a speci c network in which the parietal and perhaps lateral frontal cortices appear to be optimally situated to mediate the integration and attentional selection of motion information across modalities[12]. In audiovisual face perception, crossmodal attention in uences crossmodal binding during speech reading[13,14]. Thus, attention and audiovisual integration interact with each other in a sophisticated manner. However, feature-selective attention in audiovisual conditions and the relationship between feature-selective attention and high-level audiovisual semantic integration remain to be explored.

In a single (visual or auditory) modality, feature-selective attention may lead to selective processing of the attended features of an object in the brain[7–9,15–17]. Nobre *et al.*[8] demonstrated that ERPs are modulated by feature-selective attentios a4561 0 9 155.9.5(h)3.61al 22 Tm 2(t)-

1. ( ) Four examples of audiovisual stimuli; the red numbers indicate runs with the number task only. ( ) Time course of a trial for the runs with the number task, in which the stimuli included randomly presented numbers and videos/audios/movie clips. ( ) Time course of a trial for the runs with the gender, emotion, or bi-feature task. For both (,), the presentation of a stimulus (video/audio/movie clip) lasted 1,400 ms and was repeated four times during the rst eight seconds in a trial. A visual cue ("+") appeared at the 8th second and persisted for six seconds.

For each of the three runs with the number task, in addition to the corresponding audiovisual, visual-only, or auditory-only facial stimuli from the movie clips, numbers in red appeared sequentially at the center of the screen (see Fig. 1A). e subject's task was to attend to the numbers instead of the other stimuli (see Table 1). We designed a di cult number task for the subjects in which they were asked to nd and count the repeated numbers to ensure that they fully ignored the features of the visual-only, auditory-only, or audiovisual facial stimuli. erefore, the subjects performed this task with low accuracy, as shown in Fig. S3. At the beginning of each block, there were four seconds before the rst trial, and a short instruction in Chinese (see Table 1) was displayed on the screen in the rst two seconds (the last two seconds were used to display numbers, as indicated below). At the beginning of each trial, a visual-only, auditory-only or audiovisual facial stimulus was presented to the subject for 1,400 ms, followed by a 600-ms blank period. is two-second cycle with the same stimulus was repeated four times, followed by a six-second blank period. erefore, one trial lasted 14 seconds. In addition to the above stimuli, eight numbers in red appeared one by one at the center of the screen, each a random integer from 0 to 9. Each number lasted 900 ms, and the interval between two subsequent numbers was 350 ms. e rst number appeared 2 seconds before the beginning of this trial. e subjects were asked to nd and count the repeated numbers. A er the stimulation, a xation cross appeared on the screen. e subjects then responded by pressing the right-hand keys according to the instruction for this block (see Table 1). e xation cross changed color at the 12th second, indicating that the next trial would begin shortly (see Fig. 1B). In total, a run lasted 1,350 seconds.

e procedure for the three runs with the gender/emotion task was similar to that for the runs with the number task, except that no numbers appeared on the screen and the subjects performed a gender/emotion judgment tas (See Table 1). Speci cally, the subjects were asked to focus their attention on either the gender or the emotion o the presented stimuli (visual-only, auditory-only, or audiovisual facial stimuli; see Fig. 1A without regard to the numbers) and make a corresponding judgment (male vs. female for the gender task or crying vs. laughing for the emotion task) to each stimulus. At the beginning of each block, a short instruction (see Table 1) was displayed for four seconds on the screen. e time course of each trial was similar to that in the runs with number task (see Fig. 1C). In each trial, the subject was asked to judge the gender/emotion category of the stimulus and press th right-hand keys according to the instruction for this block.

For the three runs with the bi-feature task, the subjects were asked to simultaneously attend to both gender and emotion features (see Table 1). e experimental procedure for each run was similar to that for the runs with the gender/emotion task with the following3(lo)-c1i0364638 21e13.1((a)3.3(s).5(w)-2.6(91(o)12(r to)-c1i03Bh)3.5

voxels, time series detrending, and normalization of the time series in each block to zero mean and unit variance. All preprocessing steps were performed using SPM8[23] and custom functions in MATLAB 7.4 (MathWorks, Natick, Massachusetts, USA).

Univariate GLM analysis. is experiment included four experimental tasks (number, gender, emotion, and bi-feature). For each experimental task, three runs corresponding to the visual-only, the auditory-only, and the audiovisual stimulus conditions were performed. To con rm that audiovisual sensory integration occurred for each experimental task and determine the heteromodal areas associated with audiovisual integration, we performed voxel-wise group analysis of the fMRI data based on a mixed-e ect two-level GLM in SPM8. In particular, using the data from the three number runs, we performed GLM analysis to explore the audiovisual integration at the sensory level when the subjects fully ignored the visual-only, auditory-only, or audiovisual facial stimuli while only attending to the numbers. e GLM analysis included the following data processing. e fMRI data for each subject were subjected to a rst-level GLM, and the estimated beta coe cients across all subjects were then combined and analyzed using a second-level GLM. e following statistical criterion was used to determine brain areas for audiovisual sensory integration: $[AV > max(A,V)$ ($p < 0.05$, FWE-corrected)$] \cap [V > 0$ or $A > 0$ ($p < 0.05$, uncorrected)$]$[4,6,24–27], where $\cap$ denotes the intersection of two sets. For each subject, each task, and each stimulus condition, we also computed the percent signal changes of the pSTS/MTG clusters via region-of-interest (ROI)-based analysis (implemented by the MATLAB toolbox MarsBaR-0.4[28]). Speci cally, we identi ed the clusters consisting of signi cantly activated voxels in the bilateral pSTS/MTG via group GLM where $_{i j}$ is the angle between two pattern vec$P$ors di erentiated for di erent experimental tasks or di erent semantic features. us, audiovisual sensory integration rather than audiovisual semantic integration occurred in the identi ed heteromodal areas of the pSTS/MTG, consistent with previous results[10]

3. ( ) Le /Right: gender/emotion categories; the rst 3 rows: audiovisual, visual-only, and auditory-only stimulus conditions, respectively; the 4th row: the reproducibility ratio in the audiovisual condition minus the maximum of the reproducibility ratios in the visual-only and auditory-only conditions.

$p < 10^{-17}$, $F(3, 8) = 68.26$) (Fig. 3A–C,E–G). ere was also a signi cant interaction e ect between the two factors of stimulus condition and experimental task (gender categories: $p < 10^{-17}$, $F(6, 8) = 30.07$; emotion categories: $p < 10^{-8}$, $F(6, 8) = 10.05$). Post hoc Bonferroni-corrected paired t-tests on the stimulus conditions revealed the following: (i) for each task-relevant feature (gender categories with the gender or the bi-feature task, le panel of Fig. 3; emotion categories with the emotion or the bi-feature task, right panel of Fig. 3), the reproducibility ratios were signi cantly higher for the audiovisual stimulus condition than for the visual- or auditory-only stimulus condition (all $p < 0.001$ corrected); and (ii) for each task-irrelevant feature (gender categories with the number or the emotion task, le panel of Fig. 3; emotion categories with the number or the gender task, right panel of Fig. 3), there were no signi cant di erences between the audiovisual and the visual-only or auditory-only stimulus condition (all $p > 0.05$). Furthermore, post hoc Bonferroni-corrected paired t-tests on the experimental tasks revealed that (i) in each of the audiovisual, visual-only and auditory-only stimulus conditions, the reproducibility ratios for gender/emotion categories were signi cantly higher for each relevant task (gender categories: the gender or the bi-feature task, le panel of Fig. 3; emotion categories: the emotion or the bi-feature task, right panel of Fig. 3) than for each irrelevant task (gender categories: the number or the emotion task, le panel of Fig. 3; emotion categories: the number or the gender task, right panel of Fig. 3) (all $p < 0.05$, corrected) and that (ii) in each of the audiovisual, visual-only and auditory-only stimulus conditions, there were no signi cant di erences in the reproducibility

r ph 48 160.06626 4 Lgdala

($p < 10^{-9}$, $F(2, 8) = 36.97$ for gender categories; $p < 10^{-11}$, $F(2, 8) = 46.13$ for emotion categories). Furthermore, post hoc Bonferroni-corrected paired t-tests demonstrated that the cross-reproducibility ratios were significantly higher for the relevant task than for the irrelevant tasks (gender categories: $p < 0.001$ corrected, $t(8) = 16.23$ for gender task vs. number task; $p < 0.001$ corrected, $t(8) = 15.49$ for gender task vs. emotion task; emotion categories: $p < 0.001$ corrected, $t(8) = 16.05$ for emotion task vs. number task; $p < 0.001$ corrected, $t(8) = 14.36$ for emotion the group level (see Materials and Methods). As shown in Fig. 5, there were more functional connections from the heteromodal areas to the brain areas encoding the gender/emotion feature (Table 2/Table 3) for the relevant task (gender/emotion task) than for the irrelevant tasks (number and emotion/gender tasks). We thus observed that in the audiovisual condition, feature-selective attention enhanced the functional connectivity and thus regulated the information ows from the heteromodal areas to the brain areas encoding the attended feature. Furthermore,

results were still obtained. Second, only visual-only, auditory-only and audiovisual facial stimuli were considered in this study. In the future, we must simplify our experimental design, increase the number of subjects, and further consider non-facial stimuli to extend our conclusions.

References

1. Calvert, G. A. & esen, T. Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* **,** 191–205 (2004).
2. Campanella, S. & Belin, P. Integrating face and voice in person perception. *Trends Cogn. Sci.* **11,** 535–543 (2007).
3. Schweinberger, S. ., obertson, D. & aufmann, J. M. Hearing facial identities. *Q. J. Exp. Psych* **0,** 1446–1456 (2007).
4. Bushara, . O. *et al.* Neural correlates of cross-modal binding. *Nat. Neurosci.* **,** 190–195 (2003).
5. Macaluso, E., Frith, C. D. & Driver, J. Multisensory stimulation with or without saccades: fM I evidence for crossmodal e ects on sensory-speci c cortices that re ect multisensory location-congruence rather than tas -relevance. *NeuroImage* **2,** 414–425 (2005).
6. Macaluso, E., George, N., Dolan, ., Spence, C. & Driver, J. Spatial and temporal factors during processing of audiovisual speech: PET study. *NeuroImage* **21,** 725–732 (2004).
7. McClur in, J. W. & Optican, L. M. Primate striate and prestriate cortical neurons during discrimination. I. Simultaneous temporal encoding of information about color and pattern. *J. Neurophysiol.* **,** 481–495 (1996).
8. Nobre, A. C., ao, A. & Chelazzi, L. Selective attention to speci c features within objects: Behavioral and electrophysiological evidence. *J. Cognitive Neurosci.* **1 ,** 539–561 (2006).
9. Woodman, G. F. & Vogel, E. . Selective storage and maintenance of an object's features in visual wor ing memory. *Psychon. B. Rev.* **1 ,** 223–229 (2008).
10. Taylor, . I., Moss, H. E., Stamata is, E. A. & Tyler, L. . Binding crossmodal object features in perirhinal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 8239–8244 (2006).
11. Talsma, D., Sen ows i, D., Soto-Faraco, S. & Woldor , M. G. e multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* **1 ,** 400–410 (2010).
12. Lewis, J. W., Beauchamp, M. S. & DeYoe, E. A. A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* **10,** 873–888 (2000).
13. Joassin, F. *et al.* Cross-modal interactions between human faces and voices involved in person recognition. *Cortex* **,** 367–376 (2011).
14. Saito, D. N. *et al.* Cross-modal binding and activated attentional networ s during audio-visual speech integration: a functional M I study. *Cereb. Cortex* **1 ,** 1750–1760 (2005).
15. Ahveninen, J. *et al.* Tas -modulated "what" and "where" pathways in human auditory cortex. *Proc. Natl. Acad. Sci. USA* **103,** 14608–14613 (2006).
16. Maunsell, J. H. . & Hochstein, S. E ects of behavioral state on the stimulus selectivity of neurons in area V4 of the macaque mon ey. In: Channels in the visual nervous system: neurophysiology, psychophysics and models, (ed, Blum B), 447–470. London: Freund (1991).
17. Mirabella, G. *et al.* Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron* **,** 303–318 (2007).
18. Jeong, J. W. *et al.* Congruence of happy and sad emotion in music and faces modi es cortical audiovisual activation. *NeuroImage* **,** 2973–2982 (2011).
19. reifelts, B., Ethofer, T., Grodd, W., Erb, M. & Wildgruber, D. Audiovisual integration of emotional signals in voice and face: an event-related fM I study. *Neuroimage* **3 ,** 1445–1456 (2007).
20. Müller, V. I., Ciesli , E. C., Turets y, B. I. & Eic ho , S. B. Crossmodal interactions in audiovisual emotion processing. *Neuroimage* **0,** 553–561 (2011).
21. Müller, V. I. *et al.* Incongruence e ects in crossmodal emotional integration. *Neuroimage* **,** 2257–2266 (2011).
22. Li, Y. *et al.* Crossmodal Integration Enhances Neural epresentation of Tas - elevant Features in Audiovisual Face Perception. *Cereb. Cortex* **2 ,** 384–395 (2015).
23. Friston, . J. *et al.* Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2,** 189–210 (1994).
24. Calvert, G. A., Campbell, . & Brammer, M. J. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* **10,** 649–657 (2000).
25. Frassinetti, F., Bolognini, N. & La, d. E. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.* **1 ,** 332–343 (2002).
26. Macaluso, E. & Driver, J. Multisensory spatial interactions: a window onto functional integration in the human brain. *TRENDS Neurosci.* **2 ,** 264–271 (2005).
27. Beauchamp, M. S. Statistical criteria in FM I studies of multisensory integration. *Neuroinformatics* **3,** 93–113 (2005).
28. Brett, M., Anton, J.-L., Valabregue, . & Poline, J.-B. egion of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage* **1 ,** 1140–1141 (2002).
29. rieges orte, N., Goebel, . & Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* **103,** 3863–3868 (2006).
30. Nichols, T. & Hayasa a, S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods MEE*

Acknowledgements

is work was supported by the National Key Basic Research Program of China (973 Program) under Grant 2015CB351703, the National High-tech R&D Program of China (863 Program) under Grant 2012AA011601, the National Natural Science Foundation of China under Grants 91420302, 81471654 and 61403147, and Guangdon Natural Science Foundation under Grant 2014A030312005.

Author Contributions

Y.L. designed research and wrote the paper; J.L. and W.W. analyzed the data; B.H., T.Y. and P.L. performed th research; F.F. and P.S. revised the paper; all authors reviewed the manuscript.

Additional Information

accompanies this paper at http://www.nature.com/srep

e authors declare no competing nancial interests.

: Li, Y. *et al.* Selective Audiovisual Semantic Integration Enabled by Feature-Selective Attention. *Sci. Rep.* , 18914; doi: 10.1038/srep18914 (2016).