Separating Stimulus-Induced and Background Components of Dynamic Functional Connectivity in Naturalistic fMRI

Chee-Ming Ting, *Senior Member, IEEE*, Jeremy I. Skipper, Fuad Noman, Steven L. Small and Hernando Ombao

Abstract—We consider the challenges in extracting stimulus-related neural dynamics from other intrinsic processes and noise in naturalistic functional magnetic resonance imaging (fMRI). Most studies rely on inter-subject correlations (ISC) of low-level regional activity and neglect varying responses in individuals. We propose a novel, data-driven approach based on low-rank plus sparse (L+S) decomposition to isolate stimulus-driven dynamic changes in brain functional connectivity (FC) from the background noise, by exploiting shared network structure among subjects receiving the same naturalistic stimuli. The time-resolved multi-subject FC matrices are modeled as a sum of a low-rank component of correlated FC patterns across subjects, and a sparse component of subject-specific, idiosyncratic background activities. To recover the shared low-rank subspace, we introduce a fused version of principal component pursuit (PCP) by adding a fusion-type penalty on the differences between the columns of the low-rank matrix. The method improves the detection of stimulus-induced group-level homogeneity in the FC profile while capturing inter-subject variability. We develop an efficient algorithm via a linearized alternating direction method of multipliers to solve the fused-PCP. Simulations show accurate recovery by the fused-PCP even when a large fraction of FC edges are severely corrupted. When applied to natural fMRI data, our method reveals FC changes that were time-locked to auditory processing during movie watching, with dynamic engagement of sensorimotor systems for speech-in-noise. It also provides a better mapping to auditory content in the movie than ISC.

Index Terms—Low-rank plus sparse decomposition, dynamic functional connectivity, inter-subject correlation, fMRI.

I. INTRODUCTION

NATURALISTIC functional magnetic resonance imaging (fMRI) is an emerging approach in cognitive neuro-science, employing naturalistic stimuli (e.g., movies, spoken narratives, music, etc.) to provide an ecologically-valid paradigm that mimics real-life scenarios [1]. Complementary to traditional task-based paradigms with strictly controlled artificial stimuli, naturalistic paradigms offer a better under-standing of neural processing and network interactions in realistic contexts, where stimuli typically engage the brain in continuous integration of dynamic streams of multimodal (e.g., audiovisual) information. Resting-state fMRI, although entirely unconstrained, is vulnerable to confounds and difficult to link with ongoing cognitive states. Natural stimuli like movies have shown higher test-retest reliability than resting and task fMRI [2]. In this paper, we focus on a major challenge in naturalistic fMRI data analysis – to extract stimulus-related neural dynamics from other intrinsic and noise contributions.

Early studies use traditional methods for task fMRI such as the general linear model (GLM) to detect brain activation driven by natural stimuli, using annotations of specific features or events (e.g., faces, scenes or speech in a film [3]) as regressors. However, GLM-based analysis requires explicit *a priori* models of relevant features of the stimuli and associated neural responses as predictors, where minor model misspec-ification can result in low detection power. It is therefore effective only for simple parametric designs, but extremely rigid for capturing the richer dynamical information present in natural stimuli. Inter-subject correlation (ISC) analyses provide a powerful, data-driven alternative for handling nat-uralistic paradigms without a pre-defined response model, by leveraging the reliable, shared neural responses across different subjects when exposed to the same continuous naturalistic stimulation [4], [5]. By correlating fMRI time courses from the same regions across subjects, ISC can detect inter-subject synchronization of brain activity in specific areas activated by the stimuli, e.g., prefrontal cortex during movie viewing [6]. It was recently extended to inter-subject functional correlation (ISFC) to characterize stimulus-related functional networks by computing the correlations between all pairs of regions across subjects [7]. Use of ISFC has revealed, for example, engagement of default mode networks during narrative com-prehension [7]. The ISFC increases specificity to stimulus-locked inter-regional correlations, by filtering out intrinsic, task-unrelated neural dynamics as well as non-neuronal arti-facts (e.g., respiratory rate; head motion) that are theoretically uncorrelated across subjects. Time-resolved ISFC computed using a sliding-window technique has been used to track movie-induced dynamic changes in functional connectivity [8]. A recent study [9] explored ISC of time-courses of dynamic connectivity rather than regional activity. For a review of the ISC family of approaches, see [10]. However, one limitation of IS(F)C is that it has no inherent way to detect individual differences in neural activity, due to the averaging-out of all uncorrelated variations across subjects. Moreover, the ISFC analysis inherently relies on the Pearson correlations between

C.-M. Ting is with the School of Information Technology, Monash Uni-versity Malaysia, Subang Jaya 47500, Malaysia, and also the Biostatistics Group, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia, and also the Division of Psychology and Language Sciences, University College London, UK (e-mail: ting.cheeming@monash.edu).

J.I. Skipper is with the Department of Experimental Psychology, University College London, London WC1H, UK (e-mail: jeremy.skipper@ucl.ac.uk).

F. Noman is with the School of Information Technology, Monash University Malaysia, Subang Jaya 47500, Malaysia (e-mail: fuad.noman@monash.edu).

S.L. Small is with the School of Behavioral and Brain Sciences, University of Texas at Dallas, TX 75080, USA (e-mail: small@utdallas.edu).

H. Ombao, is with the Biostatistics Group, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia (e-mail: her-nando.ombao@kaust.edu.sa).

activity time courses of subjects, and thus how it can be generalized to other measures of brain connectivity, e.g., for directed functional networks is still unclear.

We propose a new approach based on low-rank plus sparse (L+S) decomposition to separate stimulus-induced and background components of dynamic functional connectivity (FC) in naturalistic fMRI. Our method is inspired by the successful applications of L+S decomposition in video surveillance to separate a slowly-changing background (modeled by a low-rank subspace) from moving foreground objects (sparse outliers) from a video sequence [11], [12]. We formulate a time-varying L+S model for dynamic FC (dFC), where the time-resolved, multi-subject FC matrices are represented as the sum of a low-rank component and a sparse component. The low-rank component corresponds to the similar connectivity patterns across subjects elicited by the same stimuli. The sparse component captures the corruptions on stimulus-related FC edges, arising from background activity (i.e., intrinsic processes and artifacts) that are idiosyncratic for individual subjects. We apply the robust principal component analysis [13], [14] at each time point to recover a sequence of time-resolved low-rank FC matrices from the sparse background noise, by solving the nuclear norm and ℓ_1 norm minimization (called principal component pursuit (PCP)).

We further introduce a novel L+S decomposition method called the fused PCP, by adding a new fusion-type penalty in the PCP to shrink the successive differences between the columns of the low-rank matrix towards zero. The fused L+S decomposition has two main advantages: First, it captures piecewise constancy or similarity in FC over groupings of subjects with shared response. Secondly, it provides further smoothing of background noise which enables better estimation of the low-rank structure. We develop an efficient linearized alternating direction method of multipliers (ADMM) algorithm to solve the proposed fused PCP. We evaluated its performance for L+S recovery via simulations. Application to movie-watching fMRI data reveals dFC patterns reliably related to processing of continuous auditory events.

The contributions of this work are as follows: (1) This is the first work to propose an L+S decomposition method for naturalistic fMRI, which is model-free and data-driven in detecting stimulus-induced dynamic FC. It differs from recent application of L+S method to resting-state fMRI [15], [16] which are limited to static FC analysis assuming a time-invariant L+S structure. Our proposed time-varying L+S model is a nonstationary generalization of L+S decomposition for recovering a time sequence of low-rank and sparse components to characterize changes in the common and idiosyncratic FC structure over time. Besides, [15] imposed L+S constraints on FC of individual subjects, but neglected correlated FC patterns across subjects. (2) The proposed L+S learning algorithm offers a novel way to isolate low-rank parts of correlated multi-subject FC from idiosyncratic components over time. It exploits shared neuronal activity across subjects in terms of the network connectivity directly rather than the low-level BOLD responses in ISFC. This is equivalent to denoising the FC networks instead of raw fMRI signals via ISC, and thus extracts better FC topological information. The separated

dFC components can be more reliably related to or predictive of the underlying group-level neural processing in response to the same time-varying natural stimuli, without being confounded by individual noise effects. By utilizing cross-subject reliability of FC measures captured in the low-rank part, we show that it provides a better mapping to naturalistic stimuli than the ISFC. Furthermore, our method advances a more general analysis framework applicable to other FC measures beyond the Pearson correlation in ISFC. (3) Unlike ISFC which computes a group-level FC profile related only to the stimuli, the proposed method recovers both the stimulus-related and unrelated FC patterns. This allows us to quantify confounding influence on the stimulus-induced FC, including potentially interesting stimulus-independent sources of neural dynamics, such as intrinsic processes, attentional variations and other individual differences. (4) While extracting a shared connectivity structure, our method also retains subject-specific FC patterns unavailable from ISFC, which is useful for studying individual differences in dynamic FC during naturalistic stimulation. (5) Despite the advantages of L+S modeling of multi-subject FC, use of conventional L+S decomposition as in [16] may not fully capture the considerable similarity in FC patterns across subjects when experiencing the same naturalistic stimuli. Moreover, it requires an assumption of sufficiently sparse background noise in order to recover the shared low-rank FC structure, which may be unrealistic for brain network data. Our proposed fused extension of L+S decomposition can overcome these problems by regularizing the inter-subject differences in the low-rank structure. It can encourage group-level homogeneity in FC patterns across subjects, by smoothing out the remaining stimulus-unrelated individual differences in the common connectivity structure. Furthermore, it gives a better recovery of the low-rank FC matrices in presence of dense noise (i.e., many of FC edges are severely corrupted) than the conventional L+S decomposition.

II. LOW-RANK AND SPARSE MATRIX DECOMPOSITION FOR DYNAMIC FUNCTIONAL CONNECTIVITY

A. Proposed Model

Suppose we observe a set of weighted connectivity matrices $\{\boldsymbol{\Sigma}_{ti}, t = 1, \ldots, T, i = 1, \ldots, M\}$ for multi-subject, time-dependent functional brain networks with same set of N nodes, where elements in $\boldsymbol{\Sigma}_{ti} \in \mathbb{R}^{N \times N}$ are measures of pairwise interactions between nodes in the network at time point t for ith subject. T is the number of time points and M is the number of subjects. Given the observed $\{\boldsymbol{\Sigma}_{ti}\}$, we aim to separate the task or stimulus-related brain FC dynamics from the background composed of intrinsic neural correlations, noise and other non-neuronal contributions. Our approach builds on the hypothesis that FC networks across different subjects admit a common underlying structure and exhibit similar or highly-correlated patterns, due to shared neuronal response among subjects when experiencing the same stimuli, e.g., watching the same movie in a naturalistic setting. On the other hand, the background connectivity networks are assumed to be relatively sparse compared to the stimulus-induced ones, and some confounding fluctuations such as motion-related

artifacts occur only occasionally in certain subjects. These background activities are spontaneous, not time-locked to stimuli, and thus varying considerably across subjects.

In particular, we propose a decomposition of the time-dependent connectivity matrices $\boldsymbol{\Sigma}_{ti}$ as a superposition of a low-rank and a sparse component that represent respectively the correlated stimulus-induced connectivity and the subject-varying background activities. To formulate, let $\mathbf{z}_{ti} = \mathrm{vec}(\boldsymbol{\Sigma}_{ti})$ be the vectorized version of $\boldsymbol{\Sigma}_{ti}$, and $\mathbf{Z}_t = [\mathbf{z}_{t1}, \dots, \mathbf{z}_{tM}] \in \mathbb{R}^{N^2 \times M}$ to denote the concatenation of the vectors of connectivity metrics over all subjects at time t. We consider a low-rank plus sparse (L+S) decomposition of the time-dependent matrices $\mathbf{Z}_t$

$$\mathbf{Z}_t = \mathbf{L}_t + \mathbf{S}_t, \quad \forall t = 1, \dots, T \tag{1}$$

where $\mathbf{L}_t = [\mathbf{l}_{t1}, \dots, \mathbf{l}_{tM}]$ is the low-rank matrix with $\mathrm{rank}(\mathbf{L}_t) = r_t < \min(N^2, M)$, $\mathbf{S}_t = [\mathbf{s}_{t1}, \dots, \mathbf{s}_{tM}]$ is a sparse matrix with a fraction s of non-zero entries, and $\mathbf{l}_{ti} \in \mathbb{R}^{N^2}$, $\mathbf{s}_{ti} \in \mathbb{R}^{N^2}$ are respectively the column vectors of stimulus-related and background connectivity metrics for the i-th subject at time t. The stimulus-induced connectivity that are correlated between subjects at each time t can be well-approximated by the columns of $\mathbf{L}_t$ that lie on a low-dimensional subspace spanned by common bases. For example, rank one $r_t = 1$ yields a similar connectivity pattern over all subjects. The columns of $\mathbf{S}_t$ capture the subject-specific background activities. Unlike the small Gaussian noise assumption, entries of $\mathbf{S}_t$ may have arbitrarily large magnitude with unknown sparse support, and thus capturing random, gross corruptions from the background events on the stimulus-related information in functional connectivity.

B. L+S Decomposition

1) Principal Component Pursuit: To recover the sequence of low-rank matrices $\{\mathbf{L}_t\}$ of stimulus-related connectivity from the corrupted observations $\{\mathbf{Z}_t\}$ in (1), we consider solving the following convex optimization problem

$$\min_{\{\mathbf{L}_t\}, \{\mathbf{S}_t\}} \sum_{t=1}^{T} \|\mathbf{L}_t\|_* + \lambda \sum_{t=1}^{T} \|\mathbf{S}_t\|_1$$
$$s.t. \quad \mathbf{L}_t + \mathbf{S}_t = \mathbf{Z}_t, \quad \forall t = 1, \dots, T \tag{2}$$

where $\|\mathbf{L}_t\|_* = \sum_{i=1}^{r_t} \sigma_i(\mathbf{L}_t)$ is the nuclear norm of matrix $\mathbf{L}_t$, i.e. the sum of all its singular values, $\|\mathbf{S}_t\|_1 = \sum_{ij} |s_{tij}|$ is the ℓ_1 norm of $\mathbf{S}_t$ and $\lambda > 0$ is a regularization parameter controlling the trade-off between the low-rank and sparse components. The nuclear norm induces sparsity of the vector of singular values, or equivalently for the matrix $\mathbf{L}_t$ to be low rank, while the ℓ_1 penalty imposes element-wise sparsity of $\mathbf{S}_t$. Minimization of (2) corresponds to applying the well-known PCP approach [14] at each time point. Under the incoherence conditions on the row and column subspaces of $\mathbf{L}_t$, the PCP solution can perfectly recover the low-rank and sparse components, provided that the rank of $\mathbf{L}_t$ is not too large and $\mathbf{S}_t$ is reasonably sparse [14], [17]. Many efficient and scalable algorithms such as the augmented Lagrange multipliers (ALM) method [18] have been proposed to solve this convex optimization.

2) Fused L+S Recovery: The slowly-varying or possibly constant stimulus-related connectivity patterns across subjects may not be fully explained by a low-rank structure estimated in (2). To better capture this inter-subject similarity in connectivity structure, we introduce a novel L+S decomposition by solving a fused version of PCP (fused PCP)

$$\min_{\{\mathbf{L}_t\}, \{\mathbf{S}_t\}} \sum_{t=1}^{T} \|\mathbf{L}_t\|_* + \lambda_1 \sum_{t=1}^{T} \|\mathbf{S}_t\|_1$$
$$+ \lambda_2 \sum_{t=1}^{T} \sum_{i=2}^{M} \|\mathbf{l}_{ti} - \mathbf{l}_{t(i-1)}\|_1$$
$$s.t. \quad \mathbf{L}_t + \mathbf{S}_t = \mathbf{Z}_t, \quad \forall t = 1, \dots, T \tag{3}$$

Here, the additional fused term $\sum_{i=2}^{M} \|\mathbf{l}_{ti} - \mathbf{l}_{t(i-1)}\|_1$, regularized by tuning parameter $\lambda_2 > 0$, encourages the stimulus-induced group homogeneity by penalizing the differences in the connectivity profiles $\mathbf{l}_{ti}$ between subjects. For some sufficiently large λ_2, the penalty shrinks inter-subject differences $|l_{tij} - l_{t(i-1)j}|$ in some connectivity coefficients j to 0, resulting in similar (but not necessarily identical) connectivity patterns across subjects. Clearly, the model (3) includes the original L+S decomposition (2) as a special case when $\lambda_2 = 0$.

Define a matrix $\mathbf{A} = \mathbf{D} \otimes \mathbf{I}_{N^2}$ where

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(M-1) \times M}$$

is first-order difference matrix and $\mathbf{I}_{N^2}$ is a $N^2 \times N^2$ identity matrix. We can reformulate the objective function in (3) as

$$\sum_{t=1}^{T} \|\mathbf{L}_t\|_* + \lambda_1 \sum_{t=1}^{T} \|\mathbf{S}_t\|_1 + \lambda_2 \sum_{t=1}^{T} \|\mathbf{A}\mathrm{vec}(\mathbf{L}_t)\|_1. \tag{4}$$

III. OPTIMIZATION

For the fused L+S decomposition, we develop an ADMM algorithm to solve the convex optimization problem (3) which has no closed form solution. We introduce a set of surrogate variables $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_T$ where $\boldsymbol{\alpha}_t = \mathbf{A}\mathrm{vec}(\mathbf{L}_t) \in \mathbb{R}^{N^2(M-1)}$. Then, the minimization of (4) is equivalent to

$$\min_{\{\mathbf{L}_t\}, \{\mathbf{S}_t\}, \{\boldsymbol{\alpha}_t\}} \sum_{t=1}^{T} \|\mathbf{L}_t\|_* + \lambda_1 \sum_{t=1}^{T} \|\mathbf{S}_t\|_1 + \lambda_2 \sum_{t=1}^{T} \|\boldsymbol{\alpha}_t\|_1$$
$$s.t. \quad \mathbf{L}_t + \mathbf{S}_t = \mathbf{Z}_t, \quad \mathbf{A}\mathrm{vec}(\mathbf{L}_t) = \boldsymbol{\alpha}_t, \quad \forall t = 1, \dots, T \tag{5}$$

The augmented Lagrangian function of (5) is

$$L_\mu(\{\mathbf{L}_t\}, \{\mathbf{S}_t\}, \{\boldsymbol{\alpha}_t\}, \{\widetilde{\mathbf{X}}_t\}, \{\widetilde{\mathbf{Y}}_t\})$$
$$= \sum_{t=1}^{T} \|\mathbf{L}_t\|_* + \lambda_1 \sum_{t=1}^{T} \|\mathbf{S}_t\|_1 + \lambda_2 \sum_{t=1}^{T} \|\boldsymbol{\alpha}_t\|_1$$
$$+ \sum_{t=1}^{T} \left(\mathrm{tr}\,[\widetilde{\mathbf{X}}_t^T (\mathbf{L}_t + \mathbf{S}_t - \mathbf{Z}_t)] + \widetilde{\mathbf{Y}}_t^T (\mathbf{A}\mathrm{vec}(\mathbf{L}_t) - \boldsymbol{\alpha}_t) \right)$$
$$+ \frac{\mu}{2} \sum_{t=1}^{T} \left(\|\mathbf{L}_t + \mathbf{S}_t - \mathbf{Z}_t\|_F^2 + \|\mathbf{A}\mathrm{vec}(\mathbf{L}_t) - \boldsymbol{\alpha}_t\|_2^2 \right) \tag{6}$$

where $\{\widetilde{\mathbf{X}}_t\}, \{\widetilde{\mathbf{Y}}_t\}$ are Lagrange multipliers, $\mu > 0$ is a pre-specified step-size parameter, $\|\cdot\|_2$ denotes the ℓ_2 norm and $\|\cdot\|_F$ is the Frobenius norm. In a rescaled form, (6) can be written as

$$
\begin{aligned}
&L_\mu(\{\mathbf{L}_t\}, \{\mathbf{S}_t\}, \{\boldsymbol{\alpha}_t\}, \{\mathbf{X}_t\}, \{\mathbf{Y}_t\}) \\
&= \sum_{t=1}^T \|\mathbf{L}_t\|_* + \lambda_1 \sum_{t=1}^T \|\mathbf{S}_t\|_1 + \lambda_2 \sum_{t=1}^T \|\boldsymbol{\alpha}_t\|_1 \\
&+ \frac{\mu}{2} \sum_{t=1}^T \Big(\|\mathbf{L}_t + \mathbf{S}_t - \mathbf{Z}_t + \mathbf{X}_t\|_F^2 \\
&+ \|\mathbf{A}\mathrm{vec}(\mathbf{L}_t) - \boldsymbol{\alpha}_t + \mathbf{Y}_t\|_2^2 - \|\mathbf{X}_t\|_F^2 - \|\mathbf{Y}_t\|_2^2 \Big) \quad (7)
\end{aligned}
$$

where $\mathbf{X}_t = \mu^{-1}\widetilde{\mathbf{X}}_t$ and $\mathbf{Y}_t = \mu^{-1}\widetilde{\mathbf{Y}}_t$ are rescaled Lagrange multipliers.

We can obtain the parameter estimates by minimizing the augmented Lagrangian via ADMM algorithm [19], which alternates between minimizing (7) with respect to each set of primal variables $\mathbf{L}_t, \mathbf{S}_t, \boldsymbol{\alpha}_t$ sequentially, and the updating of the dual variables $\mathbf{X}_t, \mathbf{Y}_t$. Let initialize the parameters by $(\mathbf{L}_t^0, \mathbf{S}_t^0, \boldsymbol{\alpha}_t^0, \mathbf{X}_t^0, \mathbf{Y}_t^0) = (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})$ for $t = 1, \ldots, T$. The proposed ADMM algorithm solves the following subproblems iteratively until convergence, with parameter updates at each iteration k given by

$$
\{\mathbf{L}_t^k\} = \underset{\{\mathbf{L}_t\}}{\arg\min}\, L_\mu(\{\mathbf{L}_t\}, \{\mathbf{S}_t^{k-1}\}, \{\boldsymbol{\alpha}_t^{k-1}\}, \{\mathbf{X}_t^{k-1}\}, \{\mathbf{Y}_t^{k-1}\})
$$
$$(8)$$

$$
\{\mathbf{S}_t^k\} = \underset{\{\mathbf{S}_t\}}{\arg\min}\, L_\mu(\{\mathbf{L}_t^k\}, \{\mathbf{S}_t\}, \{\boldsymbol{\alpha}_t^{k-1}\}, \{\mathbf{X}_t^{k-1}\}, \{\mathbf{Y}_t^{k-1}\})
$$
$$(9)$$

$$
\{\boldsymbol{\alpha}_t^k\} = \underset{\{\boldsymbol{\alpha}_t\}}{\arg\min}\, L_\mu(\{\mathbf{L}_t^k\}, \{\mathbf{S}_t^k\}, \{\boldsymbol{\alpha}_t\}, \{\mathbf{X}_t^{k-1}\}, \{\mathbf{Y}_t^{k-1}\})
$$
$$(10)$$

$$
\mathbf{X}_t^k = \mathbf{X}_t^{k-1} + \mathbf{L}_t^k + \mathbf{S}_t^k - \mathbf{Z}_t, \quad \forall t = 1, \ldots, T \quad (11)
$$

$$
\mathbf{Y}_t^k = \mathbf{Y}_t^{k-1} + \mathbf{A}\mathrm{vec}(\mathbf{L}_t^k) - \boldsymbol{\alpha}_t^k, \quad \forall t = 1, \ldots, T \quad (12)
$$

Next we derive the solution for each subproblem.

1) Update $\mathbf{L}_t$: To update $\{\mathbf{L}_t^k\}$, the subproblem (8) can be separated into minimizations with respect to $\mathbf{L}_t$'s at individual time points (similarly for subproblems (9)-(10)), where each $\mathbf{L}_t$ is updated independently

$$
\begin{aligned}
\mathbf{L}_t^k = \underset{\mathbf{L}_t}{\arg\min} \,&\|\mathbf{L}_t\|_* + \frac{\mu}{2}\|\mathbf{L}_t + \mathbf{S}_t^{k-1} - \mathbf{Z}_t + \mathbf{X}_t^{k-1}\|_F^2 \\
&+ \frac{\mu}{2}\|\mathbf{A}\mathrm{vec}(\mathbf{L}_t) - \boldsymbol{\alpha}_t^{k-1} + \mathbf{Y}_t^{k-1}\|_2^2.
\end{aligned}
\quad (13)
$$

Due to the non-identity matrix $\mathbf{A}$, (13) does not have a closed-form solution. Motivated by the linearized ADMM method in [20] recently applied to lasso regression [21], [22], we linearize the quadratic term $\frac{1}{2}\|\mathbf{A}\mathrm{vec}(\mathbf{L}_t) - \boldsymbol{\alpha}_t^{k-1} + \mathbf{Y}_t^{k-1}\|_2^2$ in (13) as

$$
\begin{aligned}
&\left(\mathbf{A}^T\left(\mathbf{A}\mathrm{vec}(\mathbf{L}_t^{k-1}) - \boldsymbol{\alpha}_t^{k-1} + \mathbf{Y}_t^{k-1}\right)\right)^T \left(\mathrm{vec}(\mathbf{L}_t) - \mathrm{vec}(\mathbf{L}_t^{k-1})\right) \\
&+ \frac{\nu}{2}\|\mathrm{vec}(\mathbf{L}_t) - \mathrm{vec}(\mathbf{L}_t^{k-1})\|_2^2
\end{aligned}
\quad (14)
$$

where the parameter $\nu > 0$ controls the proximity to $\mathbf{L}_t^{k-1}$. Substituting (14) into (13) and after some algebra, we obtain

---

Algorithm 1 ADMM for Fused L+S Decomposition

Input: A series of vectorized multi-subject dynamic connectivity metrics $\mathbf{Z}_1, \ldots, \mathbf{Z}_T$.

Parameters: $\lambda_1 = 1/\sqrt{\max(N^2, M)}$, $\lambda_2 > 0$, $\mu = 1/\|\mathbf{Z}_t\|_2$, $\nu > \rho(\mathbf{A}^T\mathbf{A})$ where $\rho(.)$ denotes the spectral radius.

Initialize: $(\mathbf{L}_t^0, \mathbf{S}_t^0, \boldsymbol{\alpha}_t^0, \mathbf{X}_t^0, \mathbf{Y}_t^0) = (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})$, $\forall t = 1, \ldots, T$.

1: **for** $t = 1 : T$ **do**
2:     **while** not converged **do**
3:         Compute $\mathbf{C}_t$: $\mathrm{vec}(\mathbf{C}_t) = \mathbf{c}_t$ with $\mathbf{c}_t = \mathrm{vec}(\mathbf{L}_t^{k-1}) - \mathbf{A}^T\left(\mathbf{A}\mathrm{vec}(\mathbf{L}_t^{k-1}) - \boldsymbol{\alpha}_t^{k-1} + \mathbf{Y}_t^{k-1}\right)/\nu$
4:         Compute $\mathbf{Q}_t = \frac{1}{\nu+1}(\mathbf{Z}_t - \mathbf{S}_t^{k-1} + \nu\mathbf{C}_t - \mathbf{X}_t^{k-1})$
5:         Update $\mathbf{L}_t^k = \mathcal{D}_{1/\mu(\nu+1)}(\mathbf{Q}_t)$
6:         Update $\mathbf{S}_t^k = \mathcal{S}_{\lambda_1/\mu}(\mathbf{Z}_t - \mathbf{L}_t^k - \mathbf{X}_t^{k-1})$
7:         Update $\boldsymbol{\alpha}_t^k = \mathcal{S}_{\lambda_2/\mu}(\mathbf{A}\mathrm{vec}(\mathbf{L}_t^k) + \mathbf{Y}_t^{k-1})$
8:         Update $\mathbf{X}_t^k = \mathbf{X}_t^{k-1} + \mathbf{L}_t^k + \mathbf{S}_t^k - \mathbf{Z}_t$
9:         Update $\mathbf{Y}_t^k = \mathbf{Y}_t^{k-1} + \mathbf{A}\mathrm{vec}(\mathbf{L}_t^k) - \boldsymbol{\alpha}_t^k$
10:    **end while**
11: **end for**

Output: $\widehat{\mathbf{L}}_1, \ldots, \widehat{\mathbf{L}}_T$ and $\widehat{\mathbf{S}}_1, \ldots, \widehat{\mathbf{S}}_T$.

---

the following approximate subproblem for (13)

$$
\begin{aligned}
\mathbf{L}_t^k = \underset{\mathbf{L}_t}{\arg\min} \,&\frac{1}{\mu}\|\mathbf{L}_t\|_* + \frac{1}{2}\|\mathbf{L}_t + \mathbf{S}_t^{k-1} - \mathbf{Z}_t + \mathbf{X}_t^{k-1}\|_F^2 \\
&+ \frac{\nu}{2}\|\mathbf{L}_t - \mathbf{C}_t\|_F^2
\end{aligned}
\quad (15)
$$

where $\mathbf{C}_t \in \mathbb{R}^{N^2 \times M}$ is a matrix such that $\mathrm{vec}(\mathbf{C}_t) = \mathbf{c}_t$ with $\mathbf{c}_t = \mathrm{vec}(\mathbf{L}_t^{k-1}) - \mathbf{A}^T\left(\mathbf{A}\mathrm{vec}(\mathbf{L}_t^{k-1}) - \boldsymbol{\alpha}_t^{k-1} + \mathbf{Y}_t^{k-1}\right)/\nu$. We show that (15) can be solved via the singular value thresholding (SVT) [23]. More precisely, let $\mathcal{S}_\tau(\mathbf{M})$ be the element-wise soft-thresholding (shrinkage) operator for a matrix $\mathbf{M}$ at level $\tau > 0$, $(\mathcal{S}_\tau(\mathbf{M}))_{ij} = \mathrm{sign}(m_{ij})\max(|m_{ij}| - \tau, 0)$. We also denote by $\mathcal{D}_\tau(\mathbf{M})$ the singular value thresholding operator given by $\mathcal{D}_\tau(\mathbf{M}) = \mathbf{U}\mathcal{S}_\tau(\boldsymbol{\Sigma})\mathbf{V}^T$ where $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ is the singular value decomposition (SVD) of $\mathbf{M}$. Then, $\mathbf{L}_t^k$ admits an explicit solution

$$
\mathbf{L}_t^k = \mathcal{D}_{1/\mu(\nu+1)}(\mathbf{Q}_t), \quad \mathbf{Q}_t = \frac{1}{\nu+1}(\mathbf{Z}_t - \mathbf{S}_t^{k-1} + \nu\mathbf{C}_t - \mathbf{X}_t^{k-1})
$$
$$(16)$$

The derivation of (16) is given in Supplementary Section I. Note that when $\nu \to 0$, (16) reduces to the solution of the ALM algorithm for the low-rank component [18] in the original L+S decomposition problem (2).

2) Update $\mathbf{S}_t$: The minimization in subproblem (9) with respect to $\mathbf{S}_t$ at each time point is equivalent to

$$
\mathbf{S}_t^k = \underset{\mathbf{S}_t}{\arg\min}\, \frac{\mu}{2}\|\mathbf{L}_t^k + \mathbf{S}_t - \mathbf{Z}_t + \mathbf{X}_t^{k-1}\|_F^2 + \lambda_1\|\mathbf{S}_t\|_1
$$

We can get a closed-form solution by applying the element-wise soft-thresholding operator

$$
\mathbf{S}_t^k = \mathcal{S}_{\lambda_1/\mu}(\mathbf{Z}_t - \mathbf{L}_t^k - \mathbf{X}_t^{k-1}). \quad (17)
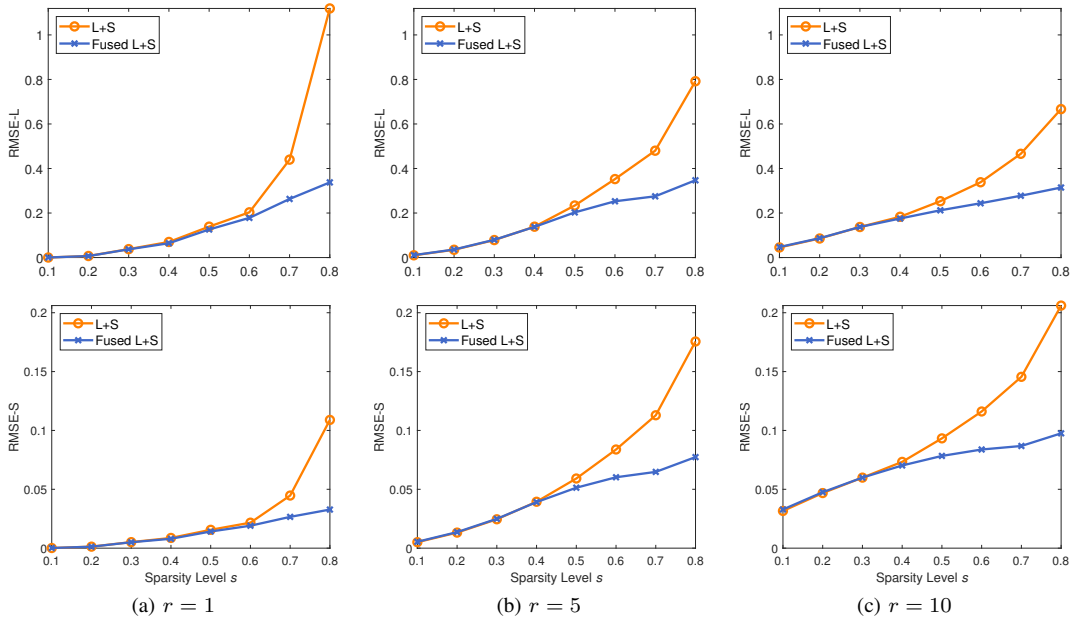$$

Fig. 1. Performance comparison of original and fused L+S decomposition algorithms in recovering low-rank and sparse structures from simulated multi-subject connectivity matrices. RMSEs of estimated low-rank (top) and sparse (bottom) components against increasing level of sparsity s for different ranks r. Results are averages over 100 replications.

3) Update $\boldsymbol{\alpha}_t$: Similarly, the subproblem (10) with respect to each $\boldsymbol{\alpha}_t$ is

$$\boldsymbol{\alpha}_t^k = \arg\min_{\boldsymbol{\alpha}_t} \frac{\mu}{2} \left\| \mathbf{A}\text{vec}(\mathbf{L}_t^k) - \boldsymbol{\alpha}_t + \mathbf{Y}_t^{k-1} \right\|_2^2 + \lambda_2 \left\| \boldsymbol{\alpha}_t \right\|_1$$

where the solution is also a soft-thresholding operation

$$\boldsymbol{\alpha}_t^k = \mathcal{S}_{\lambda_2/\mu}(\mathbf{A}\text{vec}(\mathbf{L}_t^k) + \mathbf{Y}_t^{k-1}). \quad (18)$$

We chose the parameter $\lambda_1 = 1/\sqrt{\max(N^2, M)}$, as suggested by [14] for the PCP. The tuning parameter λ_2 can be selected using cross-validation. To establish the convergence of the linearized ADMM algorithm, the parameter ν is required to satisfy $\nu > \rho(\mathbf{A}^T\mathbf{A})$ where $\rho(.)$ denotes the spectral radius [22]. We set $\nu = 1.5\rho(\mathbf{A}^T\mathbf{A})$. We set the stopping criterion as

$$\frac{\left\| (\mathbf{L}_t^k, \mathbf{S}_t^k) - (\mathbf{L}_t^{k-1}, \mathbf{S}_t^{k-1}) \right\|_F}{\max\left\{ \left\| (\mathbf{L}_t^{k-1}, \mathbf{S}_t^{k-1}) \right\|_F, 1 \right\}} \leq 10^{-6}, \ \forall t = 1, \ldots, T$$

or the maximum iteration number of 5000 is reached. The proposed ADMM algorithm for solving the fused L+S decomposition is summarized in Algorithm 1.

IV. SIMULATION

We evaluate the performance of the proposed fused L+S decomposition method via simulation studies. The goal is to assess the performance of L+S algorithms in separating the common structure from individual-specific background components in multi-subject FC networks. Here, we focus on the static snapshot of dynamic networks at a particular time point, and drop the index t for notational brevity. We generate the data matrix of concatenated connectivity metrics of weighted, undirected networks for M subjects as follows

$$\mathbf{Z} = \mathbf{B}\boldsymbol{\beta} + \mathbf{S} \quad (19)$$

where $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_M]$, $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_M]$ with $\mathbf{s}_i = \text{vec}(\boldsymbol{\Sigma}_i^s)$, $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_r] \in \mathbb{R}^{N^2 \times r}$ whose columns correspond to a set of r vectorized basis connectivity matrices $\mathbf{b}_j = \text{vec}\left(\boldsymbol{\Sigma}_j^b\right)$ that are shared across subjects, and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M]$ with $\boldsymbol{\beta}_i \in \mathbb{R}^r$ being a vector of mixing weights for subject i. To emulate the modular structure of brain networks [24] induced by stimuli or tasks, we generate each basis connectivity matrix $\boldsymbol{\Sigma}_j^b$ from a stochastic block model (SBM) [25] with two equal-sized communities with intra- and inter-community edge probabilities of 0.95 and 0.2, respectively. The edge strengths are randomly drawn from the uniform distribution $U[-1, 1]$. The subject-dependent mixing weights $\boldsymbol{\beta}_i$ are clustered according to two subgroups, which are independently sampled from normal distributions $\boldsymbol{\beta}_i \sim N(0.5\mathbf{I}, \epsilon\mathbf{I})$ for $i = 1, \ldots, M/2$ and $\boldsymbol{\beta}_i \sim N(\mathbf{0}, \epsilon\mathbf{I})$ for $i = M/2 + 1, \ldots, M$ to represent the strong and the weak response to the stimulus in the two subgroups, respectively. We set $\epsilon = 0.005$ such that the variability within each subgroup is small. The resulting matrix $\mathbf{L} = \mathbf{B}\boldsymbol{\beta}$ has rank r, and the connectivity profiles in the columns of $\mathbf{L}$ are similar within each subgroup. The background components of individual subjects $\boldsymbol{\Sigma}_i^s$ are generated independently as sparse symmetric matrices with support set uniformly chosen at random at a constant sparsity level s (fraction of non-zero entries), and whose non-zero entries are i.i.d. samples from $U[-5, 5]$ indicating large magnitude of noise in $\mathbf{Z}$. The simulations were replicated 100 times.

We compare the performance of the fused L+S decomposition with the original version of PCP in recovering $\mathbf{L}$ and $\mathbf{S}$ from the observed matrix $\mathbf{Z}$. For the fused L+S, we select the optimal parameter λ_2 over a large grid of values, which gives the best recovery performance on an independently generated validation set. To measure the accuracy of recovery, we computed the root mean squared error (RMSE)

TABLE I

COMPUTATIONAL TIME (IN SECONDS) AND NUMBER OF ITERATIONS OF PROPOSED ADMM ALGORITHM FOR FUSED L+S DECOMPOSITION ON SIMULATED MULTI-SUBJECT FC MATRICES FOR INCREASING DIMENSION AND RANK. RESULTS ARE AVERAGES OVER 100 REPLICATIONS.

| Data matrix $\mathbf{Z}$ | | Error $\varepsilon = 10^{-3}$ | | | | Error $\varepsilon = 10^{-4}$ | | | |
Size ($N^2 \times M$)	Rank r	Time (s)	# Iter	Rank $\hat{r}$	RMSE-L	Time (s)	# Iter	Rank $\hat{r}$	RMSE-L
	1	0.008	50.05	1	4.35e-03	0.011	74.92	1	4.82e-04
100×20	5	0.032	255.94	5	5.62e-02	0.242	1938.69	6	5.26e-02
	10	0.036	278.85	8	1.25e-01	0.185	1422.02	9	1.26e-01
	1	0.031	50	1	4.31e-02	0.189	320.59	1	3.59e-02
400×30	5	0.095	169.95	5	4.33e-02	0.802	1365.25	5	3.72e-02
	10	0.175	318.39	10	7.98e-02	1.772	3040.56	10.01	7.85e-02
	1	0.055	49	1	5.07e-03	0.197	181.84	1	4.52e-04
625×50	5	0.088	82.93	5	1.36e-02	0.439	416.79	5	8.31e-03
	10	0.362	317.74	10	3.64e-02	2.757	2590.46	10.99	3.15e-02
	1	0.679	49	1	3.84e-03	1.729	132.11	1	3.86e-04
10000×50	5	2.124	99.02	5	1.46e-02	6.576	502.54	5	1.06e-02
	10	6.793	309.05	10	4.57e-02	29.113	2230.26	10	4.19e-02
	1	3.981	49	1	4.95e-03	12.762	228.05	1	7.31e-04
40000×50	5	12.251	150.99	5	2.58e-02	51.139	840.03	5	2.07e-02
	10	32.343	397	10	7.10e-02	179.929	3375.88	12	6.80e-02
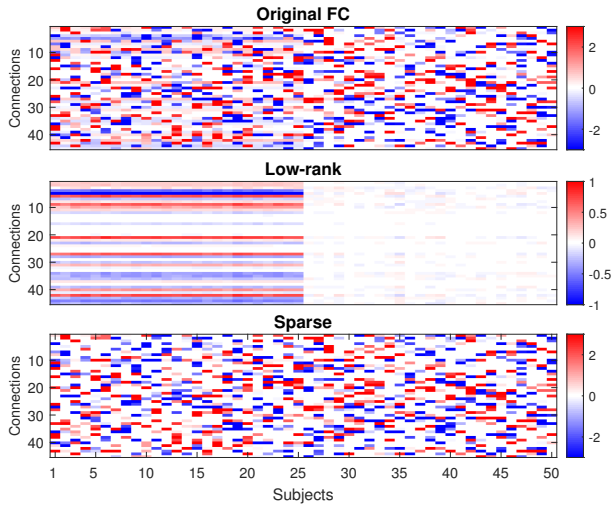
Fig. 2. L+S recovery of synthetic multi-subject FC networks. One realization of simulated matrix $\mathbf{Z}$ (top) whose columns are vectorized upper-triangular part of undirected FC matrices with $N = 10$ nodes for $M = 50$ subjects. The dimension is $N(N-1)/2 \times M = 45 \times 50$. Low-rank $\widehat{\mathbf{L}}$ (middle) and sparse $\widehat{\mathbf{S}}$ (bottom) components recovered from $\mathbf{Z}$ using the fused L+S method.
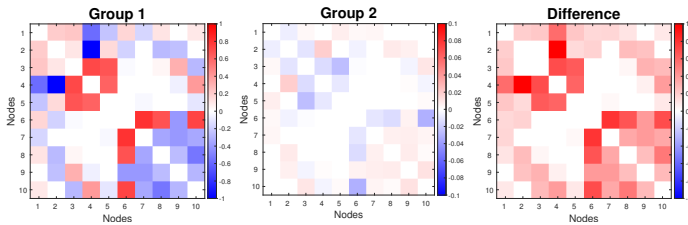

Fig. 3. Group difference in connectivity pattern identified by the low-rank components from one realization of simulated multi-subject FC matrix $\mathbf{Z}$ using the fused L+S method. Group mean FC maps (left & middle) for two groups of subjects (Group 1: $i = 1, \ldots, 25$; Group 2: $i = 26, \ldots, 50$), and their differences in terms of edge strength. The group differences shown are significantly different from zero ($p < 0.0001$, Bonferroni-corrected).

between the ground-truth and estimated low-rank and sparse components, defined by RMSE-L $= \|\widehat{\mathbf{L}} - \mathbf{L}\|_F / \|\mathbf{L}\|_F$ and RMSE-S $= \|\widehat{\mathbf{S}} - \mathbf{S}\|_F / \|\mathbf{S}\|_F$.

1) Effect of Rank & Sparsity: We investigate robustness of the methods to increasing rank r and sparsity level s of the ground-truth components. Fig. 1 plots the average RMSEs over 100 replications as a function of s (varied from 0.1 to 0.8 with an increment of 0.1) for $r = 1, 5, 10$ with fixed $N = 10$ and $M = 50$. As expected, estimation errors of both methods for the low-rank and sparse components increase with s and r, generally. The larger fraction of corrupted entries and the rank of the data matrix render the estimated $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{S}}$ more deviated from the ground-truth, in consistency with theoretical results in previous studies [14]. Both methods perform comparably when s is small. However, as s increases, the fused L+S estimator clearly outperforms the original version with substantially lower errors with a slower growth rate. This suggests the robustness of the fused L+S method under the presence of severe dense noise. It provides a better recovery of the low-rank matrix of shared connectivity patterns even if almost all of its entries are grossly corrupted by the background components, as evidenced by the relatively stable errors when $s = 0.7$ and $s = 0.8$. This is due to the additional fused penalty term which smooths out the individual-specific random background noise, when leveraging on the inter-subject similarity of the stimulus-induced connectivity profiles. Fig. 2 shows fused L+S decomposition of a simulated multi-subject connectivity matrix $\mathbf{Z}$ with underlying rank $r = 5$ and a fraction $s = 0.5$ of corrupted edges in each subject. The estimate $\widehat{\mathbf{L}}$ successfully recovers the correlated connectivity patterns across subjects with distinct sub-group shared responses from the highly-corrupted $\mathbf{Z}$, while $\widehat{\mathbf{S}}$ filters out subject-specific background noise without prior information of the locations of the corrupted edges (the support of $\mathbf{S}$). In Fig. 3, we can see apparent group differences in FC patterns between the two groups of subjects, as identified by the extracted low-rank components $\widehat{\mathbf{L}}$ in Fig. 2 (middle). This suggests the capability of the fused L+S method to produce different groupings of subjects where subjects in each group share similar connectivity pattern.
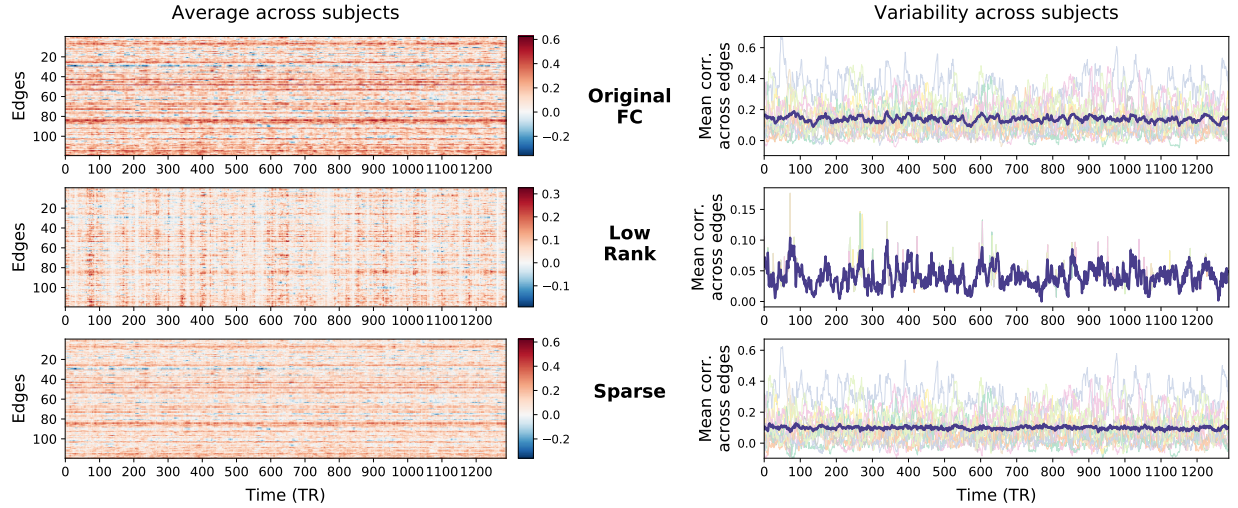
Fig. 4. Fused L+S decomposition of time-resolved, multi-subject FC from movie-watching fMRI. (Left) Original dynamic pairwise correlations between 16 ROIs and the estimated low-rank and sparse components, averaged over all subjects. (Right) Mean dynamic correlations across all edges and the estimated low-rank and sparse components. The group-average (blue lines) and individual (thin lines) time courses depict distributions of dynamic correlations over subjects. The time-varying low-rank components are temporally locked across subjects, suggesting potential shared response to stimulus sequence in the movie.

2) Computational Time: Table I shows computational results of the fused L+S algorithm for increasing matrix dimension (N, M) and rank r of the simulated multi-subject FC data. We report computational time and number of iterations required to achieve reconstruction error $\epsilon = \|\mathbf{Z} - \widehat{\mathbf{L}} - \widehat{\mathbf{S}}\|_F / \|\mathbf{Z}\|_F$. Results were obtained using MATLAB R2019b on Laptop with 2.8 GHz Intel Core i7 CPU and 16 GB memory. The proposed algorithm is fairly robust against the increased number of network nodes N and number of subjects M, as evident from the relatively stable RMSEs of the extracted low-rank components $\widehat{\mathbf{L}}$ and successful recovery of the underlying rank r. Nevertheless, we can see that longer run time was used to recover the low-rank group connectivity structure $\mathbf{L}$, as N and M increase. This is because, for an $N^2 \times M$ input matrix, our algorithm which builds on the linearized ADMM has a computational complexity of $O(N^2 M \min(N^2, M))$ per iteration required in computation of SVD for the nuclear norm regularization. Besides, more computational effort was taken to recover matrices with higher rank r. Our method requires at most $O(1/\epsilon)$ iterations to converge to an ϵ-optimal solution, as evident from more iterations needed to reach a smaller error of $\epsilon = 10^{-4}$ compared to $\epsilon = 10^{-3}$.

V. ANALYSIS OF MOVIE-WATCHING fMRI

In this section, we applied the proposed fused L+S decomposition approach to separating the stimulus-related and background components in time-varying FC from fMRI data of a group of subjects exposed to complex naturalistic stimulation from movie viewing. We tested the established hypothesis that, as speech becomes more difficult to process, e.g., as through speech-in-noise, somatosensory, motor, and auditory systems are engaged more to facilitate processing [26].

A. Data Acquisition & Stimuli

Thirteen subjects watched a television show during which 3 Tesla fMRI data were collected at the Joan & Sanford I. Weill Medical College of Cornell University by Dr Skipper (GE Medical Systems, Milwaukee, WI; TR = 1.5 s; see [27] for full details). The show, *Are you smarter than a 5th grader?*, was 32min 24sec long. Among other features, audio events were annotated at a 25ms resolution by a trained coder. These were grouped into five general categories, including *noise*, *music*, *speech* (speech that occurred in silence), *sp+noise* (speech that co-occurred with noise, including background talking, clapping, music etc.) and *silence*. The audio annotations were down-sampled to match the resolution of fMRI data by selecting the auditory category that occurred most frequently within each 1.5 s TR.

B. Preprocessing

The fMRI data were minimally preprocessed, including slice timing correction, despiking, registration, nonlinear transformation to MNI coordinate space, spatial smoothing and masking of voxels outside of the brain. We used the Freesurfer parcellation to obtain 16 regions of interest (ROIs) associated with language and auditory processing: These ROIs include the central sulcus (CS), planum polare (PP), planum temporale (PT), postcentral gyrus (PoCG), precentral gyrus (PreCG), superior temporal gyrus (STG), transverse temporal gyrus (TTG) and transverse temporal sulcus (TTS) from both left and right hemispheres. These regions are commonly associated with auditory, somatosensory and motor processing generally and speech perception more specifically. The mean BOLD time series were computed across voxels within each ROI.

C. Shared Structure in Dynamic Connectivity Across Subjects

To characterize dynamic FC, we computed time-varying correlations between the 16 ROI fMRI time series for each subject, using sliding windows of 15 TRs (22.5 s) with step size of 1 TR (1.5 s) between windows. The choice of window length of 22.5 s was based on the inverse of the lowest frequency present in the data [28] after high-pass filtering at 0.05 Hz. This window length has been used in [8] to

compute the sliding-window ISFC for movie-watching fMRI data. See Supplementary Section II.D on additional analyses on the effect of window length. At each time point t, the vectorized correlation coefficients of individual subjects are stacked as columns of matrix $\mathbf{Z}_t \in \mathbb{R}^{120 \times 13}$. Similar to fused LASSO [29], the fused L+S decomposition assumes some natural orderings of subjects in $\mathbf{L}_t$. As a preliminary step, we applied the k-means clustering on the columns of $\mathbf{L}_t$ recovered by the original non-fused PCP to estimate the subject orderings, putting subjects with similar connectivity near one another in $\mathbf{Z}_t$. We then performed the fused L+S decomposition on the multi-subject dynamic correlation matrices $\mathbf{Z}_1, \ldots, \mathbf{Z}_T$ by solving (3) using the proposed ADMM algorithm. We selected the penalty parameter λ_2 based on the performance of the resulting low-rank models in predicting the time-varying auditory stimuli (see Section IV-E). From a range of λ_2 values from 0.001 to 0.5, we identified $\lambda_2 = 0.05$ as optimal, giving the highest classification accuracy on the validation set in a leave-one-subject-out cross-validation. Fig. 4 shows the extracted time-varying low-rank and sparse components averaged across subjects and across all edges. In Fig. 4 (left), the low-rank component shows noticeable time-varying structure with higher specificity in time compared to the original dynamic correlations. It provides a better localization of dynamic changes in connectivity which were possibly elicited by the movie and time-locked to the processing of specific ongoing movie events. From Fig. 4 (right), it is apparent that the low-rank components are highly synchronized across subjects over time, while identifying some inter-subject variability possibly arising from varied responses of different individuals to the stimuli. In contrast, the sparse components show temporally uncorrelated patterns. The low-rank part is specifically sensitive to synchronized responses across subjects. This suggests that the proposed fused L+S decomposition is capable of isolating reliable FC changes driven by a correlated source across subjects (exposed to the same movie stimuli in our case). The uncorrelated background sources of dynamic FC across subjects are filtered out as the residual sparse components. These sources may correspond to subject-specific dynamic fluctuations of FC networks as well as non-neuronal artifacts unrelated to the stimulus-processing. In comparison with the time-varying ISFC and independent component analysis (ICA)-based dynamic connectivity (See Supplementary Section II.C), our method also shows a better sensitivity in detecting the shared FC dynamics across subjects (Supplementary Fig. 3).

For illustration, Fig. 5 shows decomposition for a snapshot of connectivity networks $\mathbf{Z}_t$ at time point $t = 59$ TRs (selected based on median correlation over time). The low-rank component identifies a common network structure underlying all subjects, as evident from the similar and highly-correlated connectivity patterns detected across subjects. As validated by simulation, our algorithm can reliably recover the common structure from the subject-specific background noise, which could be arbitrarily large in magnitude and contaminated most of the network correlations (e.g., the dense residuals for subject 3). The estimated rank was $\hat{r}_t = 2$, where individual networks (columns of $\widehat{\mathbf{L}}$) can be expressed as a linear combination of

2 common basis connectivity matrices with different mixing weights for each subject. See Supplementary Fig. 1 for additional example decomposition at $t = 72$ TRs.

D. Low-Rank Components Reveal Stimulus-Induced Connectivity Patterns

To test the aforementioned hypothesis, we examined whether the extracted low-rank components are related to the auditory stimuli present in the movie. We did so by fitting a multinomial logistic regression model to learn the mapping between the audio annotations and the time-resolved low-rank connectivity metrics. We set the auditory category labels as responses and the connectivity edges as predictors. The silence events were used as the reference category. To evaluate the goodness-of-fit and how well the low-rank components are temporally mapped to the auditory events, we used the fitted logistic model to decode the probability of the actual category labels at each time point. Fig. 6 shows the decoded time courses for each auditory category. For all categories, high probabilities are precisely aligned with the intervals when those categories are present, indicating accurate mapping of the low-rank components to the stimulus annotations.

In Fig. 7, we plot the mean functional connectivity maps and the estimated regression coefficient maps associated with the four auditory categories in the movie. The results are based on the low-rank components at a lag of 3 s (i.e., $\widehat{\mathbf{L}}_{t-2}$ at 2 time points before stimulus onsets at t). The latter was empirically determined to exhibit the strongest connectivity by testing at previous and future time points (at intervals of 1.5 s or 1 TR) relative to the stimulus onsets (See Supplementary Fig. 2). This lag was expected because it approximately includes to delay, rise time and peak of a canonical hemodynamic response function associated with any one stimulus lasting about 1.5 s. As shown in Fig. 7, the low-rank component detected distinct connectivity patterns, with relatively different distributions of connectivity across the different auditory environments. It provides significantly stronger stimulus-related effects compared to ISFC and ICA methods (Supplementary Fig. 4). As hypothesized, we observe a noticeable increase in connectivity when speech was accompanied by noise (*sp+noise*), compared to speech in silence. We used a group-level t-test to contrast the FC between the *sp+noise* and *speech* conditions. Fig. 8(a) shows difference in the mean FC matrices (across subjects and time points) from the low-rank component between the two conditions. The connections were thresholded at $p < 0.0001$ (Bonferroni-corrected). Specifically, there was significantly higher FC throughout regions typically involved in speech perception and language comprehension for speech-in-noise compared to speech. This included connections PT.R-TTS.R and PT.R-STG.R which showed the largest increase in FC strength. Furthermore, our method identified significant increase in connectivity between the regions of auditory and speech processing with the premotor, motor and somatosensory regions in response to speech mixed with noise, e.g., TTS.L-PoCG.R, PT.R-PreCG.R and TTG.R-PoCG.R, as shown in Fig. 8(b).

In summary, results support hypotheses about the role of 'the motor system' in speech perception [26], [30]. In
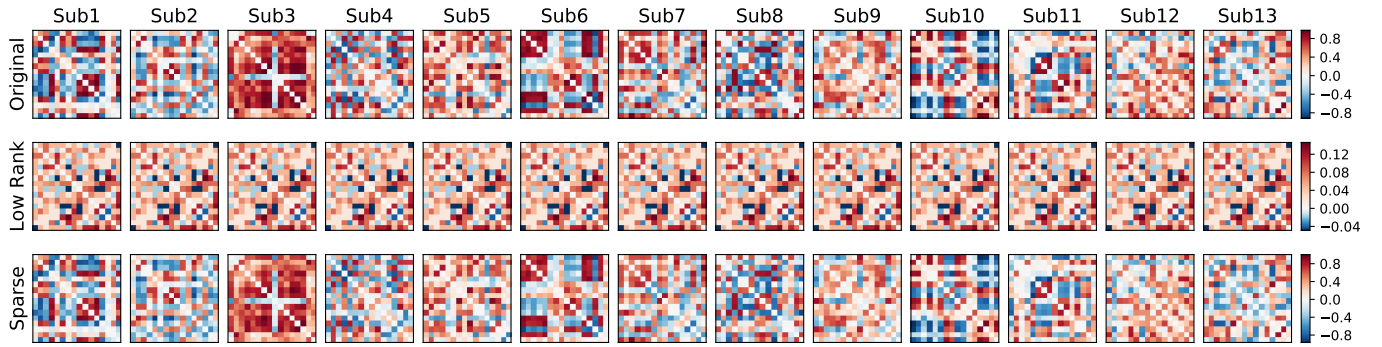
Fig. 5. Common structure in time-resolved functional connectivity across subjects. (Top) An example of original functional connectivity matrices $\mathbf{Z}_t = [\mathbf{z}_{t1}, \ldots, \mathbf{z}_{tM}]$ for $M = 13$ subjects at time window $t = 59$ TRs. (Middle) Low-rank component $\widehat{\mathbf{L}}_t$ with $\hat{r}_t = 2$ recovered by our fused L+S approach, revealing shared connectivity structure across subjects. (Bottom) Estimated sparse errors or residuals $\widehat{\mathbf{S}}_t$ correspond to individual-specific background components. The selected time window corresponds to the median value of mean correlation over all edges and subjects.
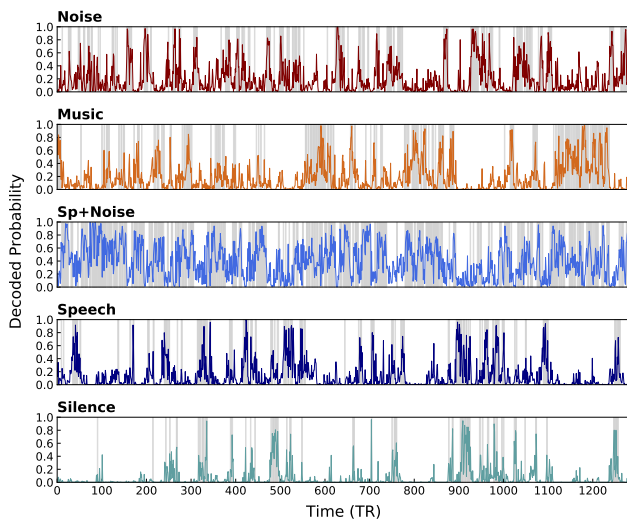
Fig. 6. Decoded probability of a specific auditory stimulus category being present in the movie over time based on the multinomial logistic regression fitted on the extracted low-rank components. Colored lines indicate time courses of decoded probabilities averaged over all subjects. Gray regions show time points when the stimulus category was actually present in the movie.

particular, a distributed set of brain regions involved in producing speech are dynamically recruited in perceiving speech, particularly in more effortful listening situations. This might be because speech production regions are engaged in predicting upcoming speech sounds. In noisy environments, this process needs to be engaged more as predictions are inherently less accurate.

E. Across-Subject Prediction of Stimulus Annotations

We trained a classifier based on multilayer fully-connected neural network (FCNN) on the time-varying low-rank FC components to evaluate any improvement in predicting the different auditory categories over the original observed version of time-varying FC metrics. For performance comparison, we also built FCNN classifiers based on the ISFC and ICA-based dynamic FC metrics. Following [7], we performed across-subject classification but using the time-resolved instead of static FC patterns to predict the stimulus events over time. We used leave-one-subject-out cross-validation. In each fold, one subject was held out for testing and the remaining $M - 1$ subjects were used to train the FCNN. We

computed the time-resolved low-rank components $\mathbf{L}_t^{(m)} = [\mathbf{l}_{t,1}, \mathbf{l}_{t,m-1}, \ldots, \mathbf{l}_{t,m+1}, \mathbf{l}_{t,M}]$ for the training set from the standard FC, $\mathbf{Z}_t^{(m)} = [\mathbf{z}_{t,1}, \mathbf{z}_{t,m-1}, \ldots, \mathbf{z}_{t,m+1}, \mathbf{z}_{t,M}]$ using the fused L+S algorithm. Since the test set is unseen from the training set, we extracted the low-rank components for the test subject m via orthogonal projection of the test data $\mathbf{l}_{t,m} = \mathbf{U}\mathbf{U}^T \mathbf{z}_{t,m}$, where $\mathbf{U}$ is a $N^2 \times r$ basis matrix estimated based on the SVD of train data $\mathbf{L}_t^{(m)}$. A 2-hidden layer FCNN with dimensions of 64 and 32 was selected over a range of model configurations, giving the best performance on the validation set in inner loop cross-validation. The model was trained using Adam optimizer, with learning rate of 0.01, reduce-factor of 0.1, 100 training epochs and a batch size of 64. Classification accuracy was computed as the percentage of time points correctly assigned to the true category labels.

Fig. 9 shows the classification results using the different time-varying FC metrics. The low-rank components significantly outperformed the standard FC, ISFC and ICA in discriminating ongoing auditory events in the movie. Both ISFC and ICA are slightly better than the standard FC. As expected, the residual sparse components, corresponding to uncorrelated effects of intrinsic neuronal processes and noise, were poorly predictive of the stimuli across subjects. The superior decoding performance suggests that the low-rank components are more stimulus-related, extracting novel information about the stimulus-induced dynamic functional changes not captured by the ISFC and ICA. Among the L+S algorithms, addition of fused penalty improves the cross-subject prediction considerably, achieving best accuracy of 78%. This may be due to the further smoothing of subject-specific fluctuations to uncover shared neuronal responses that are time-locked to the same stimulus processing. This is further evident from Fig. 10 that the fused low-rank dynamic FC can accurately decode the changes in audio events of the movie over time for all subjects, as indicated by higher accuracy of the decoded time courses compared to the time-varying ISFC. The advantage of the proposed method for naturalistic stimulus decoding was also observed when predicting word and non-word events in another movie on a dataset with more subjects (Supplementary Fig. 7).
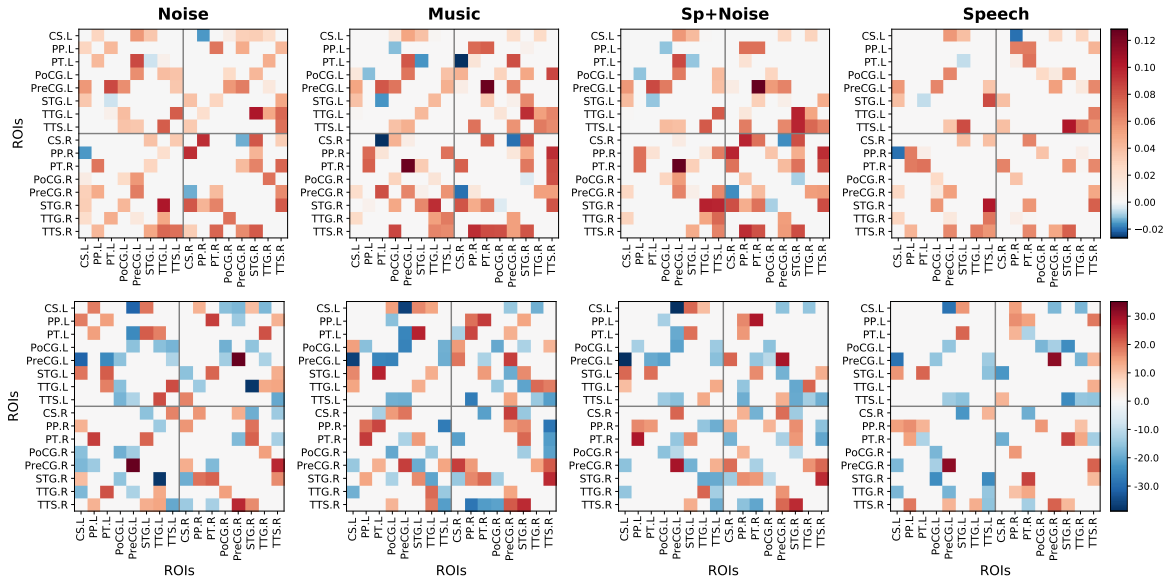
Fig. 7. Stimulus-induced connectivity networks during movie watching as revealed by the low-rank components. (Top) Average functional connectivity maps (across 13 subjects and time steps) associated with each of the four categories of auditory events in the natural movie. (Bottom) The corresponding spatial maps of regression coefficients from the fitted logistic model. Only functional connections showing significant stimulus-related effects are displayed (masked by regression coefficients with $p < 0.0001$, Bonferroni-corrected for multiple testings).

VI. DISCUSSION

We developed a novel L+S decomposition method for separating stimulus-induced and background components of dynamic FC networks during naturalistic neuroimaging, by leveraging on the time-locked nature of naturalistic stimuli. The method provides a principled way to recover both the stimulus-evoked, shared FC dynamics across individuals and the stimulus-unrelated idiosyncratic variations in individuals, respectively modeled as the low-rank and sparse structures of multi-subject dFC networks. The proposed fused PCP solved via an efficient ADMM algorithm can capture the stimulus-induced similarities in FC profiles between subjects more effectively, and its robustness to dense noise allows us to filter out corruptions of FC edges that are not necessarily sparse. Our proposed framework is general, broadly applicable to other neuroimaging data such as electroencephalography (EEG), and can incorporate other FC measures beyond simple correlations in ISC-based analyses. In an application to movie-watching fMRI, our method identified time-locked changes in FC networks across subjects, which were meaningfully related to the processing of complex, time-varying auditory events in the movie, e.g., dynamic recruitment of speech production systems in perceiving speech in noisy environments. It also revealed potentially interesting individual differences in FC patterns that may relate to behavioral appraisal of stimuli. The extracted low-rank FC components also show better prediction of the movie auditory annotations than ISFC and ICA, suggesting gain in information about the naturalistic stimuli by using shared network structure instead of synchronization of low-level regional activity. Thus, the proposed L+S recovery approach opens new opportunities for mapping brain network dynamics to stimulus features and behavior during naturalistic paradigms.

There are potential limitations of our approach. The two-dimensional matrix decomposition proposed in (1) might ne-glect the multilinear network structure by vectorizing the FC matrices. To preserve this structure, future work can consider a multi-way tensor generalization of the fused L+S method to decompose the time-varying, multi-subject FC tensor object $\mathcal{Z}_t \in \mathbb{R}^{N \times N \times M}$ into a low-rank tensor $\mathcal{L}_t$ and a sparse tensor $\mathcal{S}_t$. The optimization can build upon recent advances in robust tensor recovery [31] that minimizes the tensor nuclear norm $\sum_i \left\| \mathcal{L}_{t,(i)} \right\|_*$ (where $\mathcal{L}_{t,(i)}$ is low-rank part of mode-i unfolding of $\mathcal{Z}_t$), and ℓ_1 norm of sparse tensor, while incorporating the fusion penalty on the inter-subject differences between FC matrices (i.e., the mode-3 slices of $\mathcal{L}_t$). Another limitation is that the potentially meaningful individualized FC dynamics separated in the sparse component may be mixed with other non-neuronal artifacts. These confounding effects could be mitigated by pre-filtering steps before applying the L+S decomposition. Another solution is to extend the current model to $\mathbf{Z}_t = \mathbf{L}_t + \mathbf{S}_t + \mathbf{E}_t$ which allows for separate specification of the individual-specific neural dynamics ($\mathbf{S}_t$) and non-neuronal noise ($\mathbf{E}_t$). Further extensions are also needed to improve the scalability of the L+S learning algorithm which is computationally expensive for large-sized brain networks. Future studies can explore potential applications of L+S analysis of natural fMRI to connectome-based prediction of individual traits and behavior, and diagnosis of cognitive and affective disorders. One could investigate how the detected subject-specific FC dynamics relate to individual differences in interpretation of the complex stimuli due to different processing strategies across subjects, and between heterogeneous clinical diagnostic groups. One promising direction is to exploit the recent low-rank matrix response regression models for neuroimaging data [32], [33], to learn associations between the high-dimensional FC matrices and the behavioral and clinical outcomes. The nonparametric regression with nuclear norm regularization in [33] can also estimate low-rank representation within an FC matrix for
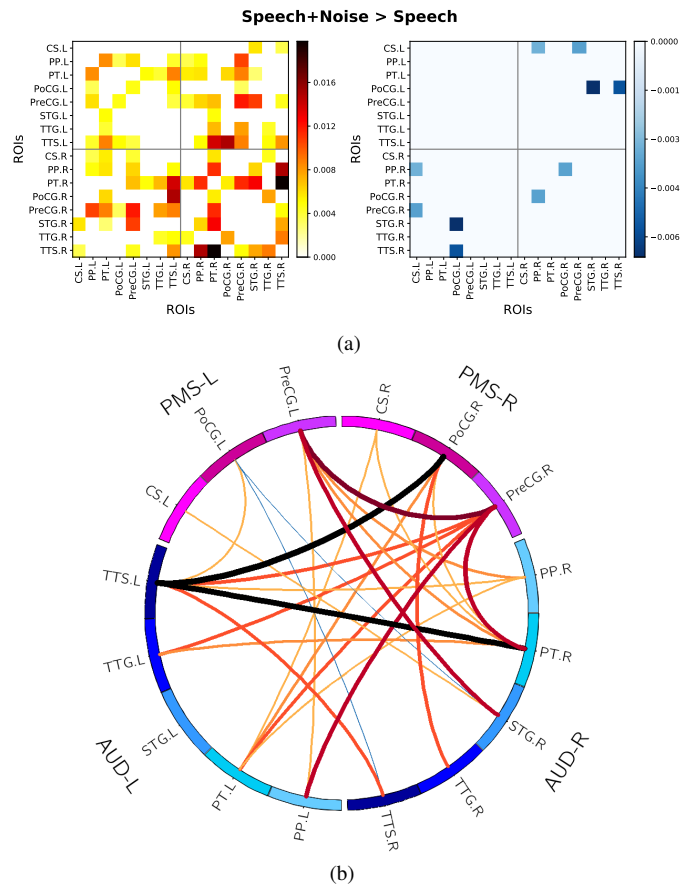
(a)

(b)

Fig. 8. Difference in functional connectivity between the speech-in-noise (*sp+noise*) and speech-only (*speech*) events as revealed by the low-rank components at time lag of 3 s (or 2 TRs). (a) Connectivity maps show increase (left) and decrease (right) in FC during *sp+noise* compared to *speech* events. Only edges with significant difference in FC strength (absolute value of correlation coefficients) are shown ($p < 0.0001$, Bonferroni-corrected). The five connections exhibiting largest increase in FC include PT.R-TTS.R, TTS.L-PoCG.R, PP.R-TTS.R, TTS.L-PT.R and PT.R-STG.R. (b) Connectogram shows increase (hot colors) and decrease (blue) in FC strength between auditory (AUD) and premotor-motor-somatosensory (PMS) regions. The thickness of links indicates the magnitude of the changes.
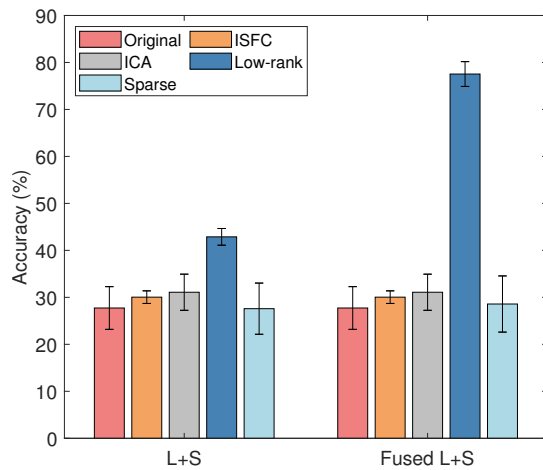

Fig. 9. Performance comparison of different time-resolved connectivity features from fMRI in predicting stimulus annotations for five categories of audio events present in a movie. The categories considered include *Noise*, *Music*, *Speech*, *Sp+Noise* and *Silence*. Results reported are accuracy scores of leave-one-subject-out classification based on fully-connected neural networks.
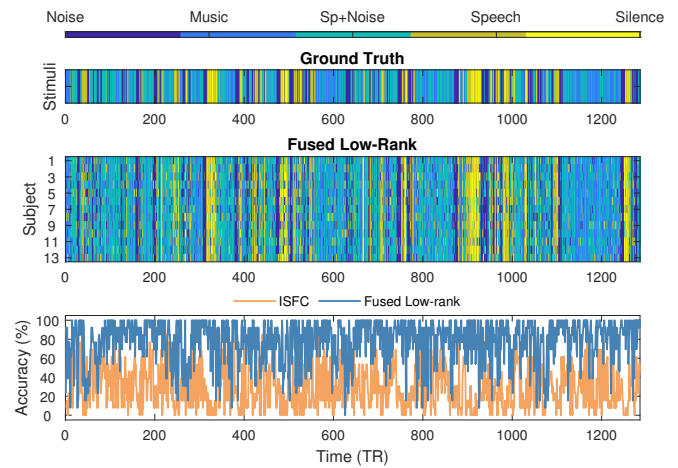

Fig. 10. Prediction of audio events in movie over time. (Top) Ground-truth stimulus annotation. (Middle) Decoded time courses using fused low-rank FC metrics for individual subjects. (Bottom) Prediction accuracy averaged over all subjects at each time point.

individual subjects, as complementary to the cross-subject low-rank structure in our method. Finally, the dynamic FC analysis can also be extended to investigate state-related changes in the network structure, e.g., dynamic community structure across subjects during naturalistic paradigms, by applying regime-switching models [34], [35] on the extracted time-resolved FC components.

REFERENCES

[1] S. Sonkusare, M. Breakspear and C. Guo, "Naturalistic stimuli in neuroscience: Critically acclaimed," *Trends Cognitive Sci.*, vol. 23, no. 8, pp. 699–714, 2019.

[2] J. Wang, et al., "Test–retest reliability of functional connectivity networks during naturalistic fmri paradigms," *Human Brain Mapp.*, vol. 38, no. 4, pp. 2226–2241, 2017.

[3] A. Bartels and S. Zeki, "Functional brain mapping during free viewing of natural scenes," *Human Brain Mapp.*, vol. 21, no. 2, pp. 75–85, 2004.

[4] U. Hasson, et al., "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, no. 5664, pp. 1634–1640, 2004.

[5] U. Hasson, R. Malach and D. J. Heeger, "Reliability of cortical activity during natural stimulation," *Trends Cognitive Sci.*, vol. 14, no. 1, pp. 40–48, 2010.

[6] I. P. Jääskeläinen, et al., "Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing," *Open Neuroimaging J.*, vol. 2, pp. 14–19, 2008.

[7] E. Simony, et al., "Dynamic reconfiguration of the default mode network during narrative comprehension," *Nature Comm.*, vol. 7, pp. 12141, 2016.

[8] T. A. Bolton, et al., "Brain dynamics in ASD during movie-watching show idiosyncratic functional integration and segregation," *Human Brain Mapp.*, vol. 39, no. 6, pp. 2391–2404, 2018.

[9] X. Di and B. B. Biswal, "Intersubject consistent dynamic connectivity during natural vision revealed by functional mri," *NeuroImage*, pp. 1–9, 2020.

[10] S. A. Nastase, et al., "Measuring shared responses across subjects using intersubject correlation," 2019.

[11] T. Bouwmans and E. H. Zahzah, "Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance," *Comput. Vision Image Understanding*, vol. 122, pp. 22–34, 2014.

[12] X. Liu, et al., "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2502–2514, 2015.

[13] J. Wright, et al., "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.

[14] E. J. Candès, et al., "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[15] P. Aggarwal and A. Gupta, "Low rank and sparsity constrained method for identifying overlapping functional brain networks," *PloS one*, vol. 13, no. 11, pp. 1–19, 2018.

[16] X. Jiang, et al., "Estimating functional connectivity networks via low-rank tensor approximation with applications to mci identification," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 7, pp. 1912–1920, 2020.

[17] V. Chandrasekaran, et al., "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optimization*, vol. 21, no. 2, pp. 572–596, 2011.

[18] Z. Lin, M. Chen and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[19] S. Boyd, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[20] X. Wang and X. Yuan, "The linearized alternating direction method of multipliers for dantzig selector," *SIAM J. Sci. Comput.*, vol. 34, no. 5, 2012.

[21] X. Li, et al., "Linearized alternating direction method of multipliers for sparse group and fused lasso models," *Comput. Stat. Data Anal.*, vol. 79, pp. 203–221, 2014.

[22] M. Li, et al., "The linearized alternating direction method of multipliers for low-rank and fused lasso matrix regression model," *J. Appl. Stat.*, pp. 1–18, 2020.

[23] J.-F. Cai, E. J. Candès and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimiz.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[24] O. Sporns and R. F. Betzel, "Modular brain networks," *Annu. Rev. Psychol.*, vol. 67, pp. 613–640, 2016.

[25] D. M. Pavlovic, et al., "Stochastic blockmodeling of the modules and core of the caenorhabditis elegans connectome," *PloS One*, vol. 9, no. 7, 2014.

[26] J. I. Skipper, J. T. Devlin and D. R. Lametti, "The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception," *Brain and language*, vol. 164, pp. 77–105, 2017.

[27] J. I. Skipper and J. D. Zevin, "Brain reorganization in anticipation of predictable words," *bioRxiv 101113*, 2017.

[28] N. Leonardi and D. Van De Ville, "On spurious and real fluctuations of dynamic functional connectivity during rest," *NeuroImage*, vol. 104, pp. 430–436, 2015.

[29] R. Tibshirani, et al., "Sparsity and smoothness via the fused lasso," *J. Roy. Statistical Soc: Series B*, vol. 67, no. 1, pp. 91–108, 2005.

[30] P. Adank, "The neural bases of difficult speech comprehension and speech production: Two activation likelihood estimation (ale) meta-analyses," *Brain and Language*, vol. 122, no. 1, pp. 42–54, 2012.

[31] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.

[32] D. Kong, et al., "L2rm: Low-rank linear regression models for high-dimensional matrix responses," *J. Amer. Statistical Assoc.*, vol. 115, no. 529, pp. 403–424, 2020.

[33] W. Hu, et al., "Nonparametric matrix response regression with application to brain imaging data analysis," *Biometrics*, pp. 1–14, 2020.

[34] C. M. Ting, et al., "Estimating dynamic connectivity states in fMRI using regime-switching factor models," *IEEE Trans. Med. Imaging*, vol. 37, no. 4, pp. 1011–1023, 2018.

[35] C. M. Ting, et al., "Detecting dynamic community structure in functional brain networks across individuals: A multilayer approach," *IEEE Trans. Med. Imaging*, vol. 40, no. 2, pp. 468–480, 2021.