Interpretability of statistical approaches in speech and language neuroscience

Sophie Bouton, Valerian Chambon, Narly Golestani, Elia Formisano, Timothée Proix, Anne-Lise Giraud

▶ **To cite this version:**

Sophie Bouton, Valerian Chambon, Narly Golestani, Elia Formisano, Timothée Proix, et al.. Interpretability of statistical approaches in speech and language neuroscience. 2023. hal-04284936

HAL Id: hal-04284936
https://hal.science/hal-04284936

Preprint submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interpretability of statistical approaches in speech and language neuroscience

Sophie Bouton[1,2], Valérian Chambon[3], Narly Golestani[4,5,6], Elia Formisano[7,8,9], Timothée Proix[10], Anne-Lise Giraud[1,10]

1. Institut Pasteur, Université Paris Cité, Inserm, Institut de l'Audition, F-75012 Paris, France
2. Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, Ecole Normale Supérieure, EHESS, CNRS, PSL University, Paris, France
3. Institut Jean Nicod, CNRS/École Normale Supérieure UMR 8129, PSL University, 75005 Paris, France
4. Section of Psychology, University of Geneva – Campus Biotech, 9 chemin des Mines, 1202 Geneva, Switzerland
5. Brain and Language Lab, Cognitive Science Hub, University of Vienna, Austria
6. Department of Behavioral and Cognitive Biology, Faculty of Life Sciences, University of Vienna, Austria
7. Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands
8. Maastricht Brain Imaging Center, Maastricht, the Netherlands
9. Maastricht Centre for Systems Biology, Maastricht University, Maastricht, the Netherlands
10. Department of Neuroscience, University of Geneva – Campus Biotech, 9 chemin des Mines, 1202 Geneva, Switzerland

Timothée Proix and Anne-Lise Giraud contributed equally to this work.

The corresponding authors are:

Sophie Bouton
E-Mail address: sophie.bouton@cnrs.fr
ORCID: https://orcid.org/0000-0001-5496-4583

Anne-Lise Giraud
E-Mail address: anne-lise.giraud@unige.ch
ORCID: https://orcid.org/0000-0002-1261-3555

Keywords. data analysis, interpretation, hypothesis-based, data-driven, speech.

Abstract

The classical view of the speech and language neural system is that of a hierarchy of interdependent modules, enabling the progressive transformation of a continuous acoustic stream into an articulated series of concepts. This modular and hierarchical view follows from the combination of lesion studies (double dissociations) and hypothesis-based factorial designs in which only a few sensory or cognitive factors are varied at a time. In the last ten years, however, data-driven explorations of large neuroimaging datasets have allowed for a more agnostic approach, and led to a whole new view where segregated hierarchically organized modules seem to give way to continuous multidimensional representations, with e.g. a distributed semantic system. While both approaches have brought about significant contributions to speech and language neuroscience, making coherent sense of them represents a substantial challenge. In this review article, we synthesize methodological and experimental findings from the speech and language neuroscience literature, dissecting strength and pitfalls of each approach and suggesting ways in which approaches could be integrated.

Main

Verification of testable hypotheses serves as the fundamental underpinning of the scientific method. Traditionally, cognitive neuroscience relies on a deductive approach, wherein theoretical models are tested through controlled experiments. However, deductive reasoning alone is often insufficient to make discoveries as the conclusions are implicitly contained in, or restricted to, the hypotheses. In contrast, inductive reasoning involves detecting regularities among observations to infer rules and propose new models. Over the past decade, inductive reasoning relying on the advent of big data and the possibility to train algorithms has increasingly influenced cognitive neuroscience, leading to the discovery of latent structures in the data such as multidimensional and spatially distributed representations. However, *data-driven* analyses alone do not provide a compelling understanding of how these representations relate to cognitive processes, and in particular of how causal they are to behavior. In this article, we outline a methodological taxonomy designed to facilitate the interpretation of results in speech and language neuroscience and foster the integration of the various statistical approaches.

Overcoming Challenges in Integrating Hypothesis-Based and Data-Driven Approaches

Traditionally, the hypothesis-based approach in neuroscience consists in relating neural responses to controlled stimuli or behaviors, assuming that only the changes in neural activity associated with one or more pre-specified variables can reliably be interpreted. This approach, based on explicit – mostly univariate – tests of cognitive or psychological models, is often criticized for being constrained to the controlled manipulation of artificial stimuli and conditions with limited generalizability. Yet, it has confirmed the notion that speech-related cognitive operations emerge from modular processing hierarchies where each step has its own computational specificity (Chevillet et al., 2011; Dewitt & Rauschecker, 2012; Formisano et al., 2003; Morillon & Baillet, 2017; Sheng et al., 2018). In the speech and language domain (hereafter referred to as speech), a modular and hierarchically organized neural system enables human speakers to learn the mapping between strings of sounds and meanings, and to combine lexical-level representations in infinite ways to convey new ideas or concepts (Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Scott & Johnsrude, 2003).

In contrast, the data-driven approach makes fewer assumptions about the nature of the system involved, and uses methods that allow the data to be more freely analyzed. Rather than constraining analyses by theoretical knowledge, relationships between features are derived from the data itself. This approach is compatible with more naturalistic stimuli,

and enables exploration of brain responses to language 'as it is used' (Hamilton & Huth, 2018). However, the use of unsupervised exploration methods and less controlled stimuli considerably obstruct the interpretation of the findings.

Reconciling the two approaches seems possible as long as we consider the underlying methods on a continuum, from pure hypothesis-based to fully unsupervised data-driven approaches (Brunton & Beyeler, 2019), rather than as mutually exclusive categories. Likewise, it requires agreeing to provisionally revisit the modular and hierarchical view of speech processing.

In what follows, we outline a taxonomy crafted to streamline the integration and interpretation of results yielded by different approaches (**Figure 1**). The taxonomy discerns four statistical approaches: (1) **Perturbation studies** reveal causal relationships between neural activity and behavior outcomes, and specify the gross functional granularity of speech processing. (2) **Univariate models** decompose global neural activity into its elementary components, such as phonemic, syllabic, and semantic features. By paralleling pre-established theories, for instance, those derived from lesion studies, these models probe the hierarchical organization of neural processing. (3) **Multi*variable* models**[1] explore the convergence and redundancy of the aforementioned modules, potentially revealing multiple feature co-processing within the hierarchy. (4) **Multi*variate* models** can be used to examine the interactions between elementary components of the processing hierarchy, thereby identifying the spatio-temporal dynamics that underpins the functional connectivity pattern – an analytical step that specifies how neural elements work together as a system.

Drawing on these distinct paradigms, we then highlight the instrumental role of recurrent models. Eschewing a compartmentalized perspective, recurrent models stand out for their capability to identify common elements among the models outlined in our taxonomy. This synthesized-oriented approach aims to harmonize these various approaches, ultimately resulting in a unified model for speech processing.

[1] There is often confusion between the terms multivariable and multivariate (Grant et al., 2019; Hidalgo & Goodman, 2013). Multivariate models refer to the relationship between several dependent variables and one or several independent variables, while multivariable models focus on several independent variables (i.e., multiple features of a stimulus) being regressed onto a single dependent variable (i.e., the neural activity). Thus, multivariable and multivariate models can be used for different purposes.

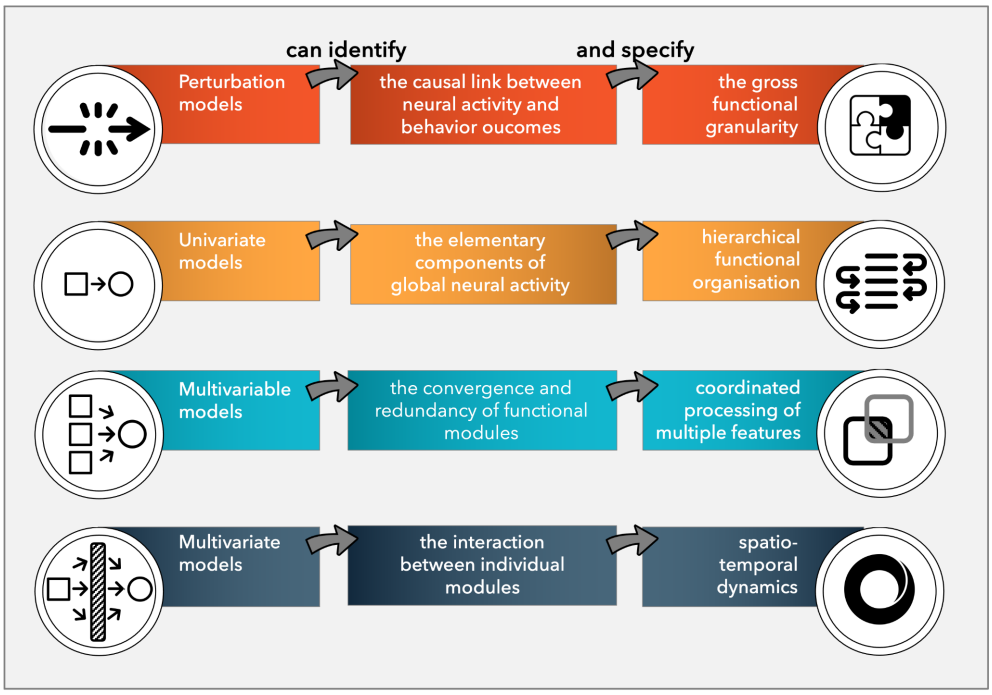

Figure 1. A taxonomy of the methods used in speech and language neuroscience to integrate interpretations derived from various statistical approaches. This interpretative guide outlines the potential benefits of using (1) perturbation models to identify causal links between neural activity and behavior and to specify gross functional granularity; (2) univariate models to decompose global neural activity into elementary components and to specify the hierarchical organization; (3) multivariable models to track the convergence and redundancy of functional modules and to specify coordinated processing of multiple features; and finally (4) multivariate models to probe interactions between individual modules and to characterize the underlying spatio-temporal dynamics. Situating each study within this taxonomy should assist in limiting conclusions to the inferential possibilities of the method used.

1. Perturbation models to address gross functional granularity

Exploring how brain lesions impact behavior stands as one of the most well-established and influential methodologies in the field of speech neuroscience. Dating back to the 19th century, case studies involving patients with focal brain damage demonstrated that specific language processes are contingent upon distinct brain regions (Broca, 1861; Wernicke, 1874) (Figure 2A). Over time, lesion-symptom mapping studies, as well as intervention and perturbation studies, further refined our knowledge of the functional neuroanatomy of language processes (Dronkers et al., 2017; Penfield & Roberts, 1959) (Figure 2B). Today, these approaches enable researchers to derive robust observations on the underlying causal relationships between neural mechanisms and their associated functions in speech processing (Peters et al., 2017; Weichwald & Jonas, 2021). Crucially, studies where perturbations are deliberately induced to provoke neural changes are key to addressing the therapeutic value of an intervention.

Perturbation studies can take advantage of chronic or acute disorders such as neurodegenerative atrophy (Chapman et al., 2023; Rogalski et al., 2011) or ischemic stroke (Bouton et al., 2018), alongside clinical interventions such as surgical lesions (Fridriksson et al., 2018; Halai et al., 2018; Hamilton et al., 2021), electrophysiological stimulation (Cardenas et al., 2020; Devlin & Watkins, 2007; Marchesotti et al., 2020; Silva et al., 2022), and cooling (Banerjee et al., 2021; Long et al., 2016). These methods, whether relying on permanent damage or temporary disruptions of neural activity, provide a direct estimate of a specific brain region's role in speech behavior (Fridriksson et al., 2018; Sperber, 2020; Vaidya et al., 2019). Importantly, the clear interpretability of their results is invaluable, especially given the complexities associated with neural processing in speech.

In particular, intervention studies have played a pivotal role in elucidating the functional granularity of the speech hierarchy, reconciling inconsistencies arising from correlation studies alone. A noteworthy example pertains to the ongoing debate regarding the functional segregation of lexico-semantic and syntactic processing across different brain modules. While early studies suggest distinct functional modules for syntax and semantics (Hickok & Poeppel, 2007; Matchin & Hickok, 2020), more recent ones highlight the concurrent involvement of a shared brain regions set (Fedorenko et al., 2020). Lesion-behavior mapping has served to reconcile these seemingly incongruous findings by revealing region-specific functional effects (Figure 2C): disorders of syntactic and semantic comprehension manifest subsequent to damages within distinct subparts of the posterior middle temporal gyrus (pMTG) (Matchin et al., 2022).

At each functional granularity level, perturbations, whether spontaneous or provoked, allow establishing a causal link between neural processing and behavioral outcomes. While this approach yields clear-cut results, it is, however, hampered by the difficulty of finding pure double dissociation cases that generalize to any new behavioral task, especially for modules higher up the speech hierarchy (Van Orden et al., 2001). Such constraints underscore the potential merit of alternative, data-driven approaches, which do not necessarily hypothesize a strict modular and sequential organization for speech processing but are poised to capture the intricacies of complex neural interactions more effectively.

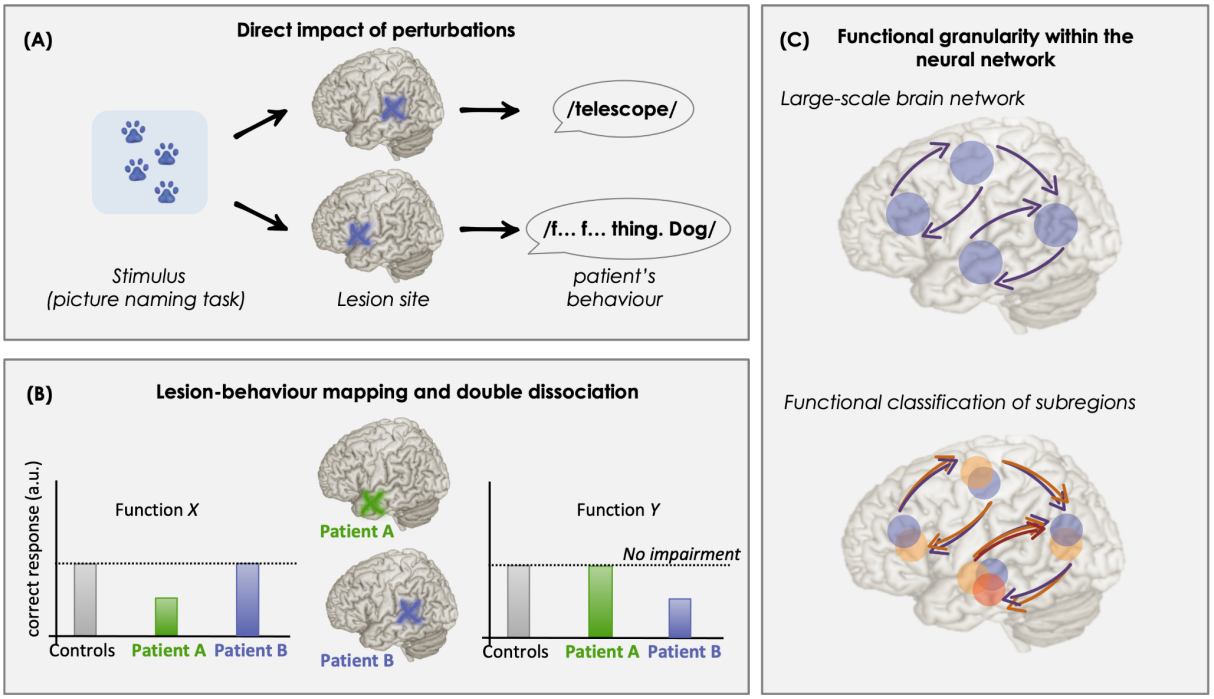

Figure 2. Leveraging perturbation studies for refined causal granularity of speech and language processing. (A) Perturbation studies directly probe the impact of local neural activity on a given behavior. e.g. damage to either the posterior superior temporal gyrus (pSTG) or the inferior frontal gyrus (IFG) can impair picture naming in different ways: a lesion in the pSTG can lead to fluent but nonsensical speech, where patients might say a word unrelated to the picture or even create non-existent words (top); a lesion in the pSTG can lead to non-fluent speech where participants might struggle to produce the name of the object in the picture, even if they know what it is (bottom). (B) A crossover double dissociation strengthens the level of evidence, by demonstrating that if damage to region A causes loss of function X but not Y, and damage to B causes loss of function in Y but not X, X and Y functions are underpinned by distinct brain modules. (C) Perturbation studies can provide fine-grained insights into functional modularity within networks of interconnected regions, e.g. by identifying and describing causal factors operating at different spatial scales within the neural network.

2. Univariate models to probe functional specialization of hierarchical modules

A first alternative for more systematic testing of speech functional modules is to relate changes in neural response to variations in a controlled stimulus, assuming that any change in the dependent variable (the neural response) is primarily explained by a controlled manipulation of the independent variable (e.g., the stimulus, the task, the context) (Figure 3A). The univariate approach is well suited to test predefined hypotheses about the relationship between specific components of cognitive and psychological models and the underlying functional neuroanatomy (Figure 3B), as well as about the functional specificity of each stage in a processing hierarchy (Figure 3C).

Univariate encoding models have been widely used to characterize auditory systems, and provided evidence that information processing proceeds sequentially across hierarchically organized areas with representations being gradually abstracted along the

speech processing hierarchy. As a result, multiple gradients (e.g., tonotopy) and processing levels (core, belt and parabelt) have been identified in the human auditory cortex (Chevillet et al., 2011; Formisano et al., 2003; Talavage et al., 2000). Notably, representations of increasingly complex speech features are organized along an antero-lateral gradient in the temporal cortex, ranging from simple tones and noise bursts to pseudowords and words (Binder, 2000; Dewitt & Rauschecker, 2012). From there, both a ventral and a dorsal stream emerge, providing inputs to the inferior frontal gyrus and reflecting different methods of processing and combining speech units (Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Scott & Johnsrude, 2003).

Univariate encoding models have also enabled a detailed characterisation of the time scales involved in the speech processing hierarchy (Brennan et al., 2012; Friederici, 2012; Scott, 2000). A time-constrained univariate analysis provided precise temporal neural mapping of multiscale processing during a phonemic categorization task (Bouton et al., 2018). Neural activity associated with low-level perceptual processing was disentangled from higher-level phonemic identification, and selective activity in temporal regions for low-level acoustic features was found 50ms before higher-order decision activity in the left inferior frontal region. In this specific case, the use of univariate time-resolved models allowed for a stepwise spatio-temporal examination of the different computational stages that constitute the speech processing hierarchy, paving the way for a comprehensive understanding of the process.

While the hypothesis-based approach, where only one feature varies at a time, enables straightforward interpretation of the results, univariate outcomes from single experiments are insufficient to capture the multi-stream, distributed nature of language processing. They also fall short in untangling the simultaneous contributions of different speech features to the examined neural activity. To achieve a more comprehensive picture of the multi-layered speech processing system, the use of methods adapted to its intrinsic complexity is necessary. This includes multivariable models, which can probe the relationships between multiple variables simultaneously.

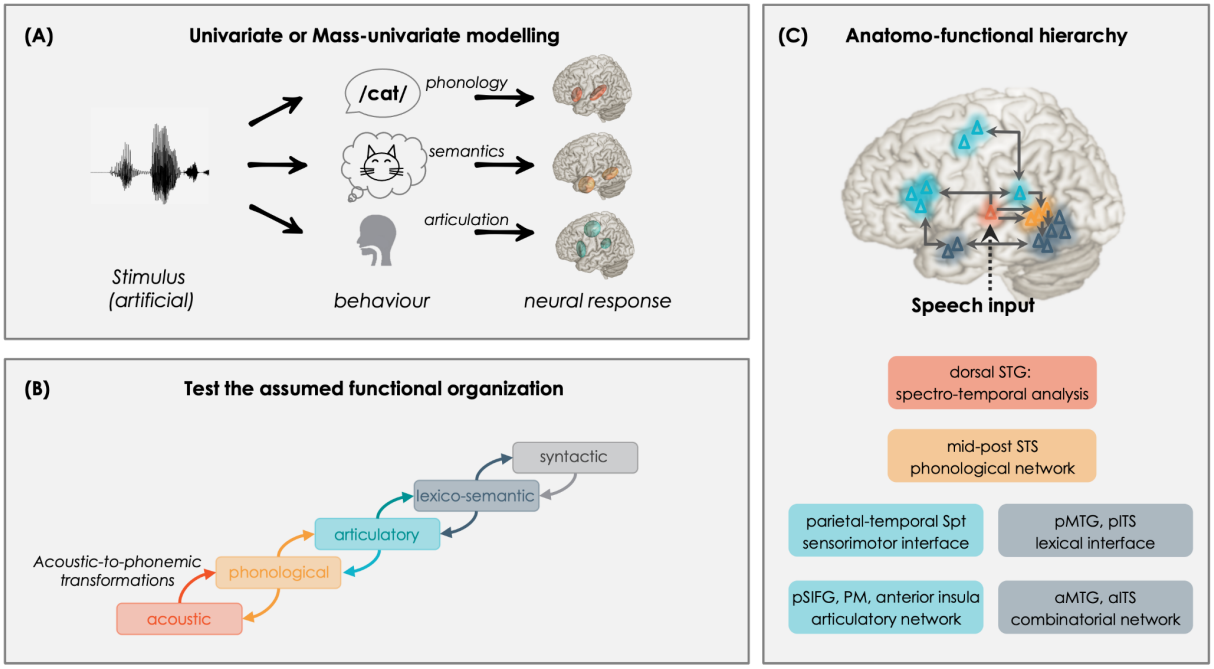

Figure 3. Probing speech and language processing hierarchies using univariate models. (A) Univariate, or mass-univariate, models establish a stimulus-response function relating neural activity to a stimulus variable, such as an articulatory feature. This hypothesis-based approach uses a few pre-specified stimulus features or behavioral variables that can be regressed in or of no interest (regressed out). (B) Univariate models imply assumptions about the relevant variables and assume functional modularity in the brain. In this view, each region or network of regions primarily encodes specific, well-specified variables. Historically, the approach derives from the study of sensory systems but is somewhat less well adapted for characterizing high-level or endogenous functions. (C) Univariate (mass-univariate) models reveal a neuronal organization, in which low-level information (i.e., spectro-temporal cues), is transformed along the cortical hierarchy to incrementally build abstract linguistic structures (Figure adapted from Hickok & Poeppel, 2007).

3. *Multivariable models to explore multiple feature co-processing*

Multivariable models have become more prevalent in speech neuroscience research due to the growing use of naturalistic speech stimuli comprising nested features such as sentences, words, or syllables, all jointly contributing to a single neural measure. Multivariable models can disentangle these simultaneous effects by fitting a function that describes how a set of particular stimulus features contributes to neural responses (Crosse et al., 2016; Di Liberto et al., 2021) (Figure 4A). Features can be imported from a priori linguistic assumptions or extracted from data-driven methods such as dimensionality reduction or large language models fitted to speech data (Caucheteux et al., 2022; Sainburg et al., 2020; Schrimpf et al., 2021). This approach has confirmed that speech comprehension relies on hierarchically organized representations, spanning from the acoustic in primary auditory areas, to phonetic, semantic and higher-level (syntactic) linguistic representations when moving away from Heschl's gyrus (Caucheteux et al.,

2021; de Heer et al., 2017) to prefrontal regions via the double dorsal and ventral stream, a result consistent with previous findings from univariate models.

In particular, multivariable models further helped uncover the spatio-temporal specificities of neural activity in the processing hierarchy (Hamilton et al., 2021; Hullett et al., 2016; Palmeri et al., 2017; Pasley et al., 2012; Santoro et al., 2014), as they accommodate the notion that there is an optimal spatio-temporal processing scale at each step considered (Gross et al., 2013; Panzeri et al., 2010; Walker et al., 2011). For example, while a spectro-temporal receptive field (STRF) is the best organization for neural activity in the auditory cortex, an STRF further tuned to phonological cues proves to be a more efficient implementation in the auditory association cortex (Santoro et al., 2017; Venezia et al., 2019). The transformation of the best model throughout the hierarchy reflects the progression from fine temporal resolution in low-level acoustic processing regions, to associative regions of the STG and STS where syllabic and word-level processing regions require integration of longer sequences, and higher-level brain regions (e.g. in the anterior temporal cortex) where semantic representations likely become atemporal (Giraud, 2020), coded with sparse spatial coding (Chang et al., 2010; Deneux et al., 2016; Van Wassenhove, 2009; Q. Zhang et al., 2019). These results highlight the importance of taking into account the hierarchical heterogeneity of functional organization, and of using functionally-informed multivariable models (Figure 4B). This description level holds promise for neuroengineering applications, such as developing speech decoders that use neural activity at a particular level of speech processing to decipher distinct components corresponding to specific neural activations (Moses et al., 2019; Tang et al., 2023).

More importantly, by enabling direct estimation of concomitant neuronal processing, multivariable models allow for a precise specification of the combined representations of stimulus features at several levels of the hierarchy: mixtures of spectral and articulatory features in the superior temporal gyrus, mixtures of articulatory and semantic features in the superior temporal sulcus (Caucheteux & King, 2022; de Heer et al., 2017), etc. Here, the choice of regressed features is critical as it determines the nature of the information being integrated. For instance, when the multiple variables to be regressed with neural data come from the fitting of a deep neural network (DNN) to behavior, the variables contain implicit information about the architecture and granularity of the hierarchical organization, i.e., *N* (inter)connected layers (Figure 4C). The number of layers in a DNN is a design choice reflecting the complexity of the task being modeled rather than any biological plausibility. There is hence a risk of smoothing out or missing anatomical (cytoarchitectonic) separations between functionally distinct brain regions when choosing the number of layers, thus creating spurious intermediate stages of processing and

representation. Although DNNs draw inspiration from biological neural networks, they are not direct replicas of biological systems, but rather extreme simplifications (Saxe et al., 2021). Fitting human behavior with DNNs may thus imply adding extra layers that have no equivalent in the brain. When seeking to map DNN layers to actual brain activity, it is therefore necessary to take into account the actual functional granularity of the cortical hierarchy.

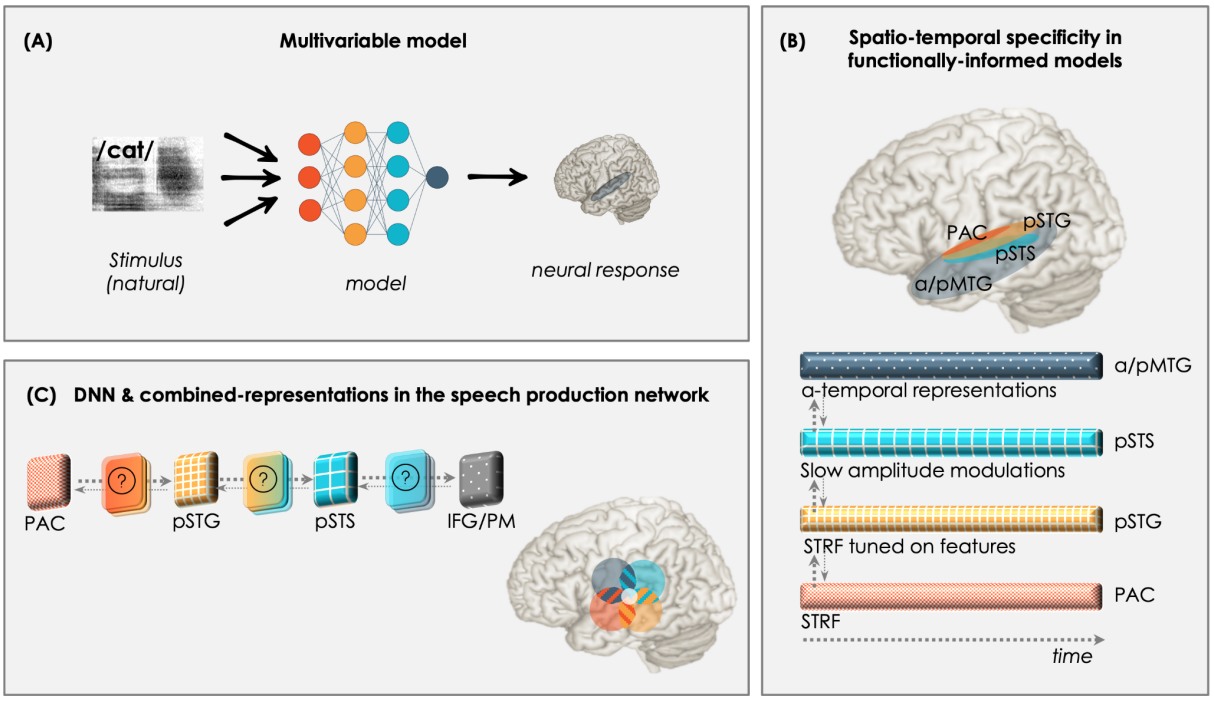

Figure 4. Neural processing involves concomitant neural processing and combined representations. (A) Multivariable models use multiple independent variables (stimulus features) to predict a single dependent variable, the neural response. (B) The multivariable model approach implies a priori assumptions about the spatial and temporal implementation of neural representations. These assumptions can be functionally informed by the specific computations performed at each level of the hierarchy. The versatility of these models allows them to address higher-order cognitive operations, so long as the inherent spatio-temporal characteristics of the associated encoding models are considered. (C) Mapping deep neural networks (DNNs) to neural data assumes concomitant processing of multiple features, revealing combined representations. The depth and complexity of DNNs are pivotal in representing linguistic complexity and abstraction. When inferring about brain-based language processing, the chosen architecture and layer count of the DNNs should be carefully considered.

4. Multivariate models to probe cross-module spatio-temporal dynamics

While multivariable models enable the simultaneous analysis of different speech (voice vs. speech content) and/or linguistic features (phonetic, phonological, lexical etc; see analysis-by-synthesis, (Poeppel & Monahan, 2011; Pulvermüller & Fadiga, 2010), identifying multi-network representations of speech requires consideration of processes distributed across brain regions and neuronal populations (Chalas et al., 2022; Correia et al., 2015; Lee et al., 2012). These patterns of distributed correlations cannot be revealed

with univariate or multivariable models (Ufer & Blank, 2023). In contrast, multivariate models, which relate one or several stimulus features to the activity of multiple simultaneous neural responses (Figure 5A), offer evidence that distributed patterns of neural activity contain discriminative information (Bishop & Nasrabadi, 2006; King et al., 2020).

Using multivariate models to show that (spatial) patterns of neural activity contain discriminative information proceeds from a different scientific approach than using them in neuroengineering to decode neuronal activity (Hebart & Baker, 2018; Weichwald et al., 2015). Decoding multivariate models aim to reconstruct stimuli from neural activity (K. Friston et al., 2008; Holdgraf et al., 2017; S. Martin et al., 2014) to help palliate deficient functions in patients with, for example, peripheral motor or cortical language diseases (Anumanchipalli et al., 2019; Moses et al., 2021; Willett et al., 2021, 2023). The use of *decoding* multivariate models in cognitive neuroscience requires a cautious approach to prevent misinterpretation of results, as the fitted decoding weights are not directly interpretable as brain activations, but depend on both the signal of interest and the nature of the noise (e.g. neural processing redundancy) present in the signal (Bouton et al., 2018; Haufe et al., 2014; Hebart & Baker, 2018; Kia et al., 2017).

Specific methods have been developed to overcome this difficulty. In the linear case, a valuable approach involves inverting fitted multivariate weights to estimate the encoding weights, enabling a direct interpretation of the resulting neural representation of the signal (Haufe et al., 2014). Applied to brain activity during natural speech listening, this method reveals that neural tracking of the speaker's voice pitch is concurrently influenced by acoustic and linguistic features (Kegler et al., 2022). This finding highlights the intricate interactions occurring during the elementary processing steps of speech perception. Another effective solution for interpreting multivariate models is to define regions of interest to spatially constrain the analysis (Çelik et al., 2019). Reliable information on the interactions between levels of elementary processing can be recovered as long as multivariate neural responses are combined at a common spatial resolution (Weichwald et al., 2015). For example, by examining the multivariate fitting performance using all voxels within each brain region of interest, low-level acoustic feature representations were shown to vary depending on whether the subject is engaged in a linguistic or a paralinguistic task (Rutten et al., 2019). This crucial finding aligns with interactive accounts of speech processing and opens up opportunities to further comprehend its dynamics.

Alternative approaches aim to improve the interpretability of multivariate analysis in the temporal domain. One such approach seeks to identify the distribution of available information at each point in time. This method can reveal neural reuse, wherein the same brain region's activity is involved sequentially for different purposes. In speech and

language processing, such neural reuse might play a crucial role in accumulating sensory evidence, for example during a dialogue until a response is triggered (Anderson, 2016; Skipper, 2014). While the use of multivariate methods is critical to ensure that neural reuse is not missed, recurrent activity is rarely causal to behavior but rather reflects downstream associations. For instance, spectro-temporal representations can be first activated during automatic acoustic processing, then used for discriminating one phoneme from another, and reused to discriminate phonological neighbors (e.g., /bear/-/pear/) (Figure 5B). Likewise, using multivariate models, letter reading was associated with a distributed network of successive and overlapping neural activities, implying sequential processing, maintaining and broadcasting increasingly rich representations across brain regions (Gwilliams & King, 2020).

Finally, at the very end of the hypothesis-based vs data-driven continuum, distributed speech components can be uncovered using completely unsupervised methods that rely on a priori measures of interest, such as neural variance, without any relation to sensory or cognitive features. Dimensionality reduction methods, such as principal component analysis, can be applied directly to neural data (Bondanelli et al., 2021; Cunningham & Yu, 2014). These methods reveal low-dimensional spaces of correlated neural activity, also known as manifolds, that account for a large part of neural variance during speech tasks. They might provide new insights about how neural activity, specific to each speech feature and distributed across several brain regions, is integrated into a common representation (Gallego et al., 2017; Perdikis et al., 2011; Stephen et al., 2023) (Figure 5C). Nevertheless a significant drawback of these methods remains the difficulty of interpreting the reduced dimensions in functionally meaningful terms.

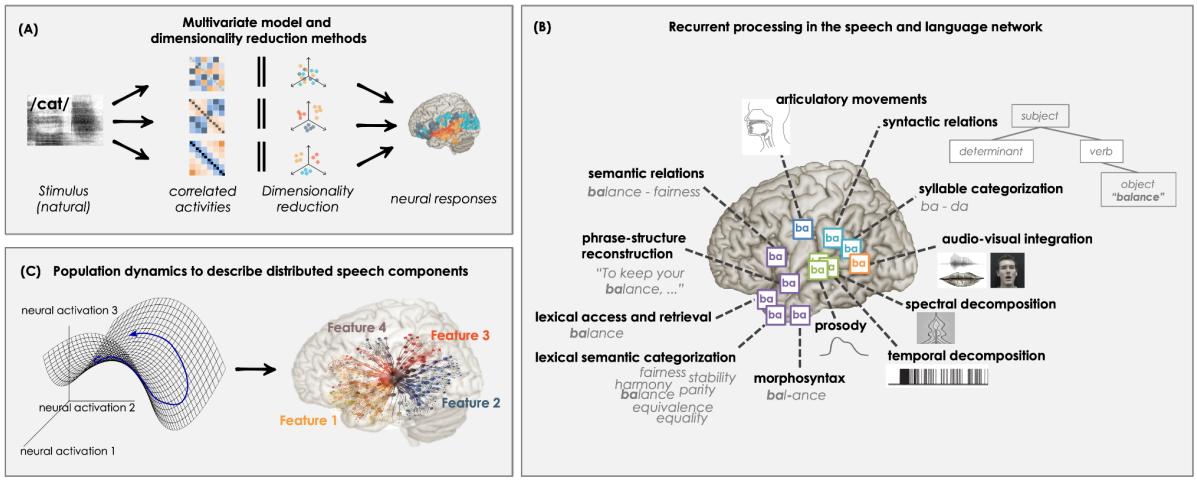

Figure 5. Speech and language processing involve distributed components that can be described through population dynamics. (A) Multivariate models rely on multiple independent variables to predict a set of dependent variables, including all possible relevant features of the stimulus. Feature selection can be achieved using methods such as correlation analyses or dimensionality reduction. (B) Being

non-selective, multivariate models can uncover processes of neural reuse. They can also identify distributed networks of sequentially organized neuronal activities. (C) Unsupervised dimensionality reduction method can help reveal underlying population dynamics (left). Projected features can be used to classify neural activity within the speech network (right).

5. Integrating hypothesis-based and data-driven methods

Used in isolation, the analytic methods presented so far enable the identification of distinct speech encoding modes and the characterization of their spatio-temporal distributions. However, these approaches can lead to contradictory conclusions about the functional architecture of speech processing. By combining these models, their individual advantages can be leveraged to get a more unified understanding of speech processing in the brain. In what follows, we highlight the complementary nature of these two approaches, but also how their respective results can inform each other to reveal new principles of neural organization, and finally how these approaches can be integrated together to provide a unified and detailed understanding of speech processing as it unfolds in natural situations.

Complementarity of statistical approaches

Understanding the functional granularity of the speech processing network poses significant challenges when relying on univariate causal models. Complex interactions among interconnected brain regions complicate these efforts (Dronkers et al., 2017; Rahimpour et al., 2019). Indeed, similar speech-deficit symptoms can stem from non-overlapping lesions in different patients (Na et al., 2022), as lesions can not only disrupt local gray-matter regions but also communication between network nodes by damaging white-matter tracts (Geschwind, 1965). Multivariate methods applied to lesion-behavior mapping help overcome these difficulties by taking into account the complex functional interactions between brain regions (Boes et al., 2015; Ivanova et al., 2021; Sperber, 2020; Y. Zhang et al., 2014) (Figure 6A). Lesion network mapping is particularly useful in understanding overall intervention's impact on the neural system by aligning brain lesions with multivariate functional networks (Boes et al., 2015; Fox, 2018; Siddiqi et al., 2022). Crucially, simultaneously introducing lesions and structural connectivity matrices as multivariable regressors can show how functional and structural networks predict behavioral scores. This method has been pivotal in identifying causal network connections and hubs, such as the temporo-parietal junction, a key region for speech fluency, auditory comprehension, speech repetition and oral naming (Yourganov et al., 2016).

Leveraging Hypothesis-Based and Data-Driven Approaches

Interpreting the functional relevance of data-driven reduced dimensions benefits greatly from hypothesis-based methods. This process entails an initial step of pinpointing neural activity associated with specific speech levels, followed by dimensionality reduction to unveil the core spatio-temporal patterns supporting these dynamics (Bouchard et al., 2013; Keshishian et al., 2023; Orepic et al., 2023). Recently this method was applied to demonstrate that the anterior STG simultaneously encodes phonetic and semantic features during natural speech perception, positioning the anterior STG as a promising hub for integrating these two levels of the speech processing hierarchy (Orepic et al., 2023). The synergy of hypothesis-based and data-driven approaches also becomes evident when hypothesis-based findings from controlled experiments are validated using data-driven methods on new datasets collected under varying experimental conditions or during natural conversations (Castellucci et al., 2022; Orepic et al., 2023) (Figure 6B). Relying primarily on hypothesis-based methods, Castellucci et al. identified a speech planning network in controlled experiments. Subsequently, they applied unsupervised dimensionality reduction tools to recover this network in natural conversations, demonstrating its consistency across different conditions.

Integration via recurrent models

To achieve a unified representation of the speech processing system, it is essential to integrate the levels identified (Lupyan & Clark, 2015). This integration must transcend the mere amalgam of speech features and brain regions through causal, multivariable, and multivariate models. Instead, it calls for consideration of how recurrent processes facilitate integration of speech signals over time (Gwilliams et al., 2022; Pulvermüller, 2018; Yi et al., 2019) . This integration requires turning to the concepts of recurrent and dynamical models, where past neural activity shapes future neural dynamics (K. J. Friston et al., 2003; Truccolo et al., 2004; Vyas et al., 2020). Recurrent models make it possible to examine multiple variables at once, and how different brain regions and speech features influence neural activity, while preserving causality, interpretability, and testability (Figure 6C). When applied to auditory cortex neural activity, recurrent models reveal how different auditory contents are represented in the brain by exposing hidden low-dimensional, dynamical structures (Bondanelli et al., 2021). Dynamic recurrent models have the intriguing feature of allowing the definition of distinct time constants for various sub-processes, thus forming a hierarchy of time scales that effectively integrates the distinct stages of the speech processing hierarchy (A. E. Martin, 2020; Perdikis et al., 2011; Pillai & Jirsa, 2017). More generally, this framework has the potential to characterize neural activity during tasks beyond speech, such as word writing or similar dynamic tasks (Perdikis et al., 2011).

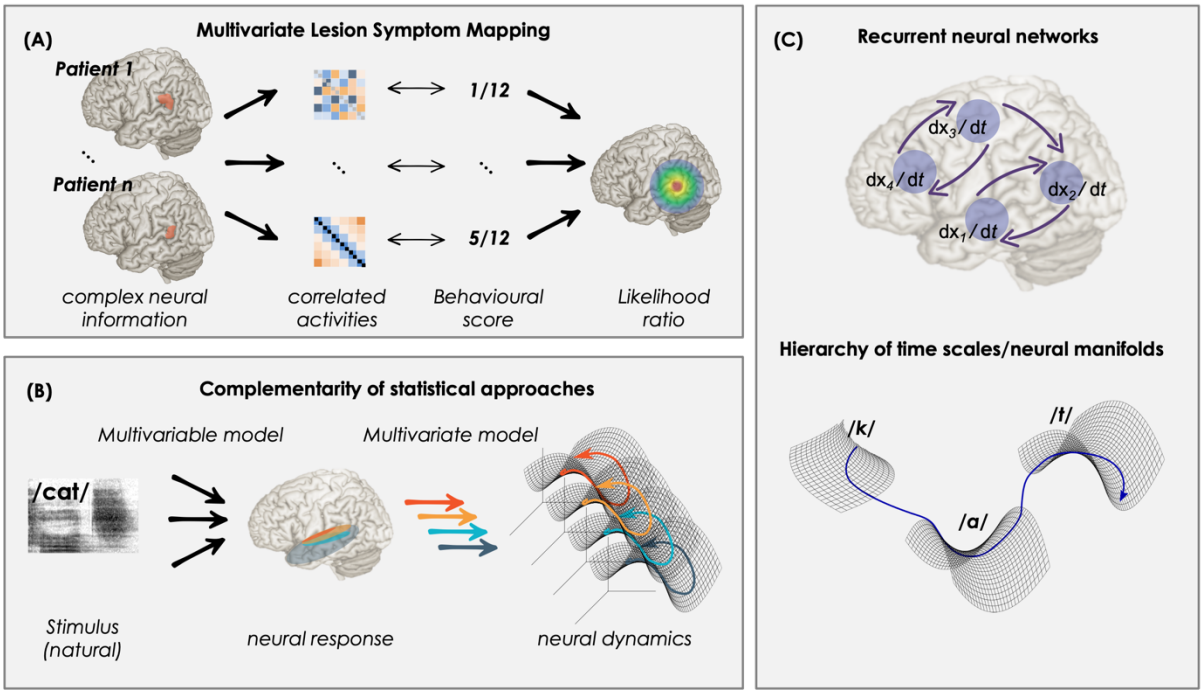

Figure 6. Speech processing involves integrating representations through recurrent activations. (A) Multivariable lesion symptom mapping, which consists of introducing multiple lesioned voxels or connectivity measurements as independent variables, allows for the consideration of the effect of the lesioned speech network on behavioral deficits. Multivariable lesion symptom mapping does not require patients to be grouped by either lesion site or behavioral cutoff, but instead makes use of continuous behavioral and lesion information. (B) Two-steps analysis helps in interpreting distributed neural patterns obtained after applying dimensionality reduction: first, an encoding multivariable model is applied to identify which neural activity relates to which speech features across the whole recording, and secondly, the resulting dimensionality is reduced to ease interpretation. (C) Recurrent models allow for setting dynamics at different time scales that integrate speech features across levels of the hierarchy.

Conclusion

We propose that the speech and language function in the human brain can be fully characterized by five cardinal approaches. These complementary approaches illustrate the strengths and limitations of hypothesis-based and data-driven methods. While perturbation models demonstrate causal links between neural processes and behavior, providing a clear understanding of the speech processing modularity, univariate models allow global neural activity to be broken down into elementary components or modules, improving our understanding of the hierarchical organization of neural processing. Multivariable models complement this approach by tracking multiple feature convergences and identifying edges with shared representations within the hierarchy. Multivariate models can reveal spatially and temporally distributed components and their

dynamics. Finally, recurrent models give an account of how different speech features are integrated over time within the speech hierarchy.

This article stresses the need to integrate the exploratory power of data-driven methods with the hypothetico-deductive approach. Indeed, data-driven research is constantly navigating between the extraordinary possibilities it offers for new discoveries and the pitfalls of random-walk science. Attention to past research and theoretical concerns can help prevent this random walk from getting lost in a vast, ultimately uninteresting space. This tension between deductive and inductive approaches is far from new in science. Two decades ago, science-fiction writer Ted Chiang wrote a premonitory article in *Nature* that purported to celebrate the 25th anniversary of the last human scientific publication (Chiang, 2000). In this provocative futuristic short-story, science is exclusively conducted by genetically improved humans, called "metahumans", and all that remains for human scientists is the interpretation of meta-human data. It is tempting to think of metahuman science as the outcome of data-/AI-driven methods, whose inner functioning lies sometimes beyond the reach of our understanding. But data are rarely, if ever, self-explanatory. They make sense to the extent that they validate or falsify theoretically informed predictions. "Human researchers", Chiang says somewhat presciently, "may indeed discern applications overlooked by metahumans, whose advantages tend to make them unaware of *our* concerns".

References

Anderson, M. L. (2016). Précis of After Phrenology: Neural Reuse and the Interactive Brain. *Behavioral and Brain Sciences*, *39*, e120. https://doi.org/10.1017/S0140525X15000631

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, *568*(7753), 493–498. https://doi.org/10.1038/s41586-019-1119-1

Banerjee, A., Egger, R., & Long, M. A. (2021). Using focal cooling to link neural dynamics and behavior. *Neuron*, *109*(16), 2508–2518. https://doi.org/10.1016/j.neuron.2021.05.029

Binder, J. R. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, *10*(5), 512–528. https://doi.org/10.1093/cercor/10.5.512

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, Issue 4). Springer.

Boes, A. D., Prasad, S., Liu, H., Liu, Q., Pascual-Leone, A., Caviness, V. S., & Fox, M. D. (2015). Network localization of neurological symptoms from focal brain lesions. *Brain*, *138*(10), 3061–3075. https://doi.org/10.1093/brain/awv228

Bondanelli, G., Deneux, T., Bathellier, B., & Ostojic, S. (2021). Network dynamics underlying OFF responses in the auditory cortex. *eLife*, *10*, e53151. https://doi.org/10.7554/eLife.53151

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–332. https://doi.org/10.1038/nature11911

Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A. G., Seeck, M., Karkar, S., Ville, D. V. D., & Giraud, A.-L. (2018). Focal versus distributed temporal cortex activity for speech sound category assignment. *PNAS*, *115*(6), E1299–E1308. https://doi.org/10.1101/133272

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*(2), 163–173. https://doi.org/10.1016/j.bandl.2010.04.002

Broca, P. (1861). Nouvelle observation d'aphémie produite par une lésion de la troisième circonvolution frontale. *Bulletins de La Société D'anatomie*, *6*(2), 398–407.

Brunton, B. W., & Beyeler, M. (2019). Data-driven models in human neuroscience and neuroengineering. *Current Opinion in Neurobiology*, *58*, 21–29. https://doi.org/10.1016/j.conb.2019.06.008

Cardenas, A. R., Behroozmand, R., Kocsis, Z., Gander, P. E., Nourski, K. V., Kovach, C. K., Ibayashi, K., Pipoly, M., Kawasaki, H., Howard, M. A., & Greenlee, J. D. (2020). *Differential causal involvement of human auditory and frontal cortices in vocal motor control* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.06.08.139881

Castellucci, G. A., Kovach, C. K., Howard, M. A., Greenlee, J. D. W., & Long, M. A. (2022). A speech planning network for interactive language use. *Nature*. https://doi.org/10.1038/s41586-021-04270-z

Caucheteux, C., Gramfort, A., & King, J.-R. (2021). Model-based analysis of brain activity reveals the hierarchy of language in 305 subjects. *arXiv*, 10. arXiv. https://doi.org/2110.06078

Caucheteux, C., Gramfort, A., & King, J.-R. (2022). Deep language algorithms predict semantic comprehension from brain activity. *Scientific Reports*, *12*(1), 16327. https://doi.org/10.1038/s41598-022-20460-9

Caucheteux, C., & King, J.-R. (2022). Brains and algorithms partially converge in natural language processing. *Communications Biology*, *5*(1), 134. https://doi.org/10.1038/s42003-022-03036-1

Çelik, E., Dar, S. U. H., Yılmaz, Ö., Keleş, Ü., & Çukur, T. (2019). Spatially informed voxelwise modeling for naturalistic fMRI experiments. *NeuroImage*, *186*, 741–757. https://doi.org/10.1016/j.neuroimage.2018.11.044

Chalas, N., Daube, C., Kluger, D. S., Abbasi, O., Nitsch, R., & Gross, J. (2022). Multivariate analysis of speech envelope tracking reveals coupling beyond auditory cortex. *NeuroImage*, *258*, 119395. https://doi.org/10.1016/j.neuroimage.2022.119395

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical Speech Representation in the Human Superior Temporal Gyrus. *Nature Neuroscience*, *13*(11), 1428–1432. https://doi.org/10.1038/nn.2641.Categorical

Chapman, C. A., Polyakova, M., Mueller, K., Weise, C., Fassbender, K., Fliessbach, K., Kornhuber, J., Lauer, M., Anderl-Straub, S., Ludolph, A., Prudlo, J., Staiger, A., Synofzik, M., Wiltfang, J., Riedl, L., Diehl-Schmid, J., Otto, M., Danek, A., FTLD Consortium Germany, … Schroeter, M. L. (2023). Structural correlates of language processing in primary progressive aphasia. *Brain Communications*, *5*(2), fcad076. https://doi.org/10.1093/braincomms/fcad076

Chevillet, M., Riesenhuber, M., & Rauschecker, J. P. (2011). Functional Correlates of the Anterolateral Processing Hierarchy in Human Auditory Cortex. *Journal of Neuroscience*, *31*(25), 9345–9352. https://doi.org/10.1523/JNEUROSCI.1448-11.2011

Chiang, T. (2000). Catching crumbs from the table. *Nature*, *405*, 517.

Correia, J. M., Jansma, B. M. B., & Bonte, M. (2015). Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions. *The Journal of Neuroscience*, *35*(45), 15015–15025. https://doi.org/10.1523/JNEUROSCI.0977-15.2015

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00604

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 1500–1509. https://doi.org/10.1038/nn.3776

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *Journal of Neuroscience*, *37*(27), 6539–6557. https://doi.org/10.1523/JNEUROSCI.3267-16.2017

Deneux, T., Kempf, A., Daret, A., Ponsot, E., & Bathellier, B. (2016). Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nature Communications*, *7*(1), 12682. https://doi.org/10.1038/ncomms12682

Devlin, J. T., & Watkins, K. E. (2007). Stimulating language: Insights from TMS. *Brain*, *130*(3), 610–622. https://doi.org/10.1093/brain/awl331

Dewitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *PNAS*, *109*(8), 505–514. https://doi.org/10.1073/pnas.1113427109

Di Liberto, G. M., Nie, J., Yeaton, J., Khalighinejad, B., Shamma, S. A., & Mesgarani, N. (2021). Neural representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage*, *227*. https://doi.org/10.1016/j.neuroimage.2020.117586

Dronkers, N. F., Ivanova, M. V., & Baldo, J. V. (2017). What Do Language Disorders Reveal about Brain–Language Relationships? From Classic Models to Network Approaches. *Journal of the International Neuropsychological Society*, *23*(9–10), 741–754. https://doi.org/10.1017/S1355617717001126

Fedorenko, E., Blank, I. A., Siegelman, M., & Mineroff, Z. (2020). Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition*, *203*, 104348. https://doi.org/10.1016/j.cognition.2020.104348

Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.-F., Ugurbil, K., & Goebel, R. (2003). Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex. *Neuron*, *40*(4), 859–869. https://doi.org/10.1016/S0896-6273(03)00669-X

Fox, M. D. (2018). Mapping Symptoms to Brain Networks with the Human Connectome. *New England Journal of Medicine*, *379*(23), 2237–2245. https://doi.org/10.1056/NEJMra1706158

Fridriksson, J., den Ouden, D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., & Bonilha, L. (2018). Anatomy of aphasia revisited. *Brain*, *141*(3), 848–862. https://doi.org/10.1093/brain/awx363

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension.

Trends in Cognitive Sciences, *16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001
Friston, K., Chu, C., Mourão-Miranda, J., Hulme, O., Rees, G., Penny, W., & Ashburner, J. (2008). Bayesian decoding of brain images. *NeuroImage*, *39*(1), 181–205. https://doi.org/10.1016/j.neuroimage.2007.08.013
Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7
Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025
Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain : A Journal of Neurology*, *88*(2), 237–294. https://doi-org.inshs.bib.cnrs.fr/10.1093/brain/88.2.237
Giraud, A.-L. (2020). Oscillations for all ¯_(ツ)_/¯? A commentary on Meyer, Sun & Martin (2020). *Language, Cognition and Neuroscience*, *35*(9), 1106–1113. https://doi.org/10.1080/23273798.2020.1764990
Grant, S. W., Hickey, G. L., & Head, S. J. (2019). Statistical primer: Multivariable regression considerations and pitfalls†. *European Journal of Cardio-Thoracic Surgery*, *55*(2), 179–185. https://doi.org/10.1093/ejcts/ezy403
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752
Gwilliams, L., & King, J.-R. (2020). Recurrent processes support a cascade of hierarchical decisions. *eLife*, *9*, 1–20. https://doi.org/10.7554/eLife.56603
Gwilliams, L., King, J.-R., Marantz, A., & Poeppel, D. (2022). Neural dynamics of phoneme sequences reveal position-invariant code for content and order. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-34326-1
Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2018). Triangulation of language-cognitive impairments, naming errors and their neural bases post-stroke. *NeuroImage: Clinical*, *17*, 465–473. https://doi.org/10.1016/j.nicl.2017.10.037
Hamilton, L. S., & Huth, A. G. (2018). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *35*(5), 573–582. https://doi.org/10.1080/23273798.2018.1499946
Hamilton, L. S., Oganian, Y., Hall, J., & Chang, E. F. (2021). Parallel and distributed encoding of speech across human auditory cortex. *Cell*, *184*(18), 4626-4639.e13. https://doi.org/10.1016/j.cell.2021.07.019
Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067
Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *NeuroImage*, *180*, 4–18. https://doi.org/10.1016/j.neuroimage.2017.08.005
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113
Hidalgo, B., & Goodman, M. (2013). Multivariate or Multivariable Regression? *American Journal of Public Health*, *103*(1), 39–40. https://doi.org/10.2105/AJPH.2012.300897
Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., Theunissen, F. E., & David, S. V. (2017). Encoding and Decoding Models in Cognitive Electrophysiology. *Frontiers in Systems Neuroscience*, *11*, 61. https://doi.org/10.3389/fnsys.2017.00061
Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. *The Journal of Neuroscience*, *36*(6), 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016
Ivanova, M. V., Herron, T. J., Dronkers, N. F., & Baldo, J. V. (2021). An empirical comparison of univariate versus multivariate methods for the analysis of brain–behavior mapping. *Human Brain Mapping*, *42*(4), 1070–1101. https://doi.org/10.1002/hbm.25278
Kegler, M., Weissbart, H., & Reichenbach, T. (2022). The neural response at the fundamental frequency of speech is modulated by word-level acoustic and linguistic information. *Frontiers in Neuroscience*, *16*, 915744. https://doi-org.inshs.bib.cnrs.fr/10.3389/fnins.2022.915744
Keshishian, M., Akkol, S., Herrero, J., Bickel, S., Mehta, A. D., & Mesgarani, N. (2023). Joint, distributed and hierarchically organized encoding of linguistic features in the human auditory cortex. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01520-0
Kia, S. M., Vega Pons, S., Weisz, N., & Passerini, A. (2017). Interpretability of Multivariate Brain Maps in Linear Brain Decoding: Definition, and Heuristic Quantification in Multivariate Analysis of MEG Time-Locked Effects. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00619
King, J.-R., Gwilliams, L., Holdgraf, C. R., Sassenhagen, J., Barachant, A., Engemann, D., Larson, E., & Gramfort, A. (2020). Encoding and Decoding Neuronal Dynamics: Methodological Framework to Uncover the Algorithms of Cognition. In *The Cognitive Neurosciences, 6th edition* (Poeppel, D., Mangun, G.R., Gazzaniga, M.S., pp. 691–702). The MIT Press.
Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical Speech Processing in Broca's Area: An fMRI Study Using Multivariate Pattern-Based Analysis. *The Journal of Neuroscience*, *32*(11), 3942–3948. https://doi.org/10.1523/JNEUROSCI.3814-11.2012
Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., Oya, H., Howard, M. A., & Greenlee, J. D. W. (2016). Functional Segregation of Cortical Regions Underlying Speech Timing and

Articulation. *Neuron*, *89*(6), 1187–1193. https://doi.org/10.1016/j.neuron.2016.01.032
Lupyan, G., & Clark, A. (2015). Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732
Marchesotti, S., Nicolle, J., Merlet, I., Arnal, L. H., Donoghue, J. P., & Giraud, A.-L. (2020). Selective enhancement of low-gamma activity by tACS improves phonemic processing and reading accuracy in dyslexia. *PLOS Biology*, *18*(9), e3000833. https://doi.org/10.1371/journal.pbio.3000833
Martin, A. E. (2020). A Compositional Neural Architecture for Language. *Journal of Cognitive Neuroscience*, *32*(8), 1407–1427. https://doi.org/10.1162/jocn_a_01552
Martin, S., Brunner, P., Holdgraf, C. R., Heinze, H., & Crone, N. E. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering*, *7*, 1–15. https://doi.org/10.3389/fneng.2014.00014
Matchin, W., Basilakos, A., Ouden, D.-B. den, Stark, B. C., Hickok, G., & Fridriksson, J. (2022). Functional differentiation in the language network revealed by lesion-symptom mapping. *NeuroImage*, *247*, 118778. https://doi.org/10.1016/j.neuroimage.2021.118778
Matchin, W., & Hickok, G. (2020). The Cortical Organization of Syntax. *Cerebral Cortex*, *30*(3), 1481–1498. https://doi.org/10.1093/cercor/bhz180
Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences*, *114*(42), E8913–E8921. https://doi.org/10.1073/pnas.1705373114
Moses, D. A., Leonard, M. K., Makin, J. G., & Chang, E. F. (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*, *10*(1), 3096. https://doi.org/10.1038/s41467-019-10994-4
Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty, M. E., Liu, P. M., Abrams, G. M., Tu-Chan, A., Ganguly, K., & Chang, E. F. (2021). Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. *New England Journal of Medicine*, *385*(3), 217–227. https://doi.org/10.1056/NEJMoa2027540
Na, Y., Jung, J., Tench, C. R., Auer, D. P., & Pyun, S.-B. (2022). Language systems from lesion-symptom mapping in aphasia: A meta-analysis of voxel-based lesion mapping studies. *NeuroImage: Clinical*, *35*, 103038. https://doi.org/10.1016/j.nicl.2022.103038
Orepic, P., Truccolo, W., Cash, S. S., Giraud, A.-L., & Proix, T. (2023). Low-dimensional neuronal population dynamics in anterior superior temporal gyrus reactivate phonetic representations during semantic processing [Preprint]. Neuroscience. https://doi.org/10.1101/2023.10.30.564638
Palmeri, T. J., Love, B. C., & Turner, B. M. (2017). Model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 59–64. https://doi.org/10.1016/j.jmp.2016.10.010
Panzeri, S., Brunel, N., Logothetis, N. K., & Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, *33*(3), 111–120. https://doi.org/10.1016/j.tins.2009.12.001
Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., & Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, *10*(1), e1001251. https://doi.org/10.1371/journal.pbio.1001251
Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton University Press.
Perdikis, D., Huys, R., & Jirsa, V. K. (2011). Time Scale Hierarchies in the Functional Organization of Complex Behaviors. *PLoS Computational Biology*, *7*(9), e1002198. https://doi.org/10.1371/journal.pcbi.1002198
Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms.* MIT Press.
Pillai, A. S., & Jirsa, V. K. (2017). Symmetry Breaking in Space-Time Hierarchies Shapes Brain Dynamics and Behavior. *Neuron*, *94*(5), 1010–1026. https://doi.org/10.1016/j.neuron.2017.05.013
Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, *26*(7), 935–951. https://doi.org/10.1080/01690965.2010.493301
Pulvermüller, F. (2018). Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology*, *160*, 1–44. https://doi.org/10.1016/j.pneurobio.2017.07.001
Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360.
Rahimpour, S., Haglund, M. M., Friedman, A. H., & Duffau, H. (2019). History of awake mapping and speech and language localization: From modules to networks. *Neurosurgical Focus*, *47*(3), E4. https://doi.org/10.3171/2019.7.FOCUS19347
Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331
Rogalski, E., Cobia, D., Harrison, T. M., Wieneke, C., Weintraub, S., & Mesulam, M.-M. (2011). Progression of language decline and cortical atrophy in subtypes of primary progressive aphasia. *Neurology*, *76*(21), 1804–1810. https://doi.org/10.1212/WNL.0b013e31821ccd3c
Rutten, S., Santoro, R., Hervais-Adelman, A., Formisano, E., & Golestani, N. (2019). Cortical encoding of speech enhances task-relevant acoustic information. *Nature Human Behaviour*, *3*(9), 974–987. https://doi.org/10.1038/s41562-019-0648-9
Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Computational Biology*, *16*(10), e1008228.

https://doi.org/10.1371/journal.pcbi.1008228
Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *PLoS Computational Biology*, *10*(1), e1003412. https://doi.org/10.1371/journal.pcbi.1003412
Santoro, R., Moerel, M., Martino, F. D., Valente, G., Ugurbil, K., Yacoub, E., & Formisano, E. (2017). Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proceedings of the National Academy of Sciences*, *114*(18), 4799–4804. https://doi.org/10.1073/pnas.1617622114
Saxe, A., Nelli, S., & Summerfield, C. (2021). If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, *22*(1), 55–67. https://doi.org/10.1038/s41583-020-00395-8
Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, *118*(45), e2105646118. https://doi.org/10.1073/pnas.2105646118
Scott, S. K. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406. https://doi.org/10.1093/brain/123.12.2400
Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100–107. https://doi.org/10.1016/S0166-2236(02)00037-1
Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., Huang, M.-X., Ding, N., & Gao, J.-H. (2018). The Cortical Maps of Hierarchical Linguistic Structures during Speech Perception. *Cerebral Cortex*, 1–9. https://doi.org/10.1093/cercor/bhy191
Siddiqi, S. H., Kording, K. P., Parvizi, J., & Fox, M. D. (2022). Causal mapping of human brain function. *Nature Reviews Neuroscience*, *23*(6), 361–375. https://doi.org/10.1038/s41583-022-00583-8
Silva, A. B., Liu, J. R., Zhao, L., Levy, D. F., Scott, T. L., & Chang, E. F. (2022). A Neurosurgical Functional Dissection of the Middle Precentral Gyrus during Speech Production. *The Journal of Neuroscience*, *42*(45), 8416–8426. https://doi.org/10.1523/JNEUROSCI.1614-22.2022
Skipper, J. I. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130297. https://doi.org/10.1098/rstb.2013.0297
Sperber, C. (2020). Rethinking causality and data complexity in brain lesion-behaviour inference and its implications for lesion-behaviour modelling. *Cortex*, *126*, 49–62. https://doi.org/10.1016/j.cortex.2020.01.004
Stephen, E. P., Li, Y., Metzger, S., Oganian, Y., & Chang, E. F. (2023). Latent neural dynamics encode temporal context in speech. *Hearing Research*, *437*, 108838. https://doi.org/10.1016/j.heares.2023.108838
Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, *150*(1–2), 225–244. https://doi.org/10.1016/S0378-5955(00)00203-3
Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*. https://doi.org/10.1038/s41593-023-01304-9
Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2004). A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. *Journal of Neurophysiology*, *93*(2), 1074–1089. https://doi.org/10.1152/jn.00697.2004
Ufer, C., & Blank, H. (2023). Multivariate analysis of brain activity patterns as a tool to understand predictive processes in speech perception. *Language, Cognition and Neuroscience*, 1–17. https://doi.org/10.1080/23273798.2023.2166679
Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., & Fellows, L. K. (2019). Lesion Studies in Contemporary Neuroscience. *Trends in Cognitive Sciences*, *23*(8), 653–671. https://doi.org/10.1016/j.tics.2019.05.009
Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, *25*(1), 111–172. https://doi.org/10.1207/s15516709cog2501_5
Van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1815–1830. https://doi.org/10.1098/rstb.2009.0023
Venezia, J. H., Thurman, S. M., Richards, V. M., & Hickok, G. (2019). Hierarchy of speech-driven spectrotemporal receptive fields in human auditory cortex. *NeuroImage*, *186*, 647–666. https://doi.org/10.1016/j.neuroimage.2018.11.049
Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation Through Neural Population Dynamics. *Annual Review of Neuroscience*, *43*(1), 249–275. https://doi.org/10.1146/annurev-neuro-092619-094115
Walker, K. M. M., Bizley, J. K., King, A. J., & Schnupp, J. W. H. (2011). Multiplexed and robust representations of sound features in auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(41), 14565–14576. https://doi.org/10.1523/JNEUROSCI.2074-11.2011
Weichwald, S., & Jonas, P. (2021). Causality in Cognitive Neuroscience: Concepts, Challenges, and Distributional Robustness. *Journal of Cognitive Neuroscience*, *33*(2), 226–247. https://doi.org/10.1162/jocn_a_01623
Weichwald, S., Meyer, T., Özdenizci, O., Schölkopf, B., Ball, T., & Grosse-Wentrup, M. (2015). Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage*, *110*, 48–59. https://doi.org/10.1016/j.neuroimage.2015.01.036
Wernicke, C. (1874). *Der aphasische Symptomenkomplex. Eine psychologische Studie auf anatomischer Basis*. Max Cohn & Weigert.

Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, *593*(7858), 249–254. https://doi.org/10.1038/s41586-021-03506-2

Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., Kamdar, F., Glasser, M. F., Hochberg, L. R., Druckmann, S., Shenoy, K. V., & Henderson, J. M. (2023). A high-performance speech neuroprosthesis. *Nature*, *620*(7976), 1031–1036. https://doi.org/10.1038/s41586-023-06377-x

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The Encoding of Speech Sounds in the Superior Temporal Gyrus. *Neuron*, *102*(6), 1096–1110. https://doi.org/10.1016/j.neuron.2019.04.023

Yourganov, G., Fridriksson, J., Rorden, C., Gleichgerrcht, E., & Bonilha, L. (2016). Multivariate Connectome-Based Symptom Mapping in Post-Stroke Patients: Networks Supporting Language and Speech. *The Journal of Neuroscience*, *36*(25), 6668–6679. https://doi.org/10.1523/JNEUROSCI.4396-15.2016

Zhang, Q., Hu, X., Hong, B., & Zhang, B. (2019). A hierarchical sparse coding model predicts acoustic feature encoding in both auditory midbrain and cortex. *PLoS Computational Biology*, *15*(2), 1–23. https://doi.org/10.1371/journal.pcbi.1006766

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom mapping using support vector regression: Multivariate Lesion Symptom Mapping. *Human Brain Mapping*, *35*(12), 5861–5876. https://doi.org/10.1002/hbm.22590