NEURAL BASIS OF SOUND-SYMBOLIC PSEUDOWORD-SHAPE CORRESPONDENCES

Deborah A. Barany[1], Simon Lacey[2,3,4], Kaitlyn L. Matthews[5,6], Lynne C. Nygaard[5] and K. Sathian[2,3,4]

[1]Department of Kinesiology, University of Georgia and Augusta University/University of Georgia Medical Partnership, Athens, GA, 30602, USA; Department of [2]Neurology, [3]Neural & Behavioral Sciences and [4]Psychology, Penn State Colleges of Medicine and Liberal Arts, Hershey, PA 17033-0859, USA; [5]Department of Psychology, Emory University, Atlanta, GA 30322, USA; [6]Present address: Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO 63130

Corresponding author:
K. Sathian
Department of Neurology
Milton S. Hershey Medical Center
Penn State College of Medicine
Hershey, PA 17033-0859, USA
Tel: 717-531-1801
Fax: 717-531-0384
Email: ksathian@pennstatehealth.psu.edu

HIGHLIGHTS

- fMRI investigation of sound-symbolic correspondences between auditory pseudowords and visual shapes
- Faster reaction times for congruent than incongruent audiovisual stimuli
- Greater activation in auditory and visual cortices for congruent stimuli
- Higher classification accuracy for congruent stimuli in language and visual areas
- Sound symbolism involves language processing and multisensory integration

ABSTRACT
Non-arbitrary mapping between the sound of a word and its meaning, termed sound symbolism, is commonly studied through crossmodal correspondences between sounds and visual shapes, e.g., auditory pseudowords, like ‘mohloh’ and ‘kehteh’, are matched to rounded and pointed visual shapes, respectively. Here, we used functional magnetic resonance imaging (fMRI) during a crossmodal matching task to investigate the hypotheses that sound symbolism (1) involves language processing; (2) depends on multisensory integration; (3) reflects embodiment of speech in hand movements. These hypotheses lead to corresponding neuroanatomical predictions of crossmodal congruency effects in (1) the language network; (2) areas mediating multisensory processing, including visual and auditory cortex; (3) regions responsible for sensorimotor control of the hand and mouth. Right-handed participants (n = 22) encountered audiovisual stimuli comprising a simultaneously presented visual shape (rounded or pointed) and an auditory pseudoword (‘mohloh’ or ‘kehteh’) and indicated via a right-hand keypress whether the stimuli matched or not. Reaction times were faster for congruent than incongruent stimuli. Univariate analysis showed that activity was greater for the congruent compared to the incongruent condition in the left primary and association auditory cortex, and left anterior fusiform/parahippocampal gyri. Multivoxel pattern analysis revealed higher classification accuracy for the audiovisual stimuli when congruent than when incongruent, in the pars opercularis of the left inferior frontal (Broca’s area), the left supramarginal, and the right mid-occipital gyri. These findings, considered in relation to the neuroanatomical predictions, support the first two hypotheses and suggest that sound symbolism involves both language processing and multisensory integration.

KEYWORDS: multisensory, sound symbolism, language, fMRI, multivoxel pattern analysis

INTRODUCTION

Sound symbolism refers to the idea that the sounds of a word resemble its meaning (Nuckolls, 1999). This idea dates back to Plato's Cratylus dialog (Ademollo, 2011), but has receded in modern times in favor of the mainstream view that sound-meaning associations are essentially arbitrary (e.g. Pinker, 1999). Recently, however, there has been a resurgence of interest in the phenomenon, fueled by the discovery of crossmodal correspondences between the sounds of words and (non-auditory) perceptual dimensions. For instance, auditory pseudowords such as 'takete' or 'kiki' are reliably matched with pointed visual shapes, whereas 'maluma' or 'bouba' are matched with rounded shapes (Köhler, 1929, 1947; Ramachandran & Hubbard, 2001). Further, the perceptual ratings of auditory pseudowords as rounded or pointed relate not only to phonological features (e.g., McCormick et al., 2015; Fort et al., 2015; Cuskley et al., 2017), but also to their acoustic properties, including spectrotemporal parameters of the pseudoword sounds and parameters of vocal roughness associated with sound production (Lacey et al., 2020).

Prior functional magnetic resonance imaging (fMRI) studies measuring the blood oxygenation level-dependent (BOLD) signal investigated the neural basis of sound symbolism using a variety of tasks across different semantic domains, without much consistency of findings. Revill et al. (2014) asked native English speakers to listen to words drawn from multiple foreign languages and match their meanings with English antonym pairs spanning domains of motion, size, and shape. In this study, greater activity in the left intraparietal sulcus for sound-symbolic compared to non-sound-symbolic words was considered to reflect crossmodal correspondences between object properties and word sounds, based on the recruitment of a nearby area during audiovisual synesthetic experiences (Neufeld et al., 2012). Visually presented Japanese mimetic words for motion and shape, when they matched visual images of actions or shapes, evoked activity in the right posterior superior temporal sulcus; this was construed as indicating conceptual integration of language and non-language sounds (Kanero et al., 2014). Judging the size of visual objects during presentation of sound symbolically incongruent, relative to congruent, pseudowords led to greater activity in the right anterior superior temporal gyrus and the left posterior middle temporal gyrus: the authors interpreted their data in terms of reliance on phonological and semantic processes (Itagaki et al., 2019). Congruent tactile hard/soft stimuli and visually presented Japanese words that were sound-symbolic for hardness/softness, relative to incongruent stimulus pairs, led to activity in areas involved in tactually assessing softness: the anterior insula and medial superior frontal gyrus bilaterally (Kitada et al., 2021). These regions also distinguished between congruent and incongruent stimuli in a multivoxel pattern analysis (MVPA) (Kitada et al., 2021), though the reported analyses did not explore whether individual stimuli could be distinguished. Collectively, these studies drew inferences about the relevant neural processes, but they were not designed to evaluate alternative explanations.

In the sound-symbolic shape domain, two previous fMRI studies presented participants with concurrent auditory pseudowords and visual shapes. Peiffer-Smadja & Cohen (2019) reported greater activity for incongruent than congruent audiovisual pairs in the dorsolateral prefrontal cortex bilaterally in a task where participants were asked to detect occasional target stimuli unrelated to the pseudoword-shape correspondence of interest, possibly reflecting greater effort during processing of the incongruent stimuli. In a prior study from our group,

McCormick et al. (2022) found that incongruent stimuli elicited stronger BOLD responses than congruent stimuli in frontoparietal regions when participants attended to the auditory input; these findings were thought to be consistent with changes in phonological processing and/or multisensory attention as a function of sound-symbolic congruency (McCormick et al., 2022). Although this study was conducted to arbitrate between Spence's (2011) accounts of crossmodal correspondences as originating in audiovisual statistical regularities, a domain-general magnitude system, or similarity of the pseudowords to real words, clear evidence did not accrue for or against any of these accounts.

In the present fMRI study, we investigated the neural correlates of sound symbolism for shape when participants were asked to explicitly assess the sound-symbolic audiovisual pseudoword-shape correspondence. Effects favoring congruent over incongruent stimuli were not observed in the two studies cited above, which used implicit matching of auditory pseudowords and visual shape (Peiffer-Smadja & Cohen, 2019; McCormick et al., 2022), but did occur during explicit matching of visually presented words to visual or tactile stimuli (Kanero et al., 2014; Kitada et al., 2021). Thus, we reasoned that an explicit judgment task, by focusing attention on whether the stimuli matched or not, would be more likely to elicit congruency effects. In addition to examining univariate BOLD effects, we also performed MVPA. We predicted that classification accuracy for specific pseudoword-shape combinations would be higher in congruent than incongruent trials and that regions distinguishing between congruent and incongruent stimuli would be distinct from those differentiating between specific stimulus pairs.

Our experimental approach allowed us to examine three hypotheses, which are not necessarily mutually exclusive, regarding the processes involved in perceiving sound-symbolic audiovisual pseudoword-shape correspondences:

Hypothesis 1 draws on a large body of psychophysical work implicating sound symbolism in language structure and function, e.g. the ubiquity of sound-symbolic associations across multiple languages (Blasi et al., 2016), the higher frequency of rounded (pointed) phonemes in words signifying rounded (pointed) objects (Sidhu et al., 2021), the facilitation of word learning (Nielsen & Dingemanse, 2021; Nygaard et al., 2009) and associative sound-shape memory (Sonier et al., 2020) by sound symbolism, the human ability to create sound-symbolic words for actions seen in locomotion videos (Saji et al., 2019), the capability of English speakers to infer Pokemon types based on sound-symbolic Japanese names (Kawahara et al., 2021), and the relative advantage for onomatopoeic words in aphasia (Meteyard et al., 2015). Such findings fit with Hypothesis 1, that *sound symbolism recruits language processing*, and lead to the neuroanatomical prediction that congruency effects would be found in classical language areas. While this hypothesis draws some support from prior neuroimaging studies (Kanero et al., 2014; Itagaki et al., 2019; see above), these studies did not evaluate alternative possibilities.

Hypothesis 2 proposes that *multisensory integrative processes underlie sound-symbolic audiovisual correspondences*, due to statistical associations between the auditory and visual stimuli (Spence, 2011; Fort & Schwartz, 2022). Due to these associations, congruent audiovisual pairings lead to multisensory integration whereas incongruent pairings do not (Spence, 2011). As reviewed above, this hypothesis is consistent with the findings of Revill et al. (2014), although this study did not seek to assess other hypotheses. It predicts congruency effects not only in areas classically thought to mediate multisensory processing, but also in

areas traditionally regarded as unisensory, i.e. visual and auditory cortex, given the wealth of evidence for widely distributed multisensory interactions in the neocortex (Ghazanfar & Schroeder, 2006).

Hypothesis 3 is rooted in the idea of acoustic similarities between pseudowords that are matched to rounded or pointed shapes and the sounds of manual actions used to draw such shapes with a sound-producing pen (Margiotoudi & Pulvermüller, 2020). Studies in the size domain suggest a different kind of relationship between articulatory and manual movements: foreign-language words or pseudowords that sound symbolically imply small or large size are associated with a precision or power grip, respectively (Vainio & Vainio, 2021, 2022; Vainio et al., 2023). These studies gave rise to Hypothesis 3, that *sound symbolism is based on embodiment of speech in hand actions*. This hypothesis, not previously tested in neuroimaging work, makes the neuroanatomical prediction that brain regions mediating motor control and related somatosensory processing for the hand and mouth are candidate loci for congruency effects.

METHODS

Participants

Twenty-four people took part in this study (14 female, 10 male; mean age = 22.6 years, SD = 3.8 years). All participants 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 written informed consent and were compensated for their time. All procedures were approved by the Emory University Institutional Review Board and complied with the Declaration of Helsinki. Participants performed three scanning sessions; in the first two, auditory pseudowords and visual shapes were presented in separate sessions (which are the subject of a separate report), and the third comprised the audiovisual task reported here. Two participants with poor behavioral performance (see *Multisensory fMRI task*) were excluded from further analyses.

Stimuli

We created sound-symbolic audiovisual stimuli, each consisting of an auditory pseudoword and a two-dimensional visual shape. Two auditory pseudowords (‘mohloh’ and ‘kehteh’) and two shapes (pointed and rounded) were chosen from large stimulus sets (McCormick et al., 2015, see below; McCormick, unpublished data, see below: Figure 1A). The pseudowords and shapes chosen were near the extremes of independent rounded and pointed dimensions based on empirical ratings for 537 pseudowords (McCormick et al., 2015) and 90 visual shapes (McCormick, unpublished data). These independent rating scales were converted to a single scale on which ‘most rounded’ = 1 and ‘most pointed’ = 7; ratings for the pseudowords ranged from 2.3 – 5.8 (median 4.2) and for the shapes from 1.6 – 6.7 (median 4.4). ‘Mohloh’ was the 12th most rounded pseudoword, rated 2.8, but we used the 20th most pointed pseudoword, ‘kehteh’ (rated 5.4), in order to get the closest match in duration. We used the 10th most rounded visual shape, rated 2.3, and the 10th most pointed visual shape, rated 6.4, to match the number of protuberances (five) between the shapes. A more detailed analysis of the pseudowords and shapes can be found in Lacey et al. (2020) and the stimuli themselves are available at https://osf.io/ekpgh/.

The pseudowords were digitally recorded in a female voice using Audacity v2.0.1 (Audacity Team, 2012), with a SHURE 5115D microphone and an EMU 0202 USB external sound card, at a 44.1 kHz sampling rate. The recordings were then processed in Sound Studio (Felt Tip Inc., NY), using standard tools and default settings, edited into separate files, amplitude-normalized, and down-sampled to a 22.05 kHz sampling rate (standard for speech). Stimulus duration was 563 ms for 'mohloh' and 546 ms for 'kehteh', resulting in a difference in duration of 17ms or approximately 3%. Because sound segments in language naturally differ in duration, these differences were retained. The visual shapes were solid gray shapes on a black background, each subtending approximately 1° of visual angle and presented at the center of the screen for 500ms. Duration of visual shape presentation was not varied to match the accompanying pseudowords, to avoid having the shape duration vary as a function of congruency and thus introduce a confound.

Stimuli were presented concurrently in audiovisual pairs (Figure 1A) that were either congruent ('mohloh'/rounded shape or 'kehteh'/pointed shape) or incongruent ('kehteh'/rounded shape or 'mohloh'/pointed shape) with respect to the crossmodal pseudoword-shape (sound-symbolic) correspondence. Stimuli were delivered via MATLAB 2014b (The MathWorks Inc., Natick, MA) to synchronize delivery with fMRI acquisition and also record responses and reaction times (RTs). 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 Sensimetrics S14 headphones (Sensimetrics Corporation, Gloucester MA).

Multisensory fMRI task

We used a jittered event-related design in which participants performed 12 runs of the audiovisual task. Each run consisted of 22 congruent trials and 22 incongruent trials, each contained within the TR of 2 s (see Image acquisition, below: Figure 1B), giving a trial sample size, across runs, of 132 for each of the four unique audiovisual pairs (see Figure 1A). This number of trials per condition is over 3 times the upper end of the typical range (10-40) and is large enough to permit condition-specific modeling of responses (Chen et al., 2022). Inter-trial intervals (ITIs), or null periods, ranged from 2 s to 30 s. We used optseq2 (https://surfer.nmr.mgh.harvard.edu/optseq/; Dale, 1999; Dale et al., 1999) to schedule trials and ITIs, in order to optimize the estimation of the hemodynamic response. Each run began and ended with a rest period of 10 s for a total run duration of 326 s. Pseudoword-shape pairs were presented in a fixed pseudorandom order within each run; the 12 run orders were counterbalanced across participants. Participants attended to each audiovisual stimulus pair and judged whether the pair was a match or a mismatch, pressing one of two buttons on a hand-held response box to indicate their response. Participants were free to make their own judgments of match/mismatch; they were not informed of our interest in studying sound symbolism. The right index and middle fingers were used to indicate match/mismatch responses, counterbalanced across subjects.

All participants, except two, judged the 'match' condition to be congruent pairs of stimuli and the 'mismatch' condition to be incongruent pairs of stimuli. These two participants responded with the opposite than expected mapping for at least one of the pseudoword-shape pairs; their data were excluded from analysis.

Image acquisition

MR scans were performed on a 3 Tesla Siemens Trio TIM whole body scanner (Siemens Medical Solutions, Malvern, PA), using a 32-channel head coil. T2*-weighted functional images were acquired using a single-shot, gradient-recalled, echoplanar imaging (EPI) sequence for BOLD contrast. For functional scans, 34 axial slices of 3.1 mm thickness were acquired using the following parameters: repetition time (TR) 2000 ms, echo time (TE) 30 ms, field of view (FOV) 200 mm, flip angle (FA) 90°, in-plane resolution 3.125×3.125 mm, and in-plane matrix 64×64. High-resolution 3D T1-weighted (T1w) anatomic images were acquired using an MPRAGE sequence (TR 2300 ms, TE 3.9ms, inversion time 1100 ms, FA 8°) comprising 176 sagittal slices of 1 mm thickness (FOV 256 mm, in-plane resolution 1×1 mm, in-plane matrix 256×256). Once magnetic stabilization was achieved in each run, the scanner triggered the computer running MATLAB and the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) so that the experiment was synchronized with scan acquisition.

Image preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* version 20.0.6, (Esteban et al., 2019; Esteban et al., 2020; RRID:SCR_016216), which is based on *Nipype* 1.4.2 (Gorgolewski et al., 2011; RRID:SCR_002502). Many internal operations of *fMRIPrep* use *Nilearn* 0.6.2 (Abraham et al., 2014, RRID:SCR_001362], mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation (https://fmriprep.readthedocs.io/en/latest/workflows.html). The T1w image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008; RRID:SCR_004757), and used as the T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM) and gray matter (GM) was performed on the brain-extracted T1w using `fast` (Zhang et al., 2001; FSL 6.0.3:b862cdd5, RRID:SCR_002823).

For each of the 12 BOLD runs per subject, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a customized implementation of *fMRIPrep*. Susceptibility distortion correction (SDC) was omitted. The BOLD reference was then co-registered to the T1w reference using `flirt` (FSL 6.0.3:b862cdd5, Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009) cost function. Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before spatiotemporal filtering using mcflirt (FSL 6.0.3:b862cdd5, Jenkinson et al., 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20190316 (Cox, 1996, RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head motion. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD.

Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), the derivative of the root mean squared variance over voxels (DVARS), and three region-wise global signals. FD and DVARS were calculated for each functional run, using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals were extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors was extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components were estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128 s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components were then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask was obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components were calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components were also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values were retained, such that the retained components' time series were sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components were dropped from consideration.

The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings were performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` from FreeSurfer.

fMRI data analysis

Statistical analyses of fMRI data were conducted using SPM12 (Wellcome Trust Center for Neuroimaging, London, UK) as implemented in *Nipype*. We used a standard General Linear Model (GLM) to estimate separate regressors (betas) for each of the four conditions in each run. Stimulus onsets were modeled as a boxcar function with duration equal to that trial's reaction time (Grinband et al., 2008) and convolved with the canonical hemodynamic response function. The six translation and rotation head-motion parameter estimates were included as additional regressors of no interest.

A univariate analysis was conducted by first smoothing the preprocessed BOLD signal in native space with an 8 mm full-width half-maximum (FWHM) isotropic Gaussian kernel prior to the first-level analysis. We then performed a linear contrast to evaluate the difference in BOLD activity between the congruent ('mohloh'/rounded shape or 'kehteh'/pointed shape) and incongruent ('kehteh'/rounded shape or 'mohloh'/pointed shape) conditions. The resulting

individual contrast maps were normalized to a standard MNI brain with 1×1×1 mm voxels and then submitted to a second-level analysis of group data with subject treated as a random factor (so that the degrees of freedom equal n-1, i.e., 21), followed by pairwise contrasts. An explicit mask based on a combined gray matter and white matter mask from segmentation of the T1w images was applied at the group level to isolate within-brain activations. Correction for multiple comparisons (p <0.05) was achieved using the topological FDR correction method, using whole brain activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$ (Chumbley et al., 2010).

Searchlight multivariate pattern analyses (MVPA) were performed using The Decoding Toolbox (TDT) in SPM (Hebart et al., 2015). The unsmoothed (Misaki et al., 2013) GLM run-wise beta estimates for each of the four conditions were first z-scored voxelwise and within each run to factor out univariate effects, avoid spurious correlations between mean estimates across runs, and improve classification accuracy (Lee & Kable, 2018; Stehr et al., 2023). We implemented two types of cross-validated, leave-one-run-out searchlight decoding analyses using spherical searchlights with a 4-voxel radius and an L2-norm regularized support vector machine classifier with $C = 1$ from the LIBSVM package (Chang & Lin, 2011). For the *within-congruency classification*, the classifier was trained to discriminate between either the two congruent stimulus pairs (i.e., ‘kehteh’/pointed shape vs. ‘mohloh’/rounded shape) or the two incongruent ones (i.e., ‘mohloh’/pointed shape vs. ‘kehteh’/rounded shape) separately (Figure 1C). To directly test whether any region showed significantly better decoding when the stimuli were congruent relative to when they were incongruent (or vice versa), we contrasted the individual classification accuracy maps for the congruent conditions with those for the incongruent conditions.

To test for specific patterns of brain activity related to congruency independent of sensory characteristics, we performed a *between-congruency classification* (Kaplan et al., 2015; Man et al., 2015). To decode the visual shapes, classifiers were trained to distinguish between two of the four conditions in which the visual shapes were different but the auditory pseudoword was the same (e.g., ‘kehteh’/pointed shape vs. ‘kehteh’/rounded shape: Figure 1D-Shape). The classifier was then tested on the two remaining conditions containing the other pseudoword (e.g., ‘mohloh’/pointed shape vs. ‘mohloh’/rounded shape: Figure 1D-Shape). Regions that represent visual properties should show decoding of visual shape independent of the accompanying auditory pseudoword (e.g., similar patterns of activity for ‘kehteh’/rounded shape and ‘mohloh’/rounded shape). In contrast, in regions where responses are driven by the congruency of crossmodal correspondences (or the associated motor responses), we should be able to decode the congruency/incongruency of the stimuli (e.g., similar patterns of activity for ‘kehteh’/pointed shape and ‘mohloh’/rounded shape). Similarly, to decode congruency/incongruency across auditory features, we trained classifiers to distinguish between auditory pseudowords when the accompanying visual shape was held constant (e.g., ‘kehteh’/pointed shape vs. ‘mohloh’/pointed shape) and then tested on the two remaining conditions containing the other shape (e.g., ‘kehteh’/rounded shape vs. ‘mohloh’/rounded shape: Figure 1D-Pseudoword). As before, we should be able to differentiate regions where auditory features drive responses from regions that represent crossmodal congruency independent of the pseudoword presented.

For both the *within-* and *between-congruency* analyses, the individual classification accuracy searchlight maps were smoothed with a 4 mm full-width half-maximum (FWHM) isotropic Gaussian kernel and normalized to MNI space before second-level analysis of group data with subject as a random factor. All analyses included a combined GM/WM explicit mask to exclude non-brain voxels. For the *within-congruency* analysis comparing differences in classification accuracy for congruent > incongruent (incongruent > congruent), only voxels exhibiting significant above-chance classification of the congruent (incongruent) conditions at the group level were included. The group-level maps were corrected for multiple comparisons using the topological FDR correction method with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$ (Chumbley et al., 2010).

All activations were localized with respect to 3D cortical anatomy with the AAL (Tzourio-Mazoyer et al., 2002; Rolls et al., 2015) and AICHA (Joliot et al., 2015) atlases; differences between the two were resolved by reference to an MRI atlas (Cho et al., 2010). Activation maps are overlaid on a group-averaged structural image normalized to MNI space (1×1×1 mm voxels).

RESULTS

Behavioral data

Table 1 shows the mean accuracy and RTs for each of the four conditions. Accuracy was not significantly different for congruent vs. incongruent trials ($t_{21} = 0.261$; $p = 0.797$; $d = 0.056$). RTs were faster when the pseudoword and shape were sound symbolically congruent compared to when they were incongruent ($t_{21} = 5.374$; $p < 0.001$; $d = 1.15$).

fMRI univariate analyses

The contrast of congruent > incongruent conditions showed higher activation in the right, and near the left, caudate nucleus, as well as in the left anterior fusiform/parahippocampal gyri, right mid-cingulate gyrus/corpus callosum, left mid-superior temporal gyrus, and left postcentral gyrus. The reverse contrast (incongruent > congruent) showed greater activation in the left precentral sulcus, left supplementary motor area (SMA), bilateral anterior insula, and right pons. These activations were cluster-corrected for multiple comparisons (cluster-forming threshold $p < 0.001$; FDR cluster-level $q < 0.05$) (Figure 2 and Table 2).

fMRI searchlight MVPA

<u>Within-congruency classification</u>

In separate decoding analyses of the congruent and incongruent conditions, a whole-brain searchlight analysis showed robust above-chance classification accuracy in visual and auditory cortices, reflecting distinct patterns for unique visual shapes and auditory pseudowords, respectively (Figure 3; Table 3). To test where decoding of shape and pseudowords was better for congruent than incongruent trials, we contrasted the individual classification accuracy maps for the congruent condition with those for the incongruent condition. This contrast was restricted to those voxels that exhibited above-chance classification in the congruent condition. The resulting difference map (cluster-corrected for multiple comparisons with cluster-forming threshold $p < 0.001$; FDR cluster-level $q < 0.05$) revealed significantly higher above-chance classification in the congruent, compared to the incongruent, conditions in the pars opercularis

of the left inferior frontal gyrus (IFG) (Broca's area), the left supramarginal gyrus (SMG), and the right mid-occipital gyrus (MOG), (Figure 4A-C respectively; Table 4). Classifier performance differences between congruent and incongruent conditions in Broca's area and the left SMG resulted from above-chance (~60% accuracy) classification for the congruent, but not the incongruent, conditions (Figure 4A & B respectively). Classification accuracies at the peak coordinate in the right MOG cluster were above chance for both congruent and incongruent conditions, but significantly higher for the congruent condition, as shown by the bar graphs in Figure 4C. Note that the right MOG cluster was part of a large swath of occipital (early visual) cortex with significant classifier accuracy in the congruent condition. No region showed significantly higher above-chance classification of the incongruent conditions after correction for multiple comparisons.

Between-congruency classification

A whole-brain searchlight analysis revealed distinct regions for processing of the visual shapes and auditory pseudowords, and for processing their crossmodal correspondences (Figure 5, Tables 5 and 6). When classifiers were trained to distinguish the two visual shapes, significant above-chance decoding of visual shape was restricted to visual cortical regions with peak local maxima in the lingual gyrus and calcarine sulcus, whereas decoding of crossmodal congruency (i.e. congruent vs. incongruent) was significantly above chance in multiple regions, including left precentral gyrus, bilateral superior frontal gyri, right pallidum, right postcentral gyrus, left hippocampus, right superior parietal lobule, right middle frontal gyrus, left posterior cingulate gyrus, and right inferior frontal gyrus (Figure 5A, Table 5). A similar pattern of results was observed for classifiers trained to distinguish between the auditory pseudowords: only bilateral auditory cortex (with peak local maxima in the superior temporal sulcus/gyrus) displayed significant above-chance pseudoword decoding, whereas there was significant above-chance decoding of crossmodal congruency/incongruency in multiple bilateral parietofrontal and temporal regions, and on the right in the parahippocampal gyrus, cerebellum, thalamus, and anterior insula (Figure 5B, Table 6). Notably, congruency decoding in parietofrontal regions was more prominent in the left hemisphere (contralateral to the index and middle fingers used to pressing the buttons indicating match or mismatch responses).

DISCUSSION

Previous fMRI studies of the sound-symbolic correspondence between auditory pseudowords and visual shapes (McCormick et al., 2022; Peiffer-Smadja & Cohen, 2019) were limited to univariate analyses of task conditions involving implicit registration of the correspondence. The present study is the first fMRI investigation of this psychophysically well-studied sound-symbolic correspondence using an explicit audiovisual matching task and incorporating MVPA to supplement conventional univariate analyses. MVPA offers insights into much finer-grained information processing than is possible with conventional univariate analyses, and has not previously been applied to the sound-symbolic pseudoword-shape correspondence, or to decode particular stimulus combinations in neuroimaging studies of sound symbolism. Our study used an event-related design and had a slightly higher participant sample size (n=22) than the prior studies, both of which employed block designs (McCormick et al., 2022: n=19; Peiffer-Smadja & Cohen, 2019: n=18). Importantly, our study also had a high number of trials

(132) per unique condition, allowing for reliable condition-level modeling of responses and making a major contribution to statistical power (Chen et al., 2022). These features of experimental design and analysis enable novel insights into the neural basis of sound symbolism.

We found some neuroanatomical support for each of the hypotheses outlined in the Introduction, based on concordance between the areas showing univariate and multivariate congruency effects and the anatomical predictions derived from the hypotheses. As predicted *a priori*, MVPA classification accuracy was higher for congruent than incongruent conditions in certain areas, but the reverse was not true. **Hypothesis 1**, proposing that sound symbolism is associated with language processing, and thus predicting congruency effects in classical language areas, was supported by MVPA revealing superior classification performance of the bisensory stimulus pairs when congruent, compared to when they were incongruent, in regions of the classical language network: the pars opercularis of the left IFG, i.e., Broca's area, and the left SMG (Figure 4A & B respectively, Table 4). Our findings also support **Hypothesis 2**, suggesting that sound symbolism depends on multisensory integration and therefore recruits sensory cortical processes for congruent stimuli, in classical multisensory areas as well as areas regarded conventionally as unisensory. Evidence for this hypothesis comes from the MVPA producing higher classifier accuracy for congruent relative to incongruent conditions in early visual cortex (Figure 4C, Table 4), and higher univariate magnitudes of the BOLD signal for the congruent than the incongruent condition in primary and association auditory cortex and in a part of higher-order visual cortex (Figure 2, Table 2). There was limited support for **Hypothesis 3**, the notion that sound symbolism arises from embodiment of speech in manual actions, and the consequent prediction of congruency effects in motor and somatosensory regions related to the hand and mouth. The only data potentially consistent with this idea were the univariate demonstration of greater BOLD signal favoring the congruent over the incongruent condition in the hand area of primary somatosensory cortex (postcentral gyrus) (Figure 2, Table 2). However, the absence of the predicted concomitant activity in regions mediating motor control weakens support for this hypothesis.

Earlier univariate fMRI studies of the pseudoword-shape correspondence (McCormick et al., 2022; Peiffer-Smadja & Cohen, 2019) did not find regions that were preferentially active for congruent relative to incongruent stimuli, perhaps owing to reliance on implicit matching tasks. Here, our explicit matching task revealed univariate activations for congruent compared to incongruent stimuli, as Kitada et al. (2021) found in touch-related regions when participants judged whether tactile stimuli were matched or mismatched with Japanese mimetic words related to tactile hardness and softness.

A valuable feature of our MVPA was testing whether activity patterns were related to low-level sensory features or to processing congruency. We observed above-chance between-congruency classification of the two visual stimuli in visual cortical areas, regardless of which auditory stimulus was paired with them. Similar above-chance classification of the two pseudoword stimuli was observed in auditory cortical areas, regardless of the accompanying visual stimulus. These findings (Figure 5, Table 5 & 6) are not surprising, but provide direct evidence that such neural processing of low-level visual/auditory features is distinct from the processing of congruency discussed below. Classification of stimulus pairs as congruent or incongruent, regardless of the particular stimuli, was above chance in multiple brain regions.

Among these regions, frontoparietal areas in the left hemisphere were prominent, perhaps attributable to decoding of finger movements (Shen et al., 2014), since each participant always used one finger to indicate that the audiovisual stimuli matched (congruent), and another finger to indicate the stimuli were mismatched (incongruent). However, loci classifying congruent vs. incongruent extended beyond the motor network and may relate to variations in attentional demand or decision processes for distinguishing congruent from incongruent stimuli. Future work is needed to isolate regions sensitive to processing congruency independent of the motor response.

Sound symbolism involves language processing

As predicted by **Hypothesis 1**, MVPA classifier accuracy on the bisensory stimulus pairs was higher for congruent pairs in Broca's area and the left SMG, both regions of the classical language network (Figure 4A & B, Table 4). In these areas, the classifier yielded above-chance performance for congruent, but not incongruent stimuli. These findings support the idea that sound symbolism is associated with language processing.

Although Broca's area is involved in syntactic processing (Fedorenko & Blank, 2020; Kemmerer, 2022; Maran et al., 2022), a number of studies also implicate it in audiovisual associations relevant to language: For instance, high-gamma power in Broca's area was enhanced by congruent audiovisual speech, with a progressive increase in coherence between auditory cortex and Broca's area over the duration of the stimulus, whereas coherence between these regions decreased over time for incongruent audiovisual speech (Lange et al., 2013). Another study found that anodal transcranial direct current stimulation over Broca's area facilitated acquisition of novel associations between pseudowords and unfamiliar visual objects (Perikova et al., 2022). An fMRI study using MVPA found that viewing of congruent lip movements facilitated classification of phonemes in Broca's area, with a larger effect for phonemes distinguished by voicing (voiced vs. unvoiced consonants) than by place of articulation (Zhang and Du, 2022). Our finding of higher classification performance for congruent, compared to incongruent, pseudoword-shape pairs in Broca's area resonates with these earlier studies and provides evidence that processing of audiovisual congruency related to sound symbolism is indeed associated with language (Blasi et al., 2016; Kawahara et al., 2021; Meteyard et al., 2015; Nielsen & Dingemanse, 2021; Nygaard et al., 2009; Saji et al., 2019; Sidhu et al., 2021; Sonier et al., 2020).

Similar reasoning may apply to the left SMG, where again classification of the bisensory stimulus pairs was better when they were congruent rather than incongruent (Figure 4B, Table 4). The SMG has been implicated in language processing, particularly phonological processes (Hartwigsen et al., 2010; Oberhuber et al., 2016). Reinforcing this line of thought, Zhang and Du (2022) found that viewing congruent lip movements facilitated classification of phonemes in the left SMG in addition to Broca's area; however, in the SMG, classifier performance was significant for place of articulation but not voicing. The left SMG also demonstrates beta power suppression for incongruent audiovisual speech (Lange et al., 2013). Curiously, a number of studies have demonstrated stronger univariate activation for incongruent than congruent stimuli in both Broca's area (Ojanen et al., 2005; Pekkola et al., 2006; Szycik et al., 2009) and the left SMG (Bernstein et al., 2009; Xie et al., 2019; McCormick et al, 2022), which may reflect greater effort associated with processing of incongruent audiovisual stimuli related to speech. Though

we did not observe increased univariate activation in Broca's area or left SMG in the present study, our MVPA findings of superior classifier performance for congruent compared to incongruent stimuli suggest that the neuronal populations recruited in these areas are more distinct for congruent than incongruent audiovisual stimuli, independent of overall activation.

Involvement of multisensory integration in sound symbolism

In the right MOG, in early visual cortex, the MVPA classifier performed above chance for both congruent and incongruent stimulus pairs, with better performance for congruent pairings (Figures 3 & 4C, Tables 3 & 4). This is consistent with multisensory convergence with a preference for congruency, indicating a role for multisensory integration in sound symbolism, in accord with **Hypothesis 2**.

Some of the univariate activations for congruent relative to incongruent stimuli in the present study were in relevant sensory cortical areas (Figure 2, Table 2): One focus was in auditory cortex in the superior temporal gyrus, including primary auditory cortex (Moerel et al., 2014) and neighboring auditory association cortex (Malinowska et al., 2017). Another was in a higher-order region of visual cortex, in the anterior fusiform gyrus (see Kliger & Yovel, 2020), extending into the parahippocampal gyrus, which has been implicated in evaluating audiovisual congruency (Diaconescu et al., 2011). We also observed preferential bilateral activation near the caudate nucleus for congruent relative to incongruent audiovisual stimuli. Although the mechanism underlying this caudate nucleus activation is unclear, a similar preference for audiovisual congruency in the caudate nuclei has been found previously (Stevenson et al., 2009; McNorgan et al, 2015). Thus, these activations may reflect multisensory convergence (e.g., Schroeder et al., 2003), with audiovisual congruency leading to stronger activity in these regions, in keeping with **Hypothesis 2**.

Embodiment of sound symbolism in manual actions

The left somatosensory cortical focus in the postcentral gyrus displayed higher univariate activity for congruent relative to incongruent stimuli (Figure 2, Table 2). Its anatomical location was in the hand representation (Martuzzi et al., 2014), medial to both the face representation (Moulton et al., 2009) and the part of somatosensory cortex activated during listening to and repeating speech (Darainy et al., 2019). In keeping with **Hypothesis 3**, activity at this focus could be due to postulated embodiment of the roundedness/pointedness of auditory pseudowords in manual actions, given similarities between the sounds of shape-signifying pseudowords and of hand actions employed in drawing corresponding shapes (Margiotoudi & Pulvermüller, 2020). However, if so, it is curious that the primary motor cortex or areas mediating motor planning were not recruited. Thus, support for **Hypothesis 3** is weak, at least in the shape domain of sound symbolism explored here. It remains possible that this hypothesis may be more applicable to the size domain of sound symbolism, based on the line of work indicating the association of pseudowords that sound symbolically imply small or large size with a precision or power grip, respectively (Vainio & Vainio, 2021, 2022; Vainio et al., 2023); this may be worth testing in future neuroimaging studies.

Processing of incongruent audiovisual stimuli

Greater activity for incongruent, compared to congruent, bimodal stimuli was found in cortex of the left precentral sulcus (premotor cortex) and the left SMA (Figure 2, Table 2), both regions involved in motor planning. Recruitment of these regions cannot be attributed solely to RT differences in the motor response since RT duration for each trial was explicitly modeled in the GLM (Grinband et al., 2008). While these activations could be construed to support **Hypothesis 3**, this interpretation would be at odds with that offered in the preceding paragraph, and the preponderance of evidence in the present study for both univariate and multivariate effects favoring the congruent condition. An alternative possibility is that greater activation in motor areas for incongruent trials could be due to the reprogramming of a preplanned response to congruent stimuli (Hartwigsen et al., 2012; Pellegrino et al., 2018).

Finally, we found higher activity for incongruent relative to congruent stimuli in the dorsal region of the anterior insula bilaterally (Figure 2, Table 2). In a meta-analytic study, this region was functionally connected with the anterior cingulate cortex and dorsolateral prefrontal cortex, and was associated with higher cognitive tasks and executive control (Chang et al., 2013). Anterior insula activation may also reflect a domain-general processing of surprise (Loued-Khenissi et al., 2020). These activations may hence be related to those reported in earlier studies using implicit sound-symbolic matching, where bimodally incongruent stimuli elicited stronger activation than congruent stimuli in prefrontal cortical areas and the intraparietal sulcus, postulated to indicate neural processes related to attention or effort (McCormick et al., 2022; Peiffer-Smadja & Cohen, 2019).

Additional research is required to test the inferences made regarding processing of incongruent stimuli, since these were not associated with well-defined predictions.

Limitations

We used an explicit judgment task to uncover neural responses related to the processing of congruency that were not observed in previous studies relying on implicit matching (Peiffer-Smadja & Cohen, 2019; McCormick et al., 2022). A caveat to this approach is that neural activity associated with the motor response required for the explicit judgment may confound detection of condition-specific effects. To address this, we adopted a standard approach used for both univariate analysis and MVPA to account for RT differences across congruent and incongruent conditions (Grinband et al., 2008; Woolgar et al., 2015). However, this method assumes that the BOLD signal scales with RT during the task, which may mask congruency-related activity that is relatively constant across trials (Mumford et al., 2023). Further, we restricted our stimuli to two pseudowords and two shapes that were distinctly associated with “roundness” or “pointedness” based on empirical ratings, but were otherwise well-matched on stimulus parameters (Lacey et al., 2020). This allowed us to clearly define pseudoword-shape correspondences for simple explicit judgments and to maximize our trial sample size for statistical efficiency (Chen et al., 2022), but may limit the generalizability of our findings. Although our interpretations of the fMRI findings rely on reverse inference, they do so in the context of a particular task and anatomical priors (Hutzler, 2014), based in hypotheses that arise from the psychophysical literature. Future replications of the present work, use of condition-rich designs, probing of hypothesized neural processes, and extension to other sound-symbolic domains is necessary to more fully understand how crossmodal congruency representations may vary across a continuum of stimuli.

Conclusions

This fMRI study provided evidence for multiple neural processes underlying sound symbolism. Most interestingly, the present study offers support for a relationship of sound symbolism to the neural processes mediating language and multisensory integration. Other relevant operations include lower-level auditory and visual cortical processing of the stimuli, somatic sensorimotor processing, and more general cognitive operations mediating attention or decisions. Additional work is needed to disentangle the contributions of these various processes to sound symbolism and to more fully understand the role of sound symbolism in language.

ACKNOWLEDGMENTS

This work was supported by grants to KS and LCN from the National Eye Institute at the NIH (R01 EY025978) and the Emory University Research Council. Support to KS from the Veterans Administration is also acknowledged. Neuroimaging data were collected at the Facility for Education and Research in Neuroscience (FERN) at Emory University. We thank Kate Pirog Revill for assistance with scanning and Sara List for data collection.

DATA AND CODE AVAILABILITY

Data are available on request to the corresponding author and on submission of a completed data sharing agreement, project outline, and agreement on any co-authorship requirements.

.

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J. et al. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 14.

Ademollo, F. (2011). *The Cratylus of Plato: A commentary.* Cambridge, UK: Cambridge University Press.

Audacity Team (2012) Audacity v2.0.1 [Computer program]. Retrieved from http://audacity.sourceforge.net/ Audacity ® software is copyright © 1999-2014 Audacity Team.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26-41.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, 37(1), 90-101.

Bernstein, L.E., Lu, Z.-L. & Jiang, J. (2008). Quantified acoustic-optical speech signal incongruity identifies cortical sites of audiovisual speech processing. *Brain Research*, 1242:172-184.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences USA,* 113: 10818–10823.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10: 433-436.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3), 1-27.

Chang, L.J., Yarkoni, T., Khaw, M.W., & Sanfey, A.G. (2013). Decoding the role of the insula in human cogniton: functional parcellation and large-scale reverse inference. *Cerebral Cortex*, 23: 739-749.

Chen, G., Pine, D. S., Brotman, M. A., Smith, A. R., Cox, R. W., Taylor, P. A., & Haller, S. P. (2022). Hyperbolic trade-off: the importance of balancing trial and subject sample sizes in neuroimaging. *NeuroImage*, 247, 118786.

Cho, Z.-H. (Editor). (2010). *7.0 Tesla MRI Brain Atlas*. New York: Springer Science+Business Media LLC.

Chumbley, J., Worsley, K., Flandin, G., & Friston, K. (2010). Topological FDR for neuroimaging. *Neuroimage*, 49(4), 3057-3064.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162-173.

Cuskley, C., Simner, J., & Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research*, 81: 119–130

Dale, A.M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8:109-114.

Dale, A.M., Greve, D.N. & Burock, M.A. (1999) Optimal Stimulus Sequences for Event-Related fMRI. *5th International Conference on Functional Mapping of the Human Brain*. Düsseldorf, Germany, June 11-16, 1999.

Darainy, M., Vahdat, S. & Ostry, D.J. (2019). Neural basis of sensorimotor plasticity in speech motor adaptation. *Cerebral Cortex*, 29: 2876-2889.

Diaconescu, A.O., Alain, C. & McIntosh, A.R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, 106: 2896-2909.

Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Ilkay Isik, A., Erramuzpe, A. et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16: 111-116.

Esteban, O., Ciric, R., Finc, K., Blair, R. W., Markiewicz, C. J., Moodie, C. A. et al. (2020). Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nature protocols,* 15(7), 2186-2202.

Fedorenko, E. & Blank, I.A. (2020). Broca's area is not a natural kind. *Trends in Cognitive Sciences*, 24:270-284.

Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are more important than vowels in the bouba-kiki effect. *Language & Speech*, 58: 247–266.

Fort, M. & Schwartz, J.-L. (2022). Resolving the bouba-kiki effect enigma by rooting iconic sound symbolism in physical properties of round and spiky objects. *Scientific Reports*, 12, 19172.

Ghazanfar, A.A. & Schroeder, C.E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10: 278-285.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, 13.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1), 63-72.

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *Neuroimage*, 43(3), 509-520.

Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S. & Siebner, H. R. (2010). Phonological decisions require both left and right supramarginal gyri*, Proceedings of the National Academy of Sciences USA*, 107: 16494–16499.

Hartwigsen, G., Bestmann, S., Ward, N. S., Woerbel, S., Mastroeni, C., Granert, O., & Siebner, H. R. (2012). Left dorsal premotor cortex and supramarginal gyrus complement each other during rapid action reprogramming. *Journal of Neuroscience*, 32(46), 16162-16171.

Hebart, M.N., Görgen, K. & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, 8: 88, doi: 10.3389/fninf.2014.00088

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage*, 84: 1061-1069.

Itagaki, S., Murai, S., & Kobayasi, K. I. (2019). Brain activity related to sound symbolism: cross-modal effect of an aurally presented phoneme on judgment of size. *Scientific Reports*, 9(1), 1-7.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143-156.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2), 825-841.

Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L. et al. (2015). AICHA: an atlas of intrinsic connectivity of homotopic areas. *Journal of Neuroscience Methods*, 254:46-59.

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: a study on Japanese mimetic words. *PLoS one*, 9(5), e97905.

Kaplan, J.T., Man, K. & Greening, S.G. (2015). Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. *Frontiers in Human Neuroscience*, 9:151, doi: 10.3389/fnhum.2015.00151

Kawahara, S., Godoy, M.C. & Kumagai, G. (2021). English speakers can infer Pokémon types based on sound symbolism. *Frontiers in Psychology*, 12, 648948.

Kemmerer, D. (2022). Revisiting the relation between syntax, action, and left BA44. *Frontiers in Human Neuroscience*, 16:923022, doi:10.3389/fnhum.2022.923022

Kitada, R., Kwon, J., Doizaki, R., Nakagawa, E., Tanigawa, T., Kajimoto, H., et al. (2021). Brain networks underlying the processing of sound symbolism related to softness perception, *Scientific Reports,* 11: 7399. doi:10.1038/s41598-021-86328-6.

Kleiner M., Brainard D. & Pelli D. (2007) "What's new in Psychtoolbox-3?" *Perception*, 36: ECVP Abstract Supplement.

Kliger, L. & Yovel, G. (2020). The functional organization of high-level visual cortex determines the representation of complex visual stimuli. *Journal of Neuroscience*, 40: 7545-7558.

Köhler, W. (1929). *Gestalt psychology.* New York: Liveright.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology.* New York: Liveright.

Lacey, S., Jamal, Y., List, S.M., McCormick, K., Sathian, K. & Nygaard, L.C. (2020). Stimulus parameters underlying sound-symbolic mapping of auditory pseudowords to visual shapes. *Cognitive Science*, 44: e12883.

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, 1(1), 76-85.

Lange, J., Christian, N. & Schnitzler, A. (2013). Audio-visual congruency alters power and coherence of oscillatory activity within and between cortical areas. *NeuroImage*, 79:111-120.

Lee, S., & Kable, J. W. (2018). Simple but robust improvement in multivoxel pattern classification. *PLoS One*, 13(11), e0207083.

Loued-Khenissi, L., Pfeuffer, A., Einhäuser, W. & Preuschoff, K. (2020). Anterior insula reflects surprise in value-based decision-making. *NeuroImage*, 210: 116549.

Malinowska, U., Crone, N.E., Lenz, F.A., Cervenka, M. & Boatman-Reich, D. (2017). Multi-regional adaptation in human auditory association cortex. *Frontiers in Human Neuroscience*, 11:247, doi: 10.3389/fnhum.2017.00247

Man, K., Damasio, A., Meyer, K. & Kaplan, J.T. (2015). Convergent and invariant object representations for sight, sound, and touch. *Human Brain Mapping*, 36:3629-3640.

Maran, M., Friederici, A.D. & Zaccarella, E. (2022). Syntax through the looking glass: A review on two-word linguistic processing across behavioral, neuroimaging and neurostimulation studies. *Neuroscience & Biobehavioral Reviews*, 142:104881.

Margiotoudi, K. & Pulvermüller, F. (2020). Action sound-shape congruencies explain sound symbolism. *Scientific Reports*, 10: 12706, doi: 10.1038/s41598-020-69528-4

Martuzzi, R., van der Zwaag, W., Farthouat, J., Gruetter, R. & Blanks, O. (2014). Human finger somatotopy in Areas 3b, 1, and 2: a 7T fMRI study using a natural stimulus. *Human Brain Mapping*, 35:213-226.

McCormick, K., Kim, J.Y., List, S. & Nygaard, L.C. (2015). Sound to meaning mappings in the bouba-kiki effect. *Proceedings 37th Annual Meeting Cognitive Science Society*, pp1565-1570.

McCormick, K., Lacey, S., Stilla, R., Nygaard, L.C. & Sathian, K. (2022). Neural basis of the sound-symbolic crossmodal correspondence between auditory pseudowords and visual shapes. *Multisensory Research,* 35: 29-78.

McNorgan, C. & Booth, J.R. (2015). Skill dependent audiovisual integration in the fusiform induces repetition suppression. *Brain & Language*, 141: 110-123.

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, 76: 264–275.

Misaki, M., Luh, W. M., & Bandettini, P. A. (2013). The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. *Journal of Neuroscience Methods*, 212: 355-361.

Moerel, M., De Martino, F. & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, 8: 225, doi: 10.3389/fnins.2014.00225

Moulton, E.A, Pense, G., Morris, S., Aiello-Lammens, M., Becerra, L. & Borsook, D. (2009). Segmentally arranged somatotopy within the face representation of human primary somatosensory cortex. *Human Brain Mapping*, 30: 757-765

Mumford, J. A., Bissett, P. G., Jones, H. M., Shim, S., Rios, J. A. H., & Poldrack, R. A. (2023). The response time paradox in functional magnetic resonance imaging analyses. *BioRxiv*, 2023-02.

Neufeld, J., Sinke, C., Dillo, W., Emrich, H.M., Szycik, G.R., Dima, D., Bleich, S., & Zedler, M. (2012). The neural correlates of coloured music: A functional MRI investigation of auditory-visual synaesthesia. *Neuropsychologia*, 50: 85-89.

Nielsen, A.K, & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech,* 64: 52-72.

Nuckolls, J.B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, 28:225-252.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112: 181–186.

Oberhuber, M., Hope, T. M. H., Seghier, M. L., Parker Jones, O., Prejawa, S., Green, D.W. & Price, C. J. (2016). Four functionally distinct regions in the left supramarginal gyrus support word processing, *Cerebral Cortex*, 26: 4212–4226.

Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I.P., Joensuu, R., Autti, T. & Sams, M. (2005). Processing of audiovisual speech in Broca's area. *NeuroImage*, 25:333-338.

Peiffer-Smadja, N., & Cohen, L. (2019). The cerebral bases of the bouba-kiki-effect. *NeuroImage*, 186: 679-689.

Pekkola, J., Laasonen, M., Ojanen, V., Autti, T., Jääskeläinen, I.P., Kujala, T. & Sams, M. (2006). Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: An fMRI study at 3T. *NeuroImage*, 29:797-807.

Pellegrino, G., Tomasevic, L., Herz, D. M., Larsen, K. M., & Siebner, H. R. (2018). Theta activity in the left dorsal premotor cortex during action re-evaluation and motor reprogramming. *Frontiers in Human Neuroscience*, 12, 364.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10: 437-442.

Perikova, E., Blagovechtchenski, E., Filippova, M., Shcherbakova, O., Kirsanov, A. & Shtyrov, Y. (2022). Anodal tDCS over Broca's area improves fast mapping and explicit encoding of novel vocabulary. *Neuropsychologia*, 168:108156

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Harper Collins, New York, NY, USA.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, 84, 320-341.

Raczkowski, D., Kalat, J. W. & Nebes, R. (1974). Reliability and validity of some handedness questionnaire items. *Neuropsychologia*, 12: 43-47.

Ramachandran, V. S. & Hubbard, E. M. (2001). Synaesthesia – a window into perception,

thought and language, *Journal of Consciousness Studies*, 8: 3–34.

Revill, K. P., Namy, L. L., DeFife, L. C. & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: evidence from fMRI and DTI. *Brain & Language,* 128:18–24.

Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, 122: 1-5.

Saji, N., Akita, K., Kantartzis, K., Kita, S. & Imai, M. (2019). Cross-linguistically shared and language-specific sound symbolism in novel words elicited by locomotion videos in Japanese and English. *PLoS ONE*, 14, e2018707.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E. et al. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, 64, 240-256.

Schroeder, C.E., Smiley, J., Fu, K.G., McGinnis, T., O'Connell, M.N. & Hackett, T. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *International Journal of Psychophysiology*, 50: 5-17.

Shen, G., Zhang, J., Wang, M., Lei, D., Yang, G., Zhang, S., & Du, X. (2014). Decoding the individual finger movements from single-trial functional magnetic resonance imaging recordings of human brain activity. *European Journal of Neuroscience*, 39(12), 2071-2082.

Sidhu, D.M., Westbury, C., Hollis, G. & Pexman, P.M. (2021). Sound symbolism shapes the English language: the maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, 28: 1390-1398.

Sonier, R.-P., Poirier, M., Guitard, D. & Saint-Aubin, J. (2020). A round Bouba is easier to remember than a curved Kiki: Sound symbolism can support associative memory. *Psychonomic Bulletin & Review*, 27, 776-782.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics,* 73, 971-95.

Stehr, D. A., Garcia, J. O., Pyles, J. A., & Grossman, E. D. (2023). Optimizing multivariate pattern classification in rapid event-related designs. *Journal of Neuroscience Methods*, 387, 109808.

Stevenson, R.A., Kim, S. & James, T.W. (2009). An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Experimental Brain Research*, 198: 183-194.

Szycik, G.R., Jansma, H. & Münte, T.F. (2009). Audiovisual integration during speech comprehension: An fMRI study comparing ROI-based and whole brain analyses. *Human Brain Mapping*, 30:1990-1999.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310-1320.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N. et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject Brain. *NeuroImage*, 15: 273-289.

Vaionio, L., Kilpeläinen, M., Wikström, A. & Vainio, M. (2023). Sound-action symbolism in relation to precision manipulation and whole-hand grasp usage. *Quarterly Journal of Experimental Psychology*, in press, doi: 10.1177/17470218231160910

Vainio, L. & Vainio, M. (2021). Sound-action symbolism. *Frontiers in Psychology*, 12: 718700, doi: 10.3389/fpsyg.2021.718700

Vainio, L. & Vainio, M. (2022). Interaction between grasping and articulation: how vowel and consonant pronunciation influences precision and power grip responses. *PLoS ONE*, 17: e0265651, doi: 10.1371/journal.pone.0265651

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible coding of task rules in frontoparietal cortex: an adaptive system for flexible cognitive control. *Journal of Cognitive Neuroscienc*e, 27(10), 1895-1911.

Xie, Y.J., Li, Y.Y., Xie, B., Xu, Y.Y. & Peng, L. (2019). The neural basis of complex audiovisual objects maintenances in working memory. *Neuropsychologia*, 133:107189.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45-57.

Zhang, L. & Du, Y. (2022). Lip movements enhance speech representations and effective connectivity in auditory dorsal stream. *NeuroImage*, 257:119311.

Table 1. Mean (standard error) accuracy and reaction times.

Stimulus pairs	Accuracy (%)	Reaction Time (ms)
Congruent		
‘mohloh’/rounded shape	95.21 (0.75)	898 (33)
‘kehteh’/pointed shape	95.73 (0.79)	832 (31)
Incongruent		
‘mohloh’/pointed shape	96.35 (0.99)	958 (31)
‘kehteh’/rounded shape	94.11 (1.76)	938 (33)

Table 2. Brain regions showing univariate activation differences for congruent vs. incongruent stimuli (whole-brain topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$). Coordinates of peak local maxima are shown for each cluster.

	Peak MNI coordinates			Cluster Level	
Anatomical Location	x	y	z	mm^3	$P_{FDR\text{-}corr}$
Congruent > incongruent					
L postcentral gyrus	-45	-19	49	163	< 0.001
L mid-superior temporal gyrus	-46	-13	2	49	0.016
L anterior fusiform/parahippocampal gyrus	-21	-34	-19	929	< 0.001
L caudate nucleus	-8	14	17	71	0.013
R mid-cingulate gyrus/corpus callosum	10	3	31	54	0.013
R caudate nucleus	9	16	12	153	< 0.001
Incongruent > congruent					
L precentral sulcus	-31	-3	52	1456	< 0.001
L supplementary motor area	-6	13	49	984	< 0.001
L anterior insula	-41	15	8	434	< 0.001
R anterior insula	38	21	7	576	< 0.001
R pons	11	-39	-32	136	< 0.001

Table 3. Brain regions showing above-chance classification of visual and auditory stimuli in congruent and incongruent conditions (whole-brain topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$). Coordinates of peak local maxima are shown for each cluster.

	Peak MNI coordinates			Cluster Level	
Anatomical Location	x	y	z	mm^3	$P_{FDR\text{-}corr}$
Congruent decoding					
L supplementary motor area	-2	0	56	4171	< 0.001
L inferior frontal gyrus, pars opercularis	-45	6	28	3263	< 0.001
L intraparietal sulcus	-41	-50	49	228	0.006
L supramarginal gyrus	-62	-30	26	13328	< 0.001
R superior frontal gyrus	20	46	25	3704	< 0.001
R intraparietal sulcus	38	-49	48	2096	< 0.001
R precuneus	6	-47	6	583	0.006
R superior temporal gyrus	57	-12	-5	10332	< 0.001
R middle temporal gyrus	64	-48	-9	122	0.042
R inferior temporal gyrus	53	-47	-10	162	0.020
R inferior temporal gyrus	41	-38	-9	152	0.023
R calcarine sulcus	19	-87	7	1080150	< 0.001
Incongruent decoding					
L superior temporal gyrus	-42	-27	18	7938	< 0.001
L precuneus	-13	-65	66	251	0.004
L middle occipital gyrus	-17	-100	4	21732	< 0.001
R superior temporal gyrus	59	-26	11	15147	< 0.001
R fusiform gyrus	38	-42	-24	450	< 0.001
R middle occipital gyrus	40	-63	15	259	0.004
R calcarine sulcus	16	-98	-1	26326	< 0.001

R cerebellum	26	-40	-35	150	0.026

Table 4. Brain regions showing higher above-chance classification of visual and auditory stimuli in the congruent conditions vs. the incongruent conditions (whole-brain topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$). Coordinates of peak local maxima are shown for each cluster. Note: no region showed greater incongruent > congruent decoding.

	Peak MNI coordinates			Cluster Level	
Anatomical Location	x	y	z	mm^3	$P_{FDR\text{-}corr}$
L inferior frontal gyrus, pars opercularis	-44	10	15	291	0.008
L supramarginal gyrus	-60	-23	26	485	0.008
R mid-occipital gyrus	19	-88	6	239	0.008

Table 5. Brain regions showing above-chance decoding of visual shape and congruency (whole-brain topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$). Coordinates of peak local maxima are shown for each cluster.

	Peak MNI coordinates			Cluster Level	
Anatomical Location	x	y	z	mm^3	$P_{FDR\text{-}corr}$
Shape decoding					
L calcarine sulcus	-11	-99	-4	20279	< 0.001
R lingual gyrus	16	-93	-6	19354	< 0.001
Congruency decoding					
L superior frontal gyrus	-17	17	41	752	< 0.001
L precentral gyrus	-35	-25	60	42895	< 0.001
L posterior cingulate gyrus	-19	-46	28	1050	< 0.001
L hippocampus	-29	-6	-23	706	< 0.001
R superior frontal gyrus	5	45	50	151	0.023
R superior frontal gyrus	19	20	44	1476	< 0.001
R middle frontal gyrus	38	41	23	832	< 0.001
R inferior frontal gyrus, pars triangularis	49	22	15	280	0.007
R inferior frontal gyrus, pars orbitalis	41	32	-2	121	0.043
R postcentral gyrus	49	-13	35	1813	< 0.001
R superior parietal lobule	20	-67	58	627	< 0.001
R pallidum	23	-7	6	928	< 0.001

Table 6. Brain regions showing above-chance decoding of auditory pseudowords and congruency (whole-brain topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$). Coordinates of peak local maxima are shown for each cluster.

	Peak MNI coordinates			Cluster Level	
Anatomical Location	x	y	z	mm^3	$P_{FDR\text{-}corr}$
Pseudoword decoding					
L superior temporal sulcus/gyrus	-54	-19	-1	720	< 0.001
R superior temporal sulcus/gyrus	63	-17	-2	5929	< 0.001
Congruency decoding					
L inferior frontal sulcus	-46	30	23	153	0.022
L orbito-frontal gyrus	-19	33	-19	253	0.003
L postcentral gyrus	-12	-36	39	80140	< 0.001
L anterior middle temporal gyrus	-57	-15	-20	300	0.002
L anterior inferior temporal gyrus	-35	-19	-23	374	0.001
L middle/inferior temporal gyrus	-48	-48	-5	958	< 0.001
L posterior inferior temporal gyrus	-35	-65	-18	1392	< 0.001
L posterior inferior temporal gyrus	-34	-86	-14	129	0.036
R postcentral gyrus	35	-29	61	17413	< 0.001
R anterior insula	33	30	6	354	0.001
R parahippocampal gyrus	28	-1	-29	2208	< 0.001
R inferior temporal gyrus	52	-50	-4	298	0.002
R thalamus	15	-10	6	15048	< 0.001
R posterior superior cerebellum	29	-72	-22	542	< 0.001
R lateral cerebellum	36	-53	-28	309	0.001

FIGURE LEGENDS

Figure 1. Task design and classification analysis scheme. (A) There were four possible stimuli, reflecting unique combinations of two visual shapes, pointed or rounded, and two auditory pseudowords, 'kehteh' or 'mohloh'. (B) Pseudoword-shape pairs were presented simultaneously and participants pressed a response button indicating whether the pair were a 'match' (i.e., congruent) or a 'mismatch' (i.e., incongruent). (C) In the within-congruency classification analysis, leave-one-run-out cross-validation was used to test for patterns of brain activity discriminating the two congruent conditions or the two incongruent conditions. (D) in the between-congruency classification analyses, classifiers were trained to distinguish between the two conditions in which the visual shapes differed but the auditory pseudoword was the same (e.g., 'kehteh'/pointed shape vs. 'kehteh'/rounded shape). The classifier was then tested on the two remaining conditions containing the other pseudoword (e.g., 'mohloh'/pointed shape vs. 'mohloh'/rounded shape). Similarly, we trained classifiers to distinguish between auditory pseudowords holding the shape constant (e.g., 'kehteh'/pointed shape vs. 'mohloh'/pointed shape) and then tested them on the other shape (e.g., 'kehteh'/rounded shape vs. 'mohloh'/rounded shape). This approach differentiates whether patterns of brain activity were specific to features of the visual/auditory stimulus, or to the congruency/incongruency of the condition.

Figure 2. Univariate results. Whole-brain map showing significant effects from the incongruent vs. congruent contrasts (topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$).

Figure 3. Whole-brain searchlight maps showing significant above-chance within-congruency classification of the two congruent conditions (congruent decoding) and the two incongruent conditions (incongruent decoding). All maps whole-brain topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$).

Figure 4. Whole-brain searchlight map showing significantly higher classification of the congruent conditions relative to the incongruent conditions and mean decoding accuracy in (A) left inferior frontal gyrus, pars opercularis (Broca's area: L IFGpo), (B) left supramarginal gyrus (L SMG), and (C) right mid-occipital gyrus (MOG). Mean decoding accuracy was extracted from the peak voxel coordinates for each of the significant clusters in A-C. No region showed significantly higher above-chance classification of the incongruent conditions. All maps whole-brain topological FDR-corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$); error bars show ± standard error of the mean.

Figure 5. Between-congruency classification. (A) Whole-brain searchlight maps showing significant above-chance decoding of either the visual shape or the congruency of the stimuli. (B) Whole-brain searchlight maps showing significant above-chance decoding of either the auditory pseudoword or the congruency of the stimuli. All maps whole-brain topological FDR-

corrected; activation clusters with cluster-forming threshold of $p < 0.001$ and FDR cluster-level correction of $q < 0.05$).

Figure 1

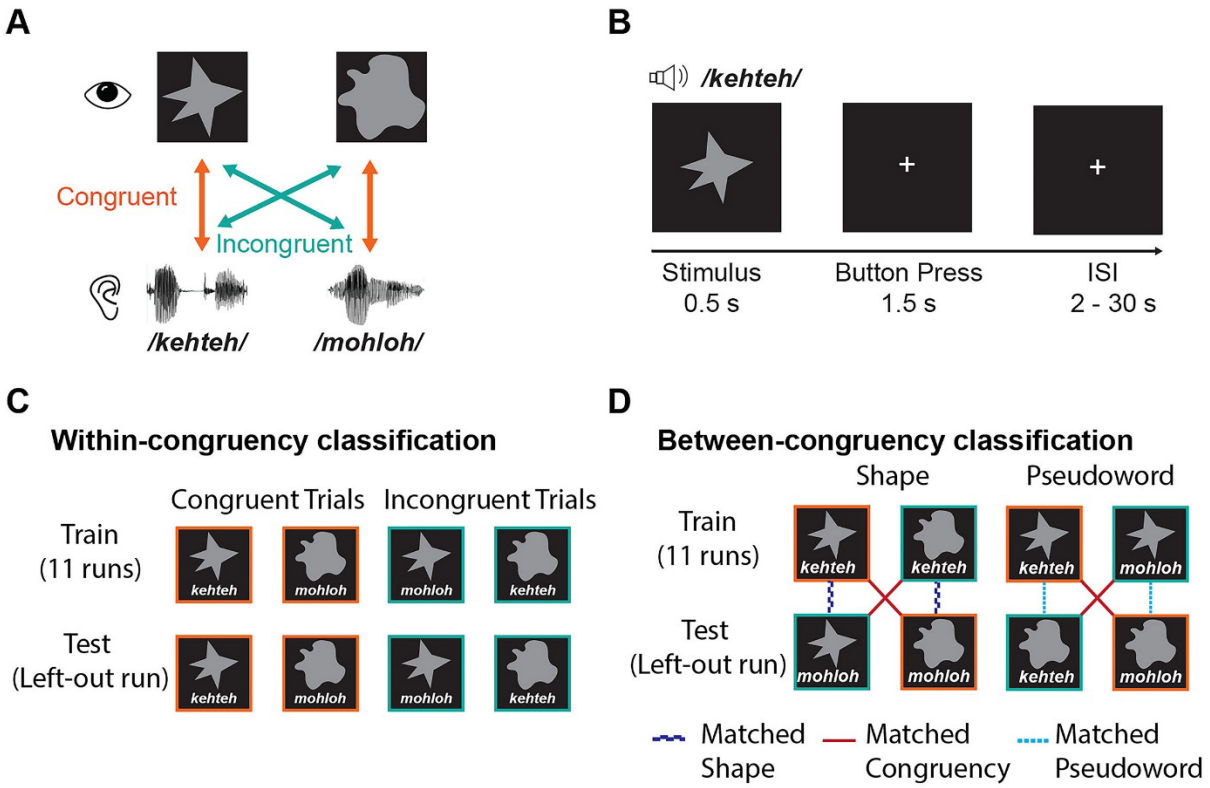

A
Congruent
Incongruent
/kehteh/
/mohloh/
B
/kehteh/
+
+
Stimulus
0.5 s
Button Press
1.5 s
ISI
2 - 30 s
C
Within-congruency classification
Congruent Trials
Incongruent Trials
Train
(11 runs)
Test
(Left-out run)
kehteh
mohloh
mohloh
kehteh
kehteh
mohloh
mohloh
kehteh
D
Between-congruency classification
Shape
Pseudoword
Train
(11 runs)
Test
(Left-out run)
kehteh
kehteh
kehteh
mohloh
mohloh
mohloh
kehteh
mohloh
Matched
Shape
Matched
Congruency
Matched
Pseudoword

Figure 2

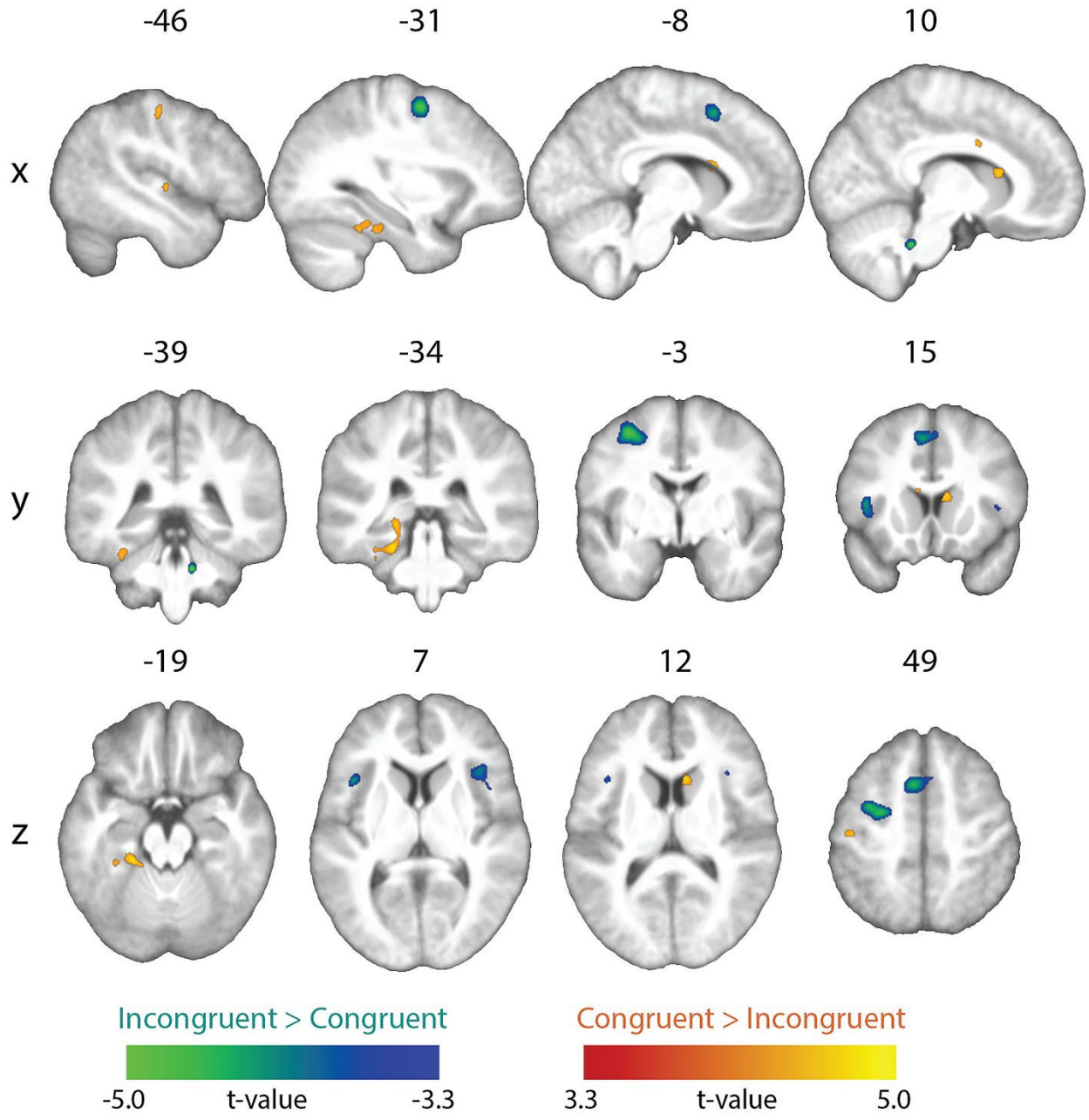

-46
-31
-8
10
x
-39
-34
-3
15
y
-19
7
12
49
z
Incongruent > Congruent
-5.0
t-value
-3.3
Congruent > Incongruent
3.3
t-value
5.0

Figure 3

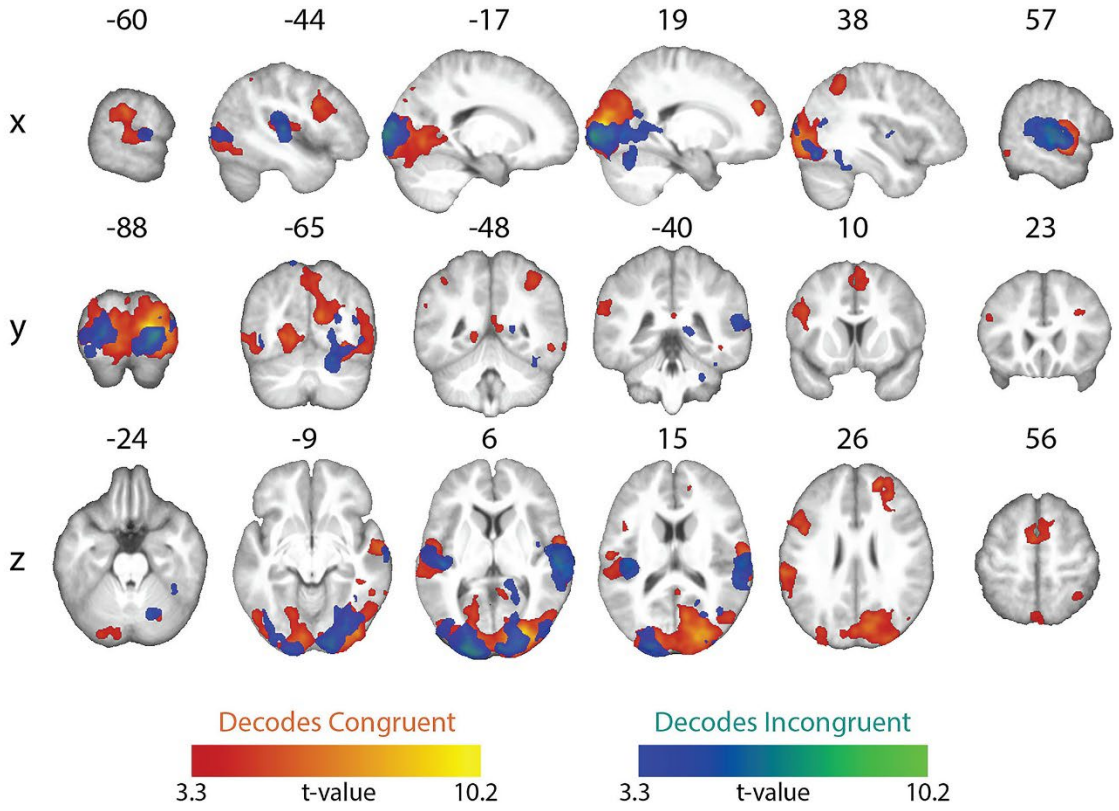

-60
-44
-17
19
38
57
x
-88
-65
-48
-40
10
23
y
-24
-9
6
15
26
56
z
Decodes Congruent
3.3
t-value
10.2
Decodes Incongruent
3.3
t-value
10.2

Figure 4

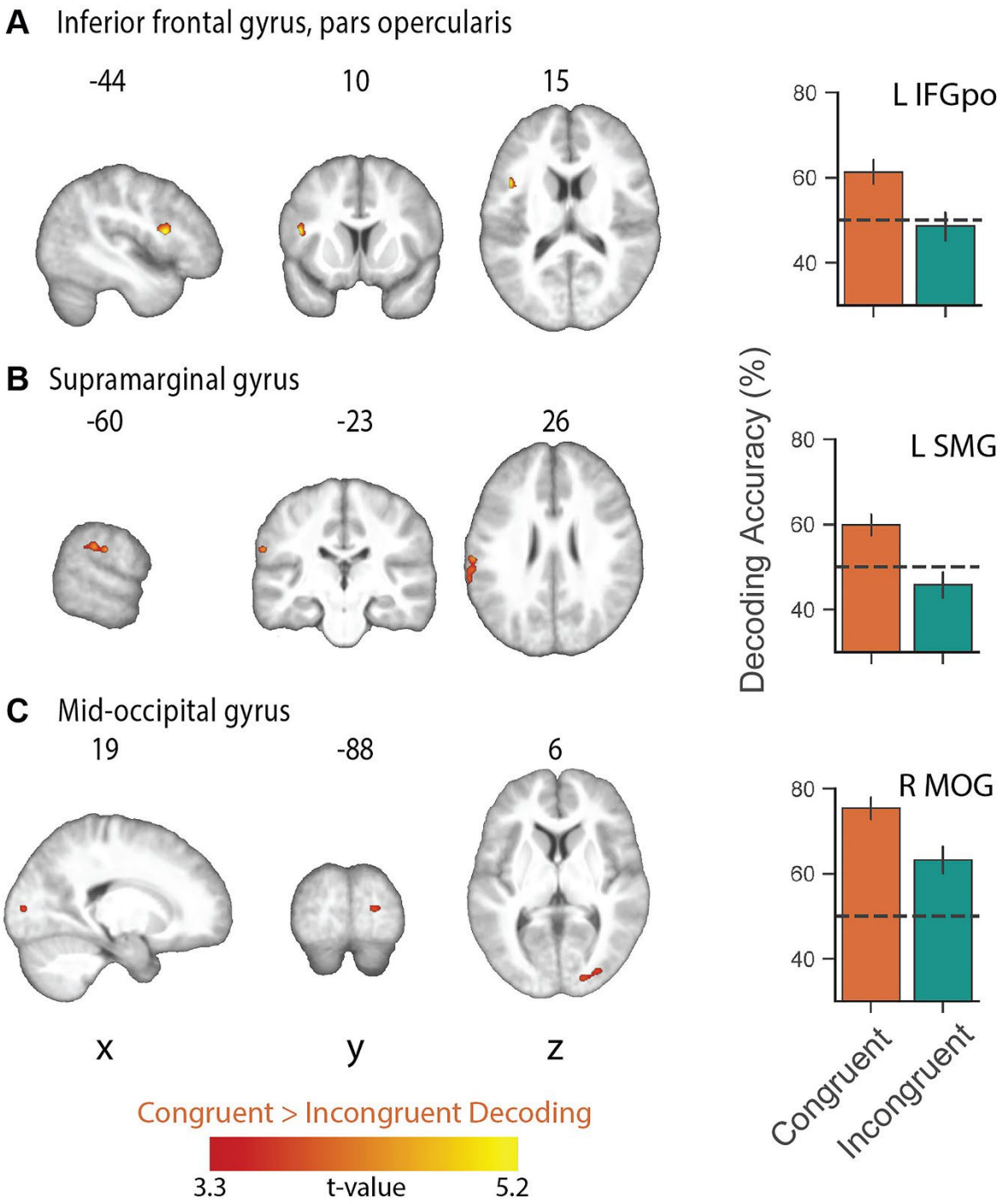
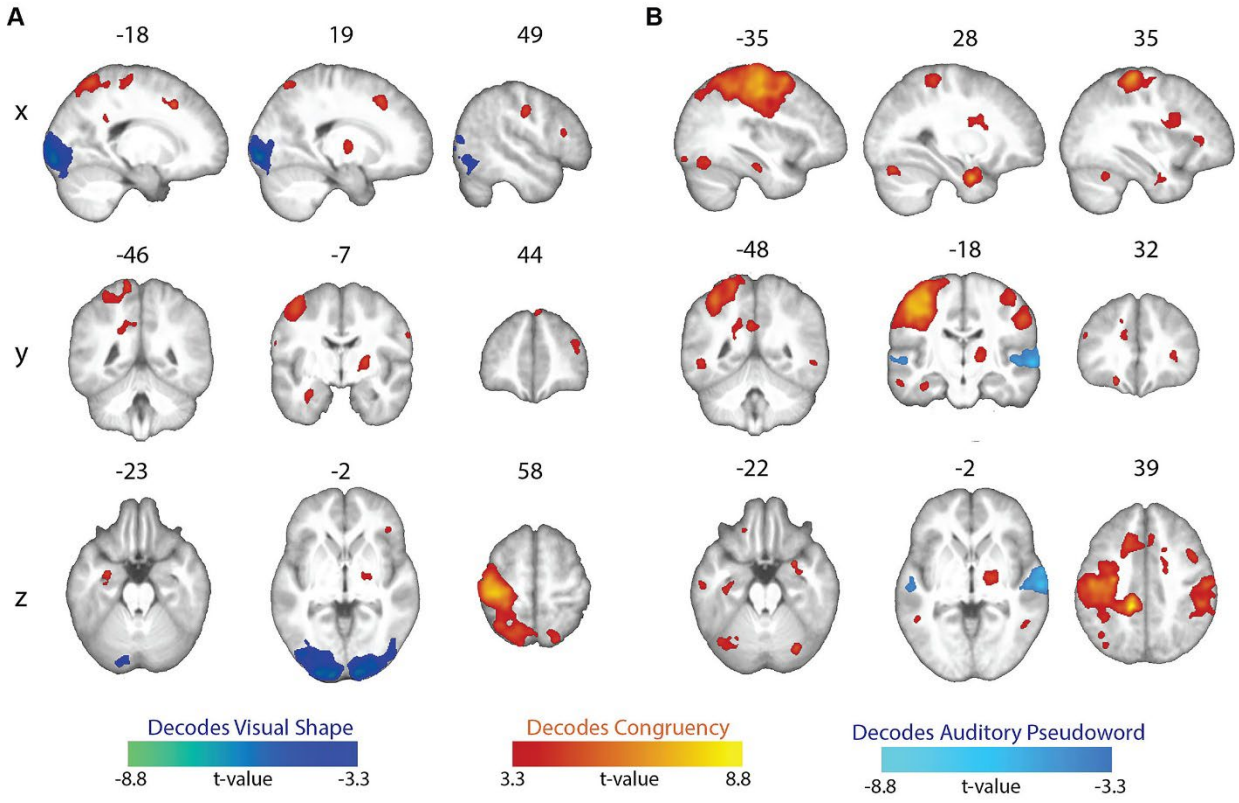

A Inferior frontal gyrus, pars opercularis
-44
10
15
B Supramarginal gyrus
-60
-23
26
C Mid-occipital gyrus
19
-88
6
x
y
z
Congruent > Incongruent Decoding
3.3
t-value
5.2
L IFGpo
L SMG
R MOG
Decoding Accuracy (%)
80
60
40
Congruent
Incongruent

Figure 5


A
B
x
y
z
-18
19
49
-35
28
35
-46
-7
44
-48
-18
32
-23
-2
58
-22
-2
39
Decodes Visual Shape
-8.8
t-value
-3.3
Decodes Congruency
3.3
t-value
8.8
Decodes Auditory Pseudoword
-8.8
t-value
-3.3