



Does music really work? The two-stage audiovisual cross-modal correspondence effect on consumers' shopping behavior

Shuai Yang¹ · Xinyu Chang¹ · Sixing Chen² · Shan Lin³ · William T. Ross Jr⁴

Accepted: 20 June 2021 / Published online: 9 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

In this research, we explore the impacts of cross-modal correspondence between sound frequency and color lightness on consumers' shopping behavior. Compared to previous studies that relied on a stable single-stage information environment, our study is based on a two-stage (i.e., elimination and choice stages) cognitive model to account for the dynamic cross-modal correspondence effect on shopping behavior. After conducting two laboratory experiments and one field experiment, we find that although consumers tend to pay more attention to light (vs. dark) products in the high (vs. low)-frequency sound condition in the elimination stage, this effect is less salient at the choice stage. We further find that consumer involvement acts as a moderator. Specifically, the correspondence effect is attenuated for highly involved consumers.

Keywords Cross-modal correspondence · Audiovisual · Two-stage cognitive model · Involvement · Eye-tracking experiment

✉ Xinyu Chang
xinyu.chang@mail.dhu.edu.cn

✉ Sixing Chen
49727500@qq.com

Shuai Yang
shuai.yang@dhu.edu.cn

Shan Lin
linsh@lewisu.edu; william.ross_jr@uconn.edu

William T. Ross, Jr
william.ross_jr@uconn.edu

Extended author information available on the last page of the article

1 Introduction

Consumers live in an environment where they receive signals simultaneously from multiple different sensory paths ranging from visual to invisible cues. Unlike separate systems, these sensory paths are closely intertwined through a process known as multisensory integration (Owens and Efros, 2018). Previous psychology research has shown that there are many interactions and interdependencies between the different senses, such as associations between color and sound (Klapetek et al., 2012) and associations between sound and shape (Spence, 2012). These links between senses are referred to as cross-modal correspondence.

Cross-modal correspondence has garnered an increasing level of attention in marketing research, mainly due to its effects on consumer judgment and decision-making (Krishna and Schwarz, 2014). One well-established mechanism for cross-modal correspondence is visual and auditory cross-modal correspondence, wherein consumers positively match high-frequency sounds with light-colored products (Spence, 2011). Cross-modal correspondence between sound frequency and color lightness can be leveraged for numerous retail applications. For example, playing high-frequency background music could help a clothing retailer increase the sales of its overstocked white shirts.

An overview of the literature in the domain of audiovisual cross-modal correspondence reveals that previous studies have focused predominantly on the effect of cross-modal correspondence on consumers' initial reactions, such as attention and information recall (Iordanescu et al., 2010; Klapetek et al., 2012; Marks, 1987; Tavassoli and Lee, 2003); however, few studies have examined its differentiated effect across different stages of the consumers' decision-making process. Bettman and Park's (1980) two-stage cognitive model has had considerable significance in the marketing modeling and consumer literature (Roberts and James, 1991). This model splits consumers' choice process into an "elimination stage" (i.e., choosing acceptable alternatives) and a "choice stage" (i.e., making a final choice). In general, consumers tend to unconsciously eliminate alternatives at the elimination stage and then use conscious processing to make purchase decisions at the choice stage (Ge et al., 2012; Song et al., 2018).

From this perspective, it is essential not only to understand whether the cross-modal correspondence between sound frequency and color lightness varies across the two stages of the shopping process but also to explore how cross-modal correspondence affects each stage of consumers' shopping behavior. To answer these questions, we use two laboratory experiments (including an eye-tracking experiment) and one field experiment to examine the relationships between cross-modal correspondence effects and consumers' shopping behavior. The results of our three studies support improvements to shopping environment design and help shape marketing strategies for stores.

2 Literature review

2.1 A two-stage cognitive model of shopping process

Bettman and Park's (1980) two-stage theory of choice has been an important theory in the domain of the decision-making process. Ge et al. (2012) summarize that the first

stage (i.e., the elimination stage) of the decision-making process involves eliminating alternatives that do not warrant serious consideration, and the second stage (i.e., the choice stage) involves identifying the best alternative among those considerations.

Not surprisingly, there are many meaningful distinctions between the elimination and choice stages. For example, there could be systematic differences in the way information is processed at different stages (Payne, 1976). Decision-makers tend to unconsciously remove alternatives from further consideration and conduct conscious assessments of the remaining alternatives when making their final choice (Bettman et al., 1998; Payne et al., 1988; Song et al., 2018). Notably, van Zee et al. (1992) find that the information used to screen the options does not have much impact on the evaluations, and vice versa.

2.2 Cross-modal correspondence effects at the elimination stage

Consumers are surrounded by multiple sensory inputs, such as visual and auditory stimuli, at all times, from text and images on electronic screens to physical product displays in retail channels. Marketing scholars have paid increasing attention to the importance of these sensory impacts on consumer behavior.

In recent years, a considerable amount of psychology research has examined perceptual matching between stimulus attributes in different sensory modes, i.e., cross-modal retrieval (e.g., Krishna, 2012; Shen and Sengupta, 2014; Spence, 2011, 2012). Specifically, some studies have shown that people feel the synergy between high-pitched (i.e., high-frequency) sounds and light-colored objects (e.g., Klapetek et al., 2012; Spence, 2011). Additionally, humans have latent preferences for certain cross-sensory combinations. The mapping of pitch-to-luminance is not uniquely human but rather constitutes a basic feature of the perceptual system (Ludwig et al., 2011). This consistency of sensory correspondence is purely abstract, rather than following any semantic consistency or any suggestive or specific location of the auditory stimulus.

Hagtvedt and Brasel (2016) find evidence through eye tracking that the effect on attention arises from the cross-modal correspondence between sound frequency and color lightness. Their results confirm that objects with a light color immediately draw increased attention in the presence of a high-frequency sound, whereas objects with a dark color attract more attention in the presence of low-frequency sounds. Thus, compared to those in the low-frequency sound condition, consumers in the high-frequency sound condition are more likely to fixate on lighter objects.

2.3 Cross-modal correspondence effects at the choice stage

To date, most research has focused on the audiovisual cross-modal correspondence effect at the elimination stage of the consumers' shopping process, wherein consumers are prone to process information unconsciously (see Table 1). Its relative effects at the choice stage where consumers make purchase decisions consciously remain unclear.

Screening at the elimination stage is accomplished by using a non-compensatory decision strategy that ignores some relevant problem information and reduces information-processing demands (Payne et al., 1988). Consumers at this stage avoid trade-offs among attributes of alternatives and unconsciously and intuitively make

Table 1 An overview of the audiovisual cross-modal correspondence effect

| Author | Moderator | Experiment | Dependent Variable | | Choice Stage | | | | Data in Multiple Stages | Differences between Two Stages |
|----------------------------|-----------------------|---------------|--------------------|-----------|--------------|------------|-----------|------------|-------------------------|--------------------------------|
| | | | Elimination Stage | | Attention | | Intention | | | |
| | | | Attention | Intention | Field data | Field data | Intention | Field data | | |
| Evans and Treisman (2010) | Direct/indirect task | Lab | Yes ^a | No | No | No | No | No | No | No |
| Hagtvedt and Brasel (2016) | - | Lab and field | Yes | No | No | No | Yes | Yes | Yes | No |
| Hecht and Reiner (2009) | - | Lab | Yes ^a | No | No | No | No | No | No | No |
| Iordanescu et al. (2010) | - | Lab | Yes | No | No | No | No | No | No | No |
| Klapetek et al. (2012) | - | Lab | Yes ^a | No | No | No | No | No | No | No |
| Marks (1987) | - | Lab | Yes ^a | No | No | No | No | No | No | No |
| Melara (1989) | - | Lab | Yes ^a | No | No | No | No | No | No | No |
| Shen and Sengupta (2014) | Goal | Lab and field | Yes | No | No | No | Yes | Yes | Yes | No |
| Tavassoli and Lee (2003) | Language | Lab | Yes | No | No | No | No | No | No | No |
| This paper | Involvement (study 3) | Lab and field | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Note: ^a According to previous literature (e.g., Hagtvedt and Brasel, 2016), response time and information recall are the results of selective attention. Thus, we treat both response time and information recall as variables that represent attention at the elimination stage

screening decisions (Beach and Terence, 1987). Specifically, consumers do not engage in conscious efforts on the screening task itself. Instead, they are willing to be exposed to relevant/irrelevant information and are more likely to rely on automatic attention effects to process such relevant/irrelevant information unconsciously.

In contrast, consumers at the choice stage are more particular about their goals and use more concrete terms to construe products than at the elimination stage (Lee and Ariely, 2006). Consumers at this stage tend to use a more effortful compensatory strategy (Gilbride and Allenby, 2006). A compensatory strategy during the choice stage is to determine whether a good value on one attribute of an alternative can compensate for a poor value on another attribute (Bettman et al., 1998). Specifically, consumers utilize conscious thought (Dijksterhuis, 2004; Dijksterhuis and Nordgren, 2006) and consciously evaluate the attributes of the remaining alternatives rather than only the auditory and visual stimuli to make a final purchase choice.

Note that sensory processes are the primary way consumers engage with the world, and sensory information represents the vital foundation for consumers' behavior and cognition (Krishna, 2012). Klapetek et al. (2012) proposed that people have a default response under cross-modal congruency. When people encounter auditory and visual stimuli, they will attend to stimuli that are synesthetically congruent before attending to incongruent stimuli, consistent with the increased target detection rates. As a default response, this cross-modal correspondence is more likely to occur on a basic and automatic level (Hagtvedt and Brasel, 2016). Thus, this kind of automatic attention effect seems especially prone to occur in the elimination phase of the two-stage cognitive model, wherein consumers are unaware of the specifics of the screening process, allow themselves to be exposed to relevant/irrelevant stimuli, and rely on automatic attention effects (e.g., cross-modal correspondence effect) to rapidly and simply process such relevant/irrelevant information to evaluate and screen alternatives. Formally, we hypothesize the following:

H1: The cross-modal correspondence effect between sound frequency and color lightness varies across different stages of the shopping process. Specifically, in the high (vs. low)-frequency sound condition, consumers are more likely to fixate longer (or click) on light (vs. dark) products, and such cross-modal correspondence effect is stronger at the elimination stage than at the choice stage.

2.4 Moderating effects of involvement

The cross-disciplinary concept of involvement, rooted in social psychology (Sherif and Cantnil, 1947), has long been a significant topic for researchers in marketing (Andrews et al., 1990). Involvement is considered to influence the complexity or extent of consumers' decision-making processes (Steinhart et al., 2013).

Consumer involvement is the perception of personal relevance related to product categories or shopping tasks and is regarded as a perceived cognitive state during the focused attention process (Chung et al., 2018). Low-involvement consumers are more likely to be persuaded by affective or peripheral information and engage in superficial processing (Petty et al., 1983), leading to less deliberate, more immediate, and nearly automatic purchase decisions (Shiv and Fedorikhin, 1999).

Involvement motivates consumers to be more concerned about making the right decision and processing all relevant information in greater detail (Puccinelli et al., 2009). When the level of consumer involvement increases, personal relevance will increase, and consumers will be more willing to apply cognitive resources to processing information (Petty et al., 1983). That is, highly involved consumers tend to engage in more detailed and conscious thoughts. As such, highly involved consumers are more likely to use a more effortful compensatory strategy to proceed with product information and make trade-offs among attributes of their alternatives. Therefore, involvement will make the cross-modal correspondence in H1 more conscious, eliminating the automatic cross-modal correspondence effect, especially among highly involved consumers. Formally, we hypothesize the following:

H2: Involvement will moderate the cross-modal correspondence effect between sound frequency and color lightness in both the elimination stage and the choice stage. Specifically, the effects of cross-modal correspondence on consumers' click intention (at the elimination stage) and purchase intention (at the choice stage) will be attenuated among highly involved consumers.

3 Study 1

The purpose of study 1 was to examine how sound frequency affects consumers' visual attention to color lightness during the two stages of the shopping process. This study was conducted via an eye-tracking laboratory experiment. Eye-tracking technology enables us to directly monitor viewers' visual attention to specific objects by providing more accurate measurements of visual attention than self-reports do. It also provides a particularly accurate simulation of consumers' shopping processing.

3.1 Method

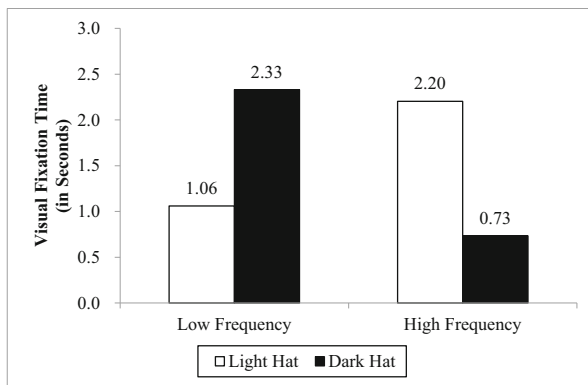
We recruited 62 students who had online shopping experiences at a university in Shanghai for study 1. Participants were randomly assigned to one of two conditions (high-frequency sound condition vs. low-frequency sound condition). Following Hagtvedt and Brasel's (2016) design, we designed high-frequency tones (approximately 1800 Hz) and low-frequency tones (approximately 120 Hz) at a predetermined volume for each experiment. Participants were first asked to indicate their initial preferred lightness from five different levels of red lightness and then were seated at the eye-tracker computer and instructed to keep their eyes on the screen. Afterward, participants were invited to an online store and shown two red hats at the same time in randomized order for 10 s; one hat was light (100% value), and the other hat was dark (60% value).

Participants were then led to the next page, which comprised a detailed textual description of hats, and were asked to consider whether to make a purchase. Both red hats had precisely the same description. This description page was displayed for 30 s. Last, participants were asked to fill in the sound frequency they perceived and background information (see details of the laboratory experiment 1's design in Appendix 1).

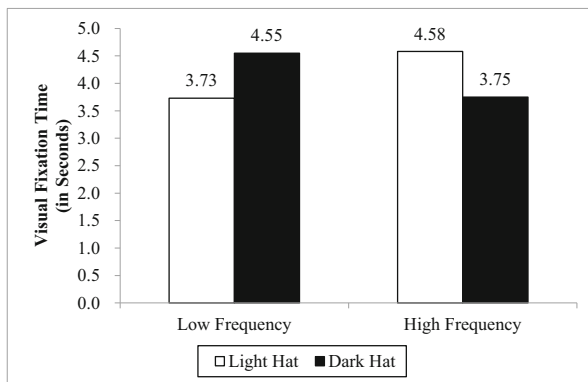
3.2 Results

Manipulation check The results showed that participants who were exposed to the high-frequency tone ($M_{\text{high}} = 4.06$) perceived a higher frequency than those who were exposed to the low-frequency tone ($M_{\text{low}} = 2.71$, $t = 4.80$, $p < 0.001$).

Hypothesis testing First, we conducted a repeated-measures ANCOVA in which sound frequency (high vs. low) was chosen as the between-subjects variable, and color lightness (light vs. dark) was chosen as the within-subject variable. Participants' initial preferred color and background information were included as covariates. The results revealed that the interaction between sound frequency and color lightness had a significant effect on participants' visual attention at the elimination stage ($F(1, 55) = 150.13$, $p < 0.001$, partial $\eta^2 = 0.73$; see Figure 1). Specifically, at the elimination stage, the light-colored hat commanded more visual attention than the dark-colored hat in the high-frequency sound condition ($M_{\text{light}} = 2.20$ s vs. $M_{\text{dark}} = 0.73$ s, $t = 10.26$, $p < 0.001$), while the dark hat commanded more visual attention than the light hat in the low-frequency sound condition ($M_{\text{light}} = 1.06$ s vs. $M_{\text{dark}} = 2.33$ s, $t = 7.65$, $p < 0.001$).



(a) Elimination (Click) Stage



(b) Choice (Purchase) Stage

Fig. 1 Influence of sound frequency on visual attention towards hats with different color lightness (study 1). **a** Elimination (click) stage. **b** Choice (purchase) stage

Second, the repeated-measures ANCOVA results showed that the sound frequency \times color lightness interaction also had a significant effect on participants' visual attention at the second stage ($F(1, 55) = 16.03, p < 0.001$, partial $\eta^2 = 0.23$; see Figure 1). Further analysis showed that at the second stage, the light hat commanded more visual attention than the dark hat in the high-frequency sound condition ($M_{\text{light}} = 4.58$ s vs. $M_{\text{dark}} = 3.75$ s, $t = 3.46, p < 0.01$), while the dark hat also commanded more visual attention than the light hat in the low-frequency sound condition ($M_{\text{light}} = 3.73$ s vs. $M_{\text{dark}} = 4.55$ s, $t = 2.44, p < 0.05$). Moreover, the sound frequency \times color lightness interaction had a stronger effect on participants' visual attention at the elimination stage than at the choice stage ($F(1, 115) = 5.53, p < 0.05$, partial $\eta^2 = 0.05$). Thus, H1 was supported.

4 Study 2

Study 2 served three purposes. First, the study aimed to confirm and generalize the cross-modal correspondence effect during the two stages of the shopping process using a different color (i.e., blue). Second, study 1 focused only on the attentional effect. Attention may affect purchase intention via several mechanisms, such as the self-perception process and a facilitation effect (Shen and Sengupta, 2014). We expect that the cross-modal correspondence effect would also influence consumers' click intention and purchase intention towards products. We thus use study 2 to provide evidence in that regard. Third, it was used to test Hypothesis 2, which postulated that the cross-modal correspondence effect during the two stages of the shopping process would be influenced by consumers' involvement.

4.1 Method

A total of 120 undergraduate students who had online shopping experience were recruited from a university in Shanghai to participate in study 2; they were compensated with snacks. Participants were randomly assigned to one of the conditions in a 2 (high-frequency sound vs. low-frequency sound) \times 2 (low involvement vs. high involvement) experimental design. The manipulation of sound was similar to study 1. Two levels of involvement (low vs. high) were designed through instructional manipulations (Puccinelli et al., 2013; Suri and Monroe, 2003).

Specifically, participants in the condition of low involvement were instructed to purchase a new T-shirt at an online store, while participants in the condition of high involvement were instructed to imagine that their university was going to launch some T-shirts for students, and they were asked to buy their new T-shirts from the online store. Similar to study 1, participants were invited to an online store where they were instructed through the pre-designed scenario to make their click and purchase decisions and then to complete an online survey (see details of the laboratory experiment 2's design in Appendix 2).

4.2 Results

Manipulation check The independent samples *t*-test results showed that the manipulations of sound frequency ($M_{\text{high}} = 4.33, M_{\text{low}} = 2.90, t = 5.57, p < 0.001$) and involvement ($M_{\text{high}} = 4.83, M_{\text{low}} = 3.15, t = 7.34, p < 0.001$) were successful.

Hypothesis testing First, the results of repeated-measures ANCOVAs again demonstrated the cross-modal correspondence effect (click: $F(1, 111) = 123.13, p < 0.001$, partial $\eta^2 = 0.53$; purchase: $F(1, 111) = 13.63, p < 0.001$, partial $\eta^2 = 0.11$). Moreover, the sound frequency \times color lightness interaction had a stronger effect on participants' click intention than on purchase intention ($F(1, 227) = 40.09, p < 0.001$, partial $\eta^2 = 0.15$). These results provided more evidence to support H1.

More importantly, the three-way interaction between sound frequency, color lightness, and involvement had a significant effect on click intention ($F(1, 111) = 53.25, p < 0.001$, partial $\eta^2 = 0.32$; see Figure 2), thus supporting H2. Specifically, the effect of the sound frequency \times color lightness interaction on click intention was weaker in the high involvement condition (for high-frequency sound condition: $M_{\text{light}} = 4.21$ vs. $M_{\text{dark}} = 3.57, t = 2.76, p < 0.05$; for low-frequency sound condition: $M_{\text{light}} = 4.19$ vs. $M_{\text{dark}} = 4.72, t = 2.50, p < 0.05$) than in the low involvement condition (for high-frequency sound condition: $M_{\text{light}} = 5.51$ vs. $M_{\text{dark}} = 2.40, t = 9.54, p < 0.001$; for low-frequency sound condition: $M_{\text{light}} = 3.01$ vs. $M_{\text{dark}} = 5.73, t = 6.65, p < 0.001$). These results are presented in Figure 2.

Furthermore, the three-way interaction between sound frequency, color lightness, and involvement had a significant effect on purchase intention ($F(1, 111) = 4.19, p < 0.05$, partial $\eta^2 = 0.04$; see Figure 2), supporting H2. The effect of the sound frequency \times color lightness interaction on purchase intention was attenuated in the high involvement condition (for high-frequency sound condition: $M_{\text{light}} = 3.40$ vs. $M_{\text{dark}} = 3.19, t = 1.13, p > 0.10$; for low-frequency sound condition: $M_{\text{light}} = 3.99$ vs. $M_{\text{dark}} = 4.22, t = 1.05, p > 0.10$) compared with the low involvement condition (for high-frequency sound condition: $M_{\text{light}} = 4.20$ vs. $M_{\text{dark}} = 3.54, t =$

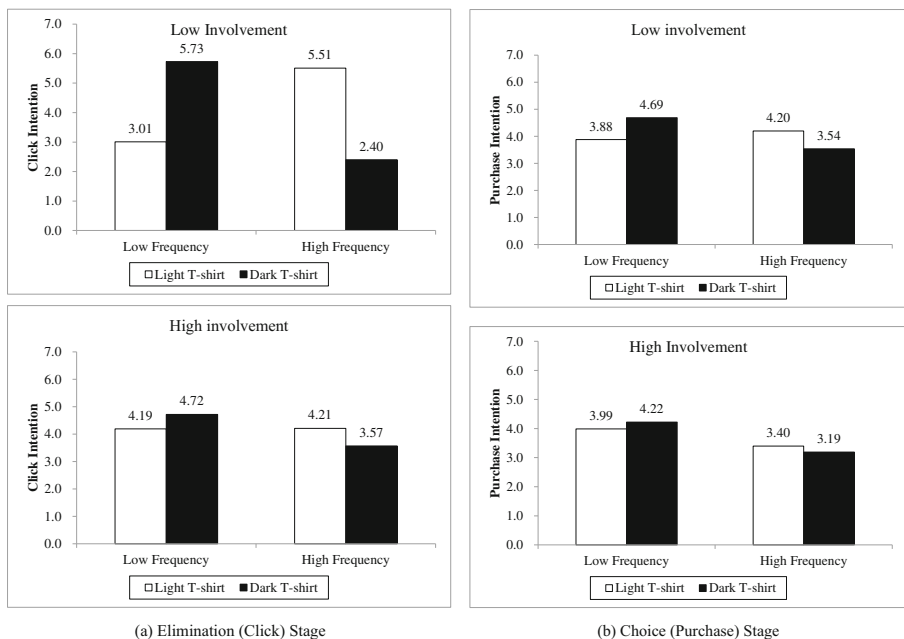


Fig. 2 Interaction of sound frequency, color lightness, and involvement (study 2). **a** Elimination (click) stage. **b** Choice (purchase) stage

2.50, $p < 0.05$; for low-frequency sound condition: $M_{\text{light}} = 3.88$ vs. $M_{\text{dark}} = 4.69$, $t = 2.45$, $p < 0.05$). Notably, in the high involvement condition, there was no significant cross-modal correspondence effect on purchase intention.

5 Study 3

5.1 Experimental design

Study 3 examines how sound frequency influences consumers' response to color lightness in the two stages of the shopping process in a real market context. We conducted this field study with the help of a start-up insole firm, which sells its products through two similar online platforms. The field study adopted a single factorial design (no music vs. high-frequency music vs. low-frequency music) to assign the use of music to one of the platforms (i.e., the treatment platform) in different weeks. That is, the designs of the treatment and control platforms are mostly the same, except for the music manipulation (see Appendix 5). Figure 3 shows that the music designs of the two platforms, demonstrating that the differences between the music designs of the two platforms existed only in the post-treatment period. The control platform was assigned “no music” during the post-treatment period, while the treatment platform was assigned “high-frequency music” or “low-frequency music” on a weekly basis. Following previous literature (e.g., Kumar and Tan, 2015; Yang and Xiong, 2019), we further conducted the randomization checks of the products on the two platforms to ensure that there were no significant differences in product prices, pre-treatment clicks, and pre-treatment conversions between the products at the two platforms ($ps > 0.1$) (see Appendix 6).

The start-up insole company tracked daily performance for each product on the two platforms and provided us with these data for our research. The data from both platforms are at the product-day level, with the same data structure. During the 6-week field experiment, there were 9546 visits to the two platforms (including 3994 visits in the pre-treatment period and 5552 visits in the post-treatment period), which resulted in 4223 clicks (1727 clicks in the pre-treatment period and 2496 clicks in the post-treatment period) and 392 conversions (136 conversions in the pre-treatment

| Platform | Pre-treatment Period | Post-treatment Period | | |
|--------------------------------|-----------------------|----------------------------------|--------------------------------------|----------------------------------|
| Control Platform (Tmall) | Weeks 1-2 No music | Weeks 3-6 No music | | |
| Treatment Platform (Taobao) | Weeks 1-2 No music | Week 3 Low-frequency music | Weeks 4-5 High-frequency music | Week 6 Low-frequency music |

Fig. 3 Illustration of the timeline and design of study 3. *Note:* The differences between the music designs of the two platforms existed only in the post-treatment period. Specifically, both platforms were assigned “no music” during the pre-treatment period (i.e., weeks 1–2). During the post-treatment period, the treatment platform was assigned “high-frequency music” (i.e., weeks 4–5) or “low-frequency music” (i.e., weeks 3 and 6) on a weekly basis, while the control platform was assigned “no music” yet (i.e., weeks 3–6)

period and 256 conversions in the post-treatment period). We utilize the click-through rate (a ratio of clicks to visits) and conversion rate (a ratio of conversions to clicks) for each product to represent consumers' behavior at the elimination and choice stages of their shopping process, respectively. Specifically, we use a difference-in-difference (DID) method to examine the cross-modal correspondence effect on a product level as follows in Eq. (1):

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Treat}_i \times \text{Post}_t + \beta_3 \text{Treat}_i \times \text{Post}_t \times \text{Correspond}_{it} + \alpha X_{it} + \mu_i + \varepsilon_{it}, \quad (1)$$

where Y_{it} represents the performance of product i at (i.e., click-through rate and conversion rate) period t . Post_t is a dummy variable indicating the pre-treatment (0) or post-treatment period (1); Treat_i indicates whether product i is in the control (0) or treatment group (1). Correspond_{it} is a dummy variable indicating whether the cross-modal correspondence between sound frequency and color lightness occurs for product i at period t ($\text{Correspond}_{it} = 0$ if a light-color (dark-color) product i is exposed to low-frequency (high-frequency) music at period t , and $\text{Correspond}_{it} = 1$ if not). We also include the effects of control variables X_{it} , such as time trend and day-of-week effects. μ_i represents the fixed effects of products, and ε_{it} is the error term. β_3 is the parameter of interest that captures the cross-modal correspondence effect on product performance. Because the product-specific fixed effects will be collinear with Treat_i , we do not include the main effect of Treat_i in Eq. (1). In addition, we only consider the three-way interactions $\text{Treat}_i \times \text{Post}_t \times \text{Correspond}_{it}$ because only observations from the treatment group in the post-treatment period vary in the levels. Specifically, Correspond_{it} matters only when $\text{Treat}_i = 1$ and $\text{Post}_t = 1$.

5.2 Estimation results

The estimation results for the click-through rate as the dependent variable are presented in Table 2, Column 1. We find strong evidence that cross-modal correspondence has a positive and significant effect on the click-through rate ($\beta_3 = 0.023$, $p < 0.01$ in Column 1). This suggests that consumers are more likely to click on light (vs. dark) products with high (vs. low)-frequency sounds in the elimination stage of their shopping process. The coefficient indicates that the cross-modal correspondence increased the products' click-through rate by 121.05% relative to the median value of 0.019.¹

Column (2) shows that cross-modal correspondence has a positive and marginally significant effect on the conversion rate ($\beta_3 = 0.032$, $p < 0.1$ in Column 2). The coefficient indicates that the cross-modal correspondence increased the products' conversion rate by 88.89% relative to the median of 0.036. Moreover, the coefficient β_3 for Column (1) is more significant than that for Column (2). This implies that although the cross-modal correspondence between sound frequency and color lightness also occurs in consumers' final choice stage, such a cross-modal correspondence effect is stronger at the elimination stage than at the final choice stage.

¹ Because of the log-normal distribution of the click-through rate and conversion rate, we followed Chesnes et al. (2017) and presented the median treatment effects.

Table 2 Model estimation results of study 3

| DV → | (1) Click-through rate Coefficient (Std. error) | (2) Conversion rate Coefficient (Std. error) |
|---------------------------|--|---|
| Treat × Post × Correspond | 0.023** (0.007) | 0.032 ^a (0.019) |
| Treat × Post | −0.029* (0.012) | −0.040* (0.017) |
| Post | 0.008 (0.007) | 0.044 ^a (0.022) |
| Time trend | 0.0001 (0.0002) | −0.001 (0.001) |
| Monday | 0.003 (0.004) | 0.017 (0.010) |
| Tuesday | 0.003 (0.006) | 0.035* (0.014) |
| Wednesday | 0.001 (0.003) | 0.031* (0.014) |
| Thursday | 0.0005 (0.004) | 0.005 (0.011) |
| Friday | 0.001 (0.004) | 0.030* (0.012) |
| Saturday | 0.0005 (0.003) | 0.021 (0.013) |
| Constant | 0.069*** (0.005) | 0.026* (0.010) |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ^a $p < 0.1$. Error terms are clustered at the product level. The coefficient of Treat × Post × Correspond captures the cross-modal correspondence effect on product performance

6 General discussion

Based on the actual behavior of participants found in the three experimental studies—one field experiment and two laboratory experiments—we reach three broad conclusions. First, sound frequency has significant effects on consumers' visual attention related to color lightness. Thus, control over sound frequency conditions allows cross-modal correspondence to guide consumers' visual attention. Second, this power is more salient at the elimination stage than at the choice stage. Finally, the cross-modal correspondence effect during the two stages of the shopping process is affected by consumers' involvement. When the instructions for participants to complete the experimental steps improved participants' level of involvement, the cross-modal correspondence effect was not found to be significant on purchase intention.

There are some alternative explanations for the differential cross-modal correspondence effect on the two stages. One possible explanation is that consumers spend much more time at the choice stage than at the elimination stage. Given that attention can be used to support higher-order information processing and thus might affect consumers' decision-making (Janiszewski et al., 2013), more time (i.e., more attention) spent at the choice stage may influence the cross-modal correspondence effect at this stage. Another possible explanation is that consumers can read more information at the choice stage than at the elimination stage, which, rather than product color, might distract consumers' attention. Ge et al. (2012) propose a weight shift mechanism: when consumers evaluate their alternatives at the choice stage, the newly introduced information at this stage about alternatives on one dimension will increase the weight that consumers attach to that dimension in their evaluation process. Thus, consumers may place more decision-making weight on the newly introduced product information

displayed at the choice stage, which may weaken the cross-modal correspondence effect in the choice stage.

Our research has several theoretical implications. First, studies have shown that cross-modal mapping occurs prior to conscious awareness of the visual stimuli (Hung et al., 2017), suggesting that audiovisual cross-modal correspondence occurs at an automatic and unconscious level. Based on this theory, we propose a new framework for how cross-modal correspondence effects impact each stage of consumers' shopping behavior. Second, we contribute to the relevant literature by identifying an essential and previously un-investigated moderating factor—consumer involvement.

Our findings also provide several important practical insights. First, our findings suggest that companies should be aware of the cross-modal correspondence effect on consumers (especially on low-involvement consumers) when developing marketing strategies. Second, marketers may want to change or enhance their advertisement targeting strategies. They should give more consideration to the interaction between hearing and vision in advertising and more effectively highlight their products or services. It is wise to combine light (vs. dark) products with high (vs. low) frequency sounds in their brand's promotion videos. For example, when marketers choose voice actors to promote their products, they may find it helpful to consider the speaker's tone. In general, women's voices are higher than men's voices, so advertisements characterized by female voices are likely to promote products with light colors more effectively.

Our studies have some limitations, and future research could be extended in several directions. First, while the randomization checks ensured that there were no significant differences between the products on the two platforms, we still observe some “imperfections” of the field experiment. For example, the website designs at the choice stage of the field experiment varied slightly between the two platforms (e.g., the color palette in the right-side table boxes). Though there is no reason to suggest that such imperfections affect our results, this provides future research opportunities to validate our findings with additional field experiments. Second, our experiments used online stores as shopping contexts. Thus, the sample of experiments is limited to online shoppers. Future research should test the generalizability of these results by repeating our research using offline shopper samples. Lastly, in addition to consumer involvement, which adjusts the influence of audiovisual cross-modal correspondence, other important but undiscovered factors may also effectively moderate the effects of cross-modal correspondence. They should therefore also be considered, such as gender, age, and situational factors in consumers' purchase decisions.

Appendix 1. Laboratory experiment design for study 1

Procedures:

- 1) Briefly explain that the objective of the laboratory experiment is to understand consumer online shopping behavior and inform participants that they will be compensated with gift certificates valued at 20 CNY for a convenience store in the university.

- 2) Ask participants whether they have an initial preferred shade of red. If yes, ask them to indicate their preferred shade from the following five different levels of red lightness.



- 3) Participants are then seated at the eye-tracker computer. The guide mentions the following to the participants: “Imagine you are going to buy a new hat and find some hats at an online store. Please keep your eyes on the screen.”²
- 4) Lead the participants to go to the online store website. Tell the participants that the store website will display two available hats to them quickly. Use the following screen that was captured from the website as an example.³

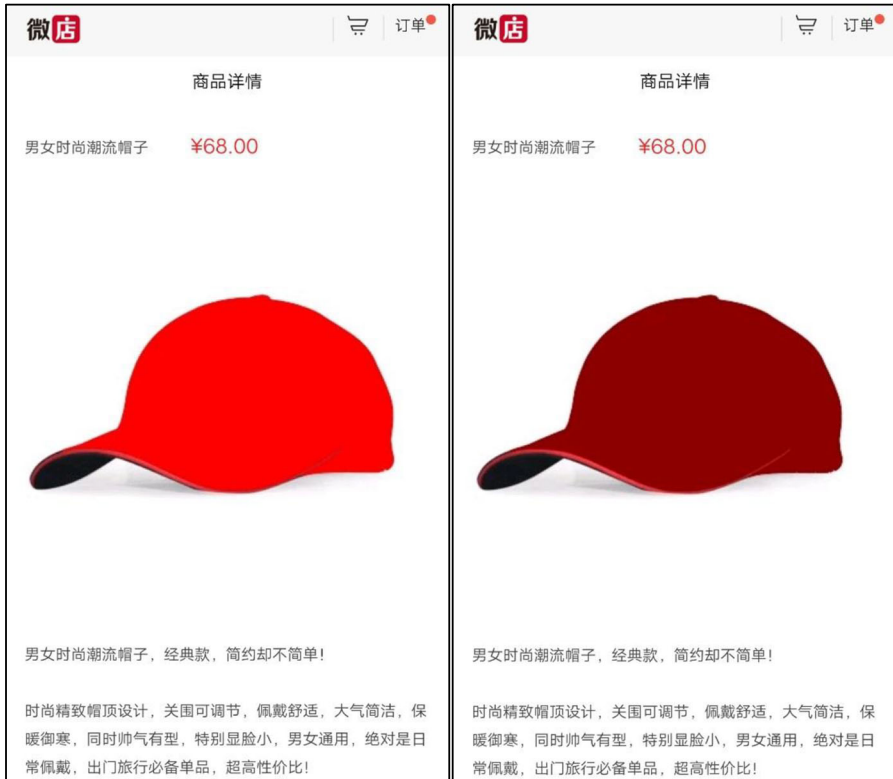


- 5) (Elimination stage) The website will show two red hats for 10 s where one is light (100% value) and the other is dark (60% value). The order of the two red hats is randomized. Each hat was priced at approximately 68 CNY (approximately \$9.81) and introduced simply without any brand information. Participants will be exposed to either a high-frequency tone (approximately 1800 Hz) or a low-frequency (approximately 120 Hz) tone of “Turkey in the Straw” at a predetermined volume displayed by the research assistant on his mobile phone.
- 6) (Choice stage) Show the participants the following screen as an example. Tell the participants that they are going to consider whether to make a purchase and remind

² The eye tracker could monitor participants’ gaze throughout the experiment and capture eye fixations according to a specified criterion. If participants’ eyes moved too quickly (i.e., less than 0.175 seconds in one area), their attention would not be recorded. In addition, participants were free to gaze anywhere on the page in the experiment, or even not look at the screen.

³ There was no “end” button on the screen, but participants could control their own eyes and would move their own eyes when they felt uninterested in the images in studies 1 and 2.

them to keep their eyes on the screen. The screen will be displayed for 30 s (there was no “end” button on the screen). Participants will still be exposed to either the high-frequency or the low-frequency tone.



- 7) Instruct them to fill in their perceived sound frequency and background information:
 - Perceived sound frequency, gender, age, monthly living expenses, online shopping frequency
- 8) End of the session. Thank the subjects for their participation.

Appendix 2. Laboratory experiment design for study 2

Procedures:

- 1) Briefly explain that the objective of the laboratory experiment is to understand consumer online shopping behavior.

- 2) Ask participants whether they have an initial preferred shade of blue. If yes, ask them to indicate their preferred shade of blue from the following five different shades of blue.



- 3) Participants in the condition of low involvement were instructed that “Imagine you are going to buy a new T-shirt and find some T-shirts at an online store.” Participants in the condition of high involvement were instructed to “Imagine that our Business and Management School is going to launch some T-shirts to the students at our school. The school has posted the T-shirts to its online store. You are going to buy a new T-shirt from the online store.”
- 4) Lead the participants to go to the online store website. Tell the participants that the store website will display two available T-shirts to them quickly. Use the following screen that was captured from the website as an example.



- 5) (Elimination stage) The website will show two blue T-shirts for 10 s where one is light (100% value) and the other is dark (60% value). The order of the two blue T-shirts is randomized. Participants will be exposed to either a high-frequency tone (approximately 1800 Hz) or a low-frequency (approximately 120 Hz) tone of “Turkey in the Straw” at a predetermined volume displayed by the research assistant on his mobile phone.
- 6) Participants were asked whether they would like to click on the light/dark blue T-shirt to obtain more information and instruct them to fill in their click intention on each T-shirt (1 = extremely unlikely, and 7 = extremely likely).

- 7) (Choice stage) Show the participants the following screen as an example. Tell the participants that they are to consider whether to make a purchase. The screen will be displayed for 30 s (there was no “end” button on the screen). Participants will still be exposed to either the high-frequency or the low-frequency tone.



- 8) Instruct the subjects to fill in their purchase intention on each T-shirt (1 = extremely unlikely, and 7 = extremely likely).
- 9) Instruct them to fill in their perceived sound frequency and background information, including perceived sound frequency, involvement, gender, age, monthly living expenses, and online shopping frequency.
- 10) End of the session. Thank the subjects for their participation.

Appendix 3. Results of models with correlated error terms in laboratory studies

To estimate the cross-modal effects on participants' visual attention in study 1, we specify the model below:

$$Y_{irt} = \alpha_0 + \alpha_1 \text{Correspondence}_{ir} + \alpha_2 \text{Preference}_i + \alpha_3 \text{Gender}_i + \alpha_4 \text{Age}_i + \alpha_5 \text{Expense}_i + \alpha_6 \text{Frequency}_i + \varepsilon_{irt}, \quad (2)$$

where Y_{irt} represents participant i 's visual attention ratio (i.e., the ratio of the fixation time to total exposure time) on hat r (0 for the light red hat and 1 for the dark red hat) on stage t (0 for the elimination stage and 1 for the choice stage). The $\text{Correspondence}_{ir}$ represents the correspondence of background music and product color that was displayed for consumer i (1 for correspondence (i.e., high-frequency music with the light red hat or the low-frequency music with the dark red hat) and 0 for non-correspondence (i.e., the other combinations of color and sound)). In the model, we also include controls for participant i 's initial color preference (Preference_i), gender (Gender_i), age (Age_i), living expenses (Expense_i) and online shopping frequency (Frequency_i).

Given that a participant's attention levels at the two stages are not independent, we follow previous literature (Agarwal et al., 2011) and correlate the error terms of the models for participant i 's attention ratio at the two stages as follows:

$$\text{where } \Sigma_\varepsilon = \begin{bmatrix} \sigma_{ARES}^2 & \sigma_{ARES,ARCS} \\ \sigma_{ARCS,ARES} & \sigma_{ARCS}^2 \end{bmatrix}, \quad (3)$$

$$\varepsilon_{irt}^{ARES}, \varepsilon_{irt}^{ARCS} \sim N(0, \Sigma_\varepsilon)$$

where ε_{irt}^{ARES} represents the error term of participant i 's attention ratio at the elimination stage, and ε_{irt}^{ARCS} represents the error term of participant i 's attention ratio at the choice stage.

Table 3 shows the results of the models with correlated error terms. The results indicate that the audiovisual cross-modal correspondence effects for the attention ratio are significant at both the elimination stage (0.136, $p < 0.001$) and the choice stage (0.027, $p < 0.001$). Moreover, the coefficient of correspondence for the attention ratio was significantly greater at the elimination stage than that at the choice stage ($\chi^2(1) = 94.92$, $p < 0.001$). These findings imply that compared to the elimination stage, the cross-modal correspondence effect on consumers' visual attention is attenuated at the choice stage.

Similarly, we estimate the cross-modal effects on participants' visual attention and the moderating effect of involvement in study 2 with correlated error terms as follows:

$$Y_{irt} = \alpha_0 + \alpha_1 \text{Correspondence}_{ir} + \alpha_2 \text{Involvement}_i + \alpha_3 \text{Correspondence}_{ir} * \text{Involvement}_i + \alpha_4 \text{Preference}_i + \alpha_5 \text{Gender}_i + \alpha_6 \text{Age}_i + \alpha_7 \text{Expense}_i + \alpha_8 \text{Frequency}_i + \varepsilon_{irt}, \quad (4)$$

where Y_{irt} represents participant i 's intention on T-shirt r (0 for the light blue T-shirt and 1 for the dark blue T-shirt) on stage t (click intention at the elimination stage and purchase intention at the choice stage). The $\text{Correspondence}_{ir}$ represents

Table 3 Results of models with correlated error terms in study 1

| DV → | (1) Elimination (click) stage Attention Ratio | (2) Choice (Purchase) stage Attention ratio |
|-----------------------|--|--|
| | Coefficient (Std. error) | Coefficient (Std. error) |
| Correspondence | 0.136*** (0.009) | 0.027*** (0.008) |
| Preference | −0.006 (0.005) | −0.004 (0.004) |
| Gender | −0.000 (0.009) | −0.005 (0.008) |
| Age | 0.000 (0.005) | −0.001 (0.004) |
| Expense | 0.005 (0.006) | −0.007 (0.006) |
| Frequency | 0.005 (0.006) | 0.006 (0.005) |
| Constant | 0.078** (0.024) | 0.144*** (0.018) |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. DV, dependent variable. The main variables of interest and their coefficient estimates are highlighted in boldface. The coefficient of Correspondence for column (1) is significantly greater than that in column (2) ($\chi^2(1) = 94.92, p < 0.001$)

the correspondence of background music and product color that was displayed for consumer i (1 for correspondence (i.e., high-frequency music with the light blue T-shirt or the low-frequency music with the dark blue T-shirt), and 0 for noncorrespondence (i.e., the other combinations of color and sound)). Involvement $_i$ represents consumer i 's perception of personal relevance related to product categories or shopping tasks (1 for high involvement and 0 for low involvement). In the model, we also include controls for participant i 's initial color preference (Preference $_i$), gender (Gender $_i$), age (Age $_i$), living expenses (Expense $_i$) and online shopping frequency (Frequency $_i$).

Given that a participant's intentions at the two stages are not independent, we correlate the error terms of the models for participant i 's intention at the two stages as follows:

$$\text{where } \Sigma_{\varepsilon} = \begin{bmatrix} \sigma_{IES}^2 & \sigma_{IES,ICS} \\ \sigma_{ICS,IES} & \sigma_{ICS}^2 \end{bmatrix}, \tag{5}$$

$$\varepsilon_{it}^{IES}, \varepsilon_{it}^{ICS} \sim N(0, \Sigma_{\varepsilon})$$

where ε_{it}^{IES} represents the error term of participant i 's click intention at the elimination stage, and ε_{it}^{ICS} represents the error term of participant i 's purchase intention at the choice stage.

Table 4 shows the results of the models with correlated error terms. The results indicate that the audiovisual cross-modal correspondence effect for intention is significant at the elimination stage (2.917, $p < 0.001$) and at the choice stage (0.733, $p < 0.01$). In addition, the coefficient of correspondence for intention at the elimination stage was significantly greater than that at the choice stage ($\chi^2(1) = 49.880, p < 0.001$). Moreover, the results also indicate that the moderating effects of involvement are negative at both the elimination stage (−2.328, $p < 0.001$) and at the choice stage (−0.511, $p < 0.1$), which provides more evidence for the robustness of our results.

Table 4 Results of models with correlated error terms in study 2

| DV → | (1) Click stage Attention ratio | (2) Purchase stage Attention ratio |
|-------------------------------------|------------------------------------|---------------------------------------|
| | Coefficient (Std. error) | Coefficient (Std. error) |
| Correspondence | 2.917*** (0.231) | 0.733** (0.231) |
| Involvement | 1.120*** (0.193) | -0.202 (0.187) |
| Involvement × Correspondence | -2.328*** (0.291) | -0.511 (0.288) |
| Preference | -0.013 (0.061) | -0.129* (0.059) |
| Gender | -0.098 (0.152) | -0.071 (0.162) |
| Age | -0.149 (0.087) | -0.205* (0.096) |
| Expense | 0.148 (0.125) | 0.083 (0.111) |
| Frequency | -0.054 (0.126) | 0.059 (0.117) |
| Constant | 2.902*** (0.426) | 4.438*** (0.393) |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. DV, dependent variable. The main variables of interest and their coefficient estimates are highlighted in boldface. The coefficient of Correspondence for column (1) is significantly greater than that in column (2) ($\chi^2(1) = 49.800, p < 0.001$)

Appendix 4. Measures of main variables

All measures were reported on a 1–7 scale (1 = extremely disagree, 7 = extremely agree).

| Construct | Item | Source |
|--|---|---|
| Visual attention ⁴ | Visual attention was measured by the total fixation duration in s within the area of interest (AIO). | Hagtvedt and Brasel (2016) |
| Click intention ($\alpha = 0.89$) | 1. I feel like clicking the item now. 2. I would like to click on the item as soon as possible. 3. I would like to click on the item right away. | Taylor and Todd (1995); Yang et al. (2017) |
| Purchase intention ($\alpha = 0.88$) | 1. I feel like buying this item now. 2. I would like to buy the item as soon as possible. 3. I would like to buy the item right away. | Taylor and Todd (1995); Song et al. (2018) |
| Consumer involvement ($\alpha = 0.87$) | 1. I perceive this item as very important. 2. I perceive this item as very significant. 3. I perceive this item as very valuable. 4. This item matters a lot to me. 5. This item means a lot to me. | Zaichkowsky (1985) |

Note: α , Cronbach's alpha

⁴ Following Hagtvedt and Brasel (2016), visual attention was measured by the total fixation duration in seconds within the area of interest (AIO). We used 0.175 s as the fixation floor. Thus, a fixation was measured when participants' eyes stayed at AIO corresponding to a hat for at least 0.175 s. The total fixation time was used for measuring participants' visual attention towards hats at the first stage ($M = 1.56, SD = 0.83$) and at the second stage ($M = 4.18, SD = 1.52$).

Appendix 5. Experimental design of field study 3



(a) The control platform



(b) The treatment platform

Fig. 4 Platforms used to conduct the field experiment. a The control platform. b The treatment platform. *Note:* Music was played only for PC customers. We thus collected data only from PC customers. In addition, there is no music stop/pause button on the platforms during the whole experimental period

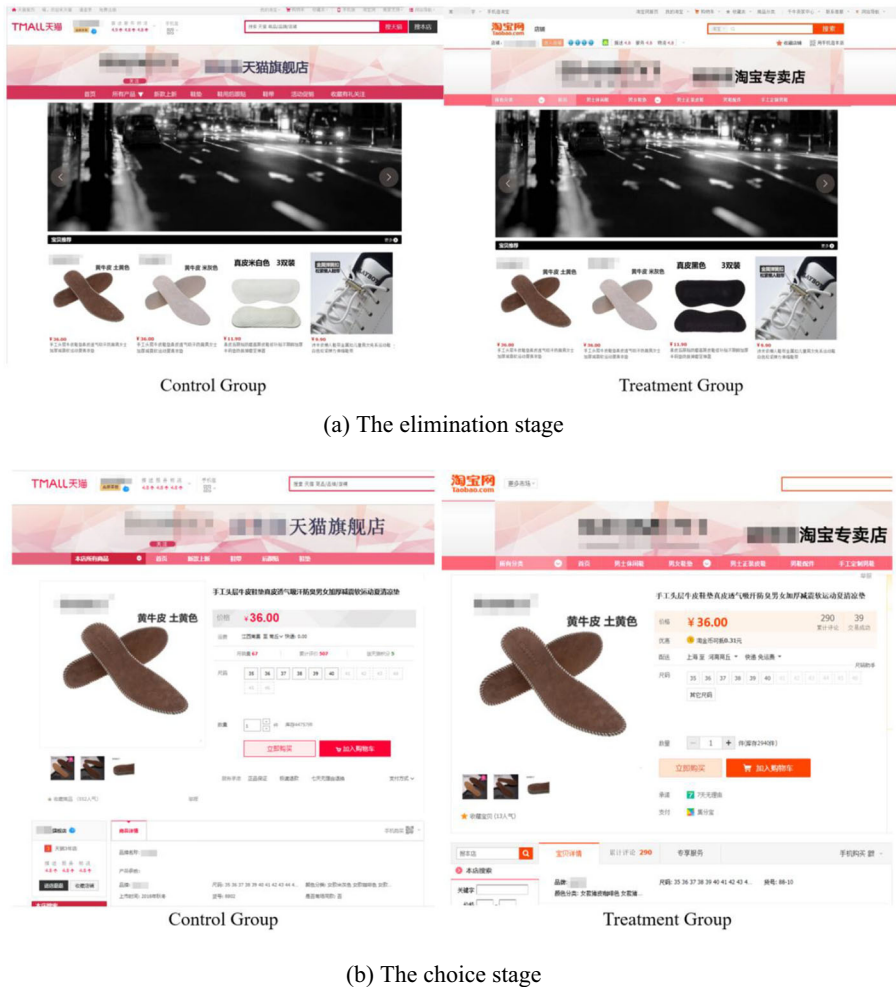


Fig. 5 Pages used to display products in elimination and choice stages. a The elimination stage. b The choice stage. *Note:* All products (with a product picture, the price, and a description) at the platform would be exposed to consumers when they visited the platform (the elimination stage). If a consumer was interested in a product, she/he would click on the product and be led to the landing page of the product's page to learn about the details (e.g., size information, materials) of the product (the choice stage). Both platforms display the product pictures on the left and details of the product on the right and bottom of the page. If the consumers took an action to purchase, a conversion was recorded

Appendix 6. Differences in prices, pre-treatment clicks, and pre-treatment conversions of treated and control products

There were 16 and 13 products at the control and treatment platforms, respectively. There are no significant differences in product prices, pre-treatment clicks, and pre-treatment conversions between the products at the two platforms ($p_s > 0.1$).

| Analysis | Product type | Mean | St. Dev. | <i>p</i> -value (<i>t</i> -value) |
|-------------|--------------|--------|----------|------------------------------------|
| Price | Treated | 9.888 | 3.665 | 0.221 (1.278) |
| | Control | 13.438 | 9.460 | |
| Clicks | Treated | 55.500 | 74.437 | 0.785 (0.276) |
| | Control | 64.538 | 101.805 | |
| Conversions | Treated | 6.375 | 9.946 | 0.300 (1.058) |
| | Control | 3.077 | 5.766 | |

Note: Numbers in the above table present the average performance of each product on platforms during the pre-treatment period. For example, on average, products at the treatment platform received 55.500 clicks and 6.375 conversions in the pre-treatment period, while products at the control platform received 64.538 clicks and 3.077 conversions in this period

Appendix 7. Robustness checks in field study 3

We note that both the click-through rate (mean: 0.070; Std.: 0.115) and conversion rate (mean: 0.049; Std.: 0.135) have a relatively high SD-vs-value ratio. This could be because the utility of the individual choice in clicking on a product or converting it to a sale follows an independent and identically distributed (*i.i.d.*) extreme value distribution rather than a normal distribution (Agarwal et al., 2011). To address this issue, we

Table 5 Robustness checks of study 3

| DV → | (1) Logit | | (2) Logarithm | | (3) Standardized | |
|----------------------------------|-----------------------------|-----------------------|-----------------------------|----------------------------------|----------------------------|----------------------------------|
| | CTR | CR | CTR | CR | CTR | CR |
| Treat × Post × Correspond | 0.442*** (0.074) | 1.976* (0.573) | 0.767*** (0.182) | 0.333^a (0.194) | 0.202** (0.058) | 0.239^a (0.139) |
| Treat × Post | 2.469*** (0.406) | 0.814 (0.184) | −0.218 (0.183) | −0.785** (0.274) | −0.257* (0.106) | −0.294* (0.129) |
| Post | 0.670 ^a (0.143) | 1.803 (0.658) | −0.434* (0.187) | 0.529 ^a (0.290) | 0.070 (0.059) | 0.326 ^a (0.160) |
| Time trend | 1.002 (0.007) | 0.984 (0.013) | 0.0001 (0.006) | −0.005 (0.007) | 0.001 (0.001) | −0.005 (0.004) |
| Monday | 1.046 (0.234) | 1.593 (0.465) | −0.268 ^a (0.158) | 0.209 (0.170) | 0.025 (0.038) | 0.126 (0.074) |
| Tuesday | 1.044 (0.228) | 2.273** (0.691) | 0.007 (0.158) | 0.258* (0.184) | 0.024 (0.050) | 0.256* (0.106) |
| Wednesday | 1.013 (0.219) | 2.131* (0.653) | −0.104 (0.158) | 0.315* (0.233) | 0.007 (0.027) | 0.228* (0.103) |
| Thursday | 1.008 (0.226) | 1.145 (0.365) | −0.378* (0.160) | −0.152 (0.172) | 0.004 (0.031) | 0.037 (0.080) |
| Friday | 1.013 (0.219) | 2.075* (0.592) | −0.063 (0.159) | 0.356* (0.177) | 0.007 (0.038) | 0.221* (0.088) |
| Saturday | 1.008 (0.216) | 1.747* (0.481) | 0.007 (0.158) | 0.203 (0.227) | 0.004 (0.023) | 0.155 (0.097) |
| Constant | 0.005*** (0.001) | 0.028*** (0.007) | −4.246*** (0.394) | −6.556*** (0.299) | −0.003 (0.042) | −0.164* (0.076) |

Notes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ^a $p < 0.1$. CTR, click-through rate; CR, conversion rate. Column (1) uses a logit model, Column (2) models the logarithm of the dependent variables (in the cases where dependent variables were zero, we added a small value of 0.0005 to enable the logarithmic transformation), and Column (3) uses the standardized values of dependent variables. Error terms are clustered at the product level. The coefficient of Treat × Post × Correspond captures the cross-modal correspondence effect on product performance

checked the robustness of the cross-modal correspondence effects in some alternative specifications (see Table 3). Column (1) followed Agarwal et al.'s (2011) approach and used a logit model to check the robustness of our results. In Column (2), we modeled the logarithm of the dependent variables, and in Column (3), we used the standardized values of dependent variables in the model. The results of Columns (1)–(3) were similar to those in Table 2; thus, our results are robust.

Acknowledgements We acknowledge the financial supports from the National Natural Science Foundation of China (no. 71972035, 71702052, 71602026), Hunan Science Foundation (2018JJ3086), the Fundamental Research Funds for the Central Universities, and DHU Distinguished Young Professor Program. We are most grateful to Tao Li for his helpful discussions in conceptualization and research design.

Funding The authors were supported by the National Natural Science Foundation of China (no. 71972035, 71702052, 71602026), Hunan Science Foundation (2018JJ3086), the Fundamental Research Funds for the Central Universities, and DHU Distinguished Young Professor Program.

References

- Agarwal, A., Hosanagar, K., & Smith, M. D. (2011). Location, location, location: an analysis of profitability of position in online advertising markets. *Journal of Marketing Research*, 48(6), 1057–1073.
- Andrews, J. C., Durvasula, S., & Akhter, S. H. (1990). A framework for conceptualizing and measuring the involvement construct in advertising research. *Journal of Advertising*, 19, 27–40.
- Beach, L. R., & Terence, R. M. (1987). Image theory: Principles, goals, and plans in decision making. *Acta Psychologica*, 66(12), 201–220.
- Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, 7(3), 234–248.
- Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25(3), 187–217.
- Chesnes, M., Dai, W., & Jin, G. Z. (2017). Banning foreign pharmacies from sponsored search: The online consumer response. *Marketing Science*, 36(6), 879–907.
- Chung, S., Kramer, T., & Wong, E. M. (2018). Do touch interface users feel more engaged? The impact of input device type on online shoppers' engagement, affect, and purchase decisions. *Psychology & Marketing*, 35(11), 795–806.
- Dijksterhuis, A. (2004). Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology*, 87(5), 586–598.
- Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 1(2), 95–109.
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10(1), 1–12.
- Ge, X., Häubl, G., & Elrod, T. (2012). What to say when: Influencing consumer choice by delaying the presentation of favorable information. *Journal of Consumer Research*, 38(6), 1004–1021.
- Gilbride, T. J., & Allenby, G. M. (2006). Estimating heterogeneous EBA and economic screening rule choice models. *Marketing Science*, 25(5), 494–509.
- Hagtvedt, H., & Brasel, S. A. (2016). Cross-modal communication: Sound frequency influences consumer responses to color lightness. *Journal of Marketing Research*, 53(4), 551–562.
- Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research*, 193(2), 307–314.
- Hung, S. M., Styles, S. J., & Hsieh, P. J. (2017). Can a word sound like a shape before you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological Science*, 28(3), 263–275.
- Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception, & Psychophysics*, 72(7), 1736–1741.
- Janiszewski, C., Kuo, A., & Tavassoli, N. T. (2013). The influence of selective attention and inattention to products on subsequent choice. *Journal of Consumer Research*, 39(6), 1258–1274.

- Klapetek, A., Ngo, M. K., & Charles, S. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics*, *74*(6), 1154–1167.
- Krishna, A. (2012). An integrative review of sensory marketing: Engaging the senses to affect perception, judgment, and behavior. *Journal of Consumer Psychology*, *22*(3), 332–351.
- Krishna, A., & Schwarz, N. (2014). Sensory marketing, embodiment, and grounded cognition: A review and introduction. *Journal of Consumer Psychology*, *24*(2), 159–168.
- Kumar, A., & Tan, Y. (2015). The demand effects of joint product advertising in online videos. *Management Science*, *61*(8), 1921–1937.
- Lee, L., & Ariely, D. (2006). Shopping goals, goal concreteness, and conditional promotions. *Journal of Consumer Research*, *33*(1), 60–70.
- Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*(51), 20661–20665.
- Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394.
- Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception & Performance*, *15*(1), 69–79.
- Owens, A., & Efros, A. A. (2018). Audio-visual scene analysis with self-supervised multisensory features. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science* (Vol. 11210).
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*(8), 252–271.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(7), 534–552.
- Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research*, *10*(2), 135–146.
- Puccinelli, N. M., Goodstein, R. C., Grewal, D., Price, R., Raghuram, P., & Stewart, D. (2009). Customer experience management in retailing: Understanding the buying process. *Journal of Retailing*, *85*(1), 15–30.
- Puccinelli, N. M., Chandrashekar, R., Grewal, D., & Suri, R. (2013). Are men seduced by red? The effect of red versus black prices on price perceptions. *Journal of Retailing*, *89*(2), 115–125.
- Roberts, J. H., & James, M. L. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, *28*(11), 429–440.
- Shen, H., & Sengupta, J. (2014). The cross-modal effect of attention on preferences: Facilitation versus impairment. *Journal of Consumer Research*, *40*(5), 885–903.
- Shiv, B., & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, *26*(3), 278–292.
- Song, Y., Chee, W., Yang, S., & Luo, X. (2018). The effectiveness of contextual competitive targeting in conjunction with promotional incentives. *International Journal of Electronic Commerce*, *22*(3), 349–385.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(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*(1), 37–54.
- Steinhart, Y., Mazursky, D., & Kamins, M. A. (2013). The process by which product availability triggers purchase. *Marketing Letters*, *24*(3), 217–228.
- Suri, R., & Monroe, K. (2003). The effects of time constraints on consumers' judgments of prices and products. *Journal of Consumer Research*, *30*(1), 92–104.
- Tavassoli, N. T., & Lee, Y. H. (2003). The differential interaction of auditory and visual advertising elements with Chinese and English. *Journal of Marketing Research*, *40*(4), 468–480.
- Taylor, S., & Todd, P. (1995). An integrated model of waste management behavior: A test of household recycling and composting intentions. *Environment and Behavior*, *27*, 603–630.
- van Zee, E. H., Paluchowski, T. F., & Beach, L. R. (1992). The effects of screening and task partitioning upon evaluations of decision options. *Journal of Behavioral Decision Making*, *5*, 1–23.
- Yang, S., & Xiong, G. (2019). Try it on! Contingency effects of virtual fitting rooms. *Journal of Management Information Systems*, *36*(3), 789–822.
- Yang, S., Song, Y. P., & Pancras, J. (2017). Matching exactly or semantically? An examination of the effectiveness of synonym-based matching strategy in Chinese paid search market. *Journal of Electronic Commerce Research*, *18*(1), 32–51.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of Consumer Research*, 12(3), 341–352.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Shuai Yang¹ · **Xinyu Chang**¹ · **Sixing Chen**² · **Shan Lin**³ · **William T. Ross Jr**⁴

¹ Glorious Sun School of Business and Management, Donghua University, Shanghai, China

² School of Management, Guangdong University of Foreign Studies South China Business College, Guangzhou, Guangdong, China

³ College of Business, Lewis University, Romeoville, USA

⁴ School of Business, University of Connecticut, Storrs, USA