Acta Psychologica 146 (2014) 73–83

Contents lists available at ScienceDirect

Acta Psychologica

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

Conceptual priming with pictures and environmental sounds

Yongju Kim [a], Anne Marie Porter [b], Paula Goolkasian [b,*]

[a] Korea Military Academy, Republic of Korea
[b] University of North Carolina at Charlotte, United States

ARTICLE INFO

Article history:
Received 30 March 2013
Received in revised form 9 December 2013
Accepted 10 December 2013
Available online 9 January 2014

Keywords:
Cross-modal
Conceptual priming
Categorization task

ABSTRACT

A series of experiments was conducted to examine conceptual priming within and across modalities with pictures and environmental sounds. In Experiment 1, we developed a new multimodal stimulus set consisting of two picture and sound exemplars that represented 80 object items. In Experiments 2, we investigated whether categorization of the stimulus items would be facilitated by picture and environmental sound primes that were derived from different exemplars of the target items; and in Experiments 3 and 4, we tested the additional influence on priming when trials were consolidated within a target modality and the inter stimulus interval was lengthened. The results demonstrated that target categorization was facilitated by the advanced presentation of conceptually related exemplars, but there were differences in effectiveness when pictures and sounds appeared as primes.

1. Introduction

This research examined a conceptual priming effect across modalities as well as within the same modality with pictures and environmental sounds. With perceptual priming the target facilitation results from the advanced presentation of some aspect of the physical stimulus, whereas in conceptual priming the perceptual format of the original stimulus is not as important because the priming is indirect and based on semantic memory. The primary concern in our research is whether categorization of pictures and environmental sounds would be similarly facilitated by the advance presentation of stimuli that were derived from different exemplars of the same items as the targets: would the sound of a dog barking facilitate categorization of a picture of a dog as a man-made or a natural thing in the same way as another picture of a dog? Would a picture of a dog facilitate categorization of a sound "woof" in the same way as the sound of another dog barking "woof"? The secondary interest is the comparison of priming effects due to the picture and environmental sound primes.

Perception is naturally accustomed to multisensory experiences that include sensory interactions and a mechanism to bind them into a coherent perceptual representation. A classic example of a binding mechanism is an object file, introduced by Kahneman, Treisman, and Gibbs (1992); and updated by Hommel (1998, 2004, 2005). They explained how features are bound to objects and enriched by object-related knowledge from long-term memory. Work by Zmigrod, Spapé, and Hommel (2009) suggests that auditory and visual features can be integrated and bound with each other and with their associated response. Their data show that feature integration operates across perceptual domains. Multimodal sensory integration occurs with an event file, which is a network of bindings that link codes of the salient features of a perceptual event. Repeated encounters with an event file produce retrieval of the event file in such a way that it may facilitate performance, if the same stimulus event is experienced; or it may interfere with the creation of new event files that share some but not all of the same features.

In the past decade, diverse experimental paradigms have been developed to investigate the structure of the cognitive network and the interactions between perceptual modalities. The semantic priming paradigm is one of the most widely used paradigms to investigate the organization of the cognitive network (Smith, Meiran, & Besner, 2000). Semantic priming refers to an improvement in responding to a stimulus, such as a word or a picture, when it is preceded by a semantically related rather than an unrelated stimulus (McNamara, 2005). Although most priming studies investigated verbal priming within a single modality, there are a few studies examining priming across sensory modalities (Easton, Greene, & Srinivas, 1997; Schneider, Engel, & Debener, 2008). Stimuli used in priming studies have also expanded from words to pictures, songs, and environmental sounds (Chen & Spence, 2010; Johnson & Halpern, 2012).

Recently, a few studies investigated whether environmental sound identification operates in a similar way to word identification across modalities as well as within a modality (Chen & Spence, 2011; Chiu & Schacter, 1995; Orgs, Lange, Dombrowski, & Heil, 2007; Schirmer, Soh, Penney, & Wyse, 2011; Stuart & Jones, 1995). For example, Chiu and Schacter (1995) found that prior encoding of brief sounds and their associated names facilitated identification of sounds while presentation of the name alone did not. Stuart and Jones (1995) found an environmental

* Corresponding author at: Department of Psychology, UNC Charlotte, 9201, University City Blvd, Charlotte, NC 28223, United States.
E-mail address: pagoolka@uncc.edu (P. Goolkasian).

0001-6918/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.12.006

sound facilitation effect with presentation of a sound and a similar sound from the same category as one of the pre-test sounds. Schirmer et al. (2011) provided evidence for a priming effect of environmental sounds by using event-related potentials, and their data suggested that, like speech, environmental sounds are processed not only perceptually but also conceptually.

More recently, Chen and Spence (2011) hypothesized that naturalistic sounds and pictures access semantic representations automatically while spoken and written words access a corresponding lexical representation. Essentially, their work extends Glaser and Glaser's (1989) picture/word model to sounds. Support for Chen and Spence's multisensory framework is provided by a series of studies that show an enhancement to picture detection when associated sounds are presented at least 346 ms prior to the presentation of a picture target. In these studies, priming effects with sounds are not evident with stimulus onset asynchronies shorter than 346 ms because auditory stimuli evolve over time. So, sufficient processing time is required to influence target detection in contrast to pictures (whose attributes appear all at once).

By investigating cross modal effects, our research builds on previous work in repetition priming (Chen & Spence, 2011; Schneider et al., 2008). Schneider et al. (2008) also used multimodal targets, but they did so in separate experiments; and their prime and target pairs were either congruent (in which case the prime was the same stimulus as the target or the same item but presented in a different modality) or incongruent (different item in either the same or different modality). In our experiments, the members of the prime/target pair were always physically different representations. For the unimodal pairs, the prime and target stimuli were different exemplars of the same item (e.g., pictures and sounds "woof" from two different dogs), while the cross modal pairs included a picture and sound representation from the same item. We did this to preclude perceptual priming effects from occurring when the unimodal pairs were used and to be sure that any difference between the unimodal and cross modal pairs were due primarily to access to semantic memory. Also, the priming effect was measured by comparing the advance presentation of picture and sound primes to neutral primes rather than incongruent primes. This way we could remove any effects that may have arisen from response interference. The neutral primes that we used were either abstract pictures or tones. Lastly, our study differed from Schneider et al. (2008) in the nature of the categorization task. They asked participants to indicate whether the target would fit in a shoebox—a size judgment task that may have required participants to visualize the target item, or may have been easier with picture rather than sound representations as targets. With categorization of the targets as man-made or natural, there is no reason to suggest any advantage when either picture or environmental targets were presented.

Our expectation, based on Chen and Spence's (2011) multisensory framework and evidence of auditory and visual feature integration into Hommel's (2004, 2005) event file, was that both picture and environmental sounds would be equally effective as conceptual primes as long as participants were provided sufficient time to process the sound primes prior to presentation of the target stimulus.

2. Experiment 1

Before we could investigate cross modal priming effects, we needed to develop a set of items that had picture and sound representations from the categories of man-made and natural things and pretest each representation to make sure that it was recognizable by the participants in our sample. Our stimulus set included two exemplars for each picture and sound item. Experiment 1 was conducted to test participant responses to each of the stimuli in the pool. Participants named the items and rated them on a 5-point scale to indicate ease of identification. Reaction times (RTs) were also collected.

2.1. Method

2.1.1. Participants

The participants were 15 (87% female) undergraduate students from the University of North Carolina at Charlotte who were at least 18 years of age ($M_{age} = 19$, $SD_{age} = 4.36$), spoke English as their primary language, had normal hearing and vision (or corrected-to-normal vision) with no history of auditory or visual impairment. They participated to obtain extra credit points toward their psychology class grade.

2.1.2. Apparatus and materials

The stimulus set consisted of 448 stimuli, which were generated by selecting 112 concepts from the following superordinate categories—man-made items and natural things, and representing each with two different pictures and two different digitized environmental sound exemplars. The pictures and digitized sounds were selected from databases and clipart files to represent common and easily recognizable environmental sounds and pictures. There were an equal number of items from each of the categories.

Sound files were 32-bit stereo WAV files (sampling rate: 22,050 Hz) taken from Marcell, Borella, Greene, Kerr, and Rogers's (2000) list of 120, and also from the Internet (http://www.freesound.org). The sounds were edited in Audacity 1.2.5 to a length of 750 ms for one exemplar of each item and 1 s for the other exemplar. These lengths were used because they were the exposure durations for the prime and target events in the follow-up experiments and we wanted to test the participant's reactions to the stimuli under conditions that would be similar to the ones used in the priming studies. Sound intensities were adjusted and delivered binaurally through headphones (Labtec Elite 825) at approximately was 65 dB SPL.

Pictures were jpeg files resized using Adobe Photoshop to 4 × 4 cm. They were selected from clip art, the Internet, and normed lists (Bonin, Peereman, Malardier, Meot, & Chalard, 2003; Rossion & Pourtois, 2004). As with the sounds, the pictures were programmed so that one exemplar of each item would be presented for 750 ms and the other for 1 s.

2.1.3. Procedure

Participants sat 30 cm from the computer screen in a well-lighted room, wore stereo headphones, and were run individually in 45-minute sessions. During each trial a fixation point appeared for 1.5 s followed by a picture or a digitized sound. Participants were asked to identify the stimulus on each trial by typing a name in the box at the bottom of the screen, and then rate how difficult it was to name the stimulus by selecting a number from a 5-point numerical scale (where 1 = easy and 5 = difficult).

There were four blocks of trials, with each consisting of a random arrangement of the 112 items. The blocks were presented in the following sequence—pictures presented for 750 ms, sounds presented for 750 ms, picture presented for 1 s followed by sounds presented for 1 s. An instruction screen preceded each block of trials, which informed the participants about the modality of the upcoming stimulus items and provided an opportunity for a rest period.

The stimuli were presented on an iMac computer with a 20″ flat screen. Stimulus presentation and data collection were controlled by SuperLab 4.5. Participant responses and reaction times (RTs) to each of the 448 trials were automatically recorded in a data file. RTs in the naming task were measured from the initial presentation of the stimulus until the first response keystroke.

2.2. Results

Responses to each item were summed across the participants and ranked based on 3 measures—the percent of participants that were able to correctly identify the stimulus (descending order), the mean rating of naming ease (ascending order), and RT to naming each item (ascending order). In scoring the accuracy of the identification response,