LETTER

doi:10.1038/nature12935

Sensory–motor transformations for speech occur bilaterally

Gregory B. Cogan[1], Thomas Thesen[2], Chad Carlson[2]†, Werner Doyle[3], Orrin Devinsky[2,3] & Bijan Pesaran[1]

Historically, the study of speech processing has emphasized a strong link between auditory perceptual input and motor production output[1–4]. A kind of 'parity' is essential, as both perception- and production-based representations must form a unified interface to facilitate access to higher-order language processes such as syntax and semantics, believed to be computed in the dominant, typically left hemisphere[5,6]. Although various theories have been proposed to unite perception and production[2,7], the underlying neural mechanisms are unclear. Early models of speech and language processing proposed that perceptual processing occurred in the left posterior superior temporal gyrus (Wernicke's area) and motor production processes occurred in the left inferior frontal gyrus (Broca's area)[8,9]. Sensory activity was proposed to link to production activity through connecting fibre tracts, forming the left lateralized speech sensory–motor system[10]. Although recent evidence indicates that speech perception occurs bilaterally[11–13], prevailing models maintain that the speech sensory–motor system is left lateralized[11,14–18] and facilitates the transformation from sensory-based auditory representations to motor-based production representations[11,15,16]. However, evidence for the lateralized computation of sensory–motor speech transformations is indirect and primarily comes from stroke patients that have speech repetition deficits (conduction aphasia) and studies using covert speech and haemodynamic functional imaging[16,19]. Whether the speech sensory–motor system is lateralized, like higher-order language processes, or bilateral, like speech perception, is controversial. Here we use direct neural recordings in subjects performing sensory–motor tasks involving overt speech production to show that sensory–motor transformations occur bilaterally. We demonstrate that electrodes over bilateral inferior frontal, inferior parietal, superior temporal, premotor and somatosensory cortices exhibit robust sensory–motor neural responses during both perception and production in an overt word-repetition task. Using a non-word transformation task, we show that bilateral sensory–motor responses can perform transformations between speech-perception- and speech-production-based representations. These results establish a bilateral sublexical speech sensory–motor system.

To investigate the sensory–motor representations that link speech perception and production, we used electrocorticography (ECoG), in which electrical recordings of neural activity are made directly from the cortical surface in a group of patients with pharmacologically intractable epilepsy. ECoG is an important electrophysiological signal recording modality that combines excellent temporal resolution with good spatial localization. Critically for this study, ECoG data contain limited artefacts due to muscle and movements during speech production compared with non-invasive methods that suffer artefacts with jaw movement[20]. Thus, using ECoG we were able to investigate directly neural representations for sensory–motor transformations using overt speech production.

Sixteen patients with subdural electrodes (see Supplementary Figs 1 and 2) implanted in the left hemisphere (6 subjects), right hemisphere (7 subjects) or both hemispheres (3 subjects) performed variants of an overt word repetition task designed to elicit sensory–motor activations (Fig. 1a, Methods and Supplementary Table 1). We observed increases in neural activity across the high gamma frequency range (60–200 Hz and above) with maximal activity across subjects between 70–90 Hz. High gamma activity reflects the spiking activity of populations of neurons during task performance[20,21]. Individual electrodes showed one of three types of task responses: sensory–motor (S-M), production (PROD), or auditory (AUD) (Fig. 1b, see Methods). We found that AUD activity was generally localized to the superior temporal gyrus and middle temporal gyrus (42 out of 57 electrodes (74%); Fig. 2a, b) and PROD activity occurred mostly in the motor and premotor corticies, somatosensory cortex, and the inferior parietal lobule (98 out of 124 electrodes (79%); Fig. 2a, b), consistent with previous models and results of speech-perception and -production studies[11,12,17]. Furthermore, electrical stimulation of PROD electrode locations resulted in orofacial movements consistent with a motor function (see Supplementary

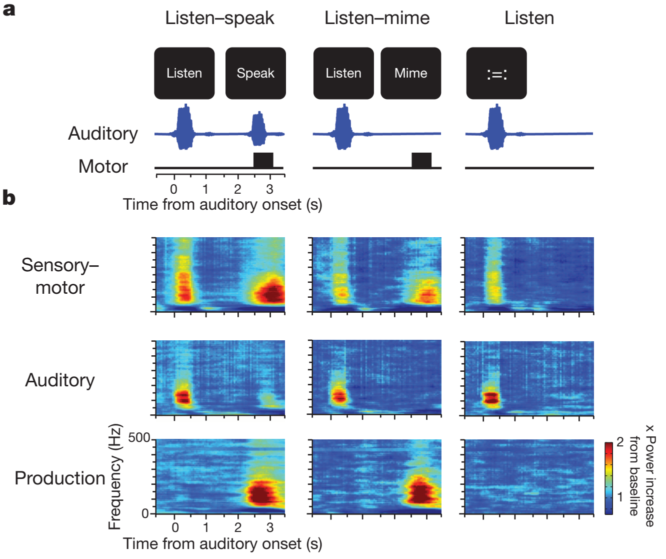

Figure 1 | Behavioural tasks and example neural activations. a, 16 Subjects were presented with an auditory consonant–vowel–consonant single-syllable word and instructed to perform one of three tasks on interleaved trials: listen–speak (listen to the word, visual prompt 'Listen', then after a 2-s delay repeat the word, visual prompt 'Speak'); listen–mime (listen to the word, visual prompt 'Listen', then after a 2-s delay, mime speaking the word, visual prompt 'Mime'); listen (passively listen to the word, visual prompt ': = :'). Auditory and motor timelines are shown. **b**, Example time–frequency spectrograms of ECoG activity normalized at each frequency to the baseline power during visual prompt. AUD, significant activity during each task epoch with auditory stimuli; PROD, significant activity during both production epochs; S-M, significant activity during the auditory and production epochs in listen–speak and listen–mime tasks.

[1]Center for Neural Science, New York University, New York, New York 10003, USA. [2]Department of Neurology, New York University School of Medicine, New York, New York 10016, USA. [3]Department of Neurosurgery, New York University School of Medicine, New York, New York 10016, USA. †Present address: Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA.

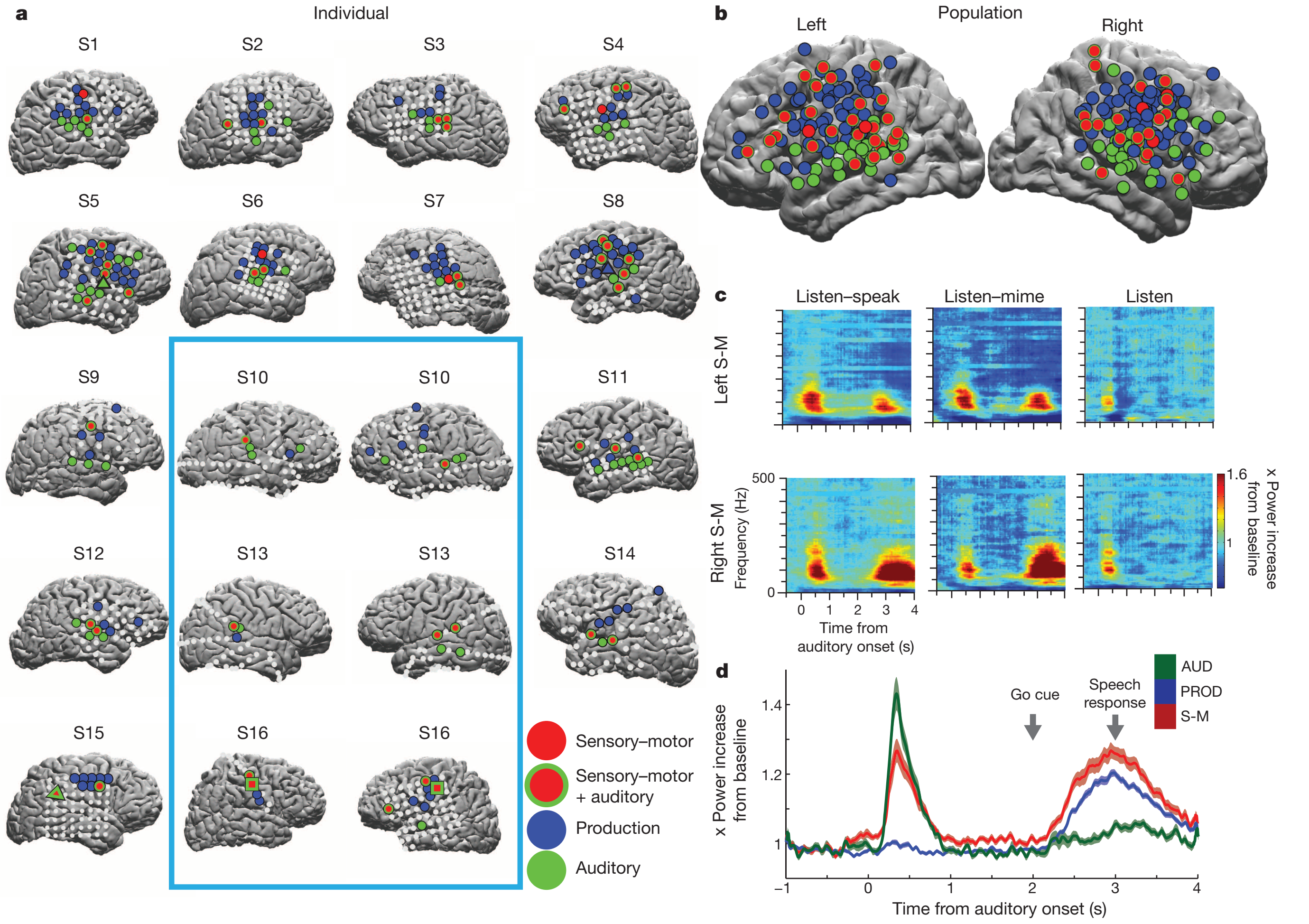

Figure 2 | Topography of neural responses and bilateral activation. **a**, Significant task-related activations within individual subject brains for left (subjects S3, S4, S7, S8, S11, S14), right (S1, S2, S5, S6, S9, S12, S15), or both (S10, S13, S16) hemispheres. Bilateral coverage is indicated by the light blue box. Electrodes with significant high gamma activity (70–90 Hz) are shown for AUD (green), PROD (blue) and S-M (red) activations. AUD and S-M activations (red with green) were often present on the same electrode. Electrodes without significant activation are shown in grey. Triangles denote example activations from Fig. 1b, and squares (S16) denote example spectrograms in Fig. 2c. **b**, Significant electrodes projected onto population average left and right hemispheres, colours as in **a**. Electrode sizes have been increased for illustrative purposes (for actual sizes see Supplementary Fig. 4). Neural spectrograms for example S-M electrodes in left and right hemispheres of S16 during listen–speak, listen–mime and listen tasks. **d**, Population average neural response profiles for each class of electrodes. Shaded regions indicate s.e.m. values across electrodes. Go cue and average production response onset are indicated by grey arrows.

Fig. 3). Critically, contrary to one of the core dogmas of brain and language, S-M activity occurred bilaterally in the supramarginal gyrus, middle temporal gyrus, superior temporal gyrus, somatosensory cortex, motor cortex, premotor cortex and inferior frontal gyrus (Fig. 2a, b, 49 electrodes; see Supplementary Table 2 and Supplementary Fig. 4) and was observed in all subjects (Fig. 2a). Of the 49 S-M sites, 45 sites showed auditory activation during the 'listen task' (Fig. 2a, b; Supplementary Figs 4 and 5; 45 out of 49 electrodes (approximately 92%)), suggesting a role in speech perception. Hemispheric dominance as determined by Wada testing did not correlate with the hemisphere of the electrode placement (χ^2 (3) = 0.92, $P = 0.34$). Importantly, in three subjects with bilateral coverage, S-M activity was present on electrodes in both hemispheres (Fig. 2a, c) and the likelihood of an electrode being a S-M site did not differ between hemispheres (Fisher's exact test, $P = 0.31$). These results demonstrate that S-M activity occurs bilaterally.

Given the evidence for bilateral S-M activity, we performed a series of analyses and experimental manipulations to test the hypothesis that bilateral S-M activity is in fact sensory–motor and represents sensory–motor transformations for speech.

One concern is that S-M activity is not due to sensory and motor processes but to sensory activation in both auditory (input) and production epochs (sound of your own voice). We observed several convergent lines of evidence that S-M activity reflects both sensory and motor processing (see Fig. 2d and Methods). First, S-M sites contain a sensory response because they responded to auditory stimulation as rapidly as AUD sites (S-M latency = 158 ms, AUD = 164 ms; see Fig. 2d). Second, S-M responses during production are not due to auditory sensory reafferent input from hearing one's own voice because responses were present during the 'listen–mime task' as well as the 'listen–speak task'. Third, S-M responses during production are not due to somatosensory reafference from moving articulators because S-M activity significantly increased within 248 ms of the production 'go' cue, whereas vocal responses occurred substantially later at 1,002 ms (±40 ms s.e.m.). Fourth, S-M production responses contain motor features because they occurred together with, and even before, PROD electrode responses (S-M = 248 ms, PROD = 302 ms, $Q = 0.03$; permutation test; see Methods). Finally, S-M activity was persistently elevated during the delay period ($P = 0.01$; see Fig. 2d, Methods), broadly consistent with planning activity, unlike PROD delay-period activity ($P = 0.64$) or AUD delay-period activity ($P = 0.53$). These results demonstrate that S-M activity cannot be simply sensory and spans both sensory and motor processes.

A related concern is that sensory–motor transformations are first carried out in the left hemisphere. If so, S-M responses in the right hemisphere could be due to communication from the left hemisphere. To test this hypothesis, we further examined latencies of S-M responses according to hemisphere. Response latencies did not differ significantly in each hemisphere in either the auditory (right hemisphere, 156 ms; left hemisphere, 182 ms; $Q = 1 \times 10^{-4}$; permutation test) or the production epoch (right hemisphere, 272 ms; left hemisphere, 268 ms; $Q = 1 \times 10^{-4}$; see Methods). Therefore, right hemisphere responses cannot be due to computations that were first carried out in the left hemisphere and the data do not support strictly lateralized sensory–motor computations.

Another concern is that S-M activity may not reflect speech processing and may also be present during simple auditory inputs and orofacial motor outputs. To test this, we employed a 'tone–move task' in one of the bilaterally implanted subjects (subject 13 (S13); see Methods). We found that S-M electrodes did not have significant sensory–motor responses during the tone–move task ($P = 0.36$, permutation test; see Supplementary Fig. 6). Thus, S-M activity is specific to mapping sounds to goal-directed vocal motor acts and is likely specific to speech (see Supplementary Discussion 1.3).

Thus far we have shown the S-M activity is bilateral, sensory–motor, and likely to be specific to speech. However, an important open question is whether S-M responses reflect the transformation that links speech perception and production and can support a unified perception–production representational interface. A specific concern is that high gamma ECoG activity may pool heterogeneous neural responses beneath the electrode. S-M responses may combine activity from neurons which encode perceptual processes active during the auditory cue and other neurons which encode production processes active during the utterance. If this is true, none of the activity necessarily reflects a sensory–motor transformation that links perception and production. To be able to rule out this alternative and demonstrate that S-M responses are involved in sensory–motor transformations, we reasoned that two requirements must be met. S-M activity must encode information about the content of the underlying speech processes, and this encoding must reflect transformative coding between the sensory input and motor output.

To test whether S-M activity encodes information about speech content, we decoded the neural activity to predict, on each trial, what the subjects heard and said. We used seven consonant–vowel–consonant words (heat, hit, hat, hoot, het, hot and hut) and trained a seven-way linear classifier to decode the neural responses (see Methods). Individual electrodes only weakly encoded speech content, but when we decoded activity pooled across groups of electrodes, we found that all three electrode groups encoded speech tokens (see Fig. 3). AUD electrodes performed best with an average classification performance of 42.7% (χ^2 (1) = 56.5, $P = 6 \times 10^{-14}$), followed by S-M electrodes, which showed performance of 33.4% (χ^2 (1) = 25.6, $P = 4 \times 10^{-7}$), and then PROD electrodes, which showed performance of 27.1% (χ^2 (1) = 11.5, $P = 7 \times 10^{-4}$). Furthermore, classification performance for S-M electrodes did not differ between the two hemispheres (left hemisphere, 29%; right hemisphere, 27%; Fisher's exact test, $P = 0.5$; Fig. 3c). Thus, bilateral S-M activity encodes information about the sensory and motor contents of speech, meeting an important requirement for sensory–motor transformations.

We next sought to test whether S-M activity can link speech perception and production by transforming auditory input into production output. The essential requirement for transformation is that neural encoding of sensory input should depend on subsequent motor output. Previous work has characterized visual–motor transformations using a transformation task in which the spatial location of a visual cue can instruct a motor response to the same or different spatial location (the 'pro–anti task')[22,23]. Sensory–motor neurons in the dorsal visual stream display different responses to the visual cue depending on the motor contingency, demonstrating a role for these neurons in the visual–motor transformation[22].

Given these predictions from animal neurophysiology, we tested four subjects as they performed an auditory–motor transformation task (the listen–speak transformation task) that employed two non-words (kig, pob) to examine whether S-M activity has a role in transformations for speech (see Fig. 4A, Supplementary Figs 7 and 8 and Methods). This task enabled us to hold the sensory and motor components constant while manipulating the transformation process itself in order to measure how the encoding of this content changed depending on how perceptual input was mapped onto production output. The use of non-words instead of words offered other advantages. Non-words enabled us to examine sublexical transformations for speech and could be designed to differ maximally in their articulatory dimensions and their neural representations (see Methods and Supplementary Discussion 1.1 and 1.2).

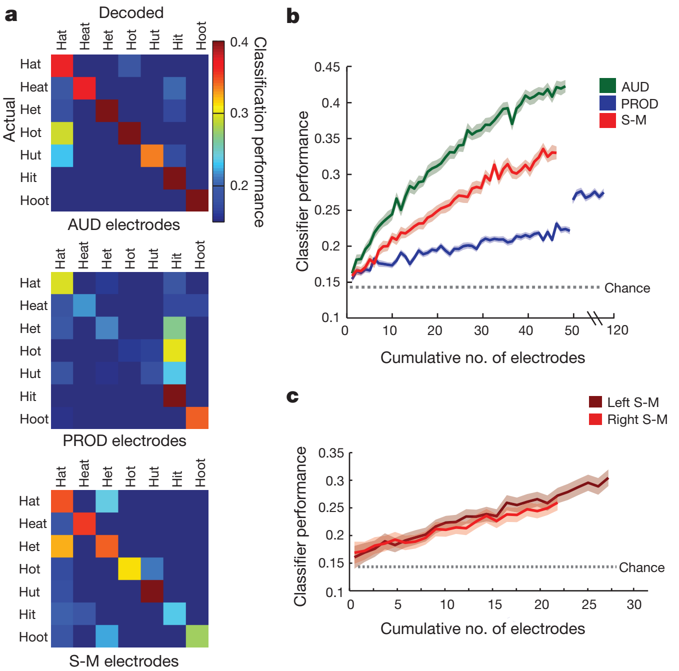

Figure 3 | Neural decoding of words. **a**, Confusion matrices show proportion identified for a seven-way linear classifier using neural responses. AUD electrodes (top), PROD electrodes (middle) and S-M electrodes (bottom) are shown. The threshold for performance is at chance level, $P = 0.14$, for the purposes of displaying the electrodes clearly. **b**, Classification performance for increasing numbers of electrodes. Chance performance is indicated by the dotted line. **c**, Classification performance for S-M electrodes in the left and right hemispheres. Methods present S-M results by response epoch.

At least three models describe how neural responses encode the task variables. If responses follow a strictly sensory model, the encoding will follow the content of the sensory inputs and confuse trial conditions in which kig is converted to kig (kig→kig) with trials in which kig is converted to pob (kig→pob), as well as trials in which pob is converted to pob (pob→pob) with trials in which pob is converted to kig (pob→kig) (see Fig. 4Ba). Conversely, responses that follow a strictly motor model will encode the production outputs, confusing kig→kig with pob→kig trials and pob→pob with kig→pob trials (see Fig. 4Bb). If S-M responses pool responses from sensory and motor neurons, the encoding will follow the sensory model during sensory input and the motor model during motor output. In contrast, S-M responses that reflect the transformation of sensory input into motor output must follow a different transformation model and encode the sensory information differently depending on the upcoming motor act (see Fig. 4Bc). Neural activity displaying this property could compute a representational transformation (see Supplementary Discussion 1.1, 1.2). If so, responses that follow a transformation model will not confuse trial conditions with either identical input or identical output. Consequently, each of the three models predicted very different patterns of neural coding.

We constructed linear classifiers to decode neural responses. As expected, AUD electrodes in the auditory epoch encoded the auditory input (Fig. 4Ba, Ca) and PROD electrodes encoded the output during

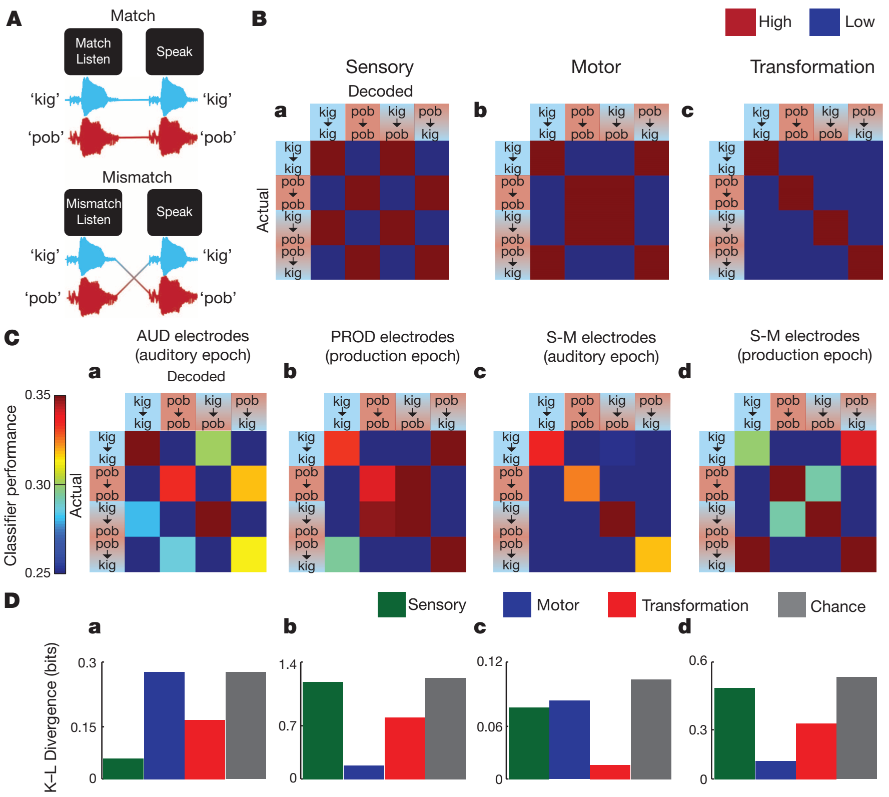

Figure 4 | Listen–speak transformation task. **A**, In the listen–speak transformation task, subjects have to transform a non-word they hear into a non-word they speak according to a simple rule. Subjects were first presented with a visual cue: 'Match Listen' or 'Mismatch Listen' that instructed the rule that determined the non-word to say in response to the non-word they heard. On 'match trials' the rule was to repeat the non-word they heard. On 'mismatch trials', the rule was to say the non-word that they did not hear. The non-words were 'kig' and 'pob'. Subjects then heard one of the two non-words, waited for a short delay, then said the appropriate non-word in response to the 'Speak' cue. There were four task conditions: kig→kig (hear 'kig' and say 'kig'); pob→pob (hear 'pob' and say 'pob'); kig→pob (hear 'kig' and say 'pob'); and pob→kig (hear 'pob' and say 'kig'). **B, a–c**, Confusion matrices predicted by the sensory, motor and transformation models with high and low classification scores. **C, a–d**, Confusion matrices during the listen–speak transformation task. **D, a–d**, Model fit quantified using a Kullback–Leibler (K–L) divergence.

the production epoch (utterance; Fig. 4Bb, Cb). However, S-M electrodes changed their encoding over the course of the trial. During the auditory epoch, S-M electrodes encoded both sensory and motor conditions concurrently, consistent with the presence of a sensory–motor transformation (Fig. 4Bc, Cc). Interestingly, during the production epoch, S-M responses no longer encoded the auditory input and encoded the production output (Fig. 4Cd), suggesting the transformation has largely been computed by that time. To quantify the comparison of different models, we used the Kullback–Leibler divergence (see Fig. 4Da–d, Methods). The results were consistent with the response patterns in the confusion matrices.

We can also rule out that the difference in S-M responses is due to a third population of neurons that selectively responds to the cue instructing how perceptual input was mapped onto production output ('match' or 'mismatch'). We ran the same linear classifier during cue presentation and found that the S-M responses did not encode the cue (χ^2 (1) = 0.08, $P = 0.78$; see Methods).

Using direct brain recordings (ECoG) and overt speech, we demonstrate that a sensory–motor system for transforming sublexical speech signals exists bilaterally. Our results are in keeping with models of speech perception that posit bilateral processing but contradict models that posit lateralized sensory–motor transformations[11,16]. Our results also highlight how S-M activity during perceptual input reflects the transformation of speech sensory input into motor output. We propose that the presence of such transformative activity demonstrates a unified sensory–motor representational interface that links speech-perception- and speech-production-based representations. Such an interface is important during speech articulation, acquisition and self-monitoring[24–26]. As right hemisphere lesions do not give rise to conduction aphasia[19,27–29], our evidence for bilateral sensory–motor transformations promotes an interesting distinction between speech and language: although sensory–motor transformations are bilateral, the computational system for higher-order language is lateralized[5,6] (see Supplementary Fig. 9). This hypothesis invokes a strong interface between sensory-based speech-perception representations and motor-based speech-production representations and suggests that deficits for conduction aphasia are more abstract and linguistic in nature. We propose that bilateral sublexical transformations could support a unification of perception- and production-based representations into a sensory–motor interface[6], drawing a distinction between the bilateral perception–production functions of speech and lateralized higher order language processes.

METHODS SUMMARY

Electrocorticographic (ECoG) recordings were obtained from 16 patients (10 females) undergoing treatment for pharmacologically resistant epilepsy. Each patient provided informed consent in accordance with the Institutional Review Board at New York University Langone Medical Center. Grid implantation was in the left hemisphere (6 subjects), right hemisphere (7 subjects) or both hemispheres (3 subjects). All 16 subjects performed an overt word repetition task (listen–speak task) as well as two control tasks (listen–mime task[30] and listen task). One subject also performed a tone–move task. Four subjects also performed a listen–speak transformation task involving non-words. ECoG recordings were made using both grid and strip electrode arrays with 2.3-mm contact size and 10-mm spacing. Spectral analysis was performed using 500-ms analysis windows with ±5-Hz frequency smoothing and a stepping size of 50 ms. Neural responses were defined as high gamma neural activity between 70 and 90 Hz and significance was assessed using a shuffling procedure. Classification analyses were carried out using a linear discriminant analysis with high gamma power spectral features.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 September; accepted 2 December 2013.

Published online 15 January 2014.

1. Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the Speech Code. *Psychol. Rev.* **74,** 431–461 (1967).
2. Liberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21,** 1–36 (1985).
3. Halle, M. & Stevens, K. N. Speech recognition: a model and a program for research. *IEEE Trans. Inf. Theory* **8,** 155–159 (1962).
4. Halle, M. & Stevens, K. N. Analysis by synthesis In (eds Wathen-Dunne, W. & Woods, L. E.) *Proceedings of seminar on speech compression and processing* Vol. 2 paper D7 1959).
5. Berwick, R. C., Friederici, A. D., Chomsky, N. & Bolhuis, J. J. Evolution, brain, and the nature of language. *Trends Cogn. Sci.* **17,** 89–98 (2013).
6. Chomsky, N. *The Minimalist Program* (MIT Press, 1995).
7. Jakobson, R. *Child Language, Aphasia and Phonological Universals* (Mouton, 1968).
8. Lichtheim, L. On aphasia. *Brain* **7,** 433–484 (1885).
9. Wernicke, C. The aphasic symptom-complex: a psychological study on an anatomical basis. *Arch. Neurol.* **22,** 280–282 (1970).
10. Geschwind, N. Disconnexion syndromes in animals and man. I. *Brain* **88,** 237–294 (1965).
11. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nature Rev. Neurosci.* **8,** 393–402 (2007).
12. Price, C. J. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. NY Acad. Sci.* **1191,** 62–88 (2010).
13. Obleser, J., Eisner, F. & Kotz, S. A. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* **28,** 8116–8123 (2008).
14. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Rev. Neurosci.* **12,** 718–724 (2009).
15. Hickok, G., Houde, J. & Rong, F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* **69,** 407–422 (2011).
16. Hickok, G., Okada, K. & Serences, J. T. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* **101,** 2725–2732 (2009).
17. Guenther, F. H. Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* **39,** 350–365 (2006).
18. Wise, R. J. S. *et al.* Separate neural subsystems within 'Wernicke's area'. *Brain* **124,** 83–95 (2001).
19. Caramazza, A., Basili, A. G., Koller, J. J. & Berndt, R. S. An investigation of repetition and language processing in a case of conduction aphasia. *Brain Lang.* **14,** 235–271 (1981).
20. Crone, N. E., Sinai, A. & Korzeniewska, A. Event-related dynamics of brain oscillations. *Prog. Brain Res.* **159,** 275–295 (2006).
21. Markowitz, D. A., Wong, Y. T., Gray, C. M. & Pesaran, B. Optimizing the decoding of movement goals from local field potentials in macaque cortex. *J. Neurosci.* **31,** 18412–18422 (2011).
22. Zhang, M. & Barash, S. Neuronal switching of sensorimotor transformations for antisaccades. *Nature* **408,** 971–975 (2000).
23. Gail, A. & Andersen, R. Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. *J. Neurosci.* **26,** 9376–9384 (2006).
24. Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S. & Houde, J. F. Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc. Natl Acad. Sci. USA* **110,** 2653–2658 (2013).
25. Oller, D. K., Eilers, R. E. & Oiler, D. K. The role of audition in infant babbling the role of audition in infant babbling. *Child Dev.* **59,** 441–449 (1988).
26. Agnew, Z. K., McGettigan, C., Banks, B. & Scott, S. K. Articulatory movements modulate auditory responses to speech. *Neuroimage* **73,** 191–199 (2013).
27. Goodglass, H., Kaplan, E. & Barresi, B. *Assessment of Aphasia and Related Disorders* (Lippincott Williams & Wilkins, 2000).
28. Damasio, H. & Damasio, A. R. The anatomical basis of conduction aphasia. *Brain* **103,** 337–350 (1980).
29. Benson, D. F. *et al.* Conduction aphasia: a clinicopathic study. *Arch. Neurol.* **28,** 339–346 (1973).
30. Murphy, K. *et al.* Cerebral areas associated with motor control of speech in humans cerebral areas associated with motor control of speech in humans. *J. Appl. Physiol.* **83,** 1438–1447 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank A. Weiss, J. MacArthur and L. Frank for developing the data acquisition hardware and software; O. Felsovalyi, E. Londen, P. Purushothaman, L. Melloni, C. Boomhaur and A. Trongnetrpunya for technical assistance; and D. Poeppel and C. Brody for comments on the manuscript. This work was supported, in part, by R03-DC010475 from the NIDCD, a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund (B.P.), a Watson Investigator Program Award from NYSTAR (B.P.), a McKnight Scholar Award (B.P.) and a Sloan Research Fellowship (B.P.).

Author Contributions G.B.C. designed the experiment, performed the research, analysed the data and wrote the manuscript. T.T. and O.D. performed the research and wrote the manuscript. C.C. and W.D. performed the research. B.P. designed the experiment, performed the research, analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.P. (bijan@nyu.edu).

METHODS

Participants. Electrocorticographic (ECoG) recordings were obtained from 16 patients (6 males, 10 females; see Supplementary Table 1) with pharmacologically resistant epilepsy undergoing clinically motivated subdural electrode recordings at the New York University School of Medicine Comprehensive Epilepsy Center. Informed Consent was obtained from each patient in accordance with the Institutional Review Board at the New York University Langone Medical Center. Patient selection for the present study followed strict criteria: first, cognitive and language abilities in the average range or above, including language and reading ability, as indicated by formal neuropsychological testing (see Supplementary Table 1); and second, normal language organization as indicated by cortical stimulation mapping, when available. In addition, only electrode contacts outside the seizure onset zone and with normal interictal activity were included in the analysis.

Behavioural tasks and recordings. All participants performed three behavioural tasks: listen–speak, listen–mime and listen (Fig. 1a). Behavioural tasks were performed while participants reclined in a hospital bed. Tasks were controlled by a computer placed on the service tray on the bed running the Presentation program (NeuroBehavioural Systems). Behavioural audio recordings were either synchronized with the neural recordings at 10 kHz (see below) or recorded on the computer and referenced to the go cue. For a subset of subjects, a video camera with built-in microphone (Sony) was positioned to monitor subject orofacial movements and utterances. Video was streamed to disk (Adobe Premier Pro. (video at 29.95 frames per s, and audio at 44.1 kHz)). Audio-visual and neural signals were synchronized video-frame-by-video-frame using an Analogue-to-Digital Video Converter (Canopus).

Listen–speak, listen–mime[30] and listen tasks were randomly interleaved on a trial-by-trial basis with at least 4 s between trials. Each trial began with a visual cue presented, followed by the auditory consonant–vowel–consonant (CVC) token 500 ms later. We used CVC words composed of the same consonants, 'h' and 't', and different vowels (hat, hit, heat, hoot, het, hut, hot). These spoken syllables span the vowel space and differ in their auditory and articulatory content. Subjects had to either listen passively (listen), repeat the syllable after a cue (listen–speak) or mime the syllable after a different cue (listen–mime, produce the appropriate mouth movements but with no vocal cord vibration[30]; see Supplementary Fig. 10). The temporal delay between the auditory cue and the movement cue was 2 s. We obtained between 49 and 166 trials per condition (within subject) and between 175 and 334 total trials per subject.

For the tone–move task (see Supplementary Fig. 6), after the listen cue was delivered, a 500-ms, 1,000-Hz sinusoidal tone (with 100-ms on and off ramps) was presented. After a short, 2-s delay another visual cue was presented (move) instructing the subject to move their articulators (tongue, lips and jaw). For one subject, these trials were randomly interleaved within blocks of the listen–speak, listen–mime and listen tasks (see above).

For the listen–speak transformation task, four subjects (see Supplementary Figs 7 and 8) were first presented with one of two visual cues: 'Match Listen' or 'Mismatch Listen'. After a delay, subjects heard one of two non-words: 'kig' (/kIg/) or 'pob' (/pɑb/). These non-words were chosen to differ maximally on their articulator dimensions: 'kig' contains a velar (back) voiceless stop consonant, followed by a high front vowel and finally a velar voiced stop consonant, and 'pob' contains a bilabial (front) voiceless stop consonant followed by a back low vowel and then a bilabial front voiced stop consonant. The tongue movement therefore goes back to front to back for 'kig' and front to back to front for 'pob'. The reason for choosing maximally different articulations was that larger articulator differences might lead to larger neural activity differences. After a short delay (randomized between 1.5 and 2 s), another visual cue was presented ('Speak') to which subjects were to respond by saying the match non-word they had heard if they had seen the initial match cue, or say the mismatch non-word if they had seen the mismatch cue. Each non-word within each condition was presented between 63 and 78 times per subject, with total trials ranging from 255 to 309 per subject. This control was carried out in separate blocks trials that alternated with blocks of the main listen–speak, listen–mime and listen tasks.

Surface reconstruction and electrode localization. To localize electrode recording sites, pre-surgical and post-surgical T1-weighted magnetic resonance imaging (MRI) scans were obtained for each patient and co-registered with each other[31]. The co-registered images were then normalized to an MNI-152 template and electrode locations were then extracted in MNI (Montreal Neurological Institute) space (projected to the surface) using the co-registered image, followed by skull stripping[32]. A three-dimensional reconstruction of each patient's brain was computed using FreeSurfer (Fig. 2; S2, S3, S4, S5, S6, S7, S8 and S10 (ref. 33). For Supplementary Table 2, Talairach coordinates were converted from MNI space using the EEG/MRI toolbox in Matlab (http://sourceforge.net/projects/eeg/, GNU General Public License).

Neural recordings and preprocessing. EEG data were recorded from intracranially implanted subdural electrodes (AdTech Medical Instrument Corp.) in patients undergoing elective monitoring of pharmacologically intractable seizures. Electrode placement was based entirely on clinical grounds for identification of seizure foci and eloquent cortex during stimulation mapping, and included grid (8 × 8 contacts), depth (1 × 8 contacts) and strip (1 × 4 to 1 × 12 contacts) electrode arrays with 10-mm inter-electrode spacing centre-to-centre. Subdural stainless steel recording grid and strip contacts were 4 mm in diameter; consequently the distance between contacts was 6 mm and they had an exposed 2.3-mm diameter recording contact.

For 7 of the 16 subjects, neural signals from up to 256 channels were amplified (×10, INA121 Burr-Brown instrumentation amplifier), bandpass filtered between 0.1–4,000 Hz and digitized at 10 kHz (NSpike, Harvard Instrumentation Laboratories) before being continuously streamed to disk for off-line analysis (custom C and Matlab code). The front-end amplifier system was powered by sealed lead acid batteries (Powersonic) and optically isolated from the subject. After acquisition, neuronal recordings were further low-pass filtered at 800 Hz and down-sampled offline to 2,000 Hz for all subsequent analysis. For the remaining 9 subjects, neural signals from up to 128 channels were recorded on a Nicolet One EEG system, bandpass-filtered between 0.5–250 Hz and digitized at 512 Hz. In some recordings, modest electrical noise was removed using line-filters centred on 60, 120 and 180 Hz (ref. 34).

Data Analysis. For activation analysis, time-frequency decomposition was performed using a multi-taper spectral analysis[34]. The power spectrum was calculated during a 500-ms analysis window with ± 5 Hz frequency smoothing stepped 50 ms between estimates. Single trials were removed from the analysis if the raw voltage exceeded eight standard deviations from the across trial pool, and noisy channels were removed from the analysis by visual inspection or if they did not contain at least 60% of the total trials after the standard deviation threshold removal.

Sensory–motor transformations were defined as activity in the gamma range (70–90 Hz) that followed the auditory stimulus as well as the production cue during both listen–speak and listen–mime (Fig. 1b). As the example responses illustrate, some electrodes showed consistent increases in activity in the high gamma band as high as 300 Hz. As the frequency extent varied across subjects, we chose to focus on the 70–90-Hz frequency range as this band showed the greatest activation consistently across all subjects. Similar results were obtained when a broader frequency range extending up to 150 Hz was analysed. Although the listen–mime condition does involve altering the motor plan (no vocal cord vibration), sensory–motor activations were based on the conjunction of activity in both the listen–speak and the listen–mime conditions. Any neural activity that was specific to the listen–mime condition and not present in 'normal' speaking conditions was therefore excluded (see Supplementary Fig. 10).

Responses were divided into three types. The first response type, auditory (AUD), was defined as containing a response that was seen within 250–750 ms following the onset of the auditory stimulus in all three conditions (Fig. 1b, top). The second response type, production (PROD), was characterized as containing a response occurring between 500–1,000 ms after the respond cue in the listen–speak and the listen–mime conditions (Fig. 1b, middle). The last response type, S-M, contained both post stimulus and a post response cue activation in both the listen–speak and the listen–mime conditions (Fig. 1b bottom). The baseline period was defined as the 500 ms preceding the auditory stimulus.

In Fig. 1b, the experimental epoch was defined as −500 ms (pre) to 3,500 ms post auditory stimulus onset. In Fig. 2c the experimental epoch was defined as −500 ms (pre) to 4,000 ms post auditory stimulus onset. The additional 500 ms was included in Fig. 2c to compensate for slightly later production responses for that the subject. Power in each frequency band was normalized to the power in the baseline period by dividing by the power at each frequency. As the neural responses had variable onset times but were on average quite long in duration, the times were chosen to sample adequately all the responses under investigation.

To assess statistical significance, the average power across trials was taken in two time regions of interest for each trial within each condition. For the listen condition, the baseline values for each trial were shuffled with the post auditory values 10,000 times to create a null distribution. For the listen–speak and the listen–mime conditions, both the post-auditory and the post-production values were shuffled 10,000 times with the baseline values to create two null distributions. Initial significance was assessed using a permutation test by comparing the actual difference between the post auditory and post production values with the shuffled data differences[35]. To correct for multiple comparisons, for all subjects, all three conditions and both analysis epochs (listen (post auditory), listen–speak (post auditory and post production) and listen–mime (post auditory and post production)) were pooled together and a false discovery rate (FDR) analysis was performed with an alpha threshold set at 0.05 (ref. 36).

The population latency analysis was performed using the baseline-corrected high gamma power response profiles for each electrode within each response class (S-M, AUD and PROD). The high gamma neural responses were first bandpass filtered (70–90 Hz) and then averaged within conditions. The listen–speak and listen–mime conditions were averaged together. As the data were recorded using two different sampling rates, the data were resampled to a 500 Hz sampling rate. To test for latencies within a response class, the latencies following either the auditory onset or the go cue were compared against the activity in the listen condition following the go cue by computing a permuted distribution for each time point. The significance values at each time point were then corrected for multiple comparisons using a FDR set with an alpha of 0.05. The first time point that was followed by at least 20 consecutive significant time points (40 ms) was taken to be the latency of the neural response. This resulted in four significant latency values. In the auditory epoch, AUD electrodes had significant neural responses at 164 ms and S-M electrodes had significant responses at 158 ms. During the production epoch, PROD electrodes had significant responses starting at 302 ms, whereas S-M electrodes had significant responses starting at 248 ms. A similar analysis was carried out comparing the left S-M electrodes with the right S-M electrodes, which resulted in four more significant latency values: right hemisphere (auditory 156 ms), left hemisphere (auditory 182 ms), right hemisphere (production 272 ms) and left hemisphere (production 268 ms). A direct comparison between these latencies within each task epoch using FDR-corrected shuffle tests (see above) revealed no significant results.

To assess whether or not during the auditory and production epochs, the S-M electrodes display significantly faster neural responses than the AUD and PROD electrodes, we repeated the permutation test, except instead of using the comparison of the task compared to the 'Listen' condition, we compared the S-M electrodes to the AUD electrodes in the auditory epoch and the S-M electrodes to the PROD electrodes in the production epoch. The results showed that whereas S-M and AUD electrodes did not differ in their latency values during the auditory epoch, S-M electrodes were significantly faster than PROD electrodes in the production epoch.

To test for power differences of the high gamma response (70–90 Hz) across hemispheres, we performed FDR-corrected permutation tests. Data were analysed by averaging a 300-ms time window, sliding 50 ms between estimates. The data were baseline-corrected (average -500 ms (pre) to 0 ms pre-stimulus activity across conditions, within electrodes) and then log-transformed before analysis. For each condition (listen–speak, listen–mime and listen) and within each hemisphere (left and right), we computed the task epoch responses by computing the average of the high gamma response during the auditory epoch (0–1,000 ms post auditory onset) and during the production epoch (0–1,500 ms post production cue onset). We then performed a series of permutation tests where we permuted the neural response across condition and/or across hemisphere, correcting for multiple comparisons using a FDR procedure. Only four tests produced significant results: listen–speak versus listen during the production epoch in each hemisphere, and listen–mime versus listen during the production epoch in each hemisphere. Furthermore, the neural responses within all conditions were not different across hemispheres (see Supplementary Fig. 11, $P > 0.05$, FDR corrected).

To assess the significant delay activation for each electrode class, a permutation test was carried out using filtered data as listed above. A permutation test was performed for each electrode class in which the average high gamma neural activity of the delay period (1–2 s post auditory onset) was compared to that of the baseline period (-1 s to -0.5 s pre auditory onset). Although PROD electrodes and AUD electrodes did not display elevated population neural activity ($P = 0.64$ and 0.53, respectively), S-M electrodes had significantly higher elevated delay activity compared to baseline ($P = 0.01$; see Fig. 2d).

Classification was performed using the single value decomposition (SVD) of the high gamma neural response (70–160-Hz, 300-ms sliding windows with an overlap of 250 ms) in either the auditory epoch (0–1,000 ms post auditory onset, AUD electrodes) or the production epoch (0–1,500 ms post go cue, PROD electrodes) or both (S-M electrodes). A linear discriminant analysis (LDA) classification was performed using a leave-one-out validation method, in which the training set consisted of all the trials of the data set except the one being tested. Note that analysing the different task epochs separately for the S-M electrodes produced classifier results that were also significantly above chance (auditory epoch, 40.2% (χ^2 (1) = 47, $P = 7 \times 10^{-12}$), production epoch, 23.2% (χ^2 (1) = 5.6, $P = 0.02$)).

To create the cumulative curves, the number of electrodes inputted into the classifier was increased linearly. To control for the variability in trial numbers, the minimum number of trials common to all subjects and electrodes was used. One-hundred iterations for each number of cumulative electrodes were performed, in which the specific trials and the specific electrodes were varied randomly and the number of SVD components was equal to the number of electrodes inputted to the classifier for the AUD and S-M electrodes, whereas five components were used for the PROD electrodes due to a lower number of components present in the PROD-electrode data.

Confusion matrix scores are simply the proportion of trails classified as the token on the horizontal axis (decoded) given that the actual trial came from the vertical axis (actual). Confusion matrices in Fig. 3a are shown for the largest number of cumulative electrodes in each electrode class.

To analyse the listen–speak transformation task responses (Fig. 4), the same decomposition (SVD) of the neural signal (70–160 Hz) was used. Note that instead of a seven-way classifier, a four-way classifier was used. Confusion matrices (Fig. 4C) are shown for the largest number of cumulative electrodes in each electrode class (AUD = 10; PROD = 19; S-M = 8). For the S-M electrodes, each response epoch (auditory, Fig. 4Cc; production, Fig. 4Cd) was analysed separately.

To measure the quality of each of the models (sensory, motor, sensory–motor or chance; Fig. 4d) we used the Kullback–Leibler divergence, which quantifies the amount of information lost in bits when Q (the model) is used to approximate P (the data):

$$D_{\mathrm{KL}}(P||Q) = \sum_i P(i) \log_2 \left(\frac{P(i)}{Q(i)}\right)$$

where P is the classification percentage for each actual or decoded pair (see above) and Q is one of the four models: sensory, motor, sensory–motor and chance. The Kullback–Leibler divergence estimates the information distance between the pattern of classification errors predicted by each model, shown in Fig. 4B and the pattern of classification errors based on neural recordings, shown in Fig. 4C. Smaller Kullback–Leibler divergence reflects more information about classification errors and improved model fit. The sensory model (Fig. 4Ba) reflects classification scores that track the auditory speech input such that in both the match and the mismatch cases, the same input will be confused with each other. Conversely, the motor model (Fig. 4Bb) reflects classification scores that track the production output so that the same outputs will be confused with one another. However, the sensory–motor model (Fig. 4Bc) will reflect both the input and output such that classifications for each of the conditions presented (kig→kig, pob→pob, kig→pob and pob→kig) will be classified correctly. Finally, the chance model will simply reflect chance performance in all cases (0.25).

Classifier analysis of the cue data ('Match Listen' versus 'Mismatch Listen') in the listen–speak transformation task was analysed on the S-M electrodes for the subjects performing the task. The same linear classifier was used as above, but was performed during the cue period (0–1,000 ms post Cue) and was two-way ('Match Listen' versus 'Mismatch Listen' cues). The results demonstrated that the classification was not significant (mean classification = 52.3%, χ^2 (1) = 0.08, $P = 0.78$). Furthermore, using the same two-way classifier between the match and mismatch condition during the auditory epoch was also not significant (mean classification = 56.4%, χ^2 (1) = 0.72, $P = 0.4$). Taken together, this indicates that the sensory–motor transformations displayed by these electrodes cannot be due to a third population of neurons that code for the visual cue.

31. Yang, A. I. *et al.* Localization of dense intracranial electrode arrays using magnetic resonance imaging. *NeuroImage* **63,** 157–165 (2012).
32. Kovalev, D. *et al.* Rapid and fully automated visualization of subdural electrodes in the presurgical evaluation of epilepsy patients. *AJNR Am. J. Neuroradiol.* **26,** 1078–1083 (2005).
33. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I: segmentation and surface reconstruction. *Neuroimage* **9,** 179–194 (1999).
34. Mitra, P. P. & Pesaran, B. Analysis of dynamic brain imaging data. *Biophys. J.* **76,** 691–708 (1999).
35. Maris, E., Schoffelen, J.-M. & Fries, P. Nonparametric statistical testing of coherence differences. *J. Neurosci. Methods* **163,** 161–175 (2007).
36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).