This file was downloaded from BI Brage,
the institutional repository (open access) at BI Norwegian Business School
http://brage.bibsys.no/bi

Multisensory Brand Search: How the Meaning of Sounds Guides Consumers' Visual Attention

Klemens M. Knöferle
BI Norwegian Business School

Pia Knöferle
Humboldt University

Carlos Velasco
University of Oxfor

Charles Spence
University of Oxford

This is the accepted and refereed manuscript to the article published in

Journal of Experimental Psychology: Applied, 22(2016)2:196-210

"This article may not exactly replicate the authoritative document published in the APA journal. It is not the copy of record."

Publisher's version available at http://dx.doi.org/10.1037/xap0000084

The publisher, American Psychological Association, allows the author to deposit the version of the article that has been accepted for publication, in an institutional repository. http://www.apa.org/pubs/authors/posting.aspx

Multisensory Brand Search:

How the Meaning of Sounds Guides Consumers' Visual Attention

Klemens M. Knoeferle[1], Pia Knoeferle[2], Carlos Velasco[3], & Charles Spence[3]

[1] *BI Norwegian Business School, Department of Marketing, Norway*

[2] *Humboldt University, Berlin, Department of German Language and Linguistics, Germany*

[3] *University of Oxford, Crossmodal Research Laboratory, Department of Experimental Psychology, United Kingdom*

Correspondence concerning this article should be addressed to Klemens M. Knoeferle, Department of Marketing, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Email: klemens.knoferle@bi.no

Acknowledgments: Parts of this research were funded by the Swiss National Science Foundation during the first author's post-doctoral studies at the University of Oxford (grant PBSGP1_141347), by the Research Fund of the Department of Marketing, BI Norwegian Business School, and by the Cognitive Interaction Technology Excellence Center 277 (DFG, German Research Council) at Bielefeld University, Germany.

Abstract

Building on models of crossmodal attention, the present research proposes that brand search is inherently multisensory, in that the consumers' visual search for a specific brand can be facilitated by semantically related stimuli that are presented in another sensory modality. A series of five experiments demonstrates that the presentation of spatially non-predictive auditory stimuli associated with products (e.g., usage sounds or product-related jingles) can crossmodally facilitate consumers' visual search for, and selection of, products. Eye-tracking data (Experiment 2) revealed that the crossmodal effect of auditory cues on visual search manifested itself not only in reaction times, but also in the earliest stages of visual attentional processing, thus suggesting that the semantic information embedded within sounds can modulate the perceptual saliency of the target products' visual representations. Crossmodal facilitation was even observed for newly-learnt associations between unfamiliar brands and sonic logos, implicating multisensory short-term learning in establishing audiovisual semantic associations. The facilitation effect was stronger when searching complex rather than simple visual displays, thus suggesting a modulatory role of perceptual load.

Keywords: *visual attention; eye-tracking; product search; crossmodal; sound; learning*

Multisensory Brand Search:

How the Meaning of Sounds Guides Consumers' Visual Attention

Searching for a product, no matter whether in a traditional retail environment or in an online store, can be challenging. A typical shopper may be exposed to as many as 1,000 different products and brands per minute whilst strolling down the aisles of the typical supermarket (Robinson 1999). In light of this overwhelming product choice, reducing the visual search time, and facilitating the customer's search for specific products (or product categories) is of great interest to marketers (see Wedel and Pieters 2007, for a review detailing the importance of visual attention in marketing). One way in which this could be achieved is by engaging/stimulating the customer's other senses. Indeed, research in the field of cognitive psychology already provides robust evidence for multisensory-integration effects on visual sensitivity and search (Chen and Spence 2010, 2011; Chen, Yeh, and Spence 2011; Klapetek, Ngo, and Spence 2012; Ngo and Spence 2010; Parise and Spence 2009; Parrott et al. 2015; see e.g., Santangelo, Ho, and Spence 2008; Spence and Driver 2004 for models of multisensory integration and crossmodal attention).

However, these studies have typically used "basic", decontextualized stimuli (e.g., rotated line segments and Gabor patches) and require hundreds of within-participant trials to obtain relatively small effects. It thus remains an open question as to whether multisensory stimulation substantially affects consumers' search behaviour in real-life settings and in the more complex human-made environments of consumer search. Indeed, in the consumer behaviour literature, the potential effects of crossmodal stimulation on visual processing have largely been neglected (but see Hagtvedt and Brasel, in press). In particular, studies examining consumers' brand search have focused on the visual modality (e.g., Atalay, Bodur, and Rasolofoarison 2012; Pieters and Warlop 1999; Van der Lans, Pieters, and Wedel 2008).

To address this gap in the literature, the present paper examines whether, and under which conditions, the *semantic* content of sounds (e.g., the crunching of crisps) facilitates people's product search for semantically related targets such as crisps, assessed by means of response time and eye gaze measures. Our focus on semantic facilitation complements and extends earlier work studying the attention-modulating effects of spatially informative (Shen and Sengupta 2014, although attention was assessed only via self-reports) and synesthetically matching sounds (Hagtvedt and Brasel, in press). In addition, we test several potential boundary conditions of semantic audiovisual facilitation: First, we assess whether such facilitation requires long-term associations between visual product representations and sounds (formed through frequent co-exposure), or whether it can also arise from short-term learning (i.e., following no more than three co-exposures to a product and its jingle). Second, realistic consumer search contexts differ in terms of their informational richness, thus prompting questions of cue integration in situations with a high versus low visual load. We thus assess the extent to which product-usage sounds modulate visual attention also (or especially) in high-load contexts, a finding that would speak to the general relevance and robustness of semantic audiovisual integration in consumer contexts. Third, we examine the extent to which any observed effects are robust in different variants of the search task and under conditions that are both realistic and "noisy" (e.g., in between-participant designs and using a virtual online store as our experimental platform).

Assuming that robust semantic audiovisual facilitation effects can be demonstrated across all of these tasks and contexts, our conclusion will be that spatially uninformative product-related sounds cue visual attention in consumer product search. From an applied perspective, this sort of conclusion would potentially encourage the use of multisensory cues in the design of advertisements, retail, and online environments, but also of multisensory user interfaces.

To motivate both our own research question and its relevance to the marketing community, we begin by summarizing the available research on crossmodal attention and multisensory learning. This research suggests that the attentional salience of a visual target can be influenced by the semantic identity of a non-visual stimulus through (experience-based) associations, but is limited in terms of applicability by not taking into account the boundary conditions mentioned above. To address these limitations, we report five experiments in which we assess how product-related sounds affect consumers' visual search performance (RTs and eye movements) in multi-product displays. We discuss the theoretical and practical implications of our findings as well as several avenues for future research.

Conceptual Background

Semantically-Associated Sounds Modulate Visual Attention

A considerable body of empirical research in cognitive psychology has demonstrated that visual attention can be modulated by crossmodal interactions, that is, inputs from other sensory modalities such as smell, touch, and sound (see Stein 2012; see also Shimojo and Shams 2001). Rather than processing the information from different sensory channels separately, the human brain constantly combines ("integrates") these inputs (Shimojo and Shams 2001) in order to reduce sensory uncertainty (Alais and Burr 2004). Due to reduced sensory uncertainty, the integration of auditory and visual events can enhance the saliency of the component stimuli, such that audiovisual events or objects capture attention more easily compared to unimodal events (e.g., Matusz and Eimer 2011; Spence 2010; Stein et al. 1996; Van der Burg et al. 2011).

What do we know about visual attention modulation through the semantic meaning of sounds? Clearly, learned associations between certain auditory and visual cues can facilitate audiovisual

integration (Chen and Spence 2010; Fiebelkorn et al. 2010, 2012; Schneider, Engel, and Debener, 2008; Suied, Bonneel, and Viaud-Delmon, 2009; see Doehrmann and Naumer 2008, for a review on semantics and multisensory information processing). For instance, Chen and Spence (2010) found that briefly (for 27 ms) presented and then masked pictures (e.g., of a dog) were more often correctly identified if they were accompanied by a semantically-related sound (e.g., a bark) than by an incongruent sound or white noise. Congruent sound facilitated image identification when the sound and picture onset simultaneously, but also when the sound was presented up to 300 ms after the picture (thus delineating a time window of semantic multisensory integration).

Sounds that are semantically congruent with a visual target (e.g., a cat's "meow") can also influence the time taken by participants to locate a visual target (Iordanescu et al. 2008). The participants located a target (e.g., the picture of a cat among four visual objects) faster when it was accompanied by a congruent sound (e.g., the "meowing" of a cat) as compared with an incongruent sound (e.g., the sound of a train), or no sound. However, this sort of crossmodal facilitation was eliminated when people looked for the word "cat" rather than for the picture of a cat. Congruent sounds therefore appeared to enhance the activation of pictorial rather than linguistic-semantic target representations.

Using eye tracking, Iordanescu et al. (2010) further showed that congruent auditory stimuli also reduce the amount of time it takes participants to make an eye saccade to the target. The participants saccaded to the picture of a dog more rapidly when listening to a barking sound (versus an incongruent sound or no sound). Interestingly, while target-congruent sounds guided the participants' initial saccades towards the target, sounds that were congruent with a nearby distractor object (henceforth "distractor-congruent" sounds) failed to elicit saccades toward the distractor. Iordanescu and her colleagues therefore proposed that these effects arise from top-down sensitization and bottom-up crossmodal interactions. The target representation might be sensitized

by a top-down signal originating in the prefrontal cortex, which is known to selectively respond to target-relevant cues (see, e.g., Duncan 2001). While a target-congruent sound could then crossmodally enhance the activation of the sensitized visual target representation, such enhancement would not apply to distractor-congruent sounds, since their visual (non-target) representations would not have been sensitized.

In summary, multisensory integration of semantically congruent images and sounds leads to a spread (or "spill over") of attention across object features in different sensory channels.[1] Certainly, event-related brain potential (ERP) and functional magnetic resonance imaging (fMRI) studies confirm this view, indicating that the bias to process objects "as a whole" extends across sensory boundaries (Busse et al. 2005; Molholm et al. 2007).

The Role of Short-Term Learning in the Crossmodal Modulation of Attention

While the above studies corroborate the idea that semantically-associated sounds can speed up the visual search for a picture when associations among the multisensory features of the objects are long-term (typically shaped through extensive experience with that object), real-world marketing often relies on a relatively small number of short-term exposures to novel audiovisual combinations of stimuli. Can such associations, established over a relatively limited number of exposures, give rise to comparable facilitation effects in visual search?

Visual recognition of an object is facilitated (hindered) after participants have been exposed once to the visual object and its congruent (incongruent) sound (as compared to a visual-only initial presentation); such learning effects are believed only to occur when the visual and auditory stimuli are congruent during initial exposure (Murray, Foxe, and Wylie 2005; Thelen and Murray 2012, 2013; Thelen, Talsma, and Murray 2015). However, Liu, Wu, and Meng (2012) observed effects of a single exposure even for ecologically unrelated stimuli: In a study phase, participants learned

ecologically unrelated audiovisual associations (e.g., four meaningless tones and pictures from the Snodgrass and Vandervart, 1980, corpus) and then completed a crossmodal priming experiment. The processing of a tone was facilitated when it was preceded by its visual associate from the study phase, as revealed by reduced amplitude N400s (an ERP signal taken to reflect the processing of semantic meaning, see Kutas and Federmeier 2011) in the ERP recording, amongst other measures.

Thus, short-term multisensory learning can facilitate subsequent object *recognition*. But would short-term semantic associations also modulate visual *attention* and search? And, if so, would this occur for audiovisual stimuli that are initially ecologically unrelated, such as products and jingles, and even in relatively complex displays with as many as 16 products?

The Role of Visual Load in Audiovisual Facilitation

With regard to questions of complexity, it is reasonable to assume real-life product search in commercial contexts occurs under varying conditions of visual complexity and load. For instance, searching for a product in a visual display that contains more (fewer) objects would typically elicit a higher (lower) visual perceptual load, thus leading to an increase in search times. Previous research suggests that visual load also moderates the audiovisual facilitation effect. Specifically, multisensory integration in humans follows the "principle of inverse effectiveness" (PoIE; Meredith and Stein 1986; Stein and Meredith 1993; Stein & Stanford 2008). The PoIE states that the enhancement of multisensory stimuli will be greatest when the constituent unisensory stimuli are ambiguous or noisy. When they are unambiguous and reliable, however, the need for multisensory integration decreases. For instance, co-occurring bursts of sound enhance human detection sensitivity for low (but not high) intensity Gabor patches (Noesselt et al. 2010). Similarly, viewing a speaker's articulatory movements improves a listener's spoken word understanding most

when the signal-to-noise ratio of the speech sounds is low (Ross, Saint-Amour, Leavitt, Javitt, & Foxe 2007).

Applied to the current question of auditory facilitation of product search, the target product of a search can be construed as the “signal”, while the distractor products constitute visual “noise”. The PoIE would thus predict a larger auditory facilitation effect in product displays with more distractor objects. This would be a situation with high noise in which participants should benefit most from cross-modal integration. Importantly, to our knowledge, the predictions of the PoIE have been examined neither for visual search nor in the context of consumer behaviour.

Overview of Experiments

We tested the implications of a multisensory approach to brand search across five experiments (four laboratory and one online experiment). Experiment 1 aimed to conceptually extend the auditory facilitation effect from simple psychological stimuli (Iordanescu et al. 2008) to a product search context. Experiment 2 assessed whether any auditory facilitation effect on visual search generalizes to product displays comprising a much larger number of different products (N = 16) and used eye tracking to evaluate auditory facilitation during early stages of processing in visual search. Experiment 3 tested whether the crossmodal facilitation would generalize to sound-brand associations learnt through short-term within-experiment co-occurrences. Finally, Experiments 4A and 4B evaluated auditory facilitation in a more everyday life situation and used a between-participants design to test the auditory facilitation effect in an online store context Experiment 4B further manipulated an important boundary condition of the effect—the complexity of the visual search display.

Experiment 1 – Do Usage-Related Product Sounds Facilitate Visual Search For Products?

Overview

Experiment 1 was designed to conceptually replicate the auditory facilitation effect reported in prior research (Iordanescu et al. 2008, 2010) in a marketing context. In order to study the effect of product-related sound on visual information processing, we used a single-target visual search task with repeated trials. This task is standard in the visual attention literature (Quinlan 2003; Treisman and Gelade 1980; Wolfe 1998), and involves active scanning of the visual environment for a target object among visual distractor objects, thus modeling actual consumer product search (Van der Lans et al. 2008). If experience-based multisensory knowledge of consumer products (e.g., the crunching of potato chips) modulates target search in a simulated shelf display with four different products, then detecting a target product (e.g., in a target present / absent task) should be faster with a matching product-usage sound (e.g., the crunching of potato chips) as compared to a silent baseline condition and an incongruent sound condition. Incongruent or distractor-congruent sound conditions should not differ from the silent condition.

Design and Stimuli

Both the experimental design and the procedure of the visual search task used in Experiment 1 were adopted from Iordanescu et al. (2008). However, instead of presenting objects from widely different categories (e.g., coins, dogs, trains) with well-established object-sound relationships, we presented different products, commonly available in the setting of a supermarket, together with their characteristic sounds (see Appendix D for details on the experimental stimuli and the instructions used in the current study).

In each trial, four product pictures were presented in a 2 × 2 search display (see Figure 1A). Each visual search display showed a virtual store shelf containing four different products (with multiple facings depending on size as in real environments, see Van der Lans et al. 2008). The four products were randomly selected from a total set of nine branded products belonging to different categories (e.g., food, beverages, personal hygiene), as found in a typical supermarket: Sparkling wine, potato chips, bacon, whipped cream, deodorant, bath gel, facial tissues, and a digital camera. One of the four products shown in each search display was the search target, while the remaining three products constituted the visual distractors. Each product picture was scaled to fit within an area of 10.5° × 9.1° visual angle.

Insert Figure 1 about here

Sound was manipulated as a within-participants factor with four levels: In each trial, the auditory stimulus was either (1) congruent with the target (target-congruent), (2) congruent with one of the distractors (distractor-congruent), (3) congruent with none of the products that were presented in the trial (incongruent), or (4) absent (no sound/control condition). Target-congruent sounds were semantically related to the target (e.g., the sound of the uncorking of a champagne bottle), distractor-congruent sounds were related to the product shown in the corner of the display opposite to the target, and incongruent sounds did not correspond to any of the products shown in the display. For each product, we sourced a typical sound either of the product itself (e.g., a camera shutter sound), of its packaging (e.g., a champagne bottle opening sound), or of a related event (e.g., a sneezing sound for the facial tissues) from a royalty-free online database of sounds (www.freesound.org; 32-bit mono, 44,100 Hz digitization). The duration of the sounds varied between 250 and 2500 ms due to the distinct nature of the recorded events. Details about the

auditory stimuli can be found in Appendix A. The sounds were played back via two loudspeakers, one on either side of the screen, at an average sound level of 70 dB.

Procedure

Fifty (student) participants at an English University participated in the study (28 female, mean age = 27.6 years, *SD* = 8.8, range 18-57 years), for which they received £5. All of the participants reported normal or corrected-to-normal vision and normal hearing.

The participants were seated in a darkened, sound-proofed experimental booth approximately 60 cm from the screen (17-inch CRT-monitor, 1280 × 1024 pixels resolution, 85 Hz refresh rate), with their chin placed on a chin rest (see Figure 1D). They received onscreen instructions about the experiment and were informed that their task was to search for products presented on-screen. They were further informed about the trial structure. On a given trial, the name of the target product (e.g., “Kellogg’s Corn Flakes”) would appear. After this, they would see a fixation cross, followed by the supermarket shelf with the four products. The participants were instructed to fixate the fixation cross whenever it was present. Once the shelf display was presented, they were instructed to indicate, as rapidly and accurately as possible, the location of the target product by pressing “f”, “j”, “v”, or “n” for the upper left, upper right, lower left, or lower right screen quadrant, respectively. The participants were also informed that the product display would sometimes be accompanied by a sound.

The experiment was self-paced, and the participants started each trial by pressing the space bar. The name of the target product (e.g., “Walkers crisps”) appeared on-screen for 1000 ms, followed by a fixation cross at the centre of the screen. After 350 ms, the visual search display, consisting of the target and three distractor products, was presented for 2500 ms. The number of times each target product appeared across conditions was counterbalanced. The distractor objects

were selected randomly from the remaining (non-target) products in each trial; locations of target and distractor products were assigned randomly in each trial. The visual search display was randomly accompanied by a target-congruent sound, a distractor-congruent sound, an incongruent sound, or else by no sound. The auditory cue preceded the visual stimulus by 100 ms; this brief asynchrony, however, was well within the temporal range known to allow multisensory integration (see, e.g., Vatakis and Spence 2010). Importantly, the sounds did not carry any spatial information regarding the location of the target item within the display, and participants reported that they perceived the sound as localized to the middle of the display. Each trial ended with a 500 ms feedback screen indicating whether or not the participant had responded correctly or not. The participants received a practice session consisting of 12 randomly selected trials (three from each sound condition). They then completed six blocks (except for four participants, who completed 4 blocks), each consisting of 36 trials, that is, nine trials for each of the four experimental conditions. The experiment lasted for approximately 15 minutes.

Results and Discussion

The total number of trials was 10,512, from which we excluded timeouts ($N = 54$) and incorrect responses ($N = 346$). We computed mean RTs by participants and condition for correct, target-present trials (trimmed minimally by 1.5% at each end of the RT distribution; see, e.g., Baayen and Milin 2010, on minimal trimming). A RM-ANOVA revealed a statistically significant difference in target localization latencies across the different sound conditions, $F(3, 147) = 2.95$, $p = .035$, $\eta_p^2 = .06$. Motivated by the present hypotheses and prior results (Iordanescu et al. 2008), planned contrasts confirmed that participants found the target product significantly more rapidly when exposed to a target-congruent sound ($M = 624$ ms, $SD = 162$ ms) rather than no sound ($M = 638$ ms, $SD = 163$ ms; $F(1, 49) = 5.60$, $p = .022$, $\eta_p^2 = .10$). Distractor-congruent sounds or incongruent

sounds did not significantly influence response times as compared to the no-sound condition (all *p*s > .53). Relative to the incongruent sound condition (*M* = 637 ms, *SD* = 164 ms), congruent sounds sped-up response times, $F(1, 49) = 4.49$, $p = .039$, $\eta_p^2 = .08$, while other conditions did not differ. The RT findings are summarized in Figure 2. There was no effect of the sound condition on response accuracy (see Appendix B for complementary analyses). Experiment 1 thus confirmed that the modulation of crossmodal attention (based on long-term semantic associations) generalizes to a marketing-relevant context.

Insert Figure 2 about here

Experiment 2 – Identifying the Attentional Processes Underlying the Crossmodal Facilitation Effect via Eye-Tracking

Overview

Experiment 2 was designed to extend the results of Experiment 1 to a multi-product display (N = 16) and to investigate the overt attentional processes underlying the facilitatory effect of sound on visual search performance. The previous experiment provided evidence that product-congruent sounds (compared to no-sound trials) can speed-up manual response times to visual targets. Manual RT measures are a standard outcome variable in visual search tasks, and valid proxies of attention (e.g., Eckstein 2011). However, they reflect not just attentional but also response-related processes, such as confirming the identity of the target object, mapping the perceptual decision to a motor response, and executing the motor response (Donders 1969; Luce 1986). To disentangle the perceptual and response-related components of the RT facilitation observed in Experiment 1,

Experiment 2 used a more direct measure of visual attention (eye movements). If long-term associations between sounds and visual representations influence participants' immediate visual attention to the target product (rather than just their selection or execution of a manual response), then they should speed-up eye movements towards the target product, and not just manual RTs.

Design and Stimuli

Since, in Experiment 1, the participants' performance was the same no matter whether the sound was congruent with a distractor or congruent with one of the products that were not presented in the trial, we decided to remove the distractor-congruent condition, thus yielding three experimental conditions: The sound could either be (1) congruent with the target, (2) absent, or (3) incongruent with the target (and congruent with a product that was not displayed).

Each visual search display contained 16 product pictures randomly arranged in a 4 × 4 matrix seen against a photorealistic shelf background (see Figure 1B). On each trial, the 16 products shown in the display were randomly selected from a set of 18 products belonging to different categories. Out of the set of 18 products, seven were potential targets, while the other 11 were used as visual distractors only throughout the experiment (since they were not linked to specific usage or consumption sounds). An example product picture had a square format and subtended 4.7° of visual angle with a centre-to-centre distance of approximately 5.4° of visual angle. The background picture (of the product shelf) subtended 24.9° of visual angle. For each of the seven potential target products, a typical usage or consumption sound was downloaded from a royalty-free online sound database (www.freesound.org; 32-bit mono, 44,100 Hz digitization). The duration of the sounds varied between 250 and 2500 ms. The auditory stimuli were presented via two loudspeakers (one to the left and the other to the right of the display); loudness was equalized and set to an average level of 70 dB.

Procedure

Fifty participants (34 female, mean age = 23.4 years, *SD* = 3.0, range 19-30 years) recruited at a large public German university took part in the experiment. Thirty were compensated with 6 € and twenty with 8 €. All participants reported normal or corrected-to-normal vision.

Upon arrival, the participants were informed about the nature of the experiment and that their eye movements would be recorded (SR Research Eyelink 1000 Desktop head-stabilized eye tracker, SR Research, Mississauga, Ontario, Canada, 9-point calibration with < 1 degree mean-average error). The participants were seated approximately 85 cm from the screen (22-inch color monitor; 1680 × 1050 pixels resolution, 120 Hz refresh rate) with their chin placed on a chin-rest.

The participants filled in demographic details and familiarized themselves with the English products by means of labeled product pictures presented on a sheet of paper. They received onscreen instructions about the experiment and were told that their task was to search for products in a supermarket shelf. They were further informed about the trial structure. On a given trial, a fixation dot would briefly appear, followed by the name of the target product (e.g., “Kellogg’s Corn Flakes”). After this, they would once again see a dot, followed by the supermarket shelf with the products. The participants were instructed to fixate the dot whenever it was present and once the shelf display was presented, they should decide as quickly and accurately as possible whether the target product was (vs. was not) on the shelf. Participants were also told that the product display would often be accompanied by a sound which they were asked to ignore. They indicated their response via a button press (“yes” vs. “no”) on a Cedrus RB834 response box. Assignment of “yes” /”no” responses to the two response buttons (left vs. right) was counterbalanced across participants. In contrast to Experiment 1, the auditory cue and the visual search display onset simultaneously.

The experiment consisted of four blocks of 84 trials (63 target-present, 21 target-absent) each, totaling 336 trials. After each block of trials, the participants took a short break and were re-calibrated. Additional calibration was performed as necessary. Overall, the experiment lasted for approximately 60 min. The participants were debriefed after the experiment.

Results and Discussion

Manual RTs. RTs were computed time-locked to display onset time. The total number of trials was 16,800 (336 trials × 50 participants) from which we excluded trials without an overt response (N = 160); we next limited the data to target-present trials (N = 12,568) and of those, we further excluded incorrect responses (N = 570, 4.54% of 12,568 target-present trials).

We computed mean RTs by participants and condition for correct, target-present trials (trimmed by 1.5% at each end of the RT distribution). These summed RTs were entered as dependent variables into a RM-ANOVA with the factor "sound" ("congruent" vs. "no sound" vs. "incongruent"). RTs were normally distributed for by-participant mean RTs summed across the three factor levels using a K-S test, $p > .09$ (Tabachnick and Fidell 2007), but not for the individual levels (rendering simple contrasts problematic). We therefore log-transformed the data prior to the ANOVA analyses (results did not change substantially when the analyses were instead conducted on the non-transformed data). Based on prior RT results and previous research (Iordanescu et al. 2008, 2010), we expected RTs in the congruent sound condition to be significantly faster than in the no-sound and the incongruent sound conditions. We did not expect any differences between the incongruent and no-sound conditions. Accordingly, we conducted three planned contrasts (simple) to follow up a main effect in the ANOVA results (congruent vs. no sound, congruent vs. incongruent sound, and incongruent vs. no sound). There was a significant main effect of condition on RTs, $F(2, 98) = 7.99$, $p = .001$, $\eta_p^2 = .14$. Simple contrasts revealed significantly faster responses

in the congruent than in the no-sound condition, $F(1, 49) = 10.17$, $p = .002$, $\eta_p^2 = .17$, but not in the incongruent relative to the no-sound condition, $F(1, 49) = 1.25$, $p = .268$, $\eta_p^2 = .03$, see Figure 3A. Responses in the congruent condition were faster than in the incongruent condition, $F(1, 49) = 11.55$, $p = .001$, $\eta_p^2 = .19$. Additional analyses of the effect of the sound condition on response accuracy are summarized in Appendix B.

Insert Figure 3 about here

Eye-movement analysis. In a first filtering step, we merged nearby fixations < 80 ms with neighboring fixations (< 1 degree). We defined 16 areas of interest (AOIs) and coded the AOI position of the target. Just as for the RT analyses, we selected those trials in which the target was present and in which the participant answered correctly ($N = 11{,}998$). We computed the following eye-movement measures for each target by condition and participants (with trimming set to 1.5% from each end of the eye-movement measures distribution): The onset latency of the first saccade to the target, the onset latency of the last saccade, and the mean ordinal sequence of the first fixation to the target. Selection of these measures was guided by the complementary insights that they provide into visual processing. Saccades are preceded by a covert attentional shift to the saccade target (Deubel and Schneider 1996; Rayner, McConkie, and Ehrlich 1978; Shepherd, Findlay, and Hockey 1986) and thus provide a good measure of the locus of attention. The onset latency of the first saccade to the target post-display onset provides insight into *initial* visual orienting (see also Iordanescu et al. 2010), and in our design reflects the initial response of visual attention to the different sounds. The onset latency of the last saccade to the target reflects comparatively *later* visual attention shifts to the target, before a participant's manual response. Differences in the sound effects on these two measures can provide insight into the crossmodal priming leading up to the

manual response. The third gaze measure, mean ordinal sequence of the first target fixation indexes the average number of fixations prior to fixating the target, a measure of interest for marketing if the goal is to direct attention to a specific product before it goes to another product (see Castelhano and Henderson 2007, for use of these measures in visual search).

The averaged eye-movement measures were subjected to a one-way RM-ANOVA with three factor levels ("congruent" vs. "no sound" vs. "incongruent"). We report Greenhouse-Geisser adjusted *p*-values when sphericity was violated. To the extent that the prior RT pattern were driven by visual attention processes, saccade onsets should be faster and mean ordinal position of the first target fixation lower in the congruent than in the no-sound and the incongruent sound conditions. We did not expect any differences between the incongruent and no-sound conditions. Accordingly, we conducted three planned contrasts (simple) to follow up a main effect in the ANOVA results (congruent vs. no sound and incongruent vs. no sound).

Eye-movement results. Analyses of the first saccade onset times to the target revealed a marginal main effect of condition, $F(2, 98) = 2.75$, $p = .069$, $\eta_p^2 = .05$, see Figure 3B. Contrasts confirmed marginally faster first saccade onset times for the congruent ($F(1,49) = 3.55$, $p = .066$, $\eta_p^2 = .07$), but not for the incongruent sound condition ($F < 1$), relative to the no-sound condition. First saccade onset times were also marginally faster in the congruent condition than the incongruent condition, $F(1, 49) = 3.78$, $p = .058$, $\eta_p^2 = .07$. The analyses of the last saccade onset times to the target revealed a clear effect of condition, $F(1.79, 87.71) = 4.15$, $p = .023$, $\eta_p^2 = .08$. Simple contrasts confirmed faster onset times of the last saccade to the target for the congruent relative to the no-sound condition ($F(1, 49) = 5.16$, $p = .028$, $\eta_p^2 = .10$), but not for the incongruent relative to the no-sound ($F < 1$) condition. Last saccade onset times were also significantly faster in the congruent condition than the incongruent condition, $F(1, 49) = 5.89$, $p = .019$, $\eta_p^2 = .11$ (see

Figure 3C). Finally, a significant effect also emerged in the mean ordinal sequence of the first fixation to the target object, $F(1.79, 87.50) = 5.26$, $p = .009$, $\eta_p^2 = .10$, see Figure 3D. Simple contrasts corroborated that the first fixation to the target occurred after fewer fixations to other products in the congruent than the no-sound condition ($F(1, 49) = 11.01$, $p = .002$, $\eta_p^2 = .18$), while this was not the case for the incongruent relative to the no-sound condition ($F < 1$). Fewer fixations to other products also were observed in the congruent relative to the incongruent condition, $F(1, 49) = 4.12$, $p = .048$, $\eta_p^2 = .08$.

Discussion. Crucially, Experiment 2 replicated the RT findings of Experiment 1 and extended them to a more complex product display. Sound affected the onset of the last saccade to the target, and, albeit less clearly, the onset of the initial saccade, revealing sound effects on overt visual attention. Of interest from a marketing perspective was that participants made fewer fixations prior to fixating the target when the sound was congruent (vs. absent or incongruent). Non-target products thus received less overt attention when a target-congruent sound was presented.

Experiment 3 – Does the Crossmodal Facilitation Effect Extend to Newly Learned Associations Between Sounds and Brands?

Overview

Thus, sound can facilitate visual search for a product when sound and product are related through prior long-term experience (Experiments 1 and 2). By contrast, the role of short-term experience for such facilitation remains unclear. Would newly learned associations between an unfamiliar brand and an unknown jingle (or a sonic logo) result in similar facilitation? Experiment

3 probed this question by establishing sound-brand associations as part of the experimental procedure and testing their effect on RTs in a subsequent visual search task.[2]

In addition, Experiment 3 extends and generalizes the findings of the previous experiments to a situation in which the target and the visual distractors all belong to a single product category (as would be the case while one stands in front of a supermarket shelf). Since products within a given product category typically share (at least some) visual features, using a single product category likely increases the perceptual similarity between the target product and the distractor products, which in turn should reduce overall search performance (see, e.g., Duncan and Humphreys 1989). Thus, Experiment 3 tests the facilitation effect in a brand-specific search setting of advertising jingles associated with different brands (rather than sounds linked to generic product categories, as was the case in Experiments 1 and 2).

Design and Stimuli

The design of the visual search task was based on Experiment 2. However, since in that experiment, the incongruent condition did not differ substantially from the no-sound condition, the incongruent sound condition was dropped, thus resulting in two within-participant conditions (congruent sound vs. no sound).

Novel visual (product depictions) and auditory stimuli (sonic logos) were created in order to eliminate the potential influence of pre-existing auditory-visual associations. In contrast to Experiments 1 and 2, we used a single product category (liquid laundry detergents). A professional designer created 19 novel designs for the packaging of laundry detergent (by varying extant packaging designs). In addition, a sound design firm created four short musical sounds (realistic sonic logos; duration range: 1700 to 2400 ms, 32-bit stereo, 48,000 Hz digitization). As in Experiment 2, each search display consisted of 16 products presented on a photorealistic shelf

background (29.1° × 27.6° of visual angle). All of the product pictures had a square format and subtended 6.2° × 6.2° of visual angle with a centre-to-centre distance of approximately 7.3° (see Figure 1C). Auditory stimuli were presented using closed headphones at an average sound level of 70 dB.

Procedure

Forty-nine student participants (31 female, mean age = 23.8 years, *SD* = 4.2, range 19-35 years) at a Norwegian university took part in the experiment, for which they received partial course credit or money (equivalent to approximately 17 USD). All of the participants reported normal hearing and normal or corrected-to-normal vision.

The participants arrived at the laboratory in groups of two to nine at a time, where they were seated in individual cubicles approximately 55 cm from the screen (22-inch CRT monitor; 1280 × 1024 pixels resolution, 60 Hz refresh rate). Each experimental session consisted of two tasks: First, the participants completed a self-paced associative learning task introduced as a memory test, which consisted of a learning phase and a test phase. In the learning phase, the participants were exposed to a series of combinations of products and sounds (4 laundry detergent brands paired with short musical sounds). After three exposures of each sound-brand combination (in a randomized order), the participants completed a matching test that required them to pair each of the previously seen brands with the previously learned associated sound. Any incorrect matching initiated an error feedback message, an additional round of exposure, and an additional matching task. Only one of the participants did not correctly match all sound-brand combinations on the first matching task, but succeeded on the second attempt after additional training.

Next, the participants started the main experiment (a visual search task). The trial structure was identical to that used in Experiment 2. Target-present/absent responses were given using the

left and right arrow keys of a standard USB keyboard. The key assignment was counterbalanced across participants. After a short practice session of eight trials with accuracy feedback displayed after each trial, the participants completed four experiment blocks of 32 trials (without feedback), totalling 128 trials per experiment session. Each block consisted of 24 target-present trials and eight target-absent trials (catch trials). Each of the four potential target products appeared equally often as a target in each of the two sound conditions; the location of the target was randomized in each trial. On average, it took the participants 13 minutes to complete the visual search task. The experiment was designed and carried out using E-Prime 2.0.

Results and Discussion

The total number of trials was 6,272. We limited the data to target-present trials (N = 4,642), from which we excluded timeouts (N = 62). Of those, we further excluded incorrect responses (N = 544). We computed mean RTs by participants and condition for correct, target-present trials (trimmed by 1.5% at each end of the RT distribution). Two participants with a mean accuracy that fell more than two standard deviations below the accuracy sample mean (which indicated random response selection) were excluded from further analyses (corrected mean accuracy 90%, range: .74-.98, SD = 0.07). Including these participants changed neither the pattern nor the significance of the RT results. A RM-ANOVA revealed a statistically significant difference in target localization latencies across the two sound conditions, $F(1, 46) = 4.68$, $p = .036$, $\eta^2_{partial} = .09$. Target-congruent sounds (M = 1176 ms, SD = 234 ms) resulted in faster search latencies than no sound (M = 1211 ms, SD = 230 ms). We also overall obtained the expected result that search latencies were longer in Experiment 3 than in either Experiments 1 or 2, thus corroborating our expectations that within-category product search is comparatively more difficult than searching for

a product in a display with products from different categories. There was no effect of the sound condition on response accuracy (see Appendix B for complementary analyses).

Experiment 3 conceptually replicated and further extended the findings of Experiments 1 and 2 using arbitrary musical (rather than usage-related) sounds. More importantly, the effect emerged even though the associative links between sounds and brands were established within the context of the study in only three exposures (short-term learning). Thus, short-term associations between advertising jingles and specific brands enabled consumers to identify these brands on the shelf more rapidly if they were cued with the jingle.

Experiment 4A – Auditory Facilitation of Brand Search in a Virtual Online Store

Overview

The goal of Experiment 4A was to extend the facilitatory effect of sound on visual brand search to a more realistic online shopping context, where the participants' task was to buy a specific target product in a virtual online store. Critically, and in contrast to Experiments 1-3 and previous studies (e.g., Iordanescu et al. 2008; 2010), Experiment 4A used a single-trial between-participants design in order to eliminate any learning effects that might arise over the repeated trials of a within-participants design and that might drive the effect. We collected the data online to test the robustness of the effect under more realistic, real-life conditions (e.g., heterogeneous screen size, audio quality, reduced attention by participants tested outside the laboratory).

Design and Stimuli

Experiment 4A used a single-factor between-participants design with three conditions: The participants performed a single-trial visual search task, which was accompanied by a target-congruent sound, a distractor-congruent sound, or no sound. The visual stimuli in Experiment 4A

resembled real stimuli that consumers might encounter in retail settings. The visual search display depicted a single page of an online store for musical instruments. The display emulated a product selection page on amazon.com (18 instruments arranged in a 3 × 6 array), which consisted of pictures of actual products sold on amazon.com, and which included product names, prices, and ratings. The target product was a violin located either on the leftmost or rightmost position of the middle row (counterbalanced across participants). A digital piano appeared in the opposite location. The position of the remaining products (various filler instruments) remained constant across the two counterbalanced versions of the search display (see Figure 1D). In order to create target-congruent and distractor-congruent sounds, the first author recorded a short musical passage on two instruments (violin and piano). Except for the difference in musical timbre, the two sounds contained identical musical material (an a-major scale ascending and descending across three octaves with equal loudness, and equal tempo; for a download link see Appendix A).

Procedure

Three-hundred MTurk participants from the United States (117 female, mean age = 33.1 years, *SD* = 10.0, range 19-72 years) participated in the experiment, for which they received 0.6 USD. On-screen instructions informed the participants about the general nature of the task (that they would “buy” a specific product in an assortment by clicking on it). They were prompted to put on their headphones or turn on their speakers (they were told that a sound might or might not accompany the buying task). The participants then saw the name of the target instrument on the screen (“Cremona violin”) and they were instructed to “buy” that instrument as quickly as possible by clicking on it. The visual search task started with an empty white screen and the onset of either the target-congruent or distractor-congruent sound (or else no sound). After 4000 ms, a horizontally centred “Start”-button appeared. In order to proceed to the search display, the participants had to

click the button, which ensured that the starting position of their mouse on the screen was standardized and equidistant from the target product and the distractor product. Once the participants clicked either the target or the distractor product, the search display was terminated. A participant's RT reflected the duration between the click on the "Start" button and the click on one of the two products.

Results and Discussion

Cases with incorrect responses (N = 10) or RTs greater than 8000 ms (N = 24) were excluded from further analyses (altogether 11.3% of the data)[3]. A one-way between-participants ANOVA revealed a statistically significant difference in target localization latencies across the different sound conditions, $F(2, 263) = 4.052$, $p = .018$, $\eta_p^2 = .03$. Planned contrasts confirmed that participants found the target product more rapidly when exposed to a target-congruent sound (M = 3569 ms, SD = 1342 ms) than to no sound (M = 3925 ms, SD = 1334 ms; $p = .077$). Distractor-congruent sounds (M = 4150 ms, SD = 1420 ms) did not significantly influence RTs as compared to the no-sound condition ($p = .279$). Additional analyses of the effect of the sound condition on response accuracy are summarized in Appendix B.

By extending the auditory facilitation effect to a between-participants setting, Experiment 4A ruled out the possibility that the effect identified in previous studies was purely an artifact of the repeated-trial structure. This is an important extension to Experiments 1-3 and previous studies (e.g., Iordanescu et al. 2008, 2010) and suggests that auditory facilitation may also be observed in real-life settings, namely, an online store. Further, Experiment 4A conceptually replicated the auditory facilitation effect for a specific product (violin) within a single product category (musical instruments).

Experiment 4B – Does Visual Clutter Moderate the Crossmodal Facilitation Effect?

Overview

Experiment 4B was designed to replicate the results of Experiment 4A in a laboratory setting, and to examine an important boundary condition of the auditory facilitation effect—the visual complexity of the search display. Note that in Experiments 1-4A, the observed mean difference between the congruent and the no sound conditions increased linearly with the complexity of the visual display—from Experiment 1 (4 products, 14 ms), through Experiment 2 (16 products, 24 ms), Experiment 3 (16 brands in a single category, 35 ms), to Experiment 4A (18 products and contextual information, 356 ms). Building on this observation and our predictions derived from the PoIE, in Experiment 4B, we directly tested the auditory facilitation effect in search displays with lower versus higher visual load/clutter.

Design and Stimuli

A 3 (sound: target-congruent, distractor-congruent, none) × 2 (visual load: low, high) between-participants design was used. For the high visual load condition, Experiment 4B used the same visual stimuli as Experiment 4A. For the low load condition, the top and bottom row of the product array were removed, resulting in a simpler search display that contained a single row of six products. The auditory stimuli were identical to the ones used in Experiment 4A and were administered using closed headphones at an average sound level of 70 dB.

Procedure

One hundred seventy-eight student participants (112 female, 156 aged 18-29 years, 14 aged 30-49 years, 8 with missing age data) at a Norwegian university participated in the experiment. The experiment took five minutes and was part of a 30-minutes data collection session, for which

the participants received the equivalent of approximately 12 USD. The participants arrived at the laboratory in groups of four to ten at a time, where they were seated in individual cubicles approximately 55 cm from the screen (22-inch CRT monitor; 1680 × 1050 pixels resolution, 60 Hz refresh rate). The instructions and the search task were identical to Experiment 4A.

Results and Discussion

As in Experiment 4A, cases with incorrect responses (N = 12) or RTs greater than 8000 ms (N = 13) were excluded from further analyses (altogether 14.0% of the data).

A two-way independent-samples ANOVA revealed an interaction between sound congruency and visual load, $F(2, 147) = 2.936$, $p = .056$, $\eta_p^2 = .04$. In the high visual load condition, participants found the target product more rapidly when exposed to a target-congruent sound (M = 3484 ms, SD = 1230 ms) as compared to no sound (M = 4075 ms, SD = 1885 ms; p = .091). Distractor-congruent sounds (M = 4445 ms, SD = 1748 ms) did not significantly influence RTs as compared to the no-sound condition (p = .328). There were no significant differences between sound conditions in the low visual load condition (M_{Tcongr} = 2134 ms, SD_{Tcongr} = 848 ms, M_{Dcongr} = 1838 ms, SD_{Dcongr} = 656 ms, M_{nosound} = 2259 ms, SD_{nosound} = 878 ms, all p's > .242). Additional analyses of the effect of the sound condition on response accuracy are summarized in Appendix B.

The results of Experiment 4B thus replicate the findings of Experiment 4A while additionally demonstrating that the facilitatory effect of sound on visual brand search is moderated by the complexity or clutter of the search display. This indicates that the facilitation effect follows the principle of inverse effectiveness, which posits that multisensory enhancement is greatest when the constituent unisensory stimuli are ambiguous or noisy. Note that the null effect for the simple search display does not contradict the results of Experiment 1, since the latter used a more powerful repeated-measures design.

General Discussion

Summary

The present study examined how sounds modulate the attentional processes implicated in searching consumer products in a shopping context. Building on insights from the crossmodal attention and the multisensory learning literatures, we proposed that sounds that are semantically associated with a particular brand or product can crossmodally facilitate (in terms of speeding-up) consumers' visual search for that brand or product.

Across five experiments, the semantic congruence between sounds and visual targets affected both how rapidly participants deployed their visual attention towards products and brands and how rapidly they executed a manual response to (i) indicate the presence (vs. absence) of a product and (ii) buy the product. First and last saccade onsets towards the visual target were faster and the number of fixations prior to inspecting the target fewer for sounds that were semantically congruent with the target products as compared to incongruent and no-sound conditions. It would seem, therefore, that a spatially non-informative product usage sound or jingle elicited earlier attention shifts to a target product, perhaps by sensitizing its visual representation.

This effect held when the associations between sounds and visual representation were long-term (formed through a lifetime of experiences with these objects), but also when the associations were short-term (learned within the experiment through three audio-visual exposure trials). Further, the observed crossmodal facilitation effect was robust across different variants of the search task. The effect reliably emerged using different response formats (indicate target position in Experiment 1, target presence vs. absence judgments in Experiments 2 and 3, clicking on the target in Experiments 4A and 4B), using different stimulus onset asynchronies (SOAs; 100 ms visual lag in

Experiment 1, synchronous onset in Experiments 2 and 3, 4000 ms visual lag in Experiments 4A and 4B), and irrespective of whether feedback was given after each trial (Experiment 1) or not (Experiments 2, 3, 4A, 4B). Importantly, the RT findings held under realistic and noisy conditions, such as when using a between-participants experimental design and a virtual online store (Experiments 4A and 4B) and when testing an online sample (in which computer set-ups may vary from one participant to another, Experiment 4A). The RT results of all five experiments suggest that the effect was enhanced in more cluttered visual search environments—this hypothesis was directly tested and confirmed in Experiment 4B.

The present study addresses several alternative explanations of the auditory facilitation effect. By demonstrating the effect in a between-participants setting (Experiments 4A and 4B), we ruled out the possibility that it was merely an artefact of the repeated-trial structure (e.g., induced by learning effects) used in previous studies (e.g., Iordanescu et al. 2008, 2010). In addition, the results indicate that the effect is not driven by a non-specific facilitatory effect of auditory stimulation. Auditory stimuli may have a general alerting effect, speeding-up participants' responses and reducing their search latencies (Petersen and Posner 2012). However, congruent sounds elicited faster search latencies than incongruent sounds in Experiments 1 and 2. Facilitation by (congruent) sounds is thus not due to increased arousal through any auditory stimulation. The design further eliminates an explanation of these crossmodal effects in terms of other features of the auditory stimuli (such as their familiarity or valence), since the presentation of the target products and sounds was counterbalanced across the experimental conditions. Similarly, potential confounds through low-level features of the visual stimuli were controlled for as the visual targets were presented equally often in each experimental condition, and the visual distractors were fully randomized.

Implications and Contributions

Our findings may be interesting to both researchers and practitioners in the context of sensory marketing (Krishna, 2012). Indeed, the present study contributes to the growing body of research on the role of multisensory representations in shaping people's product evaluation. Researchers have, for instance, investigated congruency effects of smell and touch (Krishna, Elder, and Caldara 2010), smell and vision (Lwin, Morrin, and Krishna 2010), smell and sound (Spangenberg, Grohmann, and Sprott 2005), and vision and touch (Littel and Orth 2013). Such research indicates that a product's different sensory attributes, as well as their interconnections (e.g., semantic congruence), shape the way in which we evaluate it. However, only little research in this field has focused on product usage sounds or examined attentional outcomes, let alone real-time attention measures. The present study addresses this gap by demonstrating multisensory effects on attention: Product usage sounds affected people's real-time search performance for visually displayed products. To our knowledge, only three studies in the marketing literature have examined multisensory attention processes (smell and vision: Lwin et al., in press; sound and location: Shen and Sengupta 2014; synesthetically matching sounds: Hagtvedt and Brasel, in press). The present study complements and extends the findings of these earlier studies by focusing on how semantically congruent but *spatially uninformative* product usage sounds guide people's visual attention, and by examining the boundary conditions of such effects.

Multisensory saliency maps and boundary conditions of audiovisual facilitation. The current results provide strong evidence for the view that consumers' brand search is inherently multisensory. Substantiated by eye movement data, the semantic content of a product-related sound reflexively attracted a participant's visual attention to the associated product or brand. We reason that sound can enhance the perceptual saliency of the associated product or brand, thereby guiding

the "spotlight" of visual attention (Posner, Snyder, and Davidson 1980). Our findings echo previous empirical evidence for a multisensory saliency map that weighs not only visual object features but also features from other modalities, such as audition and olfaction (see, e.g., Chen et al. 2013; Seigneuric et al. 2010).

The idea of a multisensory saliency map extends earlier research into the saliency of brands at the point of sale in unisensory settings (Van der Lans et al. 2008). In Van der Lans et al.'s study, consumers used only (one or two basic) *visual* features to localize a target brand. Our study demonstrated that not only visual but also auditory features draw attention to specific products and brands. The multisensory saliency map provides a model for the crossmodal interactions recently observed in marketing research. Complementing earlier work on the attention-modulating effects of *spatially informative* (Shen and Sengupta 2014) and synesthetically matching sounds (Hagtvedt and Brasel, in press), we assessed the role of *semantic* sound-brand associations (as a marketing-relevant factor) in the modulation of audiovisual attention. Such associations seem to guide participants' visual attention from the moment they encounter a search-relevant sound (e.g., as in Experiment 2) and seem to emerge both when searching for generic product classes and for specific brands (Experiment 3). Facilitating brand search (e.g., through sound-brand associations cued by a jingle in the background) is important for brand success given the intense competition among different brands.

Our results further speak to the generality of the effect and its boundary conditions. First, the auditory facilitation effect requires a search goal. While the present study did not directly compare conditions in which a search goal was present versus absent (since the search task would become nonsensical in the absence of a search goal), the results of Experiments 1, 4A, and 4B allow an indirect comparison based on the distractor-congruent sound condition. In these experiments, sounds that were congruent with a distractor (for which no search goal was active) did not slow

down participants' search for the target as compared to the no-sound condition. This clearly suggests that semantically congruent sounds only affect the allocation of visual attention for products if a search goal for the product has already been activated. Such goal dependency stands in contrast to the effects of spatially informative (Shen and Sengupta 2014) and synesthetically congruent (Hagtvedt and Brasel, in press) sounds, which affect choice even in the absence of a search target. If consumers want to buy a product, hearing product-relevant sounds helps them accomplish this goal and buy the product faster. But in the absence of such a goal hearing sounds congruent with a distractor product has no effect. In summary, top-down expectations thus acted as a filter on the attention-capturing effect of semantically-related sounds.

Second, crossmodal audiovisual facilitation was not limited to long-term semantic associations, but was also induced by short-term (three-trial) multisensory learning. To our knowledge, the present research is the first to demonstrate that the effects of short-term multisensory learning on subsequent object recognition (e.g., Liu et al. 2012; Thelen et al. 2015) extend to visual search. Further, semantic links were established in Experiment 3 between initially *unrelated* visual and auditory cues in the learning phase; Hearing the newly associated auditory cues then improved performance in a subsequent visual search task. Thus, while the learning of initially unrelated audiovisual cues does not increase subsequent unisensory recognition (Lehmann and Murray 2005; Thelen and Murray 2012, 2013; Thelen et al. 2015), it does seem to increase multisensory integration (i.e., visual search performance was increased when the learned sound was present in the test phase).

What mechanisms might underlie this short-term learning effect? Previous research has shown that multisensory integration is inherently adaptive, and that humans can rapidly learn to integrate sensory signals (Ernst 2007; van Attefeld, Murray, and Schroeder 2014). Importantly, the flexibility and speed with which people learn new mappings between multisensory stimuli may be

related to the variability of those signals in the environment, such that new mappings are formed more easily between stimuli with low initial mapping certainty (Ernst 2007). One reason why the new brand-jingle mappings in Experiment 3 were acquired easily may thus be that they are stimuli with low mapping certainty (i.e., they lack strong previous mappings). The short-term learning observed here portrays multisensory integration as an adaptive mechanism evolved to (a) structure the substantial variability in sound-object mappings in the multisensory input and (b) resolve many-to-many mappings between sounds and objects. Neurologically, such rapid processes of adaptation may involve short-term plasticity in several brain regions. Indeed, a recent neuroimaging study demonstrated that after 45 minutes of co-exposure to unrelated audio-visual stimuli, exposure to the auditory stimulus activated not only auditory cortex, but also visual cortex V1 (Zangenehpour and Zatorre 2010). Furthermore, short-term crossmodal association learning of unfamiliar audiovisual cues has been shown to induce plastic changes in several brain regions that are implicated in audiovisual integration (Naumer et al. 2009). Short-term plasticity as a consequence of crossmodal association learning was also observed in multisensory neurons in the superior colliculus (Yu, Stein, and Rowland 2009).

Third, the facilitatory effect of congruent sound on product search was stronger in cluttered (rather than simple) product assortments. When the participants' visual system was heavily loaded (i.e., when vision was less informative), the semantic meaning of sound guided the attentional spotlight to relevant aspects of the visual context. This finding is consistent with the PoIE, which postulates that the brain integrates incoming multimodal information as a weighted function of the informativeness of the constituent sensory channels in order to achieve optimal performance (Shams and Kim 2010). Importantly, the present study avoids several of the methodological problems that have been common in studies on the principle of inverse effectiveness, such as regression to the mean and floor/ceiling effects (Holmes 2009). Our research thus demonstrates

that the predictions of the PoIE do not only hold in (visual) detection tasks but that they also have downstream effects on visual (brand) search. In addition, the significance of the PoIE has not been recognized in the sensory marketing literature before, and we argue that perceptual load and stimulus intensity may define important boundary conditions of other recently identified crossmodal effects in marketing (e.g., Henrik and Brasel, in press; Shen and Sengupta 2014). Future research could test such moderation effects. While perceptual load in the current study was manipulated through the number of products within the search display, we speculate that other load manipulations might elicit similar results (e.g. perceptual fluency of the target, Alter and Oppenheimer 2009; target-distractor similarity, Duncan and Humphreys 1989; display layout, Janiszewski 1998).

Product search, user interface design, and choice. In a world where consumers' visual attention is heavily taxed, short-lasting, and often captured by irrelevant stimuli, techniques that facilitate product search provide a competitive advantage for marketers. Focusing on one such technique, our findings suggest that hearing auditory stimuli associated with specific products or brands may help consumers to more rapidly detect the products they are looking for among other, visually distracting products. As the presentation of product-congruent sounds also resulted in fewer fixations to other products before saccading to the target, marketers may use this effect to actively reduce consumers' attention to competitor products. Critically, the present research suggests that minimal (3-trial) exposure to audio-visual stimulus combinations, such as a brand together with its sonic logo, may be sufficient to trigger the crossmodal facilitation effect in consumers. This in turn implies that marketers can expect crossmodal facilitation effects to arise even from weakly associated audio-visual stimuli (e.g., newly learnt product-jingle combinations).

Finally, the fact that incongruent sounds did not generally hinder response execution compared to no-sound conditions implies that marketers run little risk of playing the "wrong" sound.

These effects can be leveraged in different domains. For instance, retailers may activate a purchase goal through either out-of-store or instore (i.e., auditory or visual) advertising and then play usage or brand sounds to guide consumers' attention towards the advertised products. If the speakers playing product-congruent sounds are placed close to the target product, the attention-capturing effects of semantic crossmodal facilitation and spatial crossmodal facilitation (Shen and Sengupta 2014) could be combined. Similar techniques can also be adopted in online stores. In audiovisual advertising, target-congruent sounds may be used to direct the viewers' attention to specific aspects of the visual scene. For example, playing a jingle during a commercial should guide consumers' attention towards the logo or advertised product that is embedded in the plot.

In addition to the implications for marketing, the present research also bears relevance for user interface design. In many instances, using human-machine interfaces (e.g., browsing web pages, driving cars) can be regarded as a form of goal-directed visual search, with different stimuli competing against each other for the limited visual attention of the user. Under these circumstances, auditory stimuli may be used to crossmodally facilitate visual search performance for a specific element in the interface or visual scene. In this way, the user's attention might be guided towards a product in an online shop by the associated jingle or towards a payment button by a "ka-ching" sound. Similarly, semantic audiovisual facilitation may be used to design more effective traffic warning signals. For instance, advanced car interfaces may play meaningful warning sounds (e.g., the sound of a bicycle bell) to increase the saliency of safety-relevant visual objects (e.g., a cyclist) in critical situations, thus complementing previously described techniques of cuing a driver's attention through sound (e.g., Spence and Ho 2012).

Attention could even spill over into decision processes. Product choice, for instance, can be affected by auditory semantic associations (North et al. 1997, 1999). In the seminal study by North and his colleagues (1997), French (vs. German) instrumental background music increased the sales of French (vs. German) wines. However, the mechanism underlying this behavioral priming effect has not been explored. While it is possible that the music directly affected decision-making, our findings highlight the possibility that attention-orienting processes mediate such decision processes. Once the semantic content of the music guides consumers' attention to the target bottles, the enhanced attention may have influenced the associated decision processes (see also Shimojo et al. 2003).

Avenues for Future Research

In the light of the present results, future research could study the effects of audiovisual integration on visual search in real-life shopping environments, for example by using mobile eye-tracking devices. Further, given that not all consumer products have a typical sound linked to their use or consumption, let alone a sonic logo, researchers could test whether verbal auditory cues (i.e., spoken product names) have a similar facilitatory effect on visual search performance. Following-up on this question may also help to inform earlier research reporting improved visual recognition performance for objects when they were preceded by their associated sounds, but not when preceded by their spoken names (Chen and Spence 2011).

While the current research does not focus on the effect of sound on consumer preferences or decision-making, it should nevertheless be noted that visual attention, as captured through eye movements, is a significant predictor of choice (Maughan, Gutnikov, and Stevens 2007; Pieters and Warlop 1999; Shimojo et al. 2003). The duration of visual attention to a product can affect the likelihood of selecting that product (Armel, Beaumel, and Rangel 2008; Krajbich, Armel, and

Rangel 2010). In a binary choice task between familiar options, individuals were more likely to choose the snack food option that they looked at longer, after controlling for pre-existing preferences for each option (Krajbich et al. 2010). Considering the influence of sound on visual attention demonstrated in the current article, it seems worthwhile to consider the role of multisensory influences in the traditional literature on visual attention and choice. While initial efforts in this direction concern the role of spatially informative auditory cues (Shen and Sengupta 2014), it would be particularly interesting to study the effects of semantically congruent sensory cues on the combined allocation of attention and choice.

References

Alais, David and David Burr (2004), "The Ventriloquist Effect Results from near-Optimal Bimodal Integration," *Current Biology,* 14 (3), 257-62.

Alter, Adam L. and Daniel M. Oppenheimer (2009), "Uniting the Tribes of Fluency to Form a Metacognitive Nation," *Personality and Social Psychology Review,* 13 (3), 219-35.

Alvarado, Juan C., J. William Vaughan, Terrence R. Stanford, and Barry E. Stein (2007), "Multisensory Versus Unisensory Integration: Contrasting Modes in the Superior Colliculus," *Journal of Neurophysiology,* 97(5), 3193-3205.

Armel, K. Carrie, Aurelie Beaumel, and Antonio Rangel (2008), "Biasing Simple Choices by Manipulating Relative Visual Attention," *Judgment and Decision Making,* 3 (5), 396-403.

Atalay, A. Selin, H. Onur Bodur, and Dina Rasolofoarison (2012), "Shining in the Center: Central Gaze Cascade Effect on Product Choice," *Journal of Consumer Research,* 39 (4), 848-66.

Baayen, R. H., and Petar Milin (2010), "Analyzing Reaction Times," *International Journal of Psychological Research, 3* (2), 12-28.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily (2013), "Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal," *Journal of Memory and Language,* 68 (3), 255-78.

Bates, Douglas M. and Deepayan Sarkar (2013), "lme4: Linear Mixed-Effects Models Using S4 Classes, R Package Version 1.0-5."

Busse, Laura, Kenneth C. Roberts, Roy E. Crist, Daniel H. Weissman, and Marty G. Woldorff (2005), "The Spread of Attention across Modalities and Space in a Multisensory Object," *Proceedings of the National Academy of Sciences of the United States of America,* 102 (51), 18751-56.

Castelhano, Monica S. and John M. Henderson (2007), "Initial Scene Representations Facilitate Eye Movement Guidance in Visual Search," *Journal of Experimental Psychology: Human Perception and Performance*, 33 (4), 753-63.

Chen, Kepu, Bin Zhou, Shan Chen, Sheng He, and Wen Zhou (2013), "Olfaction Spontaneously Highlights Visual Saliency Map," *Proceedings of the Royal Society B: Biological Sciences*, 280 (1768), 20131729.

Chen, Yi-Chuan and Charles Spence (2010), "When Hearing the Bark Helps to Identify the Dog: Semantically-Congruent Sounds Modulate the Identification of Masked Pictures," *Cognition*, 114 (3), 389-404.

--- (2011), "Crossmodal Semantic Priming by Naturalistic Sounds and Spoken Words Enhances Visual Sensitivity," *Journal of Experimental Psychology: Human Perception and Performance*, 37 (5), 1554-68.

Chen, Yi-Chuan, Su-Ling Yeh, and Charles Spence (2011), "Crossmodal Constraints on Human Perceptual Awareness: Auditory Semantic Modulation of Binocular Rivalry," *Frontiers in Psychology*, 2 (212).

Cousineau, Denis (2005), "Confidence Intervals in Within-Subject Designs: A Simpler Solution to Loftus and Masson's Method," *Tutorials in Quantitative Methods for Psychology*, 1 (1), 42-45.

Deubel, Heiner and Werner X. Schneider (1996), "Saccade Target Selection and Object Recognition: Evidence for a Common Attentional Mechanism," *Vision Research*, 36 (12), 1827-37.

Diaconescu, Andreea O., Claude Alain, and Anthony R. McIntosh (2011), "The Co-Occurrence of Multisensory Facilitation and Cross-Modal Conflict in the Human Brain," *Journal of Neurophysiology*, 106(6), 2896-2909.

Doehrmann, Oliver and Marcus J. Naumer (2008), "Semantics and the Multisensory Brain: How Meaning Modulates Processes of Audio-Visual Integration," *Brain Research*, 1242, 136-50.

Donders, Franciscus Cornelis (1969), "On the Speed of Mental Processes," *Acta Psychologica*, 30, 412-31.

Duncan, John (2001), "An Adaptive Coding Model of Neural Function in Prefrontal Cortex," *Nature Reviews Neuroscience*, 2 (11), 820-29.

Duncan, John and Glyn W. Humphreys (1989), "Visual Search and Stimulus Similarity," *Psychological Review*, 96 (3), 433-58.

Eckstein, Miguel P. (2011), "Visual Search: A Retrospective," *Journal of Vision*, 11 (5), 1-35.

Elder, Ryan S. and Aradhna Krishna (2010), "The Effects of Advertising Copy on Sensory Thoughts and Perceived Taste," *Journal of Consumer Research*, 36 (5), 748-56.

Ferguson, Melissa J. and Vivian Zayas (2009), "Automatic Evaluation," *Current Directions in Psychological Science*, 18 (6), 362-66.

Fiebelkorn, Ian C., John J. Foxe, and Sophie Molholm (2010), "Dual Mechanisms for the Cross-Sensory Spread of Attention: How Much Do Learned Associations Matter?," *Cerebral Cortex*, 20 (1), 109-20.

--- (2012), "Attention and Multisensory Feature Integration," in *The New Handbook of Multisensory Processing*, ed. Barry E. Stein, Cambridge, MA: MIT Press, 383-94.

Gingras, Guy, Benjamin A. Rowland, and Barry E. Stein (2009), "The Differing Impact of Multisensory and Unisensory Integration on Behavior," *Journal of Neuroscience*, 29(15), 4897-902.

Heinze, Georg (2015), "Firth's Bias Reduced Logistic Regression, R Package Version 1.21."

Hagtvedt, Henrik and S. Adam Brasel, (in press), "Crossmodal Communication: Sound Frequency Influences Consumer Responses to Color Lightness," *Journal of Marketing Research*.

Holmes, Nicholas P. (2009), "The Principle of Inverse Effectiveness in Multisensory Integration: Some Statistical Considerations," *Brain Topography*, 21 (3-4), 168-76.

Iordanescu, Lucica, Marcia Grabowecky, Steven Franconeri, Jan Theeuwes, and Satoru Suzuki (2010), "Characteristic Sounds Make You Look at Target Objects More Quickly," *Attention, Perception, & Psychophysics*, 72 (7), 1736-41.

Iordanescu, Lucica, E. Guzman-Martinez, Marcia Grabowecky, and Satoru Suzuki (2008), "Characteristic Sounds Facilitate Visual Search," *Psychonomic Bulletin & Review*, 15 (3), 548-54.

Janiszewski, Chris (1998), "The Influence of Display Characteristics on Visual Exploratory Search Behavior," *Journal of Consumer Research*, 25(3), 290-301.

Jääskeläinen, Iiro P., Jyrki Ahveninen, Mark L. Andermann, John W. Belliveau, Tommi Raij, and Mikko Sams (2011), "Short-Term Plasticity as a Neural Mechanism Supporting Memory and Attentional Functions," *Brain Research*, 1422, 66-81.

Jaeger, T. Florian (2008), "Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and Towards Logit Mixed Models," *Journal of Memory and Language*, 59 (4), 434-46.

Klapetek, Anna, Mary Ngo, and Charles Spence (2012), "Do Crossmodal Correspondences Enhance the Facilitatory Effect of Auditory Cues on Visual Search?," *Attention, Perception, & Psychophysics*, 74 (6), 1154-67.

Krajbich, Ian, Carrie Armel, and Antonio Rangel (2010), "Visual Fixations and the Computation and Comparison of Value in Simple Choice," *Nature Neuroscience*, 13 (10), 1292-98.

Krishna, Aradhna (2012), "An Integrative Review of Sensory Marketing: Engaging the Senses to Affect Perception, Judgment and Behavior," *Journal of Consumer Psychology*, 22 (3), 332-51.

Krishna, Aradhna, Ryan S. Elder, and Cindy Caldara (2010), "Feminine to Smell but Masculine to Touch? Multisensory Congruence and its Effect on the Aesthetic Experience," *Journal of Consumer Psychology*, 20 (4), 410-8.

Kutas, Marta and Kara D. Federmeier (2011), "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event Related Brain Potential (ERP)," *Annual Review of Psychology*, 62, 621-47.

Lehmann, S. and M. M. Murray (2005), "The Role of Multisensory Memories in Unisensory Object Discrimination," *Cognitive Brain Research*, 24 (2), 326-34.

Littel, Sandra and Ulrich Orth, R. (2013), "Effects of Package Visuals and Haptics on Brand Evaluations," *European Journal of Marketing*, 47 (1/2), 198-217.

Liu, Baolin, Guangning Wu, and Xianyao Meng (2012), "Cross-Modal Priming Effect Based on Short-Term Experience of Ecologically Unrelated Audio-Visual Information: An Event-Related Potential Study," *Neuroscience*, 223, 21-27.

Luce, Robert Duncan (1986), *Response Times: Their Role in Inferring Elementary Mental Organization*, New York: Oxford University Press.

Lwin, May O., Maureen Morrin, and Aradhna Krishna (2010), "Exploring the Superadditive Effects of Scent and Pictures on Verbal Recall: An Extension of Dual Coding Theory," *Journal of Consumer Psychology*, 20, 317-26.

Lwin, May O., Maureen Morrin, Chiao Sing Trinetta Chong, and Su Xin Goh (in press), "Odor Semantics and Visual Cues: What We Smell Impacts Where We Look, What We Remember, and What We Want to Buy," *Journal of Behavioral Decision Making*.

Matusz, Pawel J. and Martin Eimer (2011), "Multisensory Enhancement of Attentional Capture in Visual Search," *Psychonomic Bulletin & Review*, 18 (5), 904-09.

Maughan, Lizzie, Sergei Gutnikov, and Rob Stevens (2007), "Like More, Look More. Look More, Like More: The Evidence from Eye-Tracking," *Journal of Brand Management*, 14 (4), 335-42.

Meredith, M. Alex and Barry E. Stein (1986), "Spatial Factors Determine the Activity of Multisensory Neurons in Cat Superior Colliculus," *Brain Research*, 365(2), 350-4.

Molholm, Sophie, Antigona Martinez, Marina Shpaner, and John J. Foxe (2007), "Object ‐ Based Attention Is Multisensory: Co ‐ Activation of an Object's Representations in Ignored Sensory Modalities," *European Journal of Neuroscience*, 26 (2), 499-509.

Murray, Micah M., John J. Foxe, and Glenn R. Wylie (2005), "The Brain Uses Single-Trial Multisensory Memories to Discriminate without Awareness," *NeuroImage*, 27 (2), 473-78.

Naumer, Marcus J., Oliver Doehrmann, Notger G. Müller, Lars Muckli, Jochen Kaiser, and Grit Hein (2009), "Cortical Plasticity of Audio–Visual Object Representations," *Cerebral Cortex*, 19 (7), 1641-53.

Ngo, Mary and Charles Spence (2010), "Auditory, Tactile, and Multisensory Cues Facilitate Search for Dynamic Visual Stimuli," *Attention, Perception, & Psychophysics*, 72 (6), 1654-65.

Noesselt, Toemme, Sascha Tyll, Carsten N. Boehler, Eike Budinger, Hans-Jochen Heinze, and John Driver (2010), "Sound-Induced Enhancement of Low-Intensity Vision: Multisensory Influences on Human Sensory-Specific Cortices and Thalamic Bodies Relate to Perceptual Enhancement of Visual Detection Sensitivity," *Journal of Neuroscience*, 30(41), 13609-23.

North, Adrian C., David J. Hargreaves, and Jennifer McKendrick (1997), "In-Store Music Affects Product Choice," *Nature*, 390 (6656), 132-32.

--- (1999), "The Influence of In-Store Music on Wine Selections," *Journal of Applied Psychology*, 84 (2), 271-76.

Parise, Cesare V. and Charles Spence (2009), "“When Birds of a Feather Flock Together”: Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes," *Plos One*, 4 (5), e5664.

Parrott, Stacey, Emmanuel Guzman-Martinez, Laura Ortega, Marcia Grabowecky, Mark D. Huntington, and Satoru Suzuki (2015), "Direction of Auditory Pitch-Change Influences Visual Search for Slope from Graphs," *Perception*, 44 (7), 764-78.

Petersen, Steven E. and Michael I. Posner (2012), "The Attention System of the Human Brain: 20 Years After," *Annual Review of Neuroscience*, 35, 73-89.

Pieters, Rik and Luk Warlop (1999), "Visual Attention During Brand Choice: The Impact of Time Pressure and Task Motivation," *International Journal of Research in Marketing*, 16 (1), 1-16.

Posner, Michael I., Charles R. Snyder, and Brian J. Davidson (1980), "Attention and the Detection of Signals," *Journal of Experimental Psychology: General*, 109 (2), 160-74.

Quinlan, Philip T. (2003), "Visual Feature Integration Theory: Past, Present, and Future," *Psychological Bulletin*, 129 (5), 643-73.

Rayner, Keith, George W. McConkie, and Susan Ehrlich (1978), "Eye Movements and Integrating Information across Fixations," *Journal of Experimental Psychology: Human Perception and Performance*, 4 (4), 529-44.

Robinson, Jeffrey (1999), *The Manipulators: A Conspiracy to Make Us Buy*, London: Pocket Books.

Ross, Lars A., Dave Saint-Amour, Victoria M. Leavitt, Daniel C. Javitt, and John J. Foxe (2007), "Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments," *Cerebral Cortex*, 17(5), 1147-1153.

Salthouse, Timothy A. and Trey Hedden (2002), "Interpreting Reaction Time Measures in between-Group Comparisons," *Journal of Clinical and Experimental Neuropsychology*, 24 (7), 858-72.

Santangelo, Valerio, Cristy Ho, and Charles Spence (2008), "Capturing Spatial Attention with Multisensory Cues," *Psychonomic Bulletin & Review*, 15 (2), 398-403.

Schneider, Till R., Andreas K. Engel, and Stefan Debener (2008), "Multisensory Identification of Natural Objects in a Two-Way Crossmodal Priming Paradigm," *Experimental Psychology*, 55 (2), 121-32.

Seigneuric, Alix, Karine Durand, Tao Jiang, Jean-Yves Baudouin, and Benoist Schaal (2010), "The Nose Tells It to the Eyes: Crossmodal Associations between Olfaction and Vision," *Perception*, 39 (11), 1541-54.

Shams, Ladan and Robyn Kim (2010), "Crossmodal Influences on Visual Perception," *Physics of Life Reviews*, 7 (3), 269-84.

Shen, Hao and Jaideep Sengupta (2014), "The Crossmodal Effect of Attention on Preferences: Facilitation Versus Impairment," *Journal of Consumer Research*, 40 (5), 885-903.

Shepherd, Martin, John M. Findlay, and Robert J. Hockey (1986), "The Relationship between Eye Movements and Spatial Attention," *The Quarterly Journal of Experimental Psychology*, 38 (3), 475-91.

Shimojo, Shinsuke and Ladan Shams (2001), "Sensory Modalities Are Not Separate Modalities: Plasticity and Interactions," *Current Opinion in Neurobiology*, 11 (4), 505-09.

Shimojo, Shinsuke, Claudiu Simion, Eiko Shimojo, and Christian Scheier (2003), "Gaze Bias Both Reflects and Influences Preference," *Nature Neuroscience*, 6 (12), 1317-22.

Snodgrass, Joan G. and Mary Vanderwart (1980), "A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity," *Journal of Experimental Psychology: Human Learning and Memory*, 6 (2), 174.

Spangenberg, Eric R., Bianca Grohmann, and David E. Sprott (2005), "It's Beginning to Smell (and Sound) a Lot Like Christmas: The Interactive Effects of Ambient Scent and Music in a Retail Setting," *Journal of Business Research*, 58 (11), 1583-89.

Spence, Charles (2010), "Crossmodal Spatial Attention," *Annals of the New York Academy of Sciences (The Year in Cognitive Neuroscience)*, 1191 (1), 182-200.

--- (2013), "Just How Important Is Spatial Coincidence to Multisensory Integration? Evaluating the Spatial Rule," *Annals of the New York Academy of Sciences*, 1296 (1), 31-49.

Spence, Charles and Jon Driver, eds. (2004), *Crossmodal Space and Crossmodal Attention*, Oxford, UK: Oxford University Press.

Spence, Charles and Cristy Ho (2012), *The Multisensory Driver: Implications for Ergonomic Car Interface Design*: Ashgate Publishing, Ltd.

Stein, Barry E. (2012), *The New Handbook of Multisensory Processing*, Cambridge, MA: MIT Press.

Stein, Barry E. and M. Alex Meredith (1993), *The Merging of the Senses*, Cambridge, MA: MIT Press.

Stein, Barry E. and T R. Stanford (2008), "Multisensory Integration: Current Issues From the Perspective of the Single Neuron," *Nature Reviews Neuroscience*, 9(4), 255-266.

Stein, Barry E., Nancy London, Lee K. Wilkinson, and Donald D. Price (1996), "Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis," *Journal of Cognitive Neuroscience*, 8 (6), 497-506.

Suied, Clara, Nicholas Bonneel, and Isabelle Viaud-Delmon (2009) "Integration of Auditory and Visual Information in the Recognition of Realistic Objects," *Experimental Brain Research*, 194 (1), 91-102.

Tabachnick, Barbara G. and Linda S. Fidell (2007), *Using Multivariate Statistics*, Boston: Allyn and Bacon.

Thelen, Antonia and Micah M. Murray (2012), "Determinants of the Efficacy of Single-Trial Multisensory Learning," *Seeing and Perceiving*, 25 (1), 39.

--- (2013), "The Efficacy of Single-Trial Multisensory Memories," *Multisensory Research*, 26 (5), 483-502.

Thelen, Antonia, Durk Talsma, and Micah M. Murray (2015), "Single-Trial Multisensory Memories Affect Later Auditory and Visual Object Discrimination," *Cognition*, 138, 148-160.

Treisman, Anne M. and Garry Gelade (1980), "A Feature-Integration Theory of Attention," *Cognitive Psychology*, 12 (1), 97-136.

van Atteveldt, Nienke, Micah M. Murray, Gregor Thut, and Charles E. Schroeder (2014), "Multisensory Integration: Flexible Use of General Operations," *Neuron*, 81 (6), 1240-53.

Van der Burg, Erik, Christian N. L. Olivers, Adelbert W. Bronkhorst, and Jan Theeuwes (2008), "Pip and Pop: Nonspatial Auditory Signals Improve Spatial Visual Search," *Journal of Experimental Psychology: Human Perception and Performance*, 34 (5), 1053-65.

Van der Burg, Erik, Durk Talsma, Christian N. L. Olivers, Clayton Hickey, and Jan Theeuwes (2011), "Early Multisensory Interactions Affect the Competition among Multiple Visual Objects," *NeuroImage*, 55 (3), 1208-18.

Van der Lans, Ralf, Rik Pieters, and Michel Wedel (2008), "Competitive Brand Salience," *Marketing Science*, 27 (5), 922-31.

Vatakis, Argiro and Charles Spence (2010), "Audiovisual Temporal Integration for Complex Speech, Object-Action, Animal Call, and Musical Stimuli," in *Multisensory Object Perception in the Primate Brain*, ed. M. J. Naumer and J. Kaiser: Springer, 95-121.

Wedel, Michel and Rik Pieters (2007), "A Review of Eye-Tracking Research in Marketing," *Review of Marketing Research*, 4, 123-47.

Wolfe, Jeremy M. (1998), "What Can 1 Million Trials Tell Us About Visual Search?," *Psychological Science*, 9 (1), 33-9.

Yu, Liping, Barry E. Stein, and Benjamin A. Rowland (2009), "Adult Plasticity in Multisensory Neurons: Short-Term Experience-Dependent Changes in the Superior Colliculus," *Journal of Neuroscience*, 29 (50), 15910-22.

Zangenehpour, Shahin and Robert J. Zatorre (2010), "Crossmodal Recruitment of Primary Visual Cortex Following Brief Exposure to Bimodal Audiovisual Stimuli," *Neuropsychologia*, 48 (2), 591-600.

Footnotes

[1] Previous research has shown that multisensory integration usually leads to (super-)additive neuronal and behavioural responses, while unisensory integration effects are mostly subadditive (Alverado, Vaughan, Stanford, & Stein 2007; Gingras, Rowland, & Stein 2009).

[2] According to Diaconescu, Alain, and McIntosh (2011, p. 2896), "[c]ontinued exposure to cross-modal events sets up expectations about what a given object most likely "sounds" like [...]. The binding of familiar auditory and visual signatures is referred to as semantic, multisensory integration." In line with this definition, semantic congruence is established in Experiment 3 by means of the learning task, in which the initially "meaningless" sonic logos become associated with the brands.

[3] Slower RTs likely reflect entirely different psychological processes and/or noncompliance with the experiment instructions to click on the target product as fast as possible. A similar cut-off value (10,000 ms) was used by Van der Lans et al. (2008). The authors assume that it is realistic to complete the search task within 4000 ms. The direction or significance of the results does not change when using lower cut-off values (i.e., 7000 ms, 6000 ms).

Figure 1. (A) Example of search display for Experiment 1; (B) example of search display for Experiment 2; (C) example of search display for Experiment 3; (D) example of search display for Experiments 4A and 4B

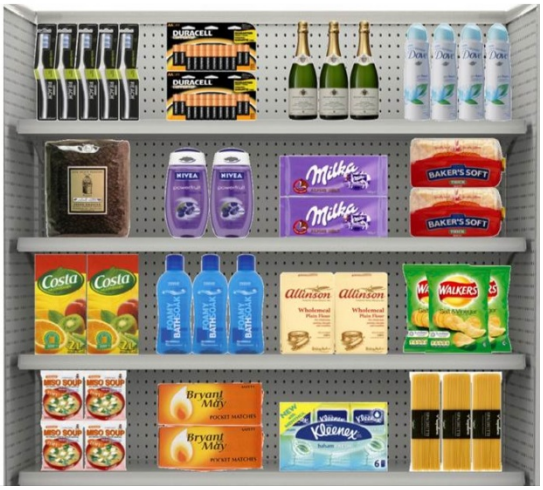
Figure 2. Experiment 1: Visual search target RTs as a function of different sound conditions (target-congruent, no sound, incongruent, distractor congruent), aggregated across participants

Figure 3. Results of Experiment 2: (A) Mean response times; (B) mean onset time of the first saccade to the target; (C) mean onset time of the last saccade to the target; (D) mean ordinal sequence of the first fixation to the target

Figure 1

A

B



C

D

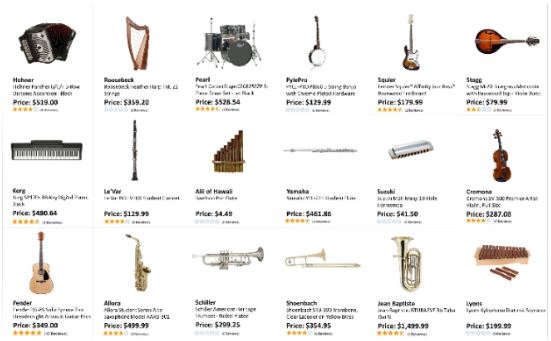
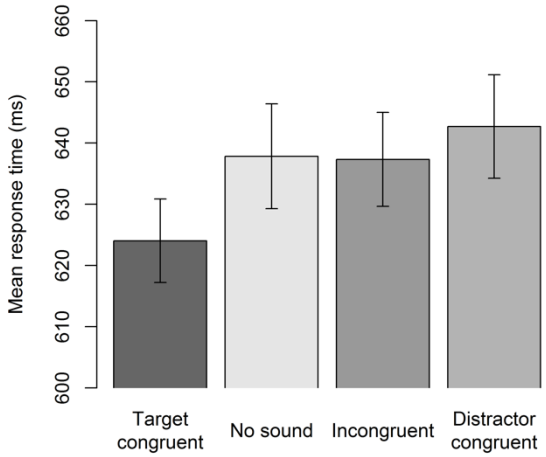
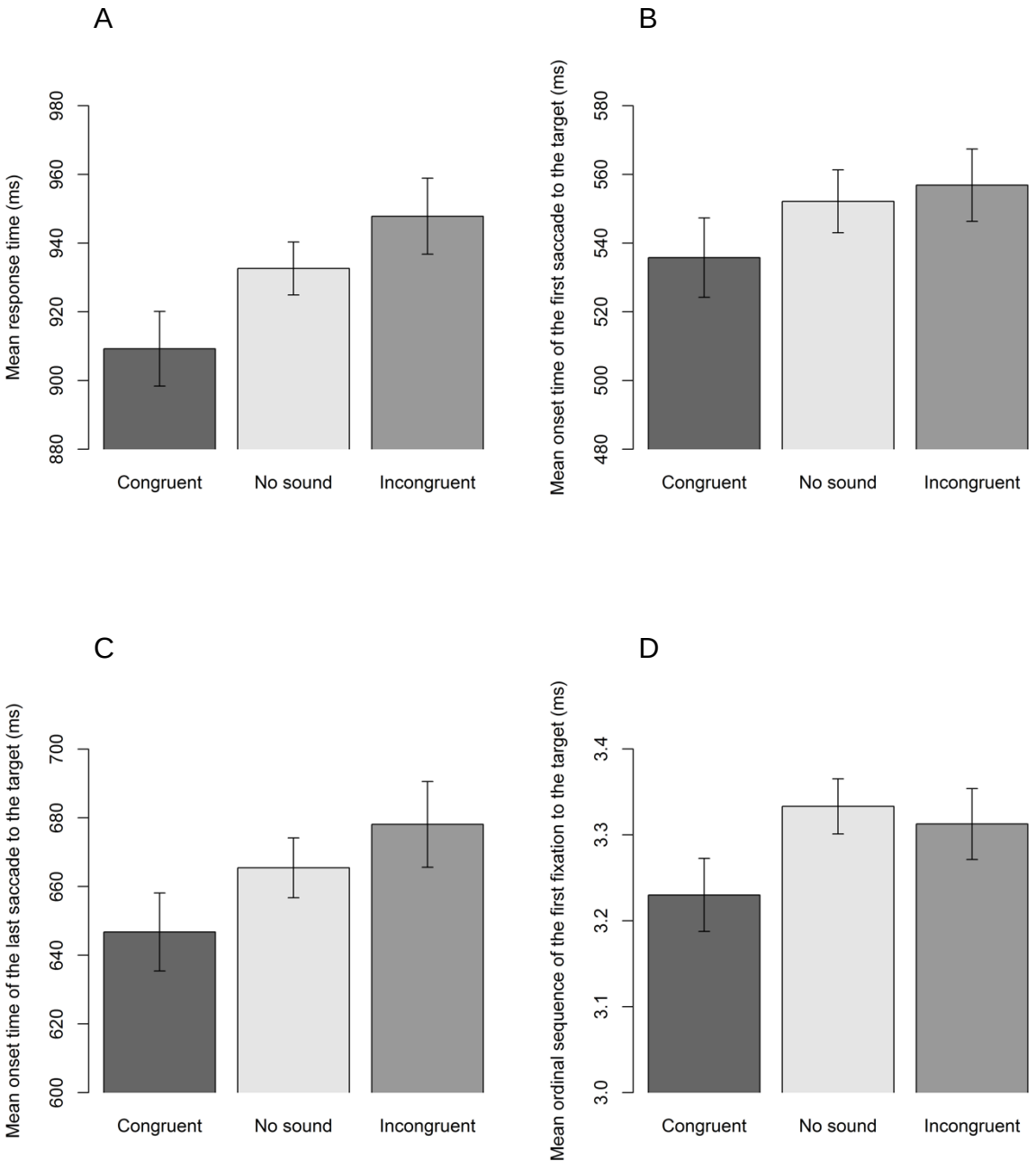


Figure 2



Note.– Error bars represent 95% within-participant confidence intervals (Cousineau 2005).

Figure 3



Note.– Error bars represent 95% within-participant confidence intervals (Cousineau 2005).