Cognition 175 (2018) 101–108

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Questioning the automaticity of audiovisual correspondences

Laura M. Getz[a,b,*], Michael Kubovy[a]

[a] University of Virginia, Department of Psychology, United States
[b] Villanova University, Psychological and Brain Sciences Department, United States

ARTICLE INFO

Keywords:
Cross-modal correspondence
Audiovisual correspondence
Automaticity
Bottom-up processing
Top-down processing

ABSTRACT

An audiovisual correspondence (AVC) refers to an observer's seemingly arbitrary yet consistent matching of sensory features across the two modalities; for example, between an auditory pitch and visual size. Research on AVCs has frequently used a speeded classification procedure in which participants are asked to rapidly classify an image when it is either accompanied by a congruent or an incongruent sound (or vice versa). When, as is typically the case, classification is faster in the presence of a congruent stimulus, researchers have inferred that the AVC is automatic and bottom-up. Such an inference is incomplete because the procedure does not show that the AVC is *not* subject to top-down influences. To remedy this problem, we devised a procedure that allows us to assess the degree of "bottom-up-ness" and "top-down-ness" in the processing of an AVC. We did this in studies of AVCs between pitch and five visual features: size, height, spatial frequency, brightness, and angularity. We find that all the AVCs we studied involve *both* bottom-up *and* top-down processing, thus undermining the prevalent generalization that AVCs are automatic.

1. Introduction

Cross-modal correspondences refer to seemingly arbitrary yet consistent associations across sensory features from different sensory modalities (for reviews, see Marks, 2004; Parise, 2016; Spence, 2011). In the present paper, we focus our attention on *audiovisual* correspondences (AVCs). For example, it has been shown that people readily associate high-pitched tones with smaller objects placed higher in space. We attempt to address the issue of automaticity by creating separate measures of "bottom-up-ness" and "top-down-ness" in our assessment of AVCs between auditory *pitch* and five visual properties: *size*, *height*, *spatial frequency*, *angularity*, and *brightness*.

A majority of past research on AVCs has used a *speeded classification* paradigm. In such experiments, participants classify a multimodal stimulus according to its value on one modality while ignoring the other modality. For instance, they might be asked to report whether a stimulus was large or small while disregarding a concurrent high or low pitch. In this case, size is called the *relevant feature* and pitch is called the *irrelevant feature*. Participants encounter two main types of trials in a typical experiment: (a) on *congruent* trials, the level of the irrelevant feature matches the level of the relevant feature (a *low* pitch with a *large* stimulus); and (b) on *incongruent*trials, the level of the irrelevant feature does not match the level of the relevant feature (a *high* pitch with a *large* stimulus).[1]

Correctly classifying the relevant feature more quickly on congruent than on incongruent trials is treated as evidence that the irrelevant feature affects the processing of the relevant feature in a bottom-up fashion. For example, Evans and Treisman (2010) argue that audiovisual correspondences are "certainly automatic and independent of attention" (p. 10) and Gallace and Spence (2006) conclude that "people cannot help but process auditory information even when it is irrelevant to their visual task" (p. 1200). It is important to note that conclusions regarding automaticity are not limited to the speeded classification paradigm; for example, Parise and Spence (2012) used a speeded implicit association task to show that auditory and visual dimensions are paired together rapidly and automatically.

However, a congruency advantage alone is inadequate to imply a purely automatic, bottom-up effect for three reasons. First, contradictory evidence exists as to the replicability of the congruency advantage. A number of studies, including our own work (see S1 Motivating Experiments) and the work of other researchers (*e.g.*, Heron, Roach, Hanson, McGraw, & Whitaker, 2012; Klein, Brennan, & Gilani, 1987) have failed to show a congruency advantage on speeded detection tasks of various AVCs.

Second, the congruency advantage itself fails to show that AVCs are *immune to top-down influences*. Although the debate regarding top-down influences on perception has centered primarily on visual as opposed to cross-modal perception (*e.g.*, Firestone & Scholl, 2016; Goldstone, de Leeuw, & Landy, 2015; Vetter & Newen, 2014), there is evidence that factors such as the stimulus situation, modality characteristics, and observer

* Corresponding author at: Villanova University, Psychological and Brain Sciences Department, 800 E. Lancaster Ave., Villanova, PA 19085, United States.
E-mail address: laura.getz@villanova.edu (L.M. Getz).

[1] Some experiments, such as Gallace and Spence (2006) include a 'neutral' or unimodal condition as well, where no irrelevant stimulus value occurs (*e.g.*, no sound is played).

https://doi.org/10.1016/j.cognition.2018.02.015
Received 24 August 2017; Received in revised form 11 February 2018; Accepted 13 February 2018
0010-0277/ © 2018 Elsevier B.V. All rights reserved.

Table 1
Consensus mapping for each audiovisual correspondence based on previous studies finding a significant congruency effect.

Visual dimension	High-pitch pairing	Low-pitch pairing	Previous experiments
Size	Small	Large	Evans and Treisman (2010), Gallace and Spence (2006), Mondloch and Maurer (2004), Spector and Maurer (2009)
Height (Elevation)	High	Low	Ben-Artzi and Marks (1995), Evans and Treisman (2010), Melara and O'Brien (1987), Patching and Quinlan (2002)
Spatial frequency	High (Narrow)	Low (Wide)	Evans and Treisman (2010)
Angularity (Sharpness)	Sharp	Rounded	Marks (1987), Maurer et al. (2012), O'Boyle and Tarte (1980), Parise and Spence (2009)
Brightness (Contrast)	Bright	Dark	Marks (1974, 1987), Martino and Marks (1999), Mondloch and Maurer (2004)

processes may affect multimodal perception as well (Chen & Spence, 2017; Welch & Warren, 1980). For example, Klapetek, Ngo, and Spence (2012) conclude that the pitch–brightness AVC operates "at a more strategic (*i.e.*, rather than at an automatic or involuntary) level" (p. 1161). Similarly, others argue that AVCs are influenced by cognitive processes rather than purely the result of perceptual encoding and contend that mappings across sensory cues are highly flexible based on prior experience (Chen & Spence, 2017; Chiou & Rich, 2012; Parise, 2016).

These seemingly contradictory conclusions point to the need to quantify the degree of automaticity in AVCs rather than choosing a side in the bottom-up vs. top-down debate (cf. Spence & Deroy, 2013). This relates to the third problem with previous research, which is that there is little consensus in the literature as to what *automaticity* really means (Moors & De Houwer, 2006; Santangelo & Spence, 2008). Though determining a theoretically and pragmatically appropriate definition is beyond the scope of this paper, we agree with previous researchers who argue that automaticity should be viewed as an umbrella term (*e.g.*, Spence & Deroy, 2013). In our work, we mean automaticity in terms of a bottom-up association between the auditory and visual modalities that exists without the necessity for intentional learning and outside the influence of attention or motivation. To that end, here we report the results of a new paradigm for assessing AVCs, which we see as a first step in answering what Spence and Deroy (2013) call "a challenge of the first order" (p. 257); namely, investigating the *degree* of "bottom-up-ness" and "top-down-ness" present in a variety of AVCs.

To achieve this goal, we created a modified version of the speeded classification task, where we manipulated the stimulus-response mapping included in the instructions to participants. This allowed us to determine whether participants could pair the corresponding dimensions in either direction without a loss in reaction time (*e.g.*, pairing high pitch with small shapes vs. pairing high pitch with large shapes). This is in line with previous work showing the importance of instructions given to participants in showing that AVC processing is at least partially goal-dependent (Chiou & Rich, 2012; Klapetek et al., 2012).

In our experiments, we jointly manipulated congruence and compatibility. We defined *congruence* according to the consensus mapping of pitch onto the visual property manipulated in that study (see Table 1). For example, in the case of the pitch–size correspondence, we consider small size to be congruent with high pitch and large size congruent with low pitch. We defined *compatibility* in reference to the instructions given on each block of trials: (a) during *compatible* blocks, the instructions pair congruent endpoints of the auditory and visual dimensions (*e.g.*, participants are told to select either the large shape/low pitch or small shape/high pitch), whereas (b) during *incompatible* blocks, the instructions are reversed and now pair incongruent endpoints (participants are told to select either the large shape/high pitch or small shape/lower pitch).

This procedure allowed us to create measures of "bottom-up-ness" (BU) and "top-down-ness" (TD) based on the participants' response speed to the various conditions. Fig. 1 shows several hypothetical outcomes for experiments using our methodology. "Bottom-up-ness" refers to the ease with which participants completed the task on compatible as opposed to incompatible blocks. Slower response speeds on incompatible blocks are evidence that it is hard to pair together the incongruent dimensions and thus show a stronger bottom-up association. Fig. 1a represents the case of a strong bottom-up effect with low top-down influence: participants are slower when given instructions asking them to pair the dimensions in the non-consensus direction on incompatible blocks. "Top-down-ness" refers to how well participants followed the instructions on compatible and incompatible blocks. If participants can just as quickly and accurately pair the dimensions in the opposite, non-consensus direction (*i.e.*, on incompatible blocks), this is evidence of a stronger top-down, goal-directed influence of the instructions. Fig. 1c represents the case of high top-down influence with little evidence of a bottom-up effect: the instructions to invert the association are followed with no cost in reaction time.

Fig. 1b represents an intermediate case on both the bottom-up and top-down dimensions. In these three cases, there is a congruency advantage on compatible blocks (showing a successful replication) and an incongruency advantage on incompatible blocks (showing a successful manipulation). Though less likely, it is not inevitable that the results will show a clear trade-off between bottom-up and top-down effects. Fig. 1d represents a case where the instructions have no effect (showing a failed manipulation): participants are always faster to respond to the congruent dimensions even when the instructions ask them to pair the dimensions in the opposite direction. Fig. 1e represents a case where the auditory and visual dimensions pair together more naturally in the *opposite* direction from what has traditionally been shown, thus showing a failure to replicate previous studies.

Having separate measures for bottom-up associations and top-down influence grants us a more direct way to *quantify* the degree of automaticity present in each correspondence, thus meaningfully adding to the debate on the cognitive penetrability of audiovisual perception.

2. Method

2.1. Participants

We recruited 179 University of Virginia undergraduates with normal or corrected-to-normal vision and normal hearing to participate in exchange for credit in an introductory psychology course ($n = 31$ for size; $n = 24$ for height; $n = 36$ for spatial frequency; $n = 38$ for angularity; $n = 50$ for brightness).

2.2. Stimuli

2.2.1. Auditory pitches

All sounds were sine tones with 10 ms rise and decay times. We used three frequency intervals: 'large' (300 Hz vs. 4500 Hz), 'octave' (440 Hz vs. 880 Hz), and 'M3' (a major third, 500 Hz vs. 630 Hz). The octave and M3 intervals were chosen to determine whether the effect previously found with the large interval generalized to smaller pitch differences.[2] We were not able to accurately measure the dB level of the sounds used, but they were manually adjusted to be equally loud across the various

[2] Smaller pitch differences (600–680 Hz, 460–820 Hz, 320–960 Hz, and 180–1100 Hz) have been used to investigate the pitch–height correspondence only (Ben-Artzi & Marks, 1995; Patching & Quinlan, 2002).

Download English Version:

https://daneshyari.com/en/article/7285358

Download Persian Version:

https://daneshyari.com/article/7285358

Daneshyari.com