (i.e., participants in the current study) are not taught any Romanization methods for Cantonese at all.

Finally, most previous studies on synaesthesia were based in English. So far, only a few studies have compared how different language backgrounds would affect the vowel-colour associations. Recently, Guillamón (2013) compared vowel-colour association between English and Arabic non-synaesthetes. She found that except strong associations between /ɑ/ and red and between /o/ and orange, there was no common association between the two languages. She suggested that variability in vowel-colour associations can be caused by linguistic, cultural, or even idiosyncratic factors. However, no acoustic measurement was conducted in her study, and she explained her findings based more on the cultural differences between the two groups of speakers. To further explore how linguistic backgrounds would affect vowel-colour associations specifically, we included three groups of non-synaesthetes: a native speaker group (Cantonese), a tone language group (Mandarin), and a non-tone group (English). They were tested with identical materials and procedures. Cantonese is unintelligible to both the Mandarin and English groups, but it is much closer to Mandarin than to English in terms of language typology, the use of lexical tone, and the presence of a front rounded vowel /y/. If the three groups exhibit different association patterns, we can conclude that language background can modulate cross-modal vowel-colour association.

II. METHOD

A. Participants

Thirty-six native speakers of Hong Kong Cantonese [15 males; mean age = 20.7 years, standard deviation (SD) = 1.74 years], 35 native speakers of Mandarin from China (8 males; mean age = 22.8 years, SD = 1.65 years), and 34 native speakers of English from USA, UK, or Australia (22 males; mean age = 20.8 years, SD = 2.06 years) participated in the experiment for course credits or payment. They were students or exchange students at The Chinese University of Hong Kong. The Mandarin speakers did not know Cantonese. Most of the English speakers had learned or were learning some Chinese (Mandarin or Cantonese), but they were at the beginner level. None of them had any hearing or neurological problems, and none reported any synaesthetic experience. All the participants had normal or corrected-to-normal vision.

B. Materials and design

Seven Cantonese vowels (/i y ɛ œ a ɔ u/) and two Cantonese tones (T1, a high-level tone [55] and T4, a low-falling tone [21]) were selected. Sixteen Cantonese initial consonants (/p p^h t t^h k k^h f s h ts ts^h m n l j w/) were also included. There were 14 stimuli for the vowel-only (V-only) condition (7 vowels × 2 tones) and 224 stimuli for the consonant-vowel (CV) condition (7 vowels × 2 tones × 16 consonants). Since all possible combinations were used, some were real words in Cantonese unavoidably: 88 out of

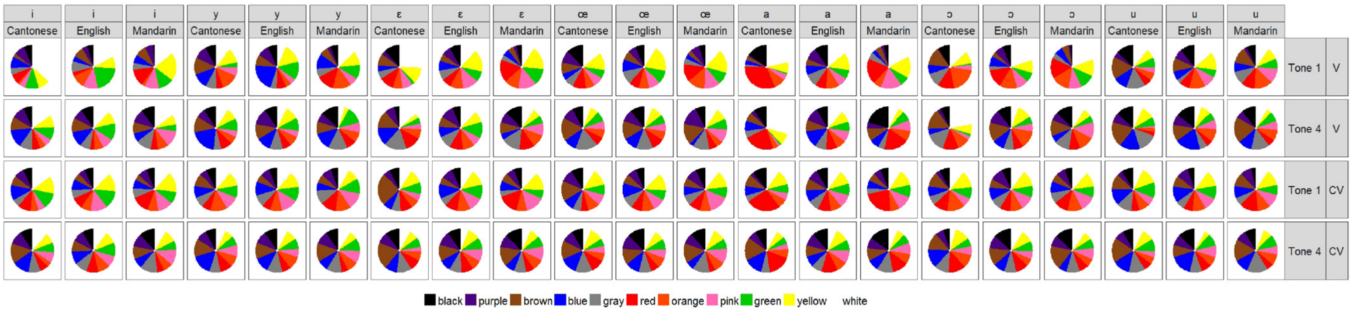

FIG. 1. (Color online) Pie charts showing colour choices for the seven vowels in different conditions by the three language groups.

the 238 stimuli (37%) were real words, which were fairly distributed among the vowels, consonants, and tones. Among the real words, only one relates to colour: /fɛ1/ "brown." This stimulus was excluded from some of the analyses (see details below). Due to similarity in language typology, 26 out of the 238 stimuli were real words in Mandarin (only with the vowels /i a u/ in tone 1, which are common to Cantonese and Mandarin). None of them is related to colour. All the stimuli were produced naturally by a female Cantonese native speaker.

We used 11 basic colours in our study: white, black, red, green, yellow, blue, brown, purple, pink, orange, and grey (Berlin and Kay, 1969). Typical RGB values for colours other than white, black, and grey were obtained by a survey conducted with 40 additional native Cantonese speakers who did not participate in this experiment. Four choices with similar hue were included for each colour, and the one with the highest votes was chosen for display in the experiment. The RGB values for the 11 colours can be found in the online supplementary material.[1] The lightness/darkness of the 11 colours was rated by the participants in a post-experiment survey containing a 7-point Likert scale. Participants were asked to rate each colour from one (extremely light) to seven (extremely dark). All but five Mandarin speakers completed the lightness rating.

C. Procedure

The experiment was a sound-colour matching task conducted individually using E-Prime 2.0 Professional (Psychology Software Tools, Sharpsburg, PA) with a desktop computer. Eleven colour squares would appear on the screen with a sound stimulus played over a headphone. The 11 squares were divided into 3 rows (4,3,4). Participants were asked to choose one colour from the 11 choices and press the corresponding button in the keyboard. The position of the colour squares on the screen was randomized for each trial. For each trial, participants had seven seconds to respond. They were encouraged to respond intuitively. The V-only syllables were repeated five times (the V-only condition) and each CV syllable appeared only once (the CV condition). There were 294 trials in total. The order of the sound stimuli was pseudorandomized so that the same syllable with two different tones would not appear consecutively.

III. RESULTS

First, we examined the colour choices of various vowels by the three language groups. Figure 1 shows the proportions of colour choices for each vowel in different conditions by the three language groups. The colour choices were arranged according to their lightness ratings (details below). A series of chi-square tests were conducted for each case in terms of condition (V vs CV), tone (T1 vs T4), and vowel (/i y ɛ œ a ɔ u/) to compare the results among different language groups. Trials of /fɛ/ in tone 1 were excluded from analysis for its strong association with brown in Cantonese. Results in Table I show significant differences between language groups for most cases except for vowels /i/, /œ/, and /y/ ($p = 0.063$ for /y/) in T4 in the V-only condition. *Post hoc*

TABLE I. Results of chi-square tests comparing colour choices of the three language groups. df = degrees of freedom.

Condition	Tone	Vowel	χ^2	df	p	Condition	Tone	Vowel	χ^2	df	p
V	T1	i	92.434	20	<0.001	CV	T1	i	49.584	20	<0.001
V	T1	y	50.726	20	<0.001	CV	T1	y	41.619	20	0.003
V	T1	ɛ	58.755	20	<0.001	CV	T1	ɛ	121.752	20	<0.001
V	T1	œ	34.369	20	0.024	CV	T1	œ	60.138	20	<0.001
V	T1	a	95.203	20	<0.001	CV	T1	a	93.374	20	<0.001
V	T1	ɔ	69.959	20	<0.001	CV	T1	ɔ	50.775	20	<0.001
V	T1	u	44.029	20	0.001	CV	T1	u	84.134	20	<0.001
V	T4	i	21.966	20	0.342	CV	T4	i	62.134	20	<0.001
V	T4	y	30.437	20	0.063	CV	T4	y	36.289	20	0.014
V	T4	ɛ	42.240	20	0.003	CV	T4	ɛ	39.666	20	<0.001
V	T4	œ	20.039	20	0.455	CV	T4	œ	58.778	20	<0.001
V	T4	a	89.201	20	<0.001	CV	T4	a	89.485	20	<0.001
V	T4	ɔ	60.705	20	<0.001	CV	T4	ɔ	94.589	20	<0.001
V	T4	u	37.209	20	0.011	CV	T4	u	66.150	20	<0.001

pairwise comparisons with Bonferroni corrections indicate that overall English and Mandarin speakers behaved similarly in most cases while both differed significantly from Cantonese speakers (details of the *post hoc* comparisons can be found in the online supplementary material).

To gain further insight into how the three language groups differ, another series of chi-square tests were conducted comparing colour choices for different vowels within each language group. The choices for /fɛ/ in tone 1 were included this time for a more comprehensive picture. To simplify the comparisons, Table II shows the highest colour choices with significant χ^2 results for various conditions (e.g., V-only in T1 with Cantonese speakers). We can see that vowel-colour mappings noted by previous studies can also be found in our data: /i/ with light colours (white/green/yellow), /a/ with red, /u/ with dark colours (brown/blue). Interestingly, /a/ was also found to associate with white in the V-only condition for Cantonese and English speakers. In addition, the effect of tone can be observed in the Cantonese and Mandarin data: T4 (a low falling tone) elicited a darker colour than T1 (a high level tone) for most vowels, and the effect was more consistent for the CV condition. The English data appeared to be less sensitive to tone differences, as there were non-significant colour associations related to tones in both conditions.

Pairwise comparisons with Bonferroni correction were conducted to further examine the strength of the vowel-colour mappings observed in Table II, which only shows the highest colour choices in each case, but the mappings may not be very robust (with 11 colours the chance level is 9%). As seven Cantonese vowels were used, any colour with significantly different pairwise comparison with three or more vowels (i.e., more than half of the vowels) was considered a strong association. This criterion was used to check the highest colour choices in Table II. In both V-only and CV conditions in Table III, there are more strong associations for Cantonese speakers than Mandarin and English speakers. It should be noted that the empty cells in Table III do not necessarily mean that there was no significant association, but that the association is not as strong as those listed (e.g., only significantly different from two other vowels). In addition to the expected mappings between /i/ and light colours, /a/ with

TABLE II. Vowel-colour associations by Cantonese, Mandarin, and English listeners in (a) V-only condition and (b) CV condition. (The numbers in parentheses show the highest percentage of choice. Empty cells indicate insignificant χ^2 results for that condition. T1 = tone 1, T4 = tone 4. The cell under T1 and T4 showed data pooled over the two tones.)

	/i/		/y/		/ɛ/		/œ/		/a/		/ɔ/		/u/	
(a) V-only	T1	T4	T1	T4	T1	T4	T1	T4	T1	T4	T1	T4	T1	T4
Cantonese	White (36.6)	White (15.3)	Brown (12.7)	Blue (20.1)	White (25.6)	Grey (16.9)	White (14.2)	Brown (17.3)	Red (31.6)	White (29.5)	Orange (19.4)	White (22.2)	Brown (16.8)	Brown (18.9)
	White (25.9)		Blue (15.9)		White (19.5)		Brown (12.9)		Red (27.2)		White (17.7)		Brown (17.8)	
Mandarin	Yellow (20.8)	White (14.9)	Orange (13.8)	Grey (14.4)	Yellow (17.3)	Grey (14.4)	Yellow (17.2)	Brown (15.6)	Red (18.9)	Black (20.8)	White (19.4)	Black (12.6)	Orange (13.5)	Brown (19.5)
	White (14.7)		Blue (12.4)		Red (14.1)		Yellow (13.0)		Red (18.7)		White (14.2)		Brown (15.9)	
English	Green (20.7)	—	Blue (21.0)	—	Yellow (16.3)	—	Yellow (18.8)	—	White (14.3)	——	White (19.8)	—	White (13.9)	——
	Green (16.7)		Blue (15.9)		Yellow (12.3)		Yellow (13.7)		White (13.0)		Orange (13.2)		Blue (13.0)	

Cantonese: T1 $\chi^2 = 307.888$, $p < 0.01$; T4 $\chi^2 = 225.692$, $p < 0.01$; overall $\chi^2 = 425.256$, $p < 0.01$.
Mandarin: T1 $\chi^2 = 123.718$, $p < 0.01$; T4 $\chi^2 = 136.287$, $p < 0.01$; overall $\chi^2 = 181.328$, $p < 0.01$.
English: T1 $\chi^2 = 146.616$, $p < 0.01$; T4 $\chi^2 = 72.159$, $p = 0.135$; overall $\chi^2 = 137.684$, $p < 0.01$.

	/i/		/y/		/ɛ/		/œ/		/a/		/ɔ/		/u/	
(b) CV	T1	T4	T1	T4	T1	T4	T1	T4	T1	T4	T1	T4	T1	T4
Cantonese	White (15.0)	Brown (14.8)	Orange (12.7)	Blue (15.5)	Brown (24.2)	Brown (19.8)	Brown (15.0)	Brown (14.1)	Red (26.0)	Brown (17.1)	Orange (13.7)	Blue (13.7)	Yellow (12.5)	Brown (18.3)
	White (13.1)		Blue (12.2)		Brown (22.0)		Brown (14.6)		Red (21.4)		Brown (12.2)		Brown (15.1)	
Mandarin	Yellow (15.4)	Grey (13.3)	Pink (12.4)	Brown (13.4)	Red (15.4)	Brown (14.3)	Pink (14.6)	Brown (14.4)	Red (22.4)	Black (15.3)	White (11.7)	Brown (17.4)	Red (14.0)	Brown (16.3)
	Yellow (12.4)		Grey (10.9)		Pink (12.2)		Pink (12.7)		Red (17.2)		Brown (11.9)		Brown (13.3)	
English	—	Green (10.1)	—	Brown (12.2)	—	Brown (12.1)	——	Black (12.9)	—	Red (12.6)	—	Black (13.4)	—	Blue (16.3)
	Green (11.6)		Blue (10.7)		Yellow (10.5)		Blue (10.8)		Red (11.3)		Yellow (10.8)		Blue (14.1)	

Cantonese: T1 $\chi^2 = 414.902$, $p < 0.01$; T4 $\chi^2 = 194.137$, $p < 0.01$; overall $\chi^2 = 477.577$, $p < 0.01$.
Mandarin: T1 $\chi^2 = 151.906$, $p < 0.01$; T4 $\chi^2 = 136.794$, $p < 0.01$; overall $\chi^2 = 211.200$, $p < 0.01$.
English: T1 $\chi^2 = 72.512$, $p = 0.129$; T4 $\chi^2 = 114.595$, $p < 0.01$; overall $\chi^2 = 137.684$, $p < 0.01$.

TABLE III. Strong vowel-colour associations in various conditions. The numbers in parentheses indicate the number of vowels that this particular association is significantly stronger than.

<table>
<tr><th></th><th colspan="2">/i/</th><th colspan="2">/y/</th><th colspan="2">/ɛ/</th><th colspan="2">/œ/</th><th colspan="2">/a/</th><th colspan="2">/ɔ/</th><th colspan="2">/u/</th></tr>
<tr><th>(a) V-only</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th></tr>
<tr><td>Cantonese</td><td>White(5)</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>Red(5)</td><td>White(5)</td><td>Orange(4)</td><td>White(3)</td><td>—</td><td>—</td></tr>
<tr><td></td><td colspan="2">White(3)</td><td colspan="2">Blue(3)</td><td colspan="2">White(3)</td><td colspan="2">—</td><td colspan="2">Red(6)</td><td colspan="2">——</td><td colspan="2">—</td></tr>
<tr><td>Mandarin</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>——</td><td>—</td><td>—</td><td>—</td><td>—</td></tr>
<tr><td></td><td colspan="2">—</td><td colspan="2">Blue(4)</td><td colspan="2">—</td><td colspan="2"></td><td colspan="2">Red(4)</td><td colspan="2">—</td><td colspan="2">Brown(3)</td></tr>
<tr><td>English</td><td>Green(4)</td><td>—</td><td>Blue(5)</td><td>—</td><td>——</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td></td><td>——</td><td>—</td><td>—</td></tr>
<tr><td></td><td colspan="2">Green(4)</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td></tr>
<tr><th></th><th colspan="2">/i/</th><th colspan="2">/y/</th><th colspan="2">/ɛ/</th><th colspan="2">/œ/</th><th colspan="2">/a/</th><th colspan="2">/ɔ/</th><th colspan="2">/u/</th></tr>
<tr><th>(b) CV</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th><th>T1</th><th>T4</th></tr>
<tr><td>Cantonese</td><td>White(3)</td><td>—</td><td>Orange(3)</td><td>—</td><td>Brown(6)</td><td>—</td><td>—</td><td>—</td><td>Red(6)</td><td>—</td><td>Orange(3)</td><td>—</td><td>—</td><td>—</td></tr>
<tr><td></td><td colspan="2">White(5)</td><td colspan="2"></td><td colspan="2">Brown(6)</td><td colspan="2">—</td><td colspan="2">Red(6)</td><td colspan="2">—</td><td colspan="2">—</td></tr>
<tr><td>Mandarin</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>Red(5)</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td></tr>
<tr><td></td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">Red(6)</td><td colspan="2">—</td><td colspan="2">—</td></tr>
<tr><td>English</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>—</td><td>Blue(4)</td></tr>
<tr><td></td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">—</td><td colspan="2">Red(3)</td><td colspan="2">—</td><td colspan="2">Blue(4)</td></tr>
</table>

red, and /u/ with dark colours, there are some new mappings. In the CV condition for Cantonese speakers, /y/ and /ɔ/ are orange, and /ɛ/ is brown (which was likely influenced by the strong association of /fɛ/ in tone 1 with brown, although Table II(b) shows that Mandarin and English speakers also preferred brown for /ɛ/ in a low tone). Interestingly, in the V-only condition, /y/ is blue for all three participant groups, while /ɛ/ is white for Cantonese speakers. The effects of language background are obvious in Tables II and III, as the same sounds elicited different mappings and strength among the participant groups. In addition, the presence of a consonant in the CV condition can alter the colour associations for some vowels even for the same group of participants, as the presence of a consonant can affect the acoustic properties of the vowels.

Second, we examined how the association between vowel and perceived lightness in colours was affected by tone, $F2$, and language background. It is generally observed in Table II that darker colours were more likely to be chosen with a low tone (T4). The mean $F0$ value of T1 is 261 Hz (SD 6.4 Hz), while that of T4 is 162 Hz (SD 5.9 Hz), collapsed across all experimental stimuli. The average difference between the two tones was about 100 Hz. The participants rated the lightness of the 11 colours on a 7-point Likert scale (1 being extremely light and 7 being extremely dark). As five Mandarin speakers did not do the lightness rating, their data were excluded. Figure 2 shows that the lightness ratings of the three participant groups were very similar across colours and across vowels, although small differences can also be found. Mean $F2$ for each vowel is given in Table IV. $F2$ formant frequencies were tracked automatically at the midpoints of the vowels in both V-only and CV conditions using a Praat script. Manual checking of tracked $F2$ formant values was conducted, and correction was done for anomalies.

Mixed effects linear regressions were performed using the lmer function in the lme4 package (Bates *et al.*, 2015) on *R* (R Core Team, 2018) to explore the effects of $F2$, tone, and language background on lightness ratings. We substituted participants' colour choices for the sound stimuli with their lightness ratings for this analysis. The base model included language (Cantonese, English, Mandarin), log $F2$, tone (T1 vs T4), condition (V vs CV) as fixed effects, by-item and by-speaker random intercepts and by-speaker random slopes for vowel (/i y ɛ œ a ɔ u/). All categorical factors were treatment coded. Two-way interactions between the fixed effects were tested through model comparison; the addition of interaction language × tone significantly improved the model, but others did not. Results in Table V(a) suggest that higher $F2$ was associated with lower lightness ratings, and that T4 elicited higher lightness rating than T1. The effect of tone was significantly larger on English ($p < 0.001$) and Mandarin ($p < 0.001$) speakers than on Cantonese speakers. The average lightness difference between the two tones was larger for Mandarin (T1: 3.44, T4: 3.98, a difference of 0.54 in the Likert scale) and English (T1: 3.48, T4: 3.98, a difference of 0.5 in the Likert scale) speakers than Cantonese speakers (T1: 3.59, T4: 3.88, a difference of 0.29 in the Likert scale).

To tease apart the effects of frontness and roundedness, separate models were built with frontness or roundedness replacing log $F2$ in the models. Specifically, frontness (front vs back) and roundedness (rounded vs unrounded) were introduced as binary categorical factors. In the frontness-lightness rating models, vowels /i, ɛ, y, œ, a/ were coded as front and vowels /u, ɔ/ were coded as back. In the roundedness-lightness rating models, vowels /y, œ, u, ɔ/ were labelled as rounded, and vowels /i, ɛ, a/ were labelled as unrounded. Similarly, two-way interactions between the fixed effects were tested through model comparison. Again, in both analyses, the interaction between language × tone was significant, while others were not. Results from both the frontness-lightness rating [Table V(b)] analysis and the roundedness-lightness rating [Table V(c)] analysis corroborate findings from the $F2$-lightness rating analysis: front vowels, in comparison with back vowels, were associated

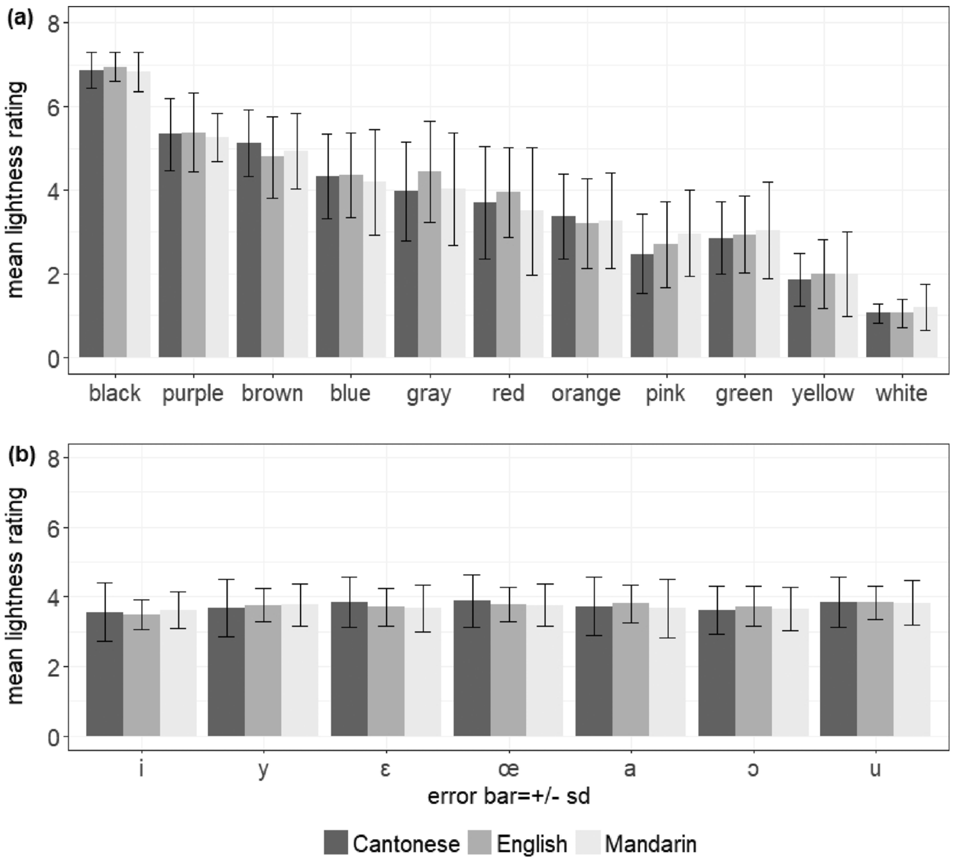

FIG. 2. Lightness ratings of the three language groups (a) across colours and (b) across vowels (1 = extremely light, 7 = extremely dark).

with lower lightness ratings (a decrease of 0.019 in rating). Rounded vowels, in comparison with unrounded vowels, were associated with higher lightness ratings (an increase of 0.081 in rating); the effect of a lower tone (T4) was stronger for English ($p < 0.001$) and Mandarin ($p < 0.001$) speakers than for Cantonese speakers.

In brief, these regression analyses consistently indicate that a high tone (T1) or a higher $F2$ (owing to frontness or unroundedness) was associated with colours that were perceived to be lighter; a low tone (T4) or a lower $F2$ (owing to backness or roundedness) was associated with colours that were perceived to be darker. Interestingly, the magnitude of the tone effect, as indicated by the coefficients, is smaller for the Cantonese speakers than for the English and Mandarin speakers.

IV. DISCUSSION

Our data clearly demonstrate non-random associations between vowels and colours in non-synaesthetic populations

TABLE IV. Mean $F2$ formant frequencies (Hz) and standard deviations of the seven vowels in two lexical tones.

vowels	Tone 1 [55]	Tone 4 [21]
i	2990 (74)	3059 (76)
y	2254 (111)	2342 (157)
ɛ	2330 (122)	2355 (96)
œ	1688 (106)	1735 (78)
a	1614 (70)	1602 (94)
ɔ	917 (84)	947 (84)
u	758 (43)	807 (73)

speaking three different languages. Our data concur well with previous findings that the associations between /a/ and red, /i/ with light colours, and /u/ with dark colours can be found across a variety of language backgrounds. A robust pitch effect is observed with a high tone eliciting lighter colours and a low tone eliciting darker colours in general. To the best of our knowledge, this was the first study on vowel-colour association involving Chinese (tone language) speakers (Simner *et al.*, 2011; Hung *et al.*, 2014 were on grapheme-colour association) and using pitch difference produced by the same voice. Our results confirm the universality of this cross-modal association by extending the findings to new languages.

In addition to the universality of vowel-colour association, our findings additionally show the language-specificity of this cross-modal association. Different colours were chosen by the language groups (Table II), and there were more robust associations for Cantonese speakers (native) than for Mandarin and especially English speakers (foreign) (Table III), which suggests that the strength of association varies according to language backgrounds. Moreover, in the regression analyses on lightness rating (Table V), consistent significant language difference was found between English and Mandarin vs Cantonese. All three language groups gave higher lightness ratings to a high tone, but the actual difference between the high and low tones was much smaller in the native Cantonese group than in the other two foreign groups. The role of learning in the cross-modal association is evident.

One aspect of the language-specificity remains unclear. We expected the tone difference to be more robust for Cantonese and Mandarin speakers (both tonal) than English

TABLE V. Results of linear mixed effects models on vowels and perceived lightness.

Predictor	Estimate	Standard error (SE)	t.value	p value
(a) Estimates of fixed effects: Model for analysing F2 and lightness ratings				
(Intercept)	4.175	0.287	14.565	<0.001
Language = English	−0.099	0.135	−0.729	0.466
Language = Mandarin	−0.115	0.140	−0.823	0.410
Log *F2*	−0.097	0.037	−2.650	0.008
Tone = T4	0.305	0.038	7.999	<0.001
Condition = CV	0.131	0.032	4.129	<0.001
Language = English × tone = T4	0.195	0.048	4.037	<0.001
Language = Mandarin × tone = T4	0.240	0.050	4.830	<0.001
(b) Estimates of fixed effects: Model for analysing frontness and lightness ratings.				
(Intercept)	3.473	0.101	34.395	<0.001
Language = English	−0.099	0.135	−0.729	0.466
Language = Mandarin	−0.115	0.140	−0.823	0.411
Front vowel = true	−0.019	0.037	−0.509	0.611
Tone = T4	0.303	0.038	7.918	<0.001
Condition = CV	0.130	0.032	4.077	<0.001
Language = English × tone = T4	0.195	0.048	4.038	<0.001
Language = Mandarin × tone = T4	0.240	0.050	4.829	<0.001
(c) Estimates of fixed effects: Model for analysing roundedness and lightness ratings				
(Intercept)	3.408	0.101	33.620	<0.001
Language = English	−0.099	0.135	−0.728	0.467
Language = Mandarin	−0.115	0.140	−0.822	0.411
Rounded vowel = true	0.081	0.037	2.192	0.028
Tone = T4	0.303	0.038	7.921	<0.001
Condition = CV	0.130	0.032	4.080	<0.001
Language = English × tone = T4	0.195	0.048	4.037	<0.001
Language = Mandarin × tone = T4	0.240	0.050	4.828	<0.001

speakers (non-tonal). The finding of a general tone-lightness effect across language groups itself is not surprising, as robust pitch-lightness effect was found in previous literature, but the smaller lightness difference between high and low tones for Cantonese speakers was unexpected. It was hypothesized that the difference would be smaller for English speakers, as they would not be as sensitive to tonal variations as tone language speakers (Cantonese and Mandarin) were. The results suggest otherwise. Both foreign groups (Mandarin and English), irrespective of whether lexical tone is used in their languages, responded to the tones with a larger lightness difference than did the native Cantonese group.

Perhaps the results can be interpreted from a different perspective. It is reasonable to assume that regardless of language background, the acoustic or psychophysical difference between the two tones (around 100 Hz) was noticeable to all three groups, as it is much larger than the just noticeable difference in pitch (Liu, 2013). As the two tones were familiar to Cantonese speakers, they did not find them very novel and, thus, felt that they did not differ as much in lightness. The two tones were clearly foreign to English speakers, and the acoustic difference between the two tones (∼100 Hz) is much larger than what English speakers are accustomed to hearing in English stressed vs unstressed syllables (∼20 Hz; Zhang *et al.*, 2008), so they felt that they sounded rather different. As for Mandarin speakers, although lexical tones were not foreign to them, the low tone T4 was because there is not such a tone in the Mandarin tone inventory. As a result, the two foreign groups responded with a larger lightness difference than the native Cantonese group did.

The above argument has parallels in major L2 speech acquisition theories. The Speech Learning Model (SLM; Flege, 1995), Perceptual Assimilation Model (PAM; Best, 1995), and the Second Language Linguistic Perception (L2LP) model (Escudero, 2009) all proposed that not all non-native sounds were of equal difficulty. Listeners perceive non-native sounds with reference to their native sounds, and phonetic similarities between non-native and native sounds are essential in predicting assimilation or discrimination performance of non-native sounds. All models suggested that if a non-native sound is phonetically similar to a native sound, then the two sounds are more difficult to discriminate (the mechanism of equivalence classification) than a non-native sound that is phonetically different from a native sound. Hence, a foreign language with many exotic sounds may become phonetically more distinguishable from the native language than one with fewer such sounds. Consequently, the heightened distinguishability alerts the learners to the peculiarities of these foreign sounds and, thus, may enhance learning in terms of orienting their attention to the sound contrasts and overall sound quality of the foreign language. The result could be, paradoxically, increased sensitivity to these foreign sound contrasts. In the present study, as the Cantonese tonal difference was not at all similar to the native sounds for Mandarin and English speakers, they might feel a larger lightness difference between the tones than Cantonese speakers did.

Some studies demonstrated that synaesthetic associations appear to be inborn, e.g., infants and young children were found to have cross-modal associations between pitch and lightness and shape (e.g., Mondlock and Maurer, 2004; Walker *et al.*, 2010). Martino and Marks (2001) suggested that cross-modal correspondences are both inborn and learned. They considered infants' matching of loudness-brightness and pitch-position as inborn, and other cross-modal correspondences between pitch and visual size developing over time as learned. Our findings supported the ideas of Martino and Marks (2001), but from a different perspective. The same association between certain vowels and certain colours found across language backgrounds shows the inborn property of synaesthetic association, while the different association strengths found among language groups illustrate the learned aspects of cross-modal perception. The language-specific association patterns found in our non-synaesthetic populations reinforce Simner's (2007) argument that linguistic synaesthesia is a psycholinguistic phenomenon, which is conceptually mediated. She highlighted the similarity between the development of linguistic synaesthesia and language acquisition in general. Humans have an innate predisposition to acquire language, but environmental factors are also influential during language development. Our cross-linguistic data can illustrate this connection very well.

Our results allow comparison of the relative importance of vowel quality and pitch difference in colour association. When individual vowels were considered (Fig. 1 and Table II), the importance appears to vary among vowels and among languages. For example, some vowels like /i a u/ have strong colour associations, while others are more affected by the pitch difference. Some conditions related to tone were not significant for English speakers. However, when the lightness rating regression analyses were considered (Table V), the pitch-lightness effect was very robust for all three groups of speakers. A low tone elicited darker colours than a high tone did consistently across vowel categories. Thus, it can be concluded that the pitch-lightness association is more robust than the vowel-colour mapping.

Previous studies suggested that the frequency of the second formant ($F2$) is significantly correlated with the light-dark dimension of colour (Marks, 1975; Moos *et al.*, 2014). The front unrounded vowel /i/ is light because it has a high $F2$, while the back rounded vowel /u/ is dark because its $F2$ is low. Nevertheless, they have not considered the two categorical phonological dimensions of vowel frontness and rounding. Our data show that the light-dark dimension is indeed related to $F2$. It is influenced by both vowel frontness and rounding. It is especially interesting to find that front rounded vowels were also darker than front unrounded vowels, so it is the continuous acoustic dimension of $F2$, instead of the categorical phonological distinctions of frontness/roundedness, which affects lightness perception. Since we are the first to try to tease apart these dimensions, it will be necessary to extend the investigation to more languages both with and without rounding contrast to corroborate our findings.

Previous studies on sound-colour mappings used only single vowels. So far, no study has systematically evaluated the sound-colour mappings of consonants. Our findings illustrated that consonants (the CV-condition) can indeed influence sound-colour mappings of vowels (V-only condition), probably because the presence of consonants both add extra acoustic information to, as well as alter some acoustic properties of vowels due to coarticulation. Some studies have shown cross-modal associations between consonants and size/shape (Shinohara and Kawahara, 2010; D'Onofrio, 2013). It is possible that there exist some non-random mappings between (properties of) consonants and colours. Further study can follow up on this possibility.

Still there remains the question as to why linguistic sounds should at all be associated with colours. As noted by Lockwood and Dingemanse (2015), many studies on sound symbolism have focused on showing that there is an effect, while the next necessary step is to explain how the effect works. Recent studies inform us that synaesthesia may have a genetic and neurophysiological basis (see the review in Simner and Hubbard, 2013), while the language-specific aspects of sound-colour mappings also tell us that learning and experience are important. The argument for a pivotal role of learning and experience in speech-colour synaesthesia is corroborated by several observations. First, according to Simner (2007), what triggers colours are abstract linguistic categories, not their concrete symbols. Hence, it is self-evident that speech-colour correspondence should reflect some learned relationship between colours and these abstract categories, which are themselves learned, rather than any "hard-wired" links between colours and the physical properties of concrete linguistic symbols. Second, linguistic frequency (how often a particular language unit is encountered in the linguistic environment) of the inducer is found to be a pervasive factor in various language-triggered synaesthesias (see the review in Simner and Hubbard, 2013). Frequent inducers are generally paired with frequent concurrents and infrequent inducers with infrequent concurrents (e.g., Rich *et al.*, 2005; Simner *et al.*, 2005; Beeli *et al.*, 2007; Hung, 2011). While the exact mechanism behind this frequency effect remains unknown, the influence of the individual's exposure to or experience with the linguistic items and colours on their pairings seems obvious and hard to ignore. Third, in their investigation on pitch-size correspondence, Brunetti *et al.* (2018) demonstrated that a medium-pitch tone when preceded by a high tone (thus, relatively low) was more likely to be paired with a large disk, whereas the same medium-pitch tone when preceded by a low tone (thus, relatively high) was more likely to be paired with a small disk. Hence, it was the individual's subjective conceptualisation of "high" vs "low" that mattered, not the absolute pitch of the tone. This observation suggests that cross-modal correspondences are relative in nature and highly sensitive to the individual's experience with the immediate environment. Taken together, the above observations provide some good support for a role of learning and experience in language-related synaesthesias and cross-modal correspondences in general.

In the literature it has been well documented that information from the different perceptual senses is integrated at a central level (Calvert and Thesen, 2004; Beauchamp, 2005; Koelewijn *et al.*, 2010; Mahoney *et al.*, 2011), and central neural activities for integration may then feed back to the more peripheral circuits and modify how they process further incoming information (Lugo *et al.*, 2008). Although the integration process is by and large automatic, it has also been shown that it can interface with non-automatic processes (Koelewijn *et al.*, 2010). The fact that cross-modal information integration communicates with our consciousness is viewed as adaptive because piecemeal perceptual information is, thus, bound to form coherent conscious experiences that better match reality. The vowel-colour association demonstrated in the present study with non-synaesthetes may provide just another example of how long-term cross-modal integration may give rise to conscious perceptual bias involving multiple senses.

Finally, there are some limitations in our study. First, only 11 focal colours were used, although some relevant studies also used 11 colours only (e.g., Rich *et al.*, 2005; Simner *et al.*, 2005; Wrembel, 2009; Root *et al.*, 2018). A wider choice of colour would allow a more comprehensive investigation. Second, some recent studies used objective measures to determine lightness/luminance of the colours (e.g., Moos *et al.*, 2014; Hamilton-Fletcher *et al.*, 2017), while we did so subjectively by asking the participants to rate the lightness of the colours. Although we believed that subjective lightness as perceived by the participants would allow us to assess the effects of lightness perception more directly, more objective measurements of the presented colours' luminance as measured by a chromameter can corroborate our findings. Third, no synaesthetes were included in our study. Although it is argued that cross-modal association of synaesthetes and non-synaesthetes share common mechanisms, the mappings for synaesthetes are more consistent and absolute than those for non-synaesthetes (Simner *et al.*, 2011; Moos *et al.*, 2014). It is possible that the pitch effect found in our study may be overridden by the more absolute mapping between vowels and colours in synaesthetes. Additionally, it will be good to further investigate the effects of language backgrounds on sound-colour association comparing synaesthetes and non-synaesthetes.

In conclusion, our study both confirmed and extended previous findings of vowel-colour associations. The language-specificity of this cross-modal association is highlighted. Future studies should explore this interesting aspect more vigorously using both synaesthetes and non-synaesthetes with improved methods. Only by doing this can we have a thorough understanding of the enigmatic phenomena of synesthesia.

ACKNOWLEDGMENTS

The authors would like to thank Lan Chen, Yanjun Yin, and Iris Jingxin Luo for their contributions to the project. They also acknowledge the Department of Linguistics and Modern Languages at The Chinese University of Hong Kong for funding support.

[1]See supplementary material at https://doi.org/10.1121/1.5096632 for the RGB values of 11 colours and detailed *post hoc* comparisons for language differences.

Baron-Cohen, S., Harrison, J., Goldstein, L. H., and Wyke, M. (**1993**). "Coloured speech perception: Is synaesthesia what happens when modularity breaks down?," Perception **22**, 419–426.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**(1), 1–48.

Beauchamp, M. S. (**2005**). "See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex," Curr. Opin. Neurobiol. **15**, 145–153.

Beeli, G., Esslen, M., and Jancke, L. (**2007**). "Frequency correlates in grapheme-color synaesthesia," Psychol. Sci. **18**, 788–792.

Berlin, B., and Kay, P. (**1969**). *Basic Color Terms: Their Universality and Evolution* (University of California Press, Berkeley, CA).

Best, C. T. (**1995**). "A direct realist perspective on cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Brunetti, R., Indraccolo, A., Del Gatto, C., Spence, C., and Santangelo, V. (**2018**). "Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification," Atten. Percept. Psychophys. **80**, 527–534.

Calvert, G. A., and Thesen, T. (**2004**). "Multisensory integration: Methodological approaches and emerging principles in the human brain," J. Physiol. **98**, 191–205.

Day, S. (**2005**). "Some demographic and socio-cultural aspects of synesthesia," in *Synesthesia: Perspectives from Cognitive Neuroscience*, edited by L. C. Robertson and N. Sagiv (Oxford University Press, Oxford), pp. 3–10.

de Thornley Head, P. (**2006**). "Synaesthesia: Pitch-colour isomorphism in RGB-space?," Cortex **42**, 164–174.

D'Onofrio, A. (**2013**). "Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm," Lang. Speech **57**, 367–393.

Escudero, P. (**2009**). "The linguistic perception of similar L2 sounds," in *Phonology in Perception*, edited by P. Boersma and S. Hamann (Mouton de Gruyter, Berlin), pp. 151–190.

Fernay, L., Reby, D., and Ward, J. (**2012**). "Visualized voices: A case study of audio-visual synesthesia," Neurocase **18**, 50–56.

Flege, J. (**1995**). "Second-language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 229–273.

Guillamón, P. M. (**2013**). "Vowel-colour symbolism in English and Arabic: A comparative study," Misc.: J. Engl. Am. Stud. **47**, 31–52.

Hamilton-Fletcher, G., Witzel, C., Reby, D., and Ward, J. (**2017**). "Sound properties associated with equiluminant colours," Multisens. Res. **30**, 337–362.

Hubbard, T. L. (**1996**). "Synesthesia-like mappings of lightness, pitch, and melodic interval," Am. J. Psychol. **109**, 219–238.

Hung, W. Y. (**2011**). "An investigation into the underlying linguistic cues of Chinese synaesthesia," Ph.D. thesis, University of Edinburgh.

Hung, W. Y., Simner, J., Shillcock, R., and Eagleman, D. M. (**2014**). "Synaesthesia in Chinese characters: The role of radical function and position," Conscious. Cogn. **24**, 38–48.

Jakobson, R. (**1962**). *Selected Writings: Phonological Studies* (Mouton, The Hague).

Koelewijn, T., Bronkhorst, A., and Theeuwes, J. (**2010**). "Attention and the multiple stages of multisensory integration: A review of audiovisual studies," Acta Psychol. **134**, 372–384.

Liu, C. (**2013**). "Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners," J. Acoust. Soc. Am. **134**, 3011–3020.

Lockwood, G., and Dingemanse, M. (**2015**). "Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism," Front. Psychol. **6**, 1246.

Lugo, J., Doti, R., Wittich, W., and Faubert, J. (**2008**). "Multisensory integration: Central processing modifies peripheral systems," Psychol. Sci. **19**, 989–997.

Maddieson, I. (**1984**). *Patterns of Sounds* (Cambridge University Press, Cambridge, UK).

Mahoney, J. R., Li, P. C., Oh-Park, M., Verghese, J., and Holtzer, R. (**2011**). "Multisensory integration across the senses in young and old adults," Brain Res. **1426**, 43–53.

Marks, L. E. (**1974**). "On associations of light and sound: The mediation of brightness, pitch, and loudness," Am. J. Psychol. **87**, 173–188.

Marks, L. E. (**1975**). "On colored-hearing synesthesia: Cross-modal translations of sensory dimensions," Psychol. Bull. **82**, 303–331.

Marks, L. E. (**1989**). "On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness," J. Exp. Psychol. Hum. Percept. Perform. **15**, 586–602.

Marks, L. E. (**2013**). "Weak synesthesia in perception and language," in *The Oxford Handbook of Synesthesia*, edited by J. Simner and E. M. Hubbard (Oxford University Press, Oxford, UK), pp. 762–789.

Marks, L. E., Ben-Artzi, E., and Lakatos, S. (**2003**). "Cross-modal interactions in auditory and visual discrimination," Int. J. Psychophysiol. **50**, 125–145.

Marks, L. E., Hammeal, R. J., and Bornstein, M. H. (**1987**). "Perceiving similarity and comprehending metaphor," Monogr. Soc. Res. Child Dev. **52**(1), 1–102.

Martino, G., and Marks, L. E. (**2001**). "Synesthesia: Strong and weak," Curr. Dir. Psychol. Sci. **10**, 61–65.

Melara, R. D. (**1989**). "Dimensional interaction between color and pitch," J. Exp. Psychol. Hum. Percept. Perform. **15**, 69–79.

Mondlock, C. J., and Maurer, D. (**2004**). "Do small white balls squeak? Pitch-object correspondences in young children," Cogn. Affect. Behav. Neurosci. **4**(2), 133–136.

Moos, A. C., Simmons, D., Simner, J., and Smith, R. (**2013**). "Color and texture associations in voice-induced synesthesia," Front. Psychol. **4**, 568.

Moos, A. C., Smith, R., Miller, S. R., and Simmons, D. (**2014**). "Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder," i-Perception **5**, 132–142.

Mroczko-Wąsowicz, A., and Nikolić, D. (**2013**). "Colored alphabets in bilingual synesthetes," in *The Oxford Handbook of Synesthesia*, edited by

J. Simner and E. M. Hubbard (Oxford University Press, Oxford), pp. 165–180.

R Core Team (**2018**). "R: A language and environment for statistical computing" (R Foundation for Statistical Computing, Vienna).

Rich, A., Bradshaw, J., and Mattingley, J. S. (**2005**). "A systematic, large-scale study of synaesthesia: Implications for the role of early experience in lexical-colour associations," Cognition **98**, 53–84.

Root, N. B., Rouw, R., Asano, M., Kim, C. Y., Melero, H., Yokosawa, K., and Ramachandran, V. S. (**2018**). "Why is the synesthete's 'A' red? Using a five-language dataset to disentangle the effects of shape, sound, semantics, and ordinality on inducer-concurrent relationships in grapheme-color synesthesia," Cortex **99**, 375–389.

Ryalls, J. H. (**1986**). "Synesthesia: A principle for the relationship between the primary colors and the cardinal vowels," Semiotica **58**, 107–121.

Shinohara, K., and Kawahara, S. (**2010**). "A cross-linguistic study of sound symbolism: The images of size," in *Berkeley Linguistics Society*, pp. 396–410.

Simner, J. (**2007**). "Beyond perception: Synaesthesia as a psycholinguistic phenomenon," Trends Cogn. Sci. **11**, 24–29.

Simner, J. (**2012**). "Defining synaesthesia," Br. J. Psychol. **103**, 1–15.

Simner, J., and Hubbard, E. M. (eds.) (**2013**). *The Oxford Handbook of Synesthesia* (Oxford University Press, Oxford, UK).

Simner, J., Hung, W. Y., and Shillcock, R. (**2011**). "Synaesthesia in a logographic language: The colouring of Chinese characters and Pinyin/Bopomo spellings," Conscious. Cogn. **20**, 1376–1392.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K., and Ward, J. (**2006**). "Synaesthesia: The prevalence of atypical cross-modal experiences," Perception **35**, 1024–1033.

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., and Oakley, D. A. (**2005**). "Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations," Cogn. Neuropsych. **22**, 1069–1085.

Walker, P., Bremner, G., Mason, U., Spring, J., Mattock, K., Slater, A., and Johnson, S. P. (**2010**). "Preverbal infants' sensitivity to synaesthetic cross-modality correspondences," Psychol. Sci. **21**(1), 21–25.

Ward, J., and Cytowic, R. (**2010**). "Synesthesia and language," in *Concise Encyclopedia of Brain and Language*, edited by H. Whitaker (Elsevier, Oxford, UK), pp. 495–500.

Ward, J., Huckstep, B., and Tsakanikos, E. (**2006**). "Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?," Cortex **42**, 264–280.

Ward, J., and Mattingley, J. S. (**2006**). "Synaesthesia: An overview of contemporary findings and controversies," Cortex **42**, 129–136.

Wrembel, M. (**2007**). "Still sounds like a rainbow—A proposal for a coloured vowel chart," in *Proceedings of the Phonetics Teaching and Learning Conference PTLC2007*, London, pp. 1–4.

Wrembel, M. (**2009**). "On hearing colours—Cross-modal associations in vowel perception in a non-synaesthetic population," Poznań Stud. Contemp. Linguist. **45**, 581–598.

Zhang, Y., Nissen, S. L., and Francis, A. L. (**2008**). "Acoustic characteristics of English lexical stress produced by native Mandarin speakers," J. Acoust. Soc. Am. **123**, 4498–4513.