発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
橋本謙二	統合失調症の認知機能障害改善薬の最新知見	Schizophrenia Frontier	13(1)	7-10	2012
橋本謙二	うつ病の病態におけるBDNF-TrkB受容体シグナル系の役割	医学のあゆみ	244(5)	471-475	2013
橋本謙二	統合失調症のNMDA受容体機能低下仮説に基づいた新規治療薬の開発状況	臨床精神医学	42(7)	927-934	2013
橋本謙二	うつ病の病態におけるグルタミン酸神経系の役割と新規治療薬の開発	日本生物学的精神医学会誌	24(3)	153-156	2013
高橋 努, 鈴木道雄	統合失調症圏のMRI研究の進歩	精神神経学雑誌	115	874-879	2013
鈴木道雄, 高橋 努	統合失調症と脳の形態変化	日本臨床	71	619-623	2013
住吉太幹, 西山志満子, 樋口悠子, 高橋 努, 松岡 理, 倉知正佳, 水上祐子, 数川 悟, 鈴木道雄	富山県における早期介入活動の実際と工夫	精神神経学雑誌	115	180-186	2013
川﨑康弘, 鈴木道雄	統合失調症を脳画像で診断するためのVBM	日本磁気共鳴医学会雑誌	32	41-47	2012
高橋 努, 中村主計, 鈴木道雄	アットリスク精神状態のMRI研究	臨床精神医学	41	1421-1426	2012
中坪太久郎, 松岡理, 古市厚志, 今村理佐, 荒井宏文, 藪田歩, 松井三枝, 鈴木道雄, 倉知正佳	統合失調症の認知機能障害に関するセッションを取り入れた家族心理教育の試み	精神療法	37	89-95	2011

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
兼田康宏，上岡義典，住吉太幹，古郡規雄，伊東　徹，樋口悠子，鈴木道雄，大森哲郎	統合失調症認知評価尺度日本語版を用いたco-primaryの測定	日本神経精神薬理学雑誌	31	259-262	2011
中村主計，高橋　努，鈴木道雄	早期統合失調症と脳の形態変化	精神科治療学	26	1421-1426	2011
鈴木道雄	統合失調症の早期介入と脳画像診断	日本精神科病院協会雑誌	29（別冊）	35-40	2011
鈴木道雄，高橋　努，川崎康弘，中村主計，高柳陽一郎	統合失調症における脳の構造画像マーカー	精神科	18	506-512	2011
高橋 努，鈴木道雄	早期精神病における脳形態変化	日本生物学的精神医学会誌	22	15-20	2011
高橋 努，鈴木道雄	特集「精神科領域における画像診断の展望」統合失調症圏のMRI研究	最新精神医学	16	269-273	2011
山田麻紀，岡崎光俊，斉藤聖，二村園恵，長房裕子，野田隆政	躁病エピソードを呈し炭酸リチウムが効果的であったてんかん患者の2症例	精神科治療学	26(2)	227-233	2011
野田隆政，安藤久美子，平林直次，大森まゆ，黒木規臣，岡田幸之	医療観察法におけるECTのインフォームド・コンセント（I.C.）の再検討	精神神経学雑誌（総会特別号）	107th	SS385-SS390	2012
富岡大，川崎真護，岩波明，野田隆政，兼子幸一，朴盛弘，三村將，中込和幸	うつ病患者のNIRSによる治療反応性と疾患鑑別への有用性 - 多施設における2時点検査の結果と診断変更症例の検討	MEDIX	58	4-9	2013

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
山田麻紀，亀井雄一，野田隆政，有馬邦正	高照度光療法により睡眠覚醒リズムと抑うつ状態が改善した概日リズム睡眠障害の1例	精神神経学雑誌	115(8)	901	2013
福田正人，吉田寿美子，杉村有司，小川勝，大渓俊幸，樋口智江，内山智恵，安井臣子	光トポグラフィー検査（NIRS）による脳機能測定	検査と技術	40	182-188	2012
鈴木道雄，川﨑康弘，高柳陽一郎，中村主計，高橋務	構造MRIによる統合失調症の補助診断の可能性（特集 当事者に届く生物学的精神医学研究 ： バイオマーカーを用いた精神疾患の客観的補助診断法の開発）	精神神経学雑誌	114(7)	807-811	2012

Ⅳ．　研究成果の刊行物・別刷

NeuroImage 85 (2014) 498–507

Contents lists available at ScienceDirect

NeuroImage

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

Neuroimaging-aided differential diagnosis of the depressive state

Ryu Takizawa [a,1], Masato Fukuda [b,*,1], Shingo Kawasaki [c], Kiyoto Kasai [a], Masaru Mimura [d,e], Shenghong Pu [f], Takamasa Noda [g], Shin-ichi Niwa [h], Yuji Okazaki [i,j]
on behalf of the Joint Project for Psychiatric Application of Near-Infrared Spectroscopy (JPSY-NIRS) Group

[a] Department of Neuropsychiatry, The University of Tokyo Graduate School of Medicine, Tokyo 113-8655, Japan
[b] Department of Psychiatry and Neuroscience, Gunma University Graduate School of Medicine, Gunma 371-8511, Japan
[c] Hitachi Medical Corporation, Tokyo 101-0021, Japan
[d] Department of Neuropsychiatry, Showa University School of Medicine, Tokyo 157-8577, Japan
[e] Department of Psychiatry, Keio University School of Medicine, Tokyo 160-8582, Japan
[f] Division of Neuropsychiatry, Department of Multidisciplinary Internal Medicine, Tottori University Faculty of Medicine, Tottori 683-8504, Japan
[g] National Centre of Neurology and Psychiatry Hospital, Tokyo 187-8551, Japan
[h] Department of Neuropsychiatry, Fukushima Medical University School of Medicine, Fukushima 960-1295, Japan
[i] Department of Psychiatry, Mie University Graduate School of Medicine, Mie 514-8507, Japan
[j] Tokyo Metropolitan Matsuzawa Hospital, Tokyo 156-0057, Japan

ARTICLE INFO

Article history:
Accepted 31 May 2013
Available online 10 June 2013

Keywords:
Neuroimaging
Near-infrared spectroscopy (NIRS)
Differential diagnosis
Depressive state
Psychiatric disorder

ABSTRACT

A serious problem in psychiatric practice is the lack of specific, objective biomarker-based assessments to guide diagnosis and treatment. The use of such biomarkers could assist clinicians in establishing differential diagnosis, which may improve specific individualised treatment. This multi-site study sought to develop a clinically suitable neuroimaging-guided diagnostic support system for differential diagnosis at the single-subject level among multiple psychiatric disorders with depressive symptoms using near-infrared spectroscopy, which is a compact and portable neuroimaging method. We conducted a multi-site, case–control replication study using two cohorts, which included seven hospitals in Japan. The study included 673 patients (women/men: 315/358) with psychiatric disorders (major depressive disorder, bipolar disorder, or schizophrenia) who manifested depressive symptoms, and 1007 healthy volunteers (530/477). We measured the accuracy of the single-subject classification in differential diagnosis among major psychiatric disorders, based on spatiotemporal characteristics of fronto-temporal cortical haemodynamic response patterns induced by a brief (<3 min) verbal fluency task. Data from the initial site were used to determine an optimal threshold, based on receiver-operator characteristics analysis, and to generate the simplest and most significant algorithm, which was validated using data from the remaining six sites. The frontal haemodynamic patterns detected by the near-infrared spectroscopy method accurately distinguished between patients with major depressive disorder (74.6%) and those with the two other disorders (85.5%; bipolar disorder or schizophrenia) that presented with depressive symptoms. These results suggest that neuroimaging-guided differential diagnosis of major psychiatric disorders developed using the near-infrared spectroscopy method can be a promising biomarker that should aid in personalised care in real clinical settings. Potential confounding effects of clinical (e.g., age, sex) and systemic (e.g., autonomic nervous system indices) variables on brain signals will need to be clarified to improve classification accuracy.

Introduction

Among non-communicable diseases, neuropsychiatric conditions, including depression, contribute most significantly to overall disability-adjusted life years (DALYs), surpassing both cardiovascular disease and cancer (Mathers and Loncar, 2006; Prince et al., 2007). Therefore, early and accurate diagnosis and treatment are critical in psychiatric disorders, for which the development of specific biomarkers is of special importance. Currently, however, the diagnostic process in psychiatry is mainly based on patients' reports of symptoms, observed behaviours and disease course. Overcoming the limitations of relying on clinical interviews alone for the diagnosis of psychiatric disorders has been a great challenge.

To complicate this issue further, the manifestation of only a major depressive episode hampers the reliable differentiation of major depressive disorder (MDD) from bipolar disorder (BP) or

* Corresponding author at: Department of Psychiatry and Neuroscience, Gunma University Graduate School of Medicine, 3-39-22 Showa-machi, Maebashi, Gunma 371-8511, Japan. Fax: +81 27 220 8187.
E-mail address: fkdpsy@med.gunma-u.ac.jp (M. Fukuda).
[1] These authors contributed equally to this paper.

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.05.126

schizophrenia (SZ) based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria alone (Zimmermann et al., 2009). Although many clinical symptoms are common to various psychiatric disorders, depressive symptoms are particularly ubiquitous in the disease process or clinical staging of various psychiatric disorders (Hafner et al., 2005). For instance, differentiation between BP presenting with depressive symptoms and unipolar MDD is a topical issue (Akiskal et al., 1995). Indeed, most patients with BP with depressive symptoms are initially diagnosed with and treated for MDD (Akiskal et al., 1995; Goldberg et al., 2001). Therefore, biomarkers that can facilitate early and accurate differentiation of BP with depressive symptoms from MDD are necessary.

In addition, depressive symptoms that fulfil the operational diagnostic criteria for a depressive episode/major depression can also occur at any stage of SZ and can contribute substantially to its associated morbidity and even mortality (an der Heiden et al., 2005). The differentiation of SZ from MDD, especially in the early stages, is also important because patients with SZ also exhibit non-psychotic and non-specific prodromal symptoms (e.g., depressive or negative symptoms and cognitive deficits) for several years before the onset of full-blown psychosis (McGorry et al., 2008). Therefore, the availability of clinically useful and cost-effective biomarkers for the differential diagnosis of major psychiatric disorders would likely enhance patient management, improve treatment/therapeutic response and lead to targeted therapies tailored to the individual (Holsboer, 2008). Despite their potential, to date, no such biomarkers have been established.

Functional imaging studies are one source of potential biomarkers (Gur et al., 2007; Phillips and Vieta, 2007); these studies have previously elucidated subtle brain abnormalities in patients with major psychiatric disorders relative to healthy control (HC) individuals and have been applied to the differential diagnosis of psychiatric disorders (e.g., to differentiate MDD from SZ, Barch et al., 2003 or BP, Almeida et al., 2009). However, some functional neuroimaging techniques are limited by the fact that, during the procedure, the individuals need to be placed in an uncomfortable or unnatural setting (e.g., lying in a supine position in a narrow gantry with the head fixed during the entire examination), for accurate measurement.

In contrast, multi-channel near-infrared spectroscopy (NIRS) using near-infrared light provides a completely non-invasive measurement of the spatiotemporal characteristics of brain function in ordinary clinical settings and allows patients to be comfortably seated in a well-lit room; therefore, it is considered a method for 'real-world neuroimaging'. Additionally, NIRS has relatively low maintenance costs and does not involve ionising radiation or objectionable noise; thus, it can be repeated as needed even for patients with psychiatric disorders. The utility and limitations of NIRS have been discussed extensively in previous reports (Ferrari and Quaresima, 2012; Obrig and Villringer, 2003; Strangman et al., 2002a). NIRS allows the measurement of haemoglobin concentration changes (1) only in the cortical surface area located immediately beneath the probes, but not in deeper brain structures, and (2) with limited spatial resolution, although it has a high temporal resolution. In NIRS, typical cortical activation represents not only decreased concentration of deoxy-haemoglobin ([deoxy-Hb]), which is considered the main source of blood oxygenation level-dependent (BOLD) contrast increase in functional magnetic resonance imaging (fMRI), but also a relatively larger increase in oxy-haemoglobin concentration ([oxy-Hb]) (Fig. 1).

The verbal fluency task (VFT) is a cognitive task that is used as a neuropsychological test or a neuroimaging task. The VFT elicits different abnormalities relevant to each diagnostic group of major psychiatric disorders (Curtis et al., 2001; Zanelli et al., 2010). We previously developed a very brief (<3 min) VFT and used it to investigate the differential fronto-temporal haemodynamic pattern between MDD and SZ (Suto et al., 2004) or MDD and BP (Kameyama et al., 2006), as well as the relationship between NIRS signals and functional impairment in SZ (Takizawa et al., 2008). We also found functional NIRS abnormalities in individuals at ultra-high risk for SZ and patients with first-episode psychosis (Koike et al., 2011). However, the clinical applicability of NIRS to the differential diagnosis of individuals remains uncertain. In this study, we extended our translational approach to replicate our previous findings (Kameyama et al., 2006; Suto et al., 2004) in a seven-site collaborative study using a large, fully independent sample set, and to evaluate the application of NIRS to psychiatric differential diagnosis in natural clinical settings.

Specifically, we used NIRS with wide coverage of the prefrontal and temporal cortices to investigate whether the frontal and temporal brain haemodynamic responses induced by cognitive activation could serve as biomarkers of underlying major psychiatric disorders with depression. To validate the reproducibility and generalisability of the results, we applied an algorithm developed using the data generated at the initial site to the test data derived from the remaining 6 sites. We hypothesised that the spatiotemporal characteristics of the haemodynamic responses detected by NIRS would not only differentiate patients with psychiatric disorders from HCs with acceptable sensitivity and specificity, but would also differentiate correctly and with a high concordance rate patients with MDD from patients with bipolar disorder and schizophrenia who present with depressive symptoms.

Material and methods

Participants

This multi-site study was performed in 7 hospitals: 6 were affiliated with universities (Fukushima, Gunma, Mie, Tokyo, Showa, and Tottori) and one was affiliated with the National Centre of Neurology and Psychiatry of Japan. The sites were situated in the Tokyo metropolitan area and in moderate-scale prefectural capital cities (Fukushima, Maebashi, Tsu and Yonago). The participants were recruited from June 2004 to June 2009, with the exception of recruitment at the initial site (Gunma University Hospital in Maebashi City), which was conducted over 6 years (March 2003 to March 2009). The ethics committees of the participating hospitals approved this collaborative study. In accordance with the Declaration of Helsinki, all participants gave written informed consent after receiving a complete explanation of the study.

Six hundred and seventy-three in-patients and out-patients with psychiatric disorders (MDD, BP and SZ), in addition to 1007 HC volunteers (Flow diagram (1)), were initially enrolled. Of note, these individuals were not the same as those included in our previous studies (Kameyama et al., 2006; Suto et al., 2004). The patients were diagnosed by experienced psychiatrists based on the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) (First et al., 1997). The HC volunteers were hospital staff members, university students and members of the general population who responded to website or newspaper advertisements in each city. The SCID non-patient edition was used to screen HC individuals. The exclusion criteria of the initial enrolment were neurological illness, traumatic brain injury with any known cognitive consequences and alcohol/substance abuse or addiction. All participants were native Japanese speakers who were capable of performing a Japanese version of the VFT easily.

On the day of NIRS measurement, the depressive symptoms of participants were evaluated using the 17-item Hamilton Rating Scale for Depression (HAMD) (Hamilton, 1960) and their psychotic and manic symptoms were evaluated using both the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1991) and the Young Mania Rating Scale (YMRS) (Young et al., 1978), respectively, by well-trained psychiatrists with no knowledge of the NIRS data. During the study, all patients with psychiatric disorders were medicated with one or more agents (anti-psychotics, anti-depressants, anxiolytics and/or anti-parkinsonian agents), with the exception of 10 drug-free patients with MDD and 5 drug-free patients with SZ.

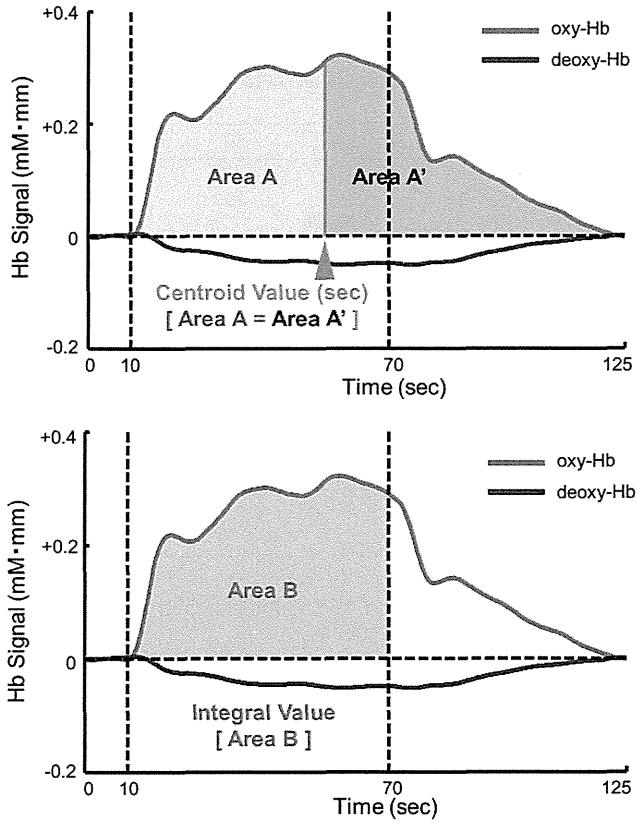


Fig. 1. Typical time-course pattern of near-infrared spectroscopy (NIRS) signals coupled with the verbal fluency task. The 'centroid value' is indicated by the time [s], which is indicated by a perpendicular line from the centroid of an NIRS signal-change area (calculated with positive change) throughout all the task periods. Oxy-Hb: oxygenated haemoglobin signal; deoxy-Hb: deoxygenated haemoglobin signal.

To minimise the influence of confounding factors, we performed group matching for age and gender among the 4 diagnostic groups using one-way analysis of variance (ANOVA) and a chi-squared test, which excluded randomly selected individuals and brought the mean age of the HC individuals and patients with MDD or SZ in closer alignment with that of the patients with BP (44.0 ± 14.9 years old (y.o.)), which was the group with the fewest individuals (Table 1 and Flow diagram (2)). For confirmation, we also analysed demographically non-matched samples that are identified in the Supplementary Material (I). The overall results were the same as those described in the main manuscript and the reduction in the total number of study participants after demographical matching did not appear to have an influence on the development of the algorithm (see Supplementary Material (I)).

Because our clinically valuable target were help-seeking unremitted patients, subsequently we excluded study participants with extremely mild symptoms (HAMD score ≤ 5, PANSS depression item score ≤ 1, PANSS negative symptom score ≤ 11, PANSS general psychopathology score ≤ 21, or PANSS positive symptom score − negative symptom score ≤ 11; the latter 3 criteria were based on the criteria from the PANSS manual for the 5th percentile of patients with mild SZ, Kay et al., 1991). We also excluded patients in a manic phase (YMRS score > 10) from the NIRS measurement; rather, we focused on patients with BP who were in the depressive phase because the different phases may produce different brain dysfunctions in patients with BP (Phillips and Vieta, 2007), and manic patients with BP were diagnosed without apparent difficulty (Flow diagram (3)).

Activation task

The activation task used in this study was similar to that used in our previous studies (Kameyama et al., 2006; Suto et al., 2004; Takizawa et al., 2008). Briefly, a VFT (letter version) was administered and NIRS signal ([oxy-Hb] and [deoxy-Hb]) changes were measured during a 10 s pre-task baseline period, a 60 s activation period and a 55 s post-task baseline period. During the activation period, the participants were instructed to utter as many Japanese words beginning with a designated syllable as possible. For the pre- and post-task baseline periods, the individuals were instructed to simply repeat Japanese vowels out loud. The total number of correct words generated during the 60 s activation period was used as the measure of task performance (Table 1).

Among the many neuropsychological tasks used for detecting neurocognitive deficits in patients with major psychiatric disorders (Barrett et al., 2009; Zanelli et al., 2010), we selected the VFT because it is an executive task that exhibits distinct differences in performance and neuroimaging data among each diagnostic group of major psychiatric disorders (Costafreda et al., 2006; Curtis et al., 2001; Zanelli et al., 2010). In addition, the VFT is easy to understand and execute; in fact, all participants generated more than one word during the VFT. Therefore, this task is suitable for translational research aimed at identifying practical biomarkers.

NIRS measurement

The NIRS apparatus and measurement procedure were described in full previously (Takizawa et al., 2008). Briefly, we used a 52-channel NIRS system (ETG-4000; Hitachi Medical Co., Tokyo, Japan). The preparation of the apparatus, including the audiovisual on-screen instructions, usually took less than 7 min and our brief version of the VFT took less than 3 min, which is less demanding for participants (10–15 min is necessary for the entire procedure).

NIRS is based on the principle that near-infrared light is preferentially absorbed by haemoglobin and less so by other tissues. Near-infrared light emitted from the skin travels into the body, is reflected and absorbed by the internal tissues and reappears on the skin. Thus, the absorption of near-infrared light reflects haemoglobin concentration ([Hb]) in the tissue placed beneath emission and detection probe pairs. Measurements taken using 2 or more wavelengths of near-infrared light enable the determination of [oxy-Hb] and [deoxy-Hb] changes because their absorptions are different at different wavelengths. The ETG-4000 system measures relative changes in [oxy-Hb] and [deoxy-Hb] using 2 wavelengths (695 and 830 nm) of infrared light, based on the modified Beer–Lambert law.(Yamashita et al., 1996) In this continuous-wave NIRS system, these [Hb] values include a differential pathlength factor (DPF); therefore, the unit of this form of NIRS measurement is mM·mm. The distance between pairs of source-detector probes was set to 3.0 cm and each measurement area located between pairs of source-detector probes was defined as one 'channel'. It is assumed that a machine in which the source-detector spacing is 3.0 cm measures points at a depth of 2–3 cm from the scalp (i.e., the surface of the cerebral cortex) (Okada and Delpy, 2003). The temporal resolution of NIRS was set to 0.1 s.

The arrangement of the probes measured relative [oxy-Hb] and [deoxy-Hb] signal changes in the bilateral prefrontal cortical area (i.e., dorso-lateral [Brodmann areas (BAs) 9 and 46], ventro-lateral [BAs 44, 45, and 47] and fronto-polar [BA 10] regions) and in the superior and middle temporal cortical surface regions, which was corroborated by a multi-individual study of anatomical cranio-cerebral correction via the international 10–20 system (Fig. 2 and Table S1) (Tsuzuki et al., 2007). However, in the 10–20 system, the anterior parts of the probes (e.g., Fpz) can be positioned precisely, whereas the position errors of more lateral probes might be increased due to inter-individual head size variability. In addition, although we initially aimed to analyse single-individual and single-channel levels in this study, studies of repeated NIRS measures have demonstrated acceptable reliability of the NIRS signal at the group and cluster levels, whereas retest reliability was unsatisfactory at the single-individual and single-channel levels (Schecklmann et al., 2008).

Table 1
Demographic and clinical characteristics of the 4 age- and gender-matched diagnostic groups at all 7 study sites.

	Major depressive disorder		Schizophrenia		Bipolar disorder		Healthy control		Group difference
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*p*-Value
n	153		136		134		590		
Age years	43.8	12.7	43.7	12.1	44.0	14.9	43.9	15.7	0.99
Gender, women/men	77/76		67/69		69/65		314/276		0.81[a]
Education, years	14.0	1.9	15.2	2.0	15.6	2.0	16.1	2.4	<0.01
Estimated premorbid IQ	106.0	10.1	103.7	11.2	106.9	8.6	107.2	10.1	0.23
Task performance	13.0	3.8	13.6	4.4	12.0	3.6	15.3	4.8	<0.01
Age at onset, years	39.2	11.3	23.4	7.4	32.9	12.4	–		
PANSS									
Positive	–		16.3	5.0	–		–		
Negative	–		21.0	6.0	–		–		
General psychopathology	–		37.0	7.6	–		–		
HAM-D	14.1	6.7	–		8.4	7.0	–		
YMRS	–		–		4.7	5.9	–		
GAF	53.9	9.7	47.3	11.4	55.5	13.3	–		

Abbreviations: IQ, intelligence quotient; PANSS, Positive and Negative Syndrome Scale; GAF, global assessment of functioning.
[a] Chi-square test was used for testing group difference. Otherwise, one-way ANOVA was used.

Therefore, instead of undertaking a full analysis at the single-individual and single-channel levels, here we performed an analysis of NIRS signals at the single-individual and cluster levels. A principal component analysis (PCA) of NIRS [oxy-Hb] signal changes in targeted fronto-temporal channels was performed at the initial study site as a preliminary analysis to capture a channel cluster of the analogous time-course pattern in HC individuals. Subsequently, the weight maps of the first and second principal component graphs were used to identify 2 cluster components.

These analyses suggested that 2 cluster components were identified and that the 2 clusters included the frontal region (11 channels) and the bilateral temporal region (10 channels each) (see Supplementary Material (II) and eFig. S1). The channels in these 2 respective regions of interest were averaged and transformed into representative 'Region 1 (R1)' and 'Region 2 (R2)' NIRS signals for each individual (Fig. 3). According to registration into the LONI Probabilistic Brain Atlas 40 (LBPA40) (Fig. 2) (see Supplementary Table S1 for LBPA40 anatomical labels) (Shattuck et al., 2008), the R1 NIRS signal consisted of signals from channels located approximately in the fronto-polar and dorsolateral prefrontal cortical regions (i.e., superior and middle frontal gyri), whereas the R2 NIRS signal consisted of signals from channels located approximately in the ventro-lateral prefrontal cortex and the superior and middle temporal cortical regions (i.e., inferior frontal gyrus and superior and middle temporal gyri).

An automatic artefact-rejection procedure (see Supplementary Material (III)) was followed and individual data were excluded when there were fewer than 6 remaining channels from each of the 2 cluster regions (Flow diagram (4)).

Statistical analyses

Taking into consideration the potential application of the technique in ordinary clinical settings and personalised care, a conservative receiver operating characteristic (ROC) analysis was performed and used to generate simple indices of NIRS signal patterns, to aid individual diagnoses.

The spatiotemporal characteristics of the frontal and temporal haemodynamic responses induced by VFT were assessed and subsequently applied to an algorithm using the simplest and fewest variables for differential diagnosis. Because previous studies have shown that the best way to differentiate patients with MDD from those with BP or SZ is to describe the time-course of changes in the NIRS signal associated with the VFT (Kameyama et al., 2006; Suto et al., 2004), we chose to create 2 simple visual indices, referred to here as 'integral value' and 'centroid value', to capture these time-course changes.

The integral value describes the size of the haemodynamic response during the 60 s activation task period, whereas the centroid value serves as an index of time-course changes throughout the task, with periods representing the timing of the haemodynamic response. The centroid value is indicated by a time shown with a perpendicular line from the centroid of the NIRS signal change area (calculated as a positive change) throughout the task periods (from 0 [s] to 125 [s]

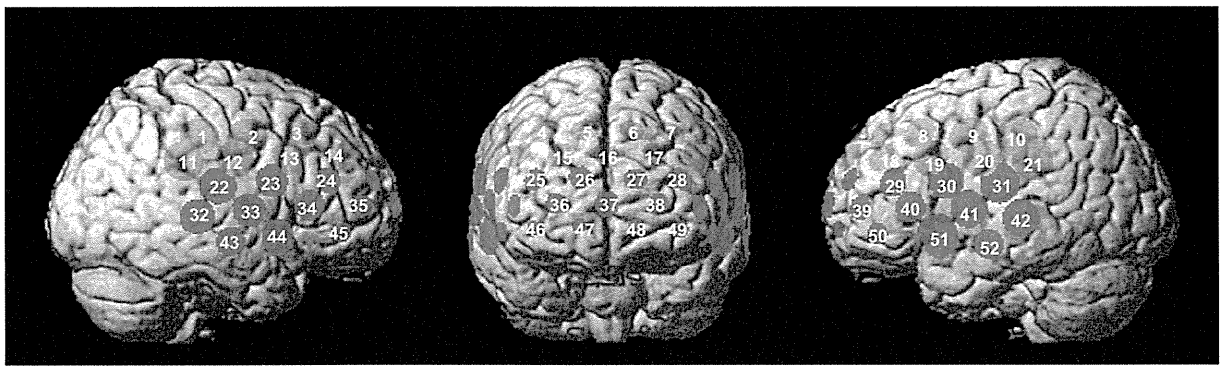


Fig. 2. Regions of interest (Regions 1 and 2) of the near-infrared spectroscopy (NIRS) signals. The locations of near-infrared spectroscopy (NIRS) measurements were probabilistically estimated and anatomically labelled in the standard brain space (LBPA40) according to Tsuzuki et al. (2007). Region 1: (ch 25–28, ch 36–38 and ch 46–49); Region 2, Right: (ch 22–24, ch 32–35 and ch 43–45); Left: (ch 29–31, ch 39–42 and ch 50–52).

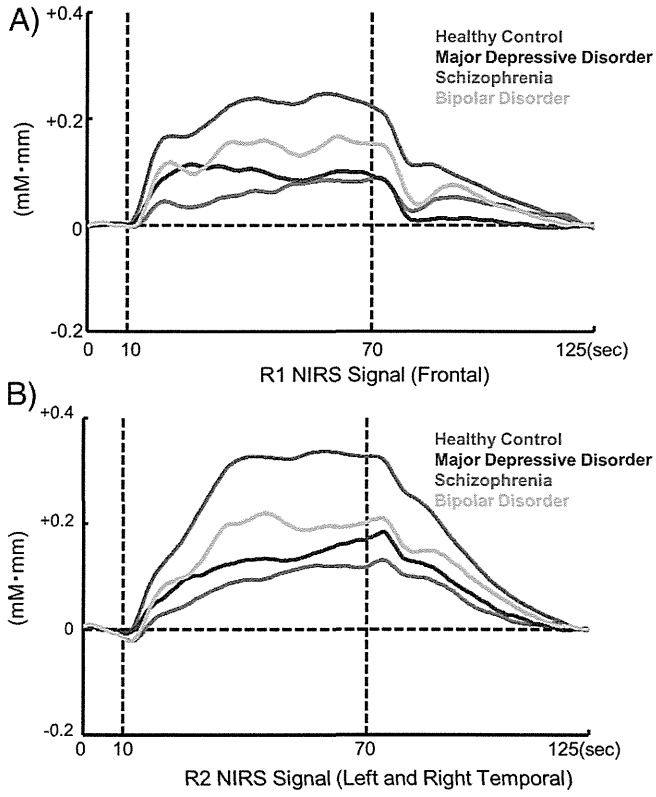


Fig. 3. Time courses of the haemodynamic responses in Region 1 (R1) and Region 2 (R2) in the 4 diagnostic groups. Panels A and B show the time courses of the haemodynamic responses in R1 and R2, respectively.

(= 10 [s] + 60 [s] + 55 [s])) (see Fig. 1). To confirm the reproducibility of each single index between the 2 measurements, a test–retest analysis (single-measure intra-class correlation (ICC) analysis using a one-way random effect model) revealed the presence of significant intra-class correlation coefficients for both the R1 and R2 integral values [r = 0.47, $p = 0.01$; r = 0.59, $p < 0.01$, respectively] and the R1 centroid value [r = 0.65, $p < 0.01$], but not for the R2 centroid value [r = 0.20, $p = 0.19$] (see Supplementary material (IV)). PCA and ICC analyses revealed that the 2 indices of NIRS analysis during VFT were acceptable at the single-individual and cluster levels. Thus, the 3 significant variables were used for further analysis.

The 2 representative R1 and R2 NIRS signals obtained from each individual were averaged separately for each type of [Hb] and the integral and centroid values were calculated using parametric statistical tests. Further analyses focused on the increases in [oxy-Hb], because these appear to reflect task-related cortical activation more directly than do decreases in [deoxy-Hb], as evidenced by the stronger correlation between the former and the blood oxygenation level-dependent signal measured by fMRI (Strangman et al., 2002b) and by the results of animal studies (Hoshi et al., 2001). As the typical [oxy-Hb] activation pattern had a positive direction (Fig. 1), data with positive [oxy-Hb] changes (i.e., data with an integral value ≥ 0) in R1 and R2 were used to create an algorithm (Flow diagram (4)). Data exhibiting negative [oxy-Hb] changes were added to the analysis and were described in the results as being appropriate. The analysis of [deoxy-Hb] changes was reported in Supplementary Material (V); however, no significant variable was found regarding [deoxy-Hb] changes.

First, as a preliminary analysis to identify the variable that differentiates patients with psychiatric diseases from HCs most robustly, the 3 variables, including both integral and centroid values of the R1 NIRS signal and the integral value of the R2 NIRS signal, were compared among all of the patients and the age- and gender-matched controls at the initial study site using ANOVA. The resulting significant variables were applied to ROC analyses at the remaining 6 sites.

Because mental health professionals in real clinical settings must differentiate patients with MDD from those with BP or SZ presenting with depression as accurately as possible, the second main analysis performed here aimed to determine the most informative variable and the optimal threshold to discriminate patients with MDD from those with non-MDD disorders. In the present study, the 3 variables, including both integral and centroid indices of R1 and the integral R2 index of the NIRS signal, were compared among patients with MDD and those with either of the other 2 disorders using ANOVA; the variables that were deemed to be significant were applied to the ROC analysis. The preliminary data from the initial site were used to determine an optimal threshold, which was then validated using the test data from the remaining 6 sites.

Third, Pearson's correlation analysis was performed between the significant variables and demographic confounding factors. Data were tested for a normal distribution using the Kolmogorov–Smirnov test. Data that were not normally distributed were analysed using Spearman's correlation analysis.

In particular, regarding clinical confounding factors, such as symptoms (HAMD, YMRS and PANSS scores) and medication doses (antidepressants: imipramine (IMP) equivalent dose; antipsychotics: chlorpromazine (CPZ) equivalent dose; anxiolytics: diazepam equivalent dose; and anti-parkinsonian drugs: biperiden equivalent dose, lithium dose, sodium valproate dose and carbamazepine dose), a stepwise multiple linear regression analysis was performed with a probability of F for conservative entry and removal criteria of 0.01 and 0.05, respectively, to elucidate the complicated relationships among these clinical confounding factors in each diagnostic group.

All data are expressed as mean and standard deviation (SD). The significance level was set to alpha = 0.05. When a difference was considered significant, we presented both the effect size (Cohen's d) and the 95% confidence interval (CI). Statistical analyses were performed using the SPSS 16.0.1J software (SPSS Inc., Tokyo, Japan).

Results

Demographic characteristics

Table 1 shows the demographic and clinical characteristics of the 4 age- and gender-matched diagnostic groups used in this study. One-way ANOVA revealed an absence of significant age differences among the groups ($p = 0.99$) and a chi-squared test showed an absence of gender differences among the groups ($p = 0.81$). In addition, the age and gender distributions among the 4 diagnostic groups were not significantly different at the initial site (Gunma University, MDD: 39.9 (11.7) y.o., 12/15; BP: 41.1 (13.2) y.o., 22/15; SZ: 40.1 (14.9) y.o., 11/20; and HC: 40.0 (4.2) y.o., 7/10) (age, $p = 0.98$; gender, $p = 0.24$) and at the other 6 sites (MDD: 44.6 (12.7) y.o., 65/61; BP: 45.1 (15.4) y.o., 47/50; SZ: 44.8 ± 11.0 y.o., 56/49; and HC: 44.0 ± 15.9 y.o., 307/266) (age, $p = 0.89$; gender, $p = 0.81$).

Preliminary test of the difference between HCs and patients

Although it was not the main theme of this study, to compare our results with those of studies of biomarkers performed only to detect functional abnormalities in patients against a control group, we also analysed the differences between HC individuals and patients to confirm the significance of the 3 variables chosen for analysis. Full analyses are described in Supplementary Material (VI).

From the analyses performed using data from the initial site, we adopted both R1 and R2 integral values as statistically significant variables for the algorithm. Thresholds were dependent on the

purpose for which the variables were used (Table 2). For example, if the optimal thresholds of the integral values of R1 and R2 derived from the initial site (73 and 104) were applied to the independent test data from the remaining 6 sites, the sensitivities were 0.73 (proportion of patients/measurement: 96/131) and 0.79 (104/131) and the specificities were 0.63 (proportion of HCs/measurement: 326/514) and 0.63 (324/514) for R1 (positive predictive value (PPV) = 0.37, negative predictive value (NPV) = 0.90) and R2 (PPV = 0.40, NPV = 0.92), respectively.

Test for differentiation of patients with unipolar MDD from those with BP and SZ

Using the preliminary data from the initial site, one-way ANOVA performed between the patients with MDD and those with one of the other 2 disorders of interest (BP or SZ) revealed a significant difference in the R1 centroid values [$F(1,53) = 9.54, p < 0.01$; $d = 0.96$, 95% CI, (0.25 to 1.62)], but not in the R1 [$F(1,53) = 0.14, p = 0.71$] or the R2 [$F(1,53) = 0.05, p = 0.83$] integral values.

As the significant R1 centroid value proved to be the most useful variable, we applied it to ROC analysis for the differentiation of patients with unipolar MDD from those with non-MDD disorders. The resulting area under the ROC curve (Az) value was 0.74 [95% CI, (0.61 to 0.87)] and the optimal threshold was 54 [s] from the extreme top left point of the ROC curve (eFig. S4).

To validate the optimal threshold calculated, we applied it to the independent test data of the remaining 6 sites, to differentiate the patients with MDD from those with SZ and BP [Az = 0.81, 95% CI, (0.74 to 0.89); $d = 1.17$, 95% CI, (0.79 to 1.54); optimal threshold = 54 [s], PPV = 0.79, NPV = 0.82; Fig. 4]. Using this threshold (54 [s]), 74.6% of the patients with MDD (proportion of patients/measurement: 41/55) and 85.5% of those with SZ or BP (65/76) were classified correctly [76.9% of BP patients (20/26) and 90.0% of SZ patients (45/50)] (Fig. 5). The ROC curves of MDD v. BP [Az = 0.74, 95% CI, (0.62 to 0.85); $d = 0.81$, 95% CI, (0.32 to 1.29); optimal threshold = 54 [s], PPV = 0.87, NPV = 0.59] and MDD v. SZ [Az = 0.86, 95% CI, (0.78 to 0.93); $d = 1.40$, 95% CI, (0.96 to 1.82); optimal threshold = 54 [s], PPV = 0.89, NPV = 0.78] are shown separately in eFig. S5.

For reference, the test performed for the differentiation between patients with BP and those SZ is shown in Supplementary Material (VII).

Correlational analysis of demographic and clinical confounding factors

Correlational analysis showed no significant correlations between any of the significant dependent variables (among the R1 and R2 integral values and the R1 centroid value of NIRS signals) and any of the demographic confounding factors [performance (number of correct words), education years and pre-morbid IQ; $p > 0.05$] for all patients with psychiatric disorders (MDD, BP and SZ).

Regarding clinical confounding factors, a stepwise regression analysis of each significant dependent variable for each disorder revealed that there was no entry clinical variable in the linear regression models, with the exception of the global assessment of functioning (GAF) score (beta = 0.50, $p < 0.01$) for the R2 integral value ($F = 10.73, p < 0.01$; $R = 0.50$, $R^2 = 0.25$, adjusted $R^2 = 0.23$) in patients with MDD, and the GAF score (beta = 0.58, $p = 0.01$) for the R2 integral value ($F = 8.43$, $p = 0.01$; $R = 0.59$, $R^2 = 0.35$, adjusted $R^2 = 0.30$) in patients with BP who exhibited depressive symptoms. Thus, only one clinical variable (i.e., GAF score) among all of the medication and clinical variables examined had a significant impact on the R2 integral values for patients with MDD or BP who exhibited depressive symptoms.

Table 2
Sensitivities and specificities of the integral values of Region 1 (R1) and Region 2 (R2) signals between healthy controls and all patients with psychiatric disorders, based on the independent data collected from the 6 additional sites.

Integral value	R1		R2	
	Sensitivity	Specificity	Sensitivity	Specificity
160	0.95	0.27	0.90	0.39
150	0.93	0.30	0.89	0.43
140	0.92	0.34	0.88	0.47
130	0.90	0.37	0.87	0.52
120	0.88	0.42	0.85	0.57
110	0.85	0.46	0.82	0.61
100	0.82	0.50	0.78	0.64
90	0.78	0.54	0.76	0.68
80	0.74	0.61	0.73	0.73
70	0.72	0.65	0.64	0.76
60	0.66	0.71	0.57	0.78
50	0.57	0.75	0.52	0.81
40	0.47	0.80	0.43	0.86
30	0.38	0.84	0.33	0.89

Discussion

The present multi-site study is the first large-scale, case–control study that demonstrates the utility of NIRS for the differential diagnosis of major psychiatric disorders. The main strengths of this study include the application of a neuroimaging biomarker in clinical practice that allows the clinically useful differential diagnosis of depressive states. The frontal centroid value, which represents the timing of frontal NIRS signal patterns, was a significant variable for differential diagnosis and the optimal threshold derived from the ROC analysis correctly discriminated patients with unipolar MDD (74.6%) from those with non-MDD disorders (85.5%; BP, 76.9% and SZ, 90.0%).

Single-individual diagnostic classification analyses among various psychiatric disorders

The present study was not only a case–control study of group comparisons, but also a study specifically designed for examining the practical utility of single-individual diagnostic classification in various psychiatric disorders. Several studies have reported the single-individual diagnostic classification of one psychiatric disorder compared with HCs by applying multivariate statistical methods (e.g.,

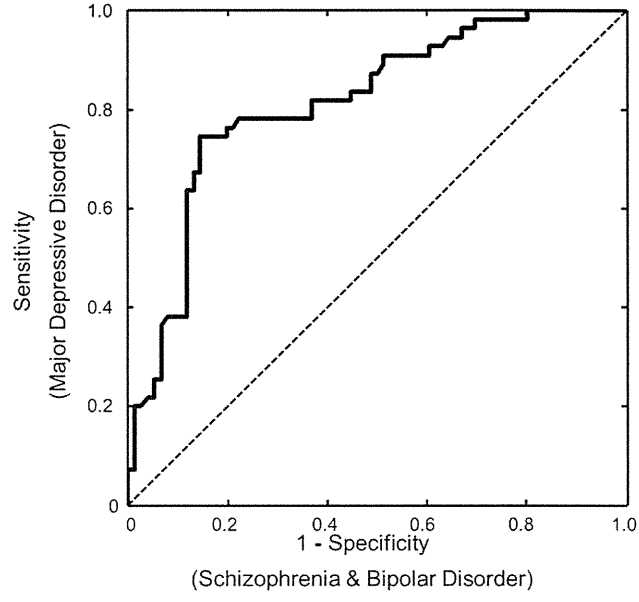


Fig. 4. Receiver operating characteristic analysis of the centroid value of Region 1 (R1) near-infrared spectroscopy signal between patients with major depressive disorder and those with either of the other 2 disorders of interest (bipolar disorder and schizophrenia) based on the independent data collected from the 6 additional sites.

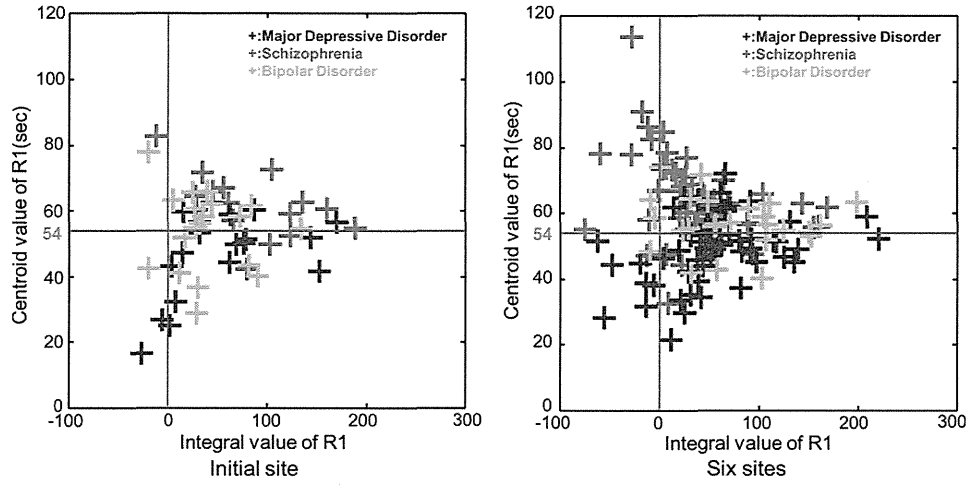


Fig. 5. Scatter plots of the centroid and integral values of Region 1 (R1) signal in the patients, both at the initial site (Gunma University) and at the 6 additional sites.

neuroanatomical pattern classification) to structural MRI data (Davatzikos et al., 2005) and NIRS data (Hahn et al., 2013) from SZ and high-risk psychosis samples (Koutsouleris et al., 2009), as well as to functional MRI data from patients with depression (Hahn et al., 2011). These studies were technically sophisticated; however, more research must be performed to test their reproducibility and generalisability in the advanced stage of clinical application, because (1) they were designed for the analysis of one diagnostic classification based on comparison to HCs, and not for differential diagnosis among multiple psychiatric disorders; and (2) they were performed using one relatively small cohort; thus, they must be replicated in another cohort including larger sample groups.

Furthermore, we will discuss briefly our results in comparison with those of other single-individual diagnostic classification studies (Davatzikos et al., 2005; Fu et al., 2008; Hahn et al., 2011; Koutsouleris et al., 2009). We used only a single variable (simple 'centroid value' of NIRS signals) and found that the classification rates (unipolar MDD: 74.6% correct classification; the 2 other disorders: 85.5% correct classification (BP, 76.9%; SZ, 90.0%)) were almost equivalent to the rates reported in the previous MRI studies using multivariate statistical methods (which had 80–90% classification rates in the patient group compared with the HC group).

To determine whether a higher disease classification rate could be achieved by using a multivariate pattern analysis (compared with that obtained using one simple variable), which was used in previous MRI studies, we confirmed the results using the multivariate pattern classification analysis described in Supplementary Material (VIII). The leave-one-out cross-validation method revealed that 4 significant variables, or even one variable (the R1 centroid value), could differentiate patients with unipolar MDD from those with either of the 2 other disorders (non-MDD) with a similar degree of mean accuracy (76.8% (unipolar MDD: 73.0% (54/74), non-MDD: 74.8% (83/111))).

Clinical importance and implications

Another clinically valuable feature of our work is that it aimed to facilitate diagnosis among patients with similar depressive symptoms, which psychiatrists often find to be a difficult task. Most BP patients with depressive symptoms are initially diagnosed with and treated for MDD (Akiskal et al., 1995; Goldberg et al., 2001). Therefore, our findings may help differentiate BP with depressive symptoms from MDD. Depressive symptoms and cognitive deficits are also common early signs of SZ (Hafner et al., 2005). Of particular clinical relevance is the observation that SZ patients with concomitant depression have a greater risk of suicide or an unfavourable disease course (an der Heiden et al., 2005). Therefore, sufficient attention must be given to the diagnosis and treatment of depression in SZ patients.

The results of the present study may draw attention to the heterogeneity observed among MDD patients. Rather than simply being misclassified, approximately 25% of patients with unipolar MDD who were classified by the system as having a non-MDD disorder may have a brain pathophysiology that is biologically different from that of the majority of MDD patients. Evidence suggests that 25–50% of individuals with recurrent major depression (particularly those in atypical early-onset or treatment-refractory subgroups) may in fact have broadly defined BP (Angst, 2007). In this study, 74.6% of the patients with MDD were classified correctly; the remaining 25.4% might include either patients who would progress to a diagnosis of one of the 2 other disorders or patients with a broadly defined BP who were diagnosed with MDD according to the DSM criteria. This explanation might be justified by the finding of a correct classification rate of 75% for patients with MDD. For practical purposes, among patients diagnosed clinically with MDD, the early suspicion of the possibility of a diagnosis of a non-MDD disorder with depression would also provide an opportunity to reduce the hazardous effects of the illness on personal, social and occupational aspects; therefore, our results should be of great clinical importance in practical applications. Thus, a prospective study aimed at elucidating the heterogeneity of unipolar MDD is required.

Advantages of the NIRS method

We used the same NIRS system (a non-invasive, portable and user-friendly device) and the same concise measurement procedure at every site; therefore, inter-site compatibility was not an issue here; however, it may be an obstacle in other neuroimaging multi-site studies. Furthermore, we used a high temporal resolution (0.1 s) in the NIRS system for measuring time-specific characteristics of dynamic prefrontal cortical functions; this enabled analyses that included more detailed time-course comparisons of NIRS signal changes. We created and adopted new variables, such as the 'centroid value', to determine the timing of the haemodynamic response (Fig. 1). The high temporal resolution of NIRS might allow not only the detection of functional abnormalities (e.g., hypofrontality), but also the capture of the specific haemodynamic activation time courses of each psychiatric disorder and aid differential diagnosis.

The practical application of biomarkers requires that they be relatively simple. The simplicity of both the test procedure and the associated data analysis is important not only for the participants, but also for their caretakers and clinicians. Therefore, rather than using complicated multivariate statistical methods, we developed a robust classification algorithm for real-time visual evaluation of patients using the simplest, and lowest number of variables on the basis of a ROC analysis. This was important because we sought to develop a psychiatric practice empowered by the initiative of patients by sharing the 'comprehensively visualised' results that can be easily recognisable by patients and caretakers, rather than results from complicated 'black-box' analyses. In addition, using the condensed VFT (<3 min) developed previously by us, we designed a diagnostic support system in a way that the results are available to clinicians in less than 15 min. The availability of such a 'comprehensively visualised' report to clinicians, patients and their caretakers at a first visit, while laying out a future treatment plan, would likely lead to a paradigm shift to a patient-centred approach in clinical psychiatry.

Limitations

The methodological aspects of the present study warrant commentary. First, most of the patients included in the study were taking medications at the time of measurement. To our knowledge, no clear evidence of the effects of medication on NIRS signals has been demonstrated. We found that none of the medications at any dose was significantly correlated with NIRS signals in this study; however, we cannot fully exclude the effects of medication on haemodynamic signals. For confirmation, the application of the algorithm described above (optimal cut-off of the R1 centroid value) to the drug-free patients exclusively, 6 out of 10 patients with MDD patients (60%) and 4 out of 5 patients with SZ (80%) were classified correctly. Second, the size of the sample included in our final analysis was substantially reduced from that initially recruited, because we tried to minimise the confounding factors of age and gender by matching the groups and excluded patients in remission, as well as patients in the manic phase (see Flow diagram). In our confirmatory analysis, we included all non-matched and in-remission patients and found that the results were quite similar, although this analysis had a lower detection power. The optimal threshold of the sample sets before demographical matching was also the same as that calculated originally. These results suggest that the reduction in the total number of study participants after demographical matching did not affect the development of the algorithm [see Supplementary Material (I)]. However, we must consider the possibility that this diagnostic support system is best suited for young and middle-aged patients with moderate or severe symptoms (e.g., aged between 23 and 65 years (mean ± 1.5 SD)). Third, a PCA of haemodynamic response performed to capture a channel cluster led to the identification of 2 cluster regions. Nonetheless, as we thought that pooling many NIRS signals together into only 2 representative regions of interest (R1 (frontopolar and dorsolateral prefrontal regions) and R2 (ventrolateral prefrontal and temporal regions)) might oversimplify the results (see the Discussion of Supplementary Material (II)), we sought to confirm the reliability of the 2 clusters by performing a test–retest analysis in a portion of the samples. We found significant ICCs for both the R1 and R2 integral values and for the R1 centroid value between 2 measurements (see Supplementary Material (IV)). Therefore, we used the two data-derived clusters that reflected a fronto-temporal haemodynamic response during VFT. Fourth, we have controlled some well-known confounders in the analyses. However, the NIRS signal might be affected by the other systemic confounders, such as autonomic function, neuroendocrine function, diet and physical activity. In addition, brain anatomical factors, such as scalp–cortex distance and frontal sinus volume, as well as genetic variants might also be potential confounders. Further studies are required to address the relationship between the NIRS signal and these confounders. If these findings are fully replicated, the development of methods of integrating confounding factors into NIRS signal in the future will be ideal. Fifth, we did not use the exclusion criterion of first-degree relatives with axis I psychiatric disorders for healthy controls. This could give a bias to the data in healthy controls, which means that some of the first-degree relatives of persons with axis I psychiatric disorders might have been included as a healthy control in the present study. However, as the same situations are assumed in real clinical settings, we daringly recruited healthy controls without applying that strict exclusion criterion.

Conclusions and future implications

In conclusion, this multi-site study provided evidence that the fronto-temporal NIRS signal may be used as a tool in assisting the diagnosis of major psychiatric disorders with depressive symptoms. Future NIRS research should be performed to study the applicability of this method to (1) the identification of a need for therapy, (2) the assessment of the efficacy of various treatments, (3) the establishment of prognostic predictions that may be clarified by longitudinal follow-up assessments of patients in various clinical stages and (4) the examination of the use of NIRS as a screening tool.

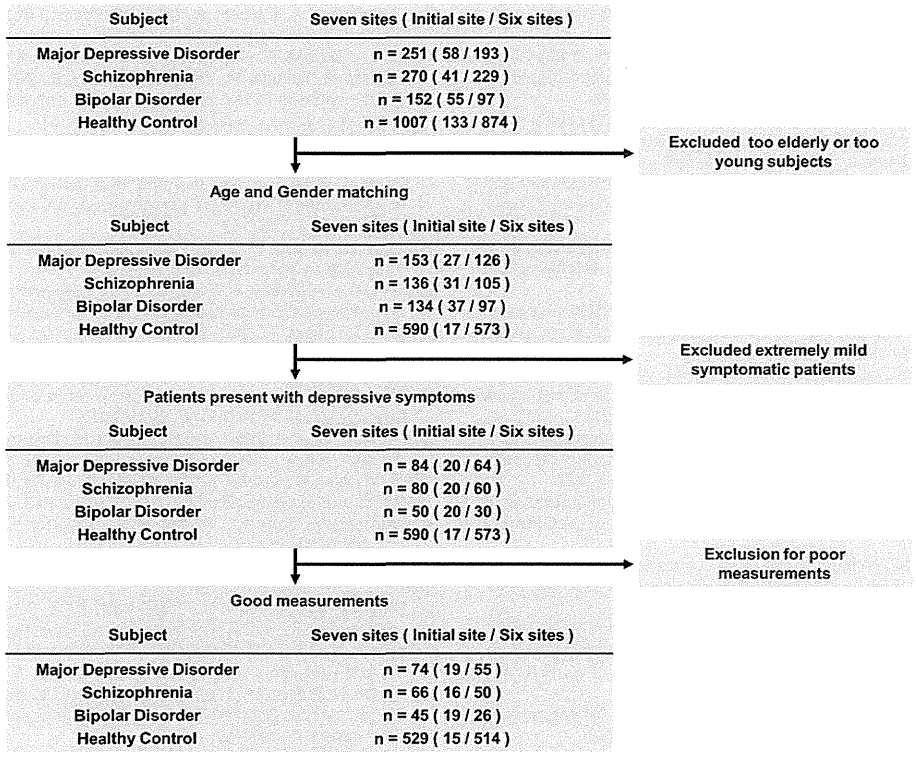
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.05.126.

Acknowledgments

The authors would like to thank all the participants in this study. The authors also thank Makoto Ito, M.D., Ph.D., Tomohiro Suto, M.D., Ph.D., Yutaka Yamagishi, M.D., Naoki Hanaoka, M.D., Ph.D., Toshimasa Sato, Ed.M. and Noriko Sakurai of Gunma University; Kohei Marumo, M.D., Ph.D. and Yuki Kawakubo, Ph.D., of The University of Tokyo; Hitomi Kobayashi, Ph.D. and Junko Motoki, Ph.D., of Showa University; Osamu Saitoh, M.D., Ph.D., Kimitaka Anami, M.D., Ph.D., Yohtaro Numachi, M.D., Ph.D., Yuji Sugimura and Masaru Ogawa of the National Centre of Neurology & Psychiatry Hospital; Daisuke Gotoh, M.D., Yojiro Sakai, M.D., Ph.D. and Emi Yoshida, M.D., of Fukushima Medical University, for technical assistance during data collection.

Role of the funding source

This study was supported in part by grants-in-aid for scientific research from the Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan (Nos. 18390318, 19659289, 20390310 and 22659209 to MF; Nos. 21249064 and 221S003 [Innovative Areas (Comprehensive Brain Science Network)] to KK; No. 17019029 [Priority Areas (Applied Genomics)] to YO; and No. 23791309 to RT), and by grants-in-aid from the Ministry of Health, Labour, and Welfare (MHLW) of Japan (H20-kokoro-ippan-001, H20-3, and H22-seishin-ippan-015 to KK). This study was also supported in part by Health and Labour Sciences Research Grants for Comprehensive Research on Disability Health and Welfare (previously, Health and Labour Science Research Grant for Research on Psychiatric and Neurological Disease and Mental Health; H20-001 to MF and H23-seishin-ippan-002 to RT) and by the Intramural Research Grants for Neurological and Psychiatric Disorders of The National Centre of Neurology and Psychiatry (previously, The Research Grant for Nervous and Mental Disorders from the MHLW; 21B-1 to MF and 20B-3 to TN). In addition, this study was supported in part by an Intramural Research Grant for Neurological and Psychiatric Disorders of NCNP (No. 23-10 to RT). A part of this study was also the result of a project entitled 'Development of biomarker candidates for social behavior', which was carried out under the Strategic Research Program for Brain Sciences by MEXT, Japan. This study was also supported in part



Flow diagram. Main analyses were based on matched samples according to the flow diagram.

by grants from the Japan Research Foundation for Clinical Pharmacology (to RT).

The funders had no role in the design of the study, data collection, data analysis, decision to publish, or preparation of the manuscript. The corresponding author had full access to all of the data in this study and takes responsibility for the integrity of the data, the accuracy of the data analysis and the decision to submit the manuscript for publication.

Author contributions

Masato Fukuda designed the experiments and organized the multi-site collaborative study. Ryu Takizawa, Masato Fukuda, Shingo Kawasaki, and Kiyoto Kasai analysed the data and wrote the first draft of the paper. The other contributors performed data acquisition and revised the first draft critically for important intellectual content. All contributors have approved the final version of the manuscript.

Contributors from the Joint Project for Psychiatric Application of Near-Infrared Spectroscopy (JPSY-NIRS) Group

Gunma University: Masato Fukuda, M.D., Ph.D., Masashi Suda, M.D., Ph.D., Yuichi Takei, M.D., Ph.D., Yoshiyuki Aoyama, M.D., Ph.D., Kosuke Narita, M.D., Ph.D., Masahiko Mikuni, M.D., Ph.D., Masaki Kameyama, M.D., Ph.D., and Toru Uehara, M.D., Ph.D.

The University of Tokyo: Ryu Takizawa, M.D., Ph.D., Kiyoto Kasai, M.D., Ph.D., Masaru Kinou, M.D., Ph.D., Shinsuke Koike, M.D. and Ayaka Ishii-Takahashi, M.D.

Hitachi Medical Corporation: Shingo Kawasaki, MS, Noriyoshi Ichikawa, BS, and Michiyuki Fujiwara, AS.

Showa University: Haruhisa Ohta, M.D., Ph.D., Hiroi Tomioka, M.D., Ph.D., Bun Yamagata, M.D., Ph.D., and Kaori Yamanaka, M.D.

Keio University: Masaru Mimura, M.D., Ph.D.

Tottori University: Shenghong Pu, Ph.D.

The National Centre of Neurology and Psychiatry: Kazuyuki Nakagome, M.D., Ph.D., Takamasa Noda, M.D., Taro Matsuda, M.D., and Sumiko Yoshida, M.D., Ph.D.

Fukushima Medical University: Shin-Ichi Niwa, M.D., Ph.D., Soichi Kono, M.D., Hirooki Yabe, M.D., Ph.D., and Sachie Miura, M.D.

Mie University and Tokyo Metropolitan Matsuzawa Hospital: Yuji Okazaki, M.D., Yukika Nishimura, Ph.D., Hisashi Tanii, M.D., Ph.D., Ken Inoue, M.D., Ph.D., Chika Yokoyama, MA, Yoichiro Takayanagi, M.D., Ph.D., Katsuyoshi Takahashi, M.D., and Mayumi Nakakita, M.D.

Conflict of interest

We would like to disclose potential conflicts of interest regarding all financial and material support for the present study. The principal investigators of each site (Masato Fukuda of Gunma University, Kiyoto Kasai of The University of Tokyo, Masaru Mimura of Keio university and Showa University, Kazuyuki Nakagome of The National Center of Neurology and Psychiatry and Tottori University, Shin-ichi Niwa of Fukushima Medical University, Yuji Okazaki of both Mie University and Tokyo Metropolitan Matsuzawa Hospital and Takamasa Noda of The National Center of Neurology and Psychiatry) have potential conflicts of interest in the submitted work. Each site has had an official contract with the Hitachi Group (Advanced Research Laboratory, Hitachi, Ltd., and The Research and Developmental Center, Hitachi Medical Corporation) for a collaborative study of the clinical application of NIRS in psychiatric disorders. For this study, the Hitachi Group provided a project grant (JPY 300,000–2,000,000 per year) and material support (temporary rental of a NIRS (Optical Topography) ETG-4000 system) for each site. Shingo Kawasaki, Noriyoshi Ichikawa and Michiyuki Fujiwara are employees of Hitachi Medical Corporation.

The other authors have no financial relationships with any organisations that might have an interest in the submitted work in previous years and no other relationships or activities that could appear to have influenced the submitted work.

References

Akiskal, H.S., Maser, J.D., Zeller, P.J., Endicott, J., Coryell, W., Keller, M., Warshaw, M., Clayton, P., Goodwin, F., 1995. Switching from 'unipolar' to bipolar II. An 11-year

prospective study of clinical and temperamental predictors in 559 patients. Arch. Gen. Psychiatry 52, 114–123.
Almeida, J.R., Versace, A., Mechelli, A., Hassel, S., Quevedo, K., Kupfer, D.J., Phillips, M.L., 2009. Abnormal amygdala–prefrontal effective connectivity to happy faces differentiates bipolar from major depression. Biol. Psychiatry 66, 451–459.
an der Heiden, W., Konnecke, R., Maurer, K., Ropeter, D., Hafner, H., 2005. Depression in the long-term course of schizophrenia. Eur. Arch. Psychiatry Clin. Neurosci. 255, 174–184.
Angst, J., 2007. The bipolar spectrum. Br. J. Psychiatry 190, 189–191.
Barch, D.M., Sheline, Y.I., Csernansky, J.G., Snyder, A.Z., 2003. Working memory and prefrontal cortex dysfunction: specificity to schizophrenia compared with major depression. Biol. Psychiatry 53, 376–384.
Barrett, S.L., Mulholland, C.C., Cooper, S.J., Rushe, T.M., 2009. Patterns of neurocognitive impairment in first-episode bipolar disorder and schizophrenia. Br. J. Psychiatry 195, 67–72.
Costafreda, S.G., Fu, C.H., Lee, L., Everitt, B., Brammer, M.J., David, A.S., 2006. A systematic review and quantitative appraisal of fMRI studies of verbal fluency: role of the left inferior frontal gyrus. Hum. Brain Mapp. 27, 799–810.
Curtis, V.A., Dixon, T.A., Morris, R.G., Bullmore, E.T., Brammer, M.J., Williams, S.C., Sharma, T., Murray, R.M., McGuire, P.K., 2001. Differential frontal activation in schizophrenia and bipolar illness during verbal fluency. J. Affect. Disord. 66, 111–121.
Davatzikos, C., Shen, D., Gur, R.C., Wu, X., Liu, D., Fan, Y., Hughett, P., Turetsky, B.I., Gur, R.E., 2005. Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch. Gen. Psychiatry 62, 1218–1227.
Ferrari, M., Quaresima, V., 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. Neuroimage 63 (2), 921–935.
First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1997. Structured Clinical Interview for DSM-IV Axis I Disorders. Biometric Research Department, New York State Psychiatric Institute, New York, U.S.A.
Fu, C.H., Mourao-Miranda, J., Costafreda, S.G., Khanna, A., Marquand, A.F., Williams, S.C., Brammer, M.J., 2008. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. Biol. Psychiatry 63, 656–662.
Goldberg, J.F., Harrow, M., Whiteside, J.E., 2001. Risk for bipolar illness in patients initially hospitalized for unipolar depression. Am. J. Psychiatry 158, 1265–1270.
Gur, R.E., Keshavan, M.S., Lawrie, S.M., 2007. Deconstructing psychosis with human brain imaging. Schizophr. Bull. 33, 921–931.
Hafner, H., Maurer, K., Trendler, G., an der Heiden, W., Schmidt, M., 2005. The early course of schizophrenia and depression. Eur. Arch. Psychiatry Clin. Neurosci. 255, 167–173.
Hahn, T., Marquand, A.F., Ehlis, A.C., Dresler, T., Kittel-Schneider, S., Jarczok, T.A., Lesch, K.P., Jakob, P.M., Mourao-Miranda, J., Brammer, M.J., Fallgatter, A.J., 2011. Integrating neurobiological markers of depression. Arch. Gen. Psychiatry 68, 361–368.
Hahn, T., Marquand, A.F., Plichta, M.M., Ehlis, A.C., Schecklmann, M.W., Dresler, T., Jarczok, T.A., Eirich, E., Leonhard, C., Reif, A., Lesch, K.P., Brammer, M.J., Mourao-Miranda, J., Fallgatter, A.J., 2013. A novel approach to probabilistic biomarker-based classification using functional near-infrared spectroscopy. Hum. Brain Mapp. 34, 1102–1114.
Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.
Holsboer, F., 2008. How can we realize the promise of personalized antidepressant medicines? Nat. Rev. Neurosci. 9, 638–646.
Hoshi, Y., Kobayashi, N., Tamura, M., 2001. Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. J. Appl. Physiol. 90, 1657–1662.
Kameyama, M., Fukuda, M., Yamagishi, Y., Sato, T., Uehara, T., Ito, M., Suto, T., Mikuni, M., 2006. Frontal lobe function in bipolar disorder: a multichannel near-infrared spectroscopy study. NeuroImage 29, 172–184.
Kay, S.R., Opler, L.A., Fiszbein, A., 1991. Positive and Negative Syndrome Scale (PANSS) Rating Manual. Multi-health Systems Inc., Toronto.
Koike, S., Takizawa, R., Nishimura, Y., Takano, Y., Takayanagi, Y., Kinou, M., Araki, T., Harima, H., Fukuda, M., Okazaki, Y., Kasai, K., 2011. Different hemodynamic response patterns in the prefrontal cortical sub-regions according to the clinical stages of psychosis. Schizophr. Res. 132, 54–61.
Koutsouleris, N., Meisenzahl, E.M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Moller, H.J., Gaser, C., 2009. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch. Gen. Psychiatry 66, 700–712.
Mathers, C.D., Loncar, D., 2006. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3, e442.
McGorry, P.D., Killackey, E., Yung, A., 2008. Early intervention in psychosis: concepts, evidence and future directions. World Psychiatry 7, 148–156.
Obrig, H., Villringer, A., 2003. Beyond the visible—imaging the human brain with light. J. Cereb. Blood Flow Metab. 23, 1–18.
Okada, E., Delpy, D.T., 2003. Near-infrared light propagation in an adult head model. II. Effect of superficial tissue thickness on the sensitivity of the near-infrared spectroscopy signal. Appl. Opt. 42, 2915–2922.
Phillips, M.L., Vieta, E., 2007. Identifying functional neuroimaging biomarkers of bipolar disorder: toward DSM-V. Schizophr. Bull. 33, 893–904.
Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M.R., Rahman, A., 2007. No health without mental health. Lancet 370, 859–877.
Schecklmann, M., Ehlis, A.C., Plichta, M.M., Fallgatter, A.J., 2008. Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency. Neuroimage 43 (1), 147–155.
Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39, 1064–1080.
Strangman, G., Boas, D.A., Sutton, J.P., 2002a. Non-invasive neuroimaging using near-infrared light. Biol. Psychiatry 52, 679–693.
Strangman, G., Culver, J.P., Thompson, J.H., Boas, D.A., 2002b. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. NeuroImage 17, 719–731.
Suto, T., Fukuda, M., Ito, M., Uehara, T., Mikuni, M., 2004. Multichannel near-infrared spectroscopy in depression and schizophrenia: cognitive brain activation study. Biol. Psychiatry 55, 501–511.
Takizawa, R., Kasai, K., Kawakubo, Y., Marumo, K., Kawasaki, S., Yamasue, H., Fukuda, M., 2008. Reduced frontopolar activation during verbal fluency task in schizophrenia: a multi-channel near-infrared spectroscopy study. Schizophr. Res. 99, 250–262.
Tsuzuki, D., Jurcak, V., Singh, A.K., Okamoto, M., Watanabe, E., Dan, I., 2007. Virtual spatial registration of stand-alone fNIRS data to MNI space. NeuroImage 34, 1506–1518.
Yamashita, Y., Maki, A., Ito, Y., Watanabe, E., Koizumi, H., 1996. Noninvasive near-infrared topography of human brain activity using intensity modulation spectroscopy. Opt. Eng. 35, 1046–1049.
Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry 133, 429–435.
Zanelli, J., Reichenberg, A., Morgan, K., Fearon, P., Kravariti, E., Dazzan, P., Morgan, C., Zanelli, C., Demjaha, A., Jones, P.B., Doody, G.A., Kapur, S., Murray, R.M., 2010. Specific and generalized neuropsychological deficits: a comparison of patients with various first-episode psychosis presentations. Am. J. Psychiatry 167, 78–85.
Zimmermann, P., Bruckl, T., Nocon, A., Pfister, H., Lieb, R., Wittchen, H.U., Holsboer, F., Angst, J., 2009. Heterogeneity of DSM-IV major depressive disorder as a consequence of subthreshold bipolarity. Arch. Gen. Psychiatry 66, 1341–1352.

NeuroImage 83 (2013) 158–173

Contents lists available at ScienceDirect

NeuroImage

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

A NIRS–fMRI investigation of prefrontal cortex activity during a working memory task

Hiroki Sato [a,*,1], Noriaki Yahata [b,e,1], Tsukasa Funane [a], Ryu Takizawa [b,f], Takusige Katura [a], Hirokazu Atsumori [a], Yukika Nishimura [b,d], Akihide Kinoshita [b], Masashi Kiguchi [a], Hideaki Koizumi [a], Masato Fukuda [c], Kiyoto Kasai [b]

[a] Hitachi, Ltd., Central Research Laboratory, Hatoyama, Saitama 350-0395, Japan
[b] Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo 113-8655, Japan
[c] Department of Psychiatry and Human Behavior, Gunma University Graduate School of Medicine, Maebashi, Gunma 371-8511, Japan
[d] Department of Youth Mental Health, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo 113-8655, Japan
[e] Global Center of Excellence (COE) Program: "Comprehensive Center of Education and Research for Chemical Biology of the Diseases", The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan
[f] MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, SE5 8AF, UK

ARTICLE INFO

Article history:
Accepted 12 June 2013
Available online 21 June 2013

Keywords:
Functional magnetic resonance imaging (fMRI)
Near-infrared spectroscopy (NIRS)
Optical topography
Blood oxygenation level dependent (BOLD)
Hemoglobin
Prefrontal cortex
Simultaneous measurement
Working memory
Finger tapping

ABSTRACT

Near-infrared spectroscopy (NIRS) is commonly used for studying human brain function. However, several studies have shown that superficial hemodynamic changes such as skin blood flow can affect the prefrontal NIRS hemoglobin (Hb) signals. To examine the criterion-related validity of prefrontal NIRS-Hb signals, we focused on the functional signals during a working memory (WM) task and investigated their similarity with blood-oxygen-level-dependent (BOLD) signals simultaneously measured by functional magnetic resonance imaging (fMRI). We also measured the skin blood flow with a laser Doppler flowmeter (LDF) at the same time to examine the effect of superficial hemodynamic changes on the NIRS-Hb signals. Correlation analysis demonstrated that temporal changes in the prefrontal NIRS-Hb signals in the activation area were significantly correlated with the BOLD signals in the gray matter rather than those in the soft tissue or the LDF signals. While care must be taken when comparing the NIRS-Hb signal with the extracranial BOLD or LDF signals, these results suggest that the NIRS-Hb signal mainly reflects hemodynamic changes in the gray matter. Moreover, the amplitudes of the task-related responses of the NIRS-Hb signals were significantly correlated with the BOLD signals in the gray matter across participants, which means participants with a stronger NIRS-Hb response showed a stronger BOLD response. These results thus provide supportive evidence that NIRS can be used to measure hemodynamic signals originating from prefrontal cortex activation.

Introduction

Human tissue is highly transparent to light in the visible to near-infrared region, so this region is called the "biological optical window." Millikan's development of a spectroscopic method for estimating the oxygen saturation of blood in human tissue (Millikan, 1942) was followed by the development of the near-infrared spectroscopy (NIRS) technique for noninvasive measurement of tissue oxygenation in the brain (Jöbsis, 1977; Jöbsis-VanderVliet et al., 1988). Particularly important has been the application of NIRS to the measurement of hemodynamic signals related to functional activation in the brain (Chance et al., 1993; Hoshi and Tamura, 1993; Kato et al., 1993; Villringer et al., 1993). The NIRS technique measures relative changes in the concentration of oxygenated hemoglobin (Hb) and deoxygenated Hb (oxy-Hb signal and deoxy-Hb signal, respectively), taking advantage of the difference in absorption coefficients between oxy-Hb and deoxy-Hb depending on the light wavelength (Wray et al., 1988).

From a theoretical viewpoint, the origin of the NIRS deoxy-Hb signal is assumed to be the same as that of the blood oxygenation level-dependent (BOLD) signal in functional magnetic resonance imaging (fMRI), which is sensitive to changes in the concentration of deoxy-Hb in local blood vessels (Ogawa et al., 1990a, 1990b). After a NIRS imaging technique (optical topography) with multiple measurement positions was introduced (Maki et al., 1995; Yamashita et al., 1996), a number of studies using NIRS were conducted in various fields as it offers several advantages. For example, its imposition of fewer body constraints facilitates measurement of the infant brain (Homae et al., 2010; Minagawa-Kawai et al., 2011; Pena et al., 2003; Sato et al., 2012; Taga et al., 2003) and of activations during

* Corresponding author. Fax: +81 49 296 6005.
E-mail address: hiroki.sato.ry@hitachi.com (H. Sato).
[1] These authors contributed equally to this work.

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.06.043

walking (Atsumori et al., 2010; Miyai et al., 2001) and communication (Cui et al., 2012; Funane et al., 2011), which are more difficult with fMRI. Moreover, NIRS is easier to administer and less expensive than fMRI, meaning that it will open up new applications.

Although NIRS has certain advantages compared to fMRI, it has a limited measurement depth and less spatial resolution. A potential problem currently drawing attention is contamination of the measurement data due to extracranial hemodynamic changes such as skin blood flow (Germon et al., 1994, 1998; Kirilina et al., 2012; Kohno et al., 2007; Takahashi et al., 2011; Toronov et al., 2001). Given the measurement principle of NIRS, it would not be surprising to find that NIRS-Hb signals contain information reflecting superficial hemodynamic changes mainly due to systemic changes. Therefore, researchers have placed special emphasis on their task paradigms and statistical analysis to ensure extraction of pure brain activity. Although this approach is valid if the assumption holds that extracranial hemodynamics is independent of the task sequence or at least negligible in obtaining brain activity, recent studies have suggested that superficial hemodynamic changes such as skin blood flow are a considerable source of the task-related signals, especially in the frontal area (Kirilina et al., 2012; Takahashi et al., 2011). For example, Takahashi et al. (2011) suggested that the majority of NIRS-Hb signal changes in the forehead reflected the skin blood flow during a verbal fluency task (VFT). The main evidence for this conclusion was derived from an experiment showing that the VFT-related signal changes disappeared when the participants manually pressed a portion of the skin in the measurement area (Takahashi et al., 2011). Moreover, Kirilina et al. (2012) showed task-evoked hemodynamic changes in the superficial forehead region during a continuous performance task. They attributed this task-evoked artifact to systemic changes in the veins draining the scalp, as shown by fMRI results in which similar task-evoked extracranial changes were localized in the superficial frontal veins (Kirilina et al., 2012). Given these results, there is concern that significant artifacts from extracranial hemodynamics in the NIRS-Hb signals may be present in a wide range of cognitive studies. However, the impact of the extracranial artifacts, including their significance and generality, has not been clarified, let alone their true origin and mechanism.

Before clarifying the effect of extracranial artifacts in NIRS measurements, the present study used simultaneous NIRS and fMRI measurements to investigate the validity of using NIRS to measure prefrontal activity during a reliable task. The BOLD signal of fMRI has relatively accurate spatial information and is proportional to the hemodynamic changes that can be measured using NIRS (Buxton et al., 2004; Toronov et al., 2003). It is therefore possible to evaluate the criterion-related validity of the NIRS-Hb signal by comparing it with the corresponding BOLD signal. Of the numerous studies using simultaneous NIRS–fMRI measurements, two recent ones have investigated prefrontal cortex (PFC) activation (Cui et al., 2011; Heinzel et al., 2013). Although they showed a wide regional and inter-individual variability in the NIRS–BOLD correlations around the PFC (Cui et al., 2011; Heinzel et al., 2013), they not fully demonstrate the correlation level in the focused-upon activation region.

Given this background, we examined in more detail to what extent the prefrontal NIRS-Hb signals showing task-related changes are correlated with the BOLD signals. To confirm the generality of the results for the PFC, sensorimotor areas were also measured using a finger tapping task (TAP). The sensorimotor activations mainly in the precentral and postcentral gyri have been well examined in a number of simultaneous NIRS–fMRI studies and have been shown to have high correlation with the signals (Huppert et al., 2006; Kleinschmidt et al., 1996; Mehagnoul-Schipper et al., 2002; Sassaroli et al., 2006; Strangman et al., 2002; Toronov et al., 2001). Of particular interest, Mehagnoul-Schipper et al. (2002) showed that the amplitudes of the deoxy-Hb signals corresponding to task-related changes were significantly correlated with those of the BOLD signals in the left motor cortex. This amplitude correlation is important because it indicates that equivalent statistical results are expected in the two modalities. However, this amplitude correlation has been reported only by Mehagnoul-Schipper et al. (2002), and it naturally remains unclear in the case of PFC activation.

Therefore, the first objective of the present study was to demonstrate a similar amplitude correlation between NIRS and BOLD in PFC activation signals. The second objective was to determine the correlation of NIRS-Hb signals not only with the BOLD signals in the gray matter (GM) layer but also with the BOLD signals in the soft tissue (ST) layer and the skin blood flow measured with a laser Doppler flowmeter (LDF). This approach should provide helpful information about the effect of superficial hemodynamic changes on the prefrontal NIRS-Hb signals. Although weaker signal intensities are expected in ST regions compared to GM regions, the usefulness of ST BOLD signals has been suggested (Heinzel et al., 2013; Kirilina et al., 2012).

We used a working memory (WM) task with a classical delayed response paradigm as a reliable cognitive task for measuring PFC activity. A region of the dorsolateral PFC, Brodmann area (BA) 46, has been shown to be involved in this task through a number of animal electrophysiological studies (Goldman-Rakic, 1987; Kubota and Niki, 1971; Tsujimoto and Postle, 2012) and human neuroimaging studies (D'Esposito, 2007; McCarthy et al., 1994, 1996). Several NIRS studies have also demonstrated PFC activation during a WM task both in children (Tsujimoto et al., 2004) and adults (Aoki et al., 2011, 2013; Sato et al., 2011a). By focusing on activity in BA 46 during the WM task, we expected to achieve more reliable signals from the cerebral cortex.

Materials and methods

Participants

Twenty-seven volunteers (23 males and 4 females, mean age of 36.4, age range of 28–48) participated in the simultaneous NIRS and fMRI measurements. All participants were native Japanese speakers and gave written informed consent to the study protocol, which was approved by the ethical committee of the Faculty of Medicine, the University of Tokyo (No. 3156-(2)). None of these participants had a history of psychiatric or neurological illness, serious head injury, or psychotropic drug use.

Experimental setup and task sequence

The participant was fitted with NIRS probe holders (see Section NIRS for more details) and wore a nonmagnetic headphone (Hitachi Advanced Systems Corp., Japan) for auditory presentation. After the participant lay down on a scanner bed, an LDF probe was centered between the eyebrows, and another one was placed over the left temple. A head coil was placed on his or her head, and mirrors were placed on the head coil to enable the participant to see a screen positioned at his or her feet.

Two task sessions were conducted for each participant: a verbal working memory (WM) task and a finger tapping (TAP) task. The functional activation signals during task performance were measured using NIRS and fMRI. A commercially available software package (E-prime, Psychology Software Tools, Inc., U.S.A.) was used to present visual and auditory stimuli, to synchronize the stimuli presentation to the fMRI scanning, to send serial commands to the NIRS system for recording the time of presentation, and to record the participant's responses for the WM task. A TAP task session was conducted after the WM task session for all participants. A verbal fluency task session was also conducted for another purpose, and the results for that session will be reported elsewhere.

Verbal working memory task

A conventional delayed-response paradigm was used as the verbal WM task (Fig. 1A). It was basically the same as that used in previous

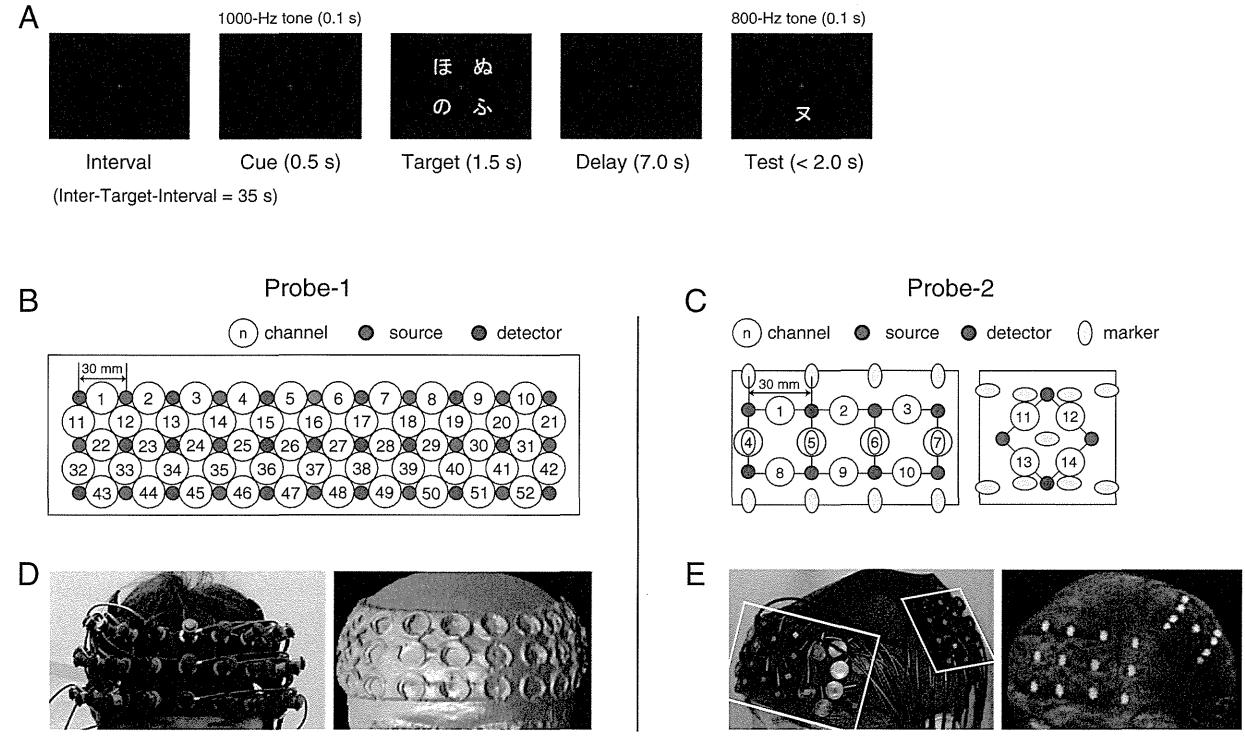


Fig. 1. Task paradigm and NIRS measurement setting. (A) Schematic diagram of WM task trial. (B) Configuration of measurement channels with unique numbers for Probe-1: there were 17 light-irradiating positions (red dots) and 16 detecting positions (blue dots). (C) Configuration of regular measurement channels (S–D distance of 30 mm) with unique numbers for Probe-2: there were six light-irradiating positions (red dots) and six detecting positions (blue dots). Positions of 21 vitamin E tablets used as markers are shown by yellow ellipses. Left part is placed on left prefrontal position; right part is placed on left parietal position. (D) Photograph of Probe-1 worn by representative participant (left) and corresponding T1-weighted image with probe holders, in which NIRS channel positions can be determined (right). (E) Photograph of Probe-2 worn by representative participant (left) and corresponding T1-weighted image with vitamin markers, in which the channel positions can be determined (right).

NIRS studies (Aoki et al., 2011, 2013; Sato et al., 2011a). Each task trial started with a 1500-ms presentation of the target stimuli for encoding (Target) on a PC display screen, which was followed by a delay of 7000 ms. A test stimulus for retrieval (Test) was then presented for 2000 ms or until the participant responded. The participant responded by pressing a button on a handheld box (Serial Response Box, Psychology Software Tools, Inc., U.S.A.) connected to the PC. The system recorded the button pressed and the reaction time. A Japanese *hiragana* character or a set of four such characters was presented as the Target, and a Japanese *katakana* character was presented as the Test. The participants were instructed to judge whether the Test character corresponded to the single Target character or to one of the four target characters and then press the appropriate button. They used the right index finger to press the "yes" button when the Test character corresponded to the Target character and used the right middle finger to press the "no" button when it did not correspond to the Target character. Because the characters presented as Target and Test were in different Japanese morphograms (i.e., *hiragana* and *katakana*), the participants made their judgments on the basis of the phonetic information conveyed by the characters, not on the basis of their form. In 12 task trials, the one-character Target (1-item) and four-character Target (4-item) conditions were randomized with the same probability. The probabilities of "yes" and "no" responses were controlled to be the same (50%). The interval between Target onset and the subsequent Target onset was set to 35 s. Only a fixation cross was presented during the interval and delay period. In addition, auditory cues (1000- and 800-Hz pure tones of 100-ms duration) were presented at the Target and Test onsets, respectively. The behavioral results for the WM task are shown in the Supplementary Materials (Table S1).

Finger tapping task

A simple finger-tapping task was used as a sensorimotor task. It was similar to that used in previous NIRS studies (Sato et al., 2005, 2006). The participants were instructed to watch a white fixation cross shown at the center of the screen. During the task period, the color of one half of the horizontal bar turned yellow at a rate of 3 Hz, and the participants were asked to tap their fingers in synchrony with the blinking timing of the horizontal bar in the fixation cross. During each task period, the fingers of one hand were repeatedly placed on the tip of the thumb in the following order: forefinger–second finger–third finger–little finger–third finger–second finger–forefinger. Each task period lasted for 15 s, and the interval between task periods was 25 s. The experiment consisted of ten task periods with the order of the right and left finger conditions alternating.

Data acquisition

NIRS

We used an optical topography system (ETG-4000, Hitachi Medical Corporation, Japan) for the NIRS measurements. The light sources consisted of continuous laser diodes with two wavelengths, 695 and 830 nm. The transmitted light (detected with avalanche photodiodes) was sampled every 100 ms. The optical fibers used for both the irradiating and detecting lights, including the points where the optical fibers were attached to the skin, are referred to as 'probes' ('source probes' and 'detecting probes').

We used two types of probe holders with different probe configurations (Figs. 1B–E). "Probe-1" was a conventional probe holder with 30-mm spacing between the source and detector (S–D) probes, covering a wide area mainly on the forehead. Seventeen light sources

and 16 detectors were arrayed in a 3 × 11 lattice pattern with 30-mm separation, forming 52 measurement positions (defined as 'channels'), one for each S–D pair (Fig. 1B). The probes were embedded in a soft silicon holder that was placed on the participant's forehead (Fig. 1D left). The probe holder is visible in a normal T1-weighted image (Fig. 1D right), so the spatial position of the probes could be registered in the MRI data. It was used for 15 participants (11 males and 4 females, mean age of 35.5, age range of 27–45).

"Probe-2" was a custom-made probe holder with both conventional S–D distance (30-mm) probes (Fig. 1C) and shorter S–D distance (15-mm) probes (not shown). The data collected with the shorter S–D distance probes were not used in the present study because it focuses on the quality of NIRS-Hb signals obtained with a conventional S–D distance of 30 mm. The shorter S–D distance results will be reported elsewhere. For the 30-mm measurements, 6 light sources and 6 detectors were arrayed in a lattice pattern, forming 14 measurement positions. Twenty-one vitamin E tablets were also attached to the probe holder as markers to identify the positions of the NIRS channels in the MRI images (Figs. 1C, E). This probe was used for the other 12 male participants (mean age of 37.7, age range of 30–48).

Data from both Probe-1 and the conventional S–D probes in Probe-2 were used for our main objective, i.e., determining the validity of activation signals in those areas commonly measured with both probe holders. In addition, we used all channel data from Probe-1 to examine the spatial characteristics of the correlation between the NIRS and BOLD signals. The experimental setup and task sequences were common to both probe holders.

Magnetic resonance imaging (MRI)

The MR imaging was performed with a Philips Achieva 3.0 T TX system (Philips Medical Systems, The Netherlands) with a 32-channel SENSE head coil. A total of 180 and 175 T2*-weighted gradient-echo echo-planar images (EPIs) were acquired while a participant underwent a single session of the TAP and WM tasks, respectively. A single EPI volume consisted of 35 3-mm-thick axial slices interspaced by a 1-mm gap, covering the entire brain. Other imaging parameters included TR = 2500 ms, TE = 30 ms, flip angle = 80°, field of view = 192 × 192 mm, and matrix = 64 × 64. Following the functional imaging, a B0 field map was acquired in the same head position (35 4-mm-thick axial slices, TR = 20 ms, TEs = 2.3/4.6 ms), which was later used to reduce image distortion caused by inhomogeneity in the magnetic field. Further, for anatomical identification of activated regions in the brain, a T1-weighted structural image was obtained (field of view = 250 × 250 mm, in-place resolution = 1.1 × 1.1 mm, 301 contiguous sagittal slices of 0.6 mm thickness, TR = 7.4 ms, TE = 3.4 ms, flip angle = 8°).

Laser-Doppler flowmetry

The skin blood flow was measured with a laser Doppler flowmeter (Microflow DSP, Oxford Optronix Ltd., UK) that had two surface probes. One was attached to the skin, centered between the eyebrows (channel 1), and the other was attached to the left temple (channel 2). The LDF analog output was converted into a digital signal with an analog-to-digital (A/D) converter (NR-2000, Keyence Corporation, Japan).

Data analysis

We primarily used MATLAB (The MathWorks, Inc., U.S.A.) for the analysis. A flow chart of the data analysis is shown in Fig. 2.

Registration of NIRS measurement area in structural MRI

To compare the temporal changes in the Hb signals from the NIRS measurements and in the BOLD signals from the fMRI measurements, we registered the NIRS channel positions in the structural MRI (Fig. 2A).

First, the T1-weighted image was coregistered to the mean functional image by using Statistical Parametric Mapping Software 8 (SPM8; Institute of Neurology, University College London, UK; run on MATLAB) (Step A-1). Then, the coordinates of the S–D pairs were obtained for each participant and each NIRS channel (NIRS-ch). This process (Step A-2) was done by manually identifying the positions of the sources and detectors in the probe holder shown in the T1-weighted image for Probe-1. For Probe-2, the position of each marker (a vitamin E tablet) was first manually identified in the T1-weighted image, and the coordinates of the sources and detectors were calculated from the marker positions. In parallel with these processes, the T1-weighted image was segmented using the "New Segment" routine in SPM8 into six components: gray matter, white matter, cerebrospinal fluid, soft tissue, bone, and air (Step A-3). Using the spatial information for the NIRS channels and the segmented MR images, we conducted two kinds of anatomical identifications of NIRS channels.

(1) Determination of BA numbers for NIRS channels
We determined the BA number for each NIRS channel to clarify the anatomical information. After we normalized the T1-weighted image with channel-position information into the Montreal Neurological Institute (MNI) space using SPM (Step A-4), we defined the measurement position for each NIRS channel as the midpoint between the source and detector. Next, we projected each measurement position onto a template of gray matter ("grey.nii" in spm8\tmp) and obtained the MNI coordinates for each participant (Step A-6). The projection point was defined as the voxel closest to the measurement position that showed an intensity greater than 0.5 in the template image. The 0.5 intensity threshold means that the voxel can be classified as being in the gray matter with a probability greater than 50% because the intensity of the template image for tissue ranges from 0 to 1 with the limitation that the sum of the intensities of all segmentation templates is 1. From the MNI coordinates, the BA number was determined (Rorden and Brett, 2000) for each projection point (Step A-8).

(2) Determination of spatial sphere for analysis (SFA)
We defined a spatial sphere of fMRI voxels for analysis (SFA), which corresponded to the measurement area for each NIRS channel. In step A-5, the photon-path-distribution function $P_n(x, y, z)$ at position r (x,y,z) was derived using the S–D coordinates and Eq. (1) (Feng et al., 1995; Sassaroli et al., 2006):

$$P_n(x,y,z) = \frac{z^2 \exp\left(-k\left[\left(x^2+y^2+z^2\right)^{1/2} + \left\{(d-x)^2+y^2+z^2\right\}^{1/2}\right]\right)}{\left(x^2+y^2+z^2\right)^{3/2}\left\{(d-x)^2+y^2+z^2\right\}^{3/2}} \times \left\{k\left(x^2+y^2+z^2\right)^{1/2}+1\right\}\left[k\left\{(d-x)^2+y^2+z^2\right\}^{1/2}+1\right] \quad (1)$$

Function P_n (x, y, z) represents the proportional probability of the photon path distribution, where a photon irradiated at (0, 0, 0) and detected at $(d, 0, 0)$ crosses the generic point (x, y, z). The effective attenuation coefficient k is defined as $\sqrt{3\mu_a\mu_s'}$, where absorption coefficient μ_a is 0.011 mm^{-1} and the reduced scattering coefficient of tissue μ_s' is 1.2 mm^{-1} (Sassaroli et al., 2006). We set a threshold of P_n $(x, y, z) > 10^{-10}$ for the SFA (Step A-7 and Fig. 3A). We used the segmented MR images to identify the gray matter (GM) and soft tissue (ST) voxels in the SFA (Step A-9) for the subsequent analysis of both the cortical BOLD and extracranial BOLD signals. We focused on the GM region, which shows the cortical responses to experimental tasks, and on the ST region, which is presumed to show the skin blood flow changes (Fig. 3B).

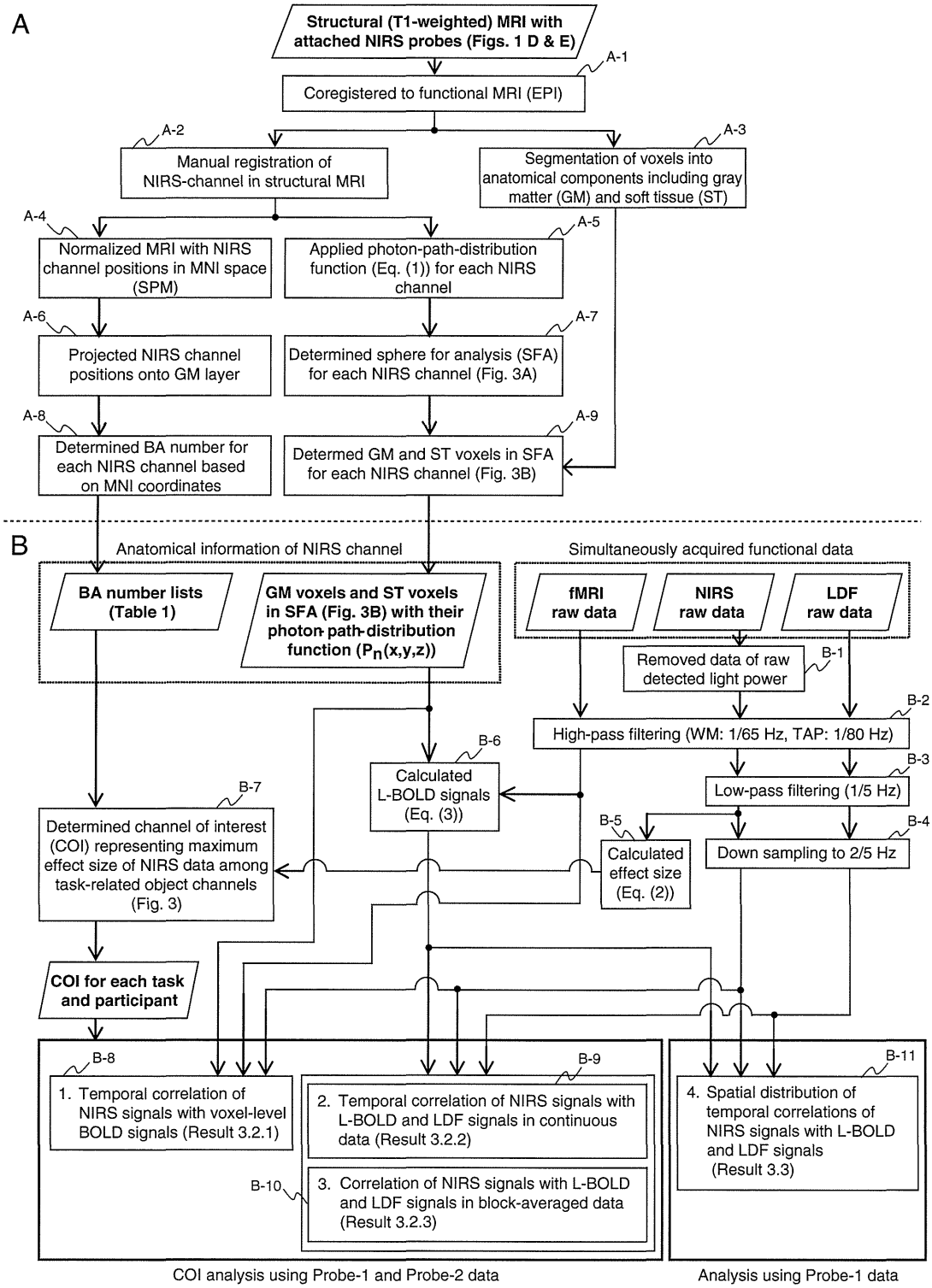


Fig. 2. Flowchart of basic analysis. Parallelograms represent input/output data, and rectangles represent processes. Each process is numbered as a step (A-1 to A-9 and B-1 to B-11) corresponding to explanation in text.

Functional signal preprocessing

The preprocessing of the simultaneously acquired fMRI, NIRS and LDF raw data are diagrammed in the lower right part of Fig. 2B.

NIRS. A portion of the NIRS signal preprocessing was performed using the plug-in-based analysis software Platform for Optical Topography Analysis Tools (developed by Hitachi, CRL; run on MATLAB). Prior to analysis, the data for channels with low detected light power were removed (Step B-1). The threshold for removal was set to a total gain of over 6667, which corresponds to the transmitted light being less than $10^{-9}\times$ the irradiation power. The main reason for the low detected light was probably poor probe attachment due to hair.

For each NIRS channel, the optical data for two wavelengths were transformed into a time series of Hb signals (oxy-Hb and deoxy-Hb

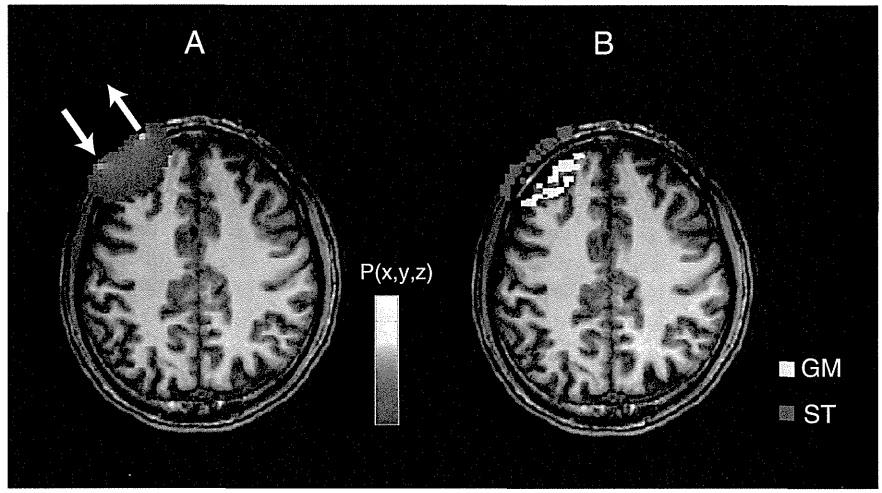


Fig. 3. Example images of spatial sphere for analysis in NIRS-channel 28 of Probe-1. (A) Proportional probability of photon path distribution as determined by simulation using photon-path-distribution function. (B) Voxels in gray matter (yellow) and those in soft tissue (red) in sphere as determined by data segmentation.

signals) on the basis of the modified Beer-Lambert law (Delpy et al., 1988; Maki et al., 1995). These signals were expressed as the product of the changes in hemoglobin concentration (mM) and optical path length (mm) in the activation region (effective optical path length). A high-pass filter was applied to the NIRS-Hb signals to remove low-frequency drift, which was longer than the twofold duration of one task sequence (WM: 1/65 Hz, TAP: 1/80 Hz) (Step B-2), and a low-pass filter was applied to the NIRS-Hb signals to remove components with a frequency higher than the sampling period in the fMRI scanning (2500 ms) (Step B-3). Finally, the filtered NIRS-Hb signals were down-sampled to the fMRI scanning frequency for comparison with the fMRI data (Step B-4).

After the filtering process, the effect size of the task-related signal change was calculated for each participant and each task (Eq. (2) and Step B-5) to determine the channel of interest. The effect size (Cohen's *d*) is similar to the contrast-to-noise ratio (CNR) used by Cui et al. (2011), which is the amplitude difference between the mean of the activation period (*mean(Act)*) and that of the control period (*mean(Con)*) divided by the pooled standard deviation (SD, σ) (Cohen, 1992):

$$d = \frac{mean(Act) - mean(Con)}{\sigma}$$
$$\sigma = \sqrt{\frac{\sum_{i}^{n_1}(y_{1i}-\bar{y}_1)^2 + \sum_{i}^{n_2}(y_{2i}-\bar{y}_2)^2\sqrt{2}}{n_1+n_2-2}}, \quad (2)$$

where y_1 and y_2 are the vectors of the raw NIRS-Hb signals in the activation and control periods, respectively. The activation period was commonly set to the interval from 5 s after onset to the task end (WM: 3.5 s, TAP: 10 s). The control period was set to the same duration (WM: 3.5 s, TAP: 10 s) before task onset.

fMRI. MR data preprocessing was performed using statistical parametric mapping software 8 (SPM8; Institute of Neurology, University College London, UK; run on MATLAB). First, raw functional images were unwarped by using the field map acquired for the participant and were spatially realigned to the mean image calculated from the entire data. We then applied to the BOLD time series the same band-pass filter used in the NIRS-Hb signal preprocessing (Step B-2). Unlike the standard preprocessing procedure, functional images were not normalized into the standard space but were instead analyzed in the participant's native space.

Using the anatomical information of NIRS channels obtained in step A-9, we calculated the photon-path-distribution average BOLD signal for each NIRS channel (Eq. (3)) as the layer BOLD (L-BOLD) signal for the GM and ST layers using a modified version of (Sassaroli et al., 2006)

$$\mathrm{L-BOLD(t)} = \frac{\int_V \mathrm{BOLD}_{raw}(x,y,z,t)P_n(x,y,z)dxdydz}{\int_V P_n(x,y,z)dxdydz}, \quad (3)$$

where L-BOLD(t) is the layer BOLD signal averaged using a weight given by the photon-path-distribution function, and BOLD_{raw} (x,y,z,t) is the raw BOLD signal for each voxel at time t and position (x,y,z) (Step B-6).

LDF. The same preprocessing as for the NIRS-Hb signals was used for the LDF signals. That is, the LDF signals were filtered using the same parameters and down-sampled to match the fMRI scanning frequency (Steps B-2, B-3 and B-4).

Selection of channel of interest (COI)

First, we defined the NIRS channels in BA 46 as the object channels for the WM task and those in BAs 1, 2, 3, 4, 5, 6, and 40 as the object channels for the TAP task (Fig. 4A). Next, we selected as the channel of interest (COI) for each task condition the object channel with the maximum effect size (mean effect size of oxy-Hb signal and deoxy-Hb signal) on the condition that the effect size exceeded 0.2 (Step B-7) (Fig. 4B). The 0.2 threshold corresponds to the "small" effect size (Cohen, 1992). Signal responses to the 4-item condition were used in the WM task, and those to the contralateral-hand tapping task were used in the TAP task. We focus on the COI for detailed analysis because our interest here is to investigate task-related NIRS-Hb signals. If we did not focus on the COI, the correlation with BOLD signals could be contaminated by the effects of noisy or non-activated channels.

Correlation analysis in COI

Using the fMRI signals in the SFA, we conducted three types of correlation analyses for COI using the data for both Probe-1 and -2.

(1) Correlation with voxel-level BOLD signals
Pearson's correlation coefficient (r) between the NIRS-Hb signal and the BOLD signal was calculated for each voxel in the GM and ST layers of the SFA (Step B-8). The significance of the correlation coefficient was determined using a false