Neural infrastructure for incremental interpretation: Interfacing sound, meaning and constraint

Yuxing Fang, Bingjiang Lyu, Benedict Vassileiou, Kamen Tsvetanov, Lorraine K. Tyler, William D. Marslen-Wilson

Centre for Speech, Language and the Brain, Dept of Psychology, University of Cambridge (yf292@cam.ac.uk, wdm10@cam.ac.uk)
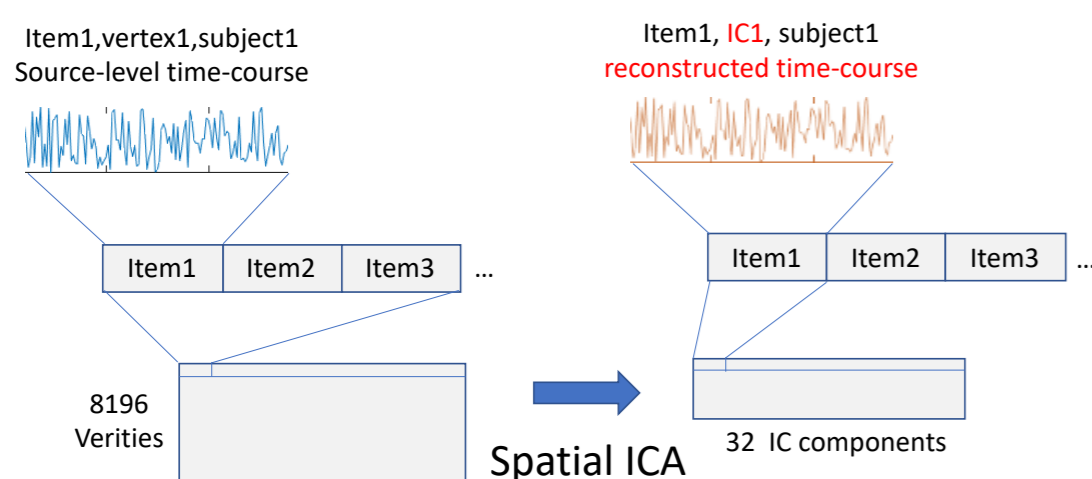
CSLB **erc** **UNIVERSITY OF CAMBRIDGE**

The remarkable speed and immediacy of human speech comprehension depends on the dynamic integration of **acoustic-phonetic cues** from the speech input with multi-dimensional **contextual constraints** to enable the early recognition of the words being heard and the integration of **lexical meaning** into the listener's incremental interpretation of the current utterance. This research probes the poorly understood neurocomputational infrastructure that supports these core **early integrative processes**. We aim to:

- Identify the critical brain regions involved;
- Determine how different cues and constraints are neurally represented,
- Measure the precise timing with which these different inputs are integrated.
- Investigate the directionality and the content of connectivity processes across the network.

To address these questions we combine source-localized EMEG measures of real-time brain activity, NLP models of linguistic form and content, multivariate Representational Similarity Analysis (RSA) and Grainger Causal Analysis (GCA) measures of connectivity with a network approach based on ICA and Lasso regression. These methods are applied to an existing source-reconstructed EMEG data set of short sentences of the form 'The elderly man ate the apple" (Lyu et al 2019), focusing on the first 250 ms after direct object onset (e.g., 'apple'), leading up to word Identification-point (IP).
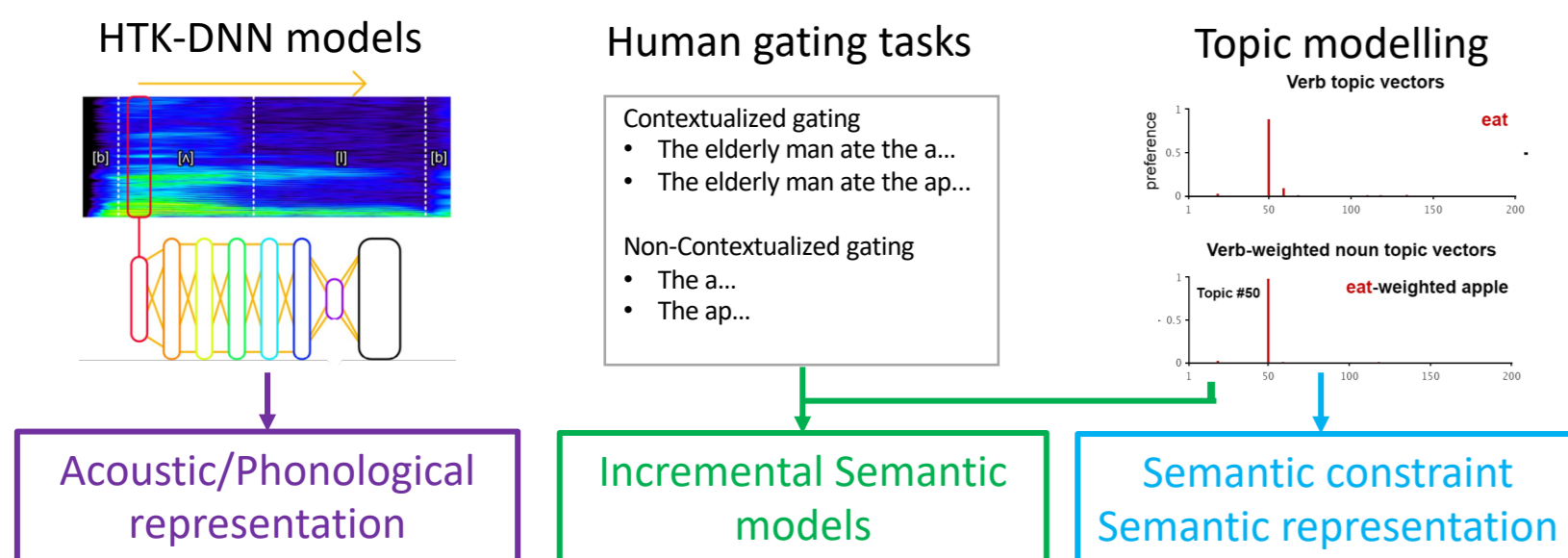
First (**Section 1**) we use an ICA whole-brain data-driven approach to decompose the source-reconstructed neural activity of the brain during the key epoch into functional spatially-independent components. In parallel (**Section 2**) we deploy ASR and NLP models of linguistic form and content in order to identify (**Section 3**) any ICA components that are sensitive to the knowledge-types involved in early integration and to determine (**Section 4**) the timing of different types of model fit across the epoch. We then (**Section 5**) use Lasso regression to decompose the selected network into major subcomponents (frontal and temporal) and then compute GCA connectivity between these components. Finally (**Sections 6 & 7**) we relate connectivity strength to phonetic and semantic models to probe the neurocomputational content of these connections.

① ICA decomposition into candidate networks

Item1,vertex1,subject1
Source-level time-course

Item1, IC1, subject1
reconstructed time-course

Item1 | Item2 | Item3 …

Item1 | Item2 | Item3 …

8196 Verities

Spatial ICA

32 IC components

To detect independent networks when processing speech, we first concatenated the source-localized signals from each item and extracted 32 independent components using ICA. To perform RSA analyses on these components, we re-constructed the signals of each component to restore the original item structure.

② Computational models for network selection and analysis

HTK-DNN models

Human gating tasks

Contextualized gating
- The elderly man ate the a…
- The elderly man ate the ap…

Non-Contextualized gating
- The a…
- The ap…

Topic modelling

Verb topic vectors

eat

Verb-weighted noun topic vectors

eat-weighted apple

Acoustic/Phonological representation

Incremental Semantic models

Semantic constraint Semantic representation

The acoustic-phonetic/phonological properties of the speech input and relevant semantic factors are derived from separate ASR and NLP models. Incremental semantic interpretation itself is captured by models of the semantic properties of the evolving word-initial cohort (estimated from gating studies) at different time-points (Gate1: 0-50ms, Gate2: 0-100ms, Gate3: 0-150ms) after word onset.

③ Identifying networks related to integration of sound, meaning, and context

The ICA process, segmenting brain activity on the basis of shared temporal coherence, generates multiple potential candidates. For a range of component numbers (illustrated here for n = 32) we used RSA to test each component for its sensitivity to the relevant model types.

The model fits for 32 IC networks

- Only the network labelled IC24 showed fit to the semantic models – most importantly to the incremental semantic models – as well as to the acoustic-phonetic and phonological models
- The IC24 network is strongly left lateralised, and implicates processing regions primarily in IFG and in posterior temporal cortex, including HG, STG and MTG
- IC24 is mirrored by a parallel RH network, IC16, with similar fronto-temporal loadings. This component only showed fit to acoustic-phonetic and phonological models

IC24 — acoustic/phonological and semantic processing

IC16 — Only acoustic/phonological processing
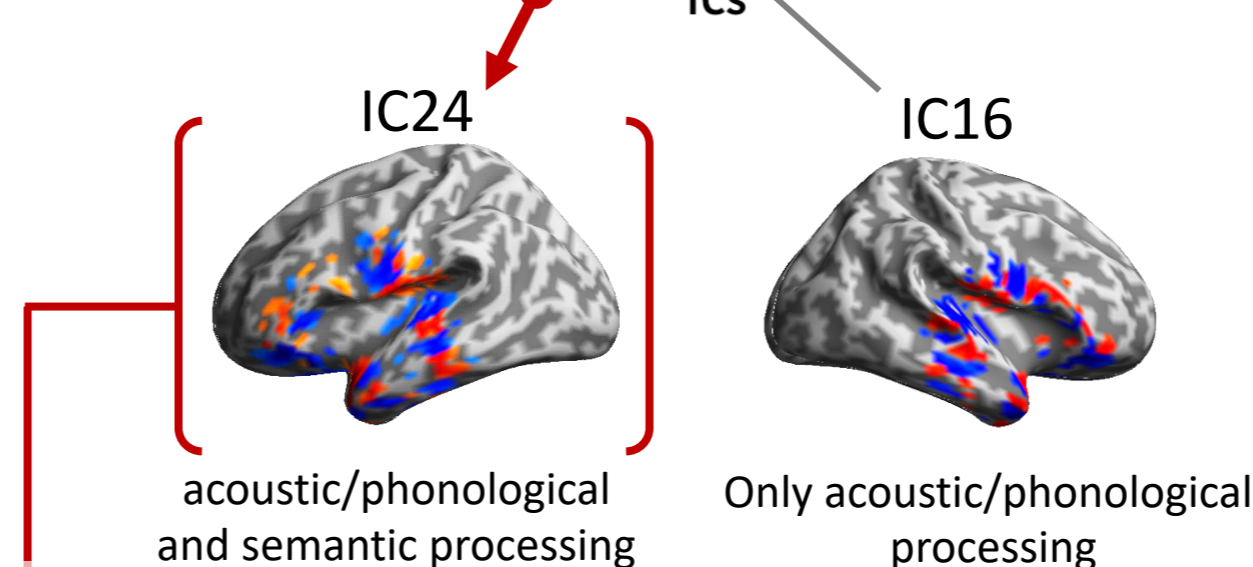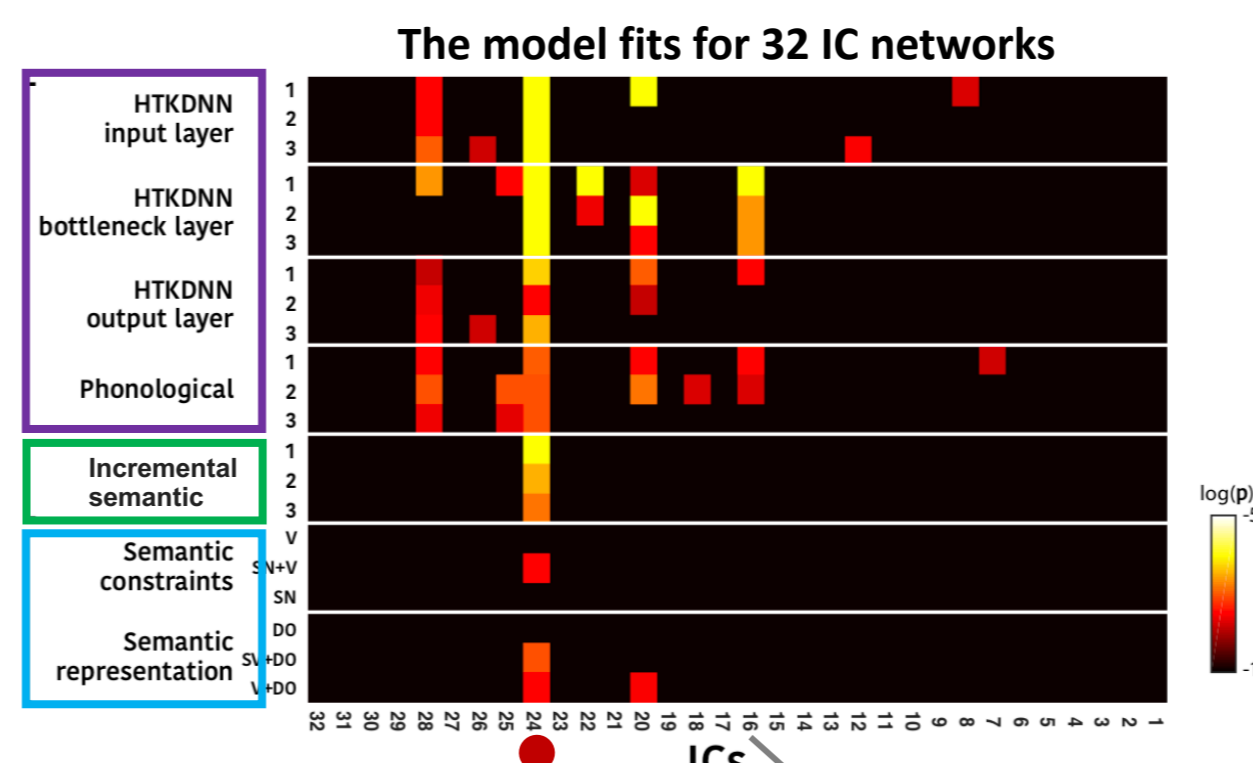
④ Temporal structure of IC24 model fits

A key question here is the timing with which incremental semantic models reflecting the word being heard start to emerge as a function of bottom-up cues from the speech input combined with contextual constraints.

- Onset of model fit (at 140 ms) for the uncontextualized incremental model is delayed relative to contextualized models (at 98 ms)
- Neither model shows model fit from word onset.
- Model fit for both terminates at noun IP (233 ms +/- 1 SD), when the word is fully identified.
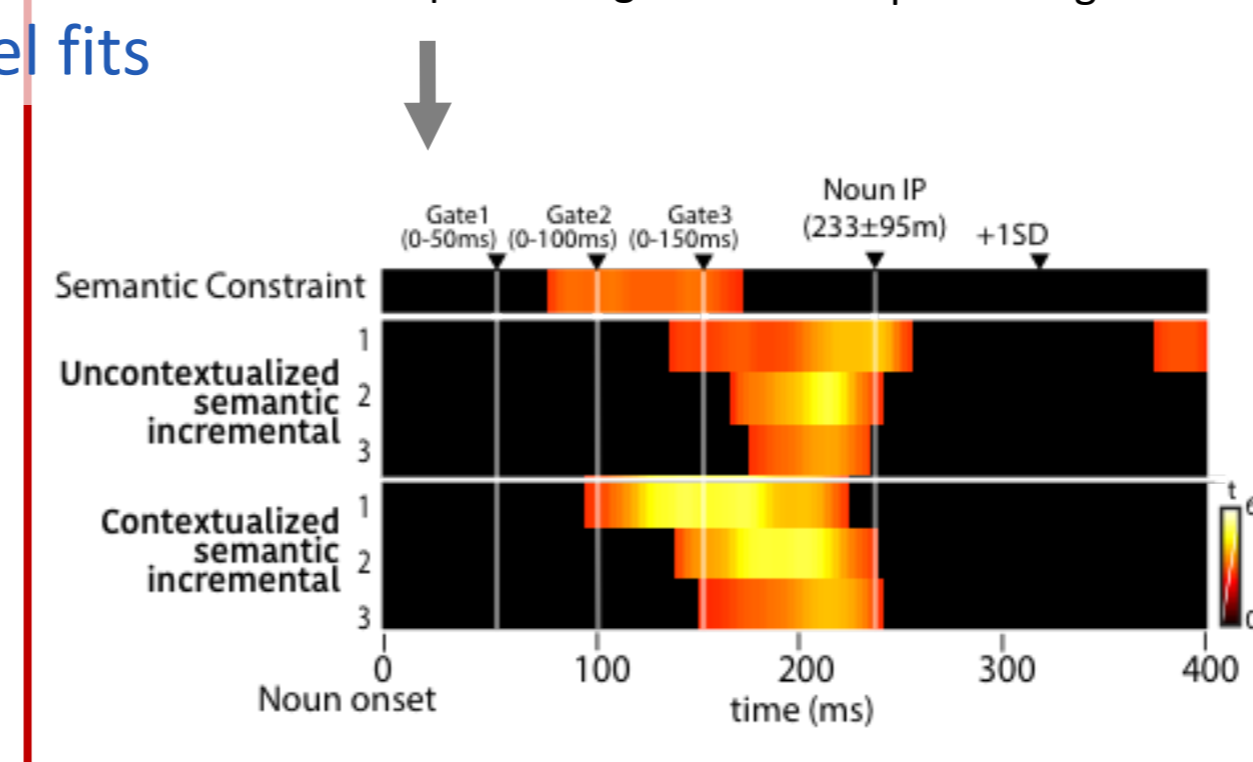
Decomposition
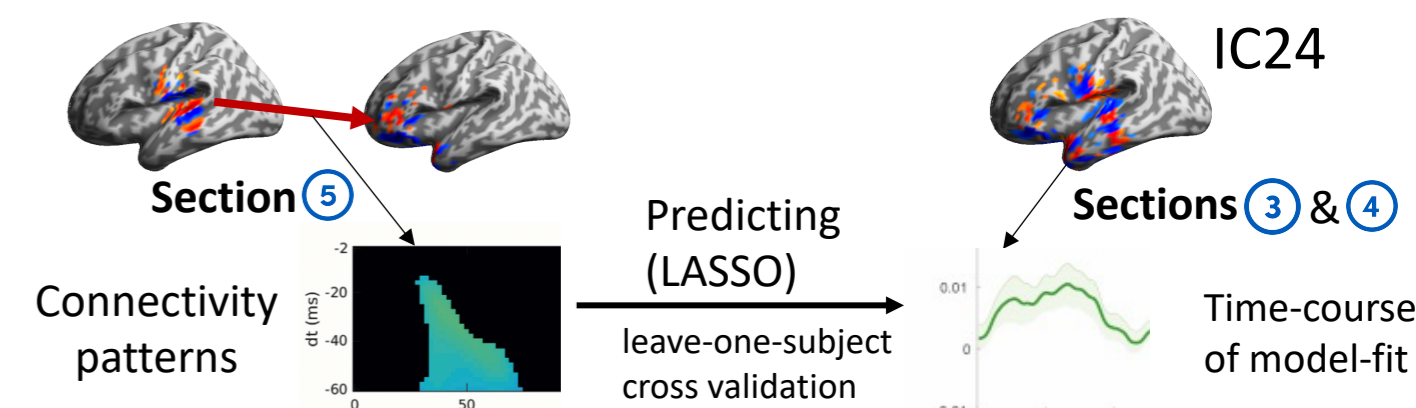
⑤ ICA network decomposition and GCA connectivity

To understand the inner structure of IC24, we used a data-driven method to determine the optimal segmentation of the network into potential subcomponents. The spatial ICA weights of IC24 can be represented as a linear combination of the sub-components (networks). The combination was solved by **lasso regression** while minimizing the number of sub-components - an IFG-based frontal network and an STG/MTG based temporal network.

The directional connectivity between these sub-networks networks was assessed using a pattern-based GCA. First we built data RDMs based on the time-varying IC signals from each network. We then calculated the correlation (ⓓ) between the RDMs from a certain time-point in Region A (ⓐ ,dt=-60ms) and Region B (ⓑ,dt = 0ms, partialling out the contribution to Region B of its own previous state. The resulting GCA-connectivity pattern captures the significant results from testing all dt combinations (from -2 to -60ms)

Frontal sub-network

Temporal sub-network

GCA connectivity

Data RDMs

Data RDMs

GCA-connectivity patterns
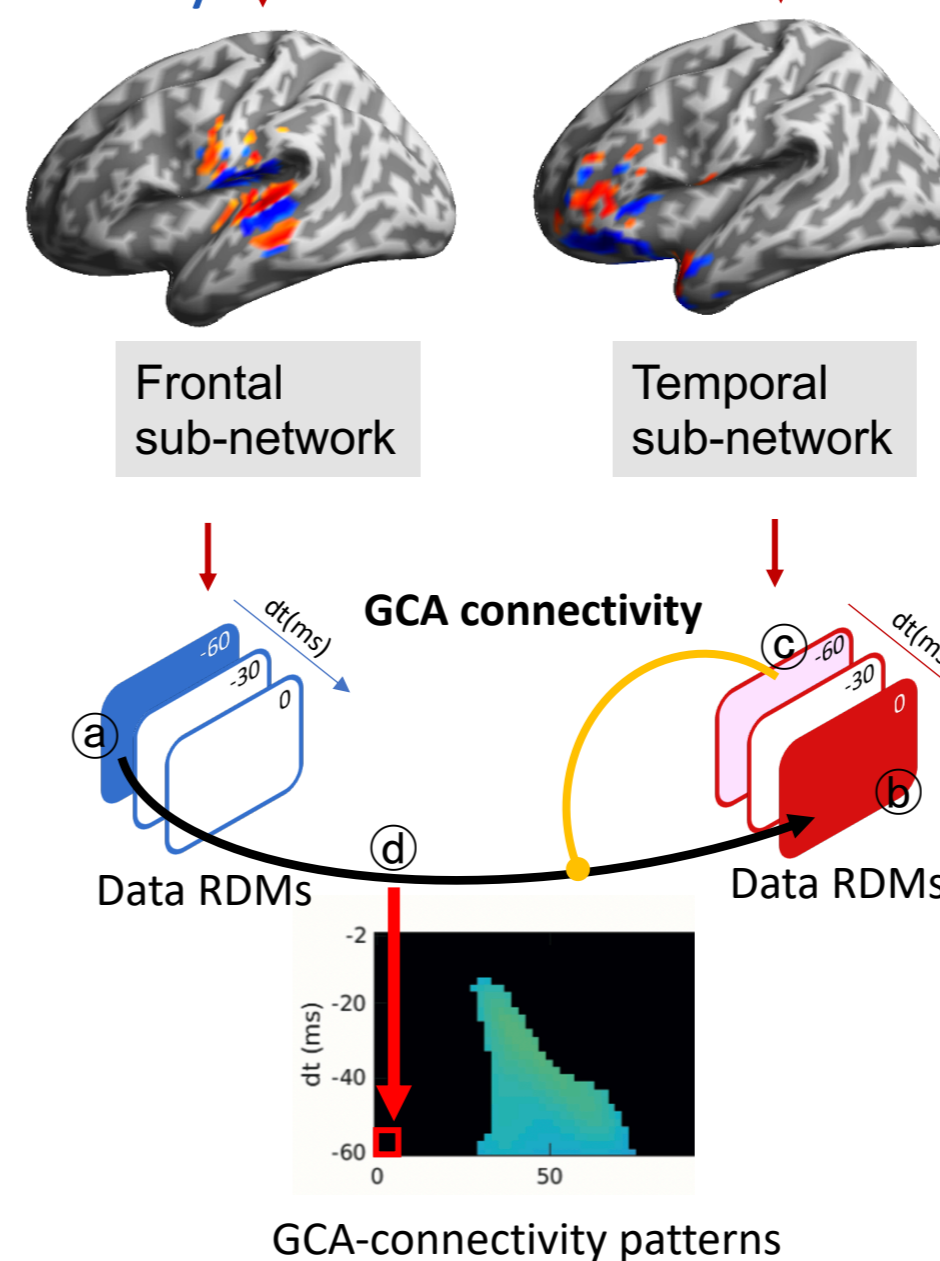
⑥ Informational content of network connectivity

To explore what kind of information is transferred across the IC24 network, we used the profile of connectivity patterns (from **section 5**) to predict the model-fits (**sections 3 & 4**) of the whole network over time. Significant model fit is interpreted as evidence about the informational role of the observed connectivity pattern

Section ⑤

Connectivity patterns

Predicting (LASSO) leave-one-subject cross validation

Sections ③ & ④

IC24

Time-course of model-fit

Applying these tests to the connectivity patterns between the two main subnetworks, we found that over the first 200 ms from word onset both acoustic-phonetic content and semantic content were bi-directionally transmitted between temporal and frontal regions.
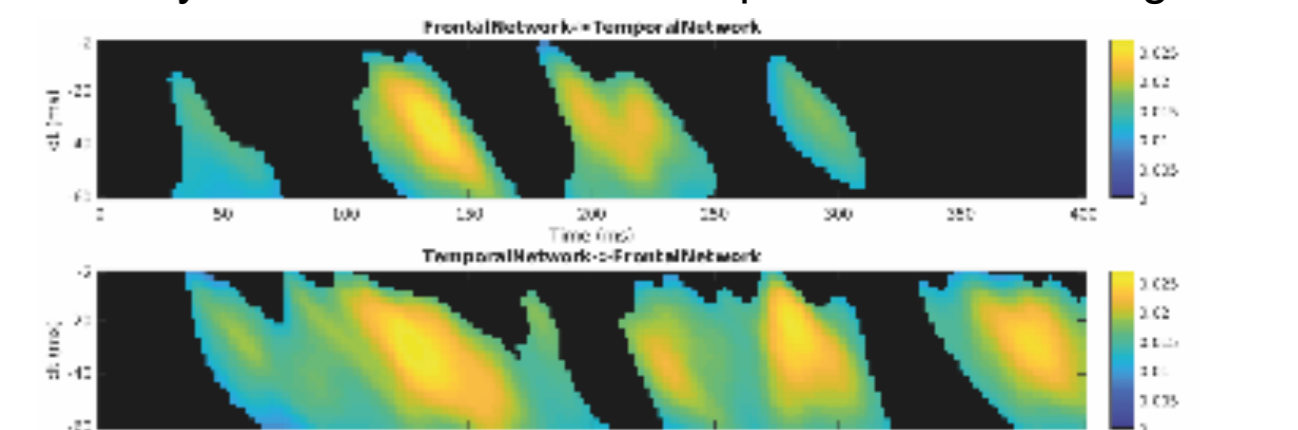
In a second set of analyses we extracted from the frontal component two subsets corresponding to BA 47 and BA 45.

BA47 — Temporal network

BA45 — Temporal network
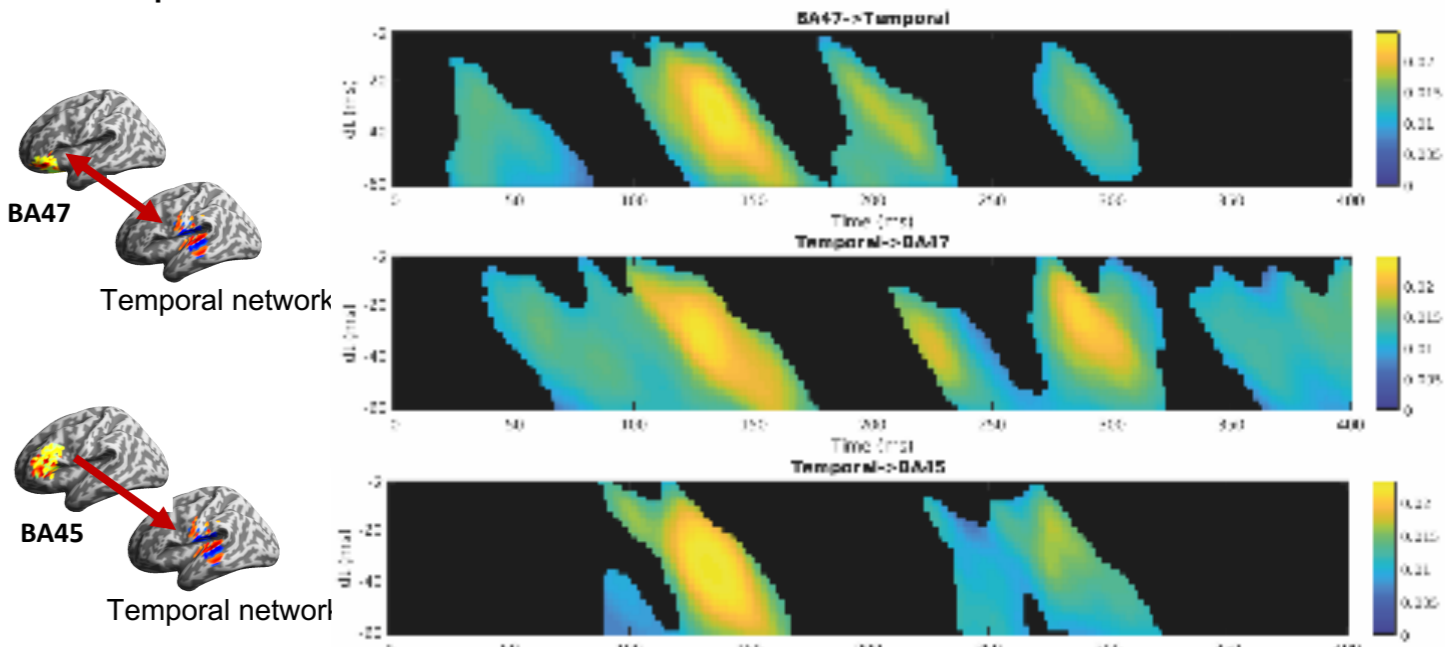
Here we see more differentiated effects. Strong bidirectional connectivity involving BA47 primarily links to semantic models, with very early outflow from BA47. Phonological information plays a minor and later role. BA45 vertices, in contrast, only show connectivity outward to temporal regions, starting relatively later and with model fit strongly dominated by acoustic-phonetic and phonological models.

Overview

- Model-based RSA evaluation of ICA components uniquely identifies a LH fronto-temporal network that integrates acoustic-phonetic cues and contextual constraints to support the identification of lexical form and meaning in the first 200 ms after word onset in sentential context

- Integration of acoustic-phonetic cues with contextual constraints is not visible at word-onset, suggesting that bottom-up cues are necessary to set the representational geometry of an analysis space with which semantic constraints can interact.

- The RH mirror component shows strong sensitivity to acoustic-phonetic and phonological models, but no systematic response to any semantic models. This indicates a qualitative difference in the hemispheric roles of parallel fronto-temporal components in the incremental interpretation of speech

- Innovative (but preliminary) analyses of the information content of directional connectivity within the LH network suggest markedly different roles for ventral and more dorsal regions of IFG

1. Lyu, B., Choi, H. S., Marslen-Wilson, W. D., Clarke, A., Randall, B., & Tyler, L. K. (2019). Neural dynamics of semantic composition. Proceedings of the National Academy of Sciences.