

<https://doi.org/10.1038/s42003-025-07466-5>

Neural representation of sensorimotor features in language-motor areas during auditory and visual perception

Check for updates

Yuanyi Zheng¹, Jianfeng Zhang¹, Yang Yang^{2,3} & Min Xu¹✉

Speech processing involves a complex interplay between sensory and motor systems in the brain, essential for early language development. Recent studies have extended this sensory-motor interaction to visual word processing, emphasizing the connection between reading and handwriting during literacy acquisition. Here we show how language-motor areas encode motoric and sensory features of language stimuli during auditory and visual perception, using functional magnetic resonance imaging (fMRI) combined with representational similarity analysis. Chinese-speaking adults completed tasks involving the perception of spoken syllables and written characters, alongside syllable articulation and finger writing tasks to localize speech-motor and writing-motor areas. We found that both language-motor and sensory areas generally encode production-related motoric features across modalities, indicating cooperative interactions between motor and sensory systems. Notably, sensory encoding within sensorimotor areas was observed during auditory speech perception, but not in visual character perception. These findings underscore the dual encoding capacities of language-motor areas, revealing both shared and distinct neural representation patterns across modalities, which may be linked to innate sensory-motor mechanisms and modality-specific processing demands. Our results shed light on the sensorimotor integration mechanisms underlying language perception, highlighting the importance of a cross-modality perspective.

Speech serves as a vital medium for real-time communication, where listening and speaking are closely interlinked. Humans are innately capable of discriminating phonetic features across languages^{1,2}, which are refined through speech production experience^{3,4}. In daily conversations, speaking and listening often occur simultaneously, fostering a continuous feedback loop between the production and perception systems^{5,6}. Theories and neuroimaging studies suggest that speech-production-related motor areas, including the ventral premotor and motor cortex in the precentral gyrus (preCG), supplementary motor area (SMA), and Broca's area in the inferior frontal gyrus (IFG), also contribute to auditory speech perception^{7,8}, although their exact role remains unclear.

Speech perception involves the precise analysis and differentiation of distinctive auditory or articulatory features. However, variability in acoustic signals complicates categorization of speech sounds into phonemes or words. One predominant perspective proposes that speech is processed as invariant articulatory gestures within motor areas to facilitate recognition, aligning with the motor theory of speech perception^{9,10}.

Research has shown that listening to speech activates the frontal motor areas involved in producing these sounds^{11–16}, suggesting potential motoric representations of articulatory gestures within these speech-motor areas during speech perception. Such production-perception mapping has been observed even in pre-babbling infants with limited articulatory abilities². Moreover, studies on songbirds^{17,18} and non-human primates¹⁹ indicate a role for vocal production systems in voice/speech perception. This evidence implies that motoric encoding for auditory language perception has an early-developmental or evolutionary basis, possibly revealed in speech-motor areas²⁰.

An alternative perspective argues that the motoric representations within speech-motor areas play a supplementary role in aiding speech discrimination during challenging situations, while fine-grained auditory processing underpins the core of speech perception^{21–24}. Extending this, recent theories propose that speech-motor areas might also encode acoustic features, complementing the auditory cortex^{5,6,25}. Consistent with this idea, an electrocorticography (ECoG) study has demonstrated that ventral

¹School of Psychology, Shenzhen University, Shenzhen, China. ²Center for Brain Science and Learning Difficulties, Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China. ³Department of Psychology, University of Chinese Academy of Sciences, Beijing, China.

✉ e-mail: xumin@szu.edu.cn

precentral cortex activity is organized by acoustic features during perception²⁶. Additionally, a functional magnetic resonance imaging (fMRI) study indicated that articulatory and acoustic information was co-represented in the ventral precentral cortex and IFG during speech perception²⁷. Research on speech production further illustrates this complex motor-sensory interaction. Zhang et al.²⁸ found that during overt and silent articulation, articulatory motoric information was represented not only in speech-motor regions such as the left anterior insula and IFG, but also in somatosensory and auditory regions like the superior temporal gyrus (STG), suggesting an interaction between motor regions and sensory pathways.

Nonetheless, acoustic information remained localized in auditory regions, particularly the STG, but not in motor areas²⁸. Yet, it remains an open question whether motoric, auditory, or both types of representations are present in speech-motor areas during perception, and how motor areas interact with sensory regions.

When language is presented visually (as in reading), do production-perception interactions occur similarly to those in auditory processing? Unlike speech, reading is a later-acquired skill through intensive training and is closely intertwined with handwriting during literacy development²⁹. This reading-writing coupling process refines the spatial representations of visual word forms^{30–33} and supports motor memory for written words^{34–37}. Consequently, it is proposed that the brain regions involved in both reading and writing may integrate to contribute to visual word recognition^{38–41}.

Recent studies have extended our understanding of the role of production-related motor regions in reading. It has been observed that visual word recognition involves not only the ventral occipitotemporal system for recognizing word shape, but also the dorsal writing-associated motor regions, such as Exner's area in the posterior part of the left superior frontal sulcus or the dorsal left middle frontal gyrus^{42–44}. Exner's area is regarded as the graphemic center⁴⁵, crucial for associating orthographic representations with handwriting motor commands^{46–48}, and it plays a significant role in facilitating visual word processing^{38,42–44}. Another important handwriting-related motor region involved in reading is the left superior parietal lobule/intraparietal sulcus (SPL/IPS), which is suggested to engage in the visual-motor sequence processing of written strokes and also contributes to reading^{30,49}.

However, the precise nature of neural representations within these writing-motor areas during visual word perception remains underexplored. The prevailing view is that the writing-motor regions encode motoric writing gestures during reading^{40,41,43}, as evidenced by the increased activation within Exner's area when viewing words or characters in stroke-by-stroke motion^{43,49,50}, or following handwriting training^{30,31,51}. Nevertheless, most of these studies might be influenced by confounding factors, such as explicit sequence processing. In contrast to motoric encoding, recent behavioral studies underscore the significance of visual-spatial information over writing-motoric knowledge in affecting reaction times for letter or character discrimination^{33,52}, suggesting a greater role of visual-form encoding⁵³. Despite these insights, it remains unclear whether writing-motor areas represent motoric, sensory, or both features, and how they associate with visual processing regions.

For a clearer picture on the encoding and integration mechanisms in writing-motor areas, representational similarity analysis (RSA) offers a promising approach. RSA characterizes the information represented in a brain region by quantifying the similarity between multi-voxel neural response patterns and the feature parameters across different stimuli, using representational dissimilarity matrices (RDMs)^{54,55}. Rothlein and Rapp (2014) applied RSA to differentiate the stroke-motoric and visual features of English letters in passive reading. They found motoric encoding in the left IPS and visual encoding in the posterior occipitotemporal and precuneus regions, suggesting motoric and visual representations in distinct regions for English letter processing⁵⁶. Nevertheless, their use of simple letter stimuli may not reflect typical processing of words or morphemes in reading. Chinese characters offer a unique window to explore motoric/sensory representations in visual word processing, as single-syllable characters could form meaningful words, serving as the basic units in natural reading.

Moreover, the complex structure of Chinese characters likely links reading acquisition in Chinese more tightly to handwriting practice^{43,44}, providing a valuable context to investigate the interaction between the representations of writing-motor and visual processing systems.

Taken together, substantial evidence suggests that language-motor areas contribute to both auditory and visual modalities of language perception, though the precise nature of these representations is not yet fully understood. Notably, these motor areas exhibit a modality-bound encoding pattern, consistent with the sensorimotor hypothesis of embodied cognition^{57–60}, which suggests that motor regions mirror the sensorimotor experience of specific language processing modalities in language perception. Specifically, speech-motor areas likely encode articulatory or acoustic information linked to speaking and listening, while writing-motor areas are suggested to encode writing programs or visual features of written words related to handwriting and reading. Nevertheless, most studies have focused on individual modalities, with little exploration of how these language-motor regions function across auditory and visual modalities. Existing studies have explored neural activation patterns between speech and reading, emphasizing distinctions between unimodal versus multimodal regions, and modality-specific effects in language comprehension^{61–65}. Despite differences in functional localization or activation strength, speech perception and reading may share consistent representational patterns, such as in processing high-level semantic information⁶⁵. However, no study has systematically compared how language-motor areas represent sensory and motoric features in auditory and visual language processing in a comparable experimental framework.

The primary aim of this study was to characterize the neural representations in speech-motor areas during auditory speech perception and in writing-motor areas during visual word perception. We focused on the phonetic and orthographic sensorimotor features tied to the articulatory and handwriting motor programs involved in producing these stimuli, as well as their auditory and visual forms. These features are fundamental to higher-order language processes like lexical and semantic processing⁶⁶. To ensure comparability, we employed consistent experimental designs and analysis protocols for both modalities, including a mixed block and event-related design for robust item-based analysis^{57,68}, and RSA to evaluate encoding patterns within language-motor regions. Using fMRI, we measured blood-oxygen-level dependent (BOLD) responses in language-motor areas during perception tasks in two modalities: auditory speech syllable perception (SP) and visual character perception (CH). To identify speech-motor and writing-motor areas, participants also performed language production localizer tasks for syllable articulation (AR) and finger writing (WR). Speech and character stimuli were categorized along motoric and sensory dimensions, enabling the construction of Representational Dissimilarity Matrices (RDMs). RSA was then applied to assess correlations between neural activity and motoric/sensory features of language stimuli⁵⁵, reflecting motoric or sensory encoding.

More importantly, a cross-modality comparison of results allowed us to explore whether the sensorimotor system generally encodes modality-bound motoric (production-related) or sensory (perception-related) features across modalities, or it exhibits distinct representation patterns for each modality. One hypothesis posits that language-motor areas share general sensorimotor encoding patterns across modalities to support recognition³⁹, consistent with the sensorimotor hypothesis of embodied cognition⁶⁰. Alternatively, motor areas may exhibit modality-specific encoding patterns to meet the unique demands of auditory and visual perception. We hypothesize that speech-motor areas may exhibit stronger motoric or sensory encoding effects in speech perception than writing-motor areas during visual character perception. Two key factors likely support this hypothesis: (1) Speech perception involves strengthened motor-sensory interaction due to evolutionary or innate production-perception-coupling mechanisms^{2,19,20}. This coupling might be further reinforced by the more frequent co-occurrence of perception and production compared to reading-writing (especially in mature adult readers); (2) Speech perception typically imposes stricter temporal demands,

necessitating rapid sensorimotor encoding^{27,69–72}, while reading generally allows more processing time and may exhibit subtler sensorimotor encoding effects.

Our multivariate RSA results revealed shared motoric representational patterns in language-motor regions across modalities (i.e., right preCG for SP task and left SFG for CH task, respectively), supporting the involvement of language-motor areas in encoding production-related features during both auditory and visual language perception. Additionally, motoric representations were also observed in sensory areas in both modalities (left STG and HG for SP task, and right SOG for CH task); this motoric encoding within sensory regions was correlated with motor regions, highlighting a dynamic sensory-motor interaction that facilitates integrative language processing across modalities. However, sensory encoding demonstrated a modality-specific representational pattern, as it was observed only in the auditory task within both motor and sensory regions (left IFG and STG), but not in the visual task. These findings align with the embodied hypothesis of language processing^{57–60} and suggest that motor areas may play differential roles in sensory encoding during auditory versus visual tasks, potentially shaped by early developmental or genetic factors and the distinct demands of each sensory modality.

Results

Language-motor areas were identified using two block-designed localizer tasks: a syllable articulation (AR) task to localize speech-motor areas (for experimental paradigm see Fig. 1a, c) and a finger writing (WR) task to localize writing-motor areas (for experimental paradigm see Fig. 1b, d; for material details see Table S1). Participants' neural (BOLD) responses and task performance were examined in two modality-specific language perception tasks. In the auditory speech perception (SP) task, participants listened to syllables or reversed speech sounds (auditory control) spoken by a male speaker, and responded to target female voices (Fig. 1e). In the visual character perception (CH) task, participants viewed Chinese characters or scrambled visual stimuli (visual control) in black font, and responded to infrequent gray-colored targets (Fig. 1f). These perception tasks utilized a mixed block and event-related design^{57,68}, enabling univariate analysis of sensorimotor regions involved in perception, as well as item-based representational similarity analysis (RSA) to investigate neural representational patterns (for analysis steps see Fig. 2). Stimuli in each task were systematically categorized based on motoric and sensory properties to ensure comparisons across modalities (for experimental paradigm see Figs. 1g, h and 2a, b; for material and value details see Tables S2–S9). This design provided a framework to examine modality-general and modality-specific encoding of motoric and sensory features in language-related motor and sensory regions (for details, see Methods).

Behavioral performance of perception tasks

The average accuracy was 94.17% (SD = 6.34%) for auditory detection and 92.22% (SD = 6.54%) for visual detection, indicating sustained attention during both tasks.

Brain activation in articulation and finger writing tasks

To explore brain activation patterns in the language production localizer tasks, we employed a general linear model (GLM) with boxcar regressors for each task. Syllable articulation and character writing were contrasted with their respective motor controls (lip movement and drawing) to generate individual-level contrast images. These contrast images were then used in group-level analyses to identify speech-motor and writing-motor areas (see Methods for the preprocessing and modeling details). Results were reported at $p < 0.05$, FDR corrected, at the whole-brain level.

The results showed that, the AR task elicited syllable-production-related activation (*vs.* lip movement) in typical frontal speech-motor regions, including the left precentral gyrus (preCG) and the triangular part of the left inferior frontal gyrus (IFGtri), as shown in Fig. 3a (voxel-level: $n = 29$, $t = 4.084$, $p < 0.05$ FDR-corrected). The WR task showed character-writing-related activation (*vs.* drawing) predominantly in parietal and

prefrontal regions, encompassing the bilateral superior frontal gyrus (SFG), left supplementary motor area (SMA), bilateral superior parietal lobules (SPL), left inferior parietal lobule (IPL), and right cerebellum, as shown in Fig. 3b (voxel-level: $n = 29$, $t = 3.432$, $p < 0.05$ FDR-corrected). The activation patterns were in line with previous studies both in speech production^{7,11,49,69,73} and writing tasks^{47,74–77}. Articulation activated more ventral frontal brain regions, whereas writing activated more dorsal frontal and parietal brain regions (refer to Table S10 for peak activation coordinates and effect sizes).

Brain activation in speech syllable and visual character perception tasks

Likewise, GLMs were conducted to investigate brain activation patterns in the perception tasks. At the individual level, spoken syllable perception was contrasted with reversed speech perception (auditory control) and rest (fixation) in SP task, and visual character perception was contrasted with scrambled character perception (visual control) and rest (fixation) in CH task. Participant-specific contrast images were then used in group-level analyses to identify regions involved in auditory and visual language perception (see Methods). Results were reported at $p < 0.01$ (uncorrected) for the “speech/character *vs.* control” contrasts to address weaker motor activations in passive perception tasks^{11,12,27,38}, and at $p < 0.05$ (FDR corrected) for the “speech/character *vs.* rest/fixation” contrasts.

As shown in Fig. 3c, in the SP task, listening to intelligible speech syllables elicited greater activation mainly in typical ventral frontal sensorimotor regions and the superior temporal regions (spoken syllable *vs.* reversed speech perception: $n = 29$, $t = 2.462$, $p < 0.01$ uncorrected; spoken syllable *vs.* rest/fixation: $n = 29$, $t = 2.832$, $p < 0.05$ FDR-corrected; refer to Table S11 for peak activation coordinates and effect sizes). Regions that were significantly activated in auditory syllable perception included the bilateral precentral and postcentral gyrus (preCG/postCG), right superior frontal gyrus (SFG), the opercular part of the bilateral inferior frontal gyrus (IFGoper), bilateral insula (Ins), bilateral supplementary motor area (SMA), bilateral superior temporal gyrus/Heschl's gyrus (STG/HG), and the right supramarginal gyrus (SMG).

As shown in Fig. 3d, in the CH task, viewing visual character activated a broader network, covering frontal, temporal, parietal, and occipital regions, during visual character perception (visual character *vs.* scrambled character perception: $n = 29$, $t = 2.462$, $p < 0.01$ uncorrected; visual character *vs.* rest/fixation: $n = 29$, $t = 3.396$, $p < 0.05$ FDR-corrected; refer to Table S12 for peak activation coordinates and effect sizes). The brain activations for visual character perception were mainly found in the bilateral precentral and postcentral gyrus (preCG/postCG), left superior/inferior frontal gyrus (SFG/IFG), bilateral middle frontal gyrus (MFG), left insula (Ins), bilateral supplementary motor area (SMA), bilateral superior/middle/inferior temporal gyrus (STG/MTG/ITG), left superior occipital gyrus (SOG), bilateral inferior occipital gyrus (IOG), bilateral lingual and fusiform (Ling/Fusi), bilateral inferior parietal lobule (IPL), and bilateral supramarginal gyrus (SMG).

The brain activation patterns for perception tasks indicate the involvement of both sensory and motor systems in language perception across auditory and visual modalities, consistent with earlier research on the neural correlates of auditory speech perception^{7,11,27} and visual character/word perception^{43,44,50}.

Representational patterns in speech syllable and visual character perception tasks

To investigate the motoric and sensory representational patterns during language perception, RSA was performed in key regions of interest (ROIs) identified from the production and perception task results. These ROIs included speech-motor areas (bilateral preCG, IFG, Insula, SMA), writing-motor areas (bilateral SFG, MFG, preCG, SPL/IPS), auditory processing areas (bilateral HG, STG), and visual processing areas (bilateral ITG-Fusiform, IOG, SOG). Neural RDMs were constructed by selecting from the top 100 most active voxels within each ROI (as defined by AAL anatomical

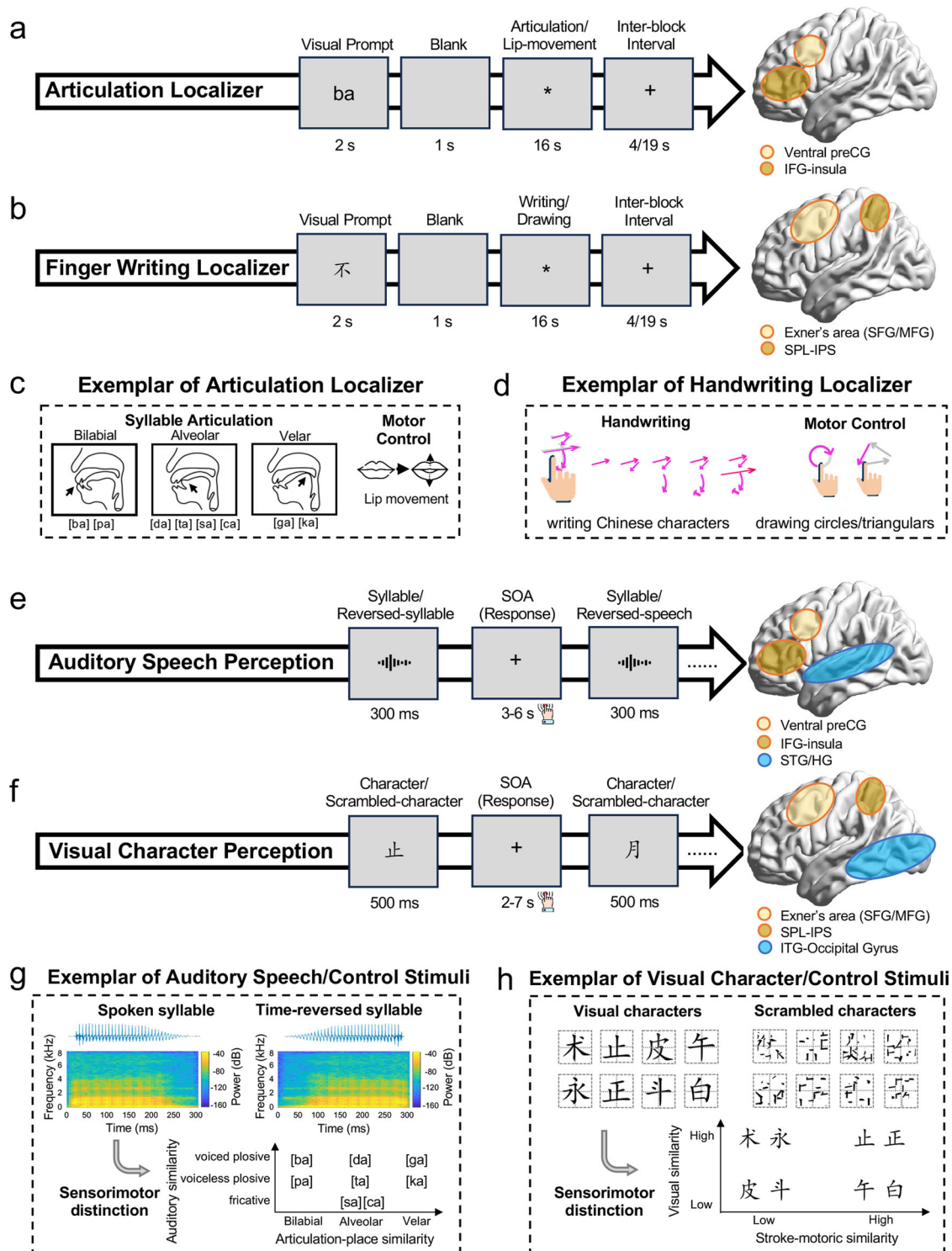


Fig. 1 | Experimental design and example stimuli. **a** Experimental design for articulation (AR) localizer task. Participants were instructed to repeatedly articulate the prompted syllables or move their lips up and down during the asterisk display. The functional localization hypotheses for the activated speech-motor areas were shown with orange circles. **b** Experimental design for finger writing (WR) localizer task. Participants were instructed to repeatedly write the prompted characters (listed in Table S1) or draw simple shapes with their right index finger during the asterisk display. The functional localization hypotheses for the activated writing-motor areas were shown with orange circles. **c** Example stimuli for AR task. **d** Example stimuli for WR task. **e** Experimental design for auditory speech perception (SP) task. Participants heard intelligible syllables or unintelligible reversed speech signals, responding to detect female voice in 1/8 of all trials. The functional localization hypotheses for

the activated speech-motor regions and auditory processing regions were shown with orange and blue circles, respectively. **f** Experimental design for visual character perception (CH) task. Participants viewed characters or scrambled characters, responding to detect gray-colored stimuli in 1/8 of all trials. The functional localization hypotheses for the activated writing-motor regions and visual processing regions were shown with orange and blue circles, respectively. **g** Example stimuli and motoric/sensory categorization for SP task. **h** Example stimuli and motoric/sensory categorization for CH task (listed in Table S3). *Abbreviations:* SOA = stimulus onset asynchrony; preCG = precentral gyrus; IFG = inferior frontal gyrus; HG = Heschl's gyrus; STG = superior temporal gyrus; SFG = superior frontal gyrus; MFG = middle frontal gyrus; SPL/IPS = superior parietal lobule/inferior parietal sulcus; ITG = inferior temporal gyrus. See also the abbreviations of brain regions in Table S14.

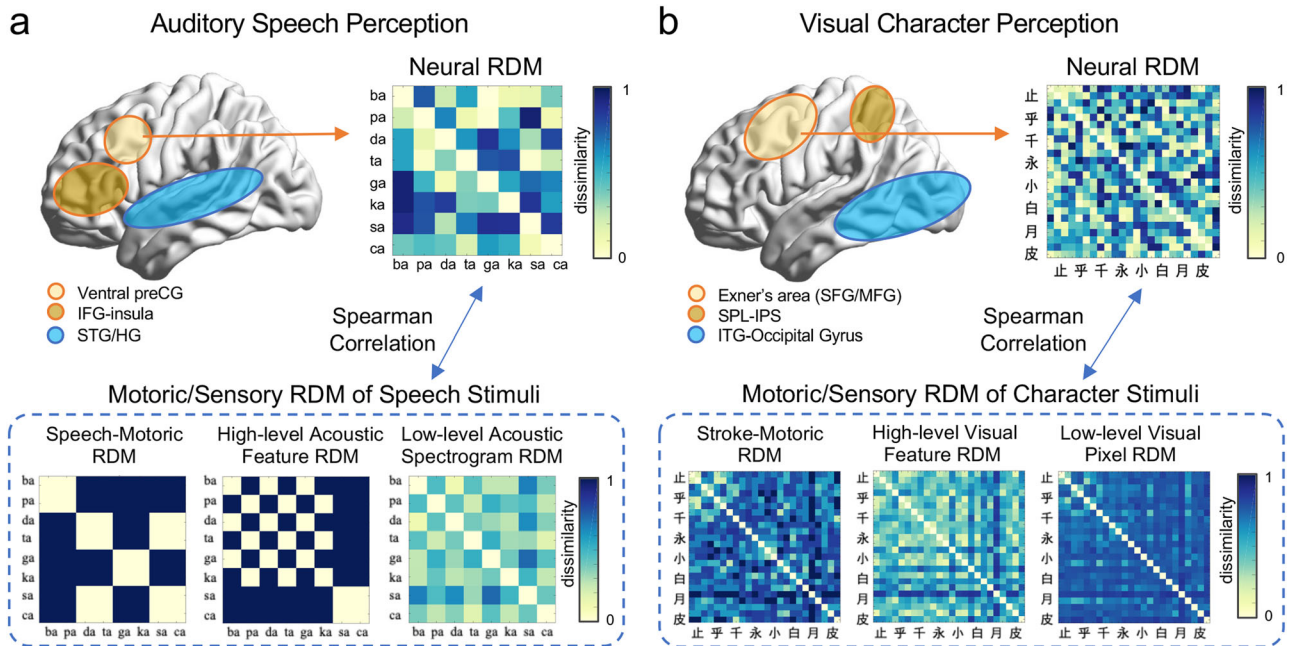


Fig. 2 | Representational similarity analysis (RSA) steps for motor and sensory regions of interest (ROIs). **a** Representational similarity analysis (RSA) steps for the auditory speech perception (SP) task. For the key speech-motor regions and auditory processing regions (orange and blue circles illustrating the hypothesized locations), neural activity patterns were assessed using item-based general linear model (GLM) *t*-maps. Activity patterns were compared across items using Pearson correlation, and the resulting values were subtracted from 1 to generate neural representational dissimilarity matrices (RDMs). These neural RDMs were then correlated with the speech-motoric RDM, the high-level acoustic feature RDM, and the low-level acoustic spectrogram RDM, respectively, to explore the contents of neural representations. Detailed values were provided in Tables S4–S6. **b** RSA steps for the visual

character perception (CH) task. For the key writing-motor regions and visual processing regions (orange and blue circles illustrating the hypothesized locations), neural activity patterns were assessed using item-based GLM *t*-maps. Activity patterns were compared across items using Pearson correlation, and the resulting values were subtracted from 1, resulting in neural RDMs. These neural RDMs were correlated with the stroke-motoric RDM, the high-level visual feature RDM, and the low-level visual pixel RDM. Detailed values were provided in Tables S7–S9. **Abbreviations:** preCG = precentral gyrus; IFG = inferior frontal gyrus; HG = Heschl’s gyrus; STG = superior temporal gyrus; SFG = superior frontal gyrus; MFG = middle frontal gyrus; SPL/IPS = superior parietal lobule/inferior parietal sulcus; ITG = inferior temporal gyrus. Also see the abbreviations of brain regions in Table S14.

templates), based on activation patterns from perception task. Predictor RDMs captured motoric and sensory features of the stimuli, including speech-motoric, high-level acoustic, and low-level acoustic features for SP task, and stroke-motoric, high-level visual, and low-level visual features for CH task (Fig. 2a, b). For each ROI, the relationship between neural and feature RDMs was assessed using a two-stage permutation test^{78,79}. At the individual level, feature RDMs were shuffled 100 times to create a null distribution of randomized correlations with neural RDMs. At the group level, one randomized correlation from each participant’s null distribution was sampled, and the group mean was calculated. This process was repeated 10,000 times to generate a null distribution of group-level correlations. *P*-values were corrected for multiple comparisons using the FDR method. The RSA results are shown in Fig. 4.

When listening to auditory syllables (Fig. 4a), speech-motoric dissimilarity predicted neural dissimilarity in right preCG ($z = 0.0990$, $p = 0.009$), left HG ($z = 0.0697$, $p = 0.0401$), and left STG ($z = 0.0775$, $p = 0.0227$). Moreover, the effect of high-level speech-acoustic dissimilarity was observed in both left IFG ($z = 0.0638$, $p = 0.0461$) and left STG regions ($z = 0.0783$, $p = 0.0248$), while the effect of low-level acoustic spectrogram dissimilarity was only found in right HG ($z = 0.0610$, $p = 0.0309$). No correlation effects were observed in other ROIs ($ps > 0.057$).

In viewing visual characters (Fig. 4b), significant correlations between stroke-motoric dissimilarity and neural dissimilarity were found in left SFG ($z = 0.0196$, $p = 0.0297$) and right SOG ($z = 0.0175$, $p = 0.0414$). No motoric effects were observed in other ROIs ($ps > 0.122$). Neither the high-level visual feature nor low-level visual pixel RDMs could predict the neural RDMs in any writing-motor ROIs or visual processing ROIs ($ps > 0.267$).

To further validate our findings, we examined the representational connectivity⁸⁰ between the motor and sensory ROIs that demonstrated

significant feature encoding during language perception tasks (Figure S1). We computed the correlations between motor-ROI neural RDMs and sensory-ROI neural RDMs using the top 100 most active voxels. Group-level significance was assessed by a permutation approach similar to RSA. Specifically, group-mean correlations between neural RDMs from different ROIs were compared against a null distribution generated from 50,000 random correlations (derived from 100 shuffled correlations per participant) for all ROI pairs. *P*-values were corrected using the Bonferroni method for multiple comparisons. For the SP task, areas involved in articulatory motoric processing exhibited significant cross-regional correlations, including left STG vs. right preCG ($z = 0.166$, $p = 0.0001$), left HG vs. right preCG ($z = 0.201$, $p = 0.0001$), and left STG vs. left HG ($z = 0.290$, $p = 0.0001$). Additionally, for high-level acoustic processing, significant correlations were observed between left IFG vs. left STG ($z = 0.223$, $p = 0.0001$). Similarly, for the CH task, areas involved in stroke-motoric processing also showed significant correlations (left SFG vs. right SOG: $z = 0.094$, $p = 0.0005$). These results further support the notion of shared motoric/sensory representations between motor and sensory regions in language perception.

Discussion

The present study investigated brain activity during auditory speech and visual character perception, corroborating existing literature on the involvement of language-motor and sensory regions in these processes, in both auditory^{7,11,14,27} and visual modalities^{43,44,50,81}. Using RSA, we observed that the language-motor areas, along with some sensory regions, exhibited a general motoric representation pattern across both modalities, while sensory representation was observed solely during auditory perception.

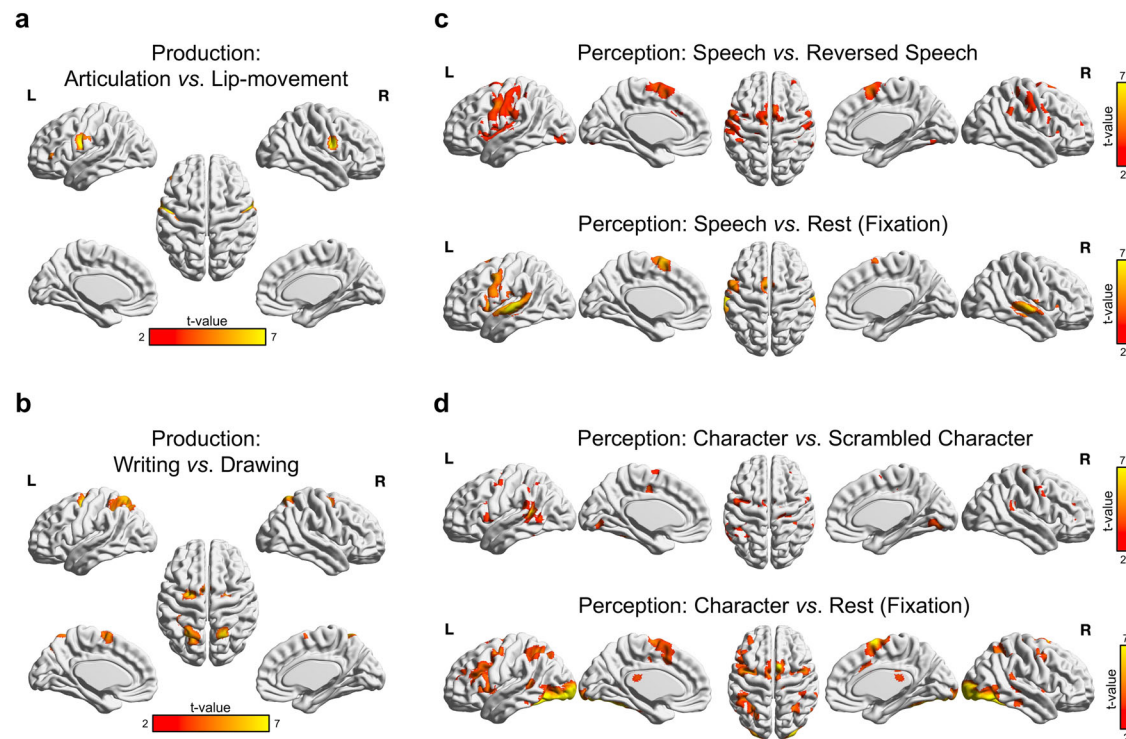


Fig. 3 | Univariate results for language production localizers and perception tasks. **a** Univariate general linear model (GLM) results for speech production (syllable articulation contrasts with lip-movement). **b** Univariate GLM results for character production (finger writing contrasts with drawing). For both **a** and **b**, brain activations (t -value maps) for language production were displayed with superior, lateral and medial surface-rendered views. These results were thresholded at voxel-corrected $p < 0.05$ with an FDR correction method for multiple comparison (cluster extent threshold = 10 voxels). The peak coordinate results and effect sizes for each cluster (*Cohen's d* and *Hedges' g*) were reported in Table S10. **c** Univariate GLM results for auditory speech perception (spoken syllable

contrasts with reversed speech or rest/fixation). **d** Univariate GLM results for visual character perception (visual character contrasts with scrambled character or rest/fixation). For both **c** and **d**, brain activations (t -value maps) for language perception were displayed with superior, lateral and medial surface-rendered views. These results were thresholded at uncorrected $p < 0.01$ at the voxel level (cluster extent threshold = 10 voxels) for the contrast between speech/character and control conditions, and at $p < 0.05$ with FDR correction for contrasts between speech/character and rest/fixation. The peak coordinate results and effect sizes for each cluster (*Cohen's d* and *Hedges' g*) were reported in Tables S11–S12. *Abbreviations:* L = left; R = right.

In speech perception, the right preCG exhibited representational patterns associated with the place of articulation features of spoken syllables. While previous research has reported potential motoric encoding in the left preCG^{11,12,26} or bilaterally^{14,27,82,83}, our findings provide direct evidence for motoric representations within speech-motor areas during speech perception using RSA⁸⁴. This result aligns with theories proposing that motor mechanisms are integral for speech perception^{9,10,85}. Notably, the motoric representation effect in the right preCG contrasts with earlier reports emphasizing left-hemisphere dominance. A plausible explanation is that the left speech-motor area assumes a dominant role in high-demand contexts^{69,73}, while the right area might compensate in low-demand conditions, such as passive listening or simple tasks. This hypothesis is consistent with prior findings that the right speech-motor area encodes phoneme-specific information in quiet but not noisy situations⁶⁹. Additionally, even during passive listening, participants might engage in implicit imitation of the heard speech, activating bilateral sensory-motor mechanisms, as demonstrated in Cogan et al.'s¹⁴ ECoG study on overt tasks like listen-speak and listen-imitation. This bilateral sensory-motor transformation potentially provides a unified interface essential for speech production, acquisition, and self-monitoring¹⁴. Future research should explore whether this lateralization pattern reflects distinct processes of sensory-motor integration in speech perception.

In visual word perception, motoric encoding was observed in the left SFG, likely corresponding to Exner's area, a critical center for handwriting^{46,47}. The SFG/Exner's area is believed to link orthographic representations and handwriting motor commands^{30,43,44,50,74,75}. Moreover, this region has shown sensitivity to orthographic irregularities, suggesting its

role in complex orthographic-motor transformations⁸⁶. The correlation between SFG activity and stroke-motoric features provides direct evidence for motoric representation in this area. Interestingly, such motoric representation was not evident in the left SPL/IPS, which has been suggested to process motoric writing-sequence information in visual word reading^{38,42,50,56,87}. Previous RSA research on English letters reported stroke-motoric sensitivity in the left IPS during reading⁵⁶, while our findings highlight SFG/Exner's area for processing stroke-motoric features in Chinese characters. This divergence may reflect differences in representational contents encoded within these writing-motor areas⁷⁴, or distinct neural mechanisms underpinning different writing systems, such as alphabetic letters (English) versus logographic characters (Chinese)⁴⁴. It is important to explore how sensorimotor neural networks adapt to the varied demands of diverse languages and writing systems to further clarify the functional roles of these regions in visual word perception.

Remarkably, motoric representations were also identified in sensory processing regions. In speech perception, motoric representations were observed in the left STG and HG, regions known for fine-grained acoustic processing during speech recognition^{27,88,89}. These findings are consistent with research indicating that auditory cortex encodes articulatory motoric information, during both speech perception²⁷ and speech production²⁸. In visual character perception, motoric representation was also detected in the right SOG. The right SOG is typically associated with intensive visual-spatial analysis in orthographic processing^{90–92}. Our result suggests that visual processing regions may encode motoric information related to handwriting, potentially facilitating visual-spatial analysis required for written characters processing⁵⁰.

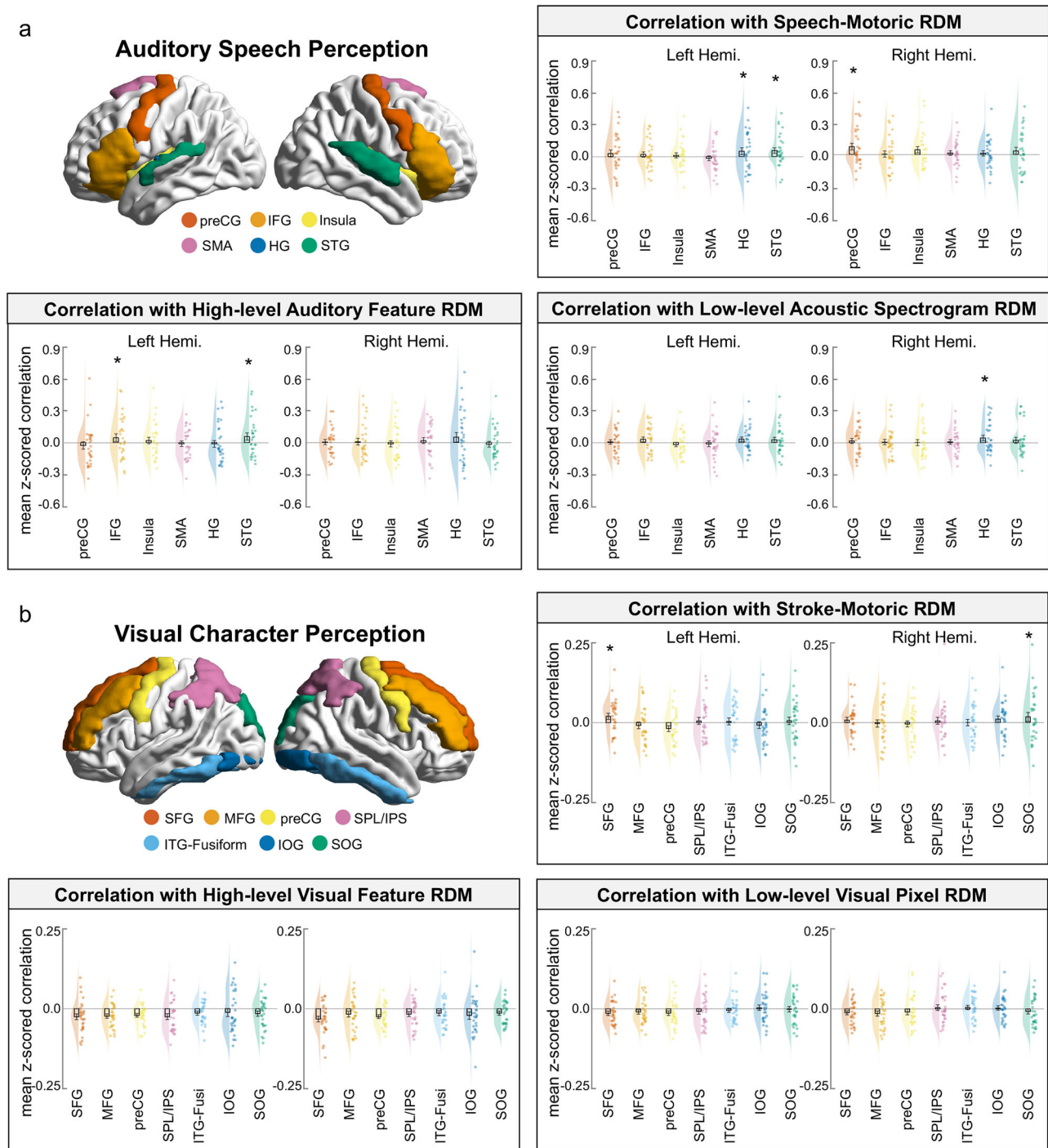


Fig. 4 | Representational similarity analysis (RSA) results within motor and sensory regions of interest (ROIs) for perception tasks. a Representational similarity analysis (RSA) results for auditory speech perception (SP) task. The localization of speech-motor and auditory processing regions of interest (ROIs) was shown with the AAL anatomical template, where the top 100 active voxels for each ROI were selected for item-based RSA. Bar plots (and error bars) indicate mean values (and standard errors) of *Fisher z*-transformed correlation effects between neural representational dissimilarity matrices (RDMs) and speech-motoric RDM, high-level acoustic feature RDM, and low-level acoustic spectrogram RDM across participants, in each ROI at left and right hemispheres. **b** RSA results for the visual character perception (CH) task. The localization of writing-motor and visual processing ROIs was shown with the AAL anatomical template, where the top 100 active voxels for each ROI were selected for item-based RSA. Bar plots (and error bars) indicate mean values

(and standard errors) of *Fisher z*-transformed correlation effects between neural RDMs and stroke-motoric RDM, high-level visual feature RDM, and low-level visual pixel RDM across participants, in each ROI at left and right hemispheres. For both **a** and **b**, a half violin plot was overlaid behind each bar plot to depict the distribution of the data, and individual data points were represented by light-colored circles to the right of the bar plots. Asterisks indicate significant correlation effects ($*p < 0.05$ with FDR correction). *Abbreviations:* Hemi. = hemisphere; preCG = precentral gyrus; IFG = inferior frontal gyrus; SMA = supplementary motor area; HG = Heschl’s gyrus; STG = superior temporal gyrus; SFG = superior frontal gyrus; MFG = middle frontal gyrus; SPL/IPS = superior parietal lobule / inferior parietal sulcus; ITG-Fusi = inferior temporal gyrus and fusiform gyrus; IOG = inferior occipital gyrus; SOG = superior occipital gyrus. Also see the abbreviations of brain regions in Table S14.

Grounded in the above findings, our study suggests that both auditory and visual language perception engage motoric representations tied to language production³⁹. These representations, observed in both motor and sensory regions, reflect a shared motoric encoding mechanism within a cooperative sensorimotor network. Furthermore, the results of cross-region representational connectivity suggest that motor areas (e.g., right preCG in SP task, or left SFG in CH task) work closely with sensory regions (e.g., left STG and HG in SP task, or right SOG in CH task) in encoding motoric features, possibly supporting the integrative/interactive processing through reciprocal connections^{27,50}. Earlier studies have demonstrated robust connections between motor and sensory systems during language perception across both auditory^{15,70,93} and visual modalities⁴⁰. Moreover, in speech perception, motor areas (e.g., IFG and preCG) have been observed to exert top-down modulatory effects on temporal auditory regions (such as STG)^{69,70}. These interactions suggest that motoric representations in sensory areas may originate from motor regions, supporting a sensorimotor integration framework rather than a purely bottom-up model of language processing.

Sensory areas may encode motoric features that are either identical or complementary to those represented in the motor areas^{26,38,43,69,73}. While motor areas possibly prioritize the processing of movement planning features, sensory areas likely play a sensory-motor transformation role, encoding abstract acoustic or visual counterparts of these movements. For example, during speech imitation, the auditory cortex encodes both acoustic formant and articulatory motoric features, indicating a bimodal representation pattern that bridges sensory input and motor output²⁸. This mechanism, rooted in feedback between motor and sensory systems during natural language communication, likely enhances both language perception and production⁹⁴. Although less explored in visual word processing, similar feedback mechanisms are supposed to exist between the visual and motor systems, as frequent transformations between visual shapes and writing programs could strengthen sensorimotor connections^{30,41}. This may be supported by findings that early-blind adults and patients who undergo surgical restoration of vision demonstrate remarkable plasticity in the visual temporal-occipital cortex, which adapts to process tactile and motoric information, compensating for visual deficits and possibly relating to multimodal feedback between visual, tactile and motor systems^{95,96}.

Moreover, this sensory-motor representation/transformation mechanism may become more pronounced under challenging conditions to facilitate the recognition of syllables/characters. For instance, Du et al.⁶⁹ found that both speech-motor regions and the auditory cortex demonstrated phoneme discriminability, with motor areas enhancing discrimination in noisy conditions to help with speech recognition. These enhanced motoric representations may also play a predictive role in aiding perception^{69,73}. Similarly, viewing handwritten words more strongly engages both visual processing and writing-motor areas compared to printed words^{36,97}. In this study, while the use of simple stimuli and passive listening minimized explicit motor engagement, it potentially reduced the need for motoric encoding or prediction. Future studies should explore the effects of task complexity and investigate the spatio-temporal connectivity between motor and sensory regions to elucidate the interplay between motoric and sensory representations under varying conditions.

The observed motoric representations within language-related motor and sensory regions align with the sensorimotor hypothesis proposed by the theory of embodied cognition^{57–60}. This theory suggests that language cognition is deeply intertwined with the bodily sensory and motor experiences in the environment, such as learning words through speaking or handwriting. Consequently, the internal representations of language stimuli may comprise production-related motoric information, which could be retrieved and activated within the language-motor system during both auditory^{11,12,26,27} and visual language perception^{43,50}. Understanding this relationship may also illuminate how children acquire language in different modalities, as they often develop perceptual language skills concurrently with motor skills, including speaking^{3,4,98} and handwriting^{30,35}.

In contrast to motoric representations, our findings reveal distinct patterns of sensory representations in motor areas for auditory and visual modalities. During speech perception, significant sensitivity was observed in the left IFG and left STG to high-level acoustic features (categorized by the manner of articulation), and in the right HG to low-level acoustic spectrogram information, consistent with previous research²⁷. However, no analogous sensory encoding effect was observed in either sensory or motor regions for visual character perception, despite previous evidence of visual encoding in occipitotemporal regions^{43,56}. Additionally, the visual task demonstrated weaker activation and feature-brain correlations compared to the auditory task.

This divergence likely arises from the inherent distinctions between auditory and visual language processing. In speech perception, sensory encoding plays a fundamental role, integrating closely with production processes^{11,12,27} from early language acquisition stages. Sensorimotor integration is evident even in pre-babbling infants, where speech perception and production systems mutually reinforce each other to support the development of robust speech perceptual abilities^{1–4,14}. Furthermore, auditory speech comprehension often demands fine sensory discrimination to overcome environmental noise, strengthening the sensorimotor network during auditory tasks⁶⁹.

In contrast, visual language processing, particularly reading comprehension, appears to rely less on sensorimotor coupling or detailed visual encoding, as evidenced by the absence of sensory representation effects and weaker activation/representation effects for visual character perception in our study. Interestingly, motoric representations were more prominent during visual character perception, contrasting with prior behavioral findings emphasizing the role of visual analysis over motoric information in word recognition^{33,52,53}, while consistent with research showing increased motor area involvement in individuals with extensive handwriting experiences^{30,31,43,44,51}. However, the overall reduction in sensorimotor encoding may reflect the gradual development of sensorimotor integration in the visual modality, which becomes more pronounced with handwriting practice, compared to the more innate production-perception coupling in speech.

Nevertheless, the observed modality differences could also be influenced by additional factors. For example, the relative simplicity of the visual task may have lowered the reliance for detailed sensory encoding, particularly given the absence of visually similar distractors or high visual interference. Moreover, stroke-motoric representations might facilitate visual-spatial processing, potentially reducing the cognitive load required for visual analysis^{36,87}. Additionally, differences in stimulus presentation between tasks could also contribute to the observed disparities. The auditory task utilized a smaller stimulus set with more repetitions, likely leading to more stable sensory and motoric representations. In contrast, the visual task involved a larger stimulus set with fewer repetitions, potentially resulting in less robust estimates of these representations. These factors suggest that while our findings highlight differences in the neural representations of motoric and sensory features across modalities, they should be interpreted within a broader context of task complexity and stimulus characteristics.

Our results contribute to the growing body of literature on the multimodal nature of language perception, emphasizing the role of motor system in supporting motoric and sensory representations across modalities^{26,27,43,50}. The motor system's engagement in language perception reflects a general mechanism of sensorimotor collaboration, yet its involvement is adaptive, likely varying with the degree of sensory-motor coupling demanded by auditory versus visual processing modalities. The pronounced sensory encoding in both motor and sensory regions during speech perception, contrasted with its absence in visual character perception, suggests modality-dependent reliance on sensorimotor integration networks. Auditory perception appears to depend more heavily on sensorimotor interactions, possibly due to its evolutionary basis or the need for precise auditory discrimination, especially under challenging conditions like noisy environments^{219,20}. These findings advance our understanding of language comprehension within sensorimotor networks and offer practical

implications for developing interventions for language disorders or optimizing language learning strategies. Additionally, our study employed a modality-comparable paradigm and used unified motoric and sensory features to explore sensorimotor encoding, providing a foundation for future cross-modality investigations.

While our study sheds light on motor regions' automatic engagement under natural perception conditions, its scope might be limited by the use of simple stimuli (monosyllabic sounds and single-radical characters) and passive perception tasks. These conditions likely have lessened sensory processing demands, resulting in weaker sensorimotor encoding effects, especially for visual characters with minimal visual interference. Motor regions may engage selectively in language processing under diverse demanding conditions^{22–24}. Future research could explore how sensorimotor representations adapt to varying perceptual difficulties, such as noisy speech or handwritten words, which could intensify motor area engagement and enhance motor-sensory connectivity^{69,97}. Additionally, the current design does not address higher-order language processes, such as semantic, syntactic, or prosodic processing, where motor areas might demonstrate significant involvement^{99–102}. Exploration of how basic sensorimotor representations in language-motor areas influences the higher-order language processes remains an exciting direction for future research. Moreover, while fMRI provides insights into the spatial patterns of sensory-motor integration, its limited temporal resolution precludes analysis of dynamic interplay between language-motor and sensory systems. Future studies using high temporal resolution techniques like MEG/EEG or multimodal MEG-fMRI could capture the fine-grained temporal-spatial dynamics of sensory-motor integration across modalities, which would further enhance our understanding of how motor and sensory systems coordinate during real-time language comprehension.

Methods

Participants

Thirty-six native Mandarin-Chinese-speaking participants (mean age = 19.75 years, SD = 2.63, range 18–30 years; 19 females, 17 males) were recruited in the current study. All participants were right-handed, reported normal hearing and normal or corrected-to-normal vision, and were neurologically healthy. The study was approved by the Ethic and Human Protection Committee of Shenzhen University. Participants provided informed written consent before the formal experiment. All ethical regulations relevant to human research participants were followed.

Experimental Design and Procedure

Participants completed four tasks: speech perception (SP), visual-character perception (CH), articulation localizer (AR) and finger writing localizer (WR) tasks. To minimize task order effects, participants were randomly assigned to one of four sequences (SP-CH-AR-WR; CH-SP-WR-AR; AR-WR-SP-CH; WR-AR-CH-SP). Response hands for SP and CH tasks were counterbalanced among participants. Each task included a brief practice to ensure comprehension of task instructions. Stimulus presentation was controlled using custom-written MATLAB scripts with Psychtoolbox¹⁰³.

Production Localizer Tasks. The language production localizer tasks aimed to identify language-motor regions of interest (ROIs). Both AR and WR tasks were performed using a block design, each comprising eight language-motor blocks and eight control blocks presented in an interleaved manner. Each block was initialized with a 2-s visual cue that prompted the stimuli for participants to speak or write, and then a 1-s blank period, followed by a 16-s response period with an asterisk on the screen. The inter-block intervals had durations of 4 s, 4 s, 4 s, and 19 s in every four blocks. A fixation cross was displayed throughout the intervals (Fig. 1a, b).

The AR localizer included eight speech-motor blocks and eight lip-movement (control) blocks. In each speech-motor block, participants saw a cue of a written consonant-vowel (CV) syllable (e.g., /ba/) on the screen and silently repeated the syllable with minimal movement during the asterisk

display. Eight CV syllables were used (i.e., /ba/, /pa/, /da/, /ta/, /ga/, /ka/, /sa/, /ca/), identical to those in SP task (see stimuli details in SP task). In each lip-movement (control) block, participants viewed a cue of a line-drawn mouth and repeatedly moved their lips up and down minimally without articulating syllables (Fig. 1a, c).

The WR localizer included eight writing blocks and eight drawing (control) blocks. In each writing block, participants repeatedly wrote the prompted Chinese characters with their right index finger in the air during the asterisk display. Although this differs slightly from natural handwriting, it was designed to isolate neural regions associated with orthographic and graphomotor processing unique to handwriting (e.g., tracing stroke trajectories) while excluding non-specific sensorimotor components (e.g., pen-holding, visual feedback)^{75,104,105}. The eight characters were selected for familiarity and ease of writing (see details in Tables S1, S2), excluding those used in CH task. In each drawing (control) block, participants repeatedly drew circles or triangles with their right index finger in the air (Fig. 1b, d).

Perception Tasks. Participants completed four SP runs and four CH runs, using a mixed block/event-related design^{67,68}. This experimental design ensured robust statistical power for detecting activation during speech and visual word perception and allowed detailed analysis of sensory-motor representations in language-motor areas with item-based RSA. To focus on motoric/sensory encoding during passive perception, a detection task (in 1/8 of trials, excluded from the main analysis) was applied to maintain participants' attention on stimuli while minimizing additional motor responses and rehearsal.

Each SP run consisted of two syllable perception blocks and two auditory control blocks, interleaved, each lasting 50 s with 15-s inter-block intervals. Ten syllable/control stimuli were presented per block in a random, temporally jittered manner following an event-related design (SOA 3–6 s with a uniform distribution). Each syllable/control stimulus was repeated twenty times across four runs. Trial numbers were optimized to balance participant engagement and data quality within fMRI session limits. More trials were used for speech stimuli to counteract scanner noise and achieve a higher signal-to-noise ratio. Participants listened to syllables spoken by a male speaker and their time-reversed control signals binaurally through earphones. Participants performed an auditory detection task, pressing a button by their left/right hand as quickly and correctly as possible when hearing a female voice in 1/8 of the trials (Fig. 1e).

Each CH run consisted of three character perception blocks and three visual control blocks, interleaved, each lasting 36 s with 15-s intervals. Eight character/control stimuli were presented per block in a random, jittered order (SOA 3–7 s with a uniform distribution). Each stimulus was displayed for 500 ms, followed by a black fixation cross on a light gray background. Each stimulus was repeated four times, once in each of the four runs. Participants focused on the visual stimuli and pressed a button upon detecting dark-gray character/control stimuli among black stimuli (in 1/8 of all trials) (Fig. 1f).

Stimuli for Perception Tasks

To isolate fundamental sensorimotor representations, simple spoken CV syllables and visually-presented Chinese characters were used in SP and CH tasks. Action-related words were excluded to avoid confounding motor-related effects from higher-level lexical/semantic processing^{99,101}.

Speech Perception Task (SP). In the SP task, eight CV syllables (/ba/, /pa/, /da/, /ta/, /ga/, /ka/, /sa/, /ca/) were selected, similar to Cheung et al.²⁶ We substituted the consonant /ʃ/ in English with /c/ in Mandarin Chinese because /c/ more closely resembles other alveolar consonants (/d/, /t/, /s/) in articulation, which are articulated with the anterior part of the tongue, compared to the retroflex fricative /ʃ/. The consonants represented distinct phonetic features based on the place of articulation (bilabial: /b/, /p/; alveolar: /d/, /t/, /s/, /c/; velar: /g/, /k/) and the manner of articulation (voiced plosives: /b/, /d/, /g/; voiceless plosives: /p/, /t/, /k/; fricatives: /s/, /c/). These distinctions enabled classification of syllables

based on both articulator-bound acoustic parameters (for place of articulation) and articulator-free acoustic characteristics (for manner of articulation)^{88,106}. The vowel /a/ was chosen for its better noise resistance in the fMRI scanner. We categorized the three place of articulation features as more indicative of speech-motoric (articulatory) characteristics, while the three manner of articulation features were considered more representative of higher-order acoustic characteristics (Fig. 1g).

Syllables were recorded by a male and a female Chinese speaker in a soundproof chamber at a sampling rate of 22,050 Hz. Post-processing in the Adobe Audition software and MATLAB codes included bandpass filtering (80–10,500 Hz), removal of acoustic transients (clicks) at onset/offset with 10-ms raised cosine ramps, length equalization to 300 ms, and matching the average root-mean-square (RMS) sound pressure level. Time-reversed versions of these speech stimuli served auditory control stimuli (Fig. 1g).

For RSA, item-based similarity matrices were computed in the motoric and auditory domains (Fig. 2a). Speech-motoric similarity was binary-coded (1=similar, 0=dissimilar) based on the place of articulation features in pairs of syllables. Auditory similarity was defined in two dimensions. The first dimension focused on the high-level acoustic feature similarity, which was binary-coded for the manner of articulation²⁶. The second dimension was low-level acoustic spectrogram similarity, which quantified the perceptual similarity of speech sounds. The spectrogram similarity between signals was assessed by computing the Pearson correlation coefficient between the absolute values of their spectrograms. Each signal's spectrogram was computed using Short-Time Fourier Transform (STFT) with a moving Hamming window (1024 points, 75% overlap), providing detailed temporal and frequency resolution for the analysis of short syllables. The resulting RDMs were obtained by subtracting the similarity matrices from 1 (Tables S4–S6). Pearson correlation analysis showed that, the articulatory speech-motoric RDM was not significantly correlated with either the higher-level auditory feature RDM ($r = -0.182, p = 0.352$) or the lower-level acoustic spectrogram RDM ($r = -0.172, p = 0.931$), whereas a moderate correlation was observed between the two acoustic features RDMs ($r = 0.394, p = 0.038$).

Visual Character Perception Task (CH). The CH task used twelve pairs of simple Chinese characters (twenty-four characters), chosen based on their varying levels of stroke-motoric similarity and visual-form similarity (Table S3). These characters were selected from a set of 211 simple characters with a single radical and no more than six strokes. Characters with multiple radicals were excluded to avoid confounding effects caused by high overlap in both visual and stroke-motoric domains, which could obscure distinctions between the targeted features of interest. Motoric representation of characters refers to the abstract, effector-independent movement sequences required to produce the character shapes^{56,107,108}. To account for stroke count, direction, and order, we quantified stroke-motoric similarity using the Levenshtein distance. This metric measures the minimum single-stroke feature edits—insertions, deletions or substitutions (see details of the eleven fundamental strokes in Chinese characters in Table S2)—needed to transform one character's stroke sequence to another's, normalized by the longer sequence length⁵². For instance, the Levenshtein distance between the character “止” (| — | —) and “正” (— | — | —) is 1 (from “止” to “正”, add — in the string) divided by 5 (the larger number of stroke features), so the stroke-motoric similarity between these two characters was $(1-1/5) = 0.8$. For visual form similarity, we quantified both high-level, font-invariant visual features and low-level, font-specific pixel features. High-level visual feature similarity was determined by the ratio of overlapping visual features to the total features of both characters. Based on the existing literature and the decomposition of Chinese character forms/strokes, twenty fundamental visual features were categorized in the current study. Nineteen features are shared among scripts in different languages, including four line types (horizontal, vertical, slanted right, and slanted left), four curve types (open to the top, bottom, left, and right), three intersection types (L, T or X), eight termination types (top, bottom, left, right, top-left, top-

right, bottom-left, bottom-right), and the number of disconnected components^{109–111}. In addition, dots were included as another type of visual feature, as dots are one of the basic components of Chinese characters. These features abstract away from specific font details, providing insight into structural and compositional visual similarities. Low-level visual similarity was assessed through a pixel-based measure to capture the spatial and geometric overlaps between characters. The pixel similarity was obtained by experimenting with various alignments to optimize the character positioning for maximum overlap, and the values were derived from the point of maximum overlap between the two characters. This measure aligns closely with early-stage visual processing, emphasizing the fine-grained distinctions critical for character recognition^{54,112,113}, and is sensitive to the specific font used in experiments (e.g., KaiTi font in our study).

Pairwise stroke-motoric and visual similarities were calculated among the full set of 211 simple characters. High similarity was defined as values exceeding the mean similarity across all character pairs by at least half a standard deviation, while low similarity was below the mean similarity minus half a standard deviation. For stroke-motoric, visual feature, and visual pixel similarities, thresholds for high values were >0.38 , >0.54 , and >0.29 , respectively, and thresholds for low values were <0.21 , <0.39 , and <0.22 , respectively. Three pairs of characters were finally selected for each combination of high/low visual similarity conditions and high/low stroke-motoric similarity conditions (twenty-four characters in total). To avoid semantic processing effects in the motor areas⁹⁹, no characters with semantic relation to body movement were chosen. Examples include “止” and “正” (high similarity in both dimensions), “木” and “永” (high visual similarity but low stroke-motoric similarity), “午” and “白” (low visual similarity but high stroke-motoric similarity), and “皮” and “斗” (low similarity in both measures) (see Tables S3–S9 and Figs. 1h & 2a). The resulting dissimilarity matrices (RDMs) were computed by subtracting the similarity matrices from 1 (Tables S7–S9). Pearson correlation analysis indicated a moderate correlation between the writing stroke-motoric RDM and both the high-level visual-feature RDM ($r = 0.406, p < 0.001$) and the low-level visual-pixel RDM ($r = 0.472, p < 0.001$), as well as between the two visual RDMs ($r = 0.392, p < 0.001$) (Fig. 1h).

The character stimuli were presented as black line drawings (approximately 120×120 pixels) in the KaiTi font. Scrambled versions of these characters served as visual control stimuli, created by dividing each character image into several 30×30 pixel sections and randomizing their positions to produce unintelligible images that retained the overall pixel density and low-level visual features of the original stimuli¹¹⁴ (Fig. 1h).

Data Acquisition and Preprocessing

Magnetic resonance imaging (MRI) images were collected using a 3 T Siemens Prisma scanner at Shenzhen University in China. Functional images were acquired using an interleaved multi-slice echo planar imaging (EPI) sequence (TR = 1000 ms, TE = 30 ms, flip angle = 35° , voxel size = $2 \times 2 \times 2 \text{ mm}^3$, FOV = $1728 \text{ mm} \times 1728 \text{ mm}$, slice thickness = 2 mm, slice number = 78), providing whole-head coverage. The slices were acquired in the axial plane. Auditory stimuli were presented using an MRI-compatible pneumatic headset. Visual stimuli were projected onto a translucent screen via an LCD projector, and participants viewed the stimuli through a mirror mounted on the head coil. Anatomical images were obtained using a T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence (TR = 2,300 ms, TE = 2.26 ms, flip angle = 8° , voxel size = $1 \times 1 \times 1 \text{ mm}^3$, FOV = $232 \text{ mm} \times 256 \text{ mm}$, slice thickness = 1 mm).

Functional images were preprocessed and analyzed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm>) in MATLAB (version R2018a, Math-Work). Initial dummy scans for signal equilibration were excluded from the analysis. The remaining images underwent slice timing correction and were realigned for motion correction by registration to the mean image. Subsequently, these images were coregistered with the T1-weighted 3D images and normalized to MNI space with cubic voxels at $2 \times 2 \times 2 \text{ mm}^3$ spatial resolution. Functional images were spatially smoothed with an 8 mm full

width at half maximum (FWHM) isotropic Gaussian kernel. Data from six participants were discarded from further analysis due to excessive head motion during fMRI, with exclusion criteria set at head motion >3 mm translation, or $>3^\circ$ rotation. The criteria were relatively lenient considering two production localizer tasks involved in the present study.

Statistics and reproducibility

All statistical analyses in this study were conducted using custom Matlab scripts. Brain activation patterns associated with production and perception tasks were examined by modeling individual participants' data at the first level and then comparing these individual results across the group to identify consistent brain activity patterns. A paired *t*-test was used to determine statistical significance for the contrasts between conditions. For the multivariate representational similarity analysis (RSA), Spearman's correlation was applied to assess the relationship between neural activity within sensorimotor ROIs and motoric/sensory features of language stimuli. Statistical significance was assessed through one-tailed permutation testing, comparing the observed neural-feature correlations with random permutations of the data.

Univariate analysis. To identify brain activation patterns during language production and perception tasks, a general linear model (GLM) approach was employed, with experimental regressors modeled as boxcar functions and convolved with a canonical hemodynamic response function (HRF). A high-pass filter with a cut-off of 128 s was applied to remove low-frequency drifts. Six head motion parameters (translation and rotation) were entered as nuisance regressors to account for variance caused by head movement. All brain activation maps were visualized with the BrainNet Viewer¹¹⁵ (<http://www.nitrc.org/projects/bnv/>). Effect sizes for the activated brain regions were computed using Cohen's *d* and Hedges' *g*. Cohen's *d* was calculated by dividing the mean effect size at each cluster's peak voxel in the group analysis by the standard deviation across participants. Hedges' *g*, which corrects for the small sample sizes in fMRI group analyses¹¹⁶, was computed using SPM 12 and the Measures of Effect Size (MES) toolbox¹¹⁷ (<https://github.com/hhentschke/measures-of-effect-size-toolbox>).

For the localizer tasks, block-based GLMs were conducted to define critical speech- and writing-motor areas, independent of the main perception tasks. At the first level, GLMs included the following regressors: (1) syllable articulation, lip movement (control), and rest/fixation period for the AR task; and (2) character writing, drawing (control), and rest/fixation period for the WR task. Speech-motor areas were identified by contrasting syllable articulation with lip movement, while writing-motor areas were identified by contrasting character writing with drawing. Participant-specific contrast images were then used to create group-level activation maps, using a voxel-level threshold of $p < 0.05$ (FDR corrected for multiple comparisons) with a cluster extent threshold of $k \geq 10$ consecutive voxels.

For the perception tasks, block-based GLMs were employed to explore key regions involved in auditory and visual language perception. At the first level, GLMs included three regressors: (1) spoken syllable perception, auditory control stimuli (reversed speech) perception, and rest/fixation for the SP task; and (2) visual character perception, visual control stimuli (scrambled character) perception, and rest/fixation for the CH task. For auditory speech perception, activation maps were obtained by contrasting spoken syllables with auditory control and rest (fixation). For visual character perception, activation maps were derived by contrasting visual character with visual control and rest (fixation). Group-level activation maps were generated from participant-specific contrasts using second-level GLM analysis. For the contrasts of "speech/character $>$ control", we used a voxel-level threshold of $p < 0.01$ uncorrected (cluster extent ≥ 10 voxels) to capture subtle neural activations in language-motor regions, which are known to exhibit weaker responses during language perception tasks^{11,12,27,38}. For the "speech/character $>$ rest (fixation)" contrasts, we applied a stricter voxel-level threshold of $p < 0.05$ FDR corrected (cluster extent ≥ 10 voxels) to

identify activations. These univariate results guided region of interest (ROI) selection for the subsequent RSA, which focused on multivoxel representational patterns rather than activation strength and was independently validated using permutation-based significance testing, ensuring the independence of results from univariate statistical thresholds.

Representational similarity analysis (RSA). The RSA was performed to investigate whether neural patterns in the language-motor cortex during speech/character perception could be predicted by the motoric or sensory features of language stimuli. Functional images were smoothed using a 4 mm FWHM kernel to enhance sensitivity in multivariate analyses^{78,118}. Separate GLMs were constructed for the two perception tasks. The pre-processed data were analyzed using an event-related design with delta function regressors. For the SP task, regressors included eight CV syllables and one auditory control condition. For the CH task, regressors included twenty-four Chinese characters and one visual control condition. High-pass filtering at 128 s and motion covariates were applied in all analyses. *T*-contrasts were generated for each stimulus type relative to control, and the resulting *t*-maps were used for RSA to enhance the detection of task-dependent effects⁵⁵.

RSA was implemented using the CoSMoMvPA toolbox¹¹⁹ (<https://www.cosmomvpa.org/>) and custom MATLAB code. ROI-based analyses focused on speech-motor regions for the SP task and writing-motor regions for the CH task. Additionally, representational activity in language-sensory areas was examined for each modality to compare motoric and sensory feature representations across language-motor and sensory regions. To address individual variability in motor and sensory region locations, we adopted a dual-step ROI-voxel selection procedure for RSA (following Patel et al., 2023)⁷⁸. First, anatomical ROIs were defined based on the guidance of group-level univariate results, and were created with the AAL templates on each participant's brain using the WFU Pickatlas Toolbox^{120,121}, covering key sensorimotor areas involved in auditory and visual language production and perception tasks. Specifically, motor ROIs were defined from the AR and WR tasks (articulation/writing vs. control), while auditory and visual ROIs were defined from group-level univariate contrasts in the SP and CH tasks (speech/character vs. control). Next, for each participant, the top 100 most active voxels within each anatomical region were selected based on their *t*-values in the "speech/character vs. control" contrast of the perception tasks. This approach ensured that RSA was focused on the most informative voxels that may engage in processing motoric or sensory features of language stimuli while accounting for individual differences in activation strength and mitigating potential voxel count biases across different ROIs. Neural representational dissimilarity matrices (RDMs) were constructed by calculating 1 minus the Pearson correlation values between activation patterns for all stimulus pairs. For the SP task, neural RDM predictors included the speech-motoric RDM, the high-level speech acoustic feature RDM, and the low-level acoustic spectrogram RDM (Fig. 2a). For the CH task, predictors included the stroke-motoric RDM, the high-level visual feature RDM, and the low-level visual pixel RDM (Fig. 2b) (see also Tables S4-S9 for maxima lists of the RDM values).

To test the relatedness between neural activity and feature RDMs, we conducted a two-stage permutation test to determine the significance of the neural-motoric or neural-sensory correlations in ROIs^{78,79}. The procedure involved the following steps: (1) At the individual level, the observed correlation (r_0) between neural RDM (top 100 voxels per ROI) and a specific motoric/sensory RDM was calculated using Spearman correlation and subsequently transformed into a Fisher *z*-score. To construct a null distribution, motoric/sensory RDM values were shuffled 100 times per ROI, and the correlation with the neural RDM was recalculated for each shuffle, yielding 100 randomized correlations (r_1) per ROI for each participant. (2) At the group level, the observed correlations were averaged across participants to obtain a group-mean correlation (mean r_0). A null distribution of group-level correlations was created using a Monte Carlo approach. Specifically, one r_1 value was randomly sampled from each participant's set of 100 shuffled correlations, and the group mean was calculated across

participants (mean r_1). This process was repeated 10,000 times to generate a set of group-mean null correlations. (3) The significance (p -value) was determined by comparing the observed group mean correlation (mean r_0) to the null distribution (mean r_1) and calculating its percentile rank. For example, if the observed correlation was ranked among the top 100 values out of 10,000 permutations, the p -value would be approximated as $100/(10,000 + 1) \approx 0.01$. This approach provided a robust evaluation of whether the observed neural-motoric or neural-sensory correlations exceeded those expected by chance. P -values were then corrected for multiple comparisons using a FDR procedure.

Sample size justification. The formal analysis included 30 participants (another 6 were excluded due to excessive head movement). This sample size aligns with standard practices in fMRI research, which typically involve 20 to 40 participants, and is generally sufficient to detect medium effect sizes in task-based analyses¹²². To further justify our sample size, we computed effect sizes for each peak significant voxel in the group-level GLMs utilizing *Cohen's d* and *Hedges' g*¹¹⁶. Additionally, a post-hoc power analysis was performed using *G*Power* software¹²³ (t -tests - Means: Difference between two dependent means, matched pairs). The results showed that the observed effect sizes for all tasks are moderate to large (*Cohen's d* > 0.48, *Hedges' g* > 0.47), with power exceeding 70%. Detailed results were provided in Tables S10–S13. These results indicate that our findings were reliable with this sample size.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The processed source data are publicly available through the Open Science Framework (OSF) repository (<https://osf.io/x6hps/>)¹²⁴. Due to ethical constraints, raw data are available for research purposes only upon request from the corresponding author. Requests for access must include a detailed project outline and may be subject to a formal data-sharing agreement to comply with data protection regulations.

Code availability

Data analysis was carried out in MATLAB (version 2018a). The custom scripts for the data analysis have been uploaded on the OSF repository (<https://osf.io/x6hps/>)¹²⁴ and are publicly available.

Received: 3 August 2024; Accepted: 3 January 2025;

Published online: 11 January 2025

References

- Choi, D., Dehaene-Lambertz, G., Peña, M. & Werker, J. F. Neural indicators of articulator-specific sensorimotor influences on infant speech perception. *Proc. Natl. Acad. Sci.* **118**, e2025043118 (2021).
- Gennari, G., Marti, S., Palu, M., Fló, A. & Dehaene-Lambertz, G. Orthogonal neural codes for speech in the infant brain. *Proc. Natl. Acad. Sci.* **118**, e2020410118 (2021).
- Chater, N., McCauley, S. M. & Christiansen, M. H. Language as skill: Intertwining comprehension and production. *J. Mem. Lang.* **89**, 244–254 (2016).
- Schmitz, J., Diaz, B., Fernandez Rubio, K. & Sebastian-Galles, N. Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities. *Lang. Cogn. Neurosci.* **33**, 527–546 (2018).
- Hickok, G., Houde, J. & Rong, F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* **69**, 407–422 (2011).
- Hickok, G. The dual stream model of speech and language processing. *Handb. Clin. Neurol.* **185**, 57–69 (2022).
- Skipper, J. I., Devlin, J. T. & Lametti, D. R. The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang.* **164**, 77–105 (2017).
- Liebenthal, E. & Möttönen, R. An interactive model of auditory-motor speech perception. *Brain Lang.* **187**, 33–40 (2018).
- Liberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21**, 1–36 (1985).
- Liberman, A. M. & Whalen, D. H. On the relation of speech to language. *Trends Cogn. Sci.* **4**, 187–196 (2000).
- Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* **7**, 701–702 (2004).
- Pulvermüller, F. et al. Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci.* **103**, 7865–7870 (2006).
- D'Ausilio, A. et al. The motor somatotopy of speech perception. *Curr. Biol.* **19**, 381–385 (2009).
- Cogan, G. B. et al. Sensory-motor transformations for speech occur bilaterally. *Nature* **507**, 94–98 (2014).
- Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y. & Poeppel, D. Speaking rhythmically can shape hearing. *Nat. Hum. Behav.* **5**, 71–82 (2021).
- Liang, B., Li, Y., Zhao, W. & Du, Y. Bilateral human laryngeal motor cortex in perceptual decision of lexical tone and voicing of consonant. *Nat. Commun.* **14**, 4710 (2023).
- Prather, J. F., Peters, S., Nowicki, S. & Mooney, R. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* **451**, 305–310 (2008).
- Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S. & Mooney, R. Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* **12**, 221–228 (2009).
- Archakov, D. et al. Auditory representation of learned sound sequences in motor regions of the macaque brain. *Proc. Natl. Acad. Sci.* **117**, 15242–15252 (2020).
- Petkov, C. I. & Jarvis, E. D. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12 (2012).
- Schomers, M. R. & Pulvermüller, F. Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Front. Hum. Neurosci.* **10**, 435 (2016).
- Lotto, A. J., Hickok, G. S. & Holt, L. L. Reflections on mirror neurons and speech perception. *Trends Cogn. Sci.* **13**, 110–114 (2009).
- Hickok, G., Costanzo, M., Capasso, R. & Miceli, G. The role of Broca's area in speech perception: Evidence from aphasia revisited. *Brain Lang.* **119**, 214–220 (2011).
- Hickok, G. The myth of mirror neurons: The real neuroscience of communication and cognition WW Norton & Company (2014).
- Lindblom, B. Role of articulation in speech perception: Clues from production. *J. Acoust. Soc. Am.* **99**, 1683–1692 (1996).
- Cheung, C., Hamilton, L. S., Johnson, K. & Chang, E. F. The auditory representation of speech sounds in human motor cortex. *Elife* **5**, e12577 (2016).
- Correia, J. M., Jansma, B. M. & Bonte, M. Decoding articulatory features from fMRI responses in dorsal speech regions. *J. Neurosci.* **35**, 15015–15025 (2015).
- Zhang, W., Liu, Y., Wang, X. & Tian, X. The dynamic and task-dependent representational transformation between the motor and sensory systems during speech production. *Cogn. Neurosci.* **11**, 194–204 (2020).
- Vinci-Booher, S., James, T. W. & James, K. H. Visual-motor functional connectivity in preschool children emerges after handwriting experience. *Trends Neurosci. Educ.* **5**, 107–120 (2016).
- James, K. H. & Engelhardt, L. The effects of handwriting experience on functional brain development in pre-literate children. *Trends Neurosci. Educ.* **1**, 32–42 (2012).

31. Cao, F. et al. Writing affects the brain network of reading in Chinese: A functional magnetic resonance imaging study. *Hum. Brain Mapp.* **34**, 1670–1684 (2013).
32. Li, J. X. & James, K. H. Handwriting generates variable visual output to facilitate symbol learning. *J. Exp. Psychol.-Gen.* **145**, 298–313 (2016).
33. Seyll, L., Wyckmans, F. & Content, A. The impact of graphic motor programs and detailed visual analysis on letter-like shape recognition. *Cognition* **205**, 104443 (2020).
34. Longcamp, M., Zerbato-Poudou, M. T. & Velay, J. L. The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychol* **119**, 67–79 (2005).
35. Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A. & Siok, W. T. Reading depends on writing. *in Chinese. Proc. Natl. Acad. Sci.* **102**, 8781–8785 (2005).
36. Longcamp, M., Boucard, C., Gilhodes, J. C. & Velay, J. L. Remembering the orientation of newly learned characters depends on the associated writing knowledge: A comparison between handwriting and typing. *Hum. Mov. Sci.* **25**, 646–656 (2006).
37. Longcamp, M. et al. Learning through hand-or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *J. Cogn. Neurosci.* **20**, 802–815 (2008).
38. James, K. H. & Gauthier, I. Letter processing automatically recruits a sensory–motor brain network. *Neuropsychologia* **44**, 2937–2949 (2006).
39. Pulvermüller, F. & Fadiga, L. Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* **11**, 351–360 (2010).
40. Perfetti, C. A. & Tan, L. H. Write to read: the brain’s universal reading and writing network. *Trends Cogn. Sci.* **17**, 56–57 (2013).
41. James, K. H. The importance of handwriting experience on the development of the literate brain. *Curr. Dir. Psychol.* **26**, 502–508 (2017).
42. Longcamp, M., Anton, J. L., Roth, M. & Velay, J. L. Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage* **19**, 1492–1500 (2003).
43. Nakamura, K. et al. Universal brain systems for recognizing word shapes and handwriting gestures during reading. *Proc. Natl. Acad. Sci.* **109**, 20762–20767 (2012).
44. Cao, F. & Perfetti, C. A. Neural signatures of the reading–writing connection: greater involvement of writing in Chinese reading than English reading. *PLoS One* **11**, e0168414 (2016).
45. Roux, F. E. et al. The graphemic/motor frontal area Exner’s area revisited. *Ann. Neurol.* **66**, 537–545 (2009).
46. Anderson, S. W., Damasio, A. R. & Damasio, H. Troubled letters but not numbers: domain specific cognitive impairments following focal damage in frontal cortex. *Brain* **113**, 749–766 (1990).
47. Purcell, J., Turkeltaub, P. E., Eden, G. F. & Rapp, B. Examining the central and peripheral processes of written word production through meta-analysis. *Front. Psychol.* **2**, 11255 (2011).
48. Longcamp, M. et al. Functional specificity in the motor system: Evidence from coupled fMRI and kinematic recordings during letter and digit writing. *Hum. Brain Mapp.* **35**, 6077–6087 (2014).
49. Zhang, Z. et al. The cortical organization of writing sequence: evidence from observing Chinese characters in motion. *Brain Struct. Func.* **226**, 1627–1639 (2021).
50. Yu, H., Gong, L., Qiu, Y. & Zhou, X. Seeing Chinese characters in action: An fMRI study of the perception of writing sequences. *Brain Lang* **119**, 60–67 (2011).
51. Mizuochi-Endo, T. et al. Graphomotor memory in Exner’s area enhances word learning in the blind. *Commun. Biol.* **4**, 443 (2021).
52. Zhai, M. & Fischer-Baum, S. Exploring the effects of knowledge of writing on reading Chinese characters in skilled readers. *J. Exp. Psychol.-Learn. Mem. Cogn.* **45**, 724–731 (2019).
53. Vinci-Booher, S. & James, K. H. Visual experiences during letter production contribute to the development of the neural systems supporting letter perception. *Dev. Sci.* **23**, e12965 (2020).
54. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 249 (2008).
55. Dimsdale-Zucker, H. R. & Ranganath, C. Representational similarity analyses: a practical guide for functional MRI applications. *Handb. Behav. Neurosci.* **28**, 509–525 (2018).
56. Rothlein, D. & Rapp, B. The similarity structure of distributed neural responses reveals the multiple representations of letters. *Neuroimage* **89**, 331–344 (2014).
57. Meyer, K. & Damasio, A. Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci* **32**, 376–382 (2009).
58. Kiefer, M. & Pulvermüller, F. Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex* **48**, 805–825 (2012).
59. Mahon, B. Z. & Hickok, G. Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychon. Bull. Rev.* **23**, 941–958 (2016).
60. Matheson, H. E. & Barsalou, L. W. Embodiment and grounding in cognitive neuroscience. *Stevens’ Handb. Exp. Psychol. Cogn. Neurosci.* **3**, 1–27 (2018).
61. Cohen, L., Jobert, A., Le Bihan, D. & Dehaene, S. Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage* **23**, 1256–1270 (2004).
62. Constable, R. T. et al. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *Neuroimage* **22**, 11–21 (2004).
63. Buchweitz, A., Mason, R. A., Tomitch, L. & Just, M. A. Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychol. Neurosci.* **2**, 111–123 (2009).
64. Labruna, L., Fernández-del-Olmo, M., Landau, A., Duqué, J. & Ivry, R. B. Modulation of the motor system during visual and auditory language processing. *Exp. Brain Res.* **211**, 243–250 (2011).
65. Deniz, F., Nunez-Elizalde, A. O., Huth, A. G. & Gallant, J. L. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *J. Neurosci.* **39**, 7722–7736 (2019).
66. Fedorenko, E., Ivanova, A. A. & Regev, T. I. The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* **25**, 289–312 (2024).
67. Petersen, S. E. & Dubis, J. W. The mixed block/event-related design. *Neuroimage* **62**, 1177–1184 (2012).
68. Visscher, K. M. et al. Mixed blocked/event-related designs separate transient and sustained activity in fMRI. *Neuroimage* **19**, 1694–1708 (2003).
69. Du, Y., Buchsbaum, B. R., Grady, C. L. & Alain, C. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci.* **111**, 7126–7131 (2014).
70. Park, H., Ince, R. A., Schyns, P. G., Thut, G. & Gross, J. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **25**, 1649–1653 (2015).
71. Park, H., Ince, R. A., Schyns, P. G., Thut, G. & Gross, J. Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. *PLoS Biol* **16**, e2006558 (2018).
72. Abbasi, O., Steingraber, N., Chalas, N., Kluger, D. S. & Gross, J. Spatiotemporal dynamics characterise spectral connectivity profiles of continuous speaking and listening. *PLoS Biol* **21**, e3002178 (2023).

73. Du, Y., Buchsbaum, B. R., Grady, C. L. & Alain, C. Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat. Commun.* **7**, 12241 (2016).
74. Planton, S., Jucla, M., Roux, F. E. & Démonet, J. F. The “handwriting brain”: a meta-analysis of neuroimaging studies of motor versus orthographic processes. *Cortex* **49**, 2772–2787 (2013).
75. Lagarrigue, A. et al. Activation of writing-specific brain regions when reading Chinese as a second language. Effects of training modality and transfer to novel characters. *Neuropsychologia* **97**, 83–97 (2017).
76. Planton, S., Longcamp, M., Péran, P., Démonet, J. F. & Jucla, M. How specialized are writing-specific brain regions? An fMRI study of writing, drawing and oral spelling. *Cortex* **88**, 66–80 (2017).
77. Vinci-Booher, S., Cheng, H. & James, K. H. An analysis of the brain systems involved with producing letters by hand. *J. Cogn. Neurosci.* **31**, 138–154 (2019).
78. Patel, T., Morales, M., Pickering, M. J. & Hoffman, P. A common neural code for meaning in discourse production and comprehension. *Neuroimage* **279**, 120295 (2023).
79. Stelzer, J., Chen, Y. & Turner, R. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage* **65**, 69–82 (2013).
80. Kriegeskorte, N. et al. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008).
81. Tan, L. H., Laird, A. R., Li, K. & Fox, P. T. Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Hum. Brain Mapp.* **25**, 83–91 (2005).
82. Hesselmann, V. et al. Discriminating the cortical representation sites of tongue and lip movement by functional MRI. *Brain Topogr.* **16**, 159–167 (2004).
83. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).
84. Grabski, K. et al. Functional MRI assessment of orofacial articulators: neural correlates of lip, jaw, larynx, and tongue movements. *Hum. Brain Mapp.* **33**, 2306–2321 (2012).
85. Galantucci, B., Fowler, C. A. & Turvey, M. T. The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* **13**, 361–377 (2006).
86. Palmis, S. et al. The impact of spelling regularity on handwriting production: a coupled fMRI and kinematics study. *Cortex* **113**, 111–127 (2019).
87. Yin, Y. & Zhang, Q. Chinese characters are read using not only visual but also writing motor information. *Psychophysiology* **58**, e13696 (2021).
88. Mesgarani, N., Cheung, C., Johnson, K. & Chang, E. F. Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–1010 (2014).
89. Giordano, B. L., Esposito, M., Valente, G. & Formisano, E. Intermediate acoustic-to-semantic representations link behavioral and neural responses to natural sounds. *Nat. Neurosci.* **26**, 664–672 (2023).
90. Kao, C. H., Chen, D. Y. & Chen, C. C. The inversion effect in visual word form processing. *Cortex* **46**, 217–230 (2010).
91. Cao, F., Brennan, C. & Booth, J. R. The brain adapts to orthography with experience: evidence from English and Chinese. *Dev. Sci.* **18**, 785–798 (2015).
92. Sussman, B. L., Reddigari, S. & Newman, S. D. The impact of inverted text on visual word processing: An fMRI study. *Brain Cogn.* **123**, 1–9 (2018).
93. Londei, A. et al. Sensory-motor brain network connectivity for speech comprehension. *Hum. Brain Mapp.* **31**, 567–580 (2010).
94. Carey, D., Miquel, M. E., Evans, B. G., Adank, P. & McGettigan, C. Vocal tract images reveal neural representations of sensorimotor transformation during speech imitation. *Cereb. Cortex.* **27**, 3064–3079 (2017).
95. Merabet, L. & Pascual-Leone, A. Neural reorganization following sensory loss: the opportunity of change. *Nat. Rev. Neurosci.* **11**, 44–52 (2010).
96. Siuda-Krzywicka, K. et al. Massive cortical reorganization in sighted Braille readers. *Elife* **5**, e10762 (2016).
97. Longcamp, M., Hlushchuk, Y. & Hari, R. What differs in visual recognition of handwritten vs printed letters? An fMRI study. *Hum. Brain Mapp.* **32**, 1250–1259 (2011).
98. Kittredge, A. K. & Dell, G. S. Learning to speak by listening: Transfer of phonotactics from perception to production. *J. Mem. Lang.* **89**, 8–22 (2016).
99. Pulvermüller, F. Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *J. Neurolinguist.* **25**, 423–459 (2012).
100. Sammler, D., Grosbras, M. H., Anwender, A., Bestelmeyer, P. E. & Belin, P. Dorsal and ventral pathways for prosody. *Curr. Biol.* **25**, 3079–3085 (2015).
101. Riccardi, N., Yourganov, G., Rorden, C., Fridriksson, J. & Desai, R. H. Dissociating action and abstract verb comprehension post-stroke. *Cortex* **120**, 131–146 (2019).
102. Matchin, W. & Hickok, G. The cortical organization of syntax. *Cereb. Cortex.* **30**, 1481–1498 (2020).
103. Brainard, D. H. & Vision, S. The psychophysics toolbox. *Spatial Vis.* **10**, 433–436 (1997).
104. Katanoda, K., Yoshikawa, K. & Sugishita, M. A functional MRI study on the neural substrates for writing. *Hum. Brain Mapp.* **13**, 34–42 (2001).
105. Sugihara, G., Kaminaga, T. & Sugishita, M. Interindividual uniformity and variety of the “Writing center”: a functional MRI study. *Neuroimage* **32**, 1837–1849 (2006).
106. Stevens, K. N. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* **111**, 1872–1891 (2002).
107. Rapp, B. & Caramazza, A. From graphemes to abstract letter shapes: levels of representation in written spelling. *J. Exp. Psychol.-Hum. Percept. Perform.* **23**, 1130 (1997).
108. Palmis, S., Danna, J., Velay, J. L. & Longcamp, M. Motor control of handwriting in the developing brain: A review. *Cogn. Neuropsychol.* **34**, 187–204 (2017).
109. Fiset, D. et al. Features for identification of uppercase and lowercase letter. *Psychol. Sci.* **19**, 1161–1168 (2008).
110. Wiley, R. W., Wilson, C. & Rapp, B. The effects of alphabet and expertise on letter perception. *J. Exp. Psychol.-Hum. Percept. Perform.* **42**, 1186 (2016).
111. Chang, L. Y., Chen, Y. C. & Perfetti, C. A. GraphCom: a multidimensional measure of graphic complexity applied to 131 written languages. *Behav. Res. Methods.* **50**, 427–449 (2018).
112. Fischer-Baum, S., Bruggemann, D., Gallego, I. F., Li, D. S. & Tamez, E. R. Decoding levels of representation in reading: A representational similarity approach. *Cortex* **90**, 88–102 (2017).
113. Liu, X. et al. The representations of Chinese characters: evidence from sublexical components. *J. Neurosci.* **42**, 135–144 (2022).
114. Saygin, Z. M. et al. Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* **19**, 1250–1255 (2016).
115. Xia, M., Wang, J. & He, Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One* **8**, e68910 (2013).
116. Gerchen, M. F., Kirsch, P. & Feld, G. B. Brain-wide inferiority and equivalence tests in fMRI group analyses: Selected applications. *Hum. Brain Mapp.* **42**, 5803–5813 (2021).

117. Hentschke, H. & Stüttgen, M. C. Computation of measures of effect size for neuroscience data sets. *Eur. J. Neurosci.* **34**, 1887–1894 (2011).
118. Hendriks, M. H., Daniels, N., Pegado, F. & Op de Beeck, H. P. The effect of spatial smoothing on representational similarity in a simple motor paradigm. *Front. Neurol.* **8**, 222 (2017).
119. Oosterhof, N. N., Connolly, A. C. & Haxby, J. V. CoSMoMPPA: multimodal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front. Neuroinformatics.* **10**, 27 (2016).
120. Maldjian, J. A., Laurienti, P. J., Kraft, R. A. & Burdette, J. H. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* **19**, 1233–1239 (2003).
121. Maldjian, J. A., Laurienti, P. J. & Burdette, J. H. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *Neuroimage* **21**, 450–455 (2004).
122. Geuter, S., Qi, G., Welsh, R. C., Wager, T. D. & Lindquist, M. A. Effect size and power in fMRI group analysis. *Biorxiv.* 295048 (2018).
123. Faul, F., Erdfelder, E., Buchner, A. & Lang, A. G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods.* **41**, 1149–1160 (2009).
124. Zheng Y. Y., Zhang J. F., Yang, Y. & Xu, M. Neural representation of sensorimotor features in language-motor areas during auditory and visual perception. <https://doi.org/10.17605/OSF.IO/X6HPS> (2024).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32171054, to M.X.; 32400865 to Y.Z.; 32271120 to Y.Y.), Shenzhen Innovation in Science and Technology Foundation for The Excellent Youth Scholars (RCYX20210706092043066, to M.X.), Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (2023SHIBS0003, to M.X.). We also would like to acknowledge the Magnetic Resonance Imaging Center at Shenzhen University.

Author contributions

Conceptualization, Y.Z., J.Z., Y.Y., and M.X.; Methodology, Y.Z., J.Z., Y.Y., and M.X.; Investigation, Y.Z.; Formal Analysis, Y.Z.; visualization, Y.Z.; Writing – Original Draft, Y.Z., and M.X.; Writing – Review & Editing, Y.Z., J.Z., Y.Y., and M.X.; Supervision, M.X.; Funding Acquisition, Y.Z., Y.Y., and M.X.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07466-5>.

Correspondence and requests for materials should be addressed to Min Xu.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Jasmine Pan. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025