Psychology & Marketing

WILEY

RESEARCH ARTICLE OPEN ACCESS

Connotative Congruency Effect Between Instrumental Timbre and Visual Design Features on Consumer Decision-Making

Tsutomu Sunaga[1] | Takeshi Moriguchi[1] | Mayuko Nishii[2] | Charles Spence[3]

[1]Faculty of Commerce, Waseda University, Tokyo, Japan | [2]Graduate School of International Management, Aoyama Gakuin University, Tokyo, Japan | [3]Department of Experimental Psychology, University of Oxford, Oxford, UK

Correspondence: Tsutomu Sunaga (sunaga@waseda.jp)

Received: 12 August 2024 | **Revised:** 9 April 2025 | **Accepted:** 12 April 2025

Funding: This work was supported by JSPS KAKENHI Grant Numbers JP20K02002 and JP19H00601.

Keywords: audio-visual congruency | connotative meaning | musical instrument | sensory marketing | timbre

ABSTRACT

Conducting online, controlled laboratory, and more realistic laboratory experiments, this study investigates how the connotative congruency between instrumental timbre and visual design affects consumer evaluation and choice behavior. The results demonstrate that consumers are more likely to respond positively to the target when the instrumental timbre of background music shares connotative meanings like bright, smooth, or heavy with a visual feature of the target. In addition, we identified that the connotative congruency effect is mediated by the sense of feeling right and excitement (i.e., high arousal positive emotion). A carefully-controlled Study 3 and Study 5, which measured behavioral outcomes in a more realistic mock-up corner drugstore, enhanced the ecological and external validity of the timbre-vision connotative congruency effects. This is the first empirical marketing study on timbre-vision congruency that covers comprehensive sets of connotative descriptions and reveals the mediating role of both feeling right and excitement.

1 | Introduction

Timbre, which refers to the quality of a sound (Wallmark 2019), is one of the most fundamental aspects of acoustic communication (Saitis et al. 2020). However, compared to other musical elements, such as pitch (Hagtvedt and Brasel 2016; Huang and Labroo 2020; Lowe and Haws 2017), loudness (Biswas et al. 2019), or tempo (Knoferle et al. 2012), marketing scholars, with a few rare exceptions (e.g., Oakes and North 2006), have typically not given much thought to timbre until recently, despite it being an important acoustic property (Puligadda and VanBergen 2023).

Studies investigating how timbre influences consumer behavior have recently started to become more popular: For instance, Melzner and Raghubir (2023) investigated how the roughness/smoothness of sound logos, which refers to a short melody created to represent a brand and convey the identity (Spence and Keller 2024; Techawachirakul et al. 2023), influenced the perceptions of ruggedness/sophistication in brand personality. Similarly, Puligadda and VanBergen (2023) demonstrated that incorporating the piano timbre in a brand's sound logo enhanced the perceived sophistication (vs. ruggedness) of a brand's personality using the flute, trumpet, piano, ukulele, and steel drums, to represent woodwinds, brass, percussion, and stringed instruments, respectively, as experimental stimuli. Meanwhile, Techawachirakul et al. (2023) found that compared to sound logos played by "feminine" instrumental timbres (including clarinet, flute, oboe, and violin timbres), US participants associated sound logos played by "masculine" ones

(including alto saxophone, trombone, trumpet, and tuba timbres) with meat-based (vs. plant-based) foods.

Although these studies provide some important findings, there are still many unresolved issues in timbre perception research. For instance, all three of the studies used sound logos as audiovisual experimental stimuli and conducted online studies using Amazon Mechanical Turk, meaning that it is unclear whether the findings are applicable to other visual stimuli than sound logos or in a real-world setting. Besides, in those studies, no actual behavior was measured. Instead, purchase intention (Melzner and Raghubir 2023) and interest in trying (Puligadda and VanBergen 2023) were assessed as downstream effects of timbre. Although Puligadda and Bergen identified a mediating role of processing fluency in the effect of congruent visual and sound logos on interest in trying the target product, it still remains unclear "how" the sensory congruency between timbre and visual elements generates such downstream effects. Lastly, Melzner and Raghubir (2023), Puligadda and VanBergen (2023), and Techawachirakul et al. (2023) investigated sound logo-vision congruency using limited sets of connotative attributes; roughness (ruggedness)/smoothness (sophistication) or gender (masculine/feminine). However, prior research (e.g., Saitis and Weinzierl 2019) has shown that timbres are not only associated with texture (e.g., smooth or rough), but also with luminance (e.g., bright or dim) and mass (e.g., heavy or light).

The present study was designed to investigate how consumers perceive instrumental timbres and materials displayed visually using more comprehensive sets of connotative descriptions (i.e., luminance, texture, and mass) and various visual stimuli. Second, to enhance the ecological validity of timbre effects, downstream effects of the sensory congruency on consumers' decision-making were measured by means of not only self-report such as evaluations and intentions, but also actual choice behaviors. This study defines sensory congruency as the degree to which the same connotative meanings (e.g., bright, smooth, and/or heavy) are shared between different sensory modalities (e.g., timbre and color). Lastly, this study investigated and identified mediators and moderators indicating how sensory congruency influences consumer decision-making.

In this study, timbre is operationally-defined as the sound quality of a particular musical instrument. That is, the auditory stimuli were exactly the same excerpts but played by different musical instruments thus they had different timbres. On the other hand, the visual features were manipulated in several ways, such as the qualities of material having the same color (Study 1), the typefaces with the same font size displayed in grayscale letters against a white background (Study 2), the colors of ballpoint pen with no perceived difference except for the ink color (Study 3) and the cushion in the advertisement (Study 4), and the product items with different flavors of the same brand (Study 5).

2 | Theoretical Background

2.1 | Audiovisual Crossmodal Correspondences

Crossmodal correspondence is the term that is increasingly being used to refer to the sometimes surprising consensual connections whereby people associate a feature, attribute, or dimension in one sensory modality, either physically present or merely imagined, with a seemingly unrelated sensory feature in a different sensory modality (Spence 2011, 2020a, 2020b). Crossmodal correspondences have been categorized into at least four distinct types: structural, statistical, semantic, and hedonic (Spence 2011, 2022). Structural (more recently renamed as physiological) correspondences (see Spence and Di Stefano 2022) are derived from common neural encoding (e.g., louder sound and bright color). Statistical correspondences are thought to result from the internalization of the crossmodal regularities of the environment or physical correlation (e.g., high-pitched sound and small size). Semantic or lexical (Spence 2022; Spence and Di Stefano 2022; Walker 2012) correspondences come from the same description of experiences involving different sensory modalities (e.g., high-pitched sound and high spatial location). Lastly, hedonic or emotionally-mediated correspondences originate in the congruence of the emotion (mood), or hedonic valence induced by, or associated with, inputs from different sensory modalities (e.g., sweet taste and rounded shape both being liked stimuli, see Spence 2020a, 2022). It should, however, be noted that these various types/explanations for the existence of crossmodal correspondences need not be treated as mutually exclusive and, in fact, a number of different explanations may jointly be involved in any given correspondence, and/or different explanations may apply to explain different crossmodal correspondences.

A large number of studies from divergent disciplines have uncovered a wide range of crossmodal correspondences between the attributes of stimuli presented in different sensory modalities (Spence 2011, 2020a, 2020b; Spence and Di Stefano 2022). Amongst them, the number of scientific research studies that have focused on crossmodal correspondences between musical attributes, including timbre and diverse visual elements, continues to grow (Anikin and Johansson 2019; Qi et al. 2020; Spence 2020a; see Table 1). For instance, Crisinel and Spence (2010) demonstrated that people exhibit a crossmodal correspondence between sweet tastes and the timbre of the piano, and between bitter tastes and the timbre of a (synthetic) trombone (see also Crisinel and Spence 2012a).

Audition and vision are similar in several important respects (see Julesz and Hirsh 1972; Spence and Di Stefano 2022). First, both have the characteristic of waves, although they are of different types. Second, both auditory and visual cues can be obtained even from far away compared to olfaction, taste (gustation), and touch (Hall 1966; Ruzeviciute et al. 2020). In other words, audition and vision are distal senses, whereas olfaction, taste, and touch are primarily proximal senses (Elder et al. 2017; Rodaway 1994). From the perspective of "stimulus belongingness," audition and vision can both be categorized as external senses, while olfactory, taste (gustation), and touch are internal senses. According to Rozin and Kalat (1972), organisms may find it easier to associate stimuli from two senses if they are both "external" (e.g., vision and auditory) than to associate (or recognize as belonging together) one external and one internal sense (such as a visual element, or stimulus, and a gustatory or tactile one). In addition, given that humans are visually dominant, one of the two senses in most crossmodal associations is likely to be visual (Elder and Krishna 2022; Hutmacher 2019).

TABLE 1 | Summary of prior studies on audiovisual correspondences.

References	Auditory property	Sound type	Visual property	Conclusions/findings
Adeli et al. (2014)	Timbre f0	Music	Shape, color, vertical position	Soft (harsh) timbres are associated with blue, green or light gray rounded shapes (red, yellow or dark gray sharp angular shapes). Timbres with both elements of softness and harshness are associated with a mixture of rounded and sharp angular shapes. Color or grayscale have no effect on timbre-shape associations. Fundamental frequency was not associated with height, grayscale or color.
Anikin and Johansson (2019)	Loudness, pitch, formants, spectral centroid, thrill	Vowel-like sound (400 ms)	Luminance, hue, saturation	Higher loudness, pitch, and spectral centroid are associated with higher saturation, while higher pitch and spectral centroid are associated with higher luminance. Meanwhile, alveolar trill sounds are perceived as having low luminance and saturation. The color hue with the same luminance and saturation does not have strong associations with any of the tested acoustic features.
Brunetti et al. (2018)	Pitch	Sinusoidal tone (300 ms)	Size	When a medium-pitch sound follows a lower-pitch sound, it is perceived as a relatively higher pitch and, as a result, congruent with a smaller object
Cheng et al. (2009)	Tempo	Music	Hue	Fast music alongside warm colors and slow music alongside cool colors enhance arousal and pleasure levels
Chiou and Rich (2012)	Pitch	Sinusoidal tone (200 ms)	Spatial location	Higher tones induce attention shifts to higher locations
Collier and Hubbard (2001)	Pitch, tempo	Tone	Brightness	Higher-pitched, faster tempo, and ascending sounds are perceived as brighter
Eitan and Granot (2006)	Dynamics, pitch contour, pitch intervals, attack rate, articulation	Music	Motion dimensions (spatial verticality, distance, speed	Musical abatements (intensifications) are associated with spatial descents (increasing speed). The musical-spatial analogies are often asymmetrical; pitch falls evoke stronger spatial descents, but pitch rises induce only weak spatial ascents.
Eitan et al. (2014)	Pitch	Sinusoid (1 s)	Size	Rise (fall) in pitch is associated with growing (shrinking) size
Evans and Treisman (2010)	Pitch	Pure/violin/piano tones (120 ms)	Vertical location, size, spatial frequency, contrast	A higher pitch is congruent with a higher visual position, smaller size, and higher spatial frequency, while contrast is not congruent with pitch. Congruency

(Continues)

TABLE 1 | (Continued)

References	Auditory property	Sound type	Visual property	Conclusions/findings
				between pitch and spatial position has the largest effects.
Hagtvedt and Brasel (2016)	Pitch	Tone, music	Color lightness	Higher frequency sounds guide visual attention toward and encourage the purchase of lighter-colored objects
Hamilton-Fletcher et al. (2017)	Pitch, loudness	Pure tone	Hue	Increases in pitch, as well as loudness, are associated with increases in chroma and shifts from blue to yellow. Sounds over 800 Hz are associated with yellow
Isbilen and Krumhansl (2016)	Tempo, mode, pitch, attack rate	Prelude from Bach's Well-tempered Clavier	Hue	Fast minor, fast major, slow minor, and slow major Preludes are associated with colors red, yellow, blue, and in the chartreuse-green-cyan range, respectively
Lidji et al. (2007)	Pitch	Tone with piano/violin timbre (500 ms)	Spatial location	Higher-pitched isolated tones are associated with higher positions regardless of musical expertise. Besides, only musicians associate both higher-pitched isolated tone and ascending melody interval with right-side and higher positions, respectively.
Lindborg and Friberg (2015)	Loudness, tempo, timbre, key, tonality	Film music	Size, hue, lightness	Music with a brighter and more variable spectrum is perceived as larger
Lowe and Haws (2017)	Pitch	Voice/music in an advertising	Size	Lower-pitched sounds are associated with larger product sizes
Melzner and Raghubir (2023, Study 5)	Timbre	Sound logo	Shape	Incongruency between the timbre of the audio and visual logos (i.e., a smooth-sounding audio logo alongside an angular visual logo or a rough-sounding audio logo alongside a round visual logo) attenuates the effect of timbre on brand personality perceptions
Palmer et al. (2016)	Tempo, note-density, mode, pitch	Piano music	Hue, lightness, saturation	Faster music in the major mode (vs. slower music in the minor mode) is associated with lighter, more saturated, yellower (warmer) colors
Parise and Spence (2009)	Pitch, source location	Pure tone (26 ms)	Size, horizontal location	Higher-pitched sounds are perceived as smaller. Spatial congruency between audio and visual stimuli enhances multisensory integration.
Puligadda and VanBergen (2023, Study 4)	Timbre	Sound logo	Design	Congruency between the timbre of a sound logo and the design of a visual logo (i.e., piano timbre alongside a sophisticated visual logo or synth bass alongside a rugged visual logo) makes assessing brand personality easier, which, in turn, leads to favorable brand evaluations

(Continues)

TABLE 1 | (Continued)

References	Auditory property	Sound type	Visual property	Conclusions/findings
Qi et al. (2020)	Instrument (timbre), pitch	Musical notes (1.5 s)	Color (hue)	Erhu (bow strings), dizi (winds), and yunluo (percussion) timbres are matched to black and red, brown, and pink and white, respectively. Higher-pitched sounds are matched to red (vs. black/gray) and brighter colors.
Rusconi et al. (2006)	Pitch	Pure tone (1 s)	Spatial height	Higher-pitched sounds are perceived as spatially higher
Sun et al. (2018)	Pitch, timbre, tempo, discontinuity	Piano music (20 s)	Hue, saturation, brightness	High- (low-) pitch is associated with red, yellow and light color (blue and dark color). Purple and orange (green and cyan) are associated with high (low) roughness. Rougher/sharper timbres are associated with lighter colors. Sharp timbres are associated with colorful colors.
Venkatesan et al. (2022)	Perceptual quality	Predicted/imagined Music	Curvilinearity, thickness	More angular (rounded) typefaces on a CD album cover induce expectations the music to sound more angular, masculine, fast, rough, happy, evil, violent, exciting, and active (feminine, slow, smooth, sad, good, gentle, calm, and passive)
Wagner et al. (1981)	Pitch	Tone	Spatial height	An ascending (descending) tone is associated with an up (down) arrow
Walker (2012)	Pitch	Verbal label (word)	Shape	Higher- (lower-)pitched sounds are perceived as angular/sharp (curved/smooth)
Walker et al. (2010)	Pitch, loudness	Sound of slide whistle (2.5 s)	Shape, movement	Higher-pitched sounds are perceived as spatially higher
Walker and Smith (1984)	Pitch	Pure tone	Size, shape, brightness, speed, spatial elevation	Higher-pitched sounds are perceived as lighter, smaller, sharper, and brighter
This study	Timbre	Background music	Visual design features	Timbre-vison connotative congruency enhances consumers' sense of feeling right and excitement, which, in turn, positively influences consumer evaluations and choices

2.2 | Instrumental Timbre

Oakes and North (2006) demonstrated that the conceptual congruency between timbre and message enhanced an advertisement's effectiveness. Specifically, the piano timbre was found to be more congruent with a radio ad for cosmetic surgery than the timbre of the church organ or the steel drum. The sound of the church organ was associated with religious events such as marriages and funerals, while the steel drum was associated with thoughts of the Caribbean (i.e., providing a sense of ethnicity). In Oakes and North (2006), the participants who were presented with background piano music exhibited better recall of the contents of the advertisement, such as the brand name and message. However, the researchers involved in the study provided no clear explanation as to why the piano timbre should have been considered as congruent with cosmetic surgery.

Meanwhile, Eitan and Rothschild (2011) found significant interactions with pitch, loudness, and vibrato. For example, people were more likely to associate the violin (vs. flute) timbre with smooth when dynamic levels of loudness were low (pianissimo), whereas they were more likely to associate the violin (vs. flute) timbre with hard when the dynamic levels of loudness were high (fortissimo) instead. Similarly, although the focus of their study was not on instrumental timbre, Imschloss and Kuehnl (2019) demonstrated that a congruent connotative association (softness) between music and haptics enhanced people's product evaluations. The results of these studies imply that in addition to the four distinct types of crossmodal correspondence identified by Spence (2011, 2022), that is, structural, statistical, semantic, and hedonic correspondences, there could be a fifth type: connotative correspondences. In this study, connotative correspondences are defined as a type of crossmodal correspondence based on the shared connotations between different sensory modalities, such as instrumental timbres and visual features.

Prior research has identified three primary categories of connotative associations with timbral qualities: Luminance (bright or dim), texture (smooth or rough), and mass (heavy or light) (Saitis and Weinzierl 2019; Saitis et al. 2020; Zacharakis et al. 2014, 2015). Table 2 shows the connotative associations of instrumental timbres that have been demonstrated by prior research.

According to Techawachirakul et al. (2023), instrumental timbres also have an association with gender; for example, people tend to perceive the sounds of the saxophone, trumpet, trombone, and tuba as masculine, while they perceive those of the flute, oboe, clarinet, and violin as feminine. These associations are thought to be established through statistical correlations; specifically, when players select an instrument to play, the former instruments are mainly chosen by men, while the latter are chosen by women (Techawachirakul et al. 2023). These co-occurrences lead people to form associations between instruments and foods; sound logos with a masculine or feminine instrument timbre are associated with meat- or plant-based food, respectively (Techawachirakul et al. 2023). It is noteworthy that the timbre associations focused on by Techawachirakul et al. (2023) were not based on the sounds themselves, but rather on their shared semantic association with gender.

Although connotative congruencies with timbral properties are undoubtedly relevant to the field of sensory marketing (e.g., Crisinel and Spence 2012b; Crisinel et al. 2013; Mahdavi et al. 2020), the prior marketing studies confined specific connotations of instrumental timbres instead of covering comprehensive connotative descriptions (i.e., luminance, texture, and mass). More importantly, the majority of prior studies have assessed timbre effects by means of self-report (i.e., by assessing participants' evaluations and intentions), thus investigating its ecological validity still remains to be tested. Understanding consumer behavior/psychology through examining the actual behavior in addition to self-reports has been called for decades, nonetheless, few studies have implemented it to date (e.g., Baumeister et al. 2007; Morales et al. 2017). The current study investigates whether and how the fifth type of audio-visual crossmodal correspondence, based on connotative congruencies, influences consumer behavior through various connotative descriptions.

2.3 | Psychological Mechanisms Underlying Sensory Congruency

2.3.1 | Emotion

Sensory (cue) congruence refers to the degree of fit among sensory elements/characteristics, reflecting how well a stimulus matches its environment across different senses (Krishna et al. 2010; Yuan et al. 2024). Eklund and Helmefalk (2022), who conducted a thorough literature review on perceived congruence amongst consumers, suggested that the results of prior studies demonstrate that congruency is multifaceted and induces various cognitive, affective, and behavioral responses. For instance, sensory congruence, rather than incongruence, tends to lead to more favorable product evaluations (see Krishna et al. 2010; Spence 2011). Many previous studies have identified emotion as a mediator of (at least a subset of) crossmodal correspondences, including certain of those involving auditory stimuli (Spence 2020a, 2020b). Note that “mediator” in those studies refers to a factor associating different sensory aspects and consequently making people form a crossmodal correspondence. For instance, using monophonic texture piano melodies based on melodies by Mozart and color patches as the stimuli, Palmer et al. (2016) demonstrated crossmodal associations between faster music in the major mode and lighter, more saturated, and warmer colors. Importantly, their data showed strong correlations between emotions (e.g., happy/sad, angry/not-angry and agitated/calm) towards the melodies and those towards the colors that the participants perceived as congruent with the melodies. In a similar vein, Lindborg and Friberg (2015) demonstrated that emotion was the dominant mediator behind associations between film music excerpts and the color/size of a color patch. Additional evidence of emotional mediation has been provided by Isbilen and Krumhansl (2016) who examined associations between music clips from Bach's the Well-Tempered Clavier, Books I and II with saturated colors.

The focus of the present study was on the mediating role in the relationship between sensory congruency and consumer responses. That is, mediators in this study refer to those factors explaining why sensory congruency typically generates a

TABLE 2 | Prior studies of connotative associations with instrumental timbres.

Instrument	Category	Connotative association
Keyboard		
Piano	Texture	Soft (Adeli et al. 2014)
Marimba	Texture	Soft (Adeli et al. 2014; Zacharakis et al. 2014)
	Luminance	Light (Zacharakis et al. 2015)
Vibraphone	Texture	Nasal (Kazazis et al. 2022)
Woodwind		
Clarinet	Texture	Nasal (Kazazis et al. 2022), Roundness (Zacharakis et al. 2014)
Flute	Luminance	Bright (Zacharakis et al. 2014)
	Texture	Smooth (Zacharakis et al. 2015), Soft (Saitis and Weinzierl 2019),
	Mass	Round (Zacharakis et al. 2014)
		Light (Zacharakis et al. 2014)
Oboe	Luminance	Bright (Kazazis et al. 2022; Saitis and Weinzierl 2019)
	Texture	Harsh (Saitis and Weinzierl 2019)
	Mass	Light (Zacharakis et al. 2014)
Saxophone	Luminance	Dim (Adeli et al. 2014)
	Texture	Soft (Adeli et al. 2014), Harsh (Adeli et al. 2014; McAdams 2019; Zacharakis et al. 2014)
	Mass	Thick (Zacharakis et al. 2014, 2015)
Stringed		
Violin	Luminance	Bright (Saitis and Weinzierl 2019)
	Texture	Less sharp, rough, hard, cold, and dry (Eitan and Rothschild 2011),
		Smooth (pianissimo) and Hard (fortissimo) (Eitan and Rothschild 2011)
Brass		
French horn	Luminance	Dark (Kazazis et al. 2022)
	Texture	Round (Zacharakis et al. 2014), Smooth (Zacharakis et al. 2015)
	Mass	Thick (Zacharakis et al. 2014)
Trumpet	Luminance	Bright (Zacharakis et al. 2014, 2015)
	Texture	Harsh (Kazazis et al. 2022; Zacharakis et al. 2014)
	Mass	Thick (Zacharakis et al. 2015)
Trombone	Luminance	Less bright (McAdams 2019)
Percussion		
Crash cymbals	Luminance	Dim (Adeli et al. 2014); Harsh (Adeli et al. 2014)
Gong	Luminance	Dim (Adeli et al. 2014); Harsh (Adeli et al. 2014)
Triangle	Luminance	Bright (Adeli et al. 2014); Harsh (Adeli et al. 2014)

favorable response amongst consumers. Sensory stimuli can enhance one another when they align within a multisensory experience (Yuan et al. 2024). Cheng et al. (2009) have demonstrated that presenting fast music and warm color simultaneously, as well as presenting slow music and cool color, enhanced the individuals' arousal and pleasure levels. The sense of pleasure emerged from sensory congruence, in turn, is attributed to the product evaluation (Fiske 1982). This is based on the theory of cognitive balance (Heider, 1958), which refers to a phenomenon in which the more harmonious or balanced situations are, the more favorably consumers respond (Flavián et al. 2021). Krishna et al. (2010) also asserted that the semantic congruence between inputs from different sensory modalities led to aesthetic pleasure. Meanwhile, Abdolmohamad Sagha et al. (2022) have shown that both sensory congruency and incongruency can enhance consumer emotions, affective experiences, and purchase intentions. The researchers demonstrated (in their Study 1) that consumers' emotions were more positive when the background color of the yellow product was incongruent (i.e., blue) than when it was congruent (i.e., yellow). In the stimuli used by Abdolmohamad Sagha et al. (2022), sensory congruence could create an ambiguous figure-ground relationship, making it difficult for their participants to recognize the target products, which could lead to their negative emotion Based on these prior studies, it is expected that the high arousal positive emotion, namely, excitement (Faseur and Geuens 2006; Kraiss et al. 2023), will mediate the effects of timbre-vison connotative congruence on favorable responses amongst consumers.

2.3.2 | Feeling Right

Another potential mediator is "feeling right." This refers to a subjective experience of feeling correct or proper that is elicited by two congruent stimuli (Kuo and Rice 2015;

Wang et al. 2022). Many consumer studies have demonstrated that a sense of "feeling right" induces favorable evaluation or judgment, amplifying positive reactions toward the evaluative target such as purchase intention, choice, and donation (Deng et al. 2019; Karataş and Gürhan-Canli 2020; Lee et al. 2010; Li and Pandelaere 2021; Wadhwa and Zhang 2015). For instance, Wang et al. (2022) revealed that feeling right mediated the effect of fit between shape and healthiness on consumer preference. Specifically, combinations of angular foods and healthiness, as well as circular foods and tastiness, were rated as feeling right and, resultantly, more likely to be chosen compared to incongruent combinations. Deng et al. (2019) demonstrated that vertical (horizontal) writing direction in advertising fits traditional/antique (modern/contemporary) brand/product perception, and those congruencies enhanced persuasiveness and favorable brand/product evaluations. In addition, the researchers confirmed that feeling right also mediated the congruency effect of writing direction and product appearance (traditional/modern) in the advertisement. Moreover, Deng et al. (2019) identified the mediating role of feeling right in the congruency effect. Similarly, Monahan and Romero (2020) found that when the motion direction in an advertisement aligns with consumers' mental representations of motion—such as moving from left to right for English-speaking audiences—trust in the advertised brand increases. The researchers also demonstrated that feeling right mediates this congruency effect. Wang et al. (2020) demonstrated that message-color congruency (i.e., health-focused messages alongside black-and-white designs and taste-focused messages alongside color designs) enhanced consumers' feeling right, which in turn, led to higher attitude toward the ad/restaurant, purchase intention, and willingness to pay.

Nelson and Hitchon (1999) demonstrated that cross-sensory headlines of an advertisement for a TV station positively influenced not only people's emotional responses (pleasantness) but also cognitive responses (number of listed thoughts toward the advertisement). Moreover, Cesario et al. (2004) suggested that feeling right leads to positive target evaluations, regardless of the affective response, meaning, it is a nonemotional experience where individuals subjectively feel correct and/or proper (Avnet et al. 2013; Camacho et al. 2003; Cesario et al. 2004; Deng et al. 2019). Therefore, although feeling right may not be completely independent of positive emotion, we posit that feeling right and excitement will parallelly mediate the effects of timbre-vision congruencies on favorable consumer responses in terms of both evaluations and choice behaviors.

Hypotheses. *Sensory congruence between an instrumental timbre and a visual attribute will induce* (*H*1) *a sense of feeling right and* (*H*2) *excitement*, *which in turn*, *will positively influence consumers'* (*H*3*a*) *evaluations and* (*H*3*b*) *choices of the target*.

2.3.3 | Processing Fluency

Although the present study does not specifically focus on processing fluency, one can argue that the congruence between timbre and visual design elements results in favorable responses because the experience is more easily understood (Biswas 2019; Biswas et al. 2021) and processed (Reber et al. 2004; Velasco et al. 2016). Processing fluency refers to the metacognitive ease or difficulty experienced during information processing (Alter and Oppenheimer 2008; Oppenheimer 2008; Reber and Unkelbach 2010). According to Hung (2000), the congruence between music and visual elements will likely enhance consumers' processing fluency. Puligadda and VanBergen (2023) demonstrated that processing fluency mediated the effect of congruence visual and sound logos on interest in trying the target product (a new shampoo). Studies in sensory marketing have also demonstrated that consumers can process auditory and visual stimuli more fluently when they are congruent with each other in terms of their crossmodal correspondences (Sunaga 2018). An experience that is highly fluent is likely to heighten the evaluation of the target stimulus (Biliciler et al. 2022; Chan and Northey 2021; Kim et al. 2009; Landwehr et al. 2013; Lee and Labroo 2004; Sundar and Noseworthy 2016). Thus, we also checked for the possibility of an alternative explanation for our results in terms of processing fluency.

2.4 | Moderating Role of Usage/Purchase Frequency

The Elaboration Likelihood Model proposed by Petty and Cacioppo (1986) suggests that experts tend to deliberately process central information, such as the quality and functions of the product, while novices tend to depend more on peripheral cues, such as background music. In fact, Nelson and Hitchon (1999) demonstrated that audiovisual crossmodal expressions (e.g., full-color sound) did not significantly enhance the effectiveness of advertising for those products where the visual and auditory senses were central to the product's functionality, such as TV stations.

Studies of sensory marketing have also suggested that the positive effects of sensory fit on consumer evaluations of co-branding became more decisive in the case of a low-involvement (vs. high-) product (Ahn et al. 2020). Moreover, people are prone to be influenced by feeling right, especially when they have little or no prior experience/expertise with the object or they are unfamiliar with it (Dunn and Schweitzer 2005; Kim et al. 2009; Vaughn et al. 2010). As Kim et al. (2009) and Vaughn et al. (2010) mentioned, experts are less likely to be influenced by sensory congruence than by the content of the message, as they tend to have more confidence in their ability to evaluate the information. While experts may possess other, more informative details that assist in their assessment, sensory inputs are often seen as more informative for individuals who lack expertise in the domain (Kim et al. 2009; Vaughn et al. 2010).

According to evolutional psychology, when individuals encounter a novel or unfamiliar event, they tend to depend more on their auditory system, which has a 360° detection capability, rather than on vision, which has limited detection abilities (Shapiro et al. 1984). This advantage of the auditory modality in alerting consumers to novel information leads to a shift in dominance from vision to audition (Fenko et al. 2010; Shapiro et al. 1984). As a result, sensitivity to timbre-visual congruency

is expected to increase because it involves a multimodal input from vision and audition rather than a unimodal input from vision. In contrast, when facing a familiar event, consumers will rely primarily on vision, which is the usually attended (i.e., dominant) modality. This reliance on vision may inhibit the utilization of the auditory modality. Taken together, frequent users or buyers of the target product category are not anticipated to be influenced by the connotative congruence effects proposed by this study.

Hypothesis 4. *The timbre-visual congruence effect will be moderated by product familiarity. Specifically, the connotative congruence effect will occur among consumers who infrequently use or purchase the target product category.*

While this study does not focus on cultural differences, people having different cultures, nationalities, languages, and environments might perceive and evaluate the same instrumental timbres differently (e.g., see Jiang et al. 2020; Qi et al. 2020; Walker 1987; Wallmark and Kendall 2018). Qi et al. (2020), exploring differences in timbre—color/taste/softness or smoothness associations between Western and Chinese instruments, suggested some of those crossmodal associations were universal. It should be noted that the auditory stimuli used in their experiments were simple musical notes tones lasting 1.5 s. Thus, Qi et al. (2020) encouraged researchers to conduct experiments in diverse countries targeting different cultures using more complex auditory stimuli such as excerpts. Many of the recent studies on timbre-vision associations have been conducted in the United States. Thus, to confirm that the proposed effects are not peculiar to a specific country, the studies were conducted outside of the US, specifically, in the United Kingdom and Japan (as a representative of regions where people live in different cultures and speak a different language from the United Kingdom). In addition, we also conducted experiments in English-speaking countries in Africa (as a representative of regions where people speak the same language as in the United Kingdom but once again live in different cultures and environments from both the United Kingdom and Japan). For an overview of studies, see Table 3 and Supporting Information S2: File 1.

3 | Study 1

3.1 | Pretests

Before Study 1, two prestudies (for detail including each sample size rationale, see Supporting Information S2: Files 2 and 3) and a control test were conducted. The purpose of Pre-test 1, whose participants were recruited from the UK, Japan, and Africa, was to explore the connotations that people have with various instrumental timbres. This study used 10 musical instruments in total as the auditory stimuli, consisting of 2 instruments each of five categories, specifically, keyboard (piano and pipe organ), woodwind (flute and clarinet), stringed (violin and guitar), brass (trumpet and French horn), and percussion (hand bells and steel drum). These five categories were based on the four instrumental categories (i.e., woodwinds, stringed, brass, and percussion instruments) that had been used in Puligadda and VanBergen (2023), and the percussion was divided into ones with and without keyboard. The participants listened to ten arpeggios (i.e., notes from a chord played in rapid succession rather than simultaneously) played by each instrument and rated them with 7-point 12 semantic differential (SD) scales such as bright/dark (see Supporting Information S2: File 2).

After pretest 1, to select those visual stimuli congruent with a specific instrumental timbre, pretest 2 was conducted in the United Kingdom (for detail, see Supporting Information S2: File 3). This study used nine images of steel, wood, velvet, satin, fur, leather, Japanese paper, hemp fabric, and polypropylene carpet, which were all shown in the dark gray (HSL; 0°, 0%, 20%, see Supporting Information S1: Web Appendix A). The participants rated them one by one using the same 12 SD scales as in pretest 1.

To identify the sensory (in)congruence between instrumental timbre and visual image, linear mixed-effects model (LMM) analyses, which are particularly useful for within-participant study designs (Magezi 2015), were used. The fixed effects of timbre/visual images generated by LMM analyses with the heterogeneity across participants as the random effects indicated that among British people, both the carpet image and the organ timbre tended to be perceived as relatively harsh, angular, rough, hard, warm, hot, full, dull, and heavy. Meanwhile, they perceived both the satin image and the hand bell timbre to be relatively cold, cool, empty, light in weight, and sharp. Based on these results, it was expected that people would feel the congruency between the carpet and the timbre of the organ as well as between satin and the timbre of the hand bells.

3.2 | Control Tests 1 and 2

To confirm that people do indeed perceive the congruence between instrumental timbre and material image as predicted based on the results of the two pretests and LMM analyses, Control study 1 was conducted (for detail including sample size rationale, see Supporting Information S2: File 4). This test revealed that, as expected, the participants perceived the timbre of the hand bell (vs. organ) as matching the visual image of satin better. The result therefore indicates that people do indeed perceive carpet and the timbre of the organ as well as satin and the timbre of the hand bells as matching.

In addition, Control test 2 (for details including sample size rationale, see Supporting Information S2: File 5) was conducted to rule out the possible alternative explanation of the matches in terms of common emotional associations as suggested by Palmer et al. (2013). This test revealed that the carpet image was not more strongly correlated with the organ (vs. hand bell) timbre in terms of a common emotional state (sad). Additionally, the satin image was not more strongly associated with the hand bell (vs. organ) timbre in terms of common emotional association. Control test 2 also revealed that the familiarity with the carpet image was not strongly correlated with that of the organ (vs. hand bell) timbre. The participants' familiarity with the satin (vs. carpet) image was not strongly correlated with that with the hand bell (vs. organ) timbre either. Thus, regarding the experimental stimuli that were used in the present study, the possibility that the congruencies between the organ timbre and

TABLE 3 | Summary of studies.

Areas		Main aims	Timbres	Visual stimuli	Outcomes
Study 1					
Pre 1	UK JPN AFR	Exploring connotations of instrumental timbres	10 instruments[a]	N/A	Music perceptions
Pre 2	UK	Selection of visual stimuli	N/A	9 materials[b]	Material perceptions
Ctrl 1	UK	Confirming timbre-image congruence	ORG, HB	CPT, ST	Perceived congruence
Ctrl 2	UK	Ruling out explanations by common emotions	ORG, HB	CPT, ST	Associations with emotion
Main	UK	Tests of H1 to H3a	ORG, HB	Book cover	EVAL, LIKE (ad, product), FR, EXC
Study 2					
	UK AFR	Tests of H1 to H3a with individual-level analyses	ORG, CL, VL, FH, HB	Typefaces	EVAL, LIKE FR, EXC
Study 3					
Pre	JPN	Selection of visual stimuli	N/A	Ballpoint-pens (ink colors)	Perceptions of ballpoint-pens
Main	JPN	Tests of Hypotheses 1, 3a, 3b, and H4	CL, HB	Red/pastel blue pens	EVAL, WTP, CHO
Study 4					
	UK	Establishing the causal process	CL, HB	Pastel blue cushion	EVAL (ad, product), WTP, FR, EXC
Study 5					
Main	JPN	Tests of H3b and H4	CL, HB	Products' packages	CHO
Ctrl	JPN	Confirming perceptions of package colors	CL, HB	Color patches	Perceived congruence

Note: Study 4 was preregistered (https://aspredicted.org/24dw-zjfg.pdf). All auditory stimuli and raw data have been stored in a project directory on the Open Science Framework (https://osf.io/6jb59/?view_only=32b87e3cdb18415ab261689a416036e8).
Abbreviations: CHO, choice; CL, clarinet; Ctrl, control test; EVAL, evaluation; EXC, excitement; FH, French horn; FR, feeling right; HB, hand bells; LIKE, liking; ORG, organ; Pre, pre-test; ST, satin; VL, violin.
[a]Piano, pipe organ, flute, clarinet, violin, guitar, trumpet, French horn, hand bells, steel drum.
[b]Steel, wood, velvet, satin, fur, leather, Japanese paper, hemp fabric, polypropylene carpet.

the carpet image, as well as between the hand bell timbre and the satin, would be based on common emotional associations or the participants' familiarity with the stimuli was ruled out.

3.3 | Methods (Main Study)

Study 1 used a 2 (timbre: organ vs. hand bells) × 2 (image: carpet vs. satin) between-participants experimental design. A total of 202 new British participants were recruited on Prolific and rewarded with 0.63 GBP for taking part in the study. The data from one participant were excluded because s/he did not follow the task instructions, specifically, the participant viewed the experimental video for a short time (17 s; $n = 201$, male = 47.2%, female = 52.8%, $M_{age} = 37.4$ years, SD = 11.20). A priori power analysis using G*Power (Faul et al. 2009) revealed that 199 observers would provide 0.80 power to detect the interaction effect ($f = 0.20$) in 2 × 2 between-participants analysis of variance (ANOVA) with the α level at 0.05. The participants viewed a promotional video for a book with background music played on the organ or hand bells. The auditory stimuli were created from a MIDI (Musical Instrument Digital Interface) file of a version of *Canon* by Johann Pachelbel. A book, titled "THE WORLD of MARKETING: How do marketers make it possible," was displayed on the screen for 21 s, and three captions were shown under the book sequentially for 7 s each: *Truly insightful and entertaining*, *Very simple and easy-to-understand*, and *A must-read for all managers and consumers*. In the carpet and satin conditions, the images of the respective materials were used for the cover design (see Supporting Information S1: Web Appendix B). We selected book cover design because it is independent of the content's quality but is influential in grabbing consumers' interest and facilitating purchase decisions. There was no vocal element in the video. The perceived loudness of the music was −15 Loudness Unit Full Scale (LUFS) in all conditions. The participants were instructed to set the volume of the speaker on their computer to a comfortable listening level and to maintain it at this level for the duration of the study.

After viewing the video, the participants indicated how much they liked it (for specific measurement items and scales, see

Supporting Information S2: File 6). Subsequently, they indicated product interest and behavioral intention (BI). Next, the processing fluency of the video clip was assessed with five dichotomous word pairs, which were averaged to form a mean fluency score (Cronbach's $\alpha = 0.871$). The participants then indicated whether the background music felt right and their liking. The participants' excitement or high arousal positive emotion (Faseur and Geuens 2006; Kraiss et al. 2023) was measured with five dichotomous word pairs (Cronbach's $\alpha = 0.889$). The responses were averaged to form an aggregate excitement score. All items above were measured using a 7-point scale. The demographic questions were presented at the end of the study.

3.4 | Results and Discussion

3.4.1 | Feeling Right, Excitement, and Product Evaluation

To test the effects of timbre-visual connotative congruency on the participants' sense of feeling right, excitement, and product evaluation, 2 (timbre: organ vs. hand bells) × 2 (image: carpet vs. satin) ANOVAs on those variables were conducted. Product evaluation scores were formed by averaging interest in the book and BI (Cronbach's $\alpha = 0.969$).

The results revealed significant interaction effects between independent variables (IVs) on feeling right ($F(1, 197) = 4.094$, $p = 0.044$, $\eta_p^2 = 0.020$) and excitement ($F(1, 197) = 6.603$, $p = 0.011$, $\eta_p^2 = 0.032$), and a borderline-significant interaction effect on product evaluation ($F(1, 197) = 3.783$, $p = 0.055$, $\eta_p^2 = 0.109$). There were no other significant effects ($Fs(1, 197) < 1.026$, $ps > 0.312$, $\eta_p^2 s < 0.006$). Pairwise comparisons revealed that the participants in the organ condition felt right more, were more excited, and evaluated the book more favorably when the image of the carpet (vs. satin) was used for the cover design, although the differences in feeling right and excitement were borderline significant (see Figure 1). Similarly, pairwise comparisons revealed that the participants in the carpet condition were more excited when the background music was played with the organ (vs. hand bells), while those in the satin condition felt right more and evaluated the book more favorably when the music was played with the hand bells (vs. organ). Again, the differences in feeling right and excitement were borderline significant. These results were supportive of our hypotheses (H1 to H3a).

3.4.2 | Mediation

An analysis was conducted with the Hayes PROCESS macro v3.5 to investigate a parallel and serial mediation model (model 80; Hayes 2018, see Figure 2), indicating that sensory congruence enhances the participants' liking of the video by increasing their experience of feeling right and excitement, which, in turn, improves their product evaluations.

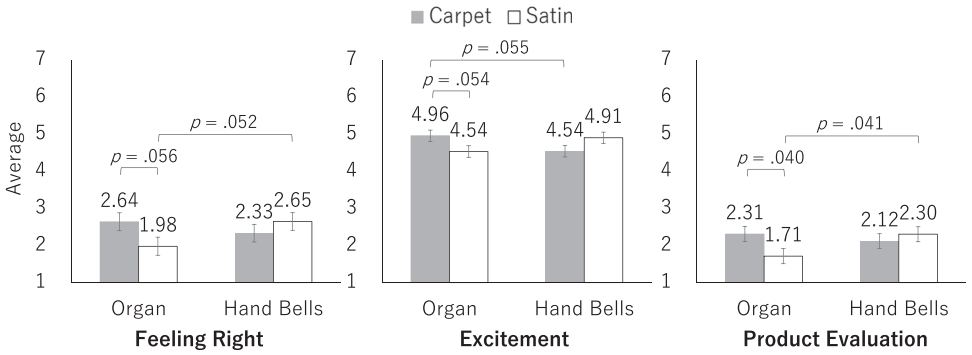

FIGURE 1 | Sensory congruence in Study 1. All p values were Bonferroni-adjusted.

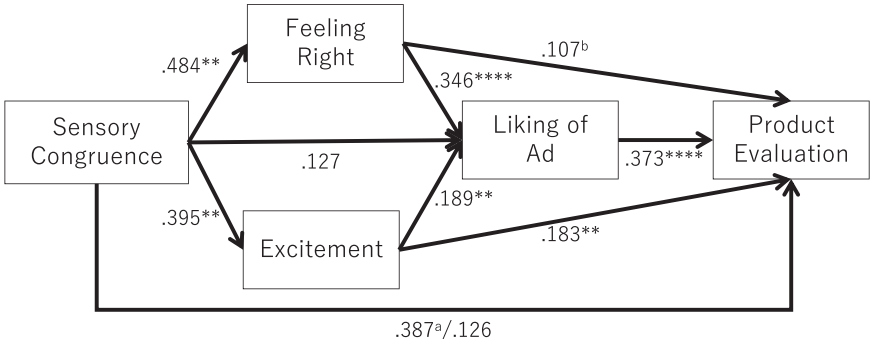

FIGURE 2 | Mediation model of sensory congruence in Study 1. Unstandardized regression coefficients are reported. The numbers to the right and left of the slash indicate the direct (i.e., excluding effects through mediators) and total (i.e., including all indirect effects through mediators) effects, respectively. Sensory congruence was coded as a dummy variable: 0 = incongruence, 1 = congruence. [a]$p = 0.057$, [b]$p = 0.080$, **$p < 0.05$, ****$p < 0.001$.

The model shown Figure 2 is consistent with prior studies on attitude towards advertisement (A_{ad}) indicating that favorable A_{ad} enhances attitude towards the target brand and purchase intention (Gardner 1985; MacKenzie and Lutz 1989; MacKenzie et al. 1986). The total effects of the sensory congruence on product evaluations were borderline significant ($B = 0.387$, $t(199) = 1.914$, $p = 0.057$). Meanwhile, the indirect effect of sensory congruence through the three mediators (i.e., feeling right, excitement, and liking of the advertisement) on product evaluation was significant ($\beta_{total} = 0.181$, 95% CI = [0.041, 0.346], see Table 4).

In addition, to check the robustness of the results, we used simple parallel mediation analyses (Model 4; Hayes 2018) removing the liking of the video from the model shown in Figure 2 (i.e., product evaluation as the DV, sensory congruence the IV, the feeling right and excitement as the mediators). The results revealed significant mediation paths of both sensory congruence–feeling right—product evaluation ($\beta = 0.079$, 95% CI = [0.002, 0.189]) and sensory congruence—excitement—product evaluation ($\beta = 0.069$, 95% CI = [0.010, 0.146], for results in detail, see Supporting Information S2: File 7). The participants in the sensory congruent (vs. incongruent) condition indicated a higher sense of feeling right ($\beta = 0.282$, $t(199) = 2.014$, $p = 0.045$) and greater excitement ($\beta = 0.395$, $t(199) = 2.580$, $p = 0.011$), which, in turn, led to a more favorable product evaluation ($\beta_{\text{feeling right}} = 0.280$, $t(197) = 4.084$, $p < 0.001$, $\beta_{excitement} = 0.193$, $t(197) = 2.793$, $p = 0.006$, see Supporting Information S2: Figure 1 in Supporting Information File S7). Therefore, H1, H2, and H3a were all supported.

Note that there was no significant main or interaction effect on liking of the "music" (Fs < 1, ps > 0.335), meaning that the participants liked the music equally, regardless of the experimental condition. It is also noteworthy that there was no difference in the processing fluency score between the various experimental conditions (Figure 3), thus ruling out fluency as a potential alternative mediator of the relationship. As mentioned earlier, feeling right is a subjective experience that can be induced by sensory congruence (Kuo and Rice 2015). On the other hand, processing fluency refers to the subjective experience of the ease of consumers' information processing instead of congruence as such. Thus, it is possible that congruency enhances consumers' feeling right but not their perceived fluency when the stimulus is so simple that people can process it easily regardless of sensory congruence.

TABLE 4 | Parallel and serial mediation results in Study 1.

Direct effects	**β[a]**	**SE**	***t* value**	***p* value**
Outcome variable: Feeling right				
Sensory Congruence	0.282	0.240	2.013	0.045
Outcome variable: Excitement				
Sensory congruence	0.359	0.153	2.580	0.011
Outcome variable: Liking of the Ad				
Sensory congruence	0.094	0.170	0.750	0.454
Feeling right	0.436	0.051	6.782	0.000
Excitement	0.152	0.080	2.355	0.020
Outcome variable: Product evaluation				
Sensory congruence	0.087	0.182	0.690	0.491
Feeling right	0.127	0.061	1.758	0.080
Excitement	0.139	0.087	2.101	0.037
Liking of the Ad	0.352	0.076	4.891	0.000
Indirect effects	**β[a]**	**SE**	**LLCI**	**ULCI**
CON → FR → PE	0.036	.0039	−0.016	0.134
CON → EXC → PE	0.050	0.030	0.001	0.120
CON → LOA → PE	0.033	0.048	−0.051	0.143
CON → FR → LOA → PE	0.043	0.027	0.001	0.107
CON → EXC → LOA → PE	0.019	0.013	0.001	0.051
Total	0.181	0.077	0.041	0.346

Note: Sensory congruence, which was coded as a dummy variable: 0 = incongruence, 1 = congruence.
Abbreviations: CON = sensory congruence, EXC = excitement, FR = feeling right LLCI = lower-limit confidence interval (95%), LOA = liking of the ad, PE = product evaluation, ULCI = upper-limit confidence interval (95%).
[a]Partially standardized coefficients.

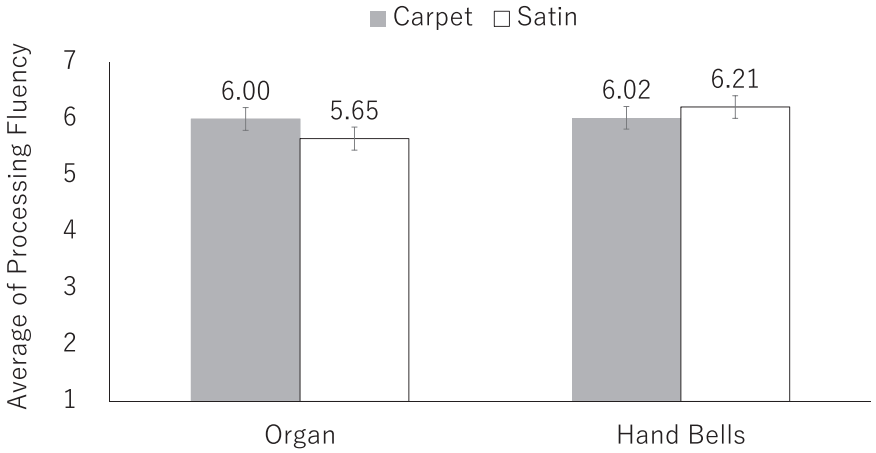

FIGURE 3 | Processing fluency score in Study 1. Neither the main ($F_{\text{timbre}}(1, 197) = 2.057$, $p = 0.153$, $\eta_p^2 = 0.010$; $F_{\text{image}}(1, 197) = 0.147$, $p = 0.701$, $\eta_p^2 = 0.001$) nor the interaction effect ($F_{\text{timbre} \times \text{image}}(1, 197) = 11.787$, $p = 0.183$, $\eta_p^2 = 0.009$) was significant.

4 | Study 2

4.1 | Methods

The aim of Study 2 was to confirm whether the results reported here could be obtained by (1) using other visual stimuli and (2) conducting an individual-level analysis. Ninety-seven British participants (male = 29.9%, female = 70.1%, $M_{\text{age}} = 35.2$ years, SD = 10.49) and 89 Africans whose first language was English living in English-speaking countries (male = 42.7%, female = 57.3%, $M_{\text{age}} = 28.7$ years, SD = 7.85) were recruited on Prolific and rewarded with 3.13 GBP for taking part in the online study. A single factor within-participants experimental design was used. A priori power analysis using G*Power (Faul et al. 2009) revealed that 88 observers would provide 0.80 power to detect an almost medium effect ($f^2 = 0.13$; Cohen 1992) in a multiple linear regression with three predictors with the alpha level at 0.05.

The participants were first randomly presented with five typefaces and rated them one by one using the same 12 SD scales (7-point) as in pretest 1 of Study 1. We selected typeface because prior studies have suggested that people have associations between typefaces and music (e.g., Venkatesan et al. 2022). Subsequently, they were randomly presented with five instrumental timbres and rated them one-by-one using the same 12 SD scales (7-point). The order of the presentation of the typefaces and timbres was counterbalanced (i.e., half of the participants were first presented with and rated five timbres and then presented with and rated five typefaces). All five of the different typefaces read "ABCDE" with the same font size. The typefaces were displayed in grayscale letters against a white background (see Supporting Information S1: Web Appendix C).

The auditory stimuli were the same five arpeggios as in pretest 1 of Study 1 (i.e., a version of Chopin's Etude in C Major Op. 10 No. 1) played by keyboard (organ), woodwind (clarinet), stringed (violin), brass (French horn), and percussion (hand bells). Once the participants had completed the rating of both typefaces and timbres, they were presented with the following instructions.

> *A company wants to develop a brand logo that people like regardless of the product category. The brand name has yet to be decided, so each logo will simply say 'ABCDE'. You will watch five short videos presenting one option each and evaluate each typeface to give the investigators feedback to help them develop a more favorable brand logo.*

In the video, five pairs of brand-logo and instrumental timbre were presented for 15 s per pair. Each experimental stimulus was randomly selected from 25 (5 brand-logos × 5 instrumental timbres) pairs. The participants indicated their evaluation (bad/good, unfavorable/favorable, negative/positive; Cronbach's $\alpha_{\text{UK}} = 0.970$, $\alpha_{\text{AFR}} = 0.961$), liking, feeling right, excitement, and processing fluency of each brand logo using a 7-point scale while listening to one of five timbres. The measurement items of liking, feeling right, excitement (Cronbach's $\alpha_{\text{UK}} = 0.899$, $\alpha_{\text{AFR}} = 0.918$), and processing fluency (Cronbach's $\alpha_{\text{UK}} = 0.945$, $\alpha_{\text{AFR}} = 0.921$) were the same as those of Study 1. Finally, the participants answered some demographic questions.

4.2 | Results and Discussion

Sensory congruence scores were calculated by participant as the difference between perceptions of brand-logo and instrumental timbre using the following equation.

$$\text{SC}_{\text{hij}} = \sum_{k=1}^{12} | \text{LP}_{\text{hik}} - \text{TP}_{\text{hjk}} |,$$

where:

SC_{hij} = sensory congruence between brand-logo i and timbre j for respondent h

LP_{hik} = perception of brand-logo i regarding dimension k for respondent h

TP_{hjk} = perception of timbre j regarding dimension k for respondent h

Thus, a higher (lower) value of the score indicates that there is a larger (smaller) distance, meaning lower (higher) congruency. The same computational method was used to measure the "self-congruency," meaning the congruency between brand personality and actual/ideal self-image (see Kressmann et al. 2006). Table 5 shows which instrumental timbre had the shortest/longest distances to, or was the most/least congruent with, each font.

Using data that merged the participants from the United Kingdom and African countries, the parallel mediation model (model 4; Hayes 2018, see Figure 4) with sensory congruence score (SC_{hij}) as the IV, feeling right and excitement toward the brand-logo as mediators, and evaluation/liking as the dependent variables (DVs) was tested to confirm the effects of the congruence between brand-logo and timbre (for the results when using data from participants in the United Kingdom and African countries separately, see Supporting Information S2: File 8).

The total effects of sensory congruence on the DVs were significant ($\beta_{evaluation} = -0.070$, $t(928) = -2.149$, $p = 0.032$, $\beta_{liking} = -0.078$, $t(928) = -2.375$, $p = 0.012$). The indirect effects of the sensory congruence through feeling right as well as excitement on both evaluation ($\beta_{feeling\ right} = -0.047$, 95% CI = [−0.080, −0.015]; $\beta_{excitement} = -0.050$, 95% CI = [−0.081, -0.020]) and liking ($\beta_{feeling\ right} = -0.048$, 95% CI = [−0.082, −0.015]; $\beta_{excitement} = -0.047$, 95% CI = [−0.078, −0.019]) were also significant (see Table 6). As Figure 4 shows, the higher the sensory congruence between brand-logo and timbre (i.e., the shorter the perceived distance between them), the more strongly the participants felt right ($\beta = -0.096$, $t(928) = -2.926$, $p = 0.004$) and excited ($\beta = -0.108$, $t(928) = -3.304$, $p = 0.001$), which in turn, led to a higher evaluation ($\beta_{feeling\ right} = 0.492$, $t(926) = 19.969$, $p < 0.001$; $\beta_{excitement} = 0.461$, $t(926) = 18.686$, $p < 0.001$) and stronger liking ($\beta_{feeling\ right} = 0.5507$, $t(926) = 20.080$, $p < 0.001$; $\beta_{excitement} = 0.440$, $t(926) = 17.421$, $p < 0.001$).

The individual-level analyses also demonstrate that brand logos are more likely to feel right to consumers and enhance consumers' excitements when they are congruent with the instrumental timbre of the background music, which, in turn, improves their evaluation and liking of the brand logo. Therefore, H1 through H3b were supported.

To rule out the alternative explanation that processing fluency might play a mediating role instead of feeling right/excitement, additional analyses were conducted using processing fluency as a mediator. The results revealed that the sensory congruence did not significantly influence the participants' processing fluency ($\beta_{UK} = -0.000$, $t(928) = -0.068$, $p = 0.946$). The outputs also failed to reveal a significant indirect effect of processing fluency on either evaluation or liking ($\beta_{evaluation} = -0.000$, 95% CI = [−0.004, 0.003]; $\beta_{liking} = -0.000$, 95% CI = [−0.003, 0.002]). Therefore, the possibility of the alternative explanation in terms of processing fluency was once again ruled out.

In contrast to Study 1, where the participants' excitement at the time was measured, here the participants were asked to indicate their emotions toward different typefaces of a brand logo in Study 2.

TABLE 5 | Instruments the most/least congruent with each brand-logo in Study 2.

		ABCDE	ABCDE	ABCDE	ABCDE	ABCDE
UK	Congruent	ORG 19.43 (7.51)	FH 19.97 (7.46)	CL 20.20 (8.17)	HB 19.13 (7.37)	ORG 18.88 (8.04)
	Incongruent	CL 26.82 (11.06)	HB 22.62 (7.23)	ORG 21.83 (7.45)	CL 21.58 (8.39)	HB 29.20 (19.37)
AFR	Congruent	ORG 21.84 (8.74)	ORG 22.19 (9.11)	CL 21.34 (7.73)	HB 22.79 (9.05)	ORG 21.06 (9.09)
	Incongruent	HB 27.06 (10.17)	HB 25.63 (9.56)	ORG 25.09 (9.83)	FH 23.62 (8.81)	HB 29.42 (11.02)

Note: The numbers represent the distances between each timbre and brand-logo while those in parenthesis indicate the standard deviation. Congruent and incongruent represent the instrumental timbre having the shortest and longest distances to (i.e., the most/least congruent with) each brand-logo, respectively.
Abbreviations: CL, clarinet; FH, French horn; HB, hand bells; ORG, organ.

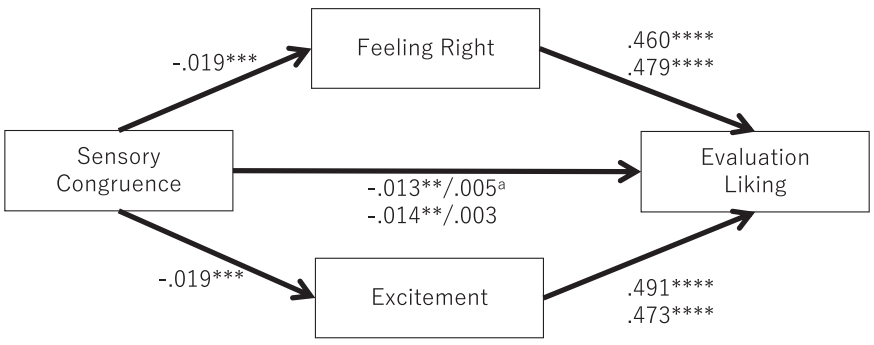

FIGURE 4 | Sensory congruence effect in Study 2. Unstandardized regression coefficients are reported. The sensory congruence score was formed by calculating the difference between perceptions of brand logo and timbre (i.e., higher values indicate lower congruency). Therefore, the negative coefficients on feeling right and excitement indicate that the higher the timbre-vison congruence, the higher the participants' feeling right/excitement. The upper and lower numbers in each pair indicate the coefficients on evaluation and liking, respectively. Numbers to either side of the slash indicate the direct and total effects, respectively. [a]$p = 0.057$, ***$p < 0.01$, ****$p < 0.001$.

TABLE 6 | Parallel mediation results in Study 2.

Direct effects	β[a]	SE	*t* value	*p* value
Outcome variable: Feeling right				
Sensory congruence	−0.096	0.006	−2.926	0.003
Outcome variable: Excitement				
Sensory congruence	0.108	0.006	−3.304	0.001
Outcome variable: Evaluation				
Sensory congruence	0.026	0.003	1.904	0.057
Feeling right	0.460	0.023	19.969	0.000
Excitement	0.491	0.026	18.686	0.000
Outcome variable: Liking				
Sensory congruence	0.018	0.003	1.281	0.200
Feeling right	0.507	0.024	20.080	0.000
Excitement	0.440	0.027	17.421	0.000
Indirect effects	β[b]	**SE**	**LLCI**	**ULCI**
CON → FR → EVL	−0.047	0.016	−0.080	−0.015
CON → EXC → EVL	−0.050	0.016	−0.081	−0.020
Total	−0.097	0.030	−0.153	−0.037
CON → FR → LIKE	−0.048	0.017	−0.078	−0.019
CON → EXC → LIKE	−0.047	0.015	−0.082	−0.015
Total	−0.096	0.030	−0.155	−0.037

Abbreviations: CON, sensory congruence; EVL, evaluation; EXC, excitement; FR, feeling right; LIKE, liking; LLCI, lower-limit confidence interval (95%); ULCI, upper-limit confidence interval (95%).
[a]Standardized coefficients.
[b]Partially standardized coefficients.

The participants in Study 2 were presented with multiple experimental stimuli (i.e., typefaces of a brand logo) and had to repeatedly answer questions to measure the mediating variables and DVs (i.e., the participants had to repeatedly indicate their excitements). However, the excitement induced by crossmodal correspondences may not be so strong and may often be unconscious. Therefore, it is difficult for the participants to catch and precisely report the subtle change in their emotional state made unconsciously within a very short time. To solve this issue, we asked participants to indicate their emotions towards different typefaces of a brand logo. The typeface of a brand logo has been demonstrated to induce consumers' affective responses (Salgado-Montejo et al. 2014; Venkatesan et al. 2022). Venkatesan et al. (2022) found that typeface curvilinearity influenced participants' emotional valence (i.e., pleasantness) in music perceived as neither round nor angular, though the effect on arousal was not significant. Thus, it is presumably easier for participants to repeatedly indicate their emotions toward different typefaces of a brand logo (vs. general emotional state) within a short space of time. Therefore, in Study 2, excitement towards different typefaces of a brand logo was used as the proxy variable of the general emotional state.

5 | Study 3

The results of pretest 1 of Study 1 indicated that for the Japanese participants, the connotative associations of the clarinet and hand bell timbres were at opposite poles for the majority of the dimensions (see Supporting Information S2: Table 1 in Supporting Information S2: File S2.2). In other words, in Japan, the clarinet and hand bell timbres had the highest and lowest average scores, respectively, for many of the connotative descriptors. Specifically, the clarinet timbre was perceived as the dullest, most angular, warmest, and hottest among the nine instrumental timbres used in the pretest 1 of Study 1. In contrast, the hand bell timbre was perceived as the brightest, sharpest, emptiest, poorest, coldest, driest, and hardest among the nine timbres. Thus, we selected these two instruments for the music stimuli in Study 3.

In Studies 1 and 2, we used visual stimuli with achromatic colors (i.e., grayscale). To investigate whether the results can be extended to chromatic colors and products whose color is a central attribute, the following experiment uses ballpoint pens as the target product. We selected ballpoint pens because there is a wide range of colors, consumers generally are not loyal to any brand, and most consumers are familiar with the category.

5.1 | Pretest

An online pretest was conducted to select those stimuli that Japanese people perceive as matching the clarinet and hand bell timbres (for detail including sample size rationale, see Supporting Information S2: File 9). The results of pairwise comparisons indicated that the ink color of the ZEBRA Surari red

(HSL: 1°, 64%, 49%) was congruent with the clarinet timbre while the color of the PILOT Juice pastel blue (HSL: 181°, 39%, 69%) was congruent with the hand bell timbre. Specifically, the former was warmer ($M_{red} = 4.92$, SD = 1.65 vs. $M_{blue} = 2.31$, SD = 1.26), more vivid ($M_{red} = 4.88$, SD = 1.45 vs. $M_{blue} = 1.77$, SD = 1.11), richer ($M_{red} = 5.08$, SD = 1.38 vs. $M_{blue} = 1.50$, SD = 0.76), and darker ($M_{red} = 3.65$, SD = 1.94 vs. $M_{blue} = 1.85$, SD = 0.93) than the latter (Bonferroni-corrected $ps < 0.005$; for F values, see Supporting Information S2: Table 8 in Supporting Information File S9). Thus, these two ballpoint pens were chosen as the target products in Study 3.

5.2 | Methods

To extend the results under more controlled conditions and test the moderating hypothesis (all hypotheses except H2), a consequential choice experiment in a lab was conducted in Osaka, Japan. The experiment was conducted in a 79 m^2 room with four ceiling loudspeakers. The illuminance on the desk was approximately 650 lx. A 3 (timbre: clarinet vs. hand bells vs. no music) × 2 (product color: red vs. pastel blue) mixed design with product color as the within-participants factor was used. One hundred and sixty-one new Japanese participants took part in the study and were rewarded with 3000 JPN (approximately 27 USD) for taking part. A priori power analysis using G*Power (Faul et al. 2009) revealed that 153 observers would provide 0.80 power to detect an almost medium interaction effect ($f = 0.22$; Cohen 1992) in a mixed ANOVA with 3 groups and 2 measurements with the α level set at 0.05. *Canon* by Johann Pachelbel played on the clarinet or hand bells was played on a loop in the room through the four ceiling loudspeakers throughout the study in the clarinet and hand bells conditions, respectively. The volume was kept at between 40 and 60 dB everywhere in the room. Meanwhile, no music was played during the study in the no music condition (i.e., the control group).

The participants were randomly assigned to one of three timbre conditions. Just as for the pretest, the participants were first instructed to write the loop line and six characters using the target product (red/pastel blue ballpoint pen) and indicate their perceptions of the product in terms of ink color and stickiness, performance, weight, and grip. Subsequently, the participants had to evaluate each pen using 7-point scales anchored by unfavorable/favorable, do not want to use/want to use, and dislike/like which were averaged to form an evaluation score (Cronbach's $\alpha = 0.968$). After that, they were asked to indicate their feeling right ("Does the ballpoint pen feel right to you?") and to indicate their willingness to pay (WTP) for the product in JPY. WTP was incorporated as an additional evaluation measure to assess it in a multifaceted manner and enhance the validity of the results. Once the participants completed their rating of one pen, they were given the other pen and answered the same questions about it. After that, they were told that they would be given one of the target products as a token of appreciation, so they should indicate which of them they would like. Subsequently, they were asked to indicate their usage frequency of each product ("How often do you use each ballpoint pen used in this study?") with a 7-point scale anchored by not at all/very often, which was followed by demographic questions. The participants took away the pen they chose at the end of the study.

In Study 3, the participants were not asked about their excitement on the assumption that the interval between evaluations of two target products was probably too short to be able to detect a meaningful change. Although the excitement could be measured after the participants complete the choice task, the procedure makes it impossible for investigators to detect which product induces the emotional state. This will be discussed later in the General Discussion.

5.3 | Results and Discussion

After completing the study, the completed questionnaires were reviewed, revealing that eight participants failed to follow the instructions. Specifically, using the pen, they wrote characters more or less than they were instructed to write. Thus, the data from the participants were excluded from the following analyses ($n = 153$, male = 47.7%, female = 52.3%, $M_{age} = 41.0$ years, SD = 11.68).

5.3.1 | Product Evaluation and WTP

To test H4, using the Hayes PROCESS model 7 (Hayes 2018), we tested the moderated mediation model with the usage frequency of the product as the moderator, sensory congruence as the IV, the feeling right of the target product as the mediator, and the evaluation as the DVs (see Figure 5). Sensory congruence was coded as 0, 1, or 2, where these values represented the control (i.e., no music), congruence, and incongruence, respectively. The result revealed a significant conditional indirect effect through feeling right in the sensory congruence condition. Specifically, the indirect effect through the feeling right of the target pen on evaluation was significant in the case of timbre-visual connotative congruence only when the participants' usage frequency of the product was low (1 standard deviation (SD) below the mean = 1.00; $B = 0.566$, 95% CI = [0.029, 1.082], see Table 7). This result indicates that the participants who had little experience using the target pen were more likely to feel right when the timbre of the background music was congruent with the color of the product (vs. no music), which in turn led to a more favorable product evaluation. However, the participants who had average or a lot of experience in using the product heightened a sense of feeling right as much as those who were not presented with any music, even though the timbre and color were congruent. Meanwhile, there were no significant conditional indirect effects through feeling right in the sensory incongruence condition. Instead, the direct effect (DE) was significant ($B = -0.200$, 95% CI = [−0.393, −0.007]), meaning that the timbre-visual connotative incongruence (vs. no music) led to lower product evaluation regardless of the usage frequency of the product. This could be obtained from the participants' lower excitement (negative emotion).

The same mediation analysis (model 7) with WTP as the DV showed similar results to the evaluation. The indirect effect through the feeling right of the target pen on WTP was

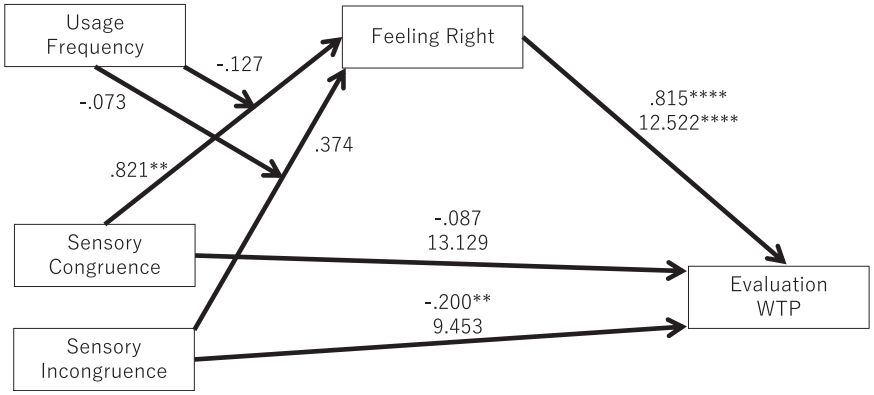

FIGURE 5 | Sensory congruence effect on product evaluation/WTP in Study 3. Unstandardized regression coefficients are reported. The control (i.e., no music) condition was specified as the reference group and was assigned a code of 0 on both congruence and incongruence dummy variables. The upper and lower numbers in each pair indicate the coefficients on the evaluation and WTP, respectively. The coefficients of Sensory (In) Congruence indicate the direct effects. $^{**}p < 0.05$, $^{****}p < 0.001$.

significant in the case of timbre-visual connotative congruence only when the participants' usage frequency of the product was average ($M = 2.89$; $B = 5.690$, 95% CI = [0.012, 13.277]) or low (1 SD below the mean = 1.00; $B = 8.689$, 95% CI = [0.412, 20.103]). This result indicates that the participants who had average or little experience using the target pen were more likely to feel right when the timbre of the background music was congruent with the color of the product (vs. no music), which in turn led to a higher WTP. However, the participants who had a lot of experience in using the product heightened their WTP as much as those who were not presented with any music, even though the timbre and color were congruent. Meanwhile, there were no significant conditional indirect or direct effects through feeling right in the sensory incongruence condition.

5.4 | Choice

To test H3b and H4, we tested the moderated mediation model (model 7; Hayes 2018) with the usage frequency of the product as the moderator, the feeling right of the blue pen as the mediator, and the choice as the DV (see Figure 6). Instrumental timbre, as an IV, was coded as 0, 1, or 2, representing the control, clarinet, and hand bells, respectively. The choice was coded as 0 or 1, representing the choice of pale blue and red pens, respectively.

The result revealed a significant conditional indirect effect of feeling right. Specifically, the indirect effect through the feeling right of the pastel blue pen on choice was significant in the hand bells condition only when the participants' usage frequency of the product was low (1 standard deviation (SD) below the mean = 1.00; $B = -0.450$, 95% CI = [−0.988, −0.038], see Table 8). This result indicates that the participants who had little experience of using the pastel blue pen were more likely to feel right when the pen was paired with the hand bell timbre (vs. no music), which in turn led to a higher probability of choosing the product (i.e., pastel blue pen). However, the participants who had average or a lot of experience in using the pastel blue pen heightened a sense of feeling right as much as those who were not presented with any music, even though it was paired with the hand bell timbre. These results imply that the effect of connotative congruence on choice through feeling right may only occur when consumers do not have much experience of using the product.

Meanwhile, the moderated mediation model (model 7; Hayes 2018) with the usage frequency of the product as the moderator, the feeling right of the red pen as the mediator, and the choice as the DV also revealed a significant conditional indirect effect of feeling right. The indirect effect of the red pen on choice was significant in the clarinet condition only when participants had a medium level of product usage frequency ($M = 3.90$; $B = 0.262$, 95% confidence interval = [0.130, 0.646], see Table 8). This means that the indirect effect was not significant for those with low or high product usage frequency. Therefore, as a whole, it is reasonable to say that H3b and H4 was partially supported.

6 | Study 4

6.1 | Methods

To establish a causal process, we move beyond measurement-of-mediator designs and instead use manipulation-of-mediator designs to directly test the effect of the mediator on the DVs (Ge 2023; Pirlott and MacKinnon 2016). Thus, Study 4 involved a direct manipulation of the mediators (excitement and feeling right) to find further support for the process we proposed. To manipulate mediators appropriately, the participants were assigned to one of the two levels of each mediator. Randomly assigning participants to mediators ensures temporal precedence and lowers the likelihood of alternative explanations about relationship between mediators and DVs (Luangrath et al. 2022; Pirlott and MacKinnon 2016).

We preregistered to recruit 850 British participants on Prolific (https://aspredicted.org/24dw-zjfg.pdf). Study 4 used a concurrent double randomization design (Pirlott and MacKinnon 2016) with a 2 (sensory congruence vs. incongruence) × 2 (feeling right: discourage vs. control) × 2 (excitement: discourage vs. control) study. A total of 839 British participants (male = 44.7%, female = 53.9%,

TABLE 7 | Moderated mediation results on evaluation/WTP in Study 3.

Direct effects	B^a	SE	*t* value	*p* value
Outcome variable: Feeling right				
Sensory congruence	0.821	0.411	2.000	0.046
Sensory incongruence	0.374	0.432	0.866	0.387
Frequency of using the product (FRQ)	0.377	0.091	4.150	0.000
Sensory congruence × FRQ	−0.127	0.122	−1.041	0.299
Sensory incongruence × FRQ	−0.073	0.124	−0.587	0.558
Outcome variable: Evaluation				
Sensory congruence	−0.087	0.098	−0.880	0.380
Sensory incongruence	−0.200	0.098	−2.044	0.042
Feeling right	0.815	0.022	37.179	0.000
Outcome variable: WTP				
Sensory congruence	13.129	10.645	1.233	0.218
Sensory incongruence	9.453	10.598	0.892	0.373
Feeling right	12.522	2.373	5.276	0.000
Indirect effects	B^a	**SE**	**LLCI**	**ULCI**
Low FRQ ($M = 1.00$)				
CON → FR → EVL	0.566	0.272	0.029	1.082
CON → FR → WTP	8.689	5.013	0.412	20.103
INCON → FR → EVL	0.246	0.288	−0.326	0.810
INCON → FR → WTP	3.773	4.716	−5.144	13.648
Average FRQ ($M = 2.89$)				
CON → FR → EVL	0.370	0.193	−0.019	0.740
CON → FR → WTP	5.690	3.430	0.012	13.277
INCON → FR → EVL	0.134	0.203	−0.262	0.544
INCON → FR → WTP	2.053	3.294	−4.160	8.890
High FRQ ($M = 4.91$)				
CON → FR → EVL	0.162	0.282	−0.383	0.724
CON → FR → WTP	2.485	4.467	−5.848	11.862
INCON → FR → EVL	0.014	0.287	−0.545	0.576
INCON → FR → WTP	0.216	4.603	−8.829	9.720

Abbreviations: CON, sensory congruence; EVL, evaluation; FR, feeling right; INCON, sensory incongruence; LLCI, lower-limit confidence interval (95%); ULCI, upper-limit confidence interval (95%).
[a]Unstandardized coefficients.

prefer not to say = 1.0%, prefer to self-describe = 0.5%, M_{age} = 38.8 years, SD = 10.75) completed the study.

The participants were asked to watch a short video (43 s) featuring a cushion whose color was pastel blue, the same as the color of the ballpoint pen used in Study 3 (i.e., HSL: 181°, 39%, 69%). We selected the cushion category because it allows for a natural incorporation of pastel blue, and it effectively implements the manipulation-of-mediator technique, which we will explain later. In the sensory congruence and incongruence condition, the participants were presented with the same excerpt as in Study 3 (i.e., *Canon* by Pachelbel) played on the hand bells and clarinet, respectively, as the background music.

To manipulate a sense of feeling right and excitement, we adopted a discouragement manipulation-of-mediator design (Pirlott and MacKinnon 2016), which was used by Mookerjee et al. (2021) and Motoki and Pathak (2023). This is also referred to as blockage manipulation, a type of experimental design that aims to manipulate the effect of a mediator (Pirlott and MacKinnon 2016). The goal is to block or 'neutralize' the mediator, preventing it from working or, more likely, reducing its impact (Pirlott and MacKinnon 2016). In Study 4, regarding a sense of feeling right, the participants were randomly assigned to one of the product descriptions. The video consisted of three slides, the first of which was the same regardless of the discouraging or control feeling right conditions: "Made from durable materials/Resilient for daily use and machine washable

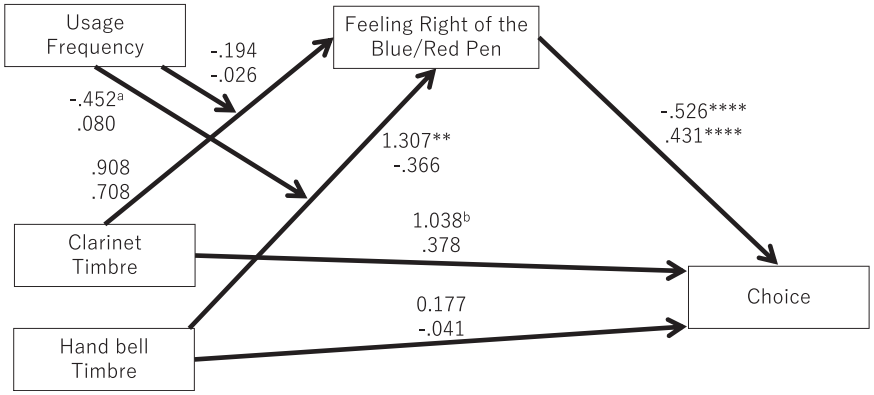

FIGURE 6 | Sensory congruence effect on choice in Study 3. Unstandardized regression coefficients are reported. The control (i.e., no music) condition was specified as the reference group and was assigned a code of 0 on both clarinet and hand bell timbre dummy variables. The choice assigned a code of 1 to the red pan and 0 to the pastel blue one. The top numbers in each pair indicate the coefficients in the case of the pastel blue pen, while the bottom numbers indicate those in the case of the red pen. [a]$p = 0.056$, [b]$p = 0.070$, **$p < 0.05$, ****$p < 0.001$.

for convenience/Simply cut the outer bag and fluff it up for instant comfort!/17 × 17 inches/Perfect fit for any armchair, sofa, or bed" (see Supporting Information S1: Appendix D). However, in the discouraging (control) feeling right condition, the product descriptions in the second slide read as follows: "Transform your living space with cosy, *warm* (*modern*) cushions/Embrace the ultimate comfort and *warmth* (*relaxation*) to every moment in your home!," which was followed the closing slide reads "Wrap Yourself in *Warmth* (*Comfort*)." There were no other differences in the visual elements (such as the images, colors, layouts, etc.) within the video between the discouraging and control feeling right conditions. People have a strong association between color hue and temperature such as warm (longer wavelength like red) and cold (shorter wavelength like blue) colors. Consequently, pale blue cushions are typically associated with coldness instead of warmth. Therefore, it is expected that consumers will feel wrong about the target product, whose color is pastel blue when it is described as warm. However, their sense of feeling right will not be influenced by the product being labelled as modern, relaxation, or comfort, descriptions that appear to be independent of sensory expectations.

Meanwhile, as for the excitement, the participants were randomly assigned the following instruction (discourage condition) or no instruction (control).

> *As you answer the following questions, please try to stay composed. Keep your emotions in check and maintain a calm mindset—this will help you respond more efficiently.*

After watching the video, the participants were asked to evaluate the video (unfavorable/favorable; uninteresting/interesting; dislike/like) and the product (the same three measurement items in Study 3) with 7-point scales, which were averaged to form a video and product evaluation scores (Cronbach's $\alpha_{video} = 0.941$, $\alpha_{product} = 0.963$) and to indicate their WTP for the product and the typical amount they usually pay for a cushion whose size was the same as the target product. WTP was incorporated as an additional evaluation measure to assess it in a multifaceted manner and enhance the validity of the results. The WTP for the target product and the typical price for a cushion were measured using a slider bar anchored with 0 and 100 GBP. Once the participants completed their rating of the video and the product, they were asked to indicate their excitement using the same items in Studies 1 and 2. Subsequently, the participants were asked about their sense of feeling right using two items adapted from Cesario et al. (2004), Wadhwa and Zhang (2015), and Wang et al. (2020) ("To what extent did you feel right while watching the video?" and "To what extent did you feel wrong while watching the video?" [reverse coded]) with 7-point scales anchored by not at all/very much. Excitement and feeling right scores were formed by averaging their items (Cronbach's $\alpha_{excitemet} = 0.889$, $r_{feeling\ right} = 0.681$). After that, the participants were asked to indicate their purchase frequency of the product category ("Do you have a lot of experience of purchasing cushions?") with a 7-point scale anchored by not at all/very often. Lastly, in addition to a question for a manipulation check ("Do you feel that the background music matched the cushions in the video?"), the participant's evaluation of the background music (bad/good; unfavorable/favorable; dislike/like; $\alpha = 0.979$), musical expertise, formal musical training, and demographic questions were measured.

6.2 | Results and Discussion

Data from 27 participants who did not meet our pre-set criteria, such as attention check failures were excluded; as a resultant, $n = 816$ (male = 44.0%, female = 54.5%, prefer not to say= 1 .0%, prefer to self-describe = 0.5%, $M_{age} = 38.8$ years, SD = 10.78). They had little or no formal musical training ($M = 0.93$ years, SD = 2.74), and most (96.1%) were neither undergraduate/graduate music students nor music professionals.

6.2.1 | Manipulation Check

We ran 2 (sensory congruence vs. incongruence) × 2 (feeling right: discourage vs. control) × 2 (excitement: discourage vs. control) analyses of covariance (ANCOVAs) with the participant's age, evaluation of the experimental music, the typical

TABLE 8 | Moderated mediation results on choice in Study 3.

Direct effects	B^a	SE	Z score	p value
Outcome variable: Feeling right to blue pen				
Clarinet timbre	0.908	0.572	1.588	0.114
Hand bell timbre	1.307	0.552	2.369	0.019
Frequency of using blue pen (FRQB)	0.492	0.171	2.369	0.005
Clarinet timbre × FRQB	−0.194	0.243	−0.797	0.427
Hand bell timbre × FRQB	−0.452	0.235	−1.926	0.056
Outcome variable: Feeling right to red pen				
Clarinet timbre	0.708	0.635	1.115	0.267
Hand bell timbre	−0.366	0.721	−0.507	0.613
Frequency of using red pen (FRQR)	0.114	0.113	1.004	0.317
Clarinet timbre × FRQR	−0.026	0.152	−0.171	0.865
Hand bell timbre × FRQR	0.080	0.165	0.482	0.631
Outcome variable: Choice of red pen				
Clarinet timbre in blue pen condition	1.038	0.573	1.812	0.070
Clarinet timbre in red pen condition	0.378	0.549	0.689	0.491
Hand bell timbre in blue pen condition	0.177	0.501	0.354	0.724
Hand bell timbre in red pen condition	−0.041	0.484	−0.084	0.933
Feeling right to blue pen	−0.526	0.129	−4.090	0.000
Feeling right to red pen	0.431	0.129	3.342	0.001
Indirect effects	B^a	**SE**	**LLCI**	**ULCI**
Low FRQ ($M_{blue} = 1.00$, $M_{red} = 1.88$)				
CL → FRB → CHOR	−0.376	0.254	−0.947	0.052
CL → FRR → CHOR	0.284	0.199	−0.070	0.730
HB → FRB → CHOR	−0.450	0.241	−0.988	−0.038
HB → FRR → CHOR	−0.093	0.214	−0.534	0.322
Average FRQ ($M_{blue} = 1.88$, $M_{red} = 3.90$)				
CL → FRB → CHOR	−0.258	0.213	−0.779	0.068
CL → FRR → CHOR	0.262	0.165	0.013	0.646
HB → FRB → CHOR	−0.240	0.194	−0.651	0.116
HB → FRR → CHOR	−0.024	0.161	−0.346	0.329
High FRQ ($M_{blue} = 3.32$, $M_{red} = 5.90$)				
CL → FRB → CHOR	−0.138	0.254	−0.691	0.329
CL → FRR → CHOR	0.239	0.242	−0.154	0.825
HB → FRB → CHOR	0.103	0.287	−0.442	0.686
HB → FRR → CHOR	0.045	0.258	−0.469	0.593

Abbreviations: CHOR, choice of the red pen; CL, clarinet timbre; FRB, feeling right to the blue pen; FRR, feeling right to the red pen; HB, hand bell timbre; LLCI, lower-limit confidence interval (95%); ULCI, upper-limit confidence interval (95%).
[a]Unstandardized coefficients.

amount they usually pay for a cushion, and purchase frequency of cushions as the covariates and the participant's perceived fit between the background music and the product, excitement, and a sense of feeling right as the DVs. The results revealed significant main effects of the sensory congruence on musical fit and excitement. The participants in the congruence (vs. incongruence) condition reported significantly higher musical fit with the product ($M_{congruence} = 4.06$, SD = 2.00, $M_{incongruence} = 3.90$, SD = 2.00; $F(1, 804) = 6.398$, $p = 0.012$, $\eta_P^2 = 0.008$) and excitement ($M_{congruence} = 5.09$, SD = 1.10, $M_{incongruence} = 4.96$, SD = 1.11; $F(1, 804) = 5.963$, $p = 0.015$, $\eta_P^2 = 0.007$). The main effects of the excitement condition on the participant's excitement ($M_{discourage} = 5.12$, SD = 1.06, $M_{control} = 4.92$, SD = 1.14; $F(1, 804) = 4.456$, $p = 0.035$, $\eta_P^2 = 0.006$) and feeling right ($M_{discourage} = 5.55$, SD = 1.32, $M_{control} = 5.22$, SD = 1.48; $F(1, 804) = 11.664$, $p < 0.001$, $\eta_P^2 = 0.014$) were also

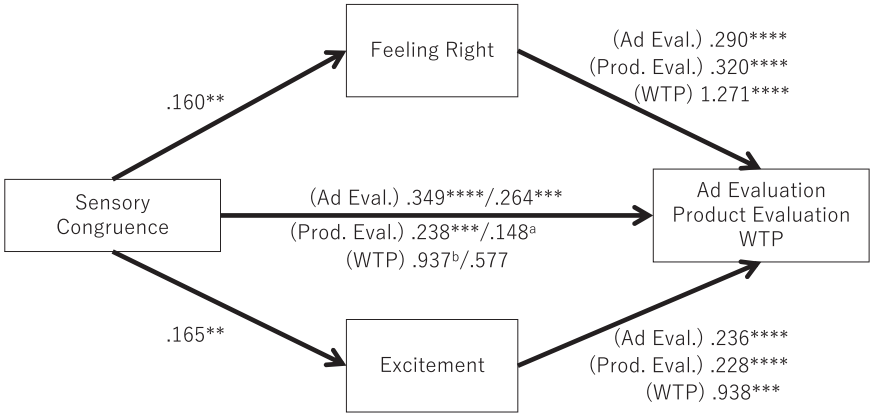

FIGURE 7 | Mediation model of sensory congruence in Study 4. Data from all the participants were used ($n = 816$). Unstandardized regression coefficients are reported. [a]$p = 0.071$, [b]$p = 0.099$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.001$.

significant, but the participants in the control (vs. discourage) condition were less excited and felt less right. As for feeling right conditions, although the participants in the control (vs. discourage) condition reported higher levels of feeling right, the difference was not significant ($M_{\text{discourage}} = 5.35$, $SD = 1.39$, $M_{\text{control}} = 5.41$, $SD = 1.44$; $F(1, 804) = 1.265$, $p = 0.261$, $\eta_P^2 = 0.002$). No other main or interaction effects were significant. These results on the excitement and feeling right were thought to be obtained, at least partially, by the sensory congruence manipulation, which was expected to enhance not only the fit between experimental music and product but also the participant's excitement and sense of feeling right.

6.2.2 | Mediating Role of Feeling Right and Excitement in Sensory Congruence Effects

To test our hypotheses (H1 to H3a), we tested the mediation model (model 4; Hayes 2018, see Figure 7) with sensory congruence as the IV, feeling right and excitement as parallel mediators, the participants' evaluations of video/product and WTP as the DVs. The model also included a variety of controls for demographic variables (age and gender), behavioral variables (purchase frequency of cushions and the typical amount in GBP they usually paid for a cushion), and psychological variables (evaluation of experimental music). Using all samples, the results revealed significant indirect effects (IEs) of the sensory congruence through both feeling right and excitement on all the DVs (for details, see Table 9).

Meanwhile, using data from discouraging feeling right conditions, meaning when feeling right were blocked (discouraged), the IEs through feeling right ($\beta_{\text{ad}} = 0.049$, 95% CI = [−0.002, 0.103]; $\beta_{\text{product}} = 0.054$, 95% CI = [−0.003, 0.119]; $\beta_{\text{WTP}} = 0.031$, 95% CI = [−0.001, 0.069]) or excitement ($\beta_{\text{ad}} = 0.017$, 95% CI = [−0.002, 0.048]; $\beta_{\text{product}} = 0.020$, 95% CI = [−0.004, 0.057]; $\beta_{\text{WTP}} = 0.010$, 95% CI = [−0.003, 0.032]) were not significant. The results were replicated in the discouraging excitement condition, meaning that the IEs through feeling right ($\beta_{\text{ad}} = 0.019$, 95% CI = [−0.033, 0.071]; $\beta_{\text{product}} = 0.021$, 95% CI = [−0.037, 0.078]; $\beta_{\text{WTP}} = 0.014$, 95% CI = [−0.025, 0.056]) or excitement ($\beta_{\text{ad}} = 0.021$, 95% CI = [−0.002, 0.060]; $\beta_{\text{product}} = 0.024$, 95% CI = [−0.003, 0.068]; $\beta_{\text{WTP}} = 0.012$, 95% CI = [−0.007, 0.044]) were not significant (for detail, see Supporting Information S2: File 10). These results suggest that the effects of sensory congruence on consumer responses such as ad/product evaluation and WTP were mediated by a sense of feeling right and excitement as shown in Figure 7, which supported H1 to H3a. Meanwhile, when a mediator was blocked, the sensory congruence effects were not significantly mediated by feeling right or excitement.

The manipulation checks indicated that our manipulations were not effective. However, the effects of sensory congruence on feeling right ($\beta = 0.161$, $t(402) = 1.883$, $p = 0.060$) and excitement ($\beta = 0.162$, $t(402) = 1.770$, $p = 0.078$) in discouraging feeling right and excitement conditions, respectively, were at borderline-significant levels. This suggests that the effects of sensory congruence on these mediators tended to be inhibited in the discourage conditions, consistent with the manipulations' intent to block these mediators. In light of this, it would seem that the discouragement manipulation-of-mediator design in Study 4 was not particularly robust. Therefore, while the results should be interpreted with caution, the replication evidence using different operationalizations of the mediators in Study 4 supported the idea that a sense of feeling right and excitement contributed to the outcomes reported here.

See also Supporting Information S2: File 11, where the results of 2 (sensory congruence vs. incongruence) × 2 (feeling right: discourage vs. control) × 2 (excitement: discourage vs. control) between participants ANCOVAs on the participants' evaluations of video/product and WTPs with the participant's age, evaluation of the music, the typical amount they usually pay for a cushion, and purchase frequency as the covariates were reported.

7 | Study 5

7.1 | Methods

Study 5 was conducted to test our hypotheses (H3b and H4) in a more realistic and controlled environment using a mock-up corner drugstore in Tokyo, Japan. A fabric softener brand (Lenor by P&G) and a potato crisps brand (Pure Potato by

TABLE 9 | Parallel mediation results in Study 4.

Direct effects	β[a]	SE	*t* value	*p* value
Outcome variable: Feeling right				
Sensory congruence	0.117	0.083	1.982	0.048
Age	0.140	0.004	4.701	0.000
Gender	0.004	0.077	0.145	0.885
Frequency of purchasing cushions	0.097	0.030	3.197	0.001
Typical amount paid for a cushion	0.009	0.003	0.294	0.769
Music evaluation	0.498	0.022	16.615	0.000
Outcome variable: Excitement				
Sensory congruence	0.145	0.071	2.267	0.024
Age	0.062	0.003	1.924	0.055
Gender	0.047	0.066	1.448	0.148
Frequency of purchasing cushions	0.200	0.026	6.044	0.000
Typical amount paid for a cushion	0.124	0.003	3.817	0.000
Music evaluation	0.289	0.018	8.888	0.000
Outcome variable: Ad (video) evaluation				
Sensory congruence	0.159	0.080	3.278	0.001
Feeling right	0.246	0.039	7.521	0.000
Excitement	0.158	0.045	5.213	0.000
Age	0.027	0.004	1.071	0.284
Gender	0.054	0.075	2.189	0.029
Frequency of purchasing cushions	−0.012	0.030	−0.452	0.652
Typical amount paid for a cushion	0.048	0.003	1.918	0.056
Music evaluation	0.459	0.024	16.135	0.000
Outcome variable: Product evaluation				
Sensory congruence	0.097	0.082	1.809	0.071
Feeling right	0.296	0.039	8.132	0.000
Excitement	0.165	0.046	4.906	0.000
Age	−0.050	0.004	−1.830	0.068
Gender	0.066	0.0763	2.434	0.015
Frequency of purchasing cushions	0.017	0.031	0.598	0.550
Typical amount paid for a cushion	0.031	0.003	1.132	0.258
Music evaluation	0.320	0.025	10.127	0.000
Outcome variable: WTP				
Sensory congruence	0.041	0.549	1.051	0.294
Feeling right	0.127	0.263	4.834	0.000
Excitement	0.074	0.310	3.027	0.003
Age	−0.044	0.026	−1.692	0.091
Gender	0.081	0.509	4.087	0.000
Frequency of purchasing cushions	0−.022	0.205	−1.060	0.289
Typical amount paid for a cushion	0.748	0.022	37.481	0.000
Music evaluation	0.126	0.165	5.532	0.000

Indirect effects	β[b]	SE	LLCI	ULCI
Sensory cong → feeling right → ad evaluation	0.048	0.015	0.001	0.059
Sensory cong → excitement → ad evaluation	0.023	0.011	0.003	0.048

(Continues)

TABLE 9 | (Continued)

Indirect effects	β^{b}	SE	LLCI	ULCI
Total	0.052	0.022	0.010	0.094
Pairwise contrasts of indirect effects	−0.001	0.015	−0.037	0.023
Sensory cong → feeling right → product Evaluation	0.035	0.018	0.001	0.070
Sensory cong → excitement → product evaluation	0.024	0.012	0.002	0.050
Total	0.058	0.025	0.010	0.108
Pairwise contrasts of indirect effects	−0.011	0.018	−0.047	0.024
Sensory cong → feeling right → WTP	0.015	0.009	0.000	0.034
Sensory cong → excitement → WTP	0.011	0.006	0.001	0.025
Total	0.026	0.011	0.004	0.049
Pairwise contrasts of indirect effects	−0.004	0.009	−0.024	0.013

Note: Sensory cong stands for sensory congruence, which was coded as a dummy variable: 0 = incongruence, 1 = congruence.
Abbreviations: LLCI, lower-limit confidence interval (95%); ULCI, upper-limit confidence interval (95%).
[a]Standardized coefficients.
[b]Partially standardized coefficients.

Koikeya) were selected as the target products, as both categories of product are commonly sold in corner drugstores in Japan. More importantly, they have items whose packaging colors are congruent with the perceptions of the auditory stimuli (i.e., timbre).

The pretest 1 of Study 1 and the pretest of Study 3 demonstrate that, compared to the hand bell timbre, the Japanese participants perceived the clarinet timbre as warmer, richer, and darker, and consequently, they felt that the clarinet timbre was more congruent with a warm, saturated, and dark color. The two target brands have items whose packaging color is red/light blue. The reds (warm color) are more saturated and darker than the light blues (cool color). Specifically, regarding Lenor, the HSL values of the red and light blue items are 356°, 67%, 56%, and 179°, 27%, 63%, respectively, while regarding Pure Potato, the HSL values of the red and light blue items are 357°, 66%, 37% and 192°, 21%, 66%, respectively (see Supporting Information S1: Appendix E). Each pair consists of the same brand's items, and consequently, shares similar visual design elements of the package other than color (i.e., layout, logo/typeface and size of product names, etc.).

Individuals aged between 20 and 59 years were recruited through the research panel of the Tokyo-based Japan Marketing Agency Co. Ltd. A total of 204 Japanese participants (male = 50.0%, female = 50.0%, $M_{age} = 44.8$ years, SD = 8.56) took part in the study and were rewarded 3000 JPN (approximately 27.27 USD). A 2 (timbre: clarinet vs. hand bells) × 2 (package color: red vs. light blue) mixed design with packaging color as the within-participants factor was conducted. A priori power analysis using G*Power (Faul et al. 2009) revealed that 197 observers would provide 0.80 power to detect a small-to-medium effect ($\omega = 0.20$; Cohen 1992) in a 2×2 χ^2 test with the alpha level at 0.05. The background auditory stimuli were the same as in Studies 3 and 4 (i.e., the Canon played by the clarinet/hand bells) and were played through the speaker (Beats Pill+ ML4M2PA/A by Beats Electronics) on a shelf around the center of the room so that the volume was kept in the range of 35–45 dB in front of the two target products' shelves (see Supporting Information S1: Appendix F). The experiment lasted 6 days, during which time the background music alternated between clarinet and handbell versions. The order of the clarinet and handbell versions was counterbalanced, meaning that each version started first on three different days during the experiment. In the mornings, the music changed every 1.5 h, while in the afternoons, it switched every 2 h. The location (right/left) on the shelf within each pair (i.e., red and light blue) of the products was counterbalanced. At the start of the study, the participants were given the following instructions:

> *Please imagine that you came to this corner drugstore to purchase potato crisps, a mask, and a fabric softener for 1,000 JPY. Please choose one of each and bring them to the checkout counter as you normally would as a consumer.*

The participants were free to walk around the store, meaning that the selection order of the three products was not controlled. 1000 JPY is approximately 9 USD. Purchasing a mask was a filler task. The prices were comparable with those of nearby corner drugstores. Specifically, the prices of Lenor and Pure Potato were 461 JPY (approximately 4.2 USD) and 122 JPY (approximately 1.1 USD), respectively, including tax. It was not until the participants came to the checkout counter to pay for the products that they had chosen that they were told that they did not actually have to pay for them.

Once they had completed the shopping task, they moved to another room to answer an online questionnaire. The questionnaire inquired about the purchase frequency of fabric softeners and potato crisps, along with several control variables. This included preferences for colors (red, dark pink, and light blue, representing the target products' package colors), scents (floral scent that is not too sweet/fresh rose and natural garden scent, corresponding to the light blue and red versions of Lenor, and flavors (salt and butter, which related to the light blue and red versions of Pure Potato). All responses were measured using a 7-point scale anchored by extremely dislike and extremely like.

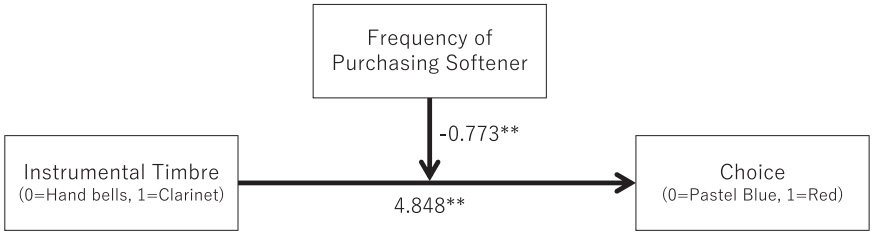

FIGURE 8 | Sensory congruence effect on choice of softener in Study 5. Unstandardized regression coefficients are reported. $^{**}p < 0.05$.

7.2 | Results

The data from one participant who did not choose either softener were excluded. No significant differences were observed in the participants' preferences for the colors, scents, and flavors amongst the experimental conditions ($|t|$s < 1.157, ps > 0.249).

The number of participants who chose the fabric softener with the red packaging was 41 (39.8%) and 35 (35.0%) in the clarinet and hand bells conditions, respectively. The difference was not significant ($\chi^2(1) = 0.500$, $p = 0.479$). To test the moderating role of the purchase frequency of the product, a simple moderation model (model 1; Hayes 2018, see Figure 8) with the purchase frequency of fabric softeners as the moderator, the instrumental timbre as the IV, and the choice as the DV, was conducted. The IV was coded as 0 or 1, representing the hand bells and clarinet, respectively, while the DV was also coded as 0 or 1, representing the choice of light blue and red, respectively. The results revealed a significant regression coefficient of instrumental timbre ($B_{\text{timbre}} = 4.848$, 95% CI = [0.427, 9.269], see Figure 8). The regression coefficient of the interaction by instrumental timbre and the purchase frequency of the category was also significant ($B_{\text{timbre} \times \text{frequency}} = -0.773$, 95% CI = [−1.500, −0.045]). Specifically, the conditional effect of instrumental timbre on the choice was significant ($B = 0.849$, 95% CI = [0.004, 1.694]) only when the participant's purchase frequency for the product category was low (1 SD below the mean = 5.18). This result is consistent with those of Study 3, indicating that the participants who had little experience buying softeners were more likely to choose a softener whose packaging color was congruent (vs. incongruent) with the timbre of background music. Meanwhile, the participants who had average or a lot of experience buying softeners did not show such a tendency. In other words, the effect of sensory congruency between color and timbre only occurs for those consumers who have relatively little experience in purchasing products in the category, which was supportive of H3b and H4.

However, the same simple moderation analysis (model 1; Hayes 2018) using data on the potato crisps revealed no significant regression coefficients on the choice of potato crisps ($B_{\text{timbre}} = 2.545$, 95% CI = [−2.569, 7.659], $B_{\text{timbre} \times \text{frequency}} = -0.404$, 95% CI = [−1.212, 0.404]). Meanwhile, using data merging the fabric softener and potato crisps together, the same simple moderation analysis revealed very similar results with those of fabric softener alone ($B_{\text{timbre}} = 4.052$, 95% CI = [0.791, 7.314], $B_{\text{timbre} \times \text{frequency}} = -0.642$, 95% CI = [−1.167, −0.116], for detail, see Supporting Information S2: File 12). Therefore, as a whole, it is reasonable to say that H3b and H4 was not supported.

7.3 | Discussion

In line with prior studies indicating that the effects of feeling right are prominent when prior experience and familiarity with a product is low (Dunn and Schweitzer 2005; Kim et al. 2009; Vaughn et al. 2010), the results of Study 3 identified a moderating role of purchase frequency of the product (category). However, the sensory congruence did not affect the choice of potato crisps. The results of the control test (for details, see Supporting Information S2: File 13) implied that this was mainly because the participants did not perceive there to be significant associations between the experimental audio stimuli and the Pure Potato's packaging colors. That said, given that people are generally familiar with potato crisps, it is possible that the participants already had a lot of experience in this category, which may have attenuated the connotative congruence effect. In fact, the participants reported that they had bought potato crisps (vs. softener) more frequently ($M_{\text{crisps}} = 6.28$, SD = 0.70, $M_{\text{softener}} = 5.99$, SD = 0.82, $t(203) = 4.756$, $p < 0.001$, $d = 0.88$).

8 | General Discussion

8.1 | Conclusions and Theoretical Implications

This study provides evidence that the connotative congruence between instrumental timbre and visually-displayed materials can affect consumers' decision-making. Perhaps surprisingly, this topic has received relatively little research interest from the field of sensory marketing until recently. The current study expands the literature on crossmodal correspondence by unveiling a novel type of correspondence and its influence on consumer behavior using comprehensive sets of descriptions. While prior related studies used sound logos as audiovisual experimental stimuli (Melzner and Raghubir 2023; Puligadda and VanBergen 2023; Techawachirakul et al. 2023), in the present study, convergent evidence was obtained using different complex auditory stimuli created from Pachelbel's Canon and Chopin's Etude and across different visual stimuli (book covers, typefaces, colors of ballpoint pens and cushions, and packaging colors) and comprehensive sets of connotative descriptions (luminance, texture, and mass). It was unclear whether the findings of those prior studies conducted online would be applicable in a real-world setting. The novelty of the present study lies in the fact that the external validity of the effects of timbre-vision connotative congruency was enhanced by measuring behavioral outcomes in the mock-up corner drugstore. In addition to aggregation level analyses, the proposed effects were tested through individual-level analyses, which also makes the current study unique.

To the best of our knowledge, no previous research has empirically demonstrated the connotative congruency—feeling right and excitement—favorable evaluation/choice chain. In other words, the present study is the first marketing study conducting multiple experiments and consistently demonstrating that timbre-vison connotative congruency can have positive effects on the consumers' sense of feeling right and excitement, which, in turn, positively influence their evaluations and choices. Although a large number of studies have demonstrated that certain crossmodal associations are derived from common emotional associations, this study revealed that an emotional state (i.e., excitement) can be the consequence as well, which has not often been tested previously. In the studies reported here, most total (i.e., X–Y) effects of the connotative congruency between instrumental timbre and visual features on decision-making were not significant. This indicates that the proposed effect may not be strong enough to directly impact consumers' decision-making. However, this does not mean that timbre is not worth paying attention in the marketing context, which was demonstrated by a highly realistic and, thus, ecologically valid study (Study 5).

In the field of marketing communications, the notion of integrated marketing communications (IMC) is popular. According to the latter, all of the elements in the promotional mix elements, such as advertising, sales promotion, publicity, and personal selling, should be integrated into a unified whole that "speaks with a single voice" (Shimp 2000). Adopting such a strategy has long been known as a key success factor (e.g., Schultz et al. 1993). The present study indicates that integrating the range of sensory elements so that they work in harmony with each other, or "integrated sensory communications (ISC)" will open a new window to improve the effectiveness of the total marketing communications effort. Although IMC has not focused on integration at the level of multisensory perception, ISC nevertheless provides a useful perspective to implement the IMC.

The study is unique in that convergent evidence was obtained from (1) both aggregate and individual level analyses and (2) behavioral outcomes as the DV in a highly realistic mock-up store. To date, the majority of marketing studies on musical timbre have investigated the effects on product evaluations (Imschloss and Kuehnl 2017, 2019; Oakes and North 2006; Sunaga et al. 2020), with relatively few studies having attempted to measure consumers' choice, as conducted in the present study.

In addition, this study ruled out an alternative explanation in terms of processing fluency (Studies 1 and 2). Some of the published research has shown that individuals can process auditory and visual stimuli more fluently when they are congruent with each other in terms of crossmodal correspondences (Spence 2012; Sunaga 2018). It is possible that the participants in the current study could process the experimental stimuli easily regardless of the congruence since the stimuli were simple. In fact, the average fluency score was relatively high ($M_{\text{study 1}} = 5.97$, SD = 1.43, $M_{\text{study 2\_UK}} = 5.22$, SD = 1.28, $M_{\text{study 2\_AFR}} = 5.16$, SD = 1.43). Consequently, that is possibly why the study did not reveal a significant mediating role of fluency. On the other hand, the results of the present study indicated that even if the stimuli were very simple, our participants experienced a "feeling of rightness" when the auditory and visual stimuli were congruent. This study revealed that congruency in terms of connotative meaning (Parncutt 2014; Walker and Walker 2012) can improve consumers' evaluations and preferences even if it does not enhance processing fluency, meaning that the effects occur even under very simple conditions.

As suggested by Cesario et al. (2004), using the parallel mediation models, the study provides evidence implying that feeling right can be differentiated from affective responses. Meanwhile, Avnet and Higgins (2006) and Schwarz (2006) both suggested that feeling right increased an individual's confidence in their own cognitive or affective reactions, which led people to make more extreme judgments. To investigate this relationship between feeling right and affective responses, using the data from Studies 1 and 2, we ran serial mediation analyses (model 6; Hayes 2018) with the timbre-vison congruence as the IV, feeling right as the first mediator, excitement as the second mediator, liking of the video as the third mediator (only in Study 1), and product evaluations as the DV. The results revealed that in Study 1, the indirect effect of sensory congruence—feeling right—excitement—liking of the video—product evaluations chain was not significant ($\beta = 0.004$, 95% CI = [−0.002, 0.012]). Similarly, another serial mediation analysis with the timbre-vison congruence as the IV, feeling right as the first mediator, excitement as the second mediator, and product evaluations as the DV also revealed that the indirect path of sensory congruence—feeling right—excitement—product evaluations chain was not significant ($\beta = 0.015$, 95% CI = [−0.000, 0.038]). Meanwhile, in Study 2, the indirect effects of sensory congruence—feeling right—excitement—evaluations/liking chain were significant ($\beta_{\text{evaluation}} = -0.036$, 95% CI = [−0.063, −0.011]; $\beta_{\text{liking}} = -0.035$, 95% CI = [−0.061, −0.011]). Therefore, it remains unclear whether the mediating role of feeling right and affective response is indeed parallel or serial, which needs further research.

8.2 | Managerial Implications

The findings of the present study will likely be useful for not only academic scholars but also marketing practitioners who consider using marketing to solve social issues, foster consumers' health and well-being, or shape prosocial behaviors (Laukkanen et al. 2022). The findings of the present study might be used to help nudge people to choose particular options. For instance, using the connotative congruence between background music and the texture of its cover, publishers could enhance consumers' evaluations of a book addressing a social issue such as racist discrimination and encourage them to choose, read, or share it. In the same vein, the findings could also be used to help nudge consumers to choose healthy foods, which tend to be associated with a poor taste. The sensory congruence between background music and the color or texture of brand-logos on the package could potentially be used to enhance the evaluation of healthy foods and unconsciously encourage people to choose, eat, or recommend them. It is noteworthy that Venkatesan et al. (2022) demonstrated that when music was perceived as having a very round or angular sound, the effect of curved typeface on the CD cover on

consumers' expectations of how pleasant the music would sound was diminished (see also Lee and Spence 2023).

The approach to assessing congruency outlined in this study may turn out to be beneficial in regard to selecting music for use in marketing communications. Brand managers should typically select musical instruments that are congruent with consumers' perception of visual elements of the brand, such as package/content color (not only hue but also saturation and lightness), logo, and visually presented material on the surface. Using background music played by an instrument (e.g., organ) that fits with the material image (e.g., carpet) and/or color (e.g., warm, saturated, dark color like mahogany) of the floor, furniture, wall, etc., a store such as a book store, mobile shop, or experiential retail like Google store, Apple store, and Nike Rise could elicit a sense of feeling right and thereafter favorable customer responses (Keller and Spence 2023; Spence and Keller 2024).

The results of the present study advance the current understanding of how congruencies between different sensory inputs affect consumers' decision-making. The moderating role of usage or purchase frequency of the product (category), which was identified in Studies 3 and 5, suggests that the timbre-vision connotative congruency effect might be prominent in non-, light, or middle (vs. heavy) users of the product, trial (vs. repeat) purchases, or novel (vs. familiar) stimuli for consumers such as new products and innovations. This partially explains why the total effects of timbre-vision congruencies on the DVs in the mediation models tended to be nonsignificant.

Although this study used music that was played by a single instrument as the auditory stimulus, we did not focus on a particular musical instrumental timbre as such, but on the congruence of sensory (connotative) associations. Therefore, it would seem likely that the approach reported here can be extended to music that is played by multiple instruments, such as orchestral music and human voices, which have also attracted little attention from marketing scholars with a few rare exceptions (e.g., Lowe and Haws 2017). This study conceptualized (and actually measured in Study 2) sensory congruency as the degree to which the same connotations are shared between different sensory modalities. The idea has been adapted in the field of self-congruency (see Kressmann et al. 2006), and the current study extended it to audio-visual congruency. The basic concept and method for the measurement of congruency reported here will not be confined to audiovisual congruency but applicable to sensory congruency in general (e.g., audio and touch, vision and olfactory, touch and taste).

8.3 | Limitations and Future Research

The study also has some limitations: Although this study (Control test 2 of Study 1) ruled out the alternative possibility that the timbre-vision congruency is obtained through emotional or familiarity mediation, one could question whether the results reported here should exclusively be attributed to the timbre-vision connotative congruency, meaning that the results could be obtained through just emotional mediation suggested by Palmer et al. (2013). Emotional mediation is often a potential factor underlying crossmodal correspondences (Spence 2020a) and will play a major role in the case of stimuli where people can easily attach emotional meaning to the stimuli (e.g., as in the case of music and fragrance) than those which are simple (e.g., pure tones and lines). It is possible that both connotative congruency and emotional mediation underly a crossmodal correspondence. In that case, we assume that the connotative congruency would come to the fore because emotional mediation is thought to occur when the perceiver cannot attribute a matching two sensory features to any other more obvious route. This is a prospective assumption but waits to be investigated.

As mentioned earlier, the participants' excitement was not measured in Study 3 due to the procedure. However, considering the aim of the lab experiment was to test the proposed effects under more controlled conditions, an experimental design that can investigate the mediating role of excitement as well should be developed. It will be possible to investigate the effects of sensory congruency on consumers' emotional states using biometric monitoring such as facial action coding system and electroencephalography (EEG) (De Keyser et al. 2021; Steele et al. 2013). These tools make it possible to measure unconscious responses of human body parameters and assess the participants' emotional states, like excitement, instead of relying on reflective questions, regardless of whether the design happens to be within- or between-participants.

Although lying outside the scope of the present study, cultural differences also deserve to be investigated. Pretest 1 of Study 1 revealed that individuals from different countries perceived the connotations of individual instrumental timbres somewhat differently. In Study 2, the effect of connotative congruence on excitement only failed to reach significance in the United Kingdom. Further research is therefore needed to investigate what explains these regional differences.

Exploring additional moderators beyond product usage or purchase frequency is promising, particularly regarding timbre-visual and other audio-visual connotative congruency effects. Superadditive multisensory interactions occur when congruent individual sensory stimuli create a greater impact on consumer perceptions and responses than the sum of these stimuli, particularly when they are weak (Spence et al. 2014). This suggests that examining the role of sensory stimulus intensity in moderating the timbre-visual connotative congruency effect would be fruitful for future research.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP20K02002 and JP19H00601.

Data Availability Statement

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/6jb59/?view_only=32b87e3cdb18415ab261689a416036e8.

References

Abdolmohamad Sagha, M., N. Seyyedamiri, P. Foroudi, and M. Akbari. 2022. "The One Thing You Need to Change Is Emotions: The Effect of Multi-Sensory Marketing on Consumer Behavior." *Sustainability* 14, no. 4: 2334.

Adeli, M., J. Rouat, and S. Molotchnikoff. 2014. “Audiovisual Correspondence Between Musical Timbre and Visual Shapes.” *Frontiers in Human Neuroscience* 8: 352.

Ahn, J., A. Kim, and Y. Sung. 2020. “The Effects of Sensory Fit on Consumer Evaluations of Co-Branding.” *International Journal of Advertising* 39, no. 4: 486–503.

Alter, A. L., and D. M. Oppenheimer. 2008. “Effects of Fluency on Psychological Distance and Mental Construal (or Why New York Is a Large City, but New York Is a Civilized Jungle).” *Psychological Science* 19, no. 2: 161–167.

Anikin, A., and N. Johansson. 2019. “Implicit Associations Between Individual Properties of Color and Sound.” *Attention, Perception, & Psychophysics* 81: 764–777.

Avnet, T., and E. T. Higgins. 2006. “How Regulatory Fit Affects Value in Consumer Choices and Opinions.” *Journal of Marketing Research* 43, no. 1: 1–10.

Avnet, T., D. Laufer, and E. T. Higgins. 2013. “Are All Experiences of Fit Created Equal?, Two Paths to Persuasion.” *Journal of Consumer Psychology* 23, no. 3: 301–316.

Baumeister, R. F., K. D. Vohs, and D. C. Funder. 2007. “Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior?” *Perspectives on Psychological Science* 2, no. 4: 396–403.

Biliciler, G., R. Raghunathan, and A. F. Ward. 2022. “Consumers as Naive Physicists: How Visual Entropy Cues Shift Temporal Focus and Influence Product Evaluations.” *Journal of Consumer Research* 48, no. 6: 1010–1031.

Biswas, D. 2019. “Sensory Aspects of Retailing: Theoretical and Practical Implications.” *Journal of Retailing* 95, no. 4: 111–115.

Biswas, D., L. I. Labrecque, and D. R. Lehmann. 2021. “Effects of Sequential Sensory Cues on Food Taste Perception: Cross-Modal Interplay Between Visual and Olfactory Stimuli.” *Journal of Consumer Psychology* 31, no. 4: 746–764.

Biswas, D., K. Lund, and C. Szocs. 2019. “Sounds Like a Healthy Retail Atmospheric Strategy: Effects of Ambient Music and Background Noise on Food Sales.” *Journal of the Academy of Marketing Science* 47, no. 1: 37–55.

Brunetti, R., A. Indraccolo, C. Del Gatto, C. Spence, and V. Santangelo. 2018. “Are Crossmodal Correspondences Relative or Absolute?, Sequential Effects on Speeded Classification.” *Attention, Perception, & Psychophysics* 80, no. 2: 527–534.

Camacho, C. J., E. T. Higgins, and L. Luger. 2003. “Moral Value Transfer From Regulatory Fit: What Feels Right Is Right and What Feels Wrong Is Wrong.” *Journal of Personality and Social Psychology* 84, no. 3: 498–510.

Cesario, J., H. Grant, and E. T. Higgins. 2004. “Regulatory Fit and Persuasion: Transfer From Feeling Right.” *Journal of Personality and Social Psychology* 86, no. 3: 388–404.

Chan, E. Y., and G. Northey. 2021. “Luxury Goods In Online Retail: How High/Low Positioning Influences Consumer Processing Fluency and Preference.” *Journal of Business Research* 132: 136–145.

Cheng, F.-F., C.-S. Wu, and D. C. Yen. 2009. “The Effect of Online Store Atmosphere on Consumer's Emotional Responses: An Experimental Study of Music and Colour.” *Behaviour & Information Technology* 28, no. 4: 323–334.

Chiou, R., and A. N. Rich. 2012. “Cross-Modality Correspondence Between Pitch and Spatial Location Modulates Attentional Orienting.” *Perception* 41, no. 3: 339–353.

Cohen, J. 1992. “A Power Primer.” *Psychological Bulletin* 112, no. 1: 155–159.

Collier, W. G., and T. L. Hubbard. 2001. “Judgments of Happiness, Brightness, Speed and Tempo Change of Auditory Stimuli Varying in Pitch and Tempo.” *Psychomusicology* 17, no. 1/2: 36–55.

Crisinel, A.-S., C. Jacquier, O. Deroy, and C. Spence. 2013. “Composing With Cross-Modal Correspondences: Music and Odors in Concert.” *Chemosensory Perception* 6: 45–52.

Crisinel, A.-S., and C. Spence. 2010. “As Bitter as a Trombone: Synesthetic Correspondences in Nonsynesthetes Between Tastes/Flavors and Musical Notes.” *Attention, Perception & Psychophysics* 72: 1994–2002.

Crisinel, A.-S., and C. Spence. 2012a. “A Fruity Note: Crossmodal Associations Between Odors and Musical Notes.” *Chemical Senses* 37: 151–158.

Crisinel, A.-S., and C. Spence. 2012b. “Assessing the Appropriateness of ‘Synaesthetic’ Messaging on Product Packaging.” *Food Quality and Preference* 26, no. 1: 45–51.

De Keyser, A., Y. Bart, X. Gu, S. Q. Liu, S. G. Robinson, and P. K. Kannan. 2021. “Opportunities and Challenges of Using Biometrics for Business: Developing a Research Agenda.” *Journal of Business Research* 136: 52–62.

Deng, X., B. Han, and L. Wang. 2019. “Up-Down Versus Left-Right: The Effect of Writing Direction Change in East Asia on Consumers' Perceptions and Advertising.” *Journal of Advertising* 48, no. 5: 437–456.

Dunn, J. R., and M. E. Schweitzer. 2005. “Feeling and Believing: The Influence of Emotion on Trust.” *Journal of Personality and Social Psychology* 88, no. 5: 736–748.

Eitan, Z., and R. Y. Granot. 2006. “How Music Moves.” *Music Perception* 23, no. 3: 221–248.

Eitan, Z., and I. Rothschild. 2011. “How Music Touches: Musical Parameters and Listeners' Audio-Tactile Metaphorical Mappings.” *Psychology of Music* 39, no. 4: 449–467.

Eitan, Z., A. Schupak, A. Gotler, and L. E. Marks. 2014. “Lower Pitch Is Larger, Yet Falling Pitches Shrink.” *Experimental Psychology* 61, no. 4: 273–284.

Eklund, A. A., and M. Helmefalk. 2022. “Congruency or Incongruency: A Theoretical Framework and Opportunities for Future Research Avenues.” *Journal of Product & Brand Management* 31, no. 4: 606–621.

Elder, R. S., and A. Krishna. 2022. “A Review of Sensory Imagery for Consumer Psychology.” *Journal of Consumer Psychology* 32, no. 2: 293–315.

Elder, R. S., A. E. Schlosser, M. Poor, and L. Xu. 2017. “So Close I Can Almost Sense It: The Interplay Between Sensory Imagery and Psychological Distance.” *Journal of Consumer Research* 44, no. 4: 877–894.

Evans, K. K., and A. Treisman. 2010. “Natural Cross-Modal Mappings Between Visual and Auditory Features.” *Journal of Vision* 10, no. 1: 6.

Faseur, T., and M. Geuens. 2006. “Different Positive Feelings Leading to Different Ad Evaluations: The Case of Coziness, Excitement, and Romance.” *Journal of Advertising* 35, no. 4: 129–142.

Faul, F., E. Erdfelder, A. Buchner, and A.-G. Lang. 2009. “Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses.” *Behavior Research Methods* 41, no. 4: 1149–1160.

Fenko, A., H. N. J. Schifferstein, and P. Hekkert. 2010. “Shifts In Sensory Dominance Between Various Stages of User-Product Interactions.” *Applied Ergonomics* 41, no. 1: 34–40.

Fiske, S. T. 1982. “Schema-Triggered Affect: Applications to Social Perception.” In *Affect and Cognition: The 17th Annual Carnegie Symposium on Cognition*, Edited by M. S. Clark and S. Fiske, 55–78. Lawrence Eribaum Associates.

Flavián, C., S. Ibáñez-Sánchez, and C. Orús. 2021. “The Influence of Scent on Virtual Reality Experiences: The Role of Aroma-Content Congruence.” *Journal of Business Research* 123: 289–301.

Gardner, M. P. 1985. “Does Attitude Toward the Ad Affect Brand Attitude Under a Brand Evaluation Set?” *Journal of Marketing Research* 22, no. 2: 192–198.

Ge, X. 2023. "Experimentally Manipulating Mediating Processes: Why and How to Examine Mediation Using Statistical Moderation Analyses." *Journal of Experimental Social Psychology* 109: 104507.

Hagtvedt, H., and S. A. Brasel. 2016. "Cross-Modal Communication: Sound Frequency Influences Consumer Responses to Color Lightness." *Journal of Marketing Research* 53, no. 4: 551–562.

Hall, E. T. 1966. *The Hidden Dimension*. Doubleday.

Hamilton-Fletcher, G., C. Witzel, D. Reby, and J. Ward. 2017. "Sound Properties Associated With Equiluminant Colours." *Multisensory Research* 30, no. 3–5: 337–362.

Hayes, A. F. 2018. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (2nd ed.) Guilford Publications.

Huang, X., and A. A. Labroo. 2020. "Cueing Morality: The Effect of High-Pitched Music on Healthy Choice." *Journal of Marketing* 84, no. 6: 130–143.

Hung, K. 2000. "Narrative Music in Congruent and Incongruent TV Advertising." *Journal of Advertising* 29, no. 1: 25–34.

Hutmacher, F. 2019. "Why Is There so Much More Research on Vision Than on Any Other Sensory Modality?" *Frontiers in Psychology* 10: 2246.

Imschloss, M., and C. Kuehnl. 2017. "Don't Ignore the Floor: Exploring Multisensory Atmospheric Congruence Between Music and Flooring in a Retail Environment." *Psychology & Marketing* 34, no. 10: 931–945.

Imschloss, M., and C. Kuehnl. 2019. "Feel the Music! Exploring the Cross-Modal Correspondence Between Music and Haptic Perceptions of Softness." *Journal of Retailing* 95, no. 4: 158–169.

Isbilen, E. S., and C. L. Krumhansl. 2016. "The Color of Music: Emotion-Mediated Associations to Bach's Well-Tempered Clavier." *Psychomusicology: Music, Mind, and Brain* 26, no. 2: 149–161.

Jiang, W., J. Liu, X. Zhang, S. Wang, and Y. Jiang. 2020. "Analysis and Modeling of Timbre Perception Features in Musical Sounds." *Applied Sciences* 10, no. 3: 789.

Julesz, B., and I. J. Hirsh. 1972. "Visual and Auditory Perception: An Essay of Comparison." In *Human Communication: A Unified View*, Edited by E. E. David and P. B. Denes, 283–340. McGraw-Hill.

Karataş, M., and Z. Gürhan-Canli. 2020. "A Construal Level Account of the Impact of Religion and God on Prosociality." *Personality and Social Psychology Bulletin* 46, no. 7: 1107–1120.

Kazazis, S., P. Depalle, and S. McAdams. 2022. "Interval and Ratio Scaling of Spectral Audio Descriptors." *Frontiers in Psychology* 13: 835401.

Keller, S., and C. Spence. 2023. "Sounds Like Branding: Cognitive Principles & Crossmodal Considerations for the Design of Successful Sonic Logos." *Expert Marketing Journal* 11, no. 2: 91–117.

Kim, H., A. R. Rao, and A. Y. Lee. 2009. "It's Time to Vote: The Effect of Matching Message Orientation and Temporal Frame on Political Persuasion." *Journal of Consumer Research* 35, no. 6: 877–889.

Knoferle, K. M., E. R. Spangenberg, A. Herrmann, and J. R. Landwehr. 2012. "It Is All in the Mix: The Interactive Effect of Music Tempo and Mode on in-Store Sales." *Marketing Letters* 23, no. 1: 325–337.

Kraiss, J. T., E. T. Bohlmeijer, and M. Schotanus-Dijkstra. 2023. "What Emotions to Encourage? The Role of High and Low Arousal Positive Emotions in Three Randomized Controlled Trials of Different Positive Psychology Interventions." *International Journal of Applied Positive Psychology* 8: 257–283.

Kressmann, F., M. J. Sirgy, A. Herrmann, F. Huber, S. Huber, and D.-J. Lee. 2006. "Direct and Indirect Effects of Self-Image Congruence on Brand Loyalty." *Journal of Business Research* 59, no. 9: 955–964.

Krishna, A., R. S. Elder, and C. Caldara. 2010. "Feminine to Smell but Masculine to Touch? Multisensory Congruence and Its Effect on the Aesthetic Experience." *Journal of Consumer Psychology* 20, no. 4: 410–418.

Kuo, A., and D. H. Rice. 2015. "The Impact of Perceptual Congruence on the Effectiveness of Cause-Related Marketing Campaigns." *Journal of Consumer Psychology* 25, no. 1: 78–88.

Landwehr, J. R., D. Wentzel, and A. Herrmann. 2013. "Product Design for the Long Run: Consumer Responses to Typical and Atypical Designs at Different Stages of Exposure." *Journal of Marketing* 77, no. 5: 92–107.

Laukkanen, T., N. Xi, H. Hallikainen, N. Ruusunen, and J. Hamari. 2022. "Virtual Technologies in Supporting Sustainable Consumption: From a Single-Sensory Stimulus to a Multi-Sensory Experience." *International Journal of Information Management* 63: 102455.

Lee, A. Y., P. A. Keller, and B. Sternthal. 2010. "Value From Regulatory Construal Fit: The Persuasive Impact of Fit Between Consumer Goals and Message Concreteness." *Journal of Consumer Research* 36, no. 5: 735–747.

Lee, A. Y., and A. A. Labroo. 2004. "The Effect of Conceptual and Perceptual Fluency on Brand Evaluation." *Journal of Marketing Research* 41, no. 2: 151–165.

Lee, B. P., and C. Spence. 2023. "Combining Typeface and Color to Prime Specific Taste Expectations." *Psychology of Aesthetics, Creativity, and the Arts*, ahead of print.

Li, Y., and M. Pandelaere. 2021. "The Denomination–Spending Matching Effect." *Journal of Business Research* 128: 338–349.

Lidji, P., R. Kolinsky, A. Lochy, and J. Morais. 2007. "Spatial Associations for Musical Stimuli: A Piano in the Head?" *Journal of Experimental Psychology: Human Perception and Performance* 33, no. 5: 1189–1207.

Lindborg, P., and A. K. Friberg. 2015. "Colour Association With Music Is Mediated By Emotion: Evidence From an Experiment Using a CIE Lab Interface and Interviews." *PLoS One* 10: e0144013.

Lowe, M. L., and K. L. Haws. 2017. "Sounds Big: The Effects of Acoustic Pitch on Product Perceptions." *Journal of Marketing Research* 54, no. 2: 331–346.

Luangrath, A. W., J. Peck, W. Hedgcock, and Y. Xu. 2022. "Observing Product Touch: The Vicarious Haptic Effect in Digital Marketing and Virtual Reality." *Journal of Marketing Research* 59, no. 2: 306–326.

MacKenzie, S. B., and R. J. Lutz. 1989. "An Empirical Examination of the Structural Antecedents of Attitude Toward the Ad in an Advertising Pretesting Context." *Journal of Marketing* 53, no. 2: 48–65.

MacKenzie, S. B., R. J. Lutz, and G. E. Belch. 1986. "The Role of Attitude Toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations." *Journal of Marketing Research* 23, no. 2: 130–143.

Magezi, D. A. 2015. "Linear Mixed-Effects Models for Within-Participant Psychology Experiments: An Introductory Tutorial and Free, Graphical User Interface (LMMgui)." *Frontiers in Psychology* 6: 2.

Mahdavi, M., B. Barbosa, Z. Oliveira, and V. Chkoniya. 2020. "Sounds of Scents: Olfactory-Auditory Correspondences in the Online Purchase Experience for Perfume." *Review of Business Management* 22, no. 4: 836–853.

McAdams, S. 2019. "The Perceptual Representation of Timbre." In *Timbre: Acoustics, Perception, and Cognition*, Edited by K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper and R. R. Fay, 23–57. Springer International Publishing.

Melzner, J., and P. Raghubir. 2023. "The Sound of Music: The Effect of Timbral Sound Quality in Audio Logos on Brand Personality Perception." *Journal of Marketing Research* 60, no. 5: 932–949.

Monahan, L., and M. Romero. 2020. "Heading the Right Way? The Influence of Motion Direction in Advertising on Brand Trust." *Journal of Advertising* 49, no. 3: 250–269.

Mookerjee, S., Y. Cornil, and J. Hoegg. 2021. “From Waste to Taste: How “Ugly” Labels Can Increase Purchase of Unattractive Produce.” *Journal of Marketing* 85, no. 3: 62–77.

Morales, A. C., O. Amir, and L. Lee. 2017. “Keeping It Real in Experimental Research: Understanding When, Where, and How to Enhance Realism and Measure Consumer Behavior.” *Journal of Consumer Research* 44, no. 2: 465–476.

Motoki, K., and A. Pathak. 2023. “The Length of Brand Names Influences the Expectation of Healthiness in Foods and Preference for Healthy Foods.” *Psychology & Marketing* 40: 1850–1862.

Nelson, M. R., and J. C. Hitchon. 1999. “Loud Tastes, Colored Fragrances, and Scented Sounds: How and When to Mix the Senses In Persuasive Communications.” *Journalism & Mass Communication Quarterly* 76, no. 2: 354–372.

Oakes, S., and A. C. North. 2006. “The Impact of Background Musical Tempo and Timbre Congruity Upon Ad Content Recall and Affective Response.” *Applied Cognitive Psychology* 20, no. 4: 505–520.

Oppenheimer, D. M. 2008. “The Secret Life of Fluency.” *Trends in Cognitive Sciences* 12, no. 6: 237–241.

Palmer, S. E., T. A. Langlois, and K. B. Schloss. 2016. “Music-to-Color Associations of Single Line Piano Melodies in Non-Synesthetes.” *Multisensory Research* 29: 157–193.

Palmer, S. E., K. B. Schloss, Z. Xu, and L. R. Prado-León. 2013. “Music-Color Associations Are Mediated by Emotion.” *Proceedings of the National Academy of Sciences* 110, no. 22: 8836–8841.

Parise, C. V., and C. Spence. 2009. “‘When Birds of a Feather Flock Together’: Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes.” *PLoS One* 4, no. 5: e5664.

Parncutt, R. 2014. “The Emotional Connotations of Major Versus Minor Tonality: One or More Origins?” *Musicae Scientiae* 18, no. 3: 324–353.

Petty, R. E., and J. T. Cacioppo. 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. Springer/Verlag.

Pirlott, A. G., and D. P. MacKinnon. 2016. “Design Approaches to Experimental Mediation.” *Journal of Experimental Social Psychology* 66: 29–38.

Puligadda, S., and N. VanBergen. 2023. “The Influence of Sound Logo Instruments on Brand Personality Perceptions: An Investigation of Brand Ruggedness and Sophistication.” *Journal of Business Research* 156: 113531.

Qi, Y., F. Huang, Z. Li, and X. Wan. 2020. “Crossmodal Correspondences in the Sounds of Chinese Instruments.” *Perception* 49, no. 1: 81–97.

Reber, R., N. Schwarz, and P. Winkielman. 2004. “Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?” *Personality and Social Psychology Review* 8, no. 4: 364–382.

Reber, R., and C. Unkelbach. 2010. “The Epistemic Status of Processing Fluency as Source for Judgments of Truth.” *Review of Philosophy and Psychology* 1, no. 4: 563–581.

Rodaway, P. 1994. *Sensuous Geographies: Body, Sense and Place*. Routledge.

Rozin, P., and J. W. Kalat. 1972. “Learning as a Situation-Specific Adaptation.” In *The Biological Boundaries of Learning*, Edited by M. E. P. Seligman and J. L. Hager, 66–96. Appleton-Century-Crofts.

Rusconi, E., B. Kwan, B. Giordano, C. Umilta, and B. Butterworth. 2006. “Spatial Representation of Pitch Height: The SMARC Effect.” *Cognition* 99, no. 2: 113–129.

Ruzeviciute, R., B. Kamleitner, and D. Biswas. 2020. “Designed to s(m)ell: When Scented Advertising Induces Proximity and Enhances Appeal.” *Journal of Marketing Research* 57, no. 2: 315–331.

Saitis, C., and S. Weinzierl. 2019. “The Semantics of Timbre.” In *Timbre: Acoustics, Perception, and Cognition*, Edited by K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper and R. R. Fay, 119–149. Springer International Publishing.

Saitis, C., S. Weinzierl, K. von Kriegstein, S. Ystad, and C. Cuskley. 2020. “Timbre Semantics Through the Lens of Crossmodal Correspondences: A New Way of Asking Old Questions.” *Acoustical Science and Technology* 41, no. 1: 365–368.

Salgado-Montejo, A., C. Velasco, J. S. Olier, J. Alvarado, and C. Spence. 2014. “Love for Logos: Evaluating the Congruency Between Brand Symbols and Typefaces and Their Relation to Emotional Words.” *Journal of Brand Management* 21, no. 7–8: 635–649.

Schwarz, N. 2006. “Feelings, Fit, and Funny Effects: A Situated Cognition Perspective.” *Journal of Marketing Research* 43, no. 1: 20–23.

Schultz, D. E., S. Tannenbaum, and R. Lauterborn. 1993. *Integrated Marketing Communications*. NTC Business Books.

Shapiro, K. L., B. Egerman, and R. M. Klein. 1984. “Effects of Arousal on Human Visual Dominance.” *Perception & Psychophysics* 35, no. 6: 547–552.

Shimp, T. 2000. *Advertising, Promotion, and Supplemental Aspects of Integrated Marketing Communications* (5th ed.). Dryden Press.

Spence, C. 2011. “Crossmodal Correspondences: A Tutorial Review.” *Attention, Perception, & Psychophysics* 73, no. 4: 971–995.

Spence, C. 2012. “Managing Sensory Expectations Concerning Products and Brands: Capitalizing on the Potential of Sound and Shape Symbolism.” *Journal of Consumer Psychology* 22, no. 1: 37–54.

Spence, C. 2020a. “Assessing the Role of Emotional Mediation in Explaining Crossmodal Correspondences Involving Musical Stimuli.” *Multisensory Research* 33, no. 1: 1–29.

Spence, C. 2020b. “Simple and Complex Crossmodal Correspondences Involving Audition.” *Acoustical Science and Technology* 41, no. 1: 6–12.

Spence, C. 2022. “Exploring Group Differences in the Crossmodal Correspondences.” *Multisensory Research* 35, no. 6: 495–536.

Spence, C., and N. Di Stefano. 2022. “Coloured Hearing, Colour Music, Colour Organs, and the Search for Perceptually Meaningful Correspondences Between Colour and Sound.” *i-Perception* 13, no. 3: 20416695221092802.

Spence, C., and S. Keller. 2024. “Sonic Branding: A Narrative Review at the Intersection of Art and Science.” *Psychology & Marketing* 41, no. 7: 1530–1548.

Spence, C., N. M. Puccinelli, D. Grewal, and A. L. Roggeveen. 2014. “Store Atmospherics: A Multisensory Perspective.” *Psychology & Marketing* 31, no. 7: 472–488.

Steele, A., D. Jacobs, C. Siefert, R. Rule, B. Levine, and C. D. Marci. 2013. “Leveraging Synergy and Emotion in a Multi-Platform World: A Neuroscience-Informed Model of Engagement.” *Journal of Advertising Research* 53, no. 4: 417–430.

Sun, X., X. Li, L. Ji, et al. 2018. “An Extended Research of Crossmodal Correspondence Between Color and Sound in Psychology and Cognitive Ergonomics.” *PeerJ* 6: e4443.

Sunaga, T. 2018. “How the Sound Frequency of Background Music Influences Consumers' Perceptions and Decision Making.” *Psychology & Marketing* 35, no. 4: 253–267.

Sunaga, T., Y. Meng, and X. Zhuang. 2020. “Interactions Between Instrumental Timbre and Consumers' Regulatory Focus.” *Journal of Business Research* 121: 1–12.

Sundar, A., and T. J. Noseworthy. 2016. “Too Exciting to Fail, Too Sincere to Succeed: The Effects of Brand Personality on Sensory Disconfirmation.” *Journal of Consumer Research* 43, no. 1: 44–67.

Techawachirakul, M., A. Pathak, K. Motoki, and G. Anne Calvert. 2023. “Sonic Branding of Meat-And Plant-Based Foods: The Role of Timbre.” *Journal of Business Research* 165: 114032.

Vaughn, L. A., A. R. Harkness, and E. K. Clark. 2010. "The Effect of Incidental Experiences of Regulatory Fit on Trust." *Personal Relationships* 17, no. 1: 57–69.

Velasco, C., A. T. Woods, O. Petit, A. D. Cheok, and C. Spence. 2016. "Crossmodal Correspondences Between Taste and Shape, and Their Implications for Product Packaging: A Review." *Food Quality and Preference* 52: 17–26.

Venkatesan, T., Q. J. Wang, and C. Spence. 2022. "Does the Typeface on Album Cover Influence Expectations and Perception of Music?" *Psychology of Aesthetics, Creativity, and the Arts* 16, no. 3: 487–503.

Wadhwa, M., and K. Zhang. 2015. "This Number Just Feels Right: The Impact of Roundedness of Price Numbers on Product Evaluations." *Journal of Consumer Research* 41, no. 5: 1172–1185.

Wagner, S., E. Winner, D. Cicchetti, and H. Gardner. 1981. "'Metaphorical' Mapping in Human Infants." *Child Development* 52, no. 2: 728–731.

Walker, P. 2012. "Cross-Sensory Correspondences and Cross Talk Between Dimensions of Connotative Meaning: Visual Angularity Is Hard, High-Pitched, and Bright." *Attention, Perception, & Psychophysics* 74: 1792–1809.

Walker, P., J. G. Bremner, U. Mason, et al. 2010. "Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences." *Psychological Science* 21, no. 1: 21–25.

Walker, P., and S. Smith. 1984. "Stroop Interference Based on the Synaesthetic Qualities of Auditory Pitch." *Perception* 13, no. 1: 75–81.

Walker, P., and L. Walker. 2012. "Size–Brightness Correspondence: Crosstalk and Congruity Among Dimensions of Connotative Meaning." *Attention, Perception, & Psychophysics* 74, no. 6: 1226–1240.

Walker, R. 1987. "The Effects of Culture, Environment, Age, and Musical Training on Choices of Visual Metaphors for Sound." *Perception & Psychophysics* 42, no. 5: 491–502.

Wallmark, Z. 2019. "Semantic Crosstalk in Timbre Perception." *Music & Science* 2: 205920431984661.

Wallmark, Z., and R. A. Kendall. 2018. "Describing Sound: The Cognitive Linguistics of Timbre." In *The Oxford Handbook of Timbre*, Edited by E. I. Dolan and A. Rehdin, 579–608. Oxford University Press.

Wang, B., S. Q. Liu, J. Kandampully, and M. Bujisic. 2020. "How Color Affects the Effectiveness of Taste- Versus Health-Focused Restaurant Advertising Messages." *Journal of Advertising* 49, no. 5: 557–574.

Wang, J., X. Zhang, and J. Jiang. 2022. "Healthy-Angular, Unhealthy-Circular: Effects of the Fit Between Shapes and Healthiness on Consumer Food Preferences." *Journal of Business Research* 139: 740–750.

Yuan, W., F. Guo, M. Li, and H. Song. 2024. "Effects of Sensory Cues on Consumers' Wine Taste Perceptions and Behavior: Evidence From a Wine-Tasting Experiment." *International Journal of Contemporary Hospitality Management* 36, no. 12: 4171–4191.

Zacharakis, A., K. Pastiadis, and J. D. Reiss. 2014. "An Interlanguage Study of Musical Timbre Semantic Dimensions and Their Acoustic Correlates." *Music Perception* 31, no. 4: 339–358.

Zacharakis, A., K. Pastiadis, and J. D. Reiss. 2015. "An Interlanguage Unification of Musical Timbre." *Music Perception* 32, no. 4: 394–412.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.