NeuroImage 142 (2016) 55–66

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Improving effect size estimation and statistical power with multi-echo fMRI and its impact on understanding the neural systems supporting mentalizing

Michael V. Lombardo [a,b,*], Bonnie Auyeung [b,c], Rosemary J. Holt [b], Jack Waldman [b], Amber N.V. Ruigrok [b], Natasha Mooney [b], Edward T. Bullmore [d], Simon Baron-Cohen [b], Prantik Kundu [e,*]

[a] Center for Applied Neuroscience, Department of Psychology, University of Cyprus, Cyprus
[b] Autism Research Centre, Department of Psychiatry, University of Cambridge, UK
[c] Department of Psychology, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, UK
[d] Brain Mapping Unit, Department of Psychiatry, University of Cambridge, UK
[e] Section on Advanced Functional Neuroimaging, Departments of Radiology & Psychiatry, Icahn School of Medicine at Mount Sinai, USA

ARTICLE INFO

Article history:
Received 21 August 2015
Accepted 9 July 2016
Available online 11 July 2016

Keywords:
Multi-echo EPI
Statistical power
Denoising
Task-fMRI
Mentalizing
Cerebellum

ABSTRACT

Functional magnetic resonance imaging (fMRI) research is routinely criticized for being statistically underpowered due to characteristically small sample sizes and much larger sample sizes are being increasingly recommended. Additionally, various sources of artifact inherent in fMRI data can have detrimental impact on effect size estimates and statistical power. Here we show how specific removal of non-BOLD artifacts can improve effect size estimation and statistical power in task-fMRI contexts, with particular application to the social-cognitive domain of mentalizing/theory of mind. Non-BOLD variability identification and removal is achieved in a biophysical and statistically principled manner by combining multi-echo fMRI acquisition and independent components analysis (ME-ICA). Without smoothing, group-level effect size estimates on two different mentalizing tasks were enhanced by ME-ICA at a median rate of 24% in regions canonically associated with mentalizing, while much more substantial boosts (40–149%) were observed in non-canonical cerebellar areas. Effect size boosting occurs via reduction of non-BOLD noise at the subject-level and consequent reductions in between-subject variance at the group-level. Smoothing can attenuate ME-ICA-related effect size improvements in certain circumstances. Power simulations demonstrate that ME-ICA-related effect size enhancements enable much higher-powered studies at traditional sample sizes. Cerebellar effects observed after applying ME-ICA may be unobservable with conventional imaging at traditional sample sizes. Thus, ME-ICA allows for principled design-agnostic non-BOLD artifact removal that can substantially improve effect size estimates and statistical power in task-fMRI contexts. ME-ICA could mitigate some issues regarding statistical power in fMRI studies and enable novel discovery of aspects of brain organization that are currently under-appreciated and not well understood.

Introduction

A common criticism of neuroscience research in general (Button et al., 2013) and functional MRI (fMRI) in particular (Yarkoni, 2009), is that studies are characteristically statistically underpowered. Low statistical power by definition means that a study will have less of a chance for detecting true effects, but also means that observed statistically significant effects are less likely to be true and will be more susceptible to the biasing impact of questionable research practices (Button et al., 2013; Ioannidis, 2005). This problem is important given the emergent 'crisis of confidence' across many domains of science (e.g., psychology, neuroscience), stemming from low frequency of replication and the pervasive nature of questionable research practices (Button et al., 2013; Ioannidis, 2005; Simmons et al., 2011).

Low statistical power can be attributed to small sample sizes, small effect sizes, or a combination of both. The general recommended solution is primarily to increase sample size (though other secondary recommendations also include increased within-subject scan time). These recommendations are pragmatic mainly because these variables are within the control of the researcher during study design. While these recommendations are important to consider (Desmond and Glover, 2002; Friston, 2012; Lindquist et al., 2013; Mumford and Nichols, 2008; Yarkoni, 2009), other considerations such as dealing with substantial sources of non-BOLD noise inherent in fMRI data also need to be evaluated before the field assumes increasing sample size

* Corresponding authors.
E-mail addresses: mvlombardo@gmail.com (M.V. Lombardo), prantik.kundu@mssm.edu (P. Kundu).

http://dx.doi.org/10.1016/j.neuroimage.2016.07.022

or scan time to be the primary or only means of increasing statistical power. These considerations are especially poignant when mandates for large-N studies and increased within-subject scan time are practically limiting due to often cited reasons such as prohibitively high imaging costs for all but the most well-funded research groups or in situations where the focus is on studying sensitive, rare, and/or less prevalent patient populations and where increasing scan time is impractical (e.g., children, patients with neuropsychiatric conditions).

On the issue of non-BOLD noise variability, it is well known that fMRI data are of variable quality. Poor and variable quality data can significantly hamper ability to achieve accurate and reproducible representations of brain organization. The poor sensitivity of fMRI often arises from high levels of subject motion (often task correlated), cardiopulmonary physiology, or other types of imaging artifact (Murphy et al., 2013). These artifacts are problematic because they are often inadequately separable from the functional blood oxygenation level dependent (BOLD) signal when using conventional fMRI methods. Given an advance in fMRI methodology that allows enhanced detection and removal of these artifacts, the situation regarding statistical power and sample size may change markedly. Such advances could create viable experimental alternatives or supplements to the recommendation for increasing sample size/scan time to boost statistical power, and concurrently make for an fMRI approach that can more reliably enable discovery of subtle but potentially key aspects of typical and atypical brain function.

In this study, we address problems related to statistical power through specific targeting of the problems related to non-BOLD artifact variability. We have applied an approach that integrates fMRI data acquisition using multi-echo EPI with the decomposition method of independent components analysis (ICA), towards more principled removal of non-BOLD signals from fMRI data without arbitrary temporal or spatial smoothing (e.g. bandpass filtering). The integration of these techniques is implemented in a pipeline called multi-echo independent components analysis or ME-ICA (Kundu et al., 2012). ME-ICA utilizes multi-echo fMRI to acquire both fMRI signal time series and their NMR signal decay, towards distinguishing functional BOLD from non-BOLD signal components based on relaxometry of their respective and differentiable signatures in the decay domain. BOLD and non-BOLD signals are markedly differentiable in data analysis of the echo time (TE) domain and independent of gross similarity in the spatial and temporal domains. BOLD-related signals specifically show linear dependence of percent signal change on TE, whereas non-BOLD signal amplitudes demonstrate TE-independence. This suggests ME-ICA could be a principled bottom-up approach towards identifying and retaining BOLD-related variability while systematically removing non-BOLD variation. ME-ICA has been successfully applied to fMRI resting state connectivity acquisition and analysis and has been shown to enable improvements regarding increased temporal signal-to-noise ratio (tSNR) and enhanced ability to remove motion and other artifactual sources of variability. ME-ICA has also been used to improve seed-based connectivity analysis, enhance specificity, and holds translational potential for use at high-field strength and within animal models (Kundu et al., 2013, 2014, 2015). ME-ICA can also be applied alongside multi-band acquisition (Olafsson et al., 2015) and has recently been applied to identify ultra-slow temporally-extended task-related responses (Evans et al., 2015). However, one very necessary yet unexamined niche within the space of uses for fMRI is within the highly utilized context of traditional task-based fMRI studies and the potential impact that ME-ICA innovations could have on effect size estimation and consequently statistical power.

Here we conduct the first assessment of how ME-ICA performs with regard to effect size estimation and statistical power in task-related activation mapping settings with block-designs. ME-ICA can be flexibly applied to both task- and resting state fMRI contexts. This unified approach is advantageous since conventional resting state and task data processing and denoising use disjoint pipelines that often may require different assumptions, analytical premises and skillsets. However, it is important and currently not well understood if and to what extent generalized non-BOLD removal as targeted by ME-ICA enhances the elucidation of task effects, compared to current task activation analysis with inline denoising based on linear artifact models and arbitrary filtering done in a study specific manner. In this study we specifically examined how ME-ICA performs against a conventional task-based imaging analysis pipeline involving baseline regression of motion parameters (i.e. TSOC+MotReg) and another prominent yet more recent task-based denoising procedure, GLMdenoise (Kay et al., 2013). We utilized two separate tasks (i.e. the 'SelfOther' and 'Stories' tasks) assessing neural systems supporting the social-cognitive domain or mentalizing and theory of mind and highlight the effects of ME-ICA in terms of effect size estimation and statistical power. We also evaluate the impact of the method on two sets of brain regions; 'canonical' regions typically highlighted as important in the neural systems for mentalizing (Frith and Frith, 2003; Lombardo et al., 2010b; Saxe and Powell, 2006; Schaafsma et al., 2015; Schurz et al., 2014; Spunt and Adolphs, 2014; van Overwalle, 2009) and 'non-canonical' regions in the cerebellum (van Overwalle et al., 2014).

Materials and methods

Participants

This study was approved by the Essex 1 National Research Ethics committee. Parents gave informed consent for their child to participate and each child also gave assent to participate. Participants were 69 adolescents (34 males, 35 females, mean age = 15.45 years, sd age = 0.99 years, range = 13.22–17.18 years) sampled from a larger cohort of individuals whose mothers underwent amniocentesis during pregnancy for clinical reasons (i.e. screening for chromosomal abnormalities). The main focus for sampling from this cohort was to study the fetal programming effects of steroid hormones on adolescent brain and behavioral development. At amniocentesis, none of the individuals screened positive for any chromosomal abnormalities and were thus considered typically developing. Upon recruitment for this particular study, we additionally checked for any self- or parent-reported neuropsychiatric conditions. One individual had a diagnosis on the autism spectrum. The remaining participants did not have any other kind of neurological or psychiatric diagnosis. Analyses were done on the full sample of 69 individuals, as analyses leaving out the one patient with an autism diagnosis did not change any of the results.

Task design

Participants were scanned using two block-design fMRI paradigms. The first paradigm, which we call the 'SelfOther' task, is a 2 × 2 within-subjects factorial design which contains two contrasts that tapped either self-referential cognition and mentalizing and was similar in nature to previously published studies (Lombardo et al., 2010a, 2010b, 2011). Briefly, participants are asked to make reflective judgments about either themselves or the British Queen that varied as either a mentalistic (e.g., "How likely are [you/the Queen] to think that it is important to keep a journal?") or physical judgment (e.g., "How likely are [you/the Queen] to have bony elbows?"). Participants make judgments on a 1–4 scale, where 1 indicated 'not at all likely' and 4 indicated 'very likely'. All stimuli are taken from Jason Mitchell's lab and have been used in prior studies on mentalizing and self-referential cognition (Jenkins et al., 2008; Mitchell et al., 2006). The SelfOther task is presented in 2 scanning runs (8:42 duration per run; 261 volumes per run). Within each scanning run there are 4 blocks per condition, and within each block there are 4 trials of 4 s duration each. Task blocks are separated from each other by a 16 s fixation block. The first 5 volumes of each run are discarded to allow for equilibration effects.

The second paradigm, which we call the 'Stories' task, is also a block-design and contains two contrasts tapping mentalizing and language