


5-2021

## Cortical Dynamics Of Language

Kiefer Forseth

Follow this and additional works at: [https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations](https://digitalcommons.library.tmc.edu/utgsbs_dissertations)

 Part of the [Cognitive Neuroscience Commons](#), [Computational Neuroscience Commons](#), [Dynamical Systems Commons](#), [Medicine and Health Sciences Commons](#), [Signal Processing Commons](#), and the [Systems Neuroscience Commons](#)

---

### Recommended Citation

Forseth, Kiefer, "Cortical Dynamics Of Language" (2021). *Dissertations and Theses (Open Access)*. 1096.  
[https://digitalcommons.library.tmc.edu/utgsbs\\_dissertations/1096](https://digitalcommons.library.tmc.edu/utgsbs_dissertations/1096)

This Dissertation (PhD) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact [digcommons@library.tmc.edu](mailto:digcommons@library.tmc.edu).

CORTICAL DYNAMICS OF LANGUAGE

by

Kiefer James Forseth, B.S., B.Mus.

APPROVED:

---

Nitin Tandon, M.D.  
Advisory Professor

---

Edgar Walters, Ph.D.

---

Xaq Pitkow, Ph.D.

---

Simon Fischer-Baum, Ph.D.

---

Valentin Dragoi, Ph.D.

APPROVED:

---

Dean, The University of Texas  
MD Anderson Cancer Center UTHHealth Graduate School of Biomedical Sciences

CORTICAL DYNAMICS OF LANGUAGE

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Kiefer James Forseth, B.S., B.Mus.

Houston, Texas

May 2021

DEDICATION

*in veritate amoenitas*



## ACKNOWLEDGEMENTS

I must first and foremost thank my loving wife, Alexandra, who has supported my academic wanderings with complete faith – even when I had doubts myself. We have shared an incredible journey and there are many more years of exploration remaining!

To my advisor and mentor, Nitin Tandon, I owe a debt of gratitude for a decade of guidance and a vibrant, personalized, and globe-trotting graduate school experience. These have truly been among the best years of my life.

Chris and Cihan carved the path for me to follow and inspired me to pursue the life of a clinician-scientist. I am immensely proud to have worked with and learned from these great peer mentors; I hope to pass on this legacy to those students who come after me.

I had a phenomenal partner throughout graduate school in Patrick who tirelessly kept the lab fundamentals humming; everything from spearheading the ever accelerating data collection efforts to vetoing my latest desired computer accessory.

To all the members of the Tandon Lab, it has been a joy to work with you and see this collective become a force on the global stage of neuroscience.

Without the daily clinical care provided by the staff of the epilepsy monitoring unit, none of this research would be possible – especially the expertise and patience of Delrick and Eugenia.

Thank you to Simon and Xaq for challenging me to broaden my academic horizons in linguistics and mathematics, pushing me beyond my comfort zone in rewarding directions. This extends as well to the many collaborators of the Tandon Lab – notably Greg Hickok and Stan Dehaene.

I have treasured my time in the MD/PhD program here in Houston with a wonderful cohort of faculty, staff, and fellow students. It is a unique bond to go through the parallel – yet wholly distinct – trials and tribulations of medical and graduate school, only to come out of it and be thrown headlong into residency.

Finally, I would be remiss not to acknowledge the great comfort and joy brought to me by my constant companions through studying and analyzing, reading and programming – my 3 dogs Ziggy, Lonnie, and Briggs.

## CORTICAL DYNAMICS OF LANGUAGE

Kiefer James Forseth, B.S., B.Mus.

Advisory Professor: Nitin Tandon, M.D.

## ABSTRACT

The human capability for fluent speech profoundly directs inter-personal communication and, by extension, self-expression. Language is lost in millions of people each year due to trauma, stroke, neurodegeneration, and neoplasms with devastating impact to social interaction and quality of life. The following investigations were designed to elucidate the neurobiological foundation of speech production, building towards a universal cognitive model of language in the brain. Understanding the dynamical mechanisms supporting cortical network behavior will significantly advance the understanding of how both focal and disconnection injuries yield neurological deficits, informing the development of therapeutic approaches.

## TABLE OF CONTENTS

APPROVAL SHEET	I
TITLE PAGE	II
DEDICATION	III
ACKNOWLEDGEMENTS	IV
ABSTRACT	V
TABLE OF CONTENTS	VI
LIST OF ILLUSTRATIONS	VII
ABBREVIATIONS	VIII
INTRODUCTION	1
GENERAL METHODS	4
NAMING WHAT YOU SEE	8
NAMING WHAT YOU HEAR	23
PREDICTION IN PERCEPTION AND PRODUCTION	32
CONCLUSIONS & FUTURE DIRECTIONS	49
REFERENCES	50
VITA	61

## LIST OF ILLUSTRATIONS

FIGURE 1: Process of registering the cortical surface with electrodes	5
FIGURE 2: Schematic for surface-based mixed-effects multilevel analysis	7
FIGURE 3: Examples of coherent and scrambled images of common objects	8
FIGURE 4: 25,810 grid and depth electrodes in left language-dominant cortex	9
FIGURE 5: Spatiotemporal atlas of picture naming	13
FIGURE 6: Models of behavior and power with psycholinguistic predictors	15
FIGURE 7: Autoregressive hidden Markov model applied to high-gamma power	17
FIGURE 8: Network activity contrast in coherent and scrambled picture naming	19
FIGURE 9: Examples of coherent and reversed auditory stimuli	23
FIGURE 10: Surface-based group fMRI of heteromodal naming	25
FIGURE 11: Surface-based group electrocorticography of heteromodal naming	27
FIGURE 12: Conjunction of surface-based electrocorticography maps	27
FIGURE 13: Surface-based group direct cortical stimulation of heteromodal naming	28
FIGURE 14: Sample 3 Hz modulated white noise stimulus with buried pure tone	32
FIGURE 15: Supratemporal depth probe trajectory and coverage map	33
FIGURE 16: Supratemporal response to low- and high-level stimuli	35
FIGURE 17: Cortical response to rhythmic white noise	37
FIGURE 18: Predictive encoding in early auditory cortex	38
FIGURE 19: Edge detection and envelope tracking during natural language speech	39
FIGURE 20: Functional dissociation of sustained and transient responses	41
FIGURE 21: Functional dissociation during listening and speaking	42
FIGURE 22: Sustained and transient responses during listening and speaking	43
FIGURE 23: Supratemporal responses in chronometric stimulation experiments	44

## ABBREVIATIONS

SDE	subdural grid electrode
sEEG	stereotactic depth electrode
SB-MEMA	surface-based mixed-effects multilevel analysis
fMRI	functional magnetic resonance imaging
BOLD	blood oxygen level dependent
gARHMM	grouped autoregressive hidden Markov model
HG/TTS	Heschl's gyrus and the transverse temporal sulcus
PT	planum temporale
NNMF	non-negative matrix factorization
ITC	inter-trial coherence

## INTRODUCTION

People can generate the name for an object and articulate that word with remarkable speed, precision, and fluency. This involves an integrated multistage process that seamlessly translates conceptual representations in the brain to acoustic output – a defining human faculty that allows for eloquent communication. The cortical substrates that support naming form a network from inferior parietal, temporal, ventrolateral prefrontal, medial frontal, insular, and sensorimotor areas. The study of this complex network has been largely mediated by analysis of lesions, neurodegenerative disease, behavioral response, functional imaging, and non-invasive electrophysiology. With these tools, much has been learned about language in the brain; however, these predominant methodologies lack the temporal or spatial resolution to discern the neural mechanisms driving networks characterized by rapid, transient dynamics with strong interactions between distributed cortical substrates. In contrast, invasive human electrocorticography affords direct access to fine spatiotemporal scale recordings of human cortical activity via subdural pial surface electrodes or penetrating depth electrodes. I integrate structural and functional imaging with electrocorticographic recordings and direct cortical stimulation to delineate the specific neuroanatomic architecture and functional organization of the human speech production network.

Traditional models of speech production agree that linguistic components are organized hierarchically and assembled sequentially – this conclusion is informed by the nature of common speech errors (e.g. word exchange)<sup>1</sup>, by chronometric studies of picture naming<sup>2</sup>, and by speech disruption motifs in aphasic patients<sup>3</sup>. However, the interaction of these components defines a fundamental dichotomy in literature between serial<sup>1,4,5</sup> and parallel<sup>6,7</sup> processing. Building on several earlier serial models<sup>1,4</sup>, Bock and Levelt proposed four processing levels: message selection, lexical selection, positional encoding, and finally phonological encoding. Critically, the flow of information in this model is unidirectional<sup>5,8</sup>. In contrast, Dell proposed a parallel processing model with spreading activation across three levels in which distinct speech units – semantic features, lexical nodes, and phonological segments – exhibit bidirectional interactions<sup>6</sup>.

The most immediate advantage of this model was a viable explanation of word and phrase blending errors (e.g. “Shout for help!” / “Yell for help!” becoming “Shell for help!”), as well as cognitive intrusions (e.g. Freudian slips). The debate between serial and parallel processing has evolved into an unresolved conflict between feedforward-only<sup>9,10</sup> and fully-interactive models<sup>11,12</sup>, a conflict perpetuated by the challenge of formalizing neurobiological correlates for linguistic processing levels and by the difficulty in characterizing the interactive behavior of cortical substrates<sup>13</sup>. Large-scale electrocorticography has the unique potential to ground this debate in empirical network dynamics, establishing a translation from psycholinguistic theory to the brain.

While psycholinguists have been primarily concerned with abstract cognitive models that govern the selection and planning of spoken words, motor control theorists have focused on developing models of articulatory execution rooted at the intersection of systems engineering and neurobiology<sup>14–16</sup>. Many such models depend critically on the pairing of feedforward (learned expectation) and feedback (sensory experience) mechanisms. For example, the directions into velocities of articulators (DIVA) model computes both an acoustic and somatosensory comparison of expected target and actual state to generate a real-time error map that monitors ongoing articulation<sup>16</sup>. In recent years, several attempts have been made to integrate psycholinguistic and motor control theories<sup>17–20</sup> culminating in the hierarchical state feedback control (HSFC) model<sup>21</sup>. This model posits that distinct internally generated acoustic and somatosensory targets correspond to syllabic and phonological features, respectively. These targets and their associated motor programs together constitute the phonological level of psycholinguistic theory, interposed between the lexical level and external acoustic output. Direct evidence for many of the functional connections hypothesized in these grand models is still absent, requiring precise quantification of relative timing between distant substrates and the robust demonstration of directional informational flow as well as causal interruption with functional ablation.

Electrocorticography presents a rare window into the inner mechanics of language in the human brain. Patients undergoing presurgical evaluation for refractory epilepsy are implanted

with intracranial electrodes and clinically observed for an extended period in the epilepsy monitoring unit. These electrodes record directly from the brain, providing physicians and scientists with powerful insights into human neurophysiology. The recorded signal constitutes the full summation of cortical oscillatory activity with millimeter spatial and millisecond temporal resolution; this allows for spectral decomposition of the temporal signal at each electrode into functional neural bands. For example, the high-gamma band (60-120 Hz) – thought to arise from focal summation of recurrent inhibitory inter-neuronal activity coupled with a surge in spike rate<sup>22-26</sup> – strongly correlates with the blood-oxygen level dependent signal from fMRI<sup>27,28</sup> and indexes local cognitive processing<sup>22,29-31</sup>. With electrocorticography, it is uniquely possible to study properties of these distinct frequency channels in local cortical regions with respect to other channels, to other regions, and to behavioral response. Additionally, these intracranial electrodes allow for direct electrical stimulation that transiently mimics focal lesions, allowing for the comparison between functional activation and lesion studies in each patient.

Despite general agreement on the enumeration of cognitive processes that lead from intention to articulation, no consensus has yet emerged governing the network architecture and dynamics that implement these processes. The overall goal of my thesis work was to discern how the speech production network in the human brain coordinates its distributed substrates to accomplish fluent verbal communication. I have studied this network with the unique advantages of electrocorticography – guided by neuroimaging – in a large cohort including both penetrating and surface electrodes: excellent spatiotemporal resolution, direct full spectrum recordings, complete cortical coverage, and focal current injection. I identify the substrates contributing to the speech production network, measure information flow between its nodes, and selectively perturb functionality. Insights into the network dynamical properties of the brain creates a new framework for the understanding of language disorders and provides generalizable perspectives on cognitive processes.



## GENERAL METHODS

All study designs were approved by the committee for the protection of human subjects at The University of Texas Health Science Center at Houston. Patients participated in all research voluntarily with informed consent. This research did not influence their clinical care.

### ELECTRODE IMPLANTATION

Two types of electrodes were implanted by my mentor, Dr. Tandon, at Memorial Hermann Hospital (Figure 1). The first method involves accessing the brain via a craniotomy, then placing grids of platinum-iridium electrodes (top-hat design; 3-mm diameter cortical contact) embedded in silastic sheets (PMT, Chanhassen, MN) at strategic locations across the lateral and ventral pial surface. The second method, increasingly favored in American neurosurgical practice, employs a surgical robot (ROSA, MedTech, France) to aid in the insertion of linear depth probes (PMT) – each with multiple platinum-iridium contacts (0.8 mm diameter, 2 mm length cylinders; separated from adjacent contacts by 1.5 to 2.43 mm) – along stereotactic trajectories. Grid and depth electrodes have distinct access to cortical structures: grid electrodes predominantly record from gyri on the lateral or ventral surface, while depth electrodes – frequently implanted bilaterally – are also able to record from cortical sulci, the medial bank of the hemisphere, and the insula. These two surgical approaches generate complementary coverage of the neuroanatomic substrates implicated in speech production.

### STRUCTURAL IMAGING

The localization of electrodes relative to the cortex is accomplished by the registration of pre- and post-operative imaging. Pre-operative structural MRI scans are obtained using a 3T whole-body MR scanner (Philips Medical Systems)<sup>32</sup>; these scans are then processed with FreeSurfer<sup>33</sup> to yield pial surface reconstructions. Post-operative CT scans are subsequently registered to the structural MRI and imaging artifacts caused by the metal electrodes are exploited to estimate position relative to the cortex in AFNI<sup>34</sup>. Grid electrode positions are manually confirmed with intra-operative photographs<sup>35</sup>.

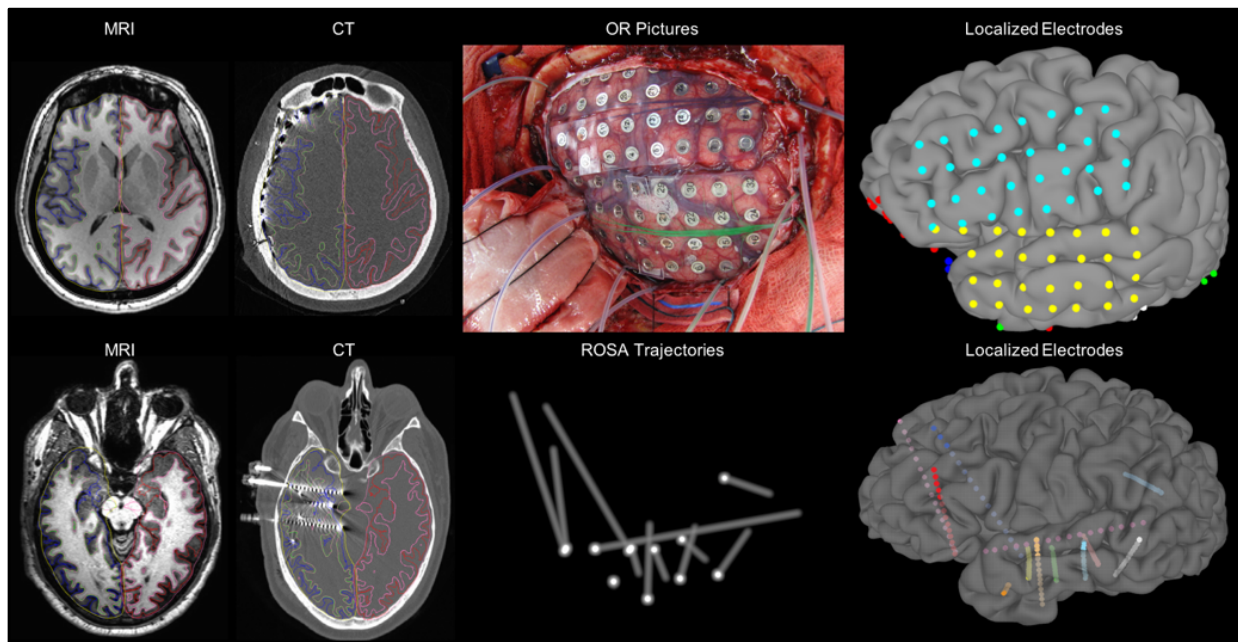


Figure 1: Process of registering the cortical surface with surface grid (top) and penetrating depth (bottom) electrodes. The final electrode representations registered to the cortical surface are shown on the right for 2 distinct patients.

#### ELECTROPHYSIOLOGY ACQUISITION

Electrocorticography data were collected with a sampling rate of either 1 kHz and bandwidth of 0.15 to 300 Hz using Neurofax (Nihon Kohden) or with a sampling rate of 2 kHz and bandwidth of 0.1 to 700 Hz using NeuroPort NSP (Blackrock Microsystems). Continuous audio recordings of each patient were performed with both an omnidirectional microphone (Audio Technica U841A, 30 to 20,000 Hz response, 73 dB SNR) placed adjacent to the presentation laptop and a cardioid lavalier microphone (Audio Technica AT898, 200 to 15,000 Hz response, 63 dB SNR) clipped to clothing near the mouth. These recordings were analyzed offline to transcribe patient responses, as well as to determine the time of articulatory onset and offset. Hemispheric language dominance was evaluated by intracarotid sodium amytal injection<sup>36</sup>, fMRI laterality index<sup>37</sup>, or cortical stimulation mapping<sup>38</sup>; the remaining patients were assumed to be left hemisphere language-dominant. Only electrodes unaffected by epileptic activity, artifacts, or electrical noise were used in subsequent analyses. Trials in which the patient answered incorrectly or did not respond were eliminated. Additionally, trials in which significant epileptic activity was observed were removed.

## INVERSE MODEL OF ELECTRODE RECORDING ZONES

Response properties of individual electrodes were mapped to patient-specific cortical models via a surface recording zone<sup>39,40</sup>. This zone both constrained the surface-based spatial registration of individual cortical models to a standard atlas and constrained the weighted estimate of cortical contributions to the observed signal at each electrode.

The definition of electrode recording zone was tailored to the type of electrode. For grid electrodes, the central coordinate of each electrode was matched to its closest node on the cortical envelope. This seed was then grown to a geodesic radius of 1.5 mm, matching the dimensions of the electrode. Each of the vertices within this region was mapped to its closest vertex on the pial surface model. These seeds were then grown along the surface to a maximum geodesic radius of 10 mm, constituting the surface electrode recording zone. For depth electrodes, the central coordinate of each electrode was simply mapped directly to the closest vertex on the pial surface model. This seed was then grown to a maximum geodesic radius of 10 mm to define the surface electrode recording zone.

For both electrode types, the inverse model within the recording zone was defined by the same piecewise inverse function. Cortex directly adjacent to electrodes (e.g. beneath grid electrodes or alongside depth electrodes) was assigned a maximal weight of 1; more distant cortex was weighted according to exponential decay with a full-width half-maximum at 2.3 mm; cortex more than 10 mm from the center was assigned a minimal weight of 0. Individual electrode statistics were subsequently propagated onto the cortical surface with this inverse function.

## SURFACE-BASED REGISTRATION TO STANDARD ATLAS

All group analysis was performed in standard space on the MNI N27 cortical surface. Electrode locations and recording zones were transformed to standard space with a nonlinear surface-based registration<sup>41,42</sup>. This registration was used to generate coverage maps, define regions of interest, and to enable group statistics at each vertex of the cortical surface in the mixed-effects multilevel analysis<sup>39,40</sup>.

## MIXED-EFFECTS MULTILEVEL ANALYSIS

To generate statistically robust and topologically precise estimates of percent change in power across the cortex, I used a surface-based mixed-effects multilevel analysis (SB-MEMA) leveraging the electrode recordings zones and nonlinear transform to standard space defined previously. This method addresses universal challenges for grouped analysis of human invasive electrophysiology: a) accurate localization and spatial localization of cortical sources<sup>39,40</sup>; b) data integration across the cohort, accounting for sparse sampling and anatomic variability<sup>41–43</sup>; c) statistical modeling of population-level effects, mitigating outlier inferences and accounting for intra- and interpatient response variability that violate the assumptions of simpler models<sup>44,45</sup>.

The model consisted of two levels: the individual and the group (Figure 2). At the individual level, an estimate of percent change in power at each electrode was fitted with a mixed-effects model informed by the sampling variance. The resulting effect and significance estimates were propagated onto the patient-specific cortical model using the surface recording zone of each electrode. These patient-specific maps then underwent surface-based registration to the standard atlas. Finally, a mixed-effects model at the group level generated the effect and significance estimates for each vertex on the MNI N27 atlas<sup>39,40</sup>.

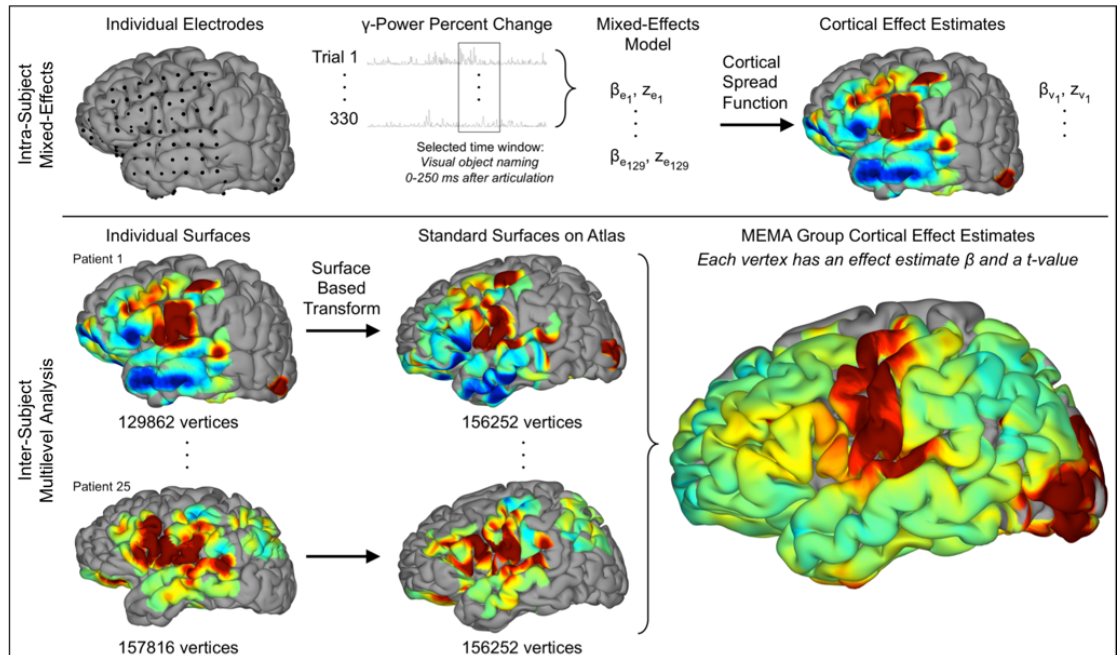
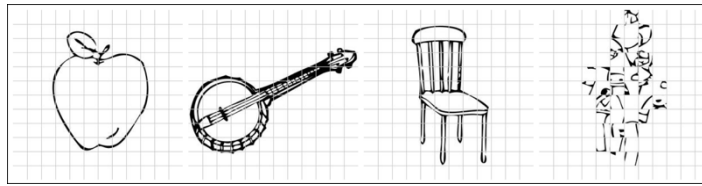


Figure 2: Schematic for surface-based mixed-effects multilevel analysis at a peri-articulatory window in 25 patients.

## NAMING WHAT YOU SEE

*What the brain does as we speak (in review)*

Forseth KJ, Pitkow X, Fischer-Baum S, and Tandon N

*Figure 3: Examples of coherent and scrambled images of common objects*

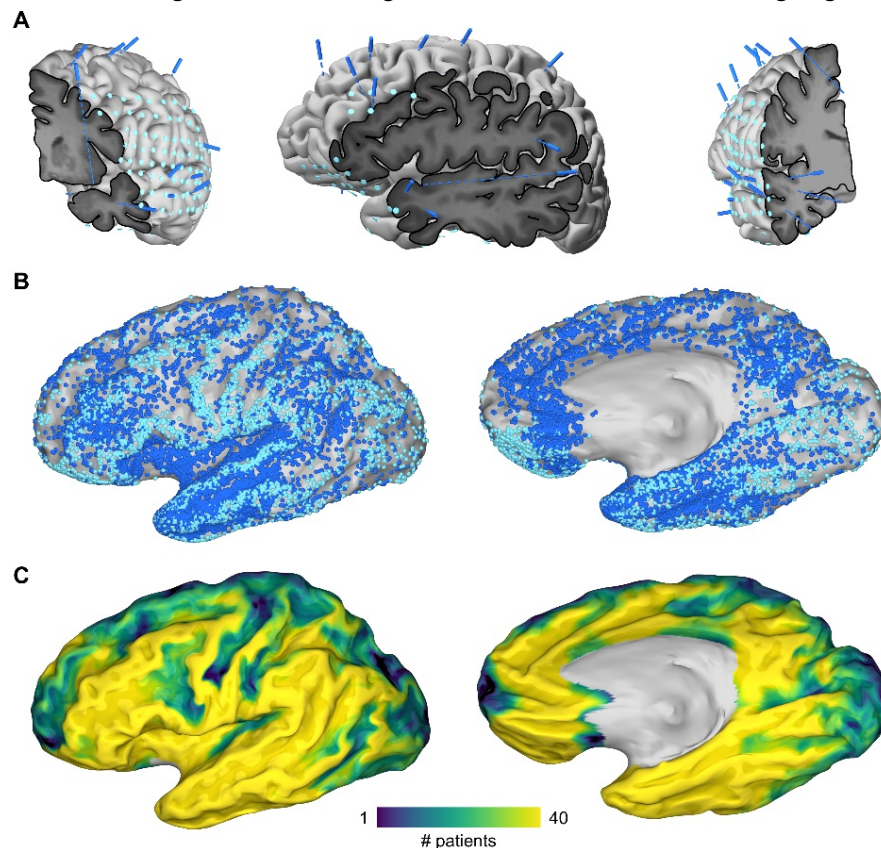
## BACKGROUND

Speech production is a defining human faculty that enables eloquent communication. The ubiquity of word generation with remarkable speed, precision, and fluency belies its complexity. Articulating even a single word requires the selection of a conceptual representation, the construction of a word form, and the execution of a complex articulatory plan with associated output monitoring. Despite general agreement on the enumeration of cognitive processes that lead from intention to articulation<sup>8</sup>, no consensus has yet emerged describing the neurobiological architecture that implements these processes.

The cortical basis of speech production has been probed by analyses of deficits secondary to lesions<sup>46</sup> and neurodegenerative diseases<sup>47</sup>, as well as by analyses of intact language systems through functional imaging<sup>48,49</sup>, structural mapping<sup>50</sup>, and noninvasive electrophysiology<sup>51</sup>. Much of this work has focused on localizing specific cognitive processes to discrete neuroanatomic substrates, yet these efforts have yielded competing interpretations – even in landmark regions like Broca’s area<sup>52</sup>. An alternative framework shifts the focus from patterns of isolated activity in separable substrates to patterns of dynamic interaction between such cortical nodes<sup>53–55</sup>. Evaluation of this theory has been hampered by limitations inherent to the predominant methodologies available for studying the neurobiology of language. These methods lack the temporal or spatial resolution to discern the neural mechanisms driving networks characterized by rapid, transient dynamics across distributed substrates. Invasive human electrocorticography uniquely affords direct, high-resolution recordings of human cortical

activity; however, such opportunities are rare and prior language studies have been limited in scale and cortical coverage<sup>56–58</sup>.

I overcame these limitations by using large-scale human electrocorticography (134 patients, 25810 electrodes, 40278 trials) to elucidate the neurobiology of language production. This cohort included both subdural surface grid and stereotactic depth electrodes, encompassing the entirety of language-dominant cortex (Figure 4). With this global perspective, I generated a comprehensive spatiotemporal atlas of a classical language paradigm: picture naming with scrambled images as a low-level control (Figure 3). I characterized functionally distinct regions within this atlas by pre- and post-articulatory encoding of established psycholinguistic variables including visual, semantic, lexical, and phonologic correlates. I further developed a grouped dynamical model to resolve discrete neural states that were distinguished by unique patterns of distributed cortical interaction. These data reveal the network architecture of speech production, informing and constraining the neurobiological instantiation of extant language models.



**Figure 4: Electrodes in left language-dominant cortex. (A)** Staged implants with grid (light blue) and depth (dark blue) electrodes, showing trajectories through frontal-insular (left), supratemporal (middle), and midtemporal (right) structures. **(B)** All grid and depth electrodes in standardized space. **(C)** Aggregate cortical coverage.

In addition to providing new insights for language production theory, this approach investigates the broad utility of linking cognitive processes to network states rather than to activity in isolated substrates<sup>59</sup>. I critically evaluate the thesis that complex behaviors comprise sequences of network states<sup>60</sup>, each defining a set of reference dynamics to coordinate the generation and transmission of information throughout the cortex<sup>61</sup>. Speech production is an ideal testbed, requiring the coordination of multiple cognitive systems and resulting in an observable behavior. Elaborating the structural and functional properties of states during speech production provides a basis for understanding the generalizable dynamical principles governing cognition<sup>62</sup>.

#### SPECIFIC METHODS

The autoregressive hidden Markov model (ARHMM) combines autoregressive (AR) stochastic linear dynamics with the discrete switching latent states of a hidden Markov model (HMM)<sup>63</sup>. This method enables single-trial analysis that does not require manual alignment of trials by picture presentation or articulation onset. Furthermore, it obviates the dubious assumption underlying all cross-trial averaging metrics that the same cognitive processes are occurring at the same times in all trials. All latent parameters – the time series of network states, their transition probabilities, and the dynamics of each state – are inferred directly from neural data. In this work, I have expanded on this framework with the implementation of a grouped training method (gARHMM) that enables the inference of a single set of generalized latent parameters across the entire patient cohort.

An AR process is a random process with temporal structure, where the current state  $x_t$  of a system is the sum of a linear combination of previous states and a stochastic innovation  $v_t$  drawn from zero-mean isotropic white noise. The linear dynamics of such a system can be described by  $N_\tau$  matrices of AR coefficients  $A_\tau$  at different time delays  $\tau$ , which can be combined into a tensor  $\mathbf{A} = \{A_\tau\}$ . The stochastic elements are specified by a covariance matrix  $Q$ .

$$\mathbf{x}_t = \sum_{\tau=1}^{N_\tau} A_\tau \mathbf{x}_{t-\tau} + Q^{1/2} \mathbf{v}_t$$

Since this model is linear, it is a poor approximation of nonlinear neural dynamics. This limitation motivated the subsequent extension of this model to include switching dynamics.

The defining property of a first-order Markov model is that transition probabilities between states depends only on the previous state. In a *hidden* Markov model, this dynamic is unobserved – each state emits observable quantities with some associated probability. *Autoregressive* hidden Markov models combine the stochastic linear dynamics of AR processes with the partial observability of a hidden Markov model. Here I use a discrete latent state  $z$  and assume autoregressive Gaussian emissions conditioned on that latent state. Each latent state  $z$  indexes a different stochastic linear process with a state-specific dynamics tensor  $A_z$  and a state-specific process noise covariance  $Q_z$ . The switching characteristics allow an ARHMM to approximate a stochastic nonlinear dynamical system.

In the context of invasive electrophysiology, the observations are high-gamma power at a fixed number of regions (e.g. visual cortex, mid-fusiform gyrus, pars triangularis, pars opercularis, subcentral gyrus, and superior temporal gyrus). The AR coefficients  $A_{z\tau jk}$  specify the Granger causal dynamical relationship between regions  $j$  and  $k$  at time lag  $\tau$  in state  $z$ . For a given state  $z$  at time  $t$ , the multivariate high-gamma power signal  $x$  is modeled as

$$\mathbf{x}_t = \sum_{\tau=1}^{N_\tau} A_{z_t\tau} \mathbf{x}_{t-\tau} + Q_{z_t}^{1/2} \mathbf{v}_t + \mu_{z_t}$$

where  $\mu_z$  is a state-dependent bias. Probabilistic inference on observed neural data determines the unobserved latent parameters: the time series of network states  $z_t$ , their transition probabilities  $\Phi_{z_{t-1}z_t}$ , and the linear dynamics parameters of each state  $\{A_z, Q_z, \mu_z\}$ .

The model was trained with iterative estimation of the state time series and state dynamics across all patients using the Baum-Welch expectation-maximization algorithm<sup>64,65</sup>. Initial conditions for  $A$  and  $Q$  were informed by the lagged correlation of multivariate AR clustering<sup>66,67</sup>. The state-dependent biases  $\mu_z$  were drawn randomly from a standard Gaussian. State transitions were sampled from a uniform prior.



A random 80% of trials from each patient were used for training the latent dynamical parameters. 10-fold validation on the training set was used to select hyperparameters. The model contains 2 hyperparameters that constrain its architecture: model order  $N_\tau$  and state number  $N_s$ . Both were evaluated with log-likelihood and AIC. The return for increasing model order plateaued after  $N_\tau = 3$ . Additional states exceeding  $N_s = 6$  trivially split the background state during inter-trial periods. I trained a second model with interaction terms forced to zero – otherwise, the architecture was preserved. This mean-only model converged to a stable solution, but its performance was inferior to the complete model featuring interactions between regions.

The remaining 20% of trials from each patient were allocated to the testing pool to assess model fit. Performance of the model, measured with log-likelihood, was equivalent on training and testing sets. In addition, I iteratively held out each patient from training to ensure that the model was not overfit to individual-level effects. I report the state sequences generated by the model for all trials in the testing set.

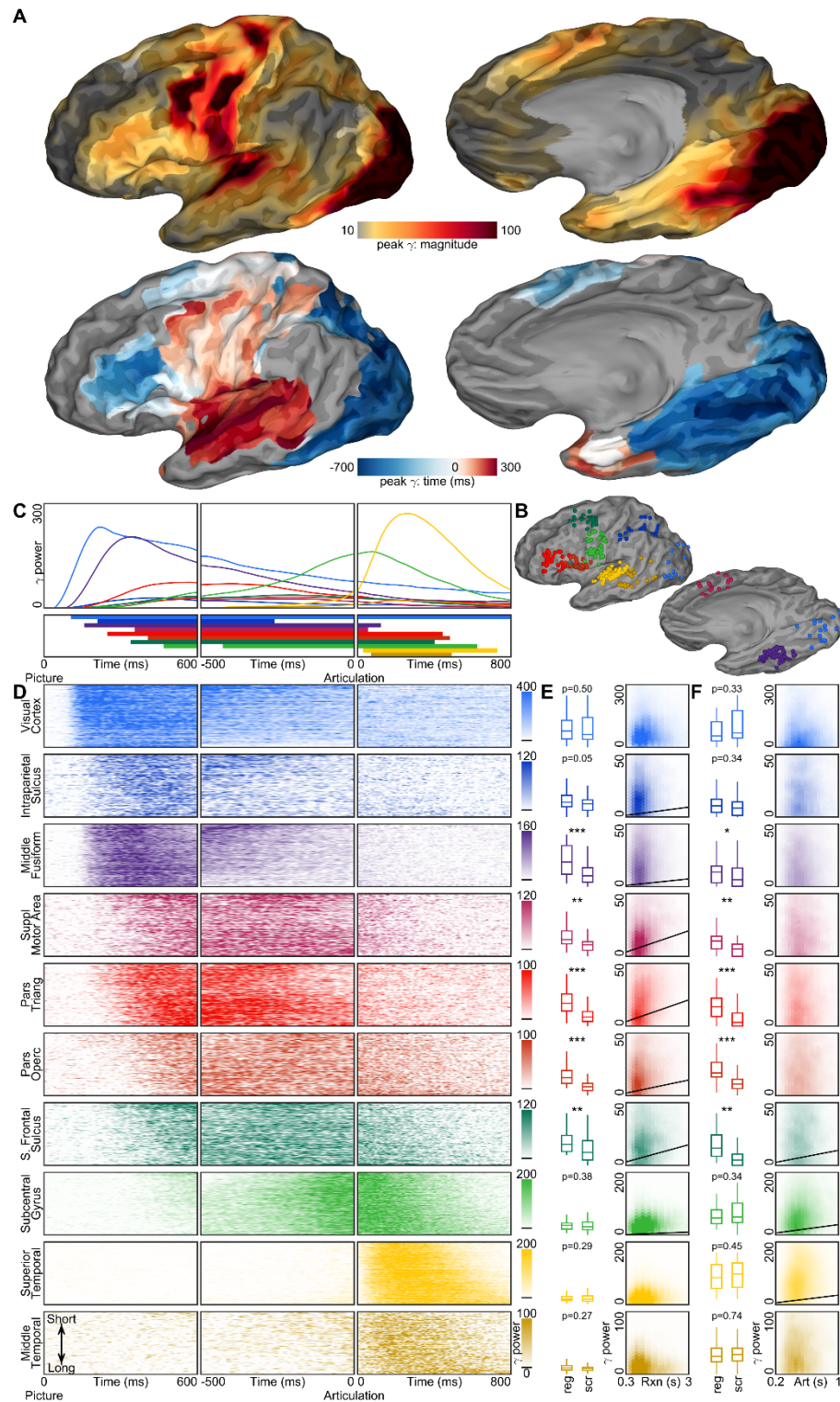
The ARHMM classifies dynamical states by the network connectivity encoded in their AR coefficients. These dynamics can be captured by partial directed coherence (PDC) in the frequency domain<sup>68</sup> – a measure of Granger causal information flow. For each state  $z$ , the pairwise PDC between regions  $j$  and  $k$  is defined as

$$\pi_{zjk}(f) = \frac{\bar{A}_{zjk}(f)}{\|\bar{A}_{zk}(f)\|}$$

where

$$\bar{A}_{zjk}(f) = \begin{cases} 1 - \sum_{\tau=1}^{N_\tau} A_{z\tau jk} e^{-2\pi i f \tau} & \text{for } j = k \\ - \sum_{\tau=1}^{N_\tau} A_{z\tau jk} e^{-2\pi i f \tau} & \text{otherwise} \end{cases}$$

represents the transfer function at frequency  $f$ . In this text, the directed graph for each state is shown with nodes representing regions and edges representing causal interactions.



**Figure 5: Spatiotemporal atlas of picture naming. (A)** SB-MEMA of peak high-gamma power relative to articulation with brightness indicating areas that were significantly more active for coherent images than scrambled images. **(B)** 10 regions of interest were delineated. **(C)** High-gamma power for stimulus and articulation aligned windows ( $p < 10^{-3}$ ). **(D)** Raster plots of single-trial high-gamma power sorted by reaction time (post-stimulus, pre-articulation) or articulatory duration (post-articulation onset). **(E)** Boxplots of average gamma power in the 4 seconds after picture presentation for coherent (left) and scrambled (right) stimuli (\*  $p < 0.05$ , \*\*  $p < 10^{-2}$ , \*\*\*  $p < 10^{-3}$ ). Scatter plots of average high-gamma power against reaction time for coherent images. Regression lines are overlaid for correlations that were both significant during coherent trials and significantly greater than during scrambled trials ( $p < 10^{-3}$ ). **(F)** Same analysis was repeated for the 1 second after articulation onset against articulatory duration.

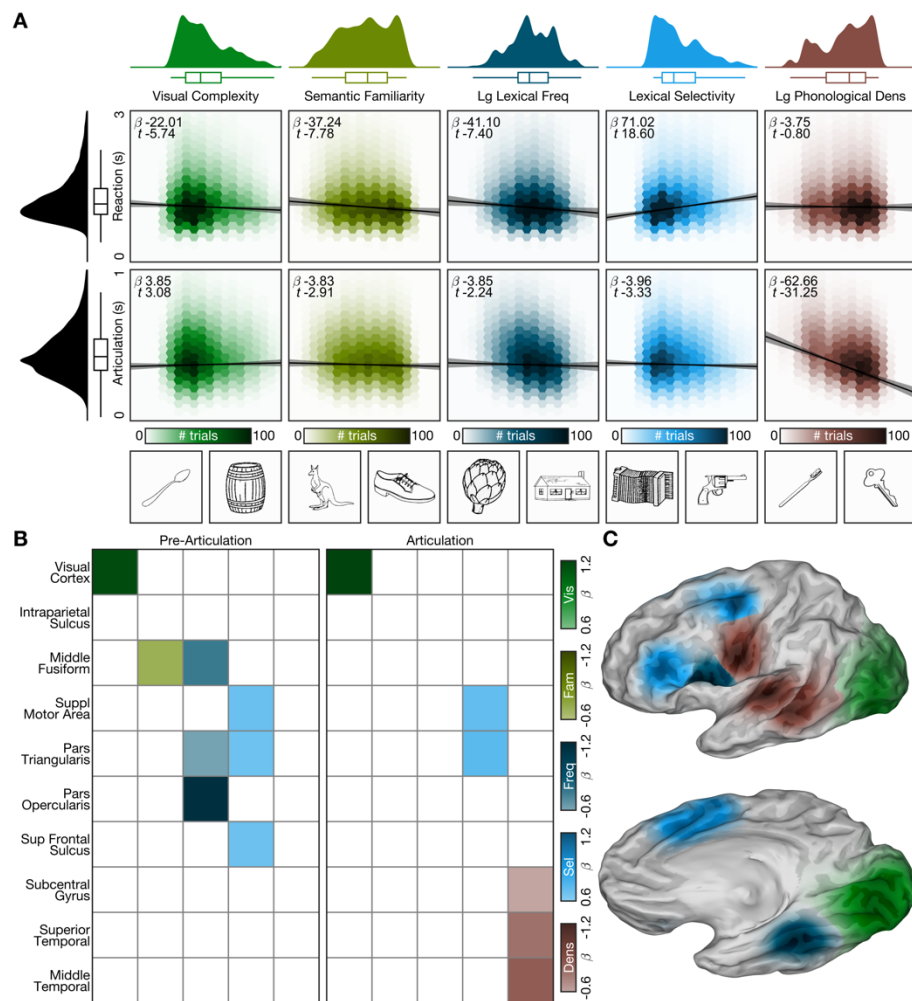
## FINDINGS

Cortical activity was integrated across the cohort to produce a spatiotemporal atlas of cued word production in the language-dominant hemisphere (Figure 2). These global mean dynamics were resolved with surface-based mixed-effects multilevel analysis of high-gamma power in narrow time windows, generating a series of effect sizes and confidence estimates for every point on the standard Colin pial surface. The resulting frames together constitute a high-resolution movie (Figure 5A). I further investigated the response properties of this distributed cortical network for naming within regions of interest constrained by the SB-MEMA atlas (Figure 5B). The mean regional response was analyzed in 3 adjacent time windows: post-stimulus, pre-articulation, and post-articulation (Figure 5C). A broad swath of language-dominant cortex was sequentially recruited during picture naming. The cortical response began in the calcarine sulcus, then spread both dorsally to the intraparietal sulcus and ventrally to the middle fusiform gyrus. Next, a complement of distinct foci in frontal cortex were engaged: pars triangularis, pars opercularis, the supplementary motor area, and the superior frontal sulcus. Significant peri-articulatory activity followed in the subcentral gyrus with dorsal extension throughout lateral sensorimotor cortex. The superior temporal and posterior middle temporal gyri were mostly quiescent in the pre-articulatory interval, but engaged throughout articulation.

I then quantified the relationship between behavior – reaction time and articulatory duration – and neural response in each region of interest for both coherent and scrambled images (Figure 5D-F). Visual cortex, the intraparietal sulcus, and the middle fusiform gyrus responded at a fixed delay from picture presentation; in contrast, frontal cortex broadly responded later during trials with longer reaction times. Of the regions that were significantly more responsive to coherent images than scrambled images, cumulative high-gamma power and reaction time were significantly correlated in the middle-fusiform gyrus ( $r = 0.082$ ,  $p < 10^{-6}$ ), pars triangularis ( $r = 0.278$ ,  $p < 10^{-6}$ ), pars opercularis ( $r = 0.166$ ,  $p < 10^{-6}$ ), the supplementary motor area ( $r = 0.082$ ,  $p < 10^{-6}$ ), and the superior frontal sulcus ( $r = 0.219$ ,  $p < 10^{-6}$ ). Cumulative post-articulatory high-gamma power and articulatory duration were significantly correlated in the superior frontal sulcus

( $r = 0.072$ ,  $p < 10^{-4}$ ), the subcentral gyrus ( $r = 0.199$ ,  $p < 10^{-6}$ ), and the superior temporal gyrus ( $r = 0.182$ ,  $p < 10^{-6}$ ). The middle fusiform gyrus, pars triangularis and opercularis, and the supplementary motor area remained significantly more engaged for coherent versus scrambled images in the post-articulatory timeframe despite having no association with articulatory duration.

This holistic view of the cortical dynamics of picture naming was uniquely afforded by large-scale intracranial electrophysiology spanning the entirety of language-dominant cortex. While there was a clear progression of activity across the cortex, many regions were jointly active for extended periods. These dynamics are inconsistent with the narrow assignment of linguistic operations to focal and isolated substrates. Instead, they may be better explained by the proposal



**Figure 6: Linear mixed-effects models of behavior and regional high-gamma power with psycholinguistic predictors. (A) Correlations with reaction time and articulation length in 19465 trials. (B) Models of regional high-gamma power response in pre-articulatory (-2150 to 0ms) and articulatory (0 to 850ms) time windows ( $p < 10^{-2}$ ). (C) Cortex attributable to electrodes in each region colored by the predictor with the largest significant t-value.**

that cognitive computation that is orchestrated across transient, distributed, and overlapping networks – a thesis I further evaluate in subsequent analyses.

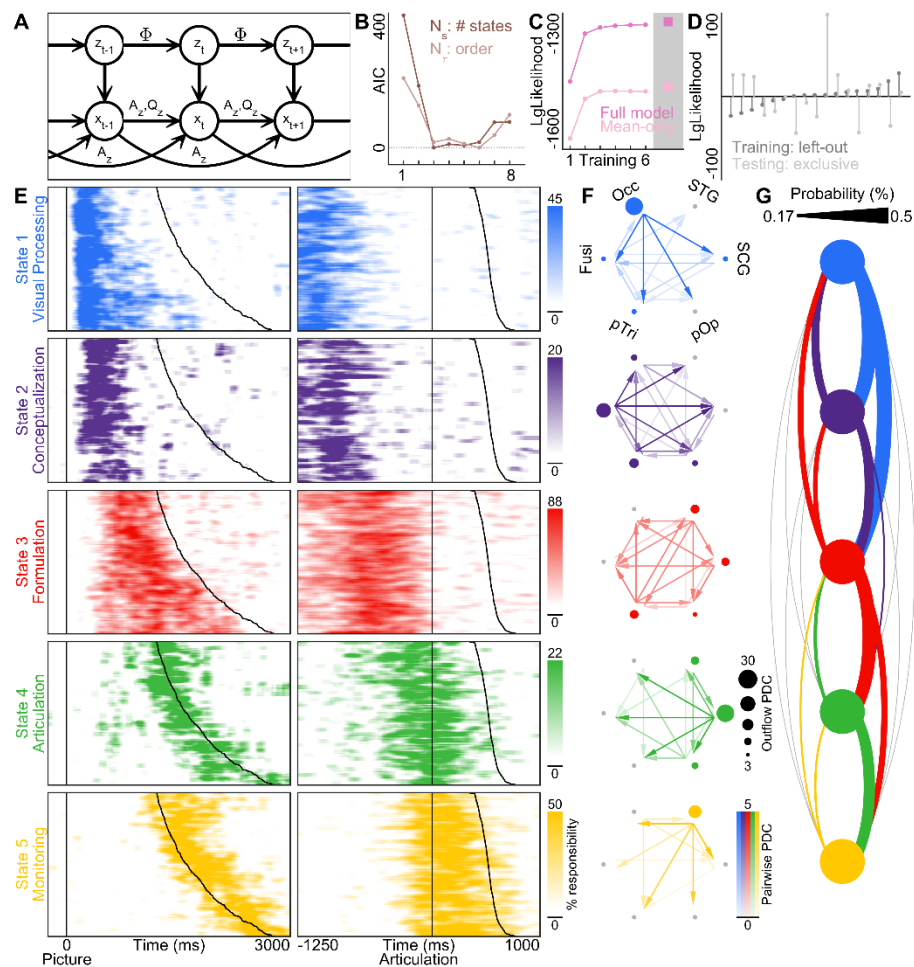
Having characterized the mean spatiotemporal extent of high-gamma power, I evaluated the trial-by-trial effects of distinct linguistic representations on regional activity (Figure 6). Linear mixed-effects models were constructed as a function of visual complexity<sup>69</sup>, semantic familiarity<sup>70</sup>, lexical frequency<sup>71</sup>, lexical selectivity<sup>70</sup>, and phonological density<sup>72</sup>. These models were validated on behavioral measures (Figure 6A): reaction time was best explained by lexical selectivity ( $\beta = 63.54\text{ms}$ ,  $p < 10^{-6}$ ), while articulation length was best explained by phonological density ( $\beta = -62.66\text{ms}$ ,  $p < 10^{-6}$ ).

I applied these models to high-gamma power in each region of interest during time windows before and after the onset of articulation. Visual cortex activity was related to visual complexity of the stimulus ( $\beta = 0.116$ ,  $p < 10^{-5}$ ), consistent with localized feature processing. In the pre-articulatory period, semantic familiarity was uniquely related to middle fusiform gyrus activity ( $\beta = -0.071$ ,  $p < 10^{-3}$ ). Lexical frequency was also encoded in the middle fusiform gyrus ( $\beta = -0.077$ ,  $p < 10^{-4}$ ), as well as in pars triangularis ( $\beta = -0.063$ ,  $p = 0.0053$ ) and opercularis ( $\beta = -0.116$ ,  $p < 10^{-4}$ ). Lexical selectivity was related to activity in pars triangularis ( $\beta = 0.069$ ,  $p < 10^{-3}$ ), the supplementary motor area ( $\beta = 0.070$ ,  $p = 0.0014$ ), and the superior frontal sulcus ( $\beta = 0.069$ ,  $p = 0.0011$ ). After the onset of articulation, phonologic density of the spoken response was encoded in perisylvian regions: subcentral ( $\beta = -0.061$ ,  $p < 10^{-4}$ ), superior temporal ( $\beta = -0.078$ ,  $p < 10^{-6}$ ), and posterior middle temporal gyri ( $\beta = -0.085$ ,  $p = 0.0043$ ).

I have established that language engages a distributed network of regions with concurrent activity and separable linguistic correlates; however, the analyses thus far assume that cognitive operations are locally computed in isolated substrates and time-locked to observable events. This assumption confounds attempts to distinguish the neurobiological correlates of interactivity that are foundational to psycholinguistic models of speech production<sup>11</sup>. To relax this assumption and directly resolve interactions between regions, I implemented a grouped autoregressive hidden Markov model (gARHMM, Figure 7A). Critically, this model learns a single set of latent

dynamical parameters across the patient population and emits state sequences of network interactions for each trial<sup>63</sup>.

17 patients had concurrent coverage of the core language network: visual cortex, middle fusiform gyrus, pars triangularis, pars opercularis, subcentral gyrus, and superior temporal gyrus. The hyperparameters – model order ( $\tau = 3$ ) and number of states ( $N_z = 6$ ) – were determined by 10-fold cross-validation (Figure 7B) of the training dataset (80% of trials uniformly sampled from all patients). Latent dynamical parameters were then optimized on the training dataset and applied to the held-out test dataset. This model performed significantly better ( $p < 10^{-6}$ ) than a



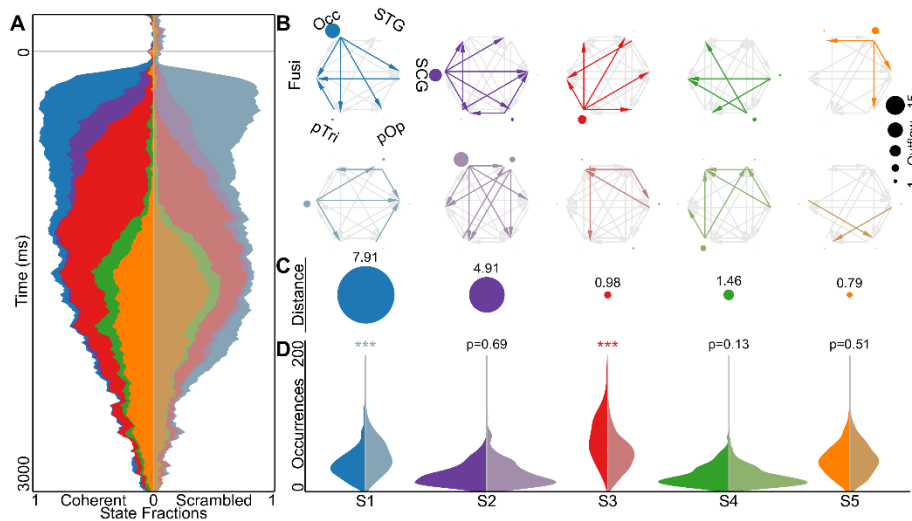
**Figure 7: Autoregressive hidden Markov model applied to single-trial high-gamma power in 17 patients with electrodes over visual cortex, middle fusiform gyrus, pars triangularis and opercularis, subcentral gyrus, and superior temporal gyrus. (A)** Latent states  $z$  and observations  $x$  evolve with autoregressive state dynamics  $A$ , regional process covariances  $Q$ , and state transition probabilities  $\Phi$ . **(B)** Hyperparameter selection with 10-fold cross-validation on the training set. **(C)** Model comparison of training and validation performance with mean-only model. **(D)** Hold-out analysis showing that performance is robust to exclusion of individual patients and generalizes well to all patients. **(E)** State sequence rasters on held-out testing data. Each active state is represented by a unique color. Trials were sorted by reaction time (left) or articulatory length (right). **(F)** Interregional dynamics specific to each state expressed as partial directed coherence (PDC, edges) between regions (nodes). Significance was evaluated via bootstrapping at  $p < 10^{-3}$ . **(G)** State transition probabilities that surpass the uniform distribution of non-self-transitions.

model of the same design but lacking interactions between regions (Figure 7C). Model training was unbiased by data from any single patient and testing generalized to data from held-out patients (Figure 7D).

I identified 6 dynamical states – 5 states during speech production and a background state between trials (Figure 7E). The 5 active states demonstrated a consistent temporal precession relative to both picture presentation and articulation. These states lend themselves in number and timing to established psycholinguistic nomenclature<sup>8</sup>: visual processing, conceptualization (activation and selection of a lexical concept), formulation (staged form encoding to produce gestural scores), articulation, and monitoring. The Granger-causal interactions between regions that define the dynamics of each state were quantified with partial directed coherence (Figure 7F). A second model trained on trials with scrambled images also identified 5 states, but with differences in network architecture and state frequency (Figure 8).

Each state featured significant contributions from a complex network of pairwise regional interactions that were essential for optimal modeling of neuronal dynamics; the most salient interactive properties of these states are described here. The first state, *visual processing*, was concentrated in the ~250 ms immediately following picture presentation and its dynamics were dominated by outflow from visual cortex. This was followed by the second state, *conceptualization*, from ~250 to ~500 ms in a distributed network organized by outflow from the middle fusiform gyrus and pars triangularis. The network architecture of these two states was significantly altered during scrambled trials (Figure 8C); in particular, the outflow from the middle fusiform gyrus was largely replaced by outflow from visual cortex. The third state, *formulation*, recruited a decentralized perisylvian network that remained engaged through the onset of articulation and accounted for the majority of variance in reaction time ( $\beta = 7.41$ ,  $p < 10^{-6}$ ). The network architecture of this state was similar for coherent and scrambled trials, but it occurred significantly more frequently in response to coherent images ( $p < 10^{-6}$ , Figure 8D). The fourth state, *articulation*, was engaged in the ~400 ms around articulatory onset with dynamics





**Figure 8: Network activity contrast in coherent and scrambled picture naming. (A)** The fraction of trials engaging a given state as a function of time from stimulus presentation for coherent (left, light colors) and scrambled (right, muted colors) images. **(B)** Pairwise interactions, quantified with partial directed coherence, and nodal outflows that were greater during coherent trials are shown in the top row; those greater during scrambled trials are shown in the bottom row. The significance of every interactional coefficient and nodal outflow ( $p < 10^{-3}$ ). **(C)** Cosine similarity of each network state pair. The first two states were different during coherent and scrambled images, while the latter three states were similar. **(D)** Distribution of state frequency ( $* p < 0.05$ ,  $** p < 10^{-2}$ ,  $*** p < 10^{-3}$ ).

dominated by outflow from subcentral gyrus. The fifth and final state, *monitoring*, was active throughout articulation relied predominantly on outflow from the superior temporal gyrus. Both *articulation* and *monitoring* featured convergent network architectures and state frequencies for coherent and scrambled images. These results are consistent with my thesis that the fundamental unit of linguistic computation in the brain is not a set of discrete functional regions, but a sequence of dynamical network states.

I also computed state transition probabilities that describe the rate of observed pairwise switching (Figure 7G). These revealed a locally interactive state switching behavior conserved across trials of varying reaction time and articulatory length. For every state, the most likely transition was to the next state. Transitions from formulation to any other state were common, while transitions directly from visual processing or conceptualization to articulation or monitoring were below chance. Together, the restricted set of cognitive state transitions and the imbricated set of interactive state dynamics ground the discreteness-interactivity axis<sup>11</sup> of speech production models<sup>3,8,10</sup> in the neurobiology of human cortex.



## DISCUSSION

Language involves the coordinated activity and interaction of diverse cortical substrates, yet prior electrocorticographic investigations have studied this broad network in fragments<sup>30,57,58,73–75</sup>. Complete coverage of the cortical surface with intracranial electrodes requires upwards of 10,000 contacts<sup>76</sup>. Furthermore, grid and depth electrodes have distinct stereotypic coverage probabilities: grid electrodes predominantly record from gyri on the lateral or ventral surface, while depth electrodes are more likely to record from cortical sulci, the ventral temporal surface, the medial bank of the hemisphere, and the insula. This unique database of intracranial recordings from both grid and depth electrodes yielded the magnitude, density, and homogeneity of coverage necessary to assemble a comprehensive spatiotemporal map of speech production encompassing the entirety of language-dominant cortex. I used this resource to finely characterize the activity at specific network nodes and the evolving interactive dynamics between nodes. Such an approach is imperative for definitively resolving the processes that lead from intention to articulation<sup>8</sup>.

Despite general agreement on the processes required to convert a picture to its spoken name<sup>3,77</sup>, the instantiation of these processes in neural dynamics is unknown. Psycholinguists and neurobiologists have long engaged in parallel approaches for the study of speech production, probing the hidden internal processes of language generation and exploring the functional contributions of discrete neuroanatomic substrates. Picture naming has demonstrated exceptional utility in these pursuits. Varying stimulus complexity of isolated representational levels – semantic, lexical, and phonological – yields behavioral data which outline the veiled architecture of speech production. Reaction time variations following challenges to the production system with level-specific interference at early and late windows suggest at least two separate steps in lexical access<sup>10,78,79</sup>. Error types and rates constrain the interactions between representational levels in formalized models of production<sup>1,4,6</sup> and, through the anatomic correlates of distinct aphasic patterns, establish connections to neurobiology<sup>80</sup>. Functional imaging in subjects with intact language, derived from measures of blood flow and metabolism,

has catalyzed the spatial categorization of nodes in a complex and distributed language network<sup>49</sup>. The breadth of recruited substrates argues against localizationist accounts of production<sup>81,82</sup> in favor of network-dependent cognitive processes<sup>21,83</sup>, but measuring internodal communication requires fine temporal scale. Electrophysiology directly captures the dynamical behavior of neural substrates with radically improved temporal resolution<sup>40,45,84</sup>. In compiling an extensive dataset of intracranial recordings during picture naming, I am able to integrate cognitive models of speech production with the neurobiology of distributed and interactive cortical networks. I reveal 5 states engaged during picture naming, remarkably concordant with the 5 cognitive stages named in the seminal work by Indefrey and Levelt: visualization, conceptualization, formulation, articulation, and monitoring<sup>8</sup>. Each of these states is dually characterized by timing within a trial-specific transition sequence and by functional network structure comprising directed information exchange.

The state characteristics inform two fundamental and disputed properties of conflicting production models: the seriality of separable cognitive processes and the interactivity between representational levels<sup>3,11,85</sup>. These findings are concordant with a concerted serial propagation of rapidly evolving network dynamics. A consistent sequence of states is observed from picture presentation through articulation and state switching is tightly constrained within pre- and post-formulation periods. Distant interactions, reflecting the unique pattern of information exchange during each state, underline the distributed basis of cognitive processes. Some states (e.g. visualization and monitoring) are directed by singular foci while others feature a balanced distribution of largely reciprocal connections (e.g. formulation). Local interactions are manifest in substrates shared between states – most notably in ventral lateral prefrontal cortex during conceptualization, formulation, and articulation. These reflect a functionally heterogeneous population of neurons contributing to the signal measured in each region. Interactions complement traditional analysis of high-gamma power by providing a fundamentally distinct measure of cortical response. This is exemplified by superior temporal gyrus which produces minimal change in power prior to articulation, but which strongly influences the pre-articulatory

formulation state<sup>86</sup>. These results comprehensively delineate the neurobiology of picture naming in language-dominant cortex, robustly motivate the use of seriality in computational models of speech production, and establish a concrete mechanism for representational interactions in language networks.

This work presents two complementary accounts of temporal dynamics. The first, mean high-gamma activity, is local and rooted in physical space; the second, network connectivity patterns, is global and defined in state space. In a similar manner to piecewise linear approximation of a curve, this second account uses the switching characteristic of the hidden Markov model to approximate the high-dimensional state space of neural dynamics<sup>87,88</sup>. Each state then represents a set of reference dynamics at informative inflections of state space. Fluctuations around these reference dynamics provide a generic mechanism by which to disseminate information in a structured manner throughout complex networks<sup>61</sup>. The pairwise measures of information flow that I quantify are thus an average of interregional exchanges, amalgamating transmissions between small groups of neurons. This perspective integrates distributed interactions<sup>89</sup> with modular cognitive processes<sup>8</sup>.

Neural state sequences comprise a powerful framework by which to model cognition. I provide empirical evidence consistent with this framework; the dynamical model identifies 5 states in speech production. Specifying fewer states results in refolding a pair of states together; specifying more states results in degenerate splitting of the baseline state. Said another way, simplifications of state space produce coarser structures that exhibit functionally similar behavior<sup>90</sup>. A hierarchically organized neural state structure may reveal increasing dynamical detail with improved observation of the system. This behavior suggests that incorporating data from additional regions, latent parameters for patient-specific network variability, and progressively finer-scale cortical recordings could result in further decomposition of the observed states (e.g. formulation might splinter into morphologic, phonologic, and phonetic encoding). Uncovering dynamical systems in the brain provides an improved understanding both of granular processes, such as picture naming, and of cognition more generally.

## NAMING WHAT YOU HEAR

*A lexical semantic hub for heteromodal naming in middle fusiform gyrus (Brain)*

Forseth KJ, Kadipasaoglu CM, Conner CR, Hickok G, Knight RT, and Tandon N

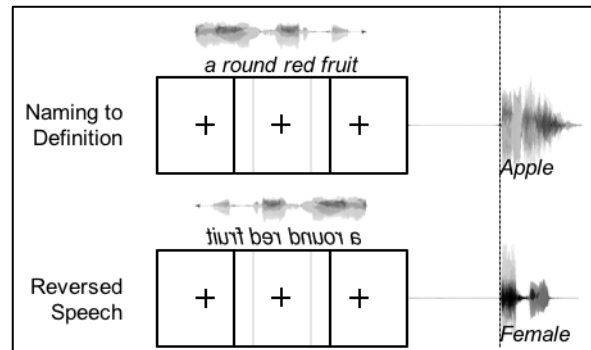


Figure 9: Coherent and reversed auditory stimuli

## BACKGROUND

Semantic memory is the understanding of objects, people, places, and ideas independent of reference to a specific instance<sup>91,92</sup>. This repository of conceptual associations is accessible by multiple lead-in processes that, when used in the service of language, drive lexical retrieval to enable speech production<sup>8</sup>. Semantic memory enables us to fluently name aloud an object that I see or hear described; yet it is easily disrupted, resulting in tip of the tongue phenomena in healthy individuals<sup>93</sup> and pervasive anomias following a variety of brain injuries<sup>94</sup>.

Prior studies of semantic memory have been largely mediated by analysis of disease<sup>47,95–97</sup>, functional imaging<sup>48,98–102</sup>, and non-invasive electrophysiology<sup>103</sup>. Vascular lesions caused by middle cerebral artery strokes largely affect perisylvian cortex<sup>104</sup>, while regions within watershed zones (e.g. ventral temporal cortex) are typically only impaired by global cerebral ischemic effects. Degenerative conditions like semantic dementia affect a diffuse set of cortical structures, challenging the precise localization of neural processes<sup>95</sup>. Experiments in healthy subjects are confronted by anatomical limitations as well: mastoid air cells result in susceptibility artifacts that degrade echo-planar imaging of ventral temporal cortex<sup>105,106</sup>, repetitive transcranial magnetic stimulation is unable to access the ventral pial surface<sup>96</sup>, and magneto-encephalography is relatively insensitive to magnetic fields oriented tangentially to the detectors<sup>107</sup>.

Direct electrocorticography with implanted intracranial electrodes yields millimeter spatial and millisecond temporal resolution of cortical activity. High-gamma activity in these recordings is especially relevant to the study of cognition<sup>22,29,30</sup>. This activity arises from the focal summation of post-synaptic currents coupled with a surge in spike rate<sup>22,23</sup>, indexing local processing<sup>24–26</sup>. High-gamma activity correlates strongly with the fMRI blood-oxygen-level dependent (BOLD) signal and, with its superior temporal resolution, can precisely characterize inter-regional timing<sup>27,28</sup>. Additionally, intracranial electrodes allow for direct electrical stimulation that transiently mimics focal lesions, enabling study of both normal functional activation and multiple lesions in each patient.

In the context of speech production tasks, I hypothesize that various input modalities feeding into the semantic memory system drive a pre-phonologic network hub to achieve lexical retrieval from semantic content. This hub should act as a convergence zone for networks supporting word production. A precise spatiotemporal delineation of cortical regions recruited for heteromodal naming tasks will isolate this lexical semantic convergence hub. I used electrocorticography in a large cohort (n = 64) of patients to categorize cortical responses during cued naming tasks using visual or auditory inputs: picture naming and naming to description. Each experiment was paired with a modality-specific non-semantic control: scrambled images and reversed speech. A subset of patients also underwent fMRI prior to surgery (n = 36) and/or direct cortical stimulation after electrode implantation (n = 30). First, SB-MEMA was used to yield a precise cortical effect estimate of high-gamma and BOLD signals across the population with electrocorticography and fMRI data, respectively. Second, these activity maps were used to direct a region-of-interest analysis that evaluated the timing, magnitude, and spectral profile of the distributed cortical response. Third, naming disruption caused by direct cortical stimulation was integrated across the cohort.

#### SPECIFIC METHODS

Functional images were collected using a gradient-recalled echo-planar imaging sequence with 33 axial slices of 3 mm thickness and an in-plane resolution of 2.75 x 2.75 mm

(TE 30 ms, TR 2015 ms, flip angle  $90^\circ$ ). Stimuli were presented in a block design with two runs of each task (8 blocks each, 136 TR volumes, 20 s of task – picture naming or naming to definition, and 14 s of control – scrambled images or reversed speech)<sup>27</sup>. Stimulus presentation was coincident with the onset of each functional image volume.

Language cortex was mapped as needed for clinical needs. Concurrent monitoring was carried out in all cases to detect any induced seizures. This procedure was conducted in 30 patients (23 with grid electrodes, 7 with depth electrodes). Trains of 50 Hz balanced 0.3 ms period square-waves were delivered to adjacent electrodes for 3-5 seconds during the task<sup>108</sup>. Stimulation was applied using a Grass S88X Stimulator with an SIU (Grass Technologies, West Warwick, RI). At each electrode pair, stimulation was begun at a current of 2 mA and increased stepwise by 1 mA until either an overt phenomenon was observed, after-discharges were induced, or the 10 mA limit was reached. Stimulation sites were defined as positive for language tasks if stimulation resulted in articulation arrest or anomia. Furthermore, stimulation sites causing movement or sensation were separately recorded. Patient responses and behavior were evaluated by clinical experts present in the room during the entire mapping session.

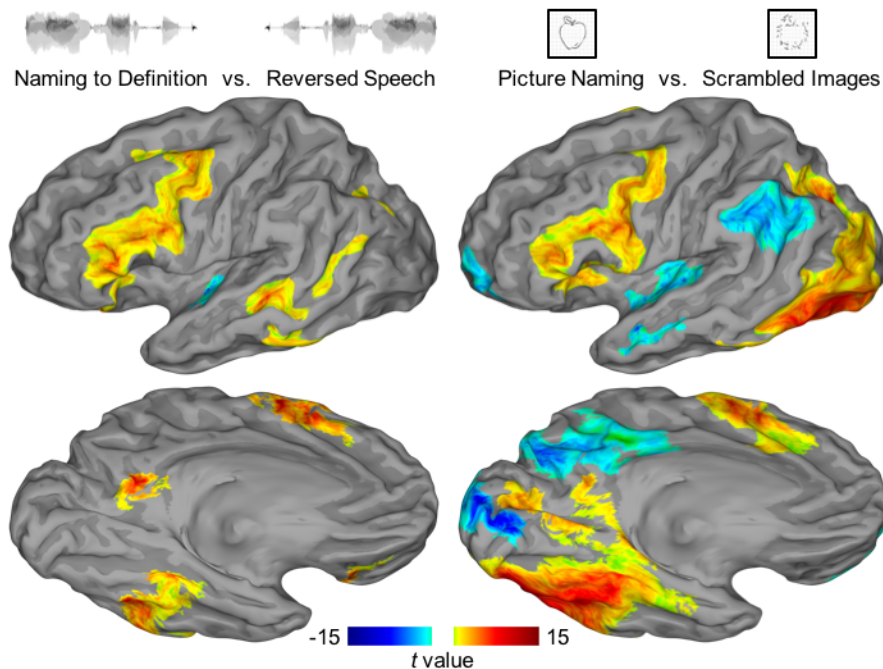


Figure 10: Surface-based group fMRI. Regions surviving a significance threshold ( $P < 0.01$ , corrected) are shown scaled by the model confidence – those with a preference for the semantic condition are shown in warm color and for the control condition in cool color.

## FINDINGS

Two group analyses were performed – one of fMRI data, the other of electrocorticography data – using SB-MEMA. The fMRI maps contrast task vs. control for both auditory and visual stimuli (Figure 10). Electrocorticography maps were computed separately for each condition relative to stimulus onset and articulation, revealing the temporal evolution of activity across the cortex (Figure 11). Finally, a conjunction of task vs. control contrasts from electrocorticography was generated to isolate areas with heteromodal high-gamma power (Figure 12).

SB-MEMA of fMRI in 36 patients (Figure 10) revealed 6 distinct regions that demonstrate significantly enhanced BOLD signal during naming to definition compared to that during reversed speech: posterior middle temporal gyrus, ventral temporal cortex (middle fusiform and inferior temporal gyri), intraparietal sulcus, precuneus, supplementary motor area, and inferior frontal gyrus. BOLD signal was also enhanced during picture naming in occipital cortex – terminating ventrally in middle fusiform gyrus, laterally in lateral occipital cortex, and dorsally in the intraparietal sulcus. Auditory and visual modalities showed overlapping regions of significance in middle fusiform gyrus, intraparietal sulcus, supplementary motor area, and inferior frontal gyrus. These results were concordant with the conjunction of contrasts generated with electrocorticography (Figure 12).

SB-MEMA of electrocorticography in 61 patients (Figure 11) during the auditory task revealed that early auditory cortex was active in the 250 ms after stimulus onset during both naming to definition and reversed speech. Naming to definition resulted in a greater magnitude and volume of activation in auditory sensory cortex than did reversed speech, consistent with the greater processing demands of phonologically structured stimuli<sup>109–112</sup>. High-gamma power was concentrated in the early window (-1000 to -500 ms) preceding articulation at posterior middle temporal gyrus, ventral temporal cortex (middle fusiform and parahippocampal gyri), intraparietal sulcus, precuneus, middle frontal gyrus, and inferior frontal gyrus (pars triangularis and opercularis). High-gamma power peaked in the late window (-500 to 0 ms) preceding articulation

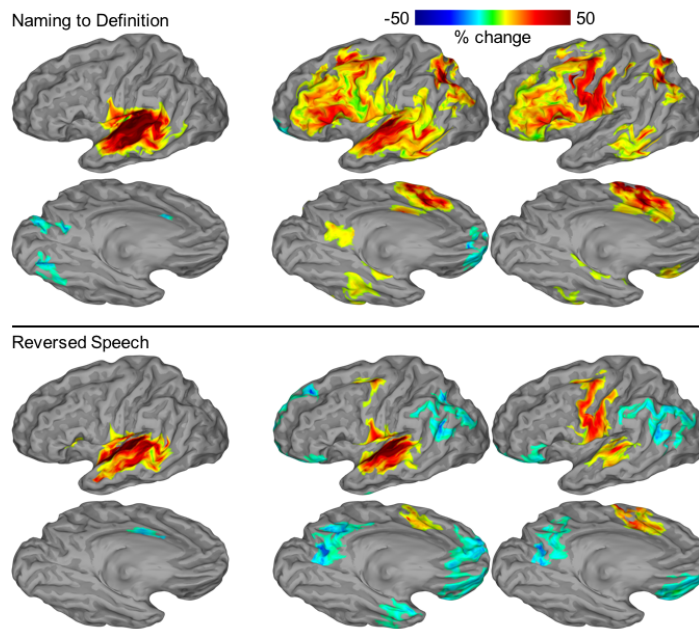


Figure 11: Surface-based group electrocorticography. Columns are organized by time windows: post-stimulus (250 ms width) or pre-articulation (500 ms width).

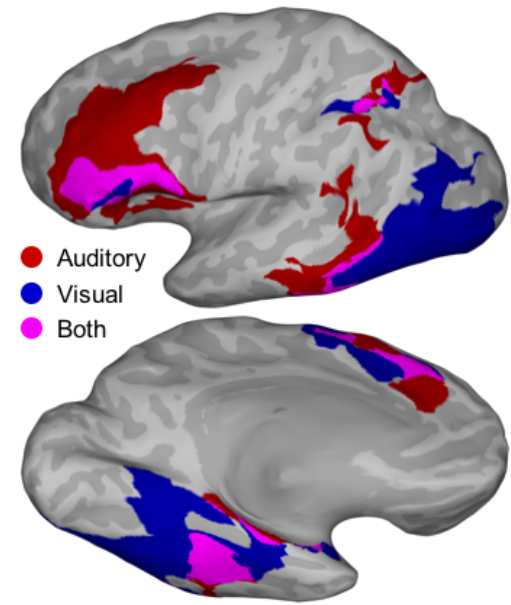


Figure 12: Conjunction of SB-MEMA maps

at the supplementary motor area. Notably, no significant high-gamma power was noted in the anterior fusiform gyrus or ventral temporal pole for any time window during naming to definition, despite significant coverage in these regions from 29 patients.

To isolate activity specific to common semantic features from activity related to modality-specific sensory processing or to articulation, I performed a conjunction of two contrasts: [naming to definition vs. reversed speech] + [picture naming vs. scrambled images] in the 1000 ms preceding articulation. Activity in 4 regions was enhanced for both auditory and visual semantic contrasts: the middle fusiform gyrus, intraparietal sulcus, supplementary motor area, and inferior frontal gyrus (Figure 12). Notably, there was no significant intersection of these heteromodal semantic contrasts in posterior middle temporal gyrus – the activity here was uniquely enhanced during the naming to definition task, while activity in lateral occipital cortex was uniquely enhanced during the picture naming task.

High-gamma power derived from electrocorticography provides correlative – but not causal – evidence for the engagement of specific regions in a cognitive process; in contrast, transient lesions induced by direct cortical stimulation provide a direct causal measure of



cognitive disruption. I compared the functional maps of lexical semantic processing from fMRI and electrocorticography with language disruption from direct cortical stimulation in 30 of these patients (Figure 13A), 23 with grid electrodes and 7 with depth electrodes.

Stimulation positive sites (i.e., sites where stimulation caused a functional deficit) specific for naming to definition were principally localized to lateral temporal cortex (Figure 13B). Those specific for picture naming were predominantly located in posterior ventral temporal cortex (Figure 13C). Sentence repetition was primarily disrupted by peri-Sylvian stimulation (Figure 13D). Motor positive sites were localized to pars opercularis and sensorimotor cortex (Figure 13E).

To isolate cortical regions where stimulation disrupted semantic processing irrespective of modality, I performed a conjunction of these functional maps: [positive for picture naming and naming to definition] + [negative for sentence repetition and sensorimotor effects] (Figure 13F). This revealed 2 loci. The first locus was the middle fusiform gyrus, which has also been characterized as the basal temporal language area<sup>113,114</sup>. This region was well-aligned with the corresponding functional locus observed in both fMRI (Figure 10) and electrocorticography (Figure 11). The second locus was posterior middle temporal gyrus, overlapping with the corresponding functional locus defined by the contrast of naming to definition and reversed speech; however, this region did not show increased functional activity for picture naming or scrambled images in any pre-articulatory time window.

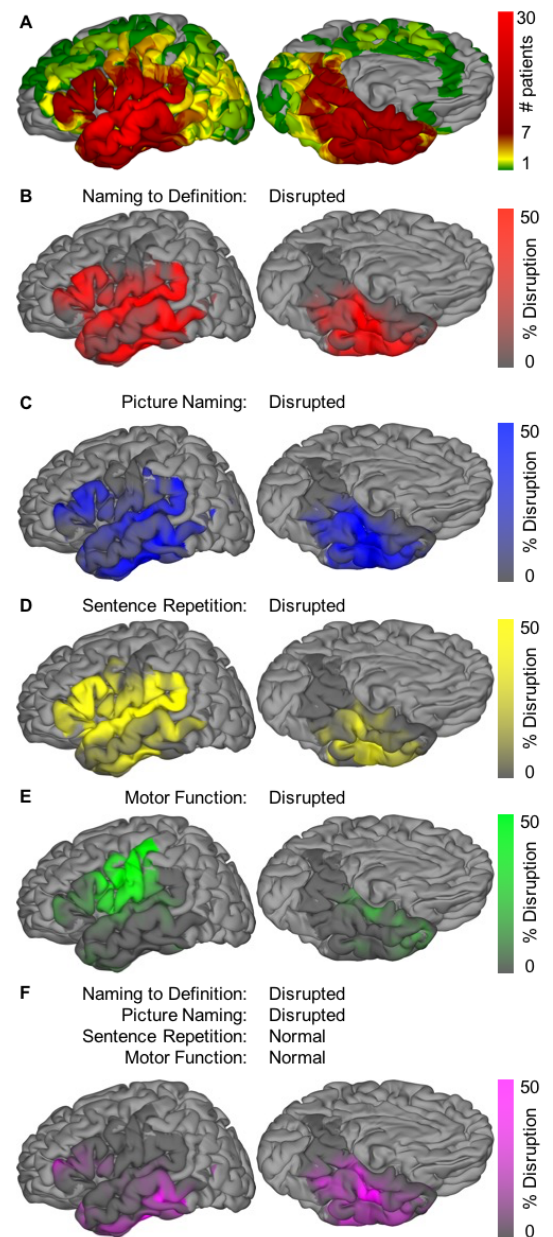


Figure 13: Surface-based group direct cortical stimulation with several distinct tasks.

## DISCUSSION

There is accumulating evidence for the involvement of ventral temporal cortex in semantic memory from neuropsychological studies<sup>47,97</sup>, electrical stimulation of cortex<sup>113,114</sup>, positron emission tomography<sup>48,98–101</sup>, magnetoencephalography<sup>103</sup>, and intracranial evoked response potentials<sup>115,116</sup>. These studies broadly implicate the entire ventral surface from the temporal pole through the fusiform gyrus. A consensus on the focal neurobiological substrate underlying a lexical semantic hub in ventral temporal cortex has yet to emerge<sup>117</sup>. Furthermore, a number of recent influential reviews disregard this region and assign semantic function solely to lateral perisylvian regions<sup>118–121</sup>.

I studied object naming using three complementary methodologies: fMRI (n = 36), electrocorticography (n = 64), and direct cortical stimulation (n = 30) during both auditory verbal and visual nonverbal stimuli, each paired with a modality-specific nonsense control. Large-scale integrated electrocorticography combining both surface and depth electrodes is particularly well-suited for the study of distributed language networks given the complete coverage of the cortical surface with high spatiotemporal resolution. These data provide compelling large-scale evidence for audio-visual cortex supporting semantic cognition in the middle fusiform gyrus.

The ventral temporal lobe is especially at risk in surgical approaches for mesial temporal lobe epilepsy. A major advantage of newer minimally invasive approaches such as laser interstitial thermal ablation is the reduction of cognitive deficits, particularly naming, for epilepsy in the language dominant hemisphere. The fact that the middle fusiform cortex is typically spared in such approaches supports its role in semantic memory<sup>122,123</sup>.

The inferior frontal gyrus is widely accepted to be involved in semantic selection and phonological processing functions<sup>112,124,125</sup>, which are thought to be segregated along an anterior-posterior axis<sup>126,127</sup>. In this study, population maps from both fMRI and electrocorticography revealed that the inferior frontal gyrus was the frontal lobe region coactive for semantic contrasts in both sensory modalities. Furthermore, disruption of this region with direct cortical stimulation resulted in heteromodal naming deficits. These deficits were also seen with stimulation during a

sentence repetition task, emphasizing the role of the inferior frontal gyrus as an interface between lexical semantic and articulatory networks. These results are consistent with pars triangularis exerting top-down control over the semantic network.

The ventral temporal cortex has been a region of interest in semantic memory since early studies of semantic dementia identified patients with intact sensory processing and deficits in conceptual knowledge<sup>47</sup>. Studies in non-human primates reveal that large numbers of auditory and visual fibers converge at the temporal pole<sup>128,129</sup>, suggesting that this region may also be an integrative locus. Subsequently, studies of direct cortical stimulation revealed that disruption of the basal temporal language area – defined as fusiform gyrus within 30-70 mm of the temporal pole – results in speech arrest<sup>114</sup>. This effect was also observed following stimulation of parahippocampal gyrus and anterior inferior temporal gyrus<sup>113</sup>. Positron emission tomography studies provided the first observations of functional activity in the ventral temporal cortex to both visual and auditory stimuli<sup>48,100,103,115,130</sup>. The atrophy and hypometabolism observed in semantic dementia has been localized to the anterior fusiform gyrus<sup>95,96</sup>.

I identified four regions with activity common to semantic conditions: middle fusiform gyrus, the intraparietal sulcus, supplementary motor area, and inferior frontal gyrus. The timing of supplementary motor area activity suggests a role in early articulatory planning while inferior frontal gyrus interfaces between lexical semantic and phonological networks. This suggests that the remaining two regions – middle fusiform gyrus and the intraparietal sulcus – are critical for semantic encoding. The spatial extent of significant BOLD signal found with fMRI and significant high-gamma power found with electrocorticography were well-aligned. Direct cortical stimulation of ventral temporal cortex revealed a middle fusiform region that produced heteromodal naming deficits. This region was slightly anterior to that identified by functional activity, but both were well within the bounds of the basal temporal language area<sup>113,114</sup>. The functional activity identified by fMRI and electrocorticography showed less substantial heteromodal overlap in the intraparietal sulcus than was found in the middle fusiform gyrus. This could be due to the focus in this study on common object naming as opposed to action naming, which may favor parietal cortex<sup>45,83,131</sup>.

The data does not support that the activity in middle fusiform gyrus and the intraparietal sulcus represents ancillary “visual imagery”<sup>132–134</sup>. First, the timing of activity in these regions shows that they engage only at the end of the auditory stimulus. Despite the serial presentation of evocative descriptors in the naming to definition task, no visual imagery occurs until its completion. Second, direct stimulation to this region causes heteromodal naming deficits – incongruous for a region performing a facultative process.

Several prominent theories hold that posterior middle temporal gyrus is a central site for semantic representation<sup>119</sup>, a secondary region engaged with inferior frontal gyrus in semantic control<sup>127,135</sup>, or a lexical interface<sup>112</sup>. Using both fMRI and electrocorticography, only naming to definition against reversed speech produced a significant semantic contrast in posterior middle temporal gyrus; however, direct cortical stimulation at posterior middle temporal gyrus disrupted both naming to definition and picture naming. These results are most consistent with lexical processing at posterior middle temporal gyrus.

Neuropsychological studies suggest a key role of the temporal pole in heteromodal semantic processing. With three distinct methods, I demonstrate no semantic-specific activity in the ventral temporal pole for either naming to definition or picture naming. First, no significant BOLD signal is reported in either fMRI semantic contrast. Second, electrocorticography further reveals that the observed artifact in ventral temporal pole is well-aligned with articulation and best explained by temporalis muscle movement<sup>136,137</sup>. Third, direct cortical stimulation of the temporal pole does not disrupt either naming to definition or picture naming. This evidence strongly suggests that the temporal pole does not support semantic memory for objects.

Semantic memory underpins our understanding of objects, people, places, and ideas. These results demonstrate that auditory and visual features processed in distinct sensory cortices converge in a shared lexical semantic network – including middle fusiform gyrus – prior to articulation. This underappreciated locus of semantic processing is at risk in resections for temporal lobe epilepsy as well as in trauma and strokes that affect the inferior temporal cortex – it may explain the range of anomie states seen in these conditions.

## PREDICTION IN PERCEPTION AND PRODUCTION

*Language prediction mechanisms in human auditory cortex (Nature Communications)*

Forseth KJ, Hickok G, Rollo PS, and Tandon N

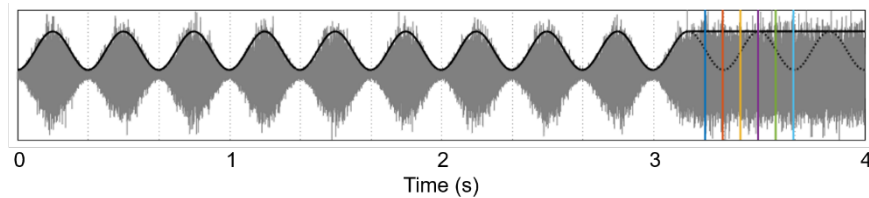


Figure 14: Sample 3 Hz modulated white noise stimulus with buried pure tone

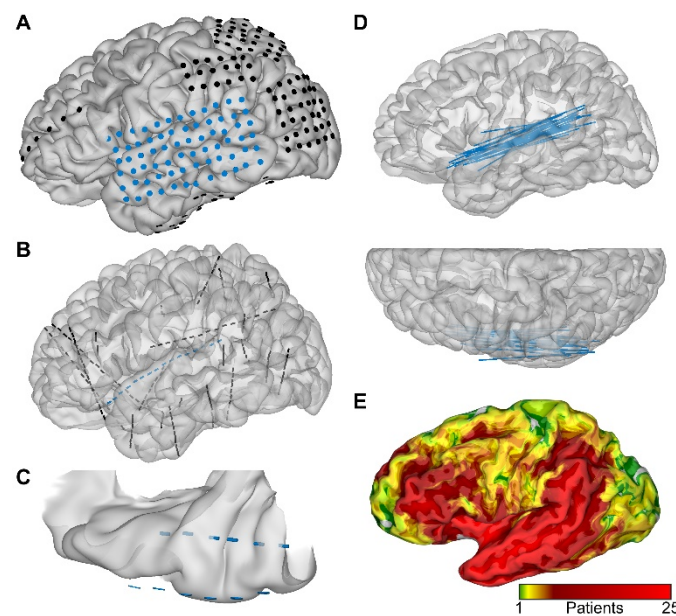
## BACKGROUND

Humans efficiently extract speech information from noisy acoustic signals and segment this into meaningful linguistic units. This complex and poorly understood process is fluidly accomplished for a wide range of voices, accents, and speaking rates<sup>138</sup>. Given the quasi-periodic and hierarchical structure of speech<sup>139</sup>, the computational load associated with its decoding can be reduced by utilizing temporal prediction<sup>140</sup>. Anticipating the arrival of salient acoustic information could enable optimal potentiation of neural networks<sup>141</sup> and discretization of the continuous signal into linguistic elements<sup>142–144</sup>. This perspective, the active sensing framework<sup>145</sup>, anticipates interactions between bottom-up sensory input and top-down predictive modulation of neuronal dynamics. Evidence for cortical entrainment – the synchronization of extrinsic quasi-periodic stimuli and intrinsic neural activity – in the auditory domain<sup>146</sup> and during speech perception<sup>147–151</sup> has driven speculation that cortical oscillations may enable temporal prediction. In addition, speech production is also thought to rely upon predictive mechanisms. Several prominent models require that the brain anticipate the sensory consequences of speech<sup>16,21</sup> and this central tenet has been buttressed by strong evidence<sup>19,152,153</sup>. It remains unclear, however, which levels of auditory cortical processing are involved in this process and where such mechanisms are instantiated in the cortex.

I elucidate the mechanisms by which auditory cortex anticipates rhythms and, further, whether such mechanisms may extend to optimize the processing of quasi-rhythmic acoustic input during language perception. To investigate the neurobiology of prediction in early auditory

cortex, I use two tasks: amplitude-modulated white noise and spoken naming to definition. The white noise stimulus comprises a rhythmic pattern followed by a constant amplitude interval; patients are tasked with detecting the occurrence of a peri-threshold tone in the latter interval. Neural encodings of prediction should uniquely persist during the latter period, while other cortical signals – including evoked response potentials and envelope tracking – would be limited to the rhythmic period. Next, I examine the cortical response to natural language speech for these same encoding signatures. Finally, to reveal causal involvement of specific neuroanatomic substrates, I apply chronometric stimulation to targeted structures during naming.

The characterization and localization of predictive mechanisms for language function requires a methodology with high temporal resolution, fine spatial resolution, and direct access to neuronal populations in human early auditory cortex. I use large-scale intracranial recordings (37 patients), focusing on depth electrodes placed along the anteroposterior extent of the supratemporal plane (Figure 15). This innovative surgical approach enables simultaneous recordings from planum polare, Heschl’s gyrus, and planum temporale. These experiments yield crucial insights into the rapid, transient dynamics of predictive timing and predictive coding – prediction of when and what<sup>154</sup> – in Heschl’s gyrus and planum temporale.



*Figure 15: Supratemporal depth probe trajectory and coverage map. (A) Grid electrodes (B) Depth electrodes (C) Dorsal view of the supratemporal plane with depth electrodes in Heschl’s gyrus and planum temporale, as well as the closest grid electrodes. (D) All anteroposterior supratemporal trajectories superimposed in Talairach space (E) Recording zone density from all electrodes (37 patients, 838 grid electrodes, 6669 depth electrodes).*

## SPECIFIC METHODS

The white stimulus comprised two periods. In the first, wideband Gaussian noise was modulated (3 Hz, 80% depth) for 3 seconds. In the second, the modulation waveform ended on the cosine phase of the next cycle to yield 833 ms of constant-amplitude noise. Furthermore, 50% of trials featured a peri-threshold tone (1 kHz, 50 ms duration, 5ms rise-decay time) that was presented at one of 5 temporal positions and at an amplitude level from 1 of 3 values. The temporal positions were separated by a quarter-cycle of the modulation frequency beginning with the constant-amplitude noise. The amplitude levels covered a range of 12 dB. On each trial, the patient was required to indicate via a key press whether a tonal signal was present during the unmodulated segment of the masking noise. All 37 patients each completed 100 trials. 7 patients also underwent testing with stimulus modulation frequencies of 5 and 7 Hz.

Non-negative matrix factorization (NNMF) is an unsupervised clustering algorithm<sup>155</sup>. This method expresses non-negative matrix  $\mathbf{A} \in R^{m \times n}$  as the product of class weight matrix  $\mathbf{W} \in R^{m \times k}$  and class archetype matrix  $\mathbf{H} \in R^{k \times n}$ , minimizing:

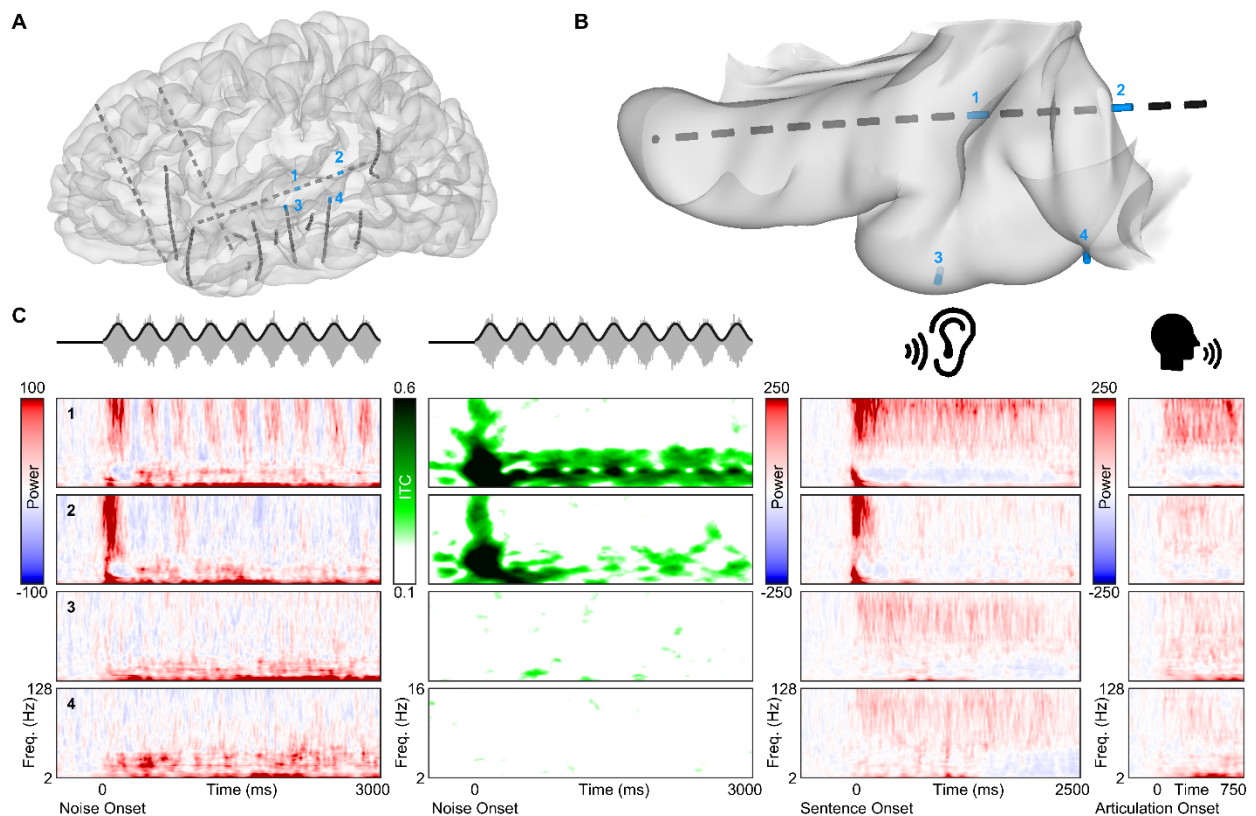
$$\|\mathbf{A} - \mathbf{WH}\|_F^2$$

The factorization rank  $k = 2$  was chosen for all analyses in this work. Two types of inputs were separately factorized: mean high-frequency power and low-frequency phase ITC. These features were calculated for the  $m$  electrodes in the supratemporal plane at  $n$  time points. Factorization thus generated a pair of class weights for each electrode and a pair of class archetypes – the basis function for each class. Class bias was defined as the difference between the class weights at each electrode. Response magnitude was defined as the sum of class weight magnitudes at each electrode.

The supratemporal plane was clinically evaluated with stimulation mapping in 3 patients. Concurrent electrocorticographic monitoring was carried out in all cases to detect any induced seizures. Trains of 50 Hz balanced 0.3 ms period square waves were delivered to adjacent electrodes for 3 seconds during the task<sup>108</sup>. Stimulation was applied using a Grass S88X

Stimulator with a SIU (Grass Technologies, West Warwick, RI). At each electrode pair, stimulation was begun at a current of 2 mA and increased stepwise by 1 mA until either an overt phenomenon was observed, after-discharges were induced, or the 10 mA limit was reached. Stimulation sites were defined as positive for language tasks if stimulation resulted in articulation arrest or anomia. Furthermore, stimulation sites causing movement or sensation were separately recorded. Patient responses and behavior were evaluated by clinical experts present in the room during the entire mapping session.

One patient underwent a pair of chronometric stimulation experiments, each performed separately at HG/TTS and PT. In the first experiment, the same clinical stimulation protocol was applied during sentence repetition at either the sentence onset or its conclusion. In the second experiment, patients attempted the auditory-cued naming task described above while single



**Figure 16:** Cortical responses in a single patient to low- and high-level auditory stimuli. **(A)** Location of all electrodes implanted in this patient relative to a model of the pial surface. **(B)** The supratemporal plane, isolated from the cortical model and viewed from above. All electrodes along the supratemporal depth probe are shown, as well as the two most superficial electrodes from probes in lateral superior temporal gyrus. Four electrodes are highlighted in blue with labels. **(C)** Neural responses at the four highlighted electrodes: (1) HG/TTS, (2) PT, (3) mediolateral superior temporal gyrus, (4) posterolateral superior temporal gyrus. The first two columns show the spectral decomposition of the power and phase response, respectively, during white noise listening. The latter two columns show the power response during natural language listening and production, respectively.

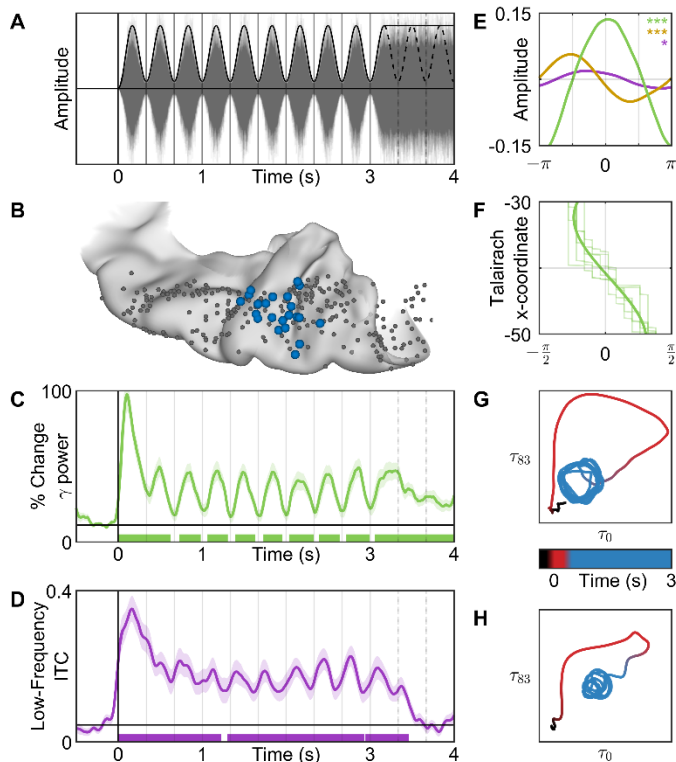


balanced 0.3 ms period pulses were triggered throughout the spoken description. These triggers were either cued by acoustic edges (HG/TTS, 31 trials; PT, 21 trials) or, as a control, uniformly distributed (18 trials). The latter condition was matched for total current delivered during the entirety of each sentence. Patient performance was quantitatively assessed in both experiments by either repetition or naming success, respectively.

## FINDINGS

I observed a sustained multispectral response of early auditory cortex during rhythmic amplitude-modulated white noise (80% depth at 3 Hz for 3 seconds, then constant amplitude for 1 second). Heschl's gyrus and the transverse temporal sulcus (HG/TTS; Figure 16A,B Electrode 1) encoded stimulus features in high-frequency power and low-frequency phase (Figure 16C). These results were robust across the patient cohort, both in high-gamma power (Figure 17C) and in low-frequency phase (Figure 17D). Following a low-latency high-magnitude broadband response to stimulus onset, HG/TTS exhibited a sustained response to subsequent acoustic pulses. Phase space trajectories of high-gamma power (Figure 17G) and low-frequency phase (Figure 17H) revealed three clearly dissociable states corresponding to rest (pre-stimulus), stimulus onset, and sustained activity (beginning with the second pulse). Electrodes in lateral superior temporal gyrus (Figure 16A,B Electrodes 3 & 4) showed no evidence of a sustained response to the white noise stimulus (Figure 16C). In contrast, a sustained response was recorded in all patients with a supratemporal depth probe in the language dominant hemisphere ( $n = 22$ ). Patients with homologous electrodes in the language non-dominant hemisphere ( $n = 5$ ) demonstrated an equivalent sustained response. This was also observed for faster modulations of the temporal envelope (5 Hz and 7 Hz).

High-gamma, beta, and low-frequency power together yielded a frequency-multiplexed encoding of acoustic envelope (Figure 17E). High-gamma power was in-phase with the stimulus, beta power was resynchronized at the trough of the stimulus, and low-frequency power was modulated by the rising slope of each pulse – the acoustic edge. These distinct and asymmetric bandlimited responses may represent separable cortical processes each engaged by the



**Figure 17: Cortical response to rhythmic white noise.** (A) 3 Hz amplitude-modulated white noise stimulus. (B) The most active electrode (blue) was selected from all electrodes (grey) in each patient with a supratemporal depth probe ( $n = 26$ ). These were used for the following analyses. (C) Average percent change in high-gamma ( $p < 0.01$ ). (D) Average absolute change in low-frequency inter-trial coherence from a pre-stimulus baseline. (E) Average amplitude of low (purple), beta (yellow), and high-gamma (green) frequencies relative to stimulus phase during the sustained response (pulses 2-9) demonstrating frequency-multiplexed encoding of acoustic envelope ( $* p < 0.05$ ,  $** p < 0.01$ ,  $*** p < 0.001$ ). (F) Spatial distribution of peak high-gamma power timing relative to stimulus phase demonstrates a traveling wave (velocity 0.1 m/sec) that begins medially at the insular boundary (top) and progresses to the lateral edge (bottom). The mean wave (dark lines) is superimposed over the wave at each pulse (light lines). (G, H) Phase space trajectory at a quarter period delay (83 ms) in high-gamma power (top) and low-frequency ITC (bottom). Time indicated by color: pre-stimulus baseline (black), red (first acoustic pulse), blue (pulses 2-9).

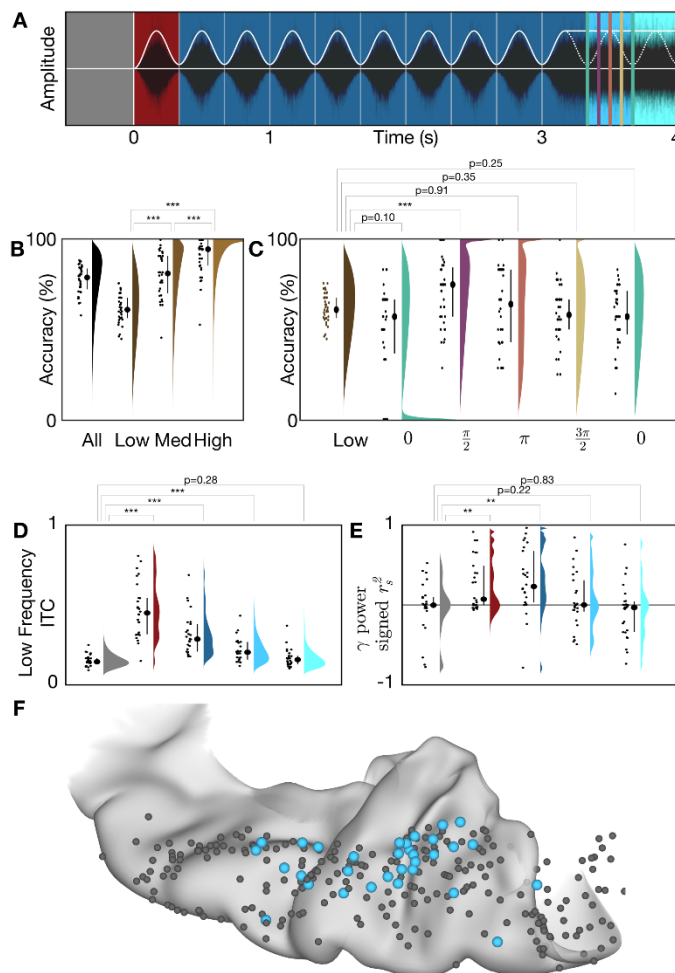
stimulus<sup>156</sup>. Low-frequency phase was reset at the acoustic edge. Spectral decomposition of the low-frequency phase response (Figure 16C) demonstrated that phase reset was constrained to the theta band. Phase reset was not observed in beta or high-gamma frequencies.

During the sustained response, I resolved the spatiotemporal topography of high-gamma power along the mediolateral extent of HG/TTS (Figure 17F). A traveling wave of cortical activity coincided with each acoustic pulse, beginning at medial HG/TTS adjacent to the inferior circular sulcus of the insula and propagating laterally across the supratemporal plane to the lip of the lateral fissure. Each wave began approximately 80 ms before the acoustic pulse maximum and ended approximately 80 ms afterwards, traversing HG/TTS at a speed of 0.1 m/s. While such spatial organization of neural activity is thought to be important for optimizing cortical computations<sup>157</sup>, prior reports of traveling waves in humans have been confined to lower frequencies<sup>158</sup> and sleep studies<sup>159</sup>. The wave observed here is considerably slower than would be expected as a simple consequence of the cochlear onset latency gradient<sup>160</sup>.

Patient performance in the tone detection task yielded evidence for the extension of the sustained response to subsequent rhythmic prediction (Figure 18A). Accuracy increased with

tone intensity, confirming that patients were engaged and that detectability was limited by the masking noise (Figure 18B). I then examined the most challenging condition – low intensity tones – for a variation in detectability modulated by temporal position (Figure 18C). If the rhythmic noise had no lasting effect on perceptual threshold, the detection rate would be equal across temporal positions. Instead, I found that detection was uniquely improved for the second temporal position, corresponding to the rising slope of the first missing pulse. Notably, this was the same acoustic feature – the rising edge – encoded by low-frequency phase reset in HG/TTS.

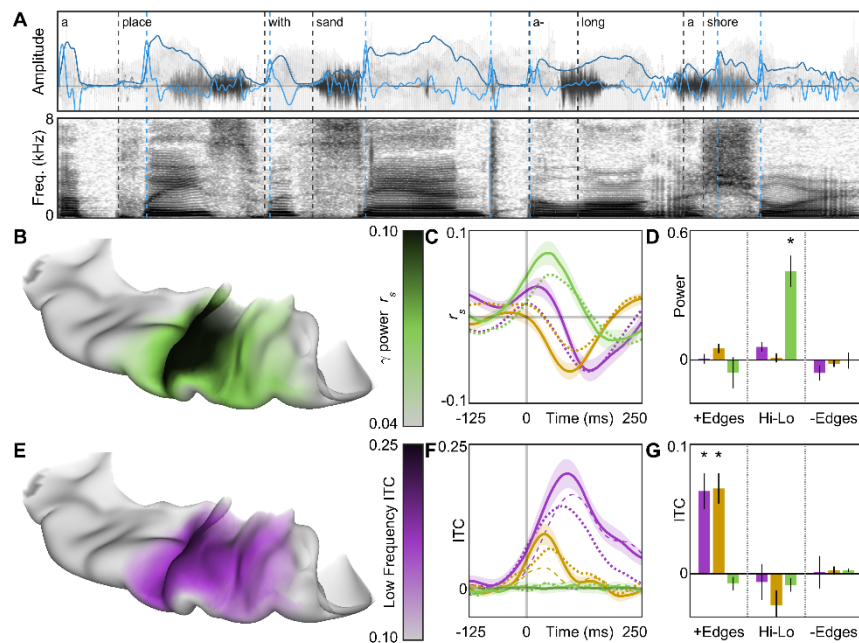
To isolate neural mechanisms supporting prediction in HG/TTS, I evaluated the persistence of the sustained response signature to a 3 Hz acoustic envelope after the stimulus rhythm ceased (Figure 18A). Low-frequency phase maintained the sustained state for one cycle after the last acoustic pulse (Figure 18D); by the second cycle, the temporal organization of cortical phase was not significantly distinct from pre-stimulus baseline. In contrast, the



**Figure 18: Low-frequency phase in early auditory cortex shows evidence of predictive encoding. (A)** The stimulus was divided into 5 intervals: baseline (grey), onset (red), sustained (dark blue), early prediction (medium blue), and late prediction (light blue). Crucially, there is no modulation of white noise amplitude in either of the prediction intervals. Tones were presented in 50% of trials at 1 of 3 intensities and 1 of 5 temporal delays. **(B)** Performance accuracy at each intensity level, grouped across delay categories. For each condition, the raw data (left), interquartile range (middle), and kernel density estimate (right) are shown (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). **(C)** Accuracies in the low-intensity condition separated by temporal delay. Behavioral performance was uniquely increased at the first “missing” acoustic pulse edge – coincident with the phase reset we observe in neural low-frequency response. **(D, E)** The same electrode group shown in Figure 1B was used for the following analyses. Violin plots demonstrating engagement of low-frequency phase and high-gamma power during each interval. The sustained response in low-frequency phase was measured as average ITC; in high-gamma power, it was measured as a signed  $r^2$  from the Spearman’s correlation with a 3 Hz sine wave. **(F)** All electrodes on the supratemporal plane ( $n = 349$ ) were evaluated for a significant sustained response during the rhythmic stimulus followed by low-frequency phase reset in the first prediction interval. Those with a significant predictive effect are shown in blue ( $n = 36$ ), predominantly found in HG/TTS.

relationship between high-gamma power and the acoustic envelope did not carry predictive information in either cycle after the last acoustic pulse (Figure 18E). Thus, prediction in early auditory cortex (Figure 18F) is best modeled by low-frequency phase reset at acoustic edges. I validated that this predictive effect is not an artifact of filter choice by replicating these findings with a variety of filter implementations. This neural mechanism is engaged within a single cycle of a rhythmic acoustic stimulus and remains active for at least one cycle afterwards. Such a neuro-computational solution for coupled perception and prediction provides a neurobiological basis for cognitive models of speech perception<sup>142–144,154</sup>.

In a second experiment, patients ( $n = 25$ ) named common objects cued by short spoken descriptions (e.g. they heard “a place with sand along a shore” and articulated “beach”)<sup>40</sup>. For each sentence (Figure 19A), I extracted a pair of key features suggested by the analysis of rhythmic white noise: acoustic envelope and edges. The former describes the instantaneous

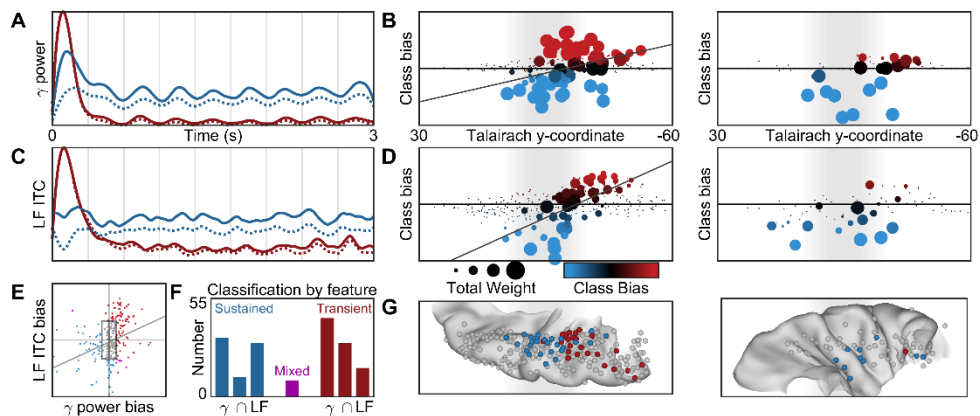


**Figure 19:** Edge detection and envelope tracking during natural language speech occurs focally in early auditory cortex. **(A)** Patients ( $n = 25$ ) listened to short sentences describing common objects. Two features were extracted: acoustic envelope (light blue) and acoustic edges (dark blue). **(B)** The peak lagged Spearman's correlation between acoustic and high-gamma envelopes. **(E)** The average low-frequency ITC following an acoustic edge. **(C)** Acoustic envelope cross correlated with low-frequency (purple), beta (yellow), and high-gamma (green) amplitudes. The effect size was slightly reduced for reversed speech (dotted lines). **(F)** Acoustic edges were better encoded by low-frequency phase than syllabic onsets (dashed lines). **(D,G)** Power and phase encodings of discrete acoustic signal events: the difference between 50 ms pre- and post-edge windows, as well as the difference between 50 ms windows centered on envelope peaks and valleys (\*  $p < 0.001$ ).

amplitude of speech, while the latter demarcates moments of rapid amplitude gain. I evaluated the engagement of neural substrates with a sustained response to the white noise stimulus during natural language speech. The cortical encoding of the speech envelope (Figure 19B) and of edges (Figure 19E) was localized to HG/TTS – the same supratemporal region that exhibited a sustained response and predictive signature for the white noise stimulus. Power in HG/TTS was significantly correlated with the acoustic envelope of speech (low-frequency,  $r_s = -0.0620$ ,  $p < 10^{-3}$ ; beta,  $r_s = -0.0632$ ,  $p < 10^{-3}$ ; high-gamma,  $r_s = 0.0738$ ,  $p < 10^{-3}$ ; Figure 5C) at a frequency-specific delay (low-frequency, 135 ms; beta, 95 ms; high-gamma, 45 ms; Figure 19C). Low-frequency phase organization in HG/TTS was significantly increased during the 125 ms following acoustic edges in speech ( $p < 10^{-3}$ ; Figure 19F). Furthermore, it was significantly greater following acoustic edges than following syllabic onsets ( $p = 0.0072$ ; Figure 19F) – a similar characteristic, but derived from and specific to speech. The correlation between high-gamma activity and the speech envelope was significantly stronger at the electrodes that best tracked the white noise envelope (Figure 17A) than at neighboring electrodes ( $p < 10^{-3}$ ). Similarly, the increase in low-frequency phase organization following acoustic edges in speech was significantly greater at electrodes demonstrating a predictive effect in the white noise task (Figure 18F) than at their neighbors ( $p < 10^{-3}$ ). These findings are concordant with the frequency-multiplexed encoding of acoustic envelope and the low-frequency phase reset at acoustic edges observed during the white noise stimulus. The neural response to acoustic edges was preserved during reversed speech, emphasizing the sublexical nature of this process.

Natural language speech recruited a much broader set of neuroanatomic substrates than white noise, including planum polare, lateral superior temporal gyrus, and superior temporal sulcus. In a patient with both surface grid and depth electrodes, only speech induced significant activity in the lateral temporal grid electrodes. This higher-order auditory cortex is presumably engaged in the processing of higher-order features (e.g. phonemes<sup>161</sup>).

Immediately posterior to HG/TTS in the planum temporale (PT; Figure 16B; Electrode 2), a distinct functional region generated a transient response to white noise (Figure 16C). This



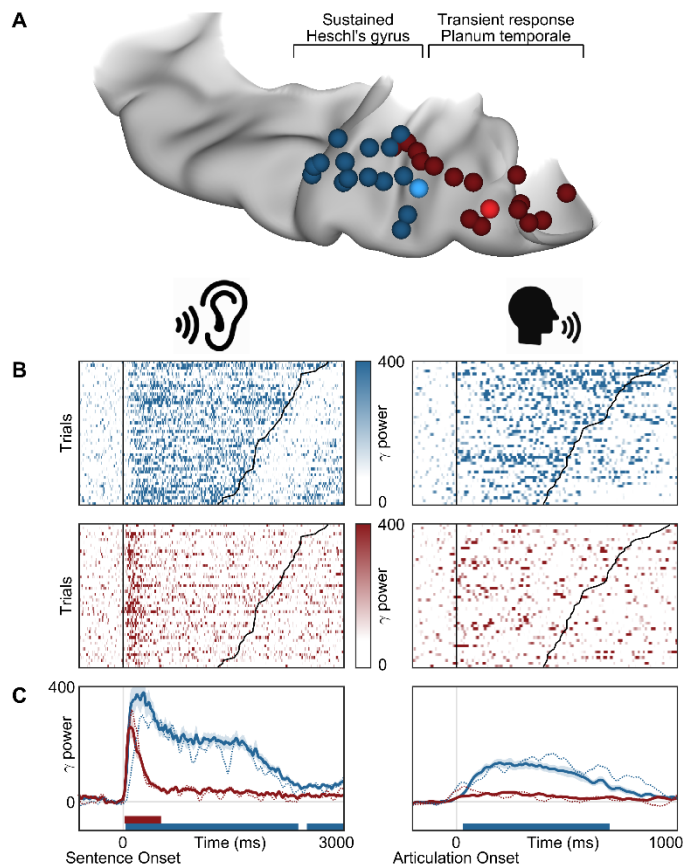
**Figure 20:** Supratemporal responses ( $n = 349$  electrodes) classified with 2-basis non-negative matrix factorization. **(A)** High-gamma power identifies a sustained (blue) and transient (red) response: normalized basis functions (dotted line) and the normalized group-average response for the top 10% of electrodes in each class (solid line). **(B)** Spatial distribution of activation (sum of class weights; point size) and bias (difference in class weights; point color) reveals anteroposterior gradient of functional response. The left panel shows electrodes in language dominant cortex; the right, in language nondominant cortex. **(C, D)** Separately, low-frequency ITC also revealed sustained and transient responses with the same spatial distribution. **(E)** The class bias determined by high-gamma power and low-frequency ITC analyses were significantly correlated. **(F)** Class biases greater than a value of 10 generated discrete classifications: sustained ( $n = 74$ ), transient ( $n = 90$ ), or mixed ( $n = 9$ ). **(G)** Electrode classifications are shown on a standard supratemporal atlas, demonstrating a clear functional split between Heschl's gyrus and planum temporale in language dominant cortex.

region featured a high-magnitude increase in high-gamma power accompanied by broadband low-frequency phase reset that returned to pre-stimulus baseline activity after a single acoustic pulse. I separated this transient response the sustained response using non-negative matrix factorization – an unsupervised clustering algorithm uninformed by anatomical position – across all supratemporal electrodes ( $n = 349$ , Figure 6A,C). This analysis revealed a distinct anteroposterior response gradient from sustained activity in HG/TTS to transient activity in PT (Figure 20B,D,G). This spatial distribution was significant for both high-gamma power (Figure 20B;  $r_s = 0.4101$ ,  $p < 10^{-4}$ ) and low-frequency phase (Figure 20D;  $r_s = 0.7356$ ,  $p < 10^{-16}$ ). Classification by both measures were strongly correlated (Figure 20E;  $r_s = 0.4188$ ,  $p < 10^{-6}$ ); only 9 of 349 electrodes showed a mixed classification (i.e. sustained bias in high-gamma power with transient bias in low-frequency phase, or the reverse; Figure 20F). The sustained response was primarily characterized by *either* high-gamma power ( $n = 33$  electrodes) or by low-frequency phase ( $n = 30$  electrodes), with only 11 electrodes engaging both measures. This effect was confirmed with analysis of the Kullback-Leibler divergence from a uniform (e.g. non-modulated) activity response. The sustained response was noted in both language dominant and non-dominant cortex, but the transient response was limited to dominant cortex (Figure 20G).



The spatial topology of early auditory cortical responses was further elucidated within a single patient who underwent two separate implants: one with surface grid electrodes and another with depth electrodes. Strong sustained encoding in HG/TTS and a robust transient response in PT were observed at electrodes along the supratemporal depth probe, but not at any subdural electrodes directly overlying superior temporal gyrus. In contrast to prior work using only surface grid electrodes<sup>162</sup>, this unique case indicates that the sustained and transient responses to sublexical features are selectively encoded in early auditory cortex – not in lateral superior temporal gyrus.

I compared neural activity in both HG/TTS and PT during listening and speaking – externally and internally generated speech. In each patient with a supratemporal depth probe, the pair of electrodes with the strongest sustained and transient responses were identified during the rhythmic white noise condition. These criteria selected electrodes in HG/TTS and PT, respectively (Figure 21A). High-gamma power in these regions was analyzed relative to sentence and articulation onset for a representative individual (Figure 21B) and across the group (Figure



21C). HG/TTS responded strongly during both listening and speaking, remaining active for the duration of each sentence and throughout articulation. PT also responded strongly following sentence onset; however, this region was quiescent during articulation.

*Figure 21: Functional dissociation in HG/TTS and PT during listening and speaking. (A) A pair of electrodes were selected in each patient for a sustained (blue) and transient (red) response to the 3 Hz amplitude-modulated white noise stimulus. One representative patient was highlighted (bright) for single-trial analysis. (B) Single-trial raster plots of the percent change in high-gamma power during speech listening and production. (C) High-gamma power averaged across trials and then across patients ( $p < 10^{-3}$ ). The mean activity from the representative individual is included as the dotted line.*

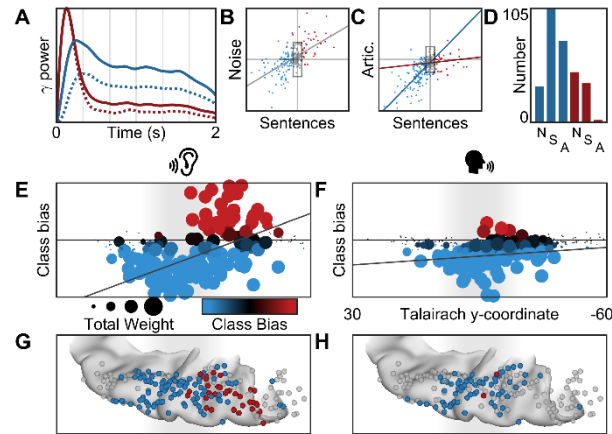


Figure 22: Supratemporal responses ( $n = 247$  electrodes) classified with 2-basis non-negative matrix factorization. **(A)** High-gamma power identifies a sustained (blue) and transient (red) response: normalized basis functions (dotted line) and the normalized group-average response for the top 10% of electrodes in each class (solid line). **(B)** The class bias determined by factorizations of electrode responses to noise and sentence listening – homogenous and structure acoustic inputs – were significantly correlated. **(C)** The factorization from sentence listening was applied to the electrode responses at articulation. Sustained class bias was significantly correlated for listening and speaking, but the transient class biases were uncorrelated. **(D)** Class biases greater than a value of 10 generated discrete classifications for speech listening (S; sustained,  $n = 105$ ; transient,  $n = 37$ ) and articulation (A; sustained,  $n = 74$ ; transient,  $n = 2$ ). **(E, F)** Spatial distribution of activation (sum of class weights; point size) and bias (difference in class weights; point color) reveals anteroposterior gradient of functional response during speech listening (left) but not articulation (right). **(G, H)** Electrodes shown on a standard supratemporal atlas reveal that the transient response is localized to PT and is uniquely suppressed during articulation.

I further characterized the spatial distribution of the transient response during speech listening and its suppression during speech production using non-negative matrix factorization. As for the white noise stimulus, high-gamma power yielded sustained and transient response types (Figure 22A) along a robust anteroposterior distribution (Figure 22E,G;  $r_s = 0.4688$ ,  $p < 10^{-10}$ ). These were strongly correlated with the class biases for white noise listening (Figure 22B;  $r_s = 0.5849$ ,  $p < 10^{-23}$ ). When this factorization was applied to high-gamma power during articulation (Figure 22C,F,H), the sustained response was preserved ( $r_s = 0.6663$ ,  $p < 10^{-16}$ ) while the transient response type was suppressed ( $r_s = 0.1094$ ,  $p = 0.3279$ ). Of the 37 electrodes demonstrating a transient response during speech listening, only 2 retained this classification during articulation (Figure 22D). The functional dissociation at PT between externally and internally generated speech is consistent with the theory of predictive coding during speech production<sup>154</sup> via motor-to-sensory feedback<sup>16,21,163</sup>.



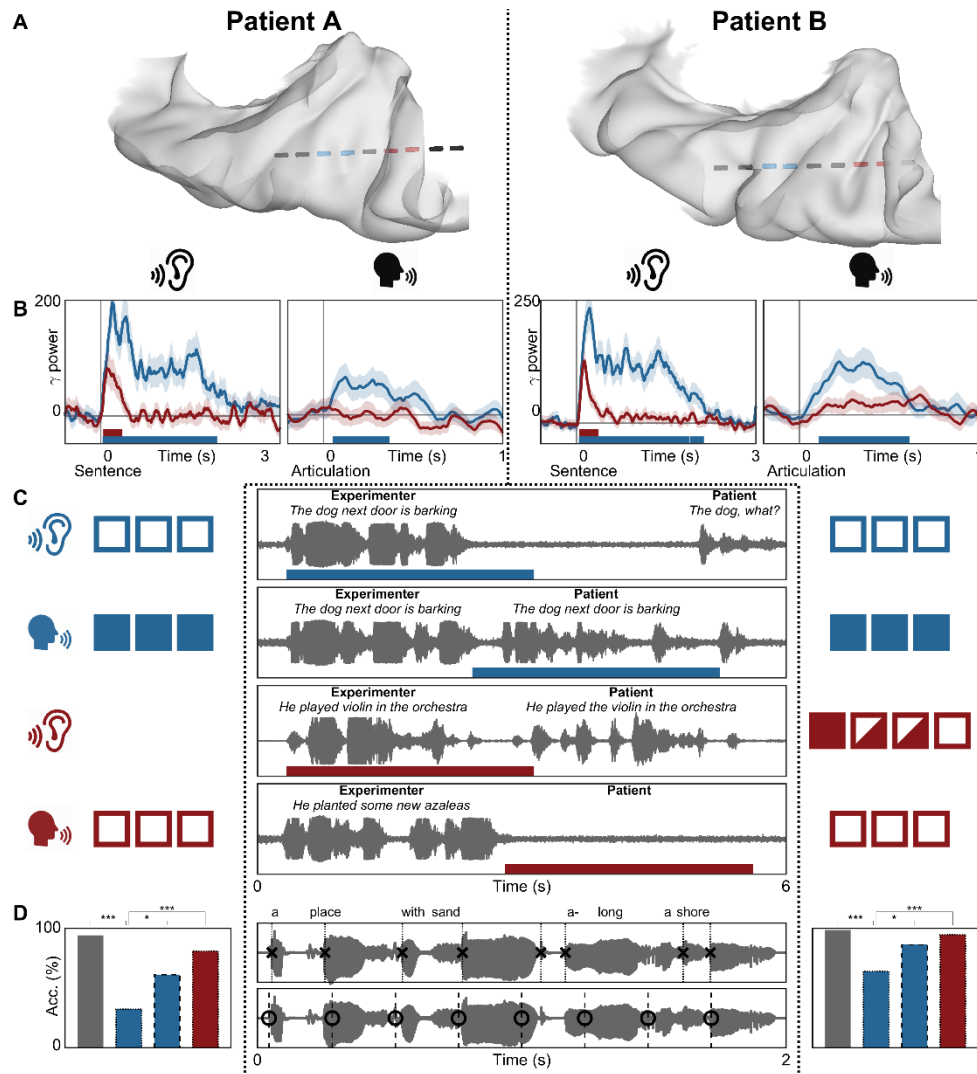


Figure 23: Supratemporal responses during two chronometric stimulation experiments. The first patient is shown on the left and the second on the right. (A) Electrode positions relative to patient-specific neuroanatomy. Two pairs of electrodes localized to HG/TTS and PT demonstrated sustained (blue) and transient responses (red), respectively. (B) Average percent change in high-gamma power during auditory naming to definition: sentence listening (left) and articulation (right). The transient response is entirely suppressed during articulation. (C) Chronometric cortical stimulation mapping with 50 Hz balanced 0.3 ms period square waves for 3 seconds during sentence repetition. Successful trials are indicated by full boxes, failed trials by empty boxes, and trials with a single error (e.g. word replacement) by half-full boxes. Stimulation was delivered either at stimulus onset (top panel) or offset (middle and bottom panel). The patient was unable to repeat the stimulus when HG/TTS was stimulated during listening, but was unaffected when the HG/TTS was stimulated during production. In contrast, stimulation of PT during production induced failure. (D) Chronometric stimulation during auditory naming to definition at either acoustic edges (x's; dotted lines) or uniformly distributed times (o's; dashed lines). Baseline accuracies were 94% and 99% (gray bar). Stimulation at acoustic edges in HG/TTS (left blue bar) and PT (red bar) resulted in naming accuracies of 32% and 81% in the first patient and 64% and 95% in the second patient. Stimulation of HG/TTS delivered uniformly throughout the stimulus (right blue bar) resulted in naming accuracies of 61% and 86% (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

In 3 patients, language mapping using direct cortical stimulation was performed along the full extent of the supratemporal plane. Current was passed between adjacent pairs of electrodes, transiently mimicking a focal lesion. Three language tasks were used to identify eloquent cortex: spoken sentence repetition, spoken naming to definition, and picture naming.

Direct cortical stimulation revealed distinct functional deficits in HG/TTS and PT. Stimulation of HG/TTS disrupted speech comprehension impacting both sentence repetition and naming to definition; however, picture naming was unaffected. Patients reported stimulation-evoked auditory phenomena including “buzzing” and “ringing.” Stimulation of PT disrupted articulation in all tasks, including picture naming. Furthermore, it evoked auditory hallucinations that included the sensation that “somebody’s talking” or “people [are] talking all around me ... like a ballpark”, the abstract ideation of “a rolling of more words”, and an “echo” like the speaker was “underwater.” These hallucinations were not induced by stimulation of the lateral superior temporal gyrus.

Prospectively, I implemented two stimulation experiments in two additional patients. The rhythmic white noise and auditory naming experiments were used to identify electrodes with sustained and transient responses in HG/TTS and PT, respectively (Figure 23A,B). In the first stimulation experiment (Figure 23C), the patients were asked to repeat a spoken sentence. The same stimulation current, frequency, and waveform used in the clinical standard above was applied at either the onset or the offset of the stimulus, disrupting processes related to comprehension or articulation, respectively. I found that stimulation of HG/TTS at stimulus onset interrupted speech perception, while the same stimulation at stimulus offset had no effect on articulation. In contradistinction, stimulation of PT at stimulus offset resulted in articulatory failure. These findings causally confirm the separable roles of HG/TTS and PT revealed by the analyses of passive electrocorticographic recordings.

In the second experiment (Figure 23D), patients performed naming to definition with concurrent chronometric stimulation. During the presentation of each spoken description, stimulation was triggered at either acoustic edges or at a uniform rate. Stimulation intensity was greatly reduced from clinical mapping parameters; only a single 500  $\mu$ s wide square wave pulse was delivered with each trigger (average 6.9 triggers per trial). Stimulation of HG/TTS at acoustic edges resulted in significantly worse task performance than equivalent stimulation of PT (32% vs. 81%,  $p < 10^{-3}$ ; 64% vs. 95%,  $p < 10^{-3}$ ). Furthermore, stimulation of HG/TTS at acoustic edges

also resulted in significantly worse task performance than uniformly distributed stimulation matched for total delivered current (32% vs. 61%,  $p = 0.0246$ , 64% vs. 86%,  $p = 0.0397$ ). Patients exhibited baseline accuracies of 94% and 99% without stimulation during the task. These results further causally corroborate the findings that HG/TTS is especially attuned to acoustic edges and that PT is not engaged for speech comprehension.

## DISCUSSION

These large-scale direct intracranial recordings and stimulations of the supratemporal plane have revealed the functional architecture of prediction in human early auditory cortex at fine-grained resolution. I demonstrate that a sustained response to speech engages a frequency-multiplexed encoding of two sublexical features: envelope and edges. I also uncover a pair of distinct neuroanatomic substrates that perform two separate types of prediction: temporal prediction in bilateral HG/TTS and speech-specific suppression in language dominant PT. The identification and characterization of these mechanisms advances the understanding of how human cortex parses continuous acoustic input for both speech perception and production.

Using electrodes positioned along the anteroposterior extent of the supratemporal plane, I localized the cortical signature of a sustained response during listening to strictly early auditory cortex: Heschl's gyrus and the transverse temporal sulcus<sup>164</sup>. This signature was considerably more complex than that suggested by prior studies, comprising a frequency-multiplexed encoding of envelope phase – distinct for rising and falling amplitudes of the same magnitude – in low-frequency, beta, and high-gamma power. I also identified a separate, concurrent encoding of acoustic edges in low-frequency phase reset. This latter encoding uniquely persisted after the rhythmic stimulus ended, consistent with the behavior of a predictive neural mechanism (and inconsistent with the behavior of evoked response potentials).

Importantly, identical cortical substrates engaged the same mechanisms during natural sentence listening. The speech envelope was tracked by bandlimited power and acoustic edges were demarcated by low-frequency phase reset. Furthermore, acoustic edges were more strongly encoded in cortical phase than syllabic onsets – a linguistic feature with similar frequency

and periodicity. This supports the assertion<sup>165</sup> that these mechanisms are driven by sublexical acoustic processing, perhaps even from subcortical regions (e.g. medial geniculate nucleus).

The organization of transient excitability states within neuronal populations has been thought of as a gating mechanism for cortical columns<sup>166–168</sup>. Discrete high excitability periods constitute windows of opportunity for input into sensory cortex, as evidenced by peri-threshold detection studies in somatosensory<sup>169</sup>, visual<sup>170</sup>, and auditory<sup>165,171,172</sup> regions. During listening, such windows might serve to segment speech to facilitate comprehension<sup>173</sup>. More generally, the temporal organization of high excitability periods could serve to minimize temporal uncertainty in stimulus processing and detection<sup>141,174</sup>. This view was corroborated by a behavioral study of responses to the same white noise stimulus used in these experiments that revealed a striking relationship between detection accuracy and the preceding rhythmic stimulus<sup>175</sup>. With the direct intracranial recordings in this study, I found that low-frequency phase reset anticipates the first missing acoustic edge. These results constitute strong evidence for neural mechanisms in early auditory cortex supporting temporal prediction, a fundamental computational element in models of speech perception<sup>142–144,154</sup>. results are consistent with the predictive encoding of when by a bandlimited complex of discrete computational channels, each arising from distinct patterns of hierarchical cortical connectivity<sup>154</sup>.

Entrainment – the synchronization of intrinsic neural oscillations with extrinsic rhythmic signals – has been suggested to have an important role in a variety of cognitive processes including attentional selection<sup>141,145,176</sup> and internal timekeeping<sup>170,177</sup>. Entrainment has also been implicated in speech perception by evidence that envelope distortions impair comprehension<sup>178,179</sup> independent of spectral content<sup>180,181</sup> and that the degree of neuro-acoustic entrainment modulates intelligibility<sup>182,183</sup>. While concurrent thalamocortical recordings are probably necessary to definitively separate entrainment from evoked response potentials, I demonstrate the neural mechanisms active during the sustained state are foundational to the hierarchy of acoustic perception – including language.

While the sustained response was constrained to Heschl's gyrus and the transverse temporal sulcus, I observed a distinct transient response in planum temporale. The transient response was characterized by a brief spike in high-gamma power and rapid reset of low-frequency phase immediately following acoustic onset. Interestingly, this response was not engaged during self-generated speech. Such preferential engagement for unexpected sound is consistent with predictive encoding during speech production<sup>16,21</sup>. Upon execution of a speech motor plan, a learned internal model generates an efference copy<sup>184,185</sup> – an expected sensory result. When the acoustic input matches this efference copy, no cortical signal is generated; however, when a mismatch occurs (e.g. externally-generated sound or speech), an error signal results<sup>154</sup>. This is precisely what I observed in the planum temporale, distinct from the sustained response in Heschl's gyrus. I corroborated these results with direct current injection at Heschl's gyrus and planum temporale; stimulation of the former area selectively disrupted speech perception, while stimulation of the latter area selectively disrupted speech production.

There is no direct evidence for internal predictive models instantiated in human cortex<sup>186</sup>. The results advance understanding of the neurobiology of predictive speech coding in two respects. First, functional studies have revealed single-unit preference in primary auditory cortex for listening or speaking in both non-humans<sup>187</sup> and humans<sup>184</sup>. It has recently been asserted that these response tunings overlap – an “intertwined mosaic of neuronal populations”<sup>188</sup> in auditory cortex. Instead, the complete anteroposterior mapping of the supratemporal plane in a large patient cohort enabled us to identify a distinct neuroanatomical organization in planum temporale. Second, several groups report cortical response suppression specific for self-generated speech<sup>188–190</sup>. I reveal two distinct modes that enable this suppression: a partial reduction of activity in Heschl's gyrus and a complete absence of the transient response in planum temporale. The stapedius reflex<sup>187</sup> does not explain the latter mode, suggesting a neural mechanism of suppression. Altogether, I provide compelling evidence for efference copies – predictive encoding of what<sup>154</sup> – and their essential role in speech production<sup>16,21</sup>.

## CONCLUSIONS & FUTURE DIRECTIONS

These experiments advance our understanding of the neurobiological foundation of speech production, building towards a universal cognitive model of language in the brain. I accomplish this aim with convergent evidence from functional and structural neuroimaging, large-scale invasive electrophysiology of multiple experimental modalities, and direct stimulation of human cortex with focal current injection.

This work provides novel insights for regions that have been marginalized in language studies. I demonstrate: a) the definitive role of middle fusiform gyrus in mediating conceptual representation, as suggested by prior work<sup>40,83</sup>, b) an expanded role of supplementary motor area for word formulation<sup>191</sup> in addition to movement planning<sup>192</sup>, consistent with its structural connections to ventrolateral frontal cortex through the aslant track<sup>193</sup>, c) a putative role for the superior frontal sulcus in networks subserving retrieval of lexical semantic information<sup>102</sup>, and d) activation of dorsal sensorimotor cortex concurrently with subcentral gyrus, concordant with secondary representations of speech articulators<sup>194</sup>.

The results thoroughly map the neurobiology of picture naming in language-dominant cortex, robustly motivate the use of seriality in computational models of speech production, and establish a concrete mechanism for representational interactions in language networks. Furthermore, I uncover evidence to better understand the integrated function of perisylvian cortex during both speech perception and production.

Quantification of dynamical systems in the brain uncovers a possible mechanistic foundation for generalized cognitive processes. By modeling both activity in isolated network nodes as well as the interdependence between nodes, we can begin to unify localizationist and connectionist accounts of cortical function. This perspective yields a powerful conceptual framework for understanding normal language function, as well as the spectrum of specific disruptions secondary to various neurological insults.

I hope to explore these new therapeutic targets with targeted and responsive multisite stimulation in my own neurosurgical practice.

## REFERENCES

1. Fromkin, V. a. The non-anomalous nature of anomalous utterances. *Speech Errors As Linguist. Evid.* **47**, 27–52 (1973).
2. Levelt, W. Speaking: From Intention to Articulation - Chapter 6. in *Speaking: From Intention to Articulation* 181–234 (1989). doi:10.1016/0166-2236(90)90151-Y
3. Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. a. Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* **104**, 801–838 (1997).
4. Garrett, M. . Levels of processing in sentence production. in *Language production* 177–200 (1980).
5. Bock, K. & Levelt, W. Language production: grammatical encoding. in *Handbook of psycholinguistics* 945–984 (1994).
6. Dell, G. S. A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* **93**, 283–321 (1986).
7. Oppenheim, G. M. & Dell, G. S. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* **106**, 528–537 (2008).
8. Indefrey, P. & Levelt, W. J. M. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144 (2004).
9. Roelofs, A. Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: comment on Rapp and Goldrick (2000). *Psychol. Rev.* **111**, 561–572; discussion 573-580 (2004).
10. Roelofs, A. The WEAVER model of word-form encoding in speech production. *Cognition* **64**, 249–284 (1997).
11. Rapp, B. & Goldrick, M. Discreteness and interactivity in spoken word production. *Psychol. Rev.* **107**, 460–499 (2000).
12. Rapp, B. & Goldrick, M. Feedback by Any Other Name Is Still Interactivity: A Reply to Roelofs (2004). *Psychol. Rev.* **111**, 573–578 (2004).
13. Hagoort, P. & Levelt, W. J. M. The speaking brain. *Science* **326**, 372–373 (2009).
14. Fairbanks, G. Systematic Research In Experimental Phonetics: A Theory Of The Speech Mechanism As A Servosystem. *J. Speech Hear. Disord.* **19**, 133–139 (1954).
15. Houde, J. F. Sensorimotor Adaptation in Speech Production. *Science (80-. )*. **279**, 1213–1216 (1998).
16. Guenther, F. H. Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* **39**, 350–365 (2006).
17. Price, C. J., Crinion, J. T. & MacSweeney, M. A generative model of speech production in Broca's and Wernicke's areas. *Front. Psychol.* **2**, (2011).
18. Tian, X. & Poeppel, D. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front. Psychol.* (2010). doi:10.3389/fpsyg.2010.00166
19. Houde, J. F. & Nagarajan, S. S. Speech Production as State Feedback Control. *Front. Hum. Neurosci.* **5**, (2011).

20. Hickok, G., Houde, J. & Rong, F. Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron* **69**, 407–422 (2011).
21. Hickok, G. Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* (2012). doi:10.1038/nrn3158
22. Lachaux, J. P., Axmacher, N., Mormann, F., Halgren, E. & Crone, N. E. High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Prog Neurobiol* **98**, 279–301 (2012).
23. Manning, J. R., Jacobs, J., Fried, I. & Kahana, M. J. Broadband Shifts in Local Field Potential Power Spectra Are Correlated with Single-Neuron Spiking in Humans. *J. Neurosci.* **29**, 13613–20 (2009).
24. Cardin, J. A. *et al.* Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459**, 663–7 (2009).
25. Logothetis, N. K. The underpinnings of the BOLD functional magnetic resonance imaging signal. *J. Neurosci.* **23**, 3963–3971 (2003).
26. Magri, C., Schridde, U., Murayama, Y., Panzeri, S. & Logothetis, N. K. The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. *J. Neurosci.* **32**, 1395–1407 (2012).
27. Conner, C. R., Ellmore, T. M., Pieters, T. a, Disano, M. a & Tandon, N. Variability of the relationship between electrophysiology and BOLD-fMRI across cortical regions in humans. *J. Neurosci.* **31**, 12855–65 (2011).
28. Mukamel, R. *et al.* Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* **309**, 951–954 (2005).
29. Crone, N. E. *et al.* Electrographic gamma activity during word production in spoken and sign language. *Neurology* **57**, 2045–2053 (2001).
30. Jacobs, J. & Kahana, M. J. Direct brain recordings fuel advances in cognitive electrophysiology. *Trends Cogn. Sci.* **14**, 162–171 (2010).
31. Fries, P. Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. *Annu. Rev. Neurosci.* **32**, 209–224 (2009).
32. Ellmore, T. M., Beauchamp, M. S., O'Neill, T. J., Dreyer, S. & Tandon, N. Relationships between essential cortical language sites and subcortical pathways. *J Neurosurg* **111**, 755–766 (2009).
33. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–94 (1999).
34. Cox, R. W. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**, 162–173 (1996).
35. Pieters, T. A., Conner, C. R. & Tandon, N. Recursive grid partitioning on a cortical surface model: an optimized technique for the localization of implanted subdural electrodes. *J. Neurosurg.* **118**, 1086–1097 (2013).
36. Wada, J. & Rasmussen, T. Intracarotid Injection of Sodium Amytal for the Lateralization of Cerebral Speech Dominance. *J. Neurosurg.* **106**, 1117–1133 (2007).
37. Tertel, K., Tandon, N. & Ellmore, T. M. Probing brain connectivity by combined analysis of diffusion MRI tractography and electrocorticography. *Comput. Biol. Med.* **41**, 1092–



- 1099 (2011).
38. Tandon, N. Cortical mapping by electrical stimulation of subdural electrodes: language areas. in *Textbook of Epilepsy Surgery* (ed. Luders, H.) 1001–1015 (McGraw Hill, 2008).
  39. Kadipasaoglu, C. M. *et al.* Surface-based mixed effects multilevel analysis of grouped human electrocorticography. *Neuroimage* **101**, 215–224 (2014).
  40. Forseth, K. J. *et al.* A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain* (2018). doi:10.1093/brain/awy120
  41. Saad, Z. S. & Reynolds, R. C. Suma. *Neuroimage* **62**, 768–773 (2012).
  42. Fischl, B., Sereno, M. I. & Dale, a M. Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System. *Neuroimage* **9**, 195–207 (1999).
  43. Argall, B. D., Saad, Z. S. & Beauchamp, M. S. Simplified intersubject averaging on the cortical surface using SUMA. *Hum. Brain Mapp.* **27**, 14–27 (2006).
  44. Chen, G., Saad, Z. S., Nath, A. R., Beauchamp, M. S. & Cox, R. W. fMRI group analysis combining effect estimates and their variances. *Neuroimage* **60**, 747–765 (2012).
  45. Conner, C. R., Chen, G., Pieters, T. A. & Tandon, N. Category specific spatial dissociations of parallel processes underlying visual naming. *Cereb. Cortex* **24**, 2741–2750 (2014).
  46. Hillis, A. E. Aphasia: Progress in the last quarter of a century. *Neurology* (2007). doi:10.1212/01.wnl.0000265600.69385.6f
  47. Warrington, E. K. The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology* **27**, 635–657 (1975).
  48. Damasio, H. ;, Grabowski, T. J., Tranel, D. ;, Hichwa, R. D. & Damasio, A. R. A neural basis for lexical retrieval. *Nature* **380**, 499–505 (1996).
  49. Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* (2012). doi:10.1016/j.neuroimage.2012.04.062
  50. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* (2016). doi:10.1038/nature18933
  51. Munding, D., Dubarry, A. S. & Alario, F. X. On the cortical dynamics of word production: a review of the MEG evidence. *Language, Cognition and Neuroscience* (2016). doi:10.1080/23273798.2015.1071857
  52. Fedorenko, E. & Blank, I. A. Broca's Area Is Not a Natural Kind. *Trends in Cognitive Sciences* (2020). doi:10.1016/j.tics.2020.01.001
  53. Bassett, D. S. & Sporns, O. Network neuroscience. *Nature Neuroscience* (2017). doi:10.1038/nn.4502
  54. Geschwind, N. Disconnexion syndromes in animals and man. *Brain* **88**, 237 (1965).
  55. Fedorenko, E. & Thompson-Schill, S. L. Reworking the language network. *Trends in Cognitive Sciences* **18**, 120–126 (2014).
  56. Crone, N. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* (1998). doi:10.1093/brain/121.12.2301

57. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).
58. Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* (80- ). **326**, 445–9 (2009).
59. Gui, P. *et al.* Assessing the depth of language processing in patients with disorders of consciousness. *Nat. Neurosci.* (2020). doi:10.1038/s41593-020-0639-1
60. Rabinovich, M. I. *Principles of brain dynamics: global state interactions.* (MIT Press, 2012).
61. Kirst, C., Timme, M. & Battaglia, D. Dynamic information routing in complex networks. *Nat. Commun.* (2016). doi:10.1038/ncomms11061
62. Friston, K. J. Transients, metastability, and neuronal dynamics. *Neuroimage* (1997). doi:10.1006/nimg.1997.0259
63. Saravani, A. G., Forseth, K. J., Tandon, N. & Pitkow, X. Dynamic brain interactions during picture naming. *eNeuro* (2019). doi:10.1523/ENEURO.0472-18.2019
64. Baum, L. E., Petrie, T., Soules, G. & Weiss, N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Ann. Math. Stat.* (1970). doi:10.1214/aoms/1177697196
65. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum Likelihood from Incomplete Data Via the EM Algorithm . *J. R. Stat. Soc. Ser. B* (1977). doi:10.1111/j.2517-6161.1977.tb01600.x
66. Morf, M., Vieira, A. & Kailath, T. Covariance Characterization by Partial Autocorrelation Matrices. *Ann. Stat.* (1978). doi:10.1214/aos/1176344208
67. Ding, M., Bressler, S. L., Yang, W. & Liang, H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: Data preprocessing, model validation, and variability assessment. *Biol. Cybern.* (2000). doi:10.1007/s004229900137
68. Baccalá, L. A. & Sameshima, K. Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.* (2001). doi:10.1007/PL00007990
69. Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* (2004). doi:10.1023/B:VISI.0000029664.99615.94
70. Snodgrass, J. G. & Vanderwart, M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* **6**, 174–215 (1980).
71. Brysbaert, M. & New, B. Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* **41**, 977–90 (2009).
72. Vaden, K., Halpin, H. & Hickok, G. Irvine Phonotactic Online Dictionary, Version 2.0. (2009).
73. Edwards, E. *et al.* Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* **50**, 291–301 (2010).
74. Engel, A. K., Moll, C. K. E. E., Fried, I. & Ojemann, G. A. Invasive recordings from the

- human brain: clinical insights and beyond. *Nat. Rev. Neurosci.* **6**, 35–47 (2005).
75. Conner, C. R., Kadipasaoglu, C. M., Shouval, H. Z., Hickok, G. & Tandon, N. Network dynamics of Broca's area during word selection. *PLoS One* (2019). doi:10.1371/journal.pone.0225756
  76. Llorens, A., Trébuchon, A., Liégeois-Chauvel, C. & Alario, F. X. Intra-cranial recordings of brain activity during language production. *Frontiers in Psychology* (2011). doi:10.3389/fpsyg.2011.00375
  77. Levelt, W. J. M., Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behavioral and Brain Sciences* (1999). doi:10.1017/S0140525X99001776
  78. Glaser, W. R. & Dünghoff, F. J. The time course of picture-word interference. *J. Exp. Psychol. Hum. Percept. Perform.* (1984). doi:10.1037/0096-1523.10.5.640
  79. Levelt, W. J. M. *et al.* The time course of lexical access in speech production: A study of picture naming. *Psychol. Rev.* (1991). doi:10.1037/0033-295X.98.1.122
  80. Luria, A. R. Towards the mechanisms of naming disturbance. *Neuropsychologia* (1973). doi:10.1016/0028-3932(73)90028-6
  81. Broca, P. Sur le siège de la faculté du langage articulé. *Bull. la Société d'anthropologie Paris* (1865). doi:10.3406/bmsap.1865.9495
  82. Penfield, W. & Roberts, L. *Speech and Brain Mechanisms*. (Princeton University Press, 1959).
  83. Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).
  84. Kadipasaoglu, C. M. *et al.* Development of grouped icEEG for the study of cognitive processing. *Front. Psychol.* **6**, 1–7 (2015).
  85. Levelt, W. J. M. Models of word production. *Trends Cogn. Sci.* **3**, 223–232 (1999).
  86. Forseth, K. J., Hickok, G., Rollo, P. S. & Tandon, N. Language prediction mechanisms in human auditory cortex. *Nat. Commun.* **11**, (2020).
  87. Linderman, S., Nichols, A., Blei, D., Zimmer, M. & Paninski, L. Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in *C. elegans*. *bioRxiv* (2019). doi:10.1101/621540
  88. Glaser, J. I., Whiteway, M., Cunningham, J. P., Paninski, L. & Linderman, S. W. Recurrent Switching Dynamical Systems Models for Multiple Interacting Neural Populations. *bioRxiv* (2020).
  89. Hinton, G. E., McClelland, J. L. & Rumelhart, D. E. *Distributed representations. Parallel Distributed Processing* (1986).
  90. Chen, G., Deng, Q., Szymczak, A., Laramée, R. S. & Zhang, E. Morse set classification and hierarchical refinement using conley index. *IEEE Trans. Vis. Comput. Graph.* (2012). doi:10.1109/TVCG.2011.107
  91. Quillian, M. R. Semantic Memory. AFCRL-66-189 (1966).
  92. Tulving, E. Episodic and semantic memory. *Organization of memory* **1**, 381–403 (1972).
  93. Brown, R. & McNeill, D. The “tip of the tongue” phenomenon. *J. Verbal Learning Verbal Behav.* **5**, 325–337 (1966).

94. Margolin, D. I., Pate, D. S., Friedrich, F. J. & Elia, E. Dysnomia in dementia and in stroke patients: Different underlying cognitive deficits. *J. Clin. Exp. Neuropsychol.* **12**, 597–612 (1990).
95. Mion, M. *et al.* What the left and right anterior fusiform gyri tell us about semantic memory. *Brain* **133**, 3256–3268 (2010).
96. Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M. & Lambon Ralph, M. A. The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: Evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cereb. Cortex* **20**, 2728–2738 (2010).
97. Nestor, P. J., Fryer, T. D. & Hodges, J. R. Declarative memory impairments in Alzheimer's disease and semantic dementia. *Neuroimage* **30**, 1010–1020 (2006).
98. Noppeney, U. & Price, C. J. A PET Study of Stimulus- and Task-Induced Semantic Processing. **935**, 927–935 (2002).
99. Sharp, D. J., Scott, S. K. & Wise, R. J. S. Retrieving meaning after temporal lobe infarction: The role of the basal language area. *Ann. Neurol.* **56**, 836–846 (2004).
100. Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E. & Wise, R. J. S. Converging Language Streams in the Human Temporal Lobe. *J. Neurosci.* **26**, 7328–7336 (2006).
101. Bright, P., Moss, H. & Tyler, L. K. Unitary vs multiple semantics: PET studies of word and picture processing. *Brain Lang.* **89**, 417–432 (2004).
102. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796 (2009).
103. Marinkovic, K. *et al.* Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* **38**, 487–497 (2003).
104. Phan, T. G., Fong, A. C., Donnan, G. A., Srikanth, V. & Reutens, D. C. Digital probabilistic atlas of the border region between the middle and posterior cerebral arteries. *Cerebrovasc. Dis.* **27**, 529–536 (2009).
105. Ojemann, J. G. *et al.* Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage* **6**, 156–67 (1997).
106. Devlin, J. T. *et al.* Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage* **11**, 589–600 (2000).
107. Hansen, P., Kringelbach, M. & Salmelin, R. *MEG: An introduction to methods. MEG: An Introduction to Methods* (2010). doi:10.1093/acprof:oso/9780195307238.001.0001
108. Tandon, N. Mapping of Human Language. *Clin. Brain Mapp.* 203–218 (2012).
109. Mesgarani, N., Cheung, C., Johnson, K. & Chang, E. F. Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–10 (2014).
110. Friederici, A. D. The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences* **16**, 262–268 (2012).
111. Chan, A. M. *et al.* Speech-specific tuning of neurons in human superior temporal gyrus. *Cereb. Cortex* **24**, 2679–2693 (2014).
112. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev.*

- Neurosci.* **8**, 393–402 (2007).
113. Burnstine, T. H. *et al.* Characterization of the basal temporal language area in patients with left temporal lobe epilepsy. *Neurology* **40**, 966–70 (1990).
  114. Lüders, H. *et al.* Basal temporal language area. *Brain* **114** ( Pt 2, 743–54 (1991).
  115. Nobre, a C., Allison, T. & McCarthy, G. Word recognition in the human inferior temporal lobe. *Nature* **372**, 260–263 (1994).
  116. Liu, H., Agam, Y., Madsen, J. R. & Kreiman, G. Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron* **62**, 281–290 (2009).
  117. Ralph, M. A. L., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience* **18**, 42–55 (2016).
  118. Catani, M. & Ffytche, D. H. The rises and falls of disconnection syndromes. *Brain* **128**, 2224–2239 (2005).
  119. Martin, A. The Representation of Object Concepts in the Brain. *Annu. Rev. Psychol.* **58**, 25–45 (2007).
  120. Mesulam, M.-M. From sensation to perception. *Brain* **121**, 1013–1052 (1998).
  121. Thompson-Schill, S. L. Neuroimaging studies of semantic memory: Inferring ‘how’ from ‘where’. *Neuropsychologia* **41**, 280–292 (2003).
  122. Hoppe, C. *et al.* Laser interstitial thermotherapy (LiTT) in epilepsy surgery. *Seizure* **48**, 45–52 (2017).
  123. Drane, D. L. *et al.* Better object recognition and naming outcome with MRI-guided stereotactic laser amygdalohippocampotomy for temporal lobe epilepsy. *Epilepsia* **56**, 101–113 (2015).
  124. Wagner, A. D., Paré-Blagojev, E. J., Clark, J. & Poldrack, R. A. Recovering Meaning. *Neuron* **31**, 329–338 (2001).
  125. Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K. & Farah, M. J. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 14792–14797 (1997).
  126. Poldrack, R. A. *et al.* Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* **10**, 15–35 (1999).
  127. Badre, D., Poldrack, R. A., Paré-Blagojev, E. J., Insler, R. Z. & Wagner, A. D. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* **47**, 907–918 (2005).
  128. Morán, M. a, Mufson, E. J. & Mesulam, M. M. Neural inputs into the temporopolar cortex of the rhesus monkey. *J. Comp. Neurol.* **256**, 88–103 (1987).
  129. Felleman, D. J. & Van Essen, D. C. Distributed hierachical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
  130. Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F. & Braun, A. R. Symbolic gestures and spoken language are processed by a common neural system. *Proc Natl Acad Sci U S A* **106**, 20664–20669 (2009).

131. Chao, L. L. & Martin, A. Representation of Manipulable Man-Made Objects in the Dorsal Stream. *Neuroimage* **12**, 478–484 (2000).
132. Ganis, G., Thompson, W. L. & Kosslyn, S. M. Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cogn. Brain Res.* **20**, 226–241 (2004).
133. Mellet, E., Tzourio, N., Denis, M. & Mazoyer, B. Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport* **9**, 803–8 (1998).
134. Ishai, A., Ungerleider, L. G. & Haxby, J. V. Distributed neural systems for the generation of visual images. *Neuron* **28**, 979–990 (2000).
135. Noppeney, U., Phillips, J. & Price, C. The neural areas that control the retrieval and selection of semantics. *Neuropsychologia* **42**, 1269–1280 (2004).
136. Otsubo, H. *et al.* High-frequency oscillations of ictal muscle activity and epileptogenic discharges on intracranial EEG in a temporal lobe epilepsy patient. *Clin. Neurophysiol.* **119**, 862–868 (2008).
137. Muthukumaraswamy, S. D. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front. Hum. Neurosci.* **7**, (2013).
138. Lotto, A. J. & Holt, L. L. *Speech perception: The view from the auditory system.* *Neurobiology of Language* (2016). doi:10.1016/B978-0-12-407794-2.00016-X
139. Rosen, S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **336**, 367–373 (1992).
140. Haegens, S. & Zion Golumbic, E. Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews* (2018). doi:10.1016/j.neubiorev.2017.12.002
141. Schroeder, C. E. & Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18 (2009).
142. Ghitza, O. Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* **2**, (2011).
143. Giraud, A.-L. L. & Poeppel, D. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–7 (2012).
144. Peelle, J. E. & Davis, M. H. Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology* **3**, (2012).
145. Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H. & Lakatos, P. Dynamics of Active Sensing and perceptual selection. *Current Opinion in Neurobiology* **20**, 172–176 (2010).
146. Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I. & Schroeder, C. E. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (80-. ).* (2008). doi:10.1126/science.1154735
147. Howard, M. F. & Poeppel, D. The neuromagnetic response to spoken sentences: Co-modulation of theta band amplitude and phase. *Neuroimage* (2012). doi:10.1016/j.neuroimage.2012.02.028
148. Lakatos, P. *et al.* An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *J. Neurophysiol.* **94**, 1904–11 (2005).

149. Luo, H., Liu, Z. & Poeppel, D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* (2010). doi:10.1371/journal.pbio.1000445
150. Kerlin, J. R., Shahin, A. J. & Miller, L. M. Attentional Gain Control of Ongoing Cortical Speech Representations in a 'Cocktail Party'. *J. Neurosci.* (2010). doi:10.1523/jneurosci.3631-09.2010
151. Zoefel, B., Costa-Faidella, J., Lakatos, P., Schroeder, C. E. & VanRullen, R. Characterization of neural entrainment to speech with and without slow spectral energy fluctuations in laminar recordings in monkey A1. *Neuroimage* (2017). doi:10.1016/j.neuroimage.2017.02.014
152. Aliu, S. O., Houde, J. F. & Nagarajan, S. S. Motor-induced suppression of the auditory cortex. *J. Cogn. Neurosci.* (2009). doi:10.1162/jocn.2009.21055
153. Okada, K., Matchin, W. & Hickok, G. Neural evidence for predictive coding in auditory cortex during speech production. *Psychon. Bull. Rev.* (2017). doi:10.3758/s13423-017-1284-x
154. Arnal, L. H. & Giraud, A. L. Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences* **16**, 390–398 (2012).
155. Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P. & Plemmons, R. J. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data Anal.* (2007). doi:10.1016/j.csda.2006.11.006
156. Fries, P. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences* (2005). doi:10.1016/j.tics.2005.08.011
157. Ermentrout, G. B. & Kleinfeld, D. Traveling electrical waves in cortex: Insights from phase dynamics and speculation on a computational role. *Neuron* (2001). doi:10.1016/S0896-6273(01)00178-7
158. Zhang, H., Watrous, A. J., Patel, A. & Jacobs, J. Theta and Alpha Oscillations Are Traveling Waves in the Human Neocortex. *Neuron* (2018). doi:10.1016/j.neuron.2018.05.019
159. Muller, L. *et al.* Rotating waves during human sleep spindles organize global patterns of activity that repeat precisely through the night. *Elife* (2016). doi:10.7554/eLife.17267.001
160. Roberts, T. P. L. & Poeppel, D. Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport* (1996). doi:10.1097/00001756-199604260-00007
161. Hickok, G. & Poeppel, D. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).
162. Hamilton, L. S., Edwards, E. & Chang, E. F. A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. *Curr. Biol.* (2018). doi:10.1016/j.cub.2018.04.033
163. Sperry, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* (1950). doi:10.1037/h0055479
164. Liegeois-Chauvel, C., Musolino, A. & Chauvel, P. Localization of the primary auditory area in man. *Brain* **114A**, 139–153 (1991).
165. Henry, M. J. & Obleser, J. Frequency modulation entrains slow neural oscillations and

- optimizes human listening behavior. *Proc. Natl. Acad. Sci.* **109**, 2009–100 (2012).
166. Buzsáki, G., Anastassiou, C. a. & Koch, C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012).
  167. Hopfield, J. J. Encoding for computation: Recognizing brief dynamical patterns by exploiting effects of weak rhythms on action-potential timing. *Proc. Natl. Acad. Sci.* (2004). doi:10.1073/pnas.0401125101
  168. Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science (80-. )*. (1996). doi:10.1126/science.273.5283.1868
  169. Monto, S., Palva, S., Voipio, J. & Palva, J. M. Very Slow EEG Fluctuations Predict the Dynamics of Stimulus Detection and Oscillation Amplitudes in Humans. *J. Neurosci.* (2008). doi:10.1523/JNEUROSCI.1910-08.2008
  170. Cravo, A. M., Rohenkohl, G., Wyart, V. & Nobre, A. C. Endogenous modulation of low frequency oscillations by temporal expectations. *J. Neurophysiol.* **106**, 2964–2972 (2011).
  171. Neuling, T., Rach, S., Wagner, S., Wolters, C. H. & Herrmann, C. S. Good vibrations: Oscillatory phase shapes perception. *Neuroimage* (2012). doi:10.1016/j.neuroimage.2012.07.024
  172. Ng, B. S. W., Schroeder, T. & Kayser, C. A Precluding But Not Ensuring Role of Entrained Low-Frequency Oscillations for Auditory Perception. *J. Neurosci.* **32**, 12268–12276 (2012).
  173. Jensen, O., Bonnefond, M. & VanRullen, R. An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences* (2012). doi:10.1016/j.tics.2012.03.002
  174. Nobre, A., Correa, A. & Coull, J. The hazards of time. *Current Opinion in Neurobiology* (2007). doi:10.1016/j.conb.2007.07.006
  175. Hickok, G., Farahbod, H. & Saberi, K. The Rhythm of Perception: Entrainment to Acoustic Rhythms Induces Subsequent Perceptual Oscillation. *Psychol. Sci.* **26**, 1006–1013 (2015).
  176. Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I. & Schroeder, C. E. Entrainment of Neuronal Attentional Selection. *Science (80-. )*. **320**, 23–25 (2008).
  177. Stefanics, G. *et al.* Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J. Neurosci.* **30**, 13578–85 (2010).
  178. Ahissar, E. & Ahissar, M. Processing of the temporal envelope of speech. in *The Auditory Cortex: A Synthesis of Human and Animal Research* 295–314 (2005). doi:10.4324/9781410613066
  179. Ghitza, O. & Greenberg, S. On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* **66**, 113–126 (2009).
  180. Drullman, R., Festen, J. M. & Plomp, R. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* **95**, 1053–1064 (1994).
  181. Smith, Z. M., Delgutte, B. & Oxenham, A. J. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87–90 (2002).



182. Ahissar, E. *et al.* Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci.* **98**, 13367–13372 (2001).
183. Peelle, J. E., Gross, J. & Davis, M. H. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* **23**, 1378–1387 (2013).
184. Creutzfeldt, O., Ojemann, G. & Lettich, E. Neuronal activity in the human lateral temporal lobe II. Responses to the subjects own voice. *Exp. brain Res.* (1989). doi:10.1007/BF00249600
185. Held, R. & Hein, A. V. Adaptation of disarranged hand-eye coordination contingent upon reafferent stimulation. *Percept. Mot. Ski.* (1958). doi:10.2466/pms.8.3.87-90
186. Keller, G. B. & Mrcsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* (2018). doi:10.1016/j.neuron.2018.10.003
187. Müller-Preuss, P. & Ploog, D. Inhibition of auditory cortical neurons during phonation. *Brain Res.* (1981). doi:10.1016/0006-8993(81)90491-1
188. Flinker, A. *et al.* Single-Trial Speech Suppression of Auditory Cortex Activity in Humans. *J. Neurosci.* (2010). doi:10.1523/JNEUROSCI.1809-10.2010
189. Towle, V. L. *et al.* ECoG gamma activity during a language task: Differentiating expressive and receptive speech areas. *Brain* (2008). doi:10.1093/brain/awn147
190. Crone, N. E., Boatman, D., Gordon, B. & Hao, L. Induced electrocorticographic gamma activity during auditory perception. *Clin. Neurophysiol.* (2001). doi:10.1016/S1388-2457(00)00545-9
191. Alario, F. X., Chainay, H., Lehericy, S. S. & Cohen, L. The role of the supplementary motor area (SMA) in word production. *Brain Res.* **1076**, 129–143 (2006).
192. Penfield, W. & Welch, K. The supplementary motor area of the cerebral cortex: A clinical and experimental study. *Arch. Neurol. Psychiatry* (1951). doi:10.1001/archneurpsyc.1951.02320090038004
193. Vergani, F. *et al.* White matter connections of the supplementary motor area in humans. *J. Neurol. Neurosurg. Psychiatry* (2014). doi:10.1136/jnnp-2013-307492
194. Stavisky, S. D. *et al.* Neural ensemble dynamics in dorsal motor cortex during speech in people with paralysis. *Elife* (2019). doi:10.7554/eLife.46015

## VITA

Kiefer James Forseth was born in Phoenix, Arizona, the son of Shannon and James Forseth. After completing his work at Brophy College preparatory, Phoenix, Arizona in 2010, he entered Rice University in Houston, Texas. In 2014, he received the degrees of Bachelor of Science with a major in electrical engineering and a specialization in digital signal processing and Bachelor of Music with a major in piano performance. Subsequently, he entered the MD/PhD program at McGovern Medical School and, in 2016, began his graduate studies at The University of Texas MD Anderson Cancer Center UTHHealth Graduate School of Biomedical Sciences.

Permanent address:

1133 John Freeman Blvd, JJL 420

Houston, TX, 77030