RESEARCH PAPER

Acta Neurobiol Exp 2024, 84: 136–152
DOI: 10.55782/ane-2024-2439

Dissociable neuronal mechanism for different crossmodal correspondence effects in humans

Carina Jaap* and Michael Rose

NeuroImage Nord, Department for Systems Neuroscience, University Medical Center Hamburg Eppendorf, Hamburg, Germany
**Email: c.jaap@uke.de*

Crossmodal correspondences (CMCs) refer to associations between seemingly arbitrary stimulus features in different sensory modalities. Pitch-size correspondences refer to the strong association of e.g., small objects with high pitches. Pitch-elevation correspondences refer to the strong association of e.g., visuospatial elevated objects with high pitches. We used functional magnetic resonance imaging (fMRI) to study the neural components, which underlie the CMCs in pitch-size and spatial pitch-elevation. This study focuses on answering the question of whether or not different CMCs are driven by similar neural mechanisms. The comparison of congruent against incongruent trials allows the estimation of CMC effects across different CMCs. The analysis of the measured neural activity in different CMCs strongly pointed toward different mechanisms which are involved in the processing of pitch-size and pitch-elevation correspondences. Differential, whole brain effects were observed within the superior parietal lobule (SPL), cerebellum and Heschls' gyrus (HG). Further, the angular gyrus (AnG), the intraparietal sulcus (IPS) and anterior cingulate cortex (ACC) were engaged in processing the CMCs but showed different effects for processing congruent compared to incongruent stimulus presentations. Within pitch-size significant effects in the AnG and ACC were found for congruent stimulus presentations whereas for pitch-elevation, significant effects in the ACC and IPS were found for incongruent stimulus presentations. In summary, the present results indicated differential neural processing in different simple audio-visual CMCs.

Key words: fMRI, crossmodal correspondence, multisensory, audio-visual, spatial, attention, perception

INTRODUCTION

Crossmodal correspondences (CMCs) refer to almost universally experienced (implicit) associations between stimulus features in different sensory modalities. A well-studied example of correspondence effect was found for pitch and visual elevation. When an object that is visually elevated in space is paired with a high-pitched tone, a stronger association of these features is observed compared to pairing the same object with a low-pitched tone (Ben-Artzi & Marks, 1995; Chiou & Rich, 2012; Evans, 2020; Evans & Treisman, 2010; Jamal et al., 2017; McCormick et al., 2018; Melara & Brien, 1987). A second pitch-based CMC is pitch and size. Presenting e.g., a small object together with a high-pitched tone resulted in a successful crossmodal correspondence in a study by Evans and Treisman (2010) and numerous other studies (Bien et al., 2012; Bonetti & Costa, 2018; Gallace & Spence, 2006; Parise & Spence, 2012).

There are several theories regarding the origin of pitch-size and pitch-elevation CMCs resulting in the assumption of different or common information-processing mechanisms for the different CMC effects. Despite numerous behavioral studies on crossmodal correspondences (Chiou & Rich, 2012; Evans & Treisman, 2010; Evans, 2020; Parise et al., 2014; Spence, 2011, 2020; Uno & Yokosawa, 2022a, 2022b), only a few studies used neuroimaging to address the neural basis of different pitch-based CMCs (McCormick et al., 2018; Sadaghiani et al., 2009).

A prominent theory about the origin of pitch-elevation correspondences is based on language process-

Received 23 October 2023, accepted 17 April 2024

ing (Parise et al., 2014; Spence, 2011, 2020; Spence & Sathian, 2020). In most cultures, the words 'high' and 'low' can describe both, the height of a pitch and the position of an object in space. This linguistic link is not described for pitch and size correspondences. Even though a language-driven cause for pitch-elevation associations is plausible (Ben-Artzi & Marks, 1999), a growing number of studies on the CMC effect between pitch and spatial-elevation raise the question if other variables than language probably cause the strong CMC of these seemingly arbitrary stimulus pairs (McCormick et al., 2018; Parise et al., 2014; Parkinson et al., 2012).

A second theory on CMCs declares that the correspondences between pitch and elevation as well as pitch and size probably arise from regularities in our natural environment that are stored in memory (Parise et al., 2014; Spence, 2011; 2020; Spence & Sathian, 2020). For example, larger bodies usually resonate lower pitches and smaller objects tend to resonate higher pitches (Parise et al., 2014). We are confronted with this regularity frequently in our daily lives. Children typically have a higher-pitched voice than adults (Lee et al., 1999) and small animals tend to make higher-pitched noises than larger animals (Bowling et al., 2017). We also tend to perceive higher-pitched tones from objects elevated in space than when on the ground (Parise et al., 2014). Following this assumption, CMCs probably have their roots in statistical regularities, i.e. naturally learned rules and assumptions from our environment (Parise et al., 2014; Spence, 2011; 2020; Spence & Sathian, 2020). If both CMCs have their origin in similar mechanisms, great activations within comparable brain regions will be measured in both pitch-size and pitch-elevation CMCs.

The third and last theory we are going to address is the theory of perceived intensity, which is also called a theory of magnitude (Spence, 2011). This theory declares that the CMC effect probably evolved from a correspondence in intensity or magnitude in the underlying neuronal structure of corresponding stimulus pairs (Spence, 2011; Spence & Sathian, 2020). The main idea underlying the magnitude in CMCs is a shared polar dimension of the stimulus pairs perceived as congruent. According to this notion, a high-pitched tone and a small visual stimulus would be situated on the same side of their respective polar dimension. Compared to incongruent stimulus pairs, congruent pairs would share 'more' in terms of intensity or magnitude (Chang & Cho, 2015). A common neural activation in terms of magnitude was found for e.g., numbers by Piazza et al. (2007) and sizes with luminance by Pinel and colleagues (2004). If pitch and size and pitch and elevation correspondences have their origin in similar coded neuronal responses, we hypothesize to find greater activations within the intraparietal sulcus (IPS) for congruent trials as a common effect in both CMCs (Humphreys & Ralph, 2015; Piazza et al., 2007; Pinel et al., 2004).

The CMC is in behavioral studies often measured *via* the reaction time (RT) differences between congruent and incongruent stimulus pairs. Thus being significant, these differences are rather small in absolute values (Chiou & Rich, 2012; Evans & Treisman, 2010). The study by Evans and Treisman, performed in 2010, included eight subjects in their pitch-size visual experiment, in which the absolute difference between congruent and incongruent trials was 14.4 ms. Their pitch-elevation visual paradigm included twelve participants and the absolute difference between congruent and incongruent RTs was 18.6 ms. Within their fMRI paradigm, McCormick and colleagues (2018) did not find significant RT differences between congruent and incongruent stimulus presentation in the pitch-elevation CMC, what may be related to the overall small size of the effect. They validated their findings outside the scanner with a behavioral task (McCormick et al., 2018). Based on these previous findings, performing a behavioral test outside the scanner appears to be an appropriate measure to validate CMC effects studied with fMRI (Koten et al., 2013; McCormick et al., 2018). In our study, we implemented a congruence classification task outside the scanner to measure the behavioral CMC effect in addition to the typically measured RTs.

Even though pitch-based correspondences are almost universally experienced and well-studied (Ben-Artzi and Marks, 1995; Bien et al., 2012; Chiou & Rich, 2012; Evans, 2020; Evans and Treisman, 2010; Gallace & Spence, 2006; Jamal et al., 2017; Marks, 1987; Spence, 2011; 2020; Zeljko et al., 2019), the evidence for the underlying neural mechanisms is still lacking. A study that used functional magnetic resonance imaging (fMRI) to examine pitch and elevation congruencies showed a probable involvement of the right angular gyrus (AnG) as well as the mid-IPS for corresponding stimulus presentations (McCormick et al., 2018).

The main focus in our study was to examine the neural basis of the processing of a pitch-elevation CMC and compare this to a pitch-size CMC while both CMCs are always in the focus. The estimation of the CMC effect can be achieved by the calculation of the difference of congruent > incongruent (C > I) presentations. The calculated difference then allows a direct comparison of the neural substrates of the CMC effect between the different CMCs. This comparison can be used to test common or different neural correlates of different CMCs focusing on the CMC effect, thus directly testing the different theoretical assumptions about the origin of the CMC effect.

If we find a common effect within the IPS for congruent > incongruent presentations, a magnitude driven CMC is likely to cause this effect. An effect within the left inferior frontal gyrus (IFG) is favorable for a CMC driven by language, which we hypothesize to find most likely for congruent > incongruent pitch-elevation presentations. However, if CMCs are based on statistical representations of our environment, we will most likely find an effect within areas common for attention and memory retrieval like the anterior cingulate cortex (ACC) or the AnG.

Although we are mainly interested in congruency effects, it cannot be excluded that effects for incongruent stimulus presentations are also part of the processing of the stimuli in our tested CMCs. Stronger effects for incongruent stimuli could be due to for example response conflict or a shift of attention (Chiou & Rich, 2012; Spence & Sathian, 2020).

Besides the question about the neural mechanisms between two different CMCs, we were interested in a probable modulation of the effect within one CMC by stimulus contrast. It has been hypothesized that the CMC effect depends on the ability to form a unique correspondence between stimulus pairs (Chiou & Rich, 2012). Therefore, we additionally measured a variant of the pitch-size CMC, in which we reduced the difference, i.e. the contrast between the stimuli to probably also reduce the CMC effect (Chiou & Rich, 2012). If the CMC effect is modulated by the contrast of the stimulus pairs, we hypothesized to find a reduced neural effect for the variant of the pitch-size CMC with a reduced difference between the stimulus pairs.

METHODS

Participants

Thirty-three mentally and physically healthy participants (21 females, age M=24.8 years, SD=3.8 years) with normal hearing and normal or corrected-to-normal vision took part in this experiment. The participants were recruited through a local online job platform. Four participants had to be excluded from the final analysis (two due to technical issues and two due to excessive movement in the scanner (>5 mm)). Therefore, the final sample size was 29 participants. All experiment protocols were approved by the Ethics Committee of the General Medical Council Hamburg (PV7022) and all our methods were carried out in accordance with relevant ethical guidelines and regulations. All participants gave their written informed consent and were paid an expense allowance of 10 €/h.

Apparatus

Inside the scanner

The stimuli were presented using Presentation® software (Version 22.01, Neurobehavioral Systems, Inc., Berkeley, CA) running on Windows 7. A mirror placed on the head coil with ~ 12 cm distance to the participant's face was used to reflect the stimulus presentation from a 40'' LCD screen with a refresh rate of 60 Hz. The auditory stimuli were presented using MR compatible in-ear head phones (MR confon). Participant responses were tracked using two MR compatible button boxes.

Outside the scanner

For a tutorial as well as the congruence classification task, Psychopy (Version 3.2.4) software running on a 15' hp laptop with Windows 10 was used to present the stimuli. A button box with two active buttons (one for each hand) was used to track the participants' responses. Auditory stimuli were presented *via* loudspeaker on both sides of the screen.

Stimuli

Black squares on a grey background were used as visual stimuli (Fig. 1A, B) and instrument tones (edited with Audacity® recording and editing software version 2.4.1) were used as acoustic stimuli. For pitch-size, squares (0° 39' .26'' & 3° 55' .03'') were presented with the sound of a piccolo flute (1225 Hz, D#/Eb6) or a double bass (73 Hz, D2). For pitch-size variant with reduced difference, squares (1° 18' .30'' & 2° 36' .79'') were presented either with a violin (588 Hz, D5) or a bassoon (149 Hz, D3). In between trials, a 0° 29' .50'', white fixation cross was presented in the middle of the screen. Small squares presented together with high-pitched tones and bigger squares presented together with low-pitched tones will be referred to as pitch-size congruent condition in the following (Fig. 1A).

For measuring the pitch-elevation CMC, a black square (1° 8' .54'') was presented either above or below a 0° 29' .50'', white fixation cross. The cross was presented in the center of the screen and the distance of the squares to the center was 3° 14' .18''. The auditory stimuli used were the same as in the variant of pitch-size with reduced difference. Squares above the fixation cross presented together with higher pitched tones and squares presented below the fixation cross together with lower tones will be referred to as congruent trials in the following (Fig. 1A).

Experimental design and procedure

We tested two different CMC types within our study, pitch-size and pitch-elevation, as well as a variant of the pitch-size CMC with reduced difference between the stimulus contrasts. A tutorial was performed by the participants before entering the scanner. Within the tutorial, the two distinct CMCs as well as the pitch-size variant with reduced difference were introduced separately to the participants. Each tutorial part for the two distinct CMCs and the pitch-size variant with reduced difference consisted of twelve trials (8 congruent; 4 incongruent). Within the tutorial, the participants were not introduced to the concept of congruence and incongruence hidden behind the stimulus pairings. The purpose of the tutorial was to familiarize the participants with the stimulus pairs and the focus was always on the visual stimuli.

After the tutorial, the participants were placed in the scanner. Before the experiment started, the volume of the acoustic stimuli were adjusted while the participants were exposed to the scanner noise. With the latter procedure we ensured a comfortable but valid presentation of the acoustic stimuli in the scanner.

We used an event-related design with jittered inter trial intervals (ITI) to present the stimuli in the scanner. In the main experiment, each participant saw all CMCs, the two distinct CMCs and the variant of pitch-size with reduced difference (Fig. 1A, B), in separate runs. The duration of a run was ~ 10 minutes. The order in which the two distinct CMCs or the variant of pitch-size with reduced difference were presented was counterbalanced between participants.

Each of the three runs, in which one of the two distinct CMCs (Fig. 1A) or the variant of pitch-size with reduced difference (Fig. 1B) was presented, consisted of 96 trials with 48 repetitions of each condition (congruent; incongruent) and 24 presentations of each stimulus pair (e.g., small square and high pitch) (Chiou & Rich, 2012). The 96 trials were presented in a pseudo-randomized order and this order was also randomized between participants. In each trial, a visual stimulus was presented simultaneously with a sound (Fig. 1C). The participants were instructed to respond to the different visual stimuli as fast and precise as possible. For small as well as elevated stimuli, the correct button press was performed with the left index finger. For large and low presented stimuli, the button press was performed with the right index finger. The audio-visual presentation lasted for 1000 ms followed by 500 ms of extended key response time. The inter-trial interval was jittered between 2000 – 8000 ms with a mean of ~ 5000 ms (Fig. 1C). Instructions were prompted on a screen in the scanner before each new run started. The participants had the opportunity to take a short break between the runs, however, they had to stay in the scanner during the short break.

A stimulus congruence classification task was performed outside the scanner following the main experiment. Within the stimulus congruence classification, the participants were instructed to classify if the audio-visual stimulus presentations match each other or not (Fig. 1D). The congruence classifications was separately performed for each distinct CMC and variant of pitch-size with reduced difference, whereby each condition (congruent; incongruent) was presented six times in a random order. The participants were instructed to classify the presented audio-visual pairs by clicking on the respective side of a scale with a computer mouse (Fig. 1D). Thereby only the ends of the scale could be clicked, no gradual adjustment was possible. The participants were instructed to classify intuitively if the audio-visual stimuli were matching or not. No feedback on the chosen pair was given. We conducted this final task to test whether the participants correctly matched the congruent and incongruent stimulus pairs in accordance with the CMC theory (Fig. 1A, B).

Behavioral data analysis

The focus of the analysis of the behavioral data was the stimulus congruence classification performed outside the scanner. All statistical tests on the behavioral data were performed in JASP (Version 0.16.1).

All congruence classifications were taken into account for the further analysis. The classifications were then divided into trials in which participants chose 'matching' and trials in which participants chose 'not matching' separately for each condition and each CMC, i.e., the two distinct CMCs (Table 1) and the variant of pitch-size with reduced difference (Table 2). We were interested in whether participants would classify our congruent stimulus pairs as matching and our incongruent stimulus pairs as not matching, i.e. whether participants show the expected classification of pairs in accordance with the CMC theory. We also wanted to know whether these classifications are dependent on the tested CMCs. Therefore, we conducted a repeated measures ANOVA to test the effect of the within-subject factors 'distinct CMCs (pitch-size & pitch-elevation),' as well as 'Classifications of congruent stimuli (congruent stimuli rated as matching & congruent stimuli rated as not matching)' on stimulus classifications. We also conducted a second repeated measures ANOVA to test the effect of the within subject factors

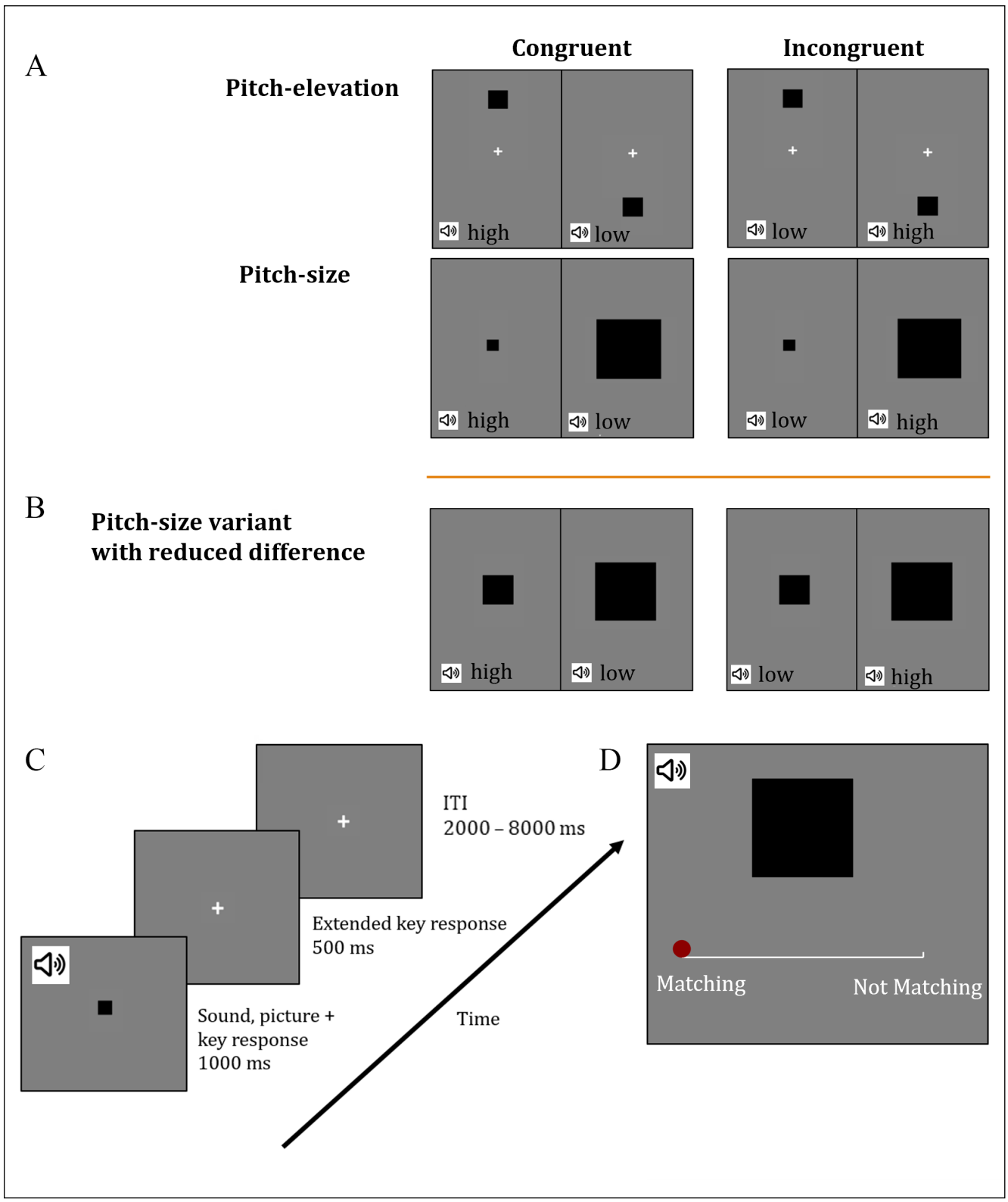

Fig. 1. (A) Overview of the congruent (left) and incongruent (right) stimulus pairs for each of the two distinct CMCs (from top to bottom: Pitch-elevation, pitch-size). (B) Overview of the congruent (left) and incongruent (right) stimulus pairs for the pitch-size variant with reduced difference between the visual and acoustic stimulus pairs. (C) Schematic sequence of events in a trial (here pitch-size). (D) Example of a trial within the post experimental test for the strength of congruence outside the scanner. Participants were asked to classify simultaneous sound and square presentations as matching or not matching without knowing the purpose of the main experiment. Only the ends of the scale could be clicked, no gradual adjustment was possible.

'pitch-size and pitch-size variant with reduced difference', as well as 'Classifications of congruent stimuli (congruent stimuli rated as matching & congruent stimuli rated as not matching)' on stimulus classifications. We also conducted these repeated measures ANOVAs for the stimulus classification performed on incongruent stimuli.

For the sake of completeness, we analyzed the RTs of the in-scanner task. Only RTs from correct trials and trials with RTs below 1000 ms were taken into account in the further analysis. We chose this threshold, as the stimulus presentation ended after 1000 ms. Furthermore, we wanted to avoid including decisions formed by e.g., complex cognitive processing or inattentiveness. We performed two repeated measures ANOVAs, one for the two distinct CMCs and another for the pitch-size CMC and its variant with reduced difference, to test whether there are differences between the RTs of the trial conditions (congruent trials & incongruent trials) and whether these differences are different between the CMCs. One repeated measures ANOVA was conducted to compare the within subject factors 'distinct CMCs (pitch-size & pitch-elevation)', as well as 'Condition (congruent trials & incongruent trials)' on RTs and the other repeated measures ANOVA was conducted to compare the effect of the within subject factors 'pitch-size and its variant with reduced difference', as well as 'Condition (congruent trials & incongruent trials)' on RTs.

Functional data acquisition

Imaging was performed on a 3-T scanner (Siemens Trio) using a 64-channel head coil. A standard gradient echo-planar imaging (EPI) T2*-sensitive sequence with 54 axial slices (2 mm thickness with .5 mm gap, voxel size=2 × 2 × 2 mm; repetition time (TR)=1640 ms and echo time (TE)=29 ms, multiband factor=2, Flip angle=70°) was acquired for functional imaging. A high resolution (1 × 1 × 1 mm voxel size) T1 weighted, three dimensional, defaced MPRAGE image (TR=7.1 ms, TE=2.98 ms, FA=9°, inversion time=1100 ms) was additionally acquired for each participant. The experiment started after the scanner reached magnetic stabilization.

Functional data analysis

Preprocessing and statistical analysis of the fMRI data were carried out in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/) on Matlab version R2020a. Image preprocessing steps included a correction for the magnetic field distortion by unwarping the images using a fieldmap, as well as motion correction with registration on the first EPI, correcting for between subject anatomical differences by normalizing images on EPI with the EPI template provided by SPM12 and smoothing the normalized images with a 6 mm (full widths half maximum; FWHM) Gaussian kernel. We did not correct for RT differences between congruent and incongruent conditions as the expected RT difference was below 100 ms (Chiou & Rich, 2012; Evans & Treisman, 2010).

The hemodynamic response for each condition (congruent; incongruent) was modelled as an event-related design (for further information see Experimental design and procedure). The six contrast images (main effects) per participant, calculated from onsets of each condition, were entered into a flexible factorial group level analysis and all statistical comparisons were estimated on the group level.

Functional data analysis of the distinct CMCs

To test for differences between the processing of the conditions in the distinct pitch-size and pitch-elevation CMCs, we estimated two interaction contrasts at the second level. To test for enhanced neural effects of congruent trials selectively in the pitch-size CMC an interaction contrast was estimated (Pitch-size (C>I) > pitch-elevation (C<I)). To test for enhanced neural effects of congruent trials selectively in the pitch-elevation CMC, another interaction contrast was estimated (Pitch-size (C<I) < pitch-elevation (C>I)). We also tested for differences in neural effects between the conditions within the distinct CMCs. Therefore, we estimated contrasts that tested for enhanced neural effects of congruent stimuli within pitch-size (C>I) and pitch-elevation (C>I), as well as contrasts that tested for enhanced neural effects of incongruent stimuli within pitch-size (C<I) and pitch-elevation (C<I) (Table 3). To test for common neural effects of congruent trials within the two distinct CMCs, a global conjunction was estimated (Pitch-size (C>I) & pitch-elevation (C>I)). Statistically significant whole brain fMRI effects were family wise error corrected (FWE, $p<0.05$).

Functional data analysis of pitch-size and its variant with reduced difference

To test for common neural effects of congruent trials within pitch-size and its variant with reduced difference, a global conjunction was estimated (Pitch-size (C>I) & pitch-size variant with reduced difference (C>I)).

To test for enhanced neural effects of congruent trials selectively in the pitch-size CMC, an interaction contrast was estimated (Pitch-size (C>I) > pitch-size variant with reduced difference (C<I)). Statistically significant whole brain fMRI effects were family wise error corrected (FWE, p<0.05).

Regions of interest

Main region of interests related to previous studies and their corresponding coordinates used in this study are the ACC [TAL: x=+/- 6, y=30, z=38] (Roelofs et al., 2006; coordinates were converted into MNI space using the implemented mni2tal tool from Yale Bioimage Suite Web (Version: 1.2.0 (2020/08/25)), the AnG [MNI: x=+/- 48, y=-64, z=34] to test for a statistical or environmental driven CMC (Humphreys & Ralph, 2015) and the IPS for a magnitude driven CMC [MNI: x=+/- 43, y=-42, z=48] (Humphreys & Ralph, 2015). We tested for language related effects within the triangular part [MNI: x=-39, y=26, z=13] of the left IFG as well as the orbital part [MNI: x=-30, y=35, z=-14] of the left IFG (Liuzzi et al., 2017). We reported significant fMRI effects of our ROIs using a sphere with a radius of 10 mm which was small volume corrected with FWE (p<0.05).

RESULTS

Behavioral results

Behavioral results of the stimulus congruence classification task

We tested separately whether the congruent and incongruent trials were overall classified as matching compared to not matching with a classification task. We also tested whether there is a difference in classifications of congruence or incongruence depending on the CMC. We tested this dependence between the distinct CMCs and between pitch-size and its variant with reduced difference. The results of the classification task, which was performed outside the scanner, showed that the congruent stimulus pairs were overall classified as matching and incongruent stimulus presentations were overall classified as not matching, i.e., the participants' classification aligns with the CMC theory (Spence, 2011). A dependence of the classification strength was observed for the classification of congruent trials between pitch-size and pitch-elevation with more congruent trials rated as matching in pitch-size compared to pitch-elevation (details in the following sections).

Stimulus congruence classification results of the two distinct CMCs

To test whether congruent stimulus pairs were significantly classified as matching by the participants (Table 1) and to test whether this classification different between the distinct CMCs, a repeated measures ANOVA with the distinct CMCs and Classifications of congruent stimuli (congruent trials rated as matching & congruent trials rated as not matching) as within subject factors was performed. For the distinct CMCs, this repeated-measures ANOVA showed a reliable effect for the factor Classifications of congruent stimuli ($F_{(1,28)}$=133.8, p<0.001, $hp2$=0.83; Table 1). This means that congruent pairs were significantly classified as matching (M=86.2%, SEM=3.9%; Table 1). An interaction of distinct CMCs × Classifications of congruent stimuli was statistically significant ($F_{(1,28)}$=7.398, p=0.011, $hp2$=0.209) and a *post-hoc* test revealed that significantly more congruent pairs were classified as matching within pitch-size compared to pitch-elevation ($pholm$=0.022). This means that a significant difference was observed for the number of congruent trials rated as matching between the pitch-size and pitch-elevation CMCs, i.e., a higher congruence classification was observed for congruent stimuli in pitch-size compared to pitch-elevation (Table 1). Furthermore, significantly more congruent pairs were classified as matching within pitch-size ($pholm$<.001) as well as within pitch-elevation ($pholm$<.001). We observed no effect for distinct CMCs ($F_{(1,28)}$=-6.010e^{-14}, p=1.0, $hp2$=-2.146e^{-15}).

To test whether incongruent stimulus pairs were significantly classified as not matching by the participants and to test whether this classification is different between the distinct CMCs this classification is different between the distinct CMCs, a repeated measures ANOVA with the distinct CMCs and Classi-

Table 1. Congruence classification of the two distinct CMCs for congruent and incongruent stimulus presentations. Mean in percent for stimulus conditions (congruent, incongruent) within each distinct CMC (pitch-size, pitch-elevation) classified as matching, i.e. congruent or not matching, i.e. incongruent. The maximum possible percentage for each presented condition (congruent; incongruent) in each distinct CMC was 100% for classifications of matching and not matching taken together.

Condition presented	Pitch-size classified as		Pitch-elevation classified as	
	matching	not matching	matching	not matching
congruent	92	8	80.5	19.5
incongruent	23.6	76.4	31	69

fications of incongruent stimuli (incongruent trials rated as matching & incongruent trials rated as not matching) as within subject factors was performed. For the distinct CMCs, this repeated-measures ANOVA showed a reliable effect for the factor Classifications of incongruent stimuli ($F_{(1,28)}$=22.34, p<.001, $hp2$=0.44). This means that incongruent pairs were significantly classified as not matching (M=72.7%, SEM=5.6%; Table 1). No effect for distinct CMCs ($F_{(1,28)}$=-4.695e^{-14}, p=1.0, $hp2$=-1.677e-15) as well as the interaction of distinct CMCs × Classifications of incongruent stimuli ($F_{(1,28)}$=1.714, p=0.2, $hp2$=0.058) was observed. This means that no significant difference was observed between pitch-size and pitch-elevation for the classification of incongruent stimuli.

Stimulus congruence classification results of the pitch-size CMC and its variant with reduced difference

To test whether congruent stimulus pairs were significantly classified as matching by the participants (Table 2) and to test whether this classification is different between pitch-size CMC and its variant with reduced difference, a repeated measures ANOVA with the within subject factors pitch-size and its variant with reduced difference and Classifications of congruent stimuli (congruent trials rated as matching & congruent trials rated as not matching) was performed. This repeated-measures ANOVA showed a reliable effect for the factor Classifications of congruent stimuli ($F_{(1,28)}$=216.87, p<.001, $hp2$=0.89; Table 2). This means that congruent pairs were significantly classified as matching (M=87.4%, SEM=3.35%; Table 2). An interaction of Pitch-size and its variant with reduced difference × Classifications of congruent stimuli was statistically significant ($F_{(1,28)}$=5.072, p=0.033, $hp2$=0.152) and a *post-hoc* test revealed that significantly more congruent pairs were classified as matching within pitch-size ($pholm$<0.001) as well as within the pitch-size variant with reduced difference ($pholm$<0.001). However, there was no significant difference observed between congruent stimuli classified as matching for pitch-size and its variant with reduced difference ($pholm$=0.066). No main effect was observed for the stimulus classifications of pitch-size and its variant with reduced difference ($F_{(1,28)}$=1.699e^{-13}.0, p=1.0, $hp2$=6.068e^{-15}). This means that no significant difference was observed between pitch-size and its variant with reduced difference for the classification of congruent stimuli.

Table 2. Congruence classification of pitch-size and its variant with reduced difference for congruent and incongruent stimulus presentations. Mean in percent for stimulus conditions (congruent, incongruent) within pitch-size and pitch-size with reduced difference classified as matching, i.e. congruent or not matching, i.e. incongruent. The maximum possible percentage for each presented condition (congruent; incongruent) in each CMC was 100% for classifications of matching and not matching taken together.

Condition presented	Pitch-size classified as		Pitch-size variant classified as	
	matching	not matching	matching	not matching
congruent	92	8	82.8	17.2
incongruent	23.6	76.4	23	77

To test whether incongruent stimulus pairs were significantly classified as not matching by the participants (Table 2) and to test whether this classification is different between the pitch-size CMC and its variant with reduced difference, a repeated measures ANOVA with pitch-size and its variant with reduced difference and Classifications of incongruent stimuli (incongruent trials rated as matching & incongruent trials rated as not matching) as within subject factors was performed. This repeated measures ANOVA showed an effect for Classifications of incongruent stimuli ($F_{(1,28)}$=33.3, p<.001, $hp2$=0.54). This means that incongruent pairs were significantly classified as not matching (M=76.7%, SEM=5.4%; Table 2). No effect was observed for pitch-size CMC and its variant with reduced difference ($F_{(1,28)}$=-5.634e^{-15}, p=1.0, $hp2$=-2.012^{-16}) as well as for the interaction of pitch-size CMC and its variant with reduced difference × Classifications of incongruent stimuli ($F_{(1,28)}$=0.01, p=0.92, $hp2$=3.55e^{-4}). This means that no significant difference was observed between pitch-size and its variant with reduced difference for the classification of incongruent stimuli.

Behavioral results of the in scanner-task recorded reaction time data

We discarded 2.25% of trials due to false or missing responses and 0.9% of trials due to slow button presses (over 1000 ms).

Error rates and reaction time data of the in scanner task for the two distinct CMCs

The error rates for pitch-size were 3.27% and for pitch-elevation 3.3%. To test whether there is a difference in RTs depending on Conditions (congruent trials & incongruent trials) and the tested distinct CMCs (pitch-size & pitch-elevation) we performed a repeated-measures ANOVA with the distinct CMCs and the two conditions as within subject factors on RTs. This ANOVA showed a main effect

for CMC Types ($F_{(1,28)}$=12.11, p=0.002, $hp2$=0.302). The RTs in the pitch-elevation CMC were significantly lower (456.3 ms; SEM=12.1 ms) compared the RTs in Pitch-size (492 ms; SEM=9.5 ms). No main effect was observed for Condition ($F_{(1,28)}$=0.044, p=0.836, $hp2$=0.002) as well as for the interaction of distinct CMCs × Condition ($F_{(1,28)}$=0.055, p=0.817, $hp2$=0.002). In summary, a significant difference in RTs was observed between the distinct CMCs, indicating a probable difference in processing the stimulus pairs dependent on the tested CMC type.

Error rates and reaction time data of the in scanner task for the pitch-size CMC and its variant

The error rates for pitch-size were 3.27% and for pitch-size with reduced difference were 2.95%. To test whether there is a difference in RTs depending on Conditions (congruent trials & incongruent trials) and the tested CMCs pitch-size and its variant with reduced difference, we performed a repeated-measures ANOVA with the pitch-size and its variant with reduced difference and the two conditions as within subject factors on RTs. This repeated-measures ANOVA showed a main effect for pitch-size and its variant with reduced difference ($F_{(1,28)}$=26.4, p<.001, $hp2$=0.485). RTs in the pitch-size CMC (492 ms; SEM=9.5 ms) were significantly lower compared to the pitch-size variant with reduced difference (526.5 ms; SEM=9.8 ms). The main effect for Condition ($F_{(1,28)}$=1.025, p=0.32, $hp2$=0.04) as well as for the interaction of pitch-size and its variant with reduced difference × Condition ($F_{(1,28)}$=1.25, p=0.27, $hp2$=0.04) were not significant. In summary, a significant difference in RTs was observed between pitch-size and its variant with reduced difference, indicating a probable difference in processing the stimulus pairs dependent on the tested pitch-size CMC.

Functional data

We were interested in finding the neural components of different CMC types as well as between pitch-size and its variant with reduced difference to evaluate different and common neural mechanisms underlying pitch-based CMC effects. To test for congruence effects in dependence of the CMCs pitch-size and pitch-elevation, interaction analyses were performed. To test whether there were common effects for congruent trials between distinct CMCs as well as between pitch-size and its variant pitch-size with reduced difference, global conjunction analyses (k>0) were performed.

Differences between pitch-size and pitch-elevation crossmodal correspondences

We did not find common congruence (C>I) effects for pitch-size and pitch-elevation CMCs when a global conjunction was performed. Albeit clear differences between the distinct CMCs were observed when we tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. The computed interaction for pitch-size (C>I) > pitch-elevation (C<I) revealed whole brain family wise error (FWE; p<0.05) corrected activations within the left superior parietal lobule (SPL; MNI coordinates: x=-26, y=-48, z=62; T=5.39, p=0.021), the left Heschls' gyrus (HG; MNI coordinates: x=-32, y=-24, z=6; T=5.33, p=0.026) and the left cerebellum (MNI coordinates: x=-22, y=-58, z=-30; T=5.22, p=0.041; Fig. 2). We observed greater neural effects for congruent trials (C>I) in pitch-size and greater neural effects for incongruent trials (C<I) in the pitch-elevation CMCs in these regions (see Fig. 2 as well as Table 3 and Table 4).

Additionally, small volume, FWE (p<0.05) corrected effects in our ROIs 10 mm spheres were observed within the right AnG (MNI coordinates: x=40, y=-66, z=32; T=3.7, p=0.03), left ACC (MNI coordinates: x=-2, y=22, z=34; T=4.13, p=0.008), right ACC (MNI coordinates: x=8, y=28, z=40; T=4.56, p=0.002) and nearly significant in the left IPS (MNI coordinates: x=-42, y=-36, z=46; T=3.51, p=0.05; Fig. 3 and Table 3). No activation of the IFG survived with the small volume correction.

The effect within the right AnG was dominated by a positive CMC effect within pitch-size (C>I; T=3.94, p=0.014 corr.). The measured effects within the right ACC were significantly greater for congruent trials in pitch-size (C>I; T=3.97, p=0.013 corr., Table 4) and for incongruent trials in pitch-elevation (C<I; T=3.61, p=0.038 corr., Table 4). The elicited effect in the IPS was driven by incongruent trials within the presentation of pitch-elevation (C<I; T=4.06, p=0.01 corr.; see Fig. 3). For the sake of completeness, we have included a table presenting the statistically significant results of the small volume corrected regions of interest for both pitch-size and pitch-elevation CMCs of both conditions (C>I, C<I; Table 4).

To evaluate a possible influence of RT differences between the pitch-size and pitch-elevation CMCs on the neural activity, we replicated the interaction analysis for pitch-size (C>I) > pitch-elevation (C<I) by including individual median RT data for each condition as a parametric modulator in the group level analysis. The results showed no substantial changes in activity neither in the whole brain nor the ROIs. All reported coordinates in the interaction analysis remained the same and the p-value remained at p<0.05 for the SPL

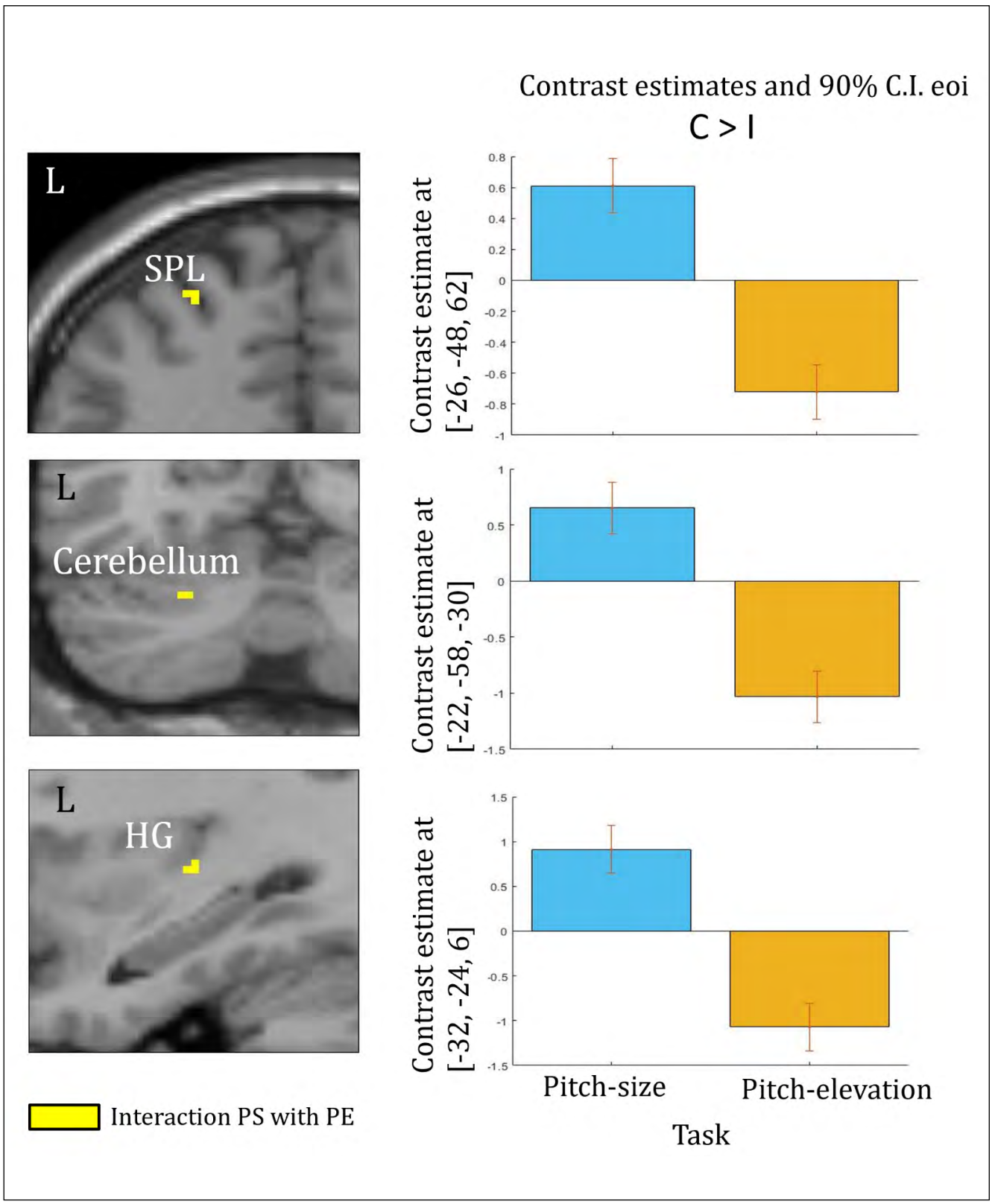

Fig. 2. Effects of the interaction for pitch-size with pitch-elevation (left), which tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. Contrast estimates for the corresponding region are displayed on the right (blue = Pitch-size; orange = Pitch-elevation). Each bar resembles the activation difference (C>I) within the specific region. Whole-brain FWE (peak-level; $p<0.05$) corrected effects were observed in the SPL, cerebellum and HG. Error bars indicate the observed standard error within each contrast. (The FWE corrected analysis with $p<0.05$ was used to create the images above).

Table 3. fMRI effects for an interaction between pitch-size and pitch-elevation. T Interaction effects within a) whole brain FWE corrected (p<0.05) peak-level effects and b) small volume, FWE corrected (p<0.05) peak-level effects within the defined ROIs. Statistically significant effects are shown and therefore no results for the interaction testing for enhanced neural effects of congruent trials selectively in pitch-elevation (pitch-size (C<I) < pitch-elevation (C>I)) are included in this table.

Pitch-size (C>I) > Pitch-elevation (C<I)					
			MNI coordinates		
Hem	Regions	T-value	x	y	z
a)					
L	Superior parietal lobule	5.39	-26	-48	62
L	Heschels' gyrus	5.33	-32	-24	6
L	Cerebellum	5.22	-22	-58	-30
b)					
R	Anterior cingulate cortex	4.56	8	28	40
L	Anterior cingulate cortex	4.13	-2	22	34
R	Angular gyrus	3.70	40	-66	32
L	Intraparietal sulcus	3.51	-42	-36	46

Table 4. Condition specific fMRI findings for pitch-size and pitch-elevation. Findings of the small volume, FWE ($p<0.05$) corrected ROIs peak-level effects for pitch-size and pitch-elevation by condition (C>I; C<I). Statistically significant effects are shown and therefore no results for pitch-size C<I and pitch-elevation C>I are included in the table.

Pitch-size					
			MNI coordinates		
Hem	Regions	T-value	x	y	z
C>I					
R	Anterior cingulate cortex	3.97	8	36	32
R	Angular gyrus	3.94	40	-68	32
Pitch-elevation					
			MNI coordinates		
Hem	Regions	T-value	x	y	z
C<I					
R	Anterior cingulate cortex	3.61	6	20	38
R	Anterior cingulate cortex	3.60	4	30	40
L	Intraparietal sulcus	4.06	-42	-36	44

(T=5.36), cerebellum (T=5.20), HG (T=5.29), right ACC (T=5.57), left ACC (T=4.11) and AnG (T=3.69). The neural activity found in the left IPS (T=3.46) stayed at a nearly significant level (p=0.058). No activation of the IFG survived with the small volume correction. These results confirm that the interaction effects between the pitch-size and pitch-elevation CMCs are independent of the observed RT differences.

Common effects for pitch-size and pitch-size with reduced difference correspondences

We were interested in to what extent a reduced difference between the visual and acoustic stimuli could modulate a CMC effect for pitch and size. No common effect was found between pitch-size and its variant with reduced difference. This means that the global conjunction over pitch-size (C>I) & pitch-size variant with reduced difference (C>I) did not lead to significant effects within the whole brain or our ROIs. We therefore concluded that there was likely a difference in neural effects between pitch-size and its variant with reduced difference. Subsequently, we examined whether there were any differences in our ROIs between pitch-size and its variant with reduced difference by conducting an interaction analysis that tested for enhanced neural effects of congruent trials selectively in pitch-size (pitch-size (C>I) > pitch-size variant with reduced difference (C<I)). A differential effect in the ACC (MNI co-ordinates: x=-2, y=30, z=26) with a T value of 3.75 was observed. However, it did not survive corrections for multiple comparisons.

DISCUSSION

In this study, the neural correlates between different CMCs were compared to evaluate a possible common or different processing of CMCs and further to obtain a possible clue regarding the theories on the origins of CMCs. Our results are in favor of different processes underlying different pitch-based CMCs as well as in favor of the theory that these CMCs are probably driven by statistical regularities from our environment.

The participants significantly classified the congruent stimulus pairs as matching compared to not matching and the incongruent stimulus pairs as not matching compared to matching in both distinct CMCs as well as the variant of pitch-size with reduced difference in the stimulus congruence classification, which was per-

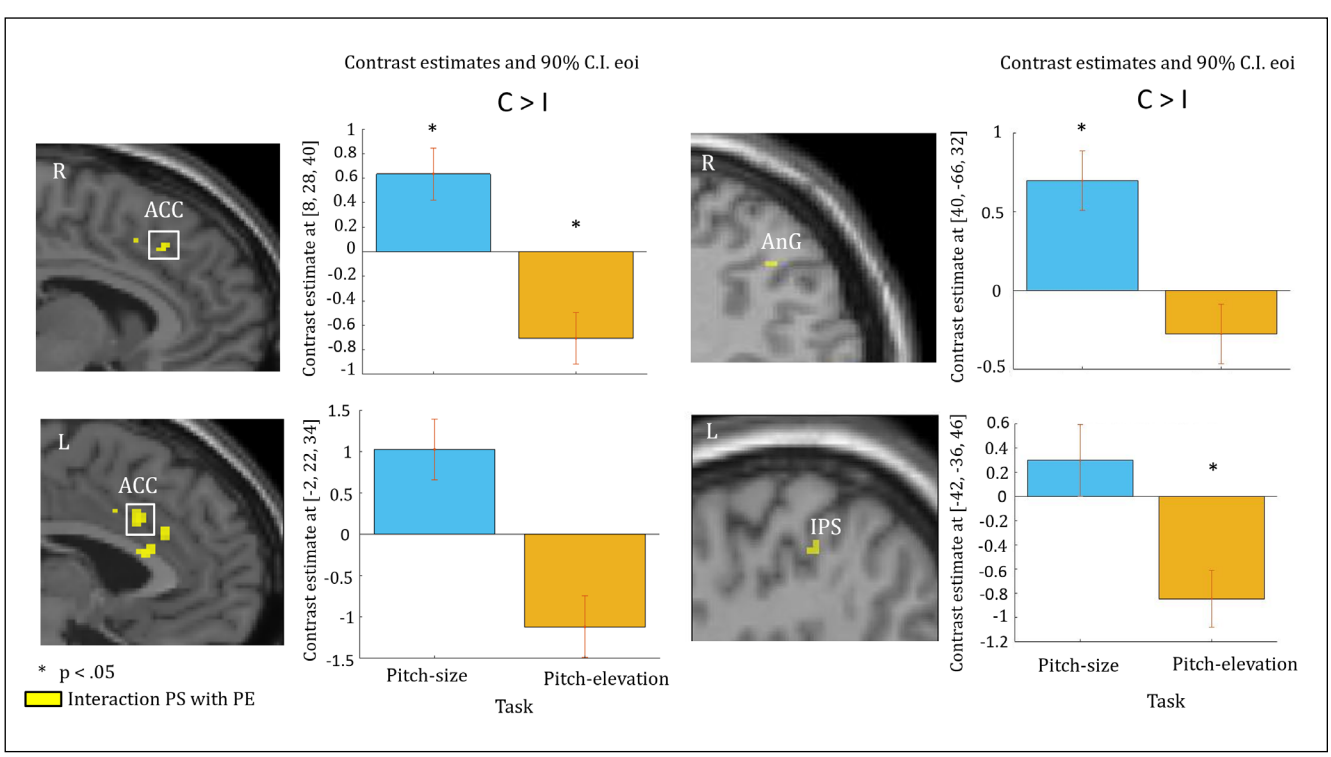

Fig. 3. Effects of the interaction for pitch-size with pitch-elevation (left), which tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. Contrast estimates for the corresponding region are displayed on the right (blue = Pitch-size; orange = Pitch-elevation). Each bar resembles the activation difference (C>I) within the specific region. FWE corrected small volume effects (peak-level; p<0.05) were observed in the ACC, AnG and IPS. Error bars indicate the observed standard error within each contrast. (For display purposes, a T value of > 4 was chosen to create the images above).

formed outside the scanner. This means that the participants classified the stimulus pairs in accordance with the CMC theory (Fig. 1A, B).

Common and different regions are involved in processing different crossmodal correspondences

Overall, no area was observed to show common effects for congruence across the distinct CMCs within this study. This finding points towards probable differences in the underlying information processing of different CMCs. Differential effects were observed in the SPL, cerebellum, HG, ACC, AnG and IPS. In all regions, the effects were different between the conditions of the two CMCs. For the pitch-size CMC, a greater activation was observed for congruent > incongruent stimulus presentations while enhanced activity was seen for congruent<incongruent stimulus pairs in the pitch-elevation CMC within identical areas. However, the stimulus congruence classification task showed clear behavioral effects, supporting the assumption that congruent stimulus presentations are overall perceived as belonging together. This finding is consistent with the possibility that the neural effects are related to the perceived congruency between the visual and acoustic stimuli in the different CMCs. Therefore, the disparate neural effects can be interpreted as correlates of different neural processes underlying the different CMCs.

The processing and integration of multimodal information is assumed to start already in the primary sensory system (Baier et al., 2006; Crottaz-Herbette & Menon, 2006; Werner & Noppeney, 2010). This likely explains our observation of effects within the HG in both CMCs. The HG, which is part of the primary acoustic system, demonstrated a modulation of acoustic processing based on the congruency of the visual stimulus in CMCs. Despite the fact that our visual stimuli were clear and informative on their own, we found that processing of the visual stimulus was clearly influenced by the acoustic stimulus (Gallace & Spence, 2006; McDonald et al., 2000; Misselhorn et al., 2016; Werner & Noppeney, 2010). This is particularly interesting since the acoustic stimulus is irrelevant in all tested CMCs. On the one hand acoustic processing was previously reported as suppressed in visually dominating tasks (Johnson & Zatorre, 2005; Schmid et al., 2011) on the other hand, the influential nature of irrelevant acoustic stimuli in visual information processing was also reported by other studies (McDonald et al., 2000; Regenbogen et al., 2018; Tonelli et al., 2017). Thus, our results are a likely indication that in CMCs, differential processing of congruent and incongruent stimuli already occurs in the early sensory systems. To our surprise, we found differential processing of congruent and incongruent stimulus combinations in the early sensory system for the different CMCs. Greater effects in the HG were found for congruent stimulus presentations in pitch-size and for incongruent stimulus presentations in pitch-elevation (Fig. 2).

The ACC was also involved in both CMCs, which probably reflects the different functions of this area in the dynamical interplay between top-down and bottom-up multisensory processing (Benedict et al., 2002; Crottaz-Herbette & Menon, 2006; Downar et al., 2002; Laurienti et al., 2003). In particular, the role of the AnG in multisensory integration is well established (Bonnici et al., 2016; Cabeza et al., 2012; Hölig et al., 2017). Further, the AnG is assumed to be crucial for the direction of attention in relation to memory (Cabeza et al., 2012; Humphreys & Ralph, 2015; Jablonowski & Rose, 2022; Regenbogen et al., 2018). The involvement of this area can be interpreted as recruitment of memory-based attention in the pitch-size CMC. We speculate that pitch-size congruent trials were mainly driven by bottom-up processes, as the AnG is part of the ventral parietal cortex (VPC) (Cabeza et al., 2012; Humphreys & Ralph, 2015; Kim, 2010). According to the attention to memory (AtoM) model postulated by Cabeza and colleagues, pitch-size congruencies can be assumed to be driven by a detection of cues as part of a memory based retrieval (Cabeza et al., 2011; 2012). According to Cabeza and colleagues, the VPC *'[...] mediates the bottom-up capture of attention by salient memory contents (bottom-up AtoM)'* (Cabeza et al., 2012, p. 6).

In contrast to pitch and size correspondences, top-down processes seem to be involved in pitch-elevation CMCs. The IPS was postulated to be part of top-down attentional processes as this region is part of the dorsal parietal cortex (Cabeza et al., 2011; 2012; Regenbogen et al., 2018). Furthermore, the IPS showed greater activations dependent on task difficulty (Regenbogen et al., 2018). As the information about the location of a visual stimulus can be assumed to be biased by the misleading acoustic stimulus within incongruent pitch-elevation trials, top-down controlled attention shifts were probably enforced within incongruent stimulus presentations in the pitch-elevation CMC (Chiou & Rich, 2012; Regenbogen et al., 2018; Salmi et al., 2009). On a theoretical level, the fMRI results supported the theory about statistical assumptions extracted from our environment as a general basis for different CMCs (Maimon et al., 2021; Pisanski et al., 2017; Spence, 2011, 2020; Spence & Sathian, 2020; Zeljko et al., 2019). This would explain the facilitation of memory driven processes, which seem to best fit the measured neural effects. As shown in other studies before, there seems to be a strong connection between pitches with certain sizes or pitches and the

spatial location of objects (Ben-Artzi & Marks, 1995; Bolam et al., 2022; Evans & Treisman, 2010; Jamal et al., 2017; Maimon et al., 2021; Pisanski et al., 2017; Tonelli et al., 2017). These connections are probably based on experiences from our daily lives (Bowling et al., 2017; Parise et al., 2014; Pisanski et al., 2017). However, this common basis result in differential neural processing in the two distinct CMCs. We can only speculate at this point that our task design allowed for an easier detection of congruent pitch-size presentations. The spatially differentiating incongruent pitch-elevation pairs on the other hand probably resulted in a confounded response detection processing (Bruns et al., 2014; Chiou & Rich, 2012; Maimon et al., 2021). Differences in processing the stimulus combinations in the two distinct CMCs may explain the lack of significant effects observed for the incongruent pitch-size and congruent pitch-elevation presentations, as well as the interaction testing for enhanced neural effects of congruent trials selectively in the pitch-elevation CMC. However, further studies are required to investigate these differences and clarify the findings.

Findings on language and magnitude driven effects in crossmodal correspondences

No statistically significant activation in our language specific ROIs was found within this study. Based on the findings from this study as well as previous research, the theories stating that pitch-elevation correspondences are driven by language, are not supported (McCormick et al., 2018; Parkinson et al., 2012). However, it is important to note that we cannot rule out any entanglements of language with other factors influencing the mechanisms underlying pitch and elevation correspondences.

When examining the theory of magnitude-driven correspondences, the effect within the left, instead of the proposed right, IPS exhibited nearly significant results in the interaction between pitch-size and pitch-elevation. Surprisingly, the effect showed significant results in relation to incongruent presentations when assessing the ROI in the main effect of pitch and elevation, contradicting our hypothesis. Hence, our data does not strongly support the existence of a shared foundation for correspondence effects driven by magnitude within or across different types of CMCs.

According to the theory of magnitude/intensity, an e.g., small visual stimulus would be located on the same side of a polar scale as a high-pitched tone on its respective scale. Thus, leading to an intrinsic matching of congruent pairs as they would be perceived as 'more' intense or 'greater' in magnitude when combined (Chang & Cho, 2015). Hence, incongruent pairs of audio-visual stimuli would be on their opposite sides of the polar scales and therefore perceived as less intense or lower in magnitude. While our findings in investigating magnitude and language effects in our chosen ROIs may be open to debate, they are in line with the relatively weak effects reported by McCormick et al. (2018) in their study on pitch-elevation correspondence.

Linking this study to a former fMRI study on pitch-elevation CMCs

It is important to note that McCormick et al. (2018) used a working memory task in their study on pitch-elevation correspondences, which might has obscured direct correspondence effects leading to no significant effects in a congruent versus incongruent fMRI contrast. Nevertheless, notable effects were observed when analyzing consecutive congruent versus consecutive incongruent trials. Even though we have to interpret these findings with caution, in regard of the magnitude and language hypothesis, both of our findings share some commonalities in terms of null or weak effects. On the other hand, we found similar effects in the right AnG, however probably originating in the pitch-size correspondence in our study. Rather than directly comparing or linking our outcomes to the study by McCormick et al. (2018), our aim was to start from a similar theoretical framework and employ similar ROIs to explore the congruence effect in different CMCs with a non-working memory related task design.

No effects of pitch and size and its variant with reduced difference

We hypothesized to find a modulation of effect strength related to the reduced difference of the contrast of pitches and sizes in our pitch-size CMCs. This hypothesis was not supported by the measured effects within this study. We did not observe common effects of congruence between pitch-size and its variant with reduced difference within our ROIs. According to the findings of a study by Chiou & Rich (2012), the mapping of acoustic and visual stimuli seems to be relative to the assignments of low and high pitches to e.g., small and large squares, however the ability to measure a CMC effect is dependent on the ability to form distinct pairs which clearly stand out from each other in the context they are presented in (Chiou & Rich, 2012; Zeljko et al., 2019). The context in which the stimuli are presented in is thought to support the formation of congruent mappings of stimuli (Chiou & Rich, 2012;

Zeljko et al., 2019). The findings in our congruence classification task led to the assumption that a reduced difference of the contrast of our implemented stimuli did not significantly weaken the mentioned ability to associate the expected stimulus pairs as congruent or incongruent. We observed a differential activity in the ACC of pitch-size compared to its variant. This effect was observed when we performed an interaction analysis that tested for enhanced neural effects of congruent trials selectively in the pitch-size CMC. However, this activity did not survive multiple comparison corrections. Nevertheless, this observation suggests that differences in processing congruency may be involved in pitch-size compared to its variant. The slowed down RTs in the pitch-size variant with reduced difference compared to the pitch-size pairs with a great difference in stimulus contrasts probably hints towards a slowed down decision process. It remains speculative whether a lack in neural modulatory effects was caused by the great similarity of the two pitch-size CMCs, as well as the higher similarity of stimuli within the variant of the pitch-size with reduced difference.

Limitations of this study

In the present study are some limitations we want to address. Instead of performing the CMC presentations in an upright position like in classical behavioral measurements, the participants lay on their back during the main experiment in our study. The perceived upright should not have inferred with the pitch-elevation matching of congruence or not matching of incongruence as the presented visual and acoustic stimuli were aligned with the body position (Harris et al., 2015), nevertheless we cannot exclude a probable influential factor of the body position on the perception of stimulus positions within the pitch-elevation part of the experiment in the scanner. While we decided to use natural stimuli instead of pure tones, which allowed for a more natural representation of crossmodal correspondences, it is worth acknowledging the potential influence of timbre on our findings. We matched the instruments and their produced tones, i.e., higher tones are played by smaller instruments and lower tones are played by larger instruments within our distinct CMCs and the pitch-size variant with reduced difference. This alignment potentially amplified the correspondence effect observed in the pitch-size CMC. Previous studies, such as Evans and Treisman (2010), successfully utilized piano and violin sounds in their indirect crossmodal correspondence tasks without encountering issues related to the instruments used. However, it is worth noting that while there are studies exploring crossmodal correspondences involving timbre and other perceptual dimensions (Adeli et al., 2014; Qi et al., 2020; Wallmark & Allen, 2020), to our knowledge, none have specifically investigated the role of timbre in pitch-size or pitch-elevation correspondences using the stimuli employed in our study. While we believe our results provide valuable insights into the mechanisms hidden behind different crossmodal correspondences, it would be beneficial to replicate our study with an alternative experimental setup to ascertain the robustness of our conclusions. It would be for example from great interest if the same results we find in this study can be replicated when utilizing the same task for e.g., different crossmodal correspondences. Regarding the null effects we observed in our study, it is important to emphasize that the absence of significant findings does not necessarily imply the absence of an effect. Further investigation with a different paradigm may provide a better understanding of the effect. Although CMCs have been successfully studied using an implicit task design in behavioral studies (Evans, 2020; Evans & Treisman, 2010; Parise & Spence, 2012, Chiou & Rich, 2012), exploring differences between the processing of CMCs in implicit and explicit task designs may be of interest for future fMRI studies. For example, a recent study on sound-symbolic CMCs (Barany et al., 2023) showed that significant congruence effects occur in an explicit design, whereas an implicit task design led to significant incongruence effects in sound-symbolic CMCs (Mccormick et al., 2021; Peiffer-Smadja & Cohen, 2019). In future studies of pitch-based CMCs, it would also be worthwhile to test other stimuli, increase the number of trials, or expand the sample size to gain a more comprehensive understanding of the neural mechanisms involved in CMC effects.

CONCLUSION

This study aimed to elucidate the neural processing of different pitch-based correspondences. The results within our study argue in favor of the idea that different mechanisms drive the integration of stimulus features within different audio-visual CMCs. The strong matching of pitches with their congruent spatial location probably led to greater top-down attentional driven processes during incongruent stimulus presentations. On the other hand, attention as well as memory retrieval seem to be crucial for pitch-size correspondences. Our results support the findings of previous studies assuming both top-down and bottom-up processes are involved in forming the CMC effect in particular and multimodal integration in general (Getz & Kubovy, 2018; Salmi et al., 2009).

ACKNOWLEDGMENTS

The present research was supported by the German Research Foundation (DFG; SFB/Transregio 169, Project B3). The authors would like to thank Marike Maack for her helpful comments on the manuscript.

M.R. and C.J. designed the study. C.J. collected the data, C.J. and M.R. analyzed the data and C.J. wrote the manuscript under supervision of M.R. All authors reviewed the manuscript.

The datasets generated and/or analyzed during this study are available on request from the corresponding author.

REFERENCES

Adeli, M., Rouat, J., & Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00352

Baier, B., Kleinschmidt, A., Müller, N. G., & Muller, N. G. (2006). Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *Journal of Neuroscience*, *26*(47), 12260–12265.

Barany, D. A., Lacey, S., Matthews, K. L., Nygaard, L. C., & Sathian, K. (2023). Neural basis of sound-symbolic pseudoword-shape correspondences. *Neuropsychologia*, *188*, 108657. https://doi.org/10.1016/j.neuropsychologia.2023.108657

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, *57*(8), 1151–1162.

Ben-Artzi, E., & Marks, L. E. (1999). Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 579–595.

Benedict, R. H. B., Shucard, D. W., Maria, M. P. S., Shucard, J. L., Abara, J. P., Coad, M. Lou, Wack, D., Sawusch, J., & Lockwood, A. (2002). Covert Auditory Attention Generates Activation in the Rostral/Dorsal Anterior Cingulate Cortex. *Journal of Cognitive Neuroscience*, *14*(4), 637–645.

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size. Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663–672. https://doi.org/10.1016/j.neuroimage.2011.06.095

Bolam, J., Boyle, S. C., Ince, R. A. A., & Delis, I. (2022). Neurocomputational mechanisms underlying cross-modal associations and their influence on perceptual decisions. *NeuroImage*, *247*, 1–32.

Bonetti, L., & Costa, M. (2018). Pitch-verticality and pitch-size cross-modal interactions. *Psychology of Music*, *46*(3), 340–356.

Bonnici, H. M., Richter, F. R., Yazar, Y., & Simons, J. S. (2016). Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. *The Journal of Neuroscience,* 36(20), 5462–5471.

Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., & Frommolt, K. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, *7*, 41070.

Bruns, P., Maiworm, M., & Röder, B. (2014). Reward expectation influences audiovisual spatial integration. *Attention, Perception & Psychophysics*, *76*(6), 1815–1827. https://doi.org/10.3758/s13414-014-0699-y

Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2011). Overlapping parietal activity in memory and perception: evidence for the attention to memory model. *Journal of cognitive neuroscience*, *23*(11), 3209–3217. https://doi.org/10.1162/jocn_a_00065

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends Cogn Sci.*, *16*(6), 338–352. https://doi.org/10.1016/j.tics.2012.04.008.

Chang, S., & Cho, Y. S. (2015). Polarity correspondence effect between loudness and lateralized response set. *Frontiers in Psychology*, *6*, 683. https://doi.org/10.3389/fpsyg.2015.00683

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*(3), 339–353. https://doi.org/10.1068/p7161

Crottaz-Herbette, S; Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response : Combined fMRI and ERP Evidence. *Journal of Cognitive Neuroscience*, *18*(5), 766–780.

Downar, J., Crawley, A. P., Mikulis, D. J., Davis, K. D., Crawley, A. P., & Mikulis, D. J. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *Journal of Psychology*, *87*, 615–620.

Evans, K. K. (2020). The role of selective attention in cross-modal interactions between auditory and visual features. *Cognition*, *196*, 104119. https://doi.org/10.1016/j.cognition.2019.104119

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 1–12. https://doi.org/10.1167/10.1.6

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception and Psychophysics*, *68*(7), 1191–1203. https://doi.org/10.3758/BF03193720

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition*, *175*, 101–108. https://doi.org/10.1016/j.cognition.2018.02.015

Harris, L. R., Carnevale, M. J., D'Amour, S., Fraser, L. E., Harrar, V., Hoover, A. E. N., Mander, C., & Pritchett, L. M. (2015). How our body influences our perception of the world. *Frontiers in Psychology*, *6*, 1–10. https://doi.org/10.3389/fpsyg.2015.00819

Hölig, C., Föcker, J., Best, A., Röder, B., & Büchel, C. (2017). Activation in the angular gyrus and in the pSTS is modulated by face primes during voice recognition. *Human Brain Mapping*, *2565*, 2553–2565. https://doi.org/10.1002/hbm.23540

Humphreys, G. F., & Ralph, M. A. L. (2015). Fusion and fission of cognitive functions in the human parietal cortex. *Cerebral Cortex*, *25*, 3547–3560. https://doi.org/10.1093/cercor/bhu198

Jablonowski, J., & Rose, M. (2022). The functional dissociation of posterior parietal regions during multimodal memory formation. *Human Brain Mapping, 43*(11), 3469-3485. https://doi.org/10.1002/hbm.25861

Jamal, Y., Lacey, S., Nygaard, L., & Sathian, K. (2017). Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception. *Multisensory Research*, *30*(3–5), 287–306. https://doi.org/10.1163/22134808-00002553

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events : Behavioral and neural correlates. *Cerebral Cortex*, *15*, 1609–1620. https://doi.org/10.1093/cercor/bhi039

Kim, H. (2010). NeuroImage Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. *NeuroImage*, *50*(4), 1648–1657. https://doi.org/10.1016/j.neuroimage.2010.01.051

Koten, J. W., Langner, R., Wood, G., & Willmes, K. (2013). Are reaction times obtained during fMRI scanning reliable and valid measures of behavior? *Experimental Brain Research*, *227*(1), 93–100. https://doi.org/10.1007/s00221-013-3488-2

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., & Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Human Brain Mapping*, *19*(4), 213–223. https://doi.org/10.1002/hbm.10112

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686

Liuzzi, A. G., Bruffaerts, R., Peeters, R., Adamczuk, K., Keuleers, E., De Deyne, S., Storms, G., Dupont, P., & Vandenberghe, R. (2017). Cross-modal representation of spoken and written word meaning in left pars triangularis. *NeuroImage*, *150*, 292–307. https://doi.org/10.1016/j.neuroimage.2017.02.032

Maimon, N. B., Lamy, D., & Eitan, Z. (2021). Space oddity: musical syntax is mapped onto visual space. *Scientific Reports*, *11*(1), 1–17. https://doi.org/10.1038/s41598-021-01393-1

Marks, L. E. (1987). On Cross-Modal Similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394. https://doi.org/10.1037/0096-1523.13.3.384

McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., & Sathian, K. (2018). Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia*, *112*, 19–30. https://doi.org/10.1016/j.neuropsychologia.2018.02.029

Mccormick, K., Lacey, S., Stilla, R., & Nygaard, L. C. (2021). Neural Basis of the Sound-Symbolic Crossmodal Correspondence Between Auditory Pseudowords and Visual Shapes. https://doi.org/10.1163/22134808-bja10060

McDonald, John J; Teder-Sälejärvi, Wolfgang A., Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*, 906–908. https://doi.org/10.1038/35038085

Melara, R. D., & Brien, T. P. O. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology*, *116*(4), 323–336.

Misselhorn, J., Daume, J., Engel, A. K., & Friese, U. (2016). A matter of attention : Crossmodal congruence enhances and impairs performance in a novel trimodal matching paradigm. *Neuropsychologia*, *88*, 113–122. https://doi.org/10.1016/j.neuropsychologia.2015.07.022

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*(3–4), 319–333. https://doi.org/10.1007/s00221-012-3140-6

Parise, C. V, Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(16), 6104–6108. https://doi.org/10.1073/pnas.1322705111

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception*, *41*(7), 854–861. https://doi.org/10.1068/p7225

Peiffer-Smadja, N., & Cohen, L. (2019). The cerebral bases of the bouba-kiki effect. *NeuroImage*, 186, 679–689. https://doi.org/10.1016/j.neuroimage.2018.11.033

Piazza, M., Pinel, P., Bihan, D. Le, & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*, 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Pinel, P., Piazza, M., Bihan, D. Le, Dehaene, S., Unit, I., & Ge, P. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*, 983–993.

Pisanski, K., Isenstein, S. G. E., Montano, K. J., O'Connor, J. J. M., & Feinberg, D. R. (2017). Low is large: spatial location and pitch interact in voice-based body size estimation. *Attention, Perception, and Psychophysics*, *79*(4), 1239–1251. https://doi.org/10.3758/s13414-016-1273-6

Qi, Y., Huang, F., Li, Z., & Wan, X. (2020). Crossmodal correspondences in the sounds of Chinese instruments. *Perception*, *49*(1), 81–97. https://doi.org/10.1177/0301006619888992

Regenbogen, C., Seubert, J., Johansson, E., Finkelmeyer, A., Andersson, P., & Lundström, J. N. (2018). The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Human Brain Mapping*, *39*(3), 1313–1326. https://doi.org/10.1002/hbm.23918

Roelofs, A., van Turennout, M., & Coles, M. G. (2006). Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(37), 13884–13889. https://doi.org/10.1073/pnas.0606265103

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *Journal of Neuroscience*, *29*(20), 6490–6499. https://doi.org/10.1523/JNEUROSCI.5437-08.2009

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., & Alho, K. (2009). Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain Research*, *1286*, 155–164. https://doi.org/10.1016/j.brainres.2009.06.083

Schmid, C., Büchel, C., & Rose, M. (2011). The neural basis of visual dominance in the context of audio-visual object processing. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2010.11.051

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics* (Vol. 73, Issue 4, pp. 971–995). https://doi.org/10.3758/s13414-010-0073-7

Spence, C. (2020). Simple and complex crossmodal correspondences involving audition. *Acoustical Science and Technology*, *41*(1), 6–12. https://doi.org/10.1250/ast.41.6

Spence, C., & Sathian, K. (2020). Audiovisual crossmodal correspondences: behavioral consequences and neural underpinnings. In Sathian, K and Ramachandran, VS (Ed.), *Multisensory Perception: From Laboratory to Clinic* (pp. 239–258). https://doi.org/10.1016/B978-0-12-812492-5.00011-5

Tonelli, A., Cuturi, L. F., & Gori, M. (2017). The influence of auditory information on visual size adaptation. *Frontiers in Neuroscience*, *11*(594). https://doi.org/10.3389/fnins.2017.00594

Uno, K., & Yokosawa, K. (2022a). Cross-modal correspondence between auditory pitch and visual elevation modulates audiovisual temporal recalibration. *Scientific Reports*, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-25614-3

Uno, K., & Yokosawa, K. (2022b). Pitch-elevation and pitch-size cross-modal correspondences do not affect temporal ventriloquism. *Attention, Perception, and Psychophysics*, 84, 1052–1063. https://doi.org/10.3758/s13414-022-02455-w

Wallmark, Z., & Allen, S. E. (2020). Preschoolers' crossmodal mappings of timbre. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-020-02015-0

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience*, *30*(7), 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010

Zeljko, M., Kritikos, A., & Grove, P. M. (2019). Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects. *Attention, Perception, and Psychophysics*, *81*(5), 1609–1623. https://doi.org/10.3758/s13414-019-01668-w