A Corollary Discharge Circuit in Human Speech

Amirhossein Khalilian-Gourtani,[1*] Ran Wang,[2] Xupeng Chen,[2]
Leyao Yu,[3] Patricia Dugan,[1] Daniel Friedman,[1] Werner Doyle,[4]
Orrin Devinsky,[1] Yao Wang,[2,3] Adeen Flinker,[1,3*]

[1]Neurology Department, New York University, New York, 10016, NY, USA
[2]Electrical and Computer Engineering Department, New York University,
Brooklyn, NY, 11201, USA
[3]Biomedical Engineering Department, New York University,
Brooklyn, 11201, NY, USA
[4]Neurosurgery Department, New York University, New York, 10016, NY, USA

*Corresponding authors; E-mail: amirhossein.khalilian@nyu.edu, adeen@nyu.edu

Abstract: When we vocalize, our brain distinguishes self-generated sounds from external ones. A corollary discharge signal supports this function in animals, however, in humans its exact origin and temporal dynamics remain unknown. We report Electrocorticographic (ECoG) recordings in neurosurgical patients and a novel connectivity approach revealing major neural communications. We find a reproducible source for corollary discharge across multiple speech production paradigms localized to ventral speech motor cortex before speech articulation. The uncovered discharge predicts the degree of auditory cortex suppression during speech, its well-documented consequence. These results reveal the human corollary discharge source and timing with far-reaching implication for speech motor-control as well as auditory hallucinations in human psychosis.

One-sentence summary: We reveal the source and timing of a corollary discharge from speech motor cortex onto auditory cortex in human speech.

1 Introduction

How does the brain dissociate self-generated stimuli from external ones? Any motor act directly activates associated sensory systems. This constant flow of sensory information, while useful as feedback, can desensitize the sensory system or be confused with external sensations (*1, 2*). A fundamental brain circuit solves this problem by transmitting a blueprint of the motor signal to sensory cortices, referred to as corollary discharge (*3*). The corollary discharge signals are established in many species and multiple sensory modalities (*1, 2, 4–6*). In the human auditory system, a corollary discharge is hypothesized to increase sensitivity to self-generated speech during production (*7–9*) and when impaired can lead to auditory hallucinations (*10, 11*).

Corollary discharge signals decrease the sensory processing load and increase sensitivity during vocalization by suppressing sensory cortices (*12, 13*). For example, an inter-neuron in the cricket motor system inhibits their auditory system during loud chirping to avoid desensitization (*2*). In non-human primates, vocalization suppresses auditory neurons, supporting an auditory corollary discharge circuit (*12*). Similarly, self-produced human speech suppresses auditory cortex (*14–17*) and schizophrenia patients with auditory hallucinations have impaired suppression (*10, 11*). In primates and humans, the source of the corollary discharge signal and its dynamics remain virtually unknown (*14, 18*).

Recent neuroscience approaches focus on the study of connectivity and information flow between cortical regions (*19–21*). However, many leverage non-invasive neuroimaging data with a limited temporal (i.e. fMRI) or spatial (i.e M/EEG) resolution and typically do not assess the directionality of information flow (e.g. functional connectivity). Here, we leverage rare human neurosurgical recordings from motor and auditory related cortical sites during speech production tasks providing both a high spatial and temporal resolution. Leveraging advanced directed connectivity approaches combined with unsupervised learning we identify the source,

target, and temporal dynamics of information flow across cortices. A discharge signal before speech onset is transmitted from ventral motor cortex to auditory cortex, and is reproducible across multiple speech tasks and patients. This directed signal from motor cortex predicts the degree of suppression across auditory sites, providing the first direct evidence for a corollary discharge signal and its circuit dynamics in humans.

2 Results

To investigate the corollary discharge (CD) signal in human speech, we employed a paradigm that directly measures the CD signal's outcome, i.e., speech induced auditory suppression (*14*). We acquired electrocorticographic (ECoG) recordings from eight neurosurgical patients while they performed an auditory word repetition task. We focused on high-gamma broadband (70-150 Hz), a marker of neural activity. We first established the temporal and spatial neural recruitment during auditory perception (Fig. 1A) and production (Fig. 1B). Neural activity commenced in auditory cortices (i.e. superior temporal gyrus; STG) shortly after stimulus onset (see 100ms in Fig. 1A, see Table S1) followed by inferior frontal cortices (Fig. 1, A and C). Prior to speech onset, activity arises in inferior frontal gyrus (IFG) followed by sensorimotor cortices (pre- and post- central gyri; Fig. 1, B and D). During articulation there is a marked suppression of auditory activity (compared with perception), a commonly reported consequence of a corollary discharge (Fig. 1, B to D). While these results raise potential candidates for a corollary discharge before speech production across frontal cortices (e.g. IFG, MFG, sensorimotor cortex), local neural activity does not elucidate the exact source and dynamics of the corollary discharge signal.

The corollary discharge signal, by nature, is a blueprint of the motor commands sent to auditory cortex. Hence, a technique is necessary which can measure both the degree of communication between brain regions as well as the direction of information flow. To this end, we

developed a directed connectivity analysis framework describing the causal predictive relationships between signals across different electrodes. Unlike previous approaches (e.g. directed transfer function (*22*), partial directed coherence (*23*)) we distilled the large neural connectivity patterns into the dominant communication patterns using an unsupervised clustering technique (orthogonal non-negative matrix factorization (*24*)). Our analysis framework (see Fig. 2 for summary and supplementary text S.5 and Fig. S1 for details) summarizes the dynamics of directed connectivity across cortical regions using a set of connectivity temporal prototypes, which depict typical temporal variation patterns of directed connectivity together with the corresponding cortical sources and targets.

We first employed our approach in a representative participant (spatiotemporal high gamma activity shown in Fig. 3, A and B), revealing three major connectivity prototypes locked to auditory stimulus onset (Fig. 3C). The first prototype (blue curve), peaking at 120 msec, showed information flow from STG onto IFG as well as speech motor cortex. The second prototype (yellow curve), peaking at 340 msec, showed information flow from STG and IFG onto speech motor cortex. The third prototype (purple curve), peaking at 690 msec, showed information flow from speech motor cortex onto STG. Unlike the high-gamma activity patterns (Fig. 3A), the directed connectivity analysis reveals the source, target, and temporal dynamics of statistically significant information flow (permutation test, $p < 0.05$, see section S.5) across cortical regions. The first two prototypes peak during early stages of auditory processing (while still hearing the auditory stimulus) and exhibit sources from auditory cortex (i.e. STG) implicating these components in auditory comprehension. However, the third prototype shows information flow from speech motor cortex onto auditory cortex and peaks before mean articulation onset (Fig. 3C horizontal violin plot). Neural information flow from motor cortex onto auditory cortex before articulation onset is consistent with the timing and directionality of a corollary discharge signal, providing the first provisional evidence for such a discharge in human speech.

To ascertain the exact timing of information flow related to speech production, we repeated our analysis framework locked to articulation onset. The first connectivity prototype replicated our previous analysis matching information flow from STG and IFG onto speech motor cortex during auditory comprehension (i.e. yellow curve locked to perception, Fig. 3C). The next prototype (purple, Fig. 3D) shows information flow from speech motor cortex onto auditory cortex peaking at -70 msec prior to articulation onset. Lastly, a third prototype (green, Fig. 3D) peaking at 210 msec shows information flow from motor cortex onto temporal and inferior frontal areas during speech articulation. Our directed connectivity analyses reveal four distinct stages during auditory repetition, two related to auditory comprehension, a pre-articulatory prototype and a speech production prototype. We then replicated this finding across eight participants (clustering the directed connectivity from all the participants together, see Fig. 3, E and F, and Fig. S2 as well as Fig. S3 for variability across participants). The timing before articulation and directionality from motor to auditory cortex of the pre-articulatory prototype (purple, Fig. 3F) establishes a corollary discharge signal replicated across participants.

To verify that the uncovered corollary discharge prototype is not specific to an auditory repetition task, we leveraged a battery of speech production tasks performed by the same participants. These tasks were designed to elicit the same set of matched words during articulation while using various stimulus modalities and word retrieval routes. Participants were instructed to name visual images, read written words, complete auditory sentences, and name auditory descriptions. By clustering the directed connectivity measures across participants for each task, we replicated the corollary discharge temporal prototype and corresponding information flow across all tasks (projected on a template brain shown in Fig. 4, A to E). To investigate how the source varied across sensorimotor cortex and tasks, we analyzed the variance of the corollary discharge's outflow weights. The source did not differ statistically across tasks (ANOVA main effect of task $F(4,369)=1.02$, $p=0.397$). However, the dorsal and ventral divisions of sen-

sorimotor cortex (precentral and postcentral gyri) were significantly different (ANOVA main effect of region F(3,369)=12.48, p=9.04e-8; no significant interaction F=0.53, p=0.896) with the majority of outflow weights originating in ventral precentral gyrus (see Fig. S4 for post-hoc analysis). To verify the robustness of the corollary discharge temporal profile across participants we clustered the directed connectivity temporal patterns for each participant and task separately. Across participants, we found similar peak timing relative to articulation onset which was not statistically significant across tasks (Kruskal-Wallis test χ^2=6.86, p=0.14), providing an overall mean estimate of, μ = -107.5 msec, and peak directed connectivity value, μ = 0.1049 (Fig. 4, G and H). Together, these results provide strong evidence across participants and retrieval routes for a corollary discharge signal peaking prior to articulation (-107.5 msec) with neural communication from ventral speech motor cortex onto auditory cortex (Fig. 4F).

To date, there was a lack of direct evidence for a corollary discharge signal and its exact location, although its consequence – a suppression of auditory cortex during speech production – has been reported (*4, 14, 25*). To verify that the novel prototype we discovered is the source of a corollary discharge we sought to show a direct link to speech induced auditory suppression. Previous human (*14*) and non-human primates (*12*) studies reported a wide range of suppressed responses in auditory cortex. We hypothesized that the degree of information flow from the corollary discharge source should predict the intensity of suppression across auditory cortex recording sites. We first quantified the level of neural activity for each electrode anatomically located within auditory cortex (i.e. within participant parcellation of superior temporal gyrus, see Fig. 5, A and B) and quantified the level of suppression by computing a normalized index (varying between 1 for completely suppressed and -1 for completely enhanced, see Fig. 5, C and D). We then quantified the degree of information flow into each electrode in auditory cortex (i.e. inflow connectivity weights see Fig. 5E; see methods S.6) from the sources in our corollary discharge prototype. To test our hypothesis, we correlated the degree of information flow

into each auditory electrode with its level of suppression and found a significant relationship (Pearson correlation coefficient r=0.430, p=1.45e-4, see Fig. 5F; see methods S.6). This further establishes that information flow from speech motor cortex before articulation promotes auditory suppression during speech. A majority of the outflow over speech motor cortex originated in ventral motor cortex (Fig. 5G). To verify the timing of this suppression signal, we repeated the correlation analysis across multiple time windows while ensuring that the information flow originated from ventral motor cortex and found significant correlations (Permutation test, $p < 0.05$; see methods S.6) peaking before speech articulation (Fig. 5H).

3 Discussion

Corollary discharge signals from motor onto auditory neural populations are a hallmark of neural circuitry in animals and convey information on impending actions (*4, 26, 27*). In humans however, the exact source and dynamics of the signal remained unknown. We leveraged the excellent spatiotemporal resolution of electrocorticography recordings from neurosurgical patients and developed a novel signal analysis framework to study this circuit in human speech. Our analysis of neural connectivity dynamics revealed four distinct stages during auditory repetition likely representing comprehension (two stages), pre-articulatory preparation, and speech production per se. A cardinal connectivity temporal prototype during the pre-articulatory stage showed information flow from speech motor cortex onto auditory cortex before articulation onset. This prototype was replicated across participants and speech production tasks even when varying stimulus modality and routes of word retrieval (i.e. repetition, naming, completion), highlighting ventral pre-central gyrus as the source of the corollary discharge. Lastly, we directly linked the information flowing from this discharge with the degree of auditory cortical suppression in each electrode. Together, these findings depict the timing, source, and consequence of the corollary discharge signal in human speech.

We designed a novel directed connectivity analysis framework to study the cortical information flow by fusing techniques that analyze the causal relationships between time-series (*22, 28–33*) and unsupervised clustering (orthogonal non-negative matrix factorization (*24*)). Only a select few previous ECoG publications have leveraged similar effective connectivity between brain regions during language tasks (*34, 35*), revealing the role of Broca's area in coordinating articulately preparation (*21*) as well as frontal orthographic processing during word reading (*36*). However, a major limitation of these approaches is the large set of revealed connections rendering the results difficult to interpret and test statistically without averaging across cortical regions. Our unsupervised clustering approach circumvents this issue by finding a select few source-and-target pairs of information flow representing the prototypical temporal dynamics of connectivity across cortex. Our approach identified cortical communication supporting speech preparation (*21*), within a single and across participants. Further, it revealed a novel corollary discharge signal, possibly obscured in the past due to limited cortical coverage as well as averaging across regions and participants.

Theoretical models of speech production all assume a corollary discharge or a motor initiation signal (*37–39*). However, the exact source and timing of this signal is variable and unclear due to the lack of direct evidence in neural recordings. Evidence for corollary discharge has been limited to two major results in both human and non-human primates: suppression of auditory cortex during vocalization (*4, 13–15, 40–44*), and enhanced auditory responses during altered feedback (*7, 8, 12, 45, 46*). Our results showed a distinct source of information flow originating in ventral pre-central gyrus before articulation irrespective of task modality. This location and timing are consistent with the hypothesized communication across theoretical frameworks, as well as neuronal suppression which has been reported prior to vocalization (*4*). In stark contrast to prior studies which failed to link motor neural responses with auditory suppression (*8, 14, 16*), our results clearly establish that the target of information flow correlates

with the degree of suppression, a critical litmus test for corollary discharge.

Neurons supporting a corollary discharge circuit have been established in the cricket (*2*), songbird (*47*), and several other mammals (*1*). In humans, a corollary discharge signal is involved in speech production as well as the ability to distinguish between self- and external-generated thoughts and actions (*10, 48*). A dysfunctional corollary discharge circuit is the major model explaining auditory hallucinations in schizophrenia patients (*10, 49–53*). Auditory suppression is impaired in schizophrenia (*10, 11, 50*), but no direct corollary discharge signal source has been identified. Our novel framework and results provide a missing link in the human auditory system and elucidate the source, target and timing of the corollary discharge circuit with major implications for dysfunction of speech and psychosis.

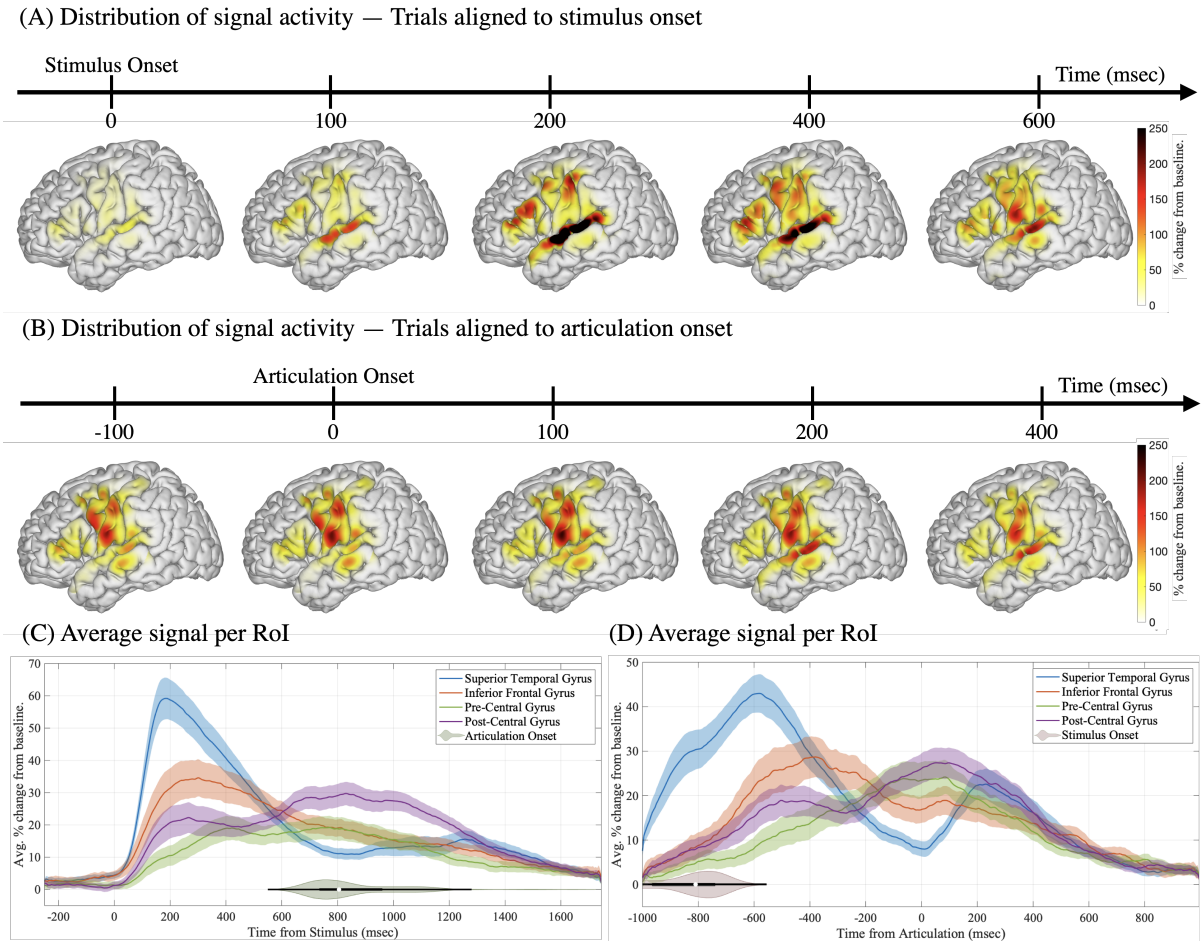

Figure 1: **Spatiotemporal distribution of neural activity during speech perception and production.**
The spatiotemporal distribution of neural activity compared to baseline when participants (N=8) listen to auditory stimuli (A) and subsequently repeat the same word (B). The color code represents percent change from pre-stimulus baseline. A region of interest approach averaging activity in superior temporal, inferior frontal, pre-central, and post-central gyri are shown when trials are aligned to (C) stimulus onset and (D) articulation onset. Shaded regions around the curves depict standard error of the mean across participants. Participants were instructed to repeat the auditory word freely when ready and the reaction time distribution is shown in a horizontal violin plot (C). The stimulus onset relative to articulation onset is shown in the horizontal violin plot in (D).

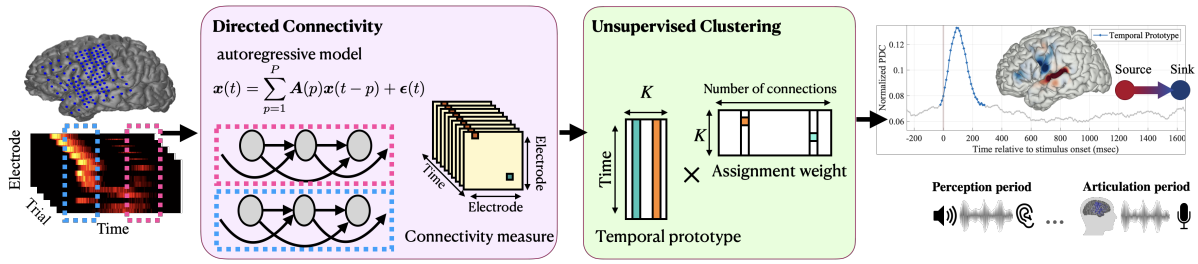

Figure 2: **Overview of the directed connectivity analysis framework.**
Electrocorticographic signals are recorded and high-gamma broadband is extracted locked to perception or production. Directed connectivity is measured across pairs of electrodes using an autoregressive model for successive overlapping time periods, providing the connectivity between electrodes as a function of time (3D matrix in purple box). The connectivity patterns are then represented by a few temporal prototypes via an unsupervised clustering technique (green box). This process reveals the major temporal connectivity patterns and their corresponding assignment weights projected on cortex (cortical sources in red and targets in blue).

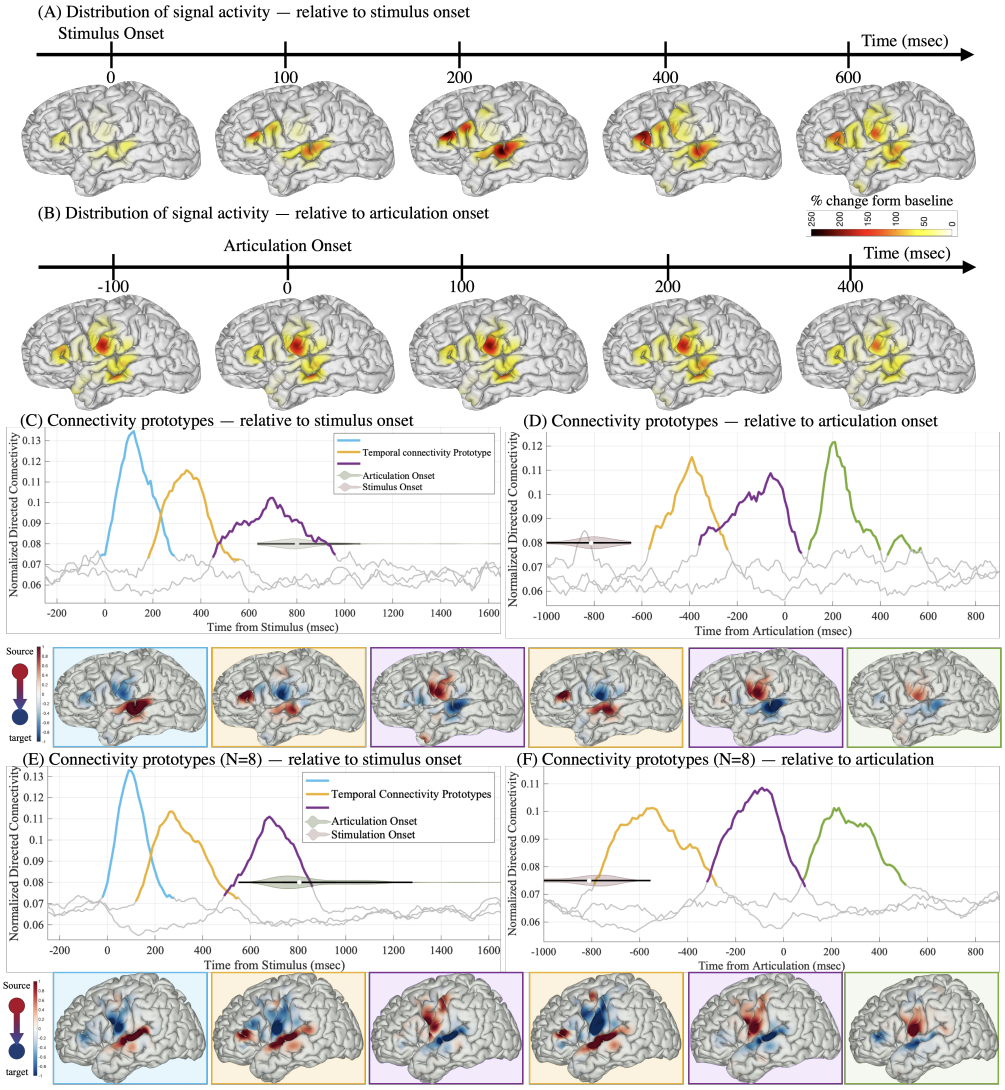

Figure 3: **Directed connectivity reveals sources and targets of neural communication during speech perception and production.**
The spatiotemporal distribution of neural activity compared to baseline in a representative participant is shown during (A) listening to auditory word stimuli and (B) subsequent word repetition. Results of directed connectivity modeling in this participant (temporal prototypes and corresponding information sources and targets) are shown locked to (C) stimulus and (D) articulation onset. Colored regions on cortex show the information source (red) and target (blue) for each prototype (colored box of the brain matches the color of the associated temporal curve in C and D). Colored segments of the curves indicate the portions of the temporal curves that are statistically significant compared to random permutation ($p < 0.05$, see section S.5 for details). (E,F) Directed connectivity analysis repeated for eight participants shows similar prototypical patterns and information sources and targets (visualized on a normal brain).

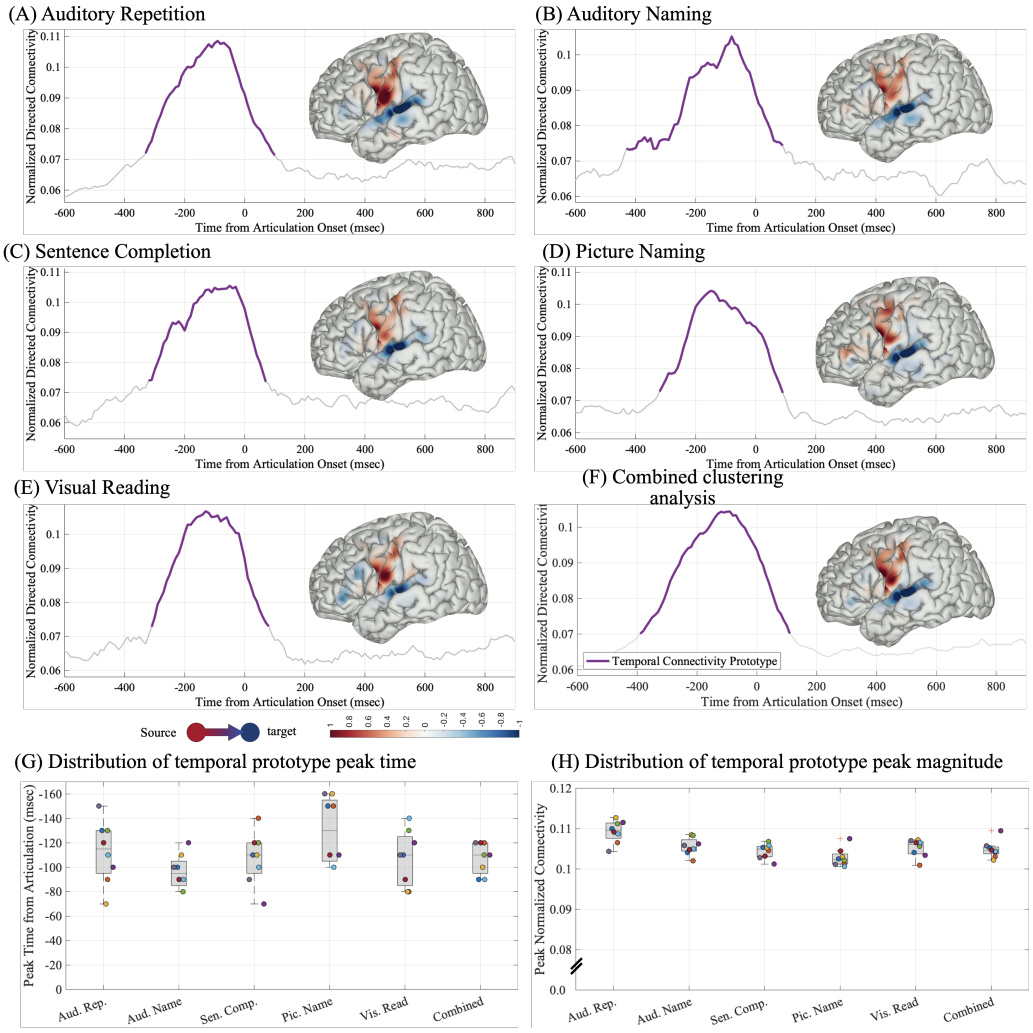

Figure 4: **Corollary discharge prototypes are replicated across tasks.**
The corollary discharge temporal prototype and its corresponding information source (red) and target (blue) are shown for all participants across five different speech production tasks (clustered separately) locked to articulation onset: (A) auditory repetition, (B) auditory naming, (C) sentence completion, (D) picture naming, and (E) visual word reading. (F) An overall representation of the corollary discharge prototype is shown by clustering the combined connectivity results from all tasks and participants. Colored regions of the curve represent statistically significant directed connectivity compared to random permutation ($p < 0.05$). The distribution across participants of the peak time (G) and magnitude (H) of the corollary discharge prototype is shown for each task (clustered for each participant separately; circles are color-coded per participant; Box-plots show minimum, maximum, quadrants, and median). The combined peak time and magnitude are established by clustering within participant but combining all the task data as shown in the last column of (G, H).

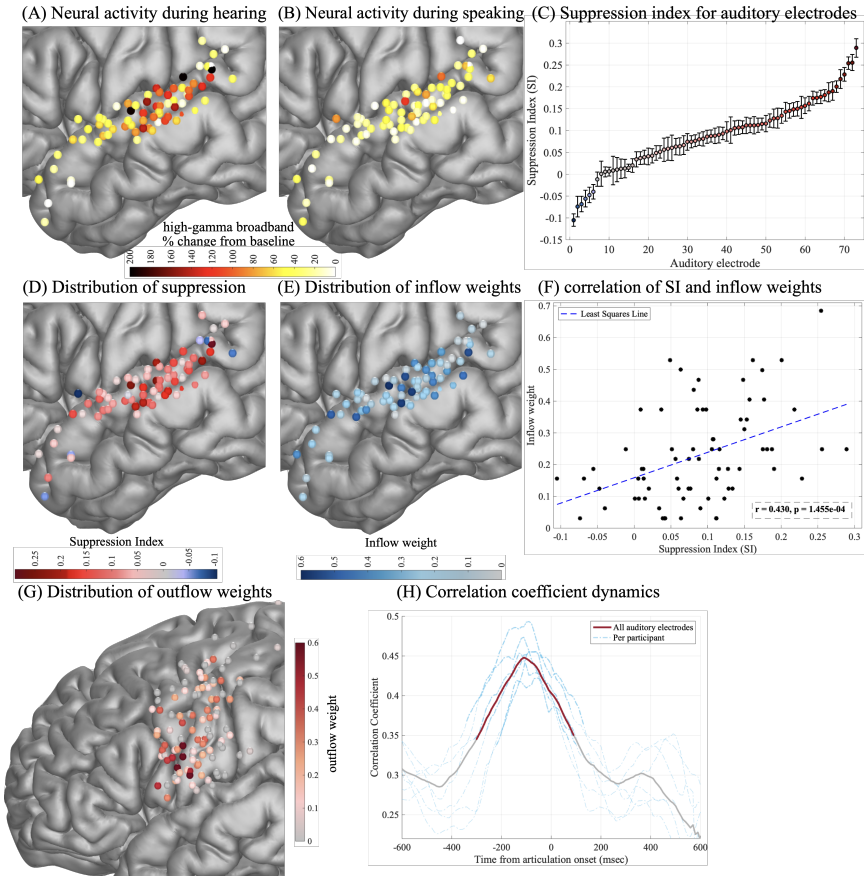

Figure 5: **Corollary discharge predicts speech-induced auditory suppression.** Average neural activity when participants (A) hear auditory stimuli and (B) speak during an auditory repetition task. (C) Most auditory responses are suppressed during speech production, quantified by a normalized suppression index (SI) shown for all active auditory electrodes (i.e. within participant anatomical label of Superior Temporal Gyrus) across all participants. Electrodes are sorted according to their mean SI over trials (error-bars represent SEM across trials) and vary from 1 (completely suppressed) to -1 (completely enhanced). Spatial distribution of mean SI for auditory electrodes across participants (D) and of the inflow weights onto auditory electrodes (E) in the corollary discharge prototype. (F) The correlation between mean SI and the inflow weights of the corollary discharge for all auditory electrodes. (G) Spatial distribution of the outflow weights (i.e. source) of the corollary discharge prototype in pre- and post-central gyri. (H) The dynamics of the correlation coefficient between mean SI and directed connectivity from sources in ventral motor cortex onto auditory cortex as a function of time (solid curve is obtained using all electrodes across participants and dashed lines show the analysis per participant; red shading denotes statistical significance from a permutation test $p < 0.05$).

References and Notes

1. T. B. Crapse, M. A. Sommer, Corollary discharge across the animal kingdom, *Nature Reviews Neuroscience* **9**, 587–600 (2008).

2. J. F. Poulet, B. Hedwig, The cellular basis of a corollary discharge, *Science* **311**, 518–522 (2006).

3. E. von Holst, The reafference principle, *The Behavioral Physiology of Animals and Man* (1973).

4. S. J. Eliades, X. Wang, Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations, *Journal of neurophysiology* **89**, 2194–2207 (2003).

5. S. Lee, I. Kruglikov, Z. J. Huang, G. Fishell, B. Rudy, A disinhibitory circuit mediates motor integration in the somatosensory cortex, *Nature neuroscience* **16**, 1662–1670 (2013).

6. J. E. Roy, K. E. Cullen, Dissociating self-generated from passively applied head motion: neural mechanisms in the vestibular nuclei, *Journal of Neuroscience* **24**, 2102–2111 (2004).

7. M. Ozker, W. Doyle, O. Devinsky, A. Flinker, A cortical network processes auditory error signals during human speech production to maintain fluency, *PLoS biology* **20**, e3001493 (2022).

8. E. F. Chang, C. A. Niziolek, R. T. Knight, S. S. Nagarajan, J. F. Houde, Human cortical sensorimotor network underlying feedback control of vocal pitch, *Proceedings of the National Academy of Sciences* **110**, 2653–2658 (2013).

9. D. Reznik, Y. Henkin, N. Schadel, R. Mukamel, Lateralized enhancement of auditory cortex activity and increased sensitivity to self-generated sounds, *Nature communications* **5**, 1–11 (2014).

10. R. Parlikar, A. Bose, G. Venkatasubramanian, Schizophrenia and corollary discharge: a neuroscientific overview and translational implications, *Clinical Psychopharmacology and Neuroscience* **17**, 170 (2019).

11. J. M. Ford, D. H. Mathalon, Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations?, *International Journal of Psychophysiology* **58**, 179–189 (2005).

12. S. J. Eliades, X. Wang, Neural substrates of vocalization feedback monitoring in primate auditory cortex, *Nature* **453**, 1102–1106 (2008).

13. J. F. Houde, S. S. Nagarajan, K. Sekihara, M. M. Merzenich, Modulation of the auditory cortex during speech: an meg study, *Journal of cognitive neuroscience* **14**, 1125–1138 (2002).

14. A. Flinker, *et al.*, Single-trial speech suppression of auditory cortex activity in humans, *Journal of Neuroscience* **30**, 16643–16650 (2010).

15. J. D. Greenlee, *et al.*, Human auditory cortical activation during self-vocalization, *PloS one* **6**, e14744 (2011).

16. V. L. Towle, *et al.*, Ecog gamma activity during a language task: differentiating expressive and receptive speech areas, *Brain* **131**, 2013–2027 (2008).

17. T. Paus, D. W. Perry, R. J. Zatorre, K. J. Worsley, A. C. Evans, Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges, *European Journal of Neuroscience* **8**, 2236–2246 (1996).

18. D. M. Schneider, A. Nelson, R. Mooney, A synaptic and circuit basis for corollary discharge in the auditory cortex, *Nature* **513**, 189–194 (2014).

19. T. Ito, L. Hearne, R. Mill, C. Cocuzza, M. W. Cole, Discovering the computational relevance of brain network organization, *Trends in cognitive sciences* **24**, 25–38 (2020).

20. D. S. Bassett, A. N. Khambhati, S. T. Grafton, Emerging frontiers of neuroengineering: a network science of brain connectivity, *Annual review of biomedical engineering* **19**, 327–352 (2017).

21. A. Flinker, *et al.*, Redefining the role of broca's area in speech, *Proceedings of the National Academy of Sciences* **112**, 2871–2875 (2015).

22. C. Wilke, L. Ding, B. He, Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function, *IEEE transactions on biomedical engineering* **55**, 2557–2564 (2008).

23. L. A. Baccalá, K. Sameshima, Partial directed coherence: a new concept in neural structure determination, *Biological cybernetics* **84**, 463–474 (2001).

24. F. Pompili, N. Gillis, P.-A. Absil, F. Glineur, Two algorithms for orthogonal nonnegative matrix factorization with application to clustering, *Neurocomputing* **141**, 15–25 (2014).

25. O. Creutzfeldt, G. Ojemann, E. Lettich, Neuronal activity in the human lateral temporal lobe, *Experimental brain research* **77**, 451–475 (1989).

26. T. A. Hackett, I. Stepniewska, J. H. Kaas, Prefrontal connections of the parabelt auditory cortex in macaque monkeys, *Brain research* **817**, 45–58 (1999).

27. A. Nelson, *et al.*, A circuit for motor cortical modulation of auditory cortical activity, *Journal of Neuroscience* **33**, 14342–14353 (2013).

28. A. K. Seth, A. B. Barrett, L. Barnett, Granger causality analysis in neuroscience and neuroimaging, *Journal of Neuroscience* **35**, 3293–3297 (2015).

29. A. Graef, *et al.*, A novel method for the identification of synchronization effects in multichannel ecog with an application to epilepsy, *Biological cybernetics* **107**, 321–335 (2013).

30. S. L. Bressler, A. K. Seth, Wiener–granger causality: a well established methodology, *Neuroimage* **58**, 323–329 (2011).

31. H. Lütkepohl, *New introduction to multiple time series analysis* (Springer Science & Business Media, 2005).

32. M. J. Kaminski, K. J. Blinowska, A new method of the description of the information flow in the brain structures, *Biological cybernetics* **65**, 203–210 (1991).

33. T. Zhang, J. Wu, F. Li, B. Caffo, D. Boatman-Reich, A dynamic directional model for effective brain connectivity using electrocorticographic (ecog) time series, *Journal of the American Statistical Association* **110**, 93–106 (2015).

34. A. Korzeniewska, C. M. Crainiceanu, R. Kuś, P. J. Franaszczuk, N. E. Crone, Dynamics of event-related causality in brain electrical activity, *Human brain mapping* **29**, 1170–1192 (2008).

35. A. G. Saravani, K. J. Forseth, N. Tandon, X. Pitkow, Dynamic brain interactions during picture naming, *eNeuro* **6** (2019).

36. M. L. Whaley, C. M. Kadipasaoglu, S. J. Cox, N. Tandon, Modulation of orthographic decoding by frontal cortex, *Journal of Neuroscience* **36**, 1173–1184 (2016).

37. F. H. Guenther, Cortical interactions underlying the production of speech sounds, *Journal of communication disorders* **39**, 350–365 (2006).

38. J. F. Houde, E. F. Chang, The cortical computations underlying feedback control in vocal production, *Current opinion in neurobiology* **33**, 174–181 (2015).

39. G. Hickok, Computational neuroanatomy of speech production, *Nature reviews neuroscience* **13**, 135–145 (2012).

40. S. J. Eliades, X. Wang, Dynamics of auditory–vocal interaction in monkey auditory cortex, *Cerebral Cortex* **15**, 1510–1523 (2005).

41. S. O. Aliu, J. F. Houde, S. S. Nagarajan, Motor-induced suppression of the auditory cortex, *Journal of cognitive neuroscience* **21**, 791–802 (2009).

42. R. Behroozmand, *et al.*, Sensory–motor networks involved in speech production and motor control: An fmri study, *Neuroimage* **109**, 418–428 (2015).

43. T. J. Whitford, Speaking-induced suppression of the auditory cortex in humans and its relevance to schizophrenia, *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **4**, 791–804 (2019).

44. P. Bäss, T. Jacobsen, E. Schröger, Suppression of the auditory n1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation, *International Journal of Psychophysiology* **70**, 137–143 (2008).

45. R. Behroozmand, C. R. Larson, Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback, *BMC neuroscience* **12**, 1–10 (2011).

46. T. H. Heinks-Maldonado, D. H. Mathalon, M. Gray, J. M. Ford, Fine-tuning of auditory cortex during speech production, *Psychophysiology* **42**, 180–190 (2005).

47. R. Mooney, J. F. Prather, The hvc microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways, *Journal of Neuroscience* **25**, 1952–1964 (2005).

48. I. Feinberg, Efference copy and corollary discharge: implications for thinking and its disorders., *Schizophrenia bulletin* **4**, 636 (1978).

49. C. Frith, Functional imaging and cognitive abnormalities., *The Lancet* **346**, 615–620 (1995).

50. T. J. Whitford, J. M. Ford, D. H. Mathalon, M. Kubicki, M. E. Shenton, Schizophrenia, myelination, and delayed corollary discharges: a hypothesis, *Schizophrenia bulletin* **38**, 486–494 (2012).

51. T. J. Whitford, M. Kubicki, M. E. Shenton, Diffusion tensor imaging, structural connectivity, and schizophrenia, *Schizophrenia research and treatment* **2011** (2011).

52. M. Spering, E. C. Dias, J. L. Sanchez, A. C. Schütz, D. C. Javitt, Efference copy failure during smooth pursuit eye movements in schizophrenia, *Journal of Neuroscience* **33**, 11779–11787 (2013).

53. S. S. Shergill, *et al.*, Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia, *JAMA psychiatry* **71**, 28–35 (2014).

54. A. Flinker, W. K. Doyle, A. D. Mehta, O. Devinsky, D. Poeppel, Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries, *Nature human behaviour* **3**, 393–405 (2019).

55. M. Zanetti, F. Bovolo, L. Bruzzone, Rayleigh-Rice mixture parameter estimation via em algorithm for change detection in multispectral images, *IEEE Transactions on Image Processing* **24**, 5004–5016 (2015).

56. L. Faes, S. Erla, G. Nollo, Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis, *Computational and mathematical methods in medicine* **2012** (2012).

57. M. Ding, S. L. Bressler, W. Yang, H. Liang, Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment, *Biological cybernetics* **83**, 35–45 (2000).

58. H. Lütkepohl, *Introduction to multiple time series analysis* (Springer Science & Business Media, 2013).

Acknowledgements

The authors would like to thank Dr. David Schneider for his valuable comments on this manuscript. **Funding:** This work is supported by National Science Foundation IIS- 1912286 and the National Institute of Health NINDS 1R01NS109367 and 1R01NS115929 grants. **Author contributions:** A.K-G. proposed and implemented the directed connectivity analysis framework and the unsupervised clustering algorithm with advisement from Y.W. and A.F.; R.W. and X.C participated in the data processing; L.Y. participated in data acquisition and preprocessing; P.D. and D.F. provided clinical care; W.D. provided neurosurgical clinical care; O.D. assisted in patient care and consent; Y.W. co-led the research project with A.F. and advised from engineering perspective; A.F. led the project, participated in all data acquisition, and advised from neuroscience perspective; A.K-G. and A.F. co-wrote the manuscript with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability** The dataset analysed during the current study is available from the corresponding author upon reasonable request. The code for processing and analysis as well as the directed connectivity analysis framework is available from the corresponding author upon request.

Supplementary Materials for

A Corollary Discharge Circuit in Human Speech

Amirhossein Khalilian-Gourtani, Ran Wang, Xupeng Chen, Leyao Yu, Patricia Dugan, Daniel Friedman, Werner Doyle, Orrin Devinsky, Yao Wang, Adeen Flinker.

Correspondence to: amirhossein.khalilian@nyu.edu, adeen@nyu.edu

S Materials and Methods

S.1 Participant information

A total of 8 neurosurgical patients (7 female, all with left hemisphere coverage, mean age: 38 with range 19 to 55 years) were implanted with electrocorticography electrodes and provided informed consent to participate in this research. All consent was obtained in writing and then requested again orally prior to the beginning of the experiment. Electrode implantation and location were guided solely by clinical requirements. Five of the participants were implanted with standard clinical electrode grid with 10 mm spaced electrodes (Ad-Tech Medical Instrument, Racine, WI). Three participants consented to a research hybrid grid implant (PMT corporation, Chanassen, MN) that included 64 additional electrodes between the standard clinical contacts (with overall 10 mm spacing and interspersed 5 mm spaced electrodes over select regions). This provided denser sampling of underlying cortex but was positioned solely based on clinical needs. The superior temporal gyrus (STG) region is sampled for all participants, and other cortical regions (including Broca's area and motor cortex) are also sampled. The study protocol was approved by the NYU Langone Medical Center Committee on Human Research.

S.2 Experiment setup

The participants were instructed to complete five tasks to pronounce the target words in response to certain auditory or visual stimuli. These five tasks consisted of: auditory repetition (i.e., repeat an auditory presented word), auditory naming (i.e., naming a word based on an auditory presented description), sentence completion (i.e., complete the last word of an auditory presented sentence), visual word reading (i.e., read out loud a visually presented written words), and picture naming (i.e., name a word based on a visually presented line drawing). All tasks consisted of the same 50 target words. For each task and trial the stimulus was randomly presented and participants produced the target word freely after which the next trial started. Each

target item appeared twice in the auditory repetition, picture naming, and visual word reading tasks and once in auditory naming and sentence completion.

S.3 Data collection and pre-processing

We rejected electrodes with epileptiform activity, line-noise artifacts, poor contact, and high amplitude-shifts. We calculated a common average reference by averaging the signal across all electrodes and subtracted from each individual electrode signal. Continuous data stream was divided into epochs based on the onset of stimulus (lock to stimulus) or onset of speech (locked to articulation). ECoG signals were preprocessed with high gamma band-pass filter (70-150 Hz). The envelope of the filtered signal was then extracted by a Hilbert transform. All signals are downsampled to 200 Hz after envelope extraction. We Z-score the signal from each electrode by the mean and standard deviation of the trials for connectivity analyses. We use units of percent change from baseline to quantify average neural activity.

S.4 Neural activity analysis

Neural activity (mainly shown in figures 1, 3A-B, and 5A-B) were measured by preprocessing the ECoG signals with high-gamma band-pass filter (70-150 Hz), extracting the envelope of the filtered signal using a Hilbert transform, and comparing to mean baseline activity 250 msec prior to stimulus onset. In Fig. 1, A and B, the broadband high-gamma signal (measured in percent change from baseline) for each electrode is averaged over participants, trials and time in a 50 msec window centered at each marked time-stamp and projected onto a normal brain with a Gaussian kernel of size 50 mm (we restrict the spread of the Gaussian kernel to within the boundaries of the associated cortical region for each electrode). Similarly, in Fig. 3, A and B, the broadband high-gamma signal is projected onto the Cortical surface model of the participant's brain. In Fig. 1, C and D, the broadband high-gamma signal is averaged over electrodes

within a region of interest and trials for each time-point and participant. We show mean and standard error of the mean across participants for each region of interest. The distribution of articulation onset relative to stimulus onset (Fig. 1C, Fig. 3C and E) and stimulus onset relative to articulation onset (Fig. 1D, Fig. 3 D and F) over trials and participants are shown as horizontal violin plots.

S.5 Directed connectivity analysis framework

For each participant, we detect a subset of electrodes that are active during the task (see supplementary text S.7 for details; see Fig. S5 for an example and Fig. S6 for electrode coverage details). For each patient and task, we use the trial information to fit a multivariate autoregressive (MVAR) model to 100 msec overlapping time windows with hops of 10 msec (see supplementary text S.8). For each window we measure the information flow (in Granger-causal sense) between pairs of electrodes by computing the partial directed coherence (PDC) (*23*) from MVAR model coefficients. The resulting connectivity measures show the information flow between any two given electrodes as a function of time. To derive the prototypical temporal connectivity patterns, we use orthogonal non-negative matrix factorization (ONMF) (*24*) as an unsupervised clustering technique. We gather the measured directed connectivity signals into a matrix (the dimensions of which are number of time-windows by number of electrode pairs) and apply the factorization, representing each column of the directed connectivity matrix as a scaled version of one of a few prototypical patterns. Consequently, our factorization reveals the temporal connectivity prototypes and the assignment of each electrode pair to one of the prototypes with a corresponding assignment weight (see supplementary text S.10, and Fig. S7 for an overview). The assignment weights associated with each particular prototype are then visualized on participant or Montreal Neurological Institute (MNI) brain with a Gaussian kernel of size 50 mm (restricting the spread to within the boundaries of the associated cortical

region for each electrode). We assess the statistical significance of prototypes with a permutation test (*54*), where we randomize the cluster assignments and recompute the prototypes for 1000 repetitions. Each time-point of the prototype is tested against the randomized projection distribution (with an alpha criterion of 0.05) and to reduce multiple comparisons error only continuous range of values (longer than 100 msec) showing statistical significance is accepted. The number of clusters (corresponding to total number prototypes) were determined based on the relative reconstruction error for different number of clusters. We empirically found that this error term plateaued at four components for the time windows considered in our experiments (see supplementary text S.10; see Fig. S8). One of the resulting prototypes always showed "noise-like" activity and was removed from analysis (see supplementary text and Fig. S2).

S.6 Speech induced auditory suppression

To quantify the level of suppression for each electrode anatomically located in the auditory cortex, we compute the average broadband high-gamma power in 300 msec after the stimulus onset (P_{hear}) and 300 msec after articulation onset (P_{speech}) in the auditory repetition task. For each auditory electrode (total of 73 across all participants) we compute the suppression index (SI) defined as $\text{SI} = \frac{P_{\text{hear}} - P_{\text{speech}}}{P_{\text{hear}} + P_{\text{speech}}}$ (varying between 1 for completely suppressed and -1 for completely enhanced; distribution shown in Fig. 5, C and D). We quantify the information inflow to each auditory electrode by the cluster assignment weight associated with the corollary discharge prototype from auditory repetition task (distribution shown in Fig. 5E). To assess the link between suppression and corollary discharge we use Pearson correlation across all auditory electrodes for all participants (Fig. 5F). To show the temporal dynamics of this correlation, for each auditory electrode we compute the average directed connectivity of the connections represented by the corollary discharge prototype sourcing from electrodes located anatomically in ventral pre-central gyrus in overlapping windows of 300 msec with hops of 10 msec. We then

correlate the suppression index and the average directed connectivity for each time-window across all auditory electrodes (solid curve in Fig. 5H). We repeat the same analysis only considering the auditory electrodes for each participant (dashed curves in Fig. 5H; two participants were removed from this analysis due to limited number of auditory electrodes). We assess the statistical significance by a permutation test (randomizing the suppression index assignment for each window a total of 1000 times and comparing each time point to this distribution; alpha of 0.05; multiple comparisons error corrected with only accepting continuous range longer than 100 msec).

Supplementary Text

In the following sections, we introduce the proposed analysis framework including selection of active electrodes, the Multivariate Autoregressive (MVAR) model, and the partial directed coherence, which is a measure of Granger causality based on the Autoregressive (AR) coefficients. We further show how clustering techniques can be utilized to analyze the temporal variation patterns of the resulting connectivity measures. An overview of the pipeline is shown in Fig. S1.

S.7 Active electrode selection

As it is common in the literature, not all the electrodes are used when fitting an MVAR model due to the fact that unrelated noisy electrodes can reduce the goodness-of-fit for the MVAR model and inclusion of more electrodes require more trials to fit. While selecting a subset of electrodes is a common practice, other studies do not report an automatic and robust criterion for electrode selection. It is desired to introduce an unsupervised automatic algorithm to select a subset of electrodes that can be considered active given the ECoG data for a task.

We denote by $x_n(r,t) \in \mathbb{R}$ the ECoG signal for electrode $n \in \{1, \cdots, N\}$ at trial $r \in$

$\{1, \cdots, R\}$ and time-step $t \in \{1, \cdots, T\}$. We aim to find a subset of electrodes which have activity related to the task. Such electrodes should have similar time series $x_n(r, t)$ over different trials whereas the electrodes in regions irrelevant to the task are likely to capture randomly fluctuating signals over different trials. Therefore, the mean signal from a relevant electrode for a task should have relatively large variation over time, whereas an irrelevant electrode tends to have a mean signal close to zero and has small temporal variation.

Motivated by this rationale, we first determine the mean signal for each electrode using $\bar{x}_n(t) = \frac{1}{R}\sum_r x_n(r, t)$. We empirically observed that further denoising this signal via wavelet-thresholding is beneficial. Let W denote the forward wavelet transform, W^T denote its inverse, and $\bar{\boldsymbol{x}}_n = [\bar{x}_n(1), \cdots, \bar{x}_n(T)]^\top$. Let $\mathcal{H}_\tau(x) = x$ for $|x| \geq \tau$ and 0 otherwise, be the hard-thresholding operator and similarly extend for vectors by applying element-wise. Then, the denoised mean signal can be represented by $\widehat{\boldsymbol{x}}_n = W^T\mathcal{H}_\tau\left(W\bar{\boldsymbol{x}}_n\right)$. We used 5 levels of *db8* wavelet transform and we set $\tau = 0.5$.

Examples of the signals $x_n(r, t)$, $\bar{x}_n(t)$, and $\widehat{x}_n(t)$ for three different electrodes are shown in Fig. S5(a). We can see that the changes of the signal $\widehat{x}_n(t)$ over time is a good measure of electrode activity. We compute the standard deviation of $\widehat{x}_n(t)$ over time by $s_n = \sqrt{\frac{1}{T-1}\sum_t(\widehat{x}_n(t) - \overline{\widehat{x}_n(t)})^2}$. Higher values of s_n indicate active electrodes and smaller values indicate the inactive ones. We want to find a threshold τ_s such that the electrode n is considered active only if $s_n > \tau_s$. A sample histogram of s_n for all $n \in \{1, \cdots, N\}$ is shown in Fig. S5(b). We find that the histogram can be modeled by a Rayleigh-Rice mixture model, where the Rayleigh component represents the noise, while the Rice component represent the signal. We set the threshold τ_s to be where the two components are equally likely. Let α indicate the probability that a sample s is from a Rayleigh distribution with parameter b^2, and $1 - \alpha$ the probability that s is from the Rice distribution with parameters ν and σ^2. Then, the distribution

of the mixture model can be written as

$$f(s|\Psi) = \alpha f_{\text{Rayl}}(s|b^2) + (1-\alpha) f_{\text{Rice}}(s|\nu, \sigma^2)$$

where $\Psi = \{\alpha, b^2, \nu, \sigma^2\}$ and

$$f_{\text{Rayl}}(s|b^2) = \frac{s}{b^2} \exp\left(-\frac{s^2}{2b^2}\right), \quad s \geq 0$$

and

$$f_{\text{Rice}}(s|\nu, \sigma^2) = \frac{s}{\sigma^2} \exp\left(-\frac{s^2+\nu^2}{2\sigma^2}\right) I_0\left(\frac{s\nu}{\sigma^2}\right), \quad s \geq 0$$

with $I_0(\cdot)$ is the 0-th order modified Bessel function of first kind. Given the samples $\{s_n\}_{n=1}^N$ we want to find the parameters of this model ($\Psi = \{\alpha, b^2, \nu, \sigma^2\}$). Expectation-Maximization (EM) algorithm can be used to estimate the model parameters Ψ (*55*). Given the fitted model parameters the threshold value τ_s is found by solving

$$\frac{1-\alpha}{\alpha} = \frac{f_{\text{Rayl}}(\tau_s|b^2)}{f_{\text{Rice}}(\tau_s|\nu, \sigma^2)}.$$

S.8 Multivariate autoregressive model

We are interested in the Granger causal relationship between different electrodes. We say signal x_i Granger causes the signal x_j if knowledge of $x_i(\tau)$ for $\tau \leq t$ improves the prediction of $x_j(t)$. In order to assess Granger causality between multichannel ECoG data, a multivariate autoregressive (MVAR) model is fitted. Let $\boldsymbol{x}(t) = [x_1(t), \cdots, x_M(t)]^T \in \mathbb{R}^M$ be the multichannel ECoG signal at time t where M indicates the total number of selected active electrodes (M total channels). The MVAR model assumes that the signal at each time point, $\boldsymbol{x}(t)$ can be estimated as a linear combination of the signal at previous time-points and a random innovation signal $\boldsymbol{\epsilon}(t)$. Consequently, we can write the MVAR formulation as

$$\boldsymbol{x}(t) = \sum_{p=1}^{P} \boldsymbol{A}(p)\boldsymbol{x}(t-p) + \boldsymbol{\epsilon}(t), \tag{1}$$

where $\boldsymbol{A}(p) \in \mathbb{R}^{M \times M}$ are coefficient matrices for which the element $a_{ij}(p)$ shows the dependency of $x_i(t)$ on $x_j(t-p)$ for electrodes $i, j = 1, \cdots, M$ and time-lags $p = 1, \cdots, P$. The random innovation signal $\boldsymbol{\epsilon}(t)$ is assumed to be composed of white uncorrelated noises with covariance matrix $\boldsymbol{Q}$. Causality relations are found when the relevant interaction is active in 1. More formally, we can say x_j Granger causes x_i if $a_{ij}(p) \neq 0$ at least for one $p \in \{1, \cdots, P\}$. This is consistent with the definition of direct causality as in (*56*). The parameters of the MVAR model in (1), $\Theta = \{\boldsymbol{A}(p), \boldsymbol{Q}\}$, are estimated using the Levinson, Wiggins, Robinson (LWR) algorithm (*57*).

After a model is fitted validation steps are required to guarantee that the obtained model satisfies the assumptions. For the estimated MVAR model to be stable, the roots of the characteristic polynomial

$$\det\left(\alpha^P \boldsymbol{I} + \alpha^{P-1}\boldsymbol{A}(1) + \cdots + \alpha \boldsymbol{A}(P-1) + \boldsymbol{A}(P)\right) = 0$$

must satisfy $|\alpha| < 1$. Another important note here is that the resulting MVAR model should have temporally uncorrelated residual $\boldsymbol{\epsilon}(t)$. This property can be checked by utilizing the Ljung-Box portmanteau test for whiteness and the Kendall's τ test for independence (*58*). We follow the recommendations in (*34*) when choosing the model order (based on AIC/BIC criterion), and rejecting correlated residuals.

Note that the AR process is a linear model and has the inherent assumption that the signal x can be modeled with a stationary process. Ding et al. propose to model short windows of signal with separate AR models to overcome this limitation (*57*). Similarly, we use short overlapping windows of 100 msec with hops of 10 msec and use the available trial data to fit an MVAR model for each window.

S.9 Measuring directed connectivity from model coefficients

We use the notion of the partial directed coherence (PDC) which is a frequency-domain approach to describing the relationships (direction of information flow) between multivariate time series (*23*). The PDC factor is defined as

$$\gamma_{ij}(f) = \frac{\bar{\boldsymbol{A}}_{ij}(f)}{\sqrt{\bar{\boldsymbol{A}}_{\cdot j}^{H}(f)\bar{\boldsymbol{A}}_{\cdot j}(f)}} \quad \text{where} \tag{2}$$

$$\bar{\boldsymbol{A}}_{ij}(f) = \begin{cases} 1 - \sum_{\tau=1}^{p} \boldsymbol{A}_{ij}(\tau)\exp\left(-2\pi i f \tau\right), & \text{if} \quad i = j \\ -\sum_{\tau=1}^{p} \boldsymbol{A}_{ij}(\tau)\exp\left(-2\pi i f \tau\right) & \text{otherwise.} \end{cases}$$

Some properties of PDC are $0 \leq |\gamma_{ij}(f)|^2 \leq 1$ and $\sum_{i=1}^{M} |\gamma_{ij}(f)|^2 = 1 \quad \forall 1 \leq j \leq M$ (normalized interaction strengths with respect to a given signal source). Note that the PDC measures the Granger causality from MVAR coefficients. When dealing with multiple windows of signal, we calculate the PDC for each window. To obtain a picture of how the connectivity changes across time-windows, we follow (*35*) and define $\gamma_{ij} = \sum_f |\gamma_{ij}(f)|$ (summing the PDC values over frequency for each window) and focus on the temporal changes of γ_{ij} over the shifted windows.

S.10 Clustering Algorithm for Directed Connectivity Temporal Patterns

For each time window $t \in \{1, \cdots, T\}$, let $\gamma_{i,j}^{t}$ be the directed connectivity, or PDC, from node $i \in \{1, \cdots, M\}$ to node $j \in \{1, \cdots, M\}$. Define the connectivity tensor $\boldsymbol{\Gamma} \in \mathbb{R}^{T \times M \times M}$, where $[\boldsymbol{\Gamma}]_{t,i,j} = \gamma_{i,j}^{t}$. To investigate the temporal pattern of the connectivity between two nodes, we reshape the connectivity tensor into a matrix such that each column represents the temporal connectivity profile between a particular pair of nodes, denoted by $\tilde{\boldsymbol{\Gamma}}_{l(i,j)} = [\gamma_{i,j}^{t}, t = 1, 2, \ldots, T]^{\top}$, where $l(i,j)$ is the index for the connection from node i to node j with $i \neq j$. We do not consider self-connection, and hence $l \in \{1, 2, \ldots, M(M-1)\}$. Define $\tilde{\boldsymbol{\Gamma}} = [\boldsymbol{\Gamma}_{l(i,j)} | l \in \{1, \cdots, M(M-1)\}] \in \mathbb{R}^{T \times (M^2 - M)}$. We want to cluster similar temporal profiles (i.e. similar

columns in $\tilde{\boldsymbol{\Gamma}}$) into the same group even if they may differ in amplitude. Given that the connectivity values are non-negative, this can be formulated as the orthogonal non-negative matrix factorization (ONMF) problem. A pictorial representation of ONMF is shown in Fig. S7.

Given the non-negative data matrix $\tilde{\boldsymbol{\Gamma}} \in \mathbb{R}_{\geq 0}^{T\times(M^2-M)}$ and a desired number of clusters K, the ONMF objective function can be written as,

$$\underset{\boldsymbol{U}\in\mathbb{R}^{T\times K},\ \boldsymbol{V}\in\mathbb{R}^{K\times(M^2-M)}}{\text{minimize}} \|\tilde{\boldsymbol{\Gamma}} - \boldsymbol{U}\boldsymbol{V}\|_F^2 \text{ subject to: } \boldsymbol{U} \geq 0,\ \boldsymbol{V} \geq 0,\ \boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}_K. \tag{3}$$

The constraints $\boldsymbol{V} \geq 0$ and $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}_k$ imply that each column of $\boldsymbol{V}$ can have at most one non-zero entry, hence the clustering nature of this objective function. Solving the clustering problem means finding a set of K disjoint clusters, $\pi_k, k = 1, 2, \ldots, K$, such that each cluster π_k contains columns in $\tilde{\boldsymbol{\Gamma}}$ that are as similar to each other as possible. The clusters should satisfy $\pi_k \subseteq \{1, 2, \ldots, M^2-M\}\ \forall k, \quad \cup_{1\leq k\leq K}\pi_k = \{1, 2, \ldots, M^2-M\}, \quad$ and $\quad \pi_k \cap \pi_l = \emptyset\ \forall k \neq l$. Cluster k is represented by its corresponding column $\boldsymbol{u}_k$ in $\boldsymbol{U}$, which we refer to as the k-th temporal connectivity prototype. The EM-ONMF algorithm (*24*) solves for the matrices $\boldsymbol{V}$ and $\boldsymbol{U}$ given the original connectivity matrix $\tilde{\boldsymbol{\Gamma}}$ and K.

If the temporal connectivity profile from node n to node m is clustered to cluster k, then it can be approximated by $\boldsymbol{\Gamma}_{l(n,m)} \approx v^k_{l(n,m)}\boldsymbol{u}_k$. We call the weight v^k_{nm} the assignment weight for prototype k from n to m. We define the inflow weight for each electrode m in cluster π_k as $\iota_{(m,\pi_k)} = \frac{\sum_n v^k_{l(n,m)}}{|\pi_k|}$, where $|\pi_k|$ indicates the number of connections in cluster π_k. Similarly we define the outflow weight for each electrode n as $o_{(n,\pi_k)} = \frac{\sum_m v^k_{l(n,m)}}{|\pi_k|}$. To visualize the source and target locations associated with each prototype, we defuse the inflow (negatively weighted; blue color) and outflow (positively weighted; red color) weight associated with each electrode into the surrounding area in the same anatomical region using a Gaussian kernel on a patient or standard brain map.

The number of clusters K (corresponding to total number prototypes) were determined

based on the relative error $\frac{\|\tilde{\mathbf{\Gamma}}-\boldsymbol{U}\boldsymbol{V}\|_F^2}{\|\tilde{\mathbf{\Gamma}}\|_F^2}$ for different K. We empirically found that this error term plateaued at $K = 4$ for the time windows considered in our experiments (see Fig. S8). One of the resulting prototypes always showed "noise-like" activity and was removed from analysis (see Fig. S2).

S.11 Supplemental Figures

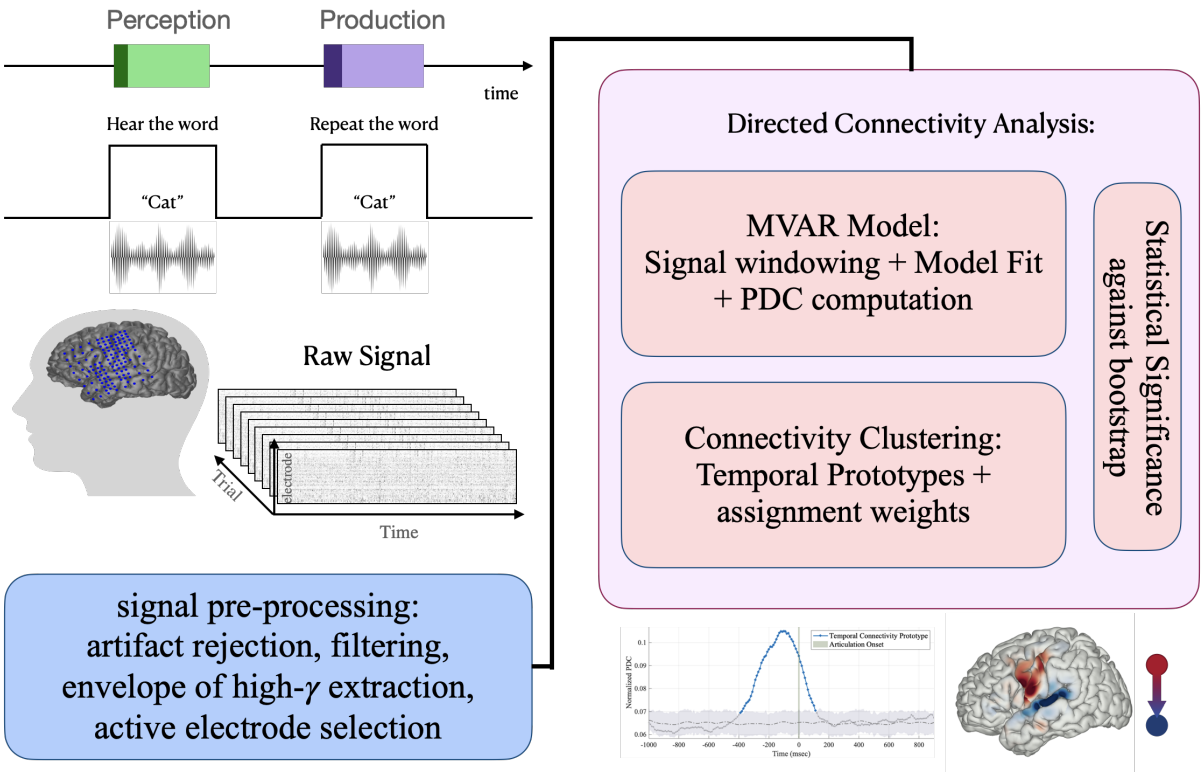

Figure S1: **Detailed overview of the proposed directed connectivity analysis framework.** Electrocorticographic signals are recorded from neurosurgical patients. We focus on the amplitude of high-gamma broadband locked to perception or production onset. Directed connectivity is measured across pairs of electrodes using an autoregressive model for successive overlapping time windows. The connectivity patterns are then represented by a few temporal prototypes via an unsupervised clustering technique. This process reveals the major temporal connectivity patterns and their corresponding assignment weights projected on cortex. We test statistical significance of each prototype against random permutation ($p < 0.05$).

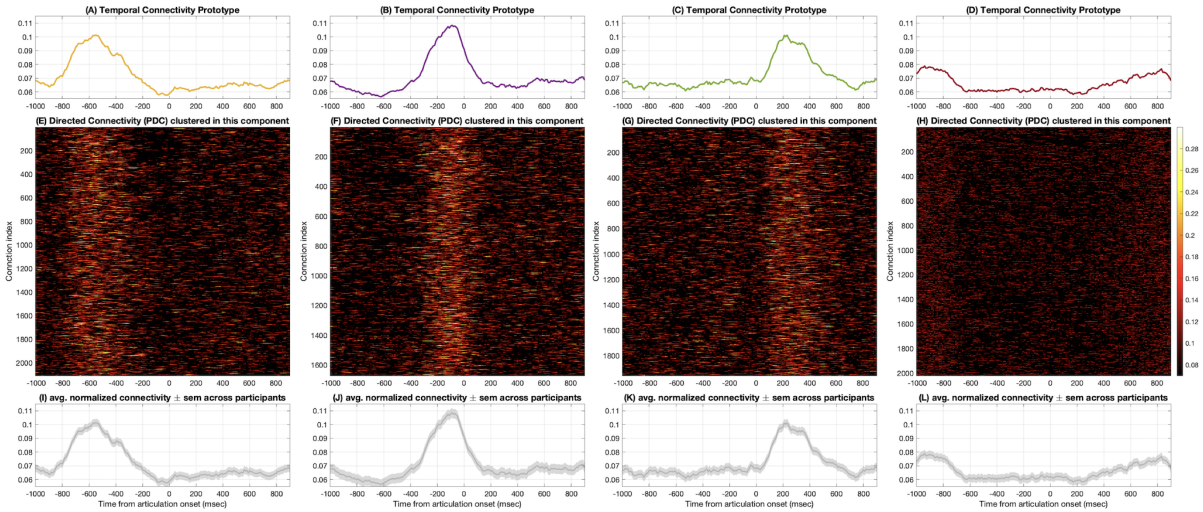

Figure S2: **Temporal connectivity profiles in auditory repetition locked to articulation** (A-D) Resulting temporal connectivity prototypes when directed connectivity profiles from auditory repetition task locked to articulation are clustered to 4 clusters for eight participants together. (E-H) Pictorial representation of temporal connectivity profiles, $\tilde{\mathbf{\Gamma}}_{\pi_k}$, assigned into cluster k, represented by the corresponding prototype in A to D. Each row in matrices E to H shows the temporal dynamics of a specific connection between two electrodes in each cluster. The rows of the matrices are ordered by participants. (I-L) The average normalized connectivity and corresponding standard error of the mean across participants (shaded region) shows robustness of the corresponding prototypes.

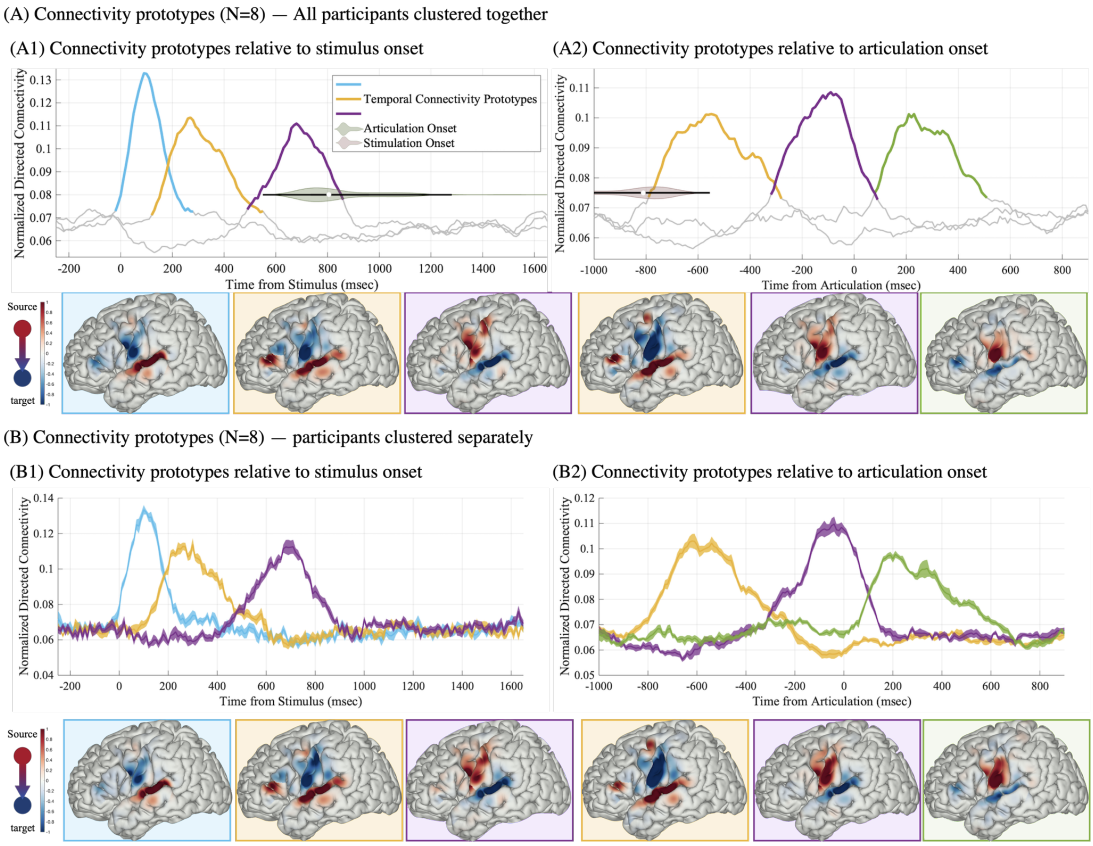

Figure S3: **Connectivity prototypes in auditory repetition; variability across participants.** (A) The connectivity prototypes and their corresponding source and target distributions when directed connectivity profiles from N=8 participants are clustered together (repeated from Fig. 3 E and F); (B) The mean of the connectivity prototypes and corresponding average source and target distribution when directed connectivity profiles from N=8 participants are clustered separately. Shaded area around the curves in B shows standard error of the mean across participants.

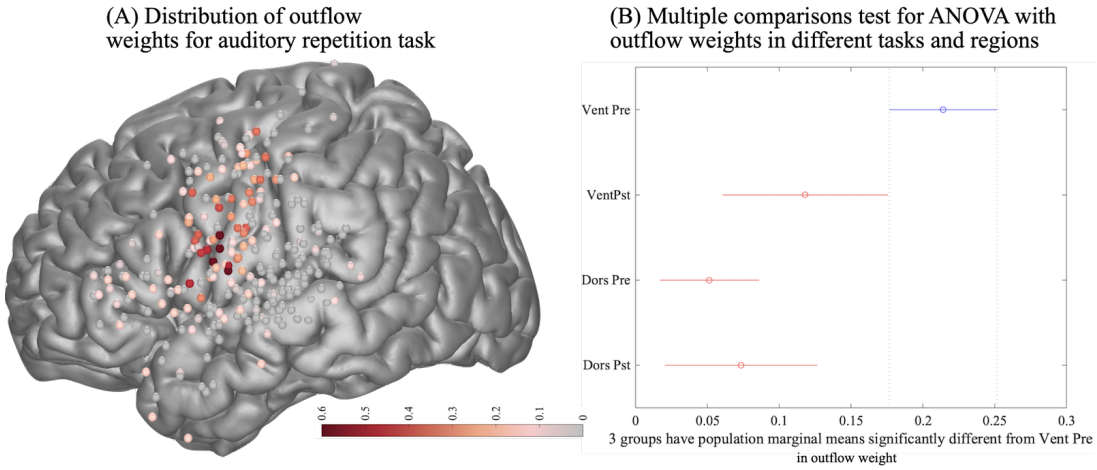

Figure S4: **Corollary discharge source.**
(A) Spatial distribution of outflow weights associated with the corollary discharge prototype for the auditory repetition task. (B) Post-hoc multiple comparisons test using the result of ANOVA test for outflow weights with tasks and anatomical region (vental/dorsal pre- and post-central areas) as effects shows ventral pre-central gyrus as the main source of outflow associated with the corollary discharge prototype.

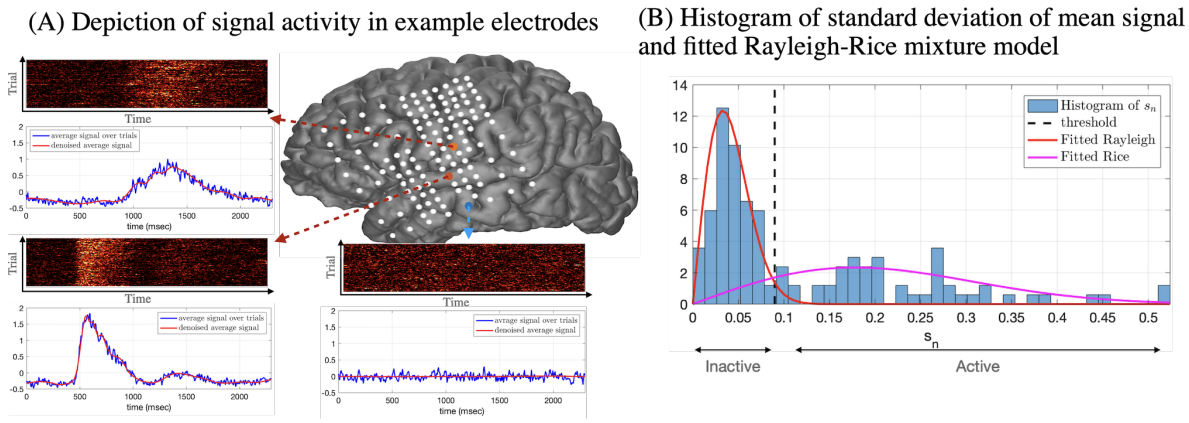

Figure S5: **Example for active electrode selection algorithm.**
(A) Pictorial depiction of the signal $x_n(r,t)$, average signal over trials $\bar{x}_n(t)$, and denoised average signal $\widehat{x}_n(t)$ for two active and one inactive electrode. (B) Histogram of temporal changes from the mean, s_n, for all the electrodes of one patient during auditory repetition task. Fitted Rayleigh-Rice mixture model and threshold are shown.

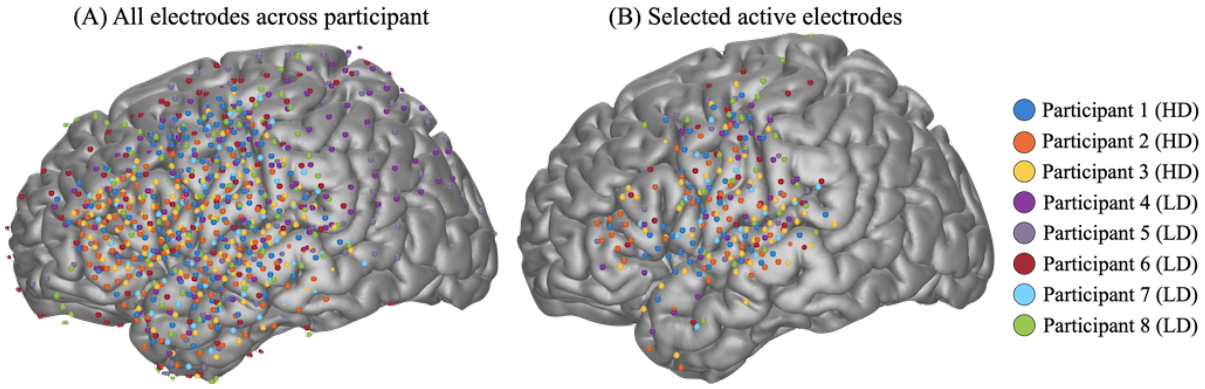

Figure S6: **Electrode coverage and selection.**
(A) Implanted electrodes for all eight participants shown on a normal brain (MNI space, color-code represents participants). Five participants were implanted with low-density (LD) grids with 10mm spacing while three participants consented to be implanted with a hybrid-density (HD) grid with 10 mm overall spacing and 5mm spacing in specific areas. (B) Resulting electrodes from the active electrode selection algorithm are shown.

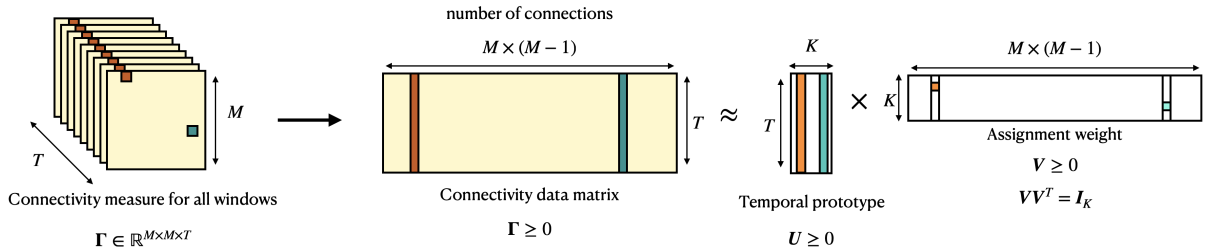

Figure S7: **Connectivity clustering algorithm.**
A pictorial representation of the temporal connectivity prototypes and their associated connections viewed as a matrix factorization problem. Note that only one element in each column in $\mathbf{V}$ can be non-zero, indicating the prototype that this column is clustered to.

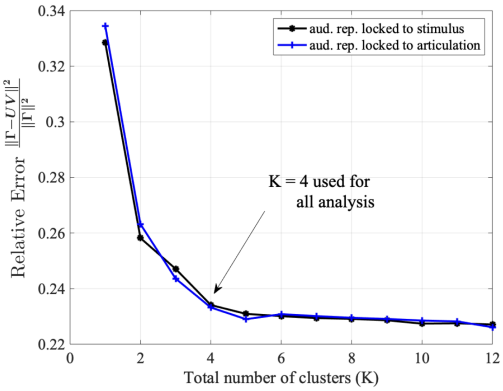

Figure S8: **Number of clusters.**
Relative recovery error, $\frac{\|\tilde{\boldsymbol{\Gamma}} - \boldsymbol{UV}\|_F^2}{\|\tilde{\boldsymbol{\Gamma}}\|_F^2}$, computed for different number of cluster components, K, and shown for auditory repetition task locked to stimulus (black) and articulation (blue). We choose $K = 4$ components in our analysis.

S.12 Supplemental Tables

Table S1: Start time of consecutive activity (above 1.96 SD of baseline; longer than 100 msec) relative to stimulus onset across electrodes in different regions of interest during the auditory repetition task (mean $\pm$ standard deviation across electrodes).

RoI	STG	IFG	Pre-central	Post-central
Time from stimulus (msec)	105 ± 37	181 ± 42	435 ± 114	845 ± 53