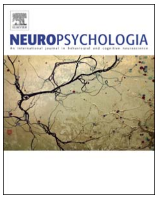
Neuropsychologia 112 (2018) 19–30

Contents lists available at ScienceDirect

Neuropsychologia

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



Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation

Kelly McCormick[a,c], Simon Lacey[a], Randall Stilla[a], Lynne C. Nygaard[c], K. Sathian[a,b,c,d,*,1]

[a] Department of Neurology, Emory University, Atlanta, GA 30322, USA
[b] Department of Rehabilitation Medicine, Emory University, Atlanta, GA 30322, USA
[c] Department of Psychology, Emory University, Atlanta, GA 30322, USA
[d] Center for Visual and Neurocognitive Rehabilitation, Atlanta VAMC, Decatur, GA 30033, USA

ARTICLE INFO

Keywords:
Multisensory
Magnitude
Semantic
Congruency effect
Attention
Working memory

ABSTRACT

Crossmodal correspondences refer to associations between otherwise unrelated stimulus features in different sensory modalities. For example, high and low auditory pitches are associated with high and low visuospatial elevation, respectively. The neural mechanisms underlying crossmodal correspondences are currently unknown. Here, we used functional magnetic resonance imaging (fMRI) to investigate the neural basis of the pitch-elevation correspondence. Pitch-elevation congruency effects were observed bilaterally in the inferior frontal and insular cortex, the right frontal eye field and right inferior parietal cortex. Independent functional localizers failed to provide strong evidence for any of three proposed mechanisms for crossmodal correspondences: semantic mediation, magnitude estimation, and multisensory integration. Instead, pitch-elevation congruency effects overlapped with areas selective for visually presented non-word strings relative to sentences, and with regions sensitive to audiovisual asynchrony. Taken together with the prior literature, the observed congruency effects are most consistent with mediation by multisensory attention.

1. Introduction

Crossmodal correspondences are near-universally experienced associations between apparently arbitrary stimulus features in different senses (Spence, 2011). For example, large and small visual size are consistently associated with low- and high-pitched sounds, respectively (Gallace and Spence, 2006; Evans and Treisman, 2010); and auditorily presented pseudowords, e.g., 'takete' and 'maluma', with pointed and rounded visual shapes, respectively (Köhler, 1929, 1947). A particularly well-known example of a crossmodal correspondence is that in which high and low auditory pitch are associated with high and low visuospatial elevation, respectively (e.g., Bernstein and Edelstein, 1971; Ben-Artzi and Marks, 1995; Evans and Treisman, 2010; Lacey et al., 2016; Jamal et al., 2017). Crossmodal correspondences often occur between stimulus properties that are correlated in nature. Thus, they could render information processing more efficient and facilitate integrating sensory data into unified representations (Spence, 2011). However, in some cases, crossmodal correspondences may lead to false cues: for example, although formant frequencies of animal vocalizations are inversely related to body size, many species can make atypically low sounds as a defensive strategy to exaggerate their perceived size (Fitch, 2000). The neural basis for crossmodal correspondences is unknown. Here, we focus on the audiovisual pitch-elevation correspondence and its relation to three postulated mechanisms: semantic processing, magnitude estimation, and multisensory integration.

One hypothesis for the pitch-elevation correspondence is that it may be semantically mediated because, at least in Western culture, the words 'high' and 'low' can describe both elevation and pitch (Spence, 2011; Walker et al., 2012). Some studies support this suggestion of polysemy: for instance, Shor (1975) reported Stroop interference between the auditorily presented words "high" or "low" and tones that were high- or low-pitched. In addition, these words, in relation to pitch

---

Abbreviations: **Directional:** a, anterior; med, medial; p, posterior; v, ventral; **Anatomical:** AG, angular gyrus; AOS, anterior occipital sulcus; calcS, calcarine sulcus; CS, central sulcus; cingG, cingulate gyrus; cingS, cingulate sulcus; collatS, collateral sulcus; FG, fusiform gyrus; FO, frontal operculum; Ins, insula; IOG, inferior occipital gyrus; IOS, intra-occipital sulcus; IPS, intraparietal sulcus; ITG, inferior temporal gyrus; ITS, inferior temporal sulcus; LG, lingual gyrus; MFG, middle frontal gyrus; MOG, middle occipital gyrus; OG, orbital gyrus; poCG, postcentral gyrus; po, pars opercularis of Broca's area; poCS, postcentral sulcus; POF, parieto-occipital fissure; preCS, precentral sulcus; preCG, precentral gyrus; precun, precuneus; preSMA, pre-supplementary motor area; pt, pars triangularis of Broca's area; SFG, superior frontal gyrus; SFS, superior frontal sulcus; SMG, supramarginal gyrus; SOG, superior occipital gyrus; SPG, superior parietal gyrus; STS, superior temporal sulcus. All other abbreviations are as in the main text.

[*] Correspondence to: Department of Neurology, Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, PA 17033-0859, USA.
E-mail address: ksathian@pennstatehealth.psu.edu (K. Sathian).

[1] Present address: Department of Neurology, Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey, PA 17033-0859, USA.

https://doi.org/10.1016/j.neuropsychologia.2018.02.029
Received 25 August 2017; Received in revised form 22 February 2018; Accepted 26 February 2018

and elevation, generated Garner interference and congruency effects for all combinations of these high/low pairs of words, tones and elevations (Ben-Artzi and Marks, 1999). Such polysemic mediation might ultimately reflect involvement of perceptual simulations evoked by the presented words, as proposed by grounded cognition accounts (e.g. Barsalou, 2008). However, other studies argue against the semantic mediation hypothesis: some non-Western cultures show the pitch-elevation correspondence despite not using spatial language to describe auditory pitch (e.g., Parkinson et al., 2012), and pre-linguistic infants are sensitive to the pitch-elevation correspondence (Walker et al., 2010; Dolscheid et al., 2014). Nonetheless, the hypothesis of semantic mediation of at least some crossmodal correspondences bears testing. If this explanation is valid, we might expect to see activity in the left hemisphere lexical-semantic network (Fedorenko et al., 2010, 2011) during processing of the pitch-elevation correspondence.

Alternatively, the pitch-elevation correspondence may arise from statistical regularities in the natural environment, given that higher-pitched sounds tend to emanate from higher spatial locations than lower-pitched sounds, which tend to occur at low spatial locations (Cabrera and Morimoto, 2007; Parise et al., 2014). This may be related to the strong tendency for vertical localizations of sounds to be strongly biased by their frequency (Cabrera and Morimoto, 2007; Roffler and Butler, 1968; Parise et al., 2014; Pratt, 1930). Thus, the crossmodal correspondence of auditory pitch and visual elevation might be related to multisensory integration based on the crossmodal co-occurrence of sounds and their visualized sources. In support of this view is our recent finding that the crossmodal pitch-elevation correspondence interacts strongly with auditory elevation (Jamal et al., 2017). This suggests that the pitch-elevation correspondence might have its roots in multisensory integration of these naturally occurring signals. If so, neural activity underlying this crossmodal correspondence might co-localize with activity related to multisensory integration, e.g. in the superior temporal sulcus (STS) when audiovisual synchrony (Beauchamp, 2005a, 2005b; van Atteveldt et al., 2007; Stevenson et al., 2010; Marchant et al., 2012; Noesselt et al., 2012; Erickson et al., 2014) or audiovisual identity (Sestieri et al., 2006; Erickson et al., 2014) are manipulated, or in regions such as the intraparietal sulcus (IPS) when audiovisual spatial congruency is manipulated (Sestieri et al., 2006).

Finally, certain crossmodal correspondences may have their basis in representations of magnitude (Lourenco and Longo, 2011). On this account, crossmodal correspondences could arise because the features involved, for example, size and loudness, involve polar dimensions of magnitude where one end is ‘more than’ the other (Smith and Sera, 1992). Thus, high pitch and high elevation may be associated because they are both on the same end of a polar dimension. In this case, we would expect to see activity related to the crossmodal pitch-elevation correspondence in the IPS, an area involved in processing both numerical and non-numerical (e.g., luminance) magnitude (Sathian et al., 1999; Eger et al., 2003; Walsh, 2003; Pinel et al., 2004; Piazza et al., 2004, 2007; Sokolowksi et al., 2017).

The approach we took in the present study was to use functional magnetic resonance imaging (fMRI) to investigate cerebral cortical localization of congruency effects related to the crossmodal correspondence between auditory pitch and visual elevation. In order to test the relevance of the proposed mechanisms outlined above, we conducted three independent localizers in the same individuals: a semantic localizer, a magnitude estimation localizer, and a multisensory synchrony localizer.

2. Methods

2.1. Participants

Twenty participants took part in this study but two were later excluded for excessive movement in the scanner (> 1.5 mm), leaving a final sample of 18 (9 male, 9 female; mean age 24 years, 9 months). All were right-handed based on the validated subset of the Edinburgh handedness inventory (Raczkowski et al., 1974) and reported normal hearing and normal, or corrected-to-normal, vision. All participants gave informed consent and were compensated for their time. All procedures were approved by the Emory University Institutional Review Board.

2.2. Procedures

2.2.1. General

Participants performed the pitch-elevation congruency scans first, and then underwent three localizer scans to test potential mechanisms underlying the pitch-elevation correspondence. After these scans, they performed a behavioral task to determine the strength of their crossmodal pitch-elevation correspondence. This fixed order was followed to avoid potential priming effects of the localizer and behavioral tasks on the pitch-elevation scans. The order of the localizer tasks was also fixed, progressing from the one perceived as most difficult to the easiest: participants did the magnitude estimation localizer first, then the temporal synchrony localizer, and finally the semantic localizer. Each localizer comprised two runs with a fixed stimulus order; the order of runs was counterbalanced across participants. Nine out of 18 participants completed the pitch-elevation and localizer scans in a single session. The other 9 took part in a pitch-size study as well as the current pitch-elevation study; for these participants, experimental and localizer scans were performed in separate sessions (the experimental scans were first in all cases) and the inter-session interval was approximately 1–2 days. All experiments were presented via Presentation software (Neurobehavioral Systems Inc., Albany CA) which allowed synchronization of scan acquisition with experiments and also recorded responses and response times (RTs). Behavioral data were analyzed in IBM SPSS v23 (IBM Corporation, Armonk NY) and effect sizes (Cohen's d) were calculated using the online tool provided by Lenhard and Lenhard (2016); for ease of comparison between t and F statistics, the partial η^2 values provided by SPSS for effect sizes in ANOVA were transformed to Cohen's d using the same tool.

2.2.2. Pitch-elevation fMRI task

The auditory stimuli were low- or high-pitched pure tones (180 Hz and 1440 Hz respectively) of 200 ms duration with a 20 ms on/off ramp. The visual stimulus was a gray circle (RGB values 240, 240, 240) subtending approximately 1° of visual angle with its center approximately 4.2° above (high) or below (low) a central fixation cross. These stimuli were combined to form audiovisual triplets of 1000 ms duration (200 ms on, 200 ms off), comprising three repetitions of identical stimuli that were either congruent (high pitch/high elevation or low pitch/low elevation) or incongruent (high pitch/low elevation or low pitch/high elevation: Fig. 1a) with respect to the crossmodal pitch-elevation correspondence. A mirror angled over the head coil enabled participants to see the visual stimuli projected onto a screen placed in the rear magnet aperture. Auditory stimuli were presented via scanner-compatible headphones.

Because perceived loudness varies with frequency (Moore, 2012; Suzuki and Takeshima, 2004), we matched the high- and low-pitched stimuli for perceived loudness. Once positioned in the scanner and fitted with earplugs and headphones, each participant listened to the high-pitched tone at a range of amplitudes and selected the loudest tone that was still comfortable. The selected high-pitched tone was then compared to a range of low-pitched tones similarly varying in amplitude; participants selected the low-pitched tone that they perceived as matching the high-pitched tone in loudness. Participants chose high-pitched tones that ranged from approximately 95 to 102 dB SPL and low-pitched tones ranging from approximately 85 to 92 dB SPL at the headphones, before the 33 dB noise reduction by the earplugs. The high-pitched tones were, on average, 10 dB SPL greater in intensity than the low-pitched tones, but were judged equally loud by participants;