The Temporal Efficiency Paradox: Fast Brains React, Slow Brains Understand

Cuicui Jiang[1], **Xiaoyue Ding**[3], **Xingyue Liu**[4], **Jiaxuan Wei**[4], **Xiaoyan Li**[2,*]
[1]School of Computer Science, Inner Mongolia University, Hohhot, China
[2]Department of Automation, Tsinghua University, Beijing, China
[3]Beijing Normal University, Beijing, China
[4]School of Artificial Intelligence, Taiyuan University of Technology, Jinzhong, China

Abstract

The cortical temporal hierarchy—sensory regions process rapidly while association regions operate slowly—is well-documented, yet whether slow processing enables deeper integration remains debated. We analyzed fMRI data from four participants viewing ∼65 hours of naturalistic movies, computing Temporal Dynamics Speed (TDS) and encoding performance across seven functional networks. Networks differed significantly in TDS ($F(6, 993) = 18.42$, $p < .001$): Frontoparietal fastest (1.20), Ventral Attention slowest (0.64), Default Mode intermediate (0.78). We hypothesized slower networks would benefit more from extended temporal integration; however, this was not supported: temporal averaging impaired encoding across all networks (negative TWG, $p < .001$), with no relationship between TDS and integration benefit. Critically, LSTM models outperformed ridge regression—even with concatenated temporal features—by 47–57% uniformly across all networks ($p < .001$). These findings resolve the paradox: all networks benefit from temporal integration, but only when sequential dependencies are explicitly modeled. Code: https://github.com/xiaoyanLi629/temporal-efficiency-paradox.

Keywords: temporal dynamics; cortical hierarchy; LSTM; naturalistic neuroscience; encoding models

Introduction

Different brain regions process information at characteristically different temporal scales, forming what researchers have termed the cortical temporal hierarchy (Hasson et al., 2008; Murray et al., 2014). Sensory regions exhibit rapid temporal dynamics, with neural activity fluctuating at relatively high frequencies and responding to immediate environmental changes. In contrast, association regions, particularly those comprising the Default Mode Network (DMN), exhibit markedly slower temporal dynamics, integrating information across extended time periods (Honey et al., 2012; Lerner et al., 2011). This hierarchical organization has been documented through autocorrelation analysis, temporal receptive window estimation, and power spectral analysis (Chaudhuri et al., 2015; Gao et al., 2020). Recent work has further demonstrated that these intrinsic timescales are functionally dynamic and shaped by cortical microarchitecture (Gao et al., 2020), that global waves of activity propagate along this temporal hierarchy in coordination with arousal fluctuations (Raut et al., 2021), and that a nested hierarchy of neural states underlies event segmentation during naturalistic viewing (Geerligs et al., 2022; Lugtmeijer et al., 2025). However, while the existence of temporal hierarchies is well-established, the functional significance of slow processing remains incompletely understood. Previous studies have primarily characterized temporal hierarchies descriptively, but few have directly tested whether slow processing translates into measurable encoding advantages during naturalistic cognition.

A prominent theoretical interpretation proposes that slow temporal dynamics may be computationally necessary for certain cognitive operations (Hasson et al., 2015; Kiebel et al., 2008). Cognitive functions such as narrative comprehension, theory of mind, and future planning represent inherently temporal integration problems that require maintaining and connecting information across extended time periods, which cannot be solved through rapid, reactive processing alone. From this perspective, slow dynamics may not represent inefficiency but rather optimization for computational depth; what appears inefficient from a speed-focused viewpoint may actually represent efficiency from a comprehension-focused viewpoint (Chen et al., 2017). This theoretical framework generates a testable prediction: if slow dynamics reflect optimization for temporal integration, then networks with slower dynamics should benefit more from extended temporal context when encoding complex, naturalistic stimuli.

The Default Mode Network (DMN) presents a particularly illuminating test case. Despite consuming approximately 20% of the brain's metabolic budget at rest (Raichle & Mintun, 2006), the DMN exhibits slow dynamics while supporting sophisticated integrative functions including narrative comprehension, self-referential processing, and social cognition (Andrews-Hanna et al., 2014; Buckner et al., 2008; Menon, 2023; Spreng et al., 2009). Recent work has characterized the DMN as a hub where idiosyncratic self-representations meet shared social understanding (Yeshurun et al., 2021), and has shown that slow-processing regions detect event boundaries in continuous experience (Baldassano et al., 2017). This raises a fundamental question: does slow processing confer functional advantages that justify its metabolic cost?

The present research addresses this question using naturalistic movie-watching fMRI data. We asked: (1) Do functional networks differ in temporal dynamics speed? (2) Do networks with slower dynamics benefit more from extended temporal integration in encoding models? We hypothesized that slower networks would show greater encoding improvement with extended temporal context. Additionally, we explored whether

*Corresponding author: xiaoyanli629@tsinghua.edu.cn

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

functional connectivity stability relates to temporal dynamics.

Methods

Participants and Data

Data were obtained from the Algonauts 2025 Challenge (Gifford et al., 2024), derived from the Courtois NeuroMod project (Boyle et al., 2020). Four healthy adults (ages 31–47; 2 female) viewed approximately 65 hours of naturalistic movie content: *Friends* seasons 1–6 (55 hours) and four additional films (10 hours). This intensive within-subject design (>15 hours per participant) provides high statistical power while cross-subject replication ensures generalizability (Hasson et al., 2010).

fMRI Acquisition and Preprocessing

Functional MRI data were acquired on a Siemens Prisma 3T scanner using a multiband echo-planar imaging sequence (TR = 1.49 s, TE = 30 ms, flip angle = 52°, multiband factor = 4, voxel size = 2 mm isotropic, 60 slices). Preprocessing was performed using fMRIPrep v20.2.0 (Esteban et al., 2019) and included: motion correction via rigid-body registration to the mean functional image; slice-timing correction; co-registration to the T1-weighted anatomical image using boundary-based registration; spatial normalization to MNI152NLin2009cAsym template via nonlinear registration; and confound regression removing 6 motion parameters, their temporal derivatives, framewise displacement, and aCompCor components from white matter and cerebrospinal fluid. High-motion volumes (framewise displacement > 0.5 mm) were censored. Data were parcellated using the Schaefer 2018 atlas comprising 1000 cortical regions organized into seven canonical functional networks: Visual, Somatomotor, Dorsal Attention, Ventral Attention, Limbic, Frontoparietal, and Default Mode (Schaefer et al., 2018).

Stimulus Feature Extraction

Visual features were extracted using SlowFast 3D ResNet (Feichtenhofer et al., 2019), audio features as MFCCs, and language features using BERT embeddings (Devlin et al., 2019). Features were temporally aligned with fMRI accounting for hemodynamic delay (4–6 s) and reduced via PCA to 520 dimensions per time point (250 visual, 250 language, 20 audio).

Temporal Dynamics Speed (TDS)

TDS quantifies the relative dominance of fast versus slow neural activity fluctuations. For each brain region, the fMRI time series was detrended and z-score normalized. Power spectral density $P(f)$ was computed using Welch's method ($f_s = 1/\text{TR} \approx 0.67$ Hz; segment length = min(256, N/4); 50% overlap). TDS was calculated per region as:

$$\text{TDS} = \frac{P_{\text{high}}}{P_{\text{low}}} \tag{1}$$

where $P_{\text{high}} = \int_{0.07}^{0.25} P(f)\,df$ and $P_{\text{low}} = \int_{0.01}^{0.07} P(f)\,df$. Higher TDS indicates relatively greater high-frequency power. TDS was computed for each of the 1000 regions and then averaged within each network. The 0.07 Hz boundary follows conventions in fMRI spectral analysis and was validated via sensitivity analysis at 0.05, 0.07, and 0.09 Hz cutoffs, all yielding consistent network orderings.

TDS differs from prior intrinsic timescale metrics: autocorrelation decay measures how long a signal "remembers" past states (Murray et al., 2014), while temporal receptive windows reflect inter-subject agreement at different timescales (Hasson et al., 2008). TDS captures a complementary property—the spectral power balance—and may therefore yield different network orderings. These metrics are conceptually related but not mathematically equivalent (see Limitations).

Encoding Models

We employed two encoding approaches to test whether networks benefit from temporal integration (RQ2).

Ridge Regression and Temporal Window Gain (TWG). Ridge regression predicted fMRI responses from stimulus features: $\hat{y} = X\beta$ where $\beta = (X^\top X + \lambda I)^{-1} X^\top y$, with λ selected via nested 5-fold cross-validation ($\lambda \in \{0.1, 1, 10, 100, 1000\}$). For TWG analysis, features within each temporal window were averaged into a single vector. Encoding accuracy was evaluated using Pearson r. TWG was computed as:

$$\text{TWG} = r_{\text{long}} - r_{\text{instant}} \tag{2}$$

where r_{long} and r_{instant} are encoding accuracies for long (25 TRs: 20 past + 5 future, ~37 s) and instant (1 TR, ~1.5 s) windows. Positive TWG indicates benefit from extended temporal context.

LSTM Networks. To test whether temporal integration benefits depend on the modeling approach, we compared ridge regression with LSTM networks. Separate models were trained for each participant, network, and sampled region (10 regions per network). LSTM architecture: 2 stacked layers (hidden dimension = 64, dropout = 0.3), followed by a fully connected layer with ReLU activation, taking the final hidden state h_T as input:

$$h_t = \text{LSTM}(x_t, h_{t-1}) \tag{3}$$

$$\hat{y} = \text{FC}(h_T) \tag{4}$$

Models were trained for 30 epochs with Adam optimizer (learning rate = 0.001) on an 80/20 chronological train/test split, preserving temporal order to prevent data leakage. MSE loss was used for training; encoding accuracy was evaluated on the held-out test set using Pearson r. Sequence length was 20 TRs (~30 s). For a stringent comparison, ridge regression ($\alpha = 1000$) used the same 20 TR sequences with features concatenated across time (yielding $20 \times 520 = 10{,}400$ input dimensions), providing Ridge with full access to temporal information without sequential modeling. Ridge used 3-fold cross-validation. LSTM improvement was computed as $(\text{LSTM}_r - \text{Ridge}_r)/\text{Ridge}_r \times 100\%$.

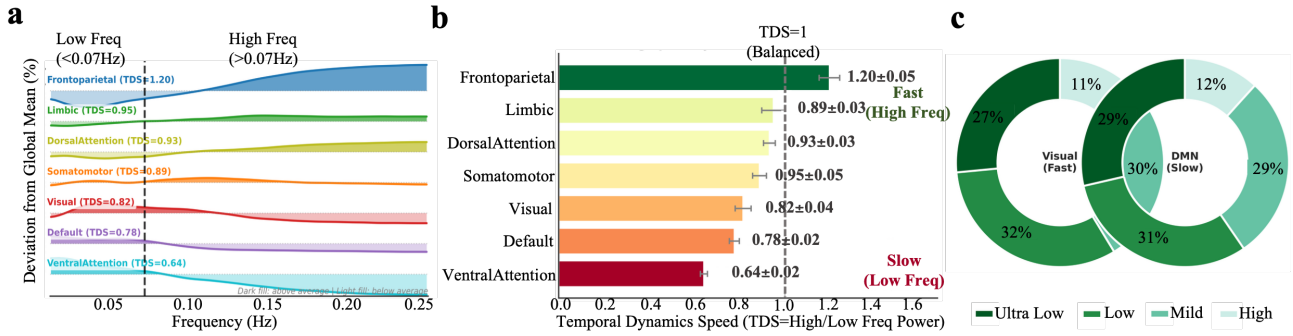

Figure 1: **Temporal Dynamics Hierarchy Across Cortical Networks.** (a) Network-specific spectral signatures showing deviation from global mean PSD. Dark shading: above-average power; light shading: below-average. Networks ordered by TDS (fastest to slowest); vertical dashed line marks 0.07 Hz boundary. (b) TDS values with error bars (SEM, $n = 4$). (c) Frequency band composition for Visual and Default Mode networks.

Dynamic Connectivity Stability (DCS)

As an exploratory analysis, DCS quantifies temporal consistency of within-network functional connections using sliding-window correlation (window = 30 TRs ≈ 45 s, step = 10 TRs). For computational efficiency, 20 regions were randomly sampled per network; robustness was confirmed across window sizes (20, 30, 40, 50 TRs). For each window t, within-network FC_t was the mean pairwise Pearson r (upper triangle, excluding diagonal). DCS was defined as:

$$\text{DCS} = 1 - \text{CV}(FC) = 1 - \frac{\sigma_{FC}}{\mu_{FC}} \tag{5}$$

where CV is the coefficient of variation, σ_{FC} is the standard deviation, and μ_{FC} is the mean of connectivity across windows. Higher DCS values indicate more temporally stable connections.

Statistical Analysis

Network-level metrics were computed by averaging across regions within each network. Bootstrap resampling (1000 iterations for TDS, 500 for DCS) provided 95% confidence intervals. Effect sizes were quantified using Cohen's d for between-network comparisons. Sensitivity analyses examined robustness to parameter choices: frequency band cutoffs (0.05, 0.07, 0.09 Hz) for TDS and window sizes (20, 30, 40, 50 TRs) for DCS. All analyses were implemented in Python 3.9 using NumPy, SciPy, scikit-learn, PyTorch, and nilearn. Analysis code is available at https://github.com/xiaoyanLi629/temporal-efficiency-paradox.

Results

Temporal Dynamics Hierarchy Across Networks (RQ1)

Brain networks exhibited systematically different temporal dynamics during naturalistic viewing (Figure 1), consistent with cortical temporal hierarchy (Hasson et al., 2008; Murray et al., 2014). A one-way ANOVA confirmed significant between-network differences in TDS ($F(6, 993) = 18.42$, $p < .001$, $\eta^2 = 0.10$).

TDS values reveal a 1.9-fold range across networks (0.64 to 1.20; Figure 1b). Frontoparietal Network exhibited the fastest dynamics (TDS = 1.20, SEM = 0.05), significantly faster than Ventral Attention Network (TDS = 0.64, SEM = 0.02; permutation $p = .028$, Cohen's $d = 6.09$). The Default Mode Network (TDS = 0.78, SEM = 0.02) falls among the slower networks. Networks deviate from the global average in characteristic ways (Figure 1a): Frontoparietal shows enhanced high-frequency power, while Ventral Attention shows enhanced low-frequency power. These findings establish that the brain maintains a hierarchy of temporal processing speeds.

Multi-Timescale Encoding Analysis (RQ2)

We next tested whether networks with slower dynamics benefit from extended temporal integration. Contrary to expectations, all seven networks showed negative Temporal Window Gain (Figure 2a–d), indicating that encoding accuracy decreased with longer temporal windows (one-sample t-test: $t(6) = -8.73$, $p < .001$). TWG ranged from −0.029 (Somatomotor) to −0.015 (Frontoparietal), with all networks showing non-overlapping confidence intervals from zero (Figure 2b). The correlation between TDS and TWG was non-significant ($r = 0.52$, $p = .23$; Figure 2d).

LSTM vs Ridge: Preserving Temporal Structure Benefits All Networks

These negative TWG findings raised a question: does temporal integration fail because extended context is inherently unhelpful, or because temporal averaging specifically destroys useful structure? To disentangle these possibilities, we compared LSTM networks (which process the sequence step-by-step, learning temporal dependencies) against a stringent ridge regression baseline that receives the same 20 TR features *con-*

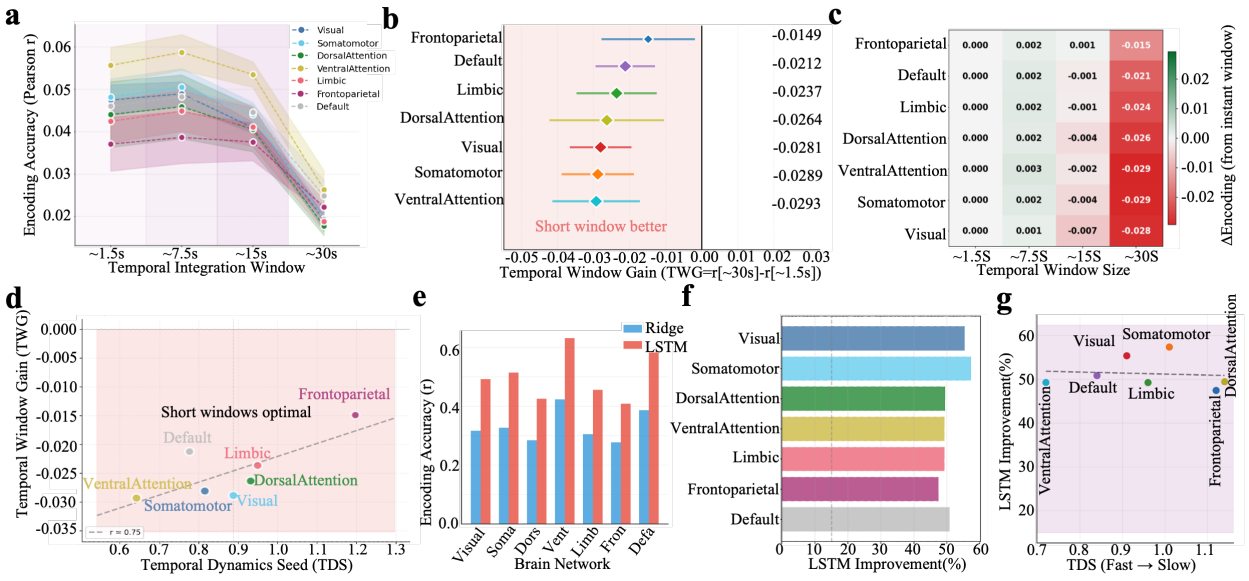

Figure 2: **Encoding Model Analysis: Temporal Integration and LSTM Performance. Top row (a–d): Temporal Window Gain analysis.** (a) Encoding accuracy (Pearson r) across temporal integration windows (~1.5s to ~37s) for each network; shaded regions indicate SEM. (b) Forest plot of TWG values with 95% confidence intervals; all networks show negative TWG, indicating impaired encoding with longer windows. (c) Heatmap of encoding accuracy change from instant window across networks and temporal windows. (d) Relationship between TDS and TWG; dashed line shows non-significant correlation ($r = 0.52$, $p = .23$). **Bottom row (e–g): LSTM vs Ridge comparison.** (e) Encoding accuracy comparison showing LSTM (red) consistently outperforms Ridge with concatenated features (blue) across all networks. (f) LSTM improvement percentage by network; Somatomotor shows largest gain (+57.4%). (g) Relationship between TDS and LSTM improvement; dashed line shows non-significant correlation ($r = 0.18$, $p = .70$).

catenated rather than averaged—thus preserving all temporal information while lacking the ability to model sequential dependencies.

LSTM models outperformed ridge regression across all networks (paired t-test: $t(6) = 12.4$, $p < .001$), with LSTM (red) consistently exceeding Ridge (blue) encoding accuracy for every network (Figure 2e). Somatomotor Network showed the largest improvement (+57.4%), followed by Visual (+55.4%) and Default Mode (+50.8%; Figure 2f). Even Frontoparietal Network, with the fastest temporal dynamics, benefited from LSTM (+47.5%). Slow networks showed comparable improvements: Ventral Attention improved by 49.3% and Default Mode by 50.8%, indicating that these networks benefit from temporal integration when temporal structure is preserved.

The correlation between TDS and LSTM improvement was weak ($r = 0.18$, $p = .70$; Figure 2g), indicating that all networks benefit from sequential modeling regardless of their intrinsic dynamics. Notably, the LSTM advantage over ridge regression—which had access to the same temporal information via feature concatenation—demonstrates that the benefit arises specifically from learning sequential dependencies, not merely from having more input features. The negative TWG reflects the destructive effect of temporal averaging, while the uniform LSTM advantage reveals that all networks contain exploitable temporal structure.

Dynamic Connectivity Stability

Beyond temporal dynamics speed, we examined functional connectivity stability (Figure 3a shows representative single-subject data). Connectivity fluctuations differed across networks (Kruskal-Wallis $H = 45.3$, $p < .001$). Visual Network showed the highest stability (DCS = 0.57), while Frontoparietal showed the lowest (DCS = 0.30; Figure 3b; see Table 1). The correlation between TDS and DCS was non-significant ($r = -0.32$, $p = .48$; Figure 3c), suggesting connectivity stability reflects a distinct dimension of network organization.

Integrated Network Profiles

Table 1 summarizes temporal metrics and encoding performance. Radar charts (Figure 3d) reveal distinct network profiles: fast networks show high TDS but variable connectivity stability; slow networks show low TDS with higher stability. The anatomical distribution (Figure 4) shows fast processing regions concentrated in frontoparietal cortices, while slow regions distribute across ventral attention and default mode areas. Network TDS rankings (Figure 5) range from Frontoparietal (TDS $\approx$ 1.2) to Ventral Attention (TDS $\approx$ 0.64).

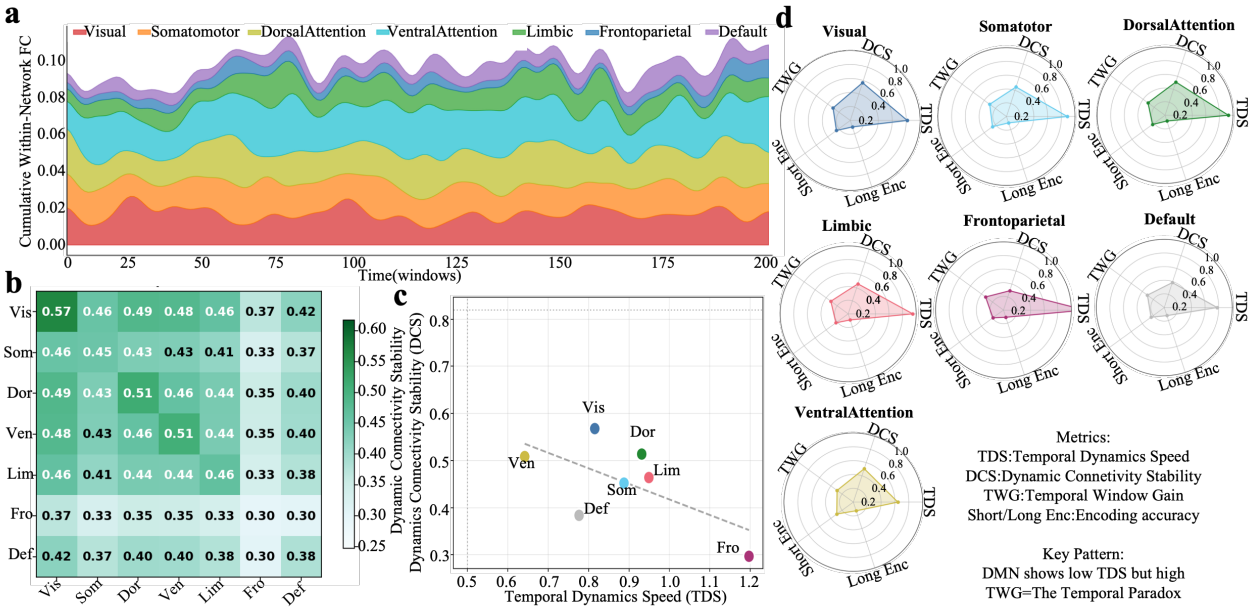

Figure 3: **Dynamic Connectivity Stability and Network Profiles.** (a) Stream graph of cumulative within-network functional connectivity (mean pairwise Pearson r within each network) over time; data from a representative participant (sub-01) viewing *Friends*, showing first 200 windows (~50 min). (b) Inter-network DCS stability matrix. (c) TDS vs DCS scatter plot; dashed line shows non-significant correlation ($r = -0.32$, $p = .48$). (d) Radar charts of network profiles integrating TDS, DCS, TWG, and encoding accuracy.

Table 1: **Summary of Network Temporal Properties and Encoding Performance**

Network	TDS	DCS	TWG	Ridge r	LSTM r	LSTM Gain
Frontoparietal	1.20	0.30	−0.015	0.277	0.408	+47.5%
Limbic	0.95	0.46	−0.024	0.305	0.456	+49.3%
Dorsal Attention	0.93	0.51	−0.026	0.285	0.425	+49.5%
Somatomotor	0.89	0.45	−0.029	0.327	0.514	+57.4%
Visual	0.82	0.57	−0.028	0.317	0.492	+55.4%
Default	0.78	0.38	−0.021	0.386	0.582	+50.8%
Ventral Attention	0.64	0.51	−0.029	0.423	0.632	+49.3%

Note. TDS = Temporal Dynamics Speed; DCS = Dynamic Connectivity Stability; TWG = Temporal Window Gain (ridge with averaged features, long [25 TRs] minus instant [1 TR]); Ridge/LSTM r = encoding accuracy (20 TR window; Ridge uses concatenated features); LSTM Gain = improvement over Ridge. Separate models trained per subject per network. Networks sorted by TDS.

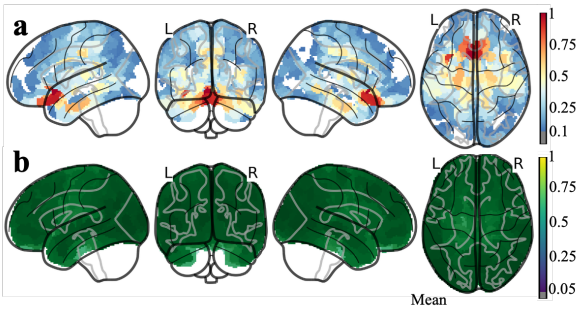

Figure 4: **Brain Glass Visualization of Temporal Dynamics.** (a) Temporal Dynamics Speed (TDS) across cortex: fast (red/warm) vs slow (blue/cool). (b) Network Temporal Hierarchy, color-coded by network position in temporal gradient.

Discussion

Summary of Findings

This study yielded three main findings. First, functional networks differ systematically in temporal dynamics speed (RQ1). Second, our hypothesis that slower networks benefit more from temporal integration was not supported: temporal averaging impaired encoding across all networks, and LSTM models improved encoding by 47–57% over ridge regression with concatenated features, demonstrating that modeling sequential dependencies—not merely accessing temporal information—is the key factor. Third, connectivity stability is uncorrelated with TDS, representing a distinct dimension of network organization.

Our TDS-based hierarchy diverges from canonical unimodal-to-transmodal gradients (Hasson et al., 2008; Murray et al., 2014): Frontoparietal exhibited the fastest dynamics, consistent with rapid cognitive control (Badre & Nee, 2018); Ventral Attention the slowest; and DMN intermediate, sup-

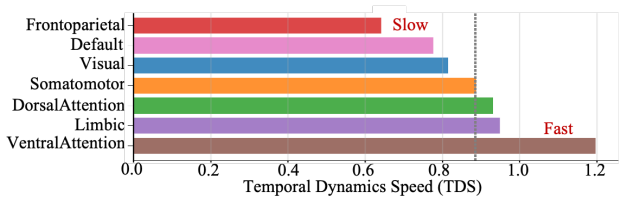

Figure 5: **Network TDS Ranking.** Temporal Dynamics Speed (TDS) for each functional network, ordered from slowest (Ventral Attention, TDS = 0.64) to fastest (Frontoparietal, TDS = 1.20). Dashed vertical line indicates TDS = 1.0, separating networks with predominantly slow dynamics (TDS < 1) from those with fast dynamics (TDS > 1). Only Frontoparietal Network exceeds this threshold, indicating predominantly high-frequency activity.

porting its reconceptualization as dynamically flexible (Gao et al., 2020). DCS was uncorrelated with TDS (e.g., Visual shows moderate TDS but high stability; Frontoparietal shows fast dynamics but low stability), suggesting networks are characterized by independent temporal properties rather than a single speed gradient.

The Temporal Integration Paradox Resolved

Two results jointly resolve the apparent paradox. First, temporal averaging impaired encoding (negative TWG), showing that collapsing temporal information is destructive. Second, LSTM outperformed ridge regression even when Ridge received the same temporal information via feature concatenation—demonstrating that the critical factor is not merely *having* temporal information, but *modeling sequential dependencies* within it. This two-level distinction—averaging destroys information; concatenation preserves it but cannot exploit temporal order—reveals that recurrent, order-sensitive computation is what drives encoding gains. This view aligns with hierarchical predictive coding (Friston, 2010; Kiebel et al., 2008), where each cortical level generates predictions based on temporal sequences, and with evidence that slow cortical dynamics support information accumulation through recurrent processing (Hasson et al., 2015; Honey et al., 2012).

The uniform LSTM benefit across all networks, regardless of intrinsic dynamics speed, suggests that naturalistic stimuli contain temporal structure at multiple scales that all networks must process—fast networks exploiting short-range dependencies (e.g., visual motion continuity) and slow networks exploiting long-range dependencies (e.g., narrative coherence). This is consistent with a parallel multi-timescale architecture rather than a single speed gradient, and aligns with evidence that recurrent architectures substantially outperform static encoding models (Oota et al., 2023). From a practical standpoint, encoding models should preserve temporal dependencies rather than aggregate features across time.

Limitations and Future Directions

Several limitations should be noted. First, while core findings replicate across all four participants, the small sample (N=4) limits generalizability of precise TDS rankings and effect sizes; quantitative estimates require replication in larger cohorts. Second, TDS captures spectral power balance, which is methodologically distinct from autocorrelation decay (Murray et al., 2014) and temporal receptive windows (Hasson et al., 2008); divergent network orderings likely reflect this difference. TDS was also computed from task-state data, so comparison with resting-state would clarify whether differences reflect intrinsic or task-specific properties. Third, negative TWG may partly reflect HRF nonlinearities, though the LSTM advantage over concatenation-based Ridge argues against information loss as the sole explanation. Fourth, BERT's bidirectionality means language features already incorporate future context, potentially reducing the LSTM advantage for language-sensitive regions; future work should examine causal language models. Fifth, Somatomotor and Visual networks showed numerically larger LSTM gains (+55–57% vs. +47–49%), possibly reflecting higher signal-to-noise ratios. Finally, network-level analysis may obscure within-network heterogeneity.

Future extensions include temporal scrambling controls to isolate sequential order contributions, comparison with other sequence architectures (GRU, Transformer), validation of TDS against autocorrelation-based metrics, and testing with diverse naturalistic stimuli beyond narrative film.

Conclusions

Functional networks exhibit a clear temporal hierarchy, but slower networks do not benefit more from temporal integration. Instead, LSTM models improved encoding by 47–57% over ridge regression with concatenated features across all networks uniformly, resolving an apparent paradox: temporal integration benefits encoding universally, but only when sequential dependencies are explicitly modeled. Temporal *structure*, not merely temporal *extent*, accounts for substantial neural response variance during naturalistic viewing.

References

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the new York Academy of Sciences*, *1316*(1), 29–52.

Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in cognitive sciences*, *22*(2), 170–188.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709–721.

Boyle, J. A., Pinsard, B., Boukhdhir, A., Belleville, S., Brambatti, S., Chen, J., Cohen-Adad, J., Cyr, A., Fuente Rainville, P., & Bellec, P. (2020). The courtois project on neuronal modelling-first data release. *26th annual meeting of the organization for human brain mapping*.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the new York Academy of Sciences*, *1124*(1), 1–38.

Chaudhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H., & Wang, X.-J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron*, *88*(2), 419–431.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience*, *20*(1), 115–125.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 4171–4186.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). Fmriprep: A robust preprocessing pipeline for functional mri. *Nature methods*, *16*(1), 111–116.

Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). Slowfast networks for video recognition. *Proceedings of the IEEE/CVF international conference on computer vision*, 6202–6211.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature reviews neuroscience*, *11*(2), 127–138.

Gao, R., Van den Brink, R. L., Pfeffer, T., & Voytek, B. (2020). Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *elife*, *9*, e61277.

Geerligs, L., Gözükara, D., Oetringer, D., Campbell, K. L., van Gerven, M., & Güçlü, U. (2022). A partially nested cortical hierarchy of neural states underlies event segmentation in the human brain. *elife*, *11*, e77430.

Gifford, A. T., Bersch, D., St-Laurent, M., Pinsard, B., Boyle, J., Bellec, L., Oliva, A., Roig, G., & Cichy, R. M. (2024). The algonauts project 2025 challenge: How the human brain makes sense of multimodal movies. *arXiv preprint arXiv:2501.00504*.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in cognitive sciences*, *19*(6), 304–313.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, *14*(1), 40–48.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of neuroscience*, *28*(10), 2539–2550.

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., Doyle, W. K., Rubin, N., Heeger, D. J., & Hasson, U. (2012). Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*, *76*(2), 423–434.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS computational biology*, *4*(11), e1000209.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of neuroscience*, *31*(8), 2906–2915.

Lugtmeijer, S., Oetringer, D., Geerligs, L., & Campbell, K. L. (2025). Temporal dedifferentiation of neural states with age during naturalistic viewing. *Communications Biology*, *8*(1), 1390.

Menon, V. (2023). 20 years of the default mode network: A review and synthesis. *Neuron*, *111*(16), 2469–2487.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience*, *17*(12), 1661–1663.

Oota, S. R., Chen, Z., Gupta, M., Bapi, R. S., Jobard, G., Alexandre, F., & Hinaut, X. (2023). Deep neural networks and brain alignment: Brain encoding and decoding (survey). *arXiv preprint arXiv:2307.10246*.

Raichle, M. E., & Mintun, M. A. (2006). Brain work and brain imaging. *Annu. Rev. Neurosci.*, *29*(1), 449–476.

Raut, R. V., Snyder, A. Z., Mitra, A., Yellin, D., Fujii, N., Malach, R., & Raichle, M. E. (2021). Global waves synchronize the brain's functional systems with fluctuating arousal. *Science advances*, *7*(30), eabf2709.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, *28*(9), 3095–3114.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of cognitive neuroscience*, *21*(3), 489–510.

Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: Where the idiosyncratic self meets the shared social world. *Nature reviews neuroscience*, *22*(3), 181–192.