ORIGINAL ARTICLE

Near-Infrared Spectroscopy and Imaging for Investigating Stroke Rehabilitation: Test-Retest Reliability and Review of the Literature

Gary Strangman, PhD, Richard Goldstein, PhD, Scott L. Rauch, MD, Joel Stein, MD

ABSTRACT. Strangman G, Goldstein R, Rauch SL, Stein J. Near-infrared spectroscopy and imaging for investigating stroke rehabilitation: test-retest reliability and review of the literature. Arch Phys Med Rehabil 2006;87(12 Suppl 2):S12-9.

Objectives: To review the use of near-infrared spectroscopy (NIRS) in stroke rehabilitation and to evaluate NIRS test-retest reliability within-session on a motor control task commonly used in neuroimaging of stroke recovery.

Design: Cohort study.

Setting: Hospital-based research laboratory.

Participants: Nineteen healthy control subjects (age range, 22−55y).

Interventions: Subjects performed 2 experimental runs of a finger-opposition task in a block-design paradigm (finger opposition alternated with a fixation rest period) while undergoing multichannel NIRS and physiologic monitoring.

Main Outcome Measure: Reliability coefficients (Pearson *r*) for oxyhemoglobin (O_2Hb) and deoxyhemoglobin (HHb) correlated amplitude modulations across measurement channels during individual blocks and block averages.

Results: Correlations between single blocks (ie, 16-s slices of data) exhibited a correlation intercept of .33±.09 for O_2Hb. This value was minimally decreased by increasing lag between compared blocks (slope, −.012; P=.019) but was substantially enhanced by averaging across blocks (within-run slope, .11; between-run slope, .044). Correlations using 64 seconds of data reached 0.6. Results for HHb were virtually identical.

Conclusions: NIRS modulations were repeatable even when comparing very short segments of data. When averaging longer data segments, the test-retest correspondences compared favorably to neuroimaging using other modalities. This suggests that NIRS is a reliable tool for longitudinal stroke rehabilitation and recovery studies.

Key Words: Hemodynamics; Motor skills; Oxyhemoglobins; Rehabilitation; Spectroscopy, near infrared.

NEUROIMAGING OFFERS GREAT promise in revealing both the brain regions and the temporal evolution of the neural activity underlying stroke rehabilitation and recovery. Traditional methods for monitoring brain function—functional magnetic resonance imaging (fMRI), positron emission tomography, electroencephalography, and magnetoencephalography—have provided considerable initial insight into the functional brain changes associated with recovery from stroke.[1] Despite tremendous advances, however, most current technologies still impose strict constraints on the subject or the environment. For example, these techniques still generally involve (1) a confining or restrictive monitoring environment, (2) subject head immobilization, (3) high cost, (4) a nonportable and obtrusive apparatus, and (5) techniques that are invasive and/or use ionizing radiation. As a result, brain monitoring at the bedside, at home, in remote locations, or in tasks that require substantial motion—for example, gait training—is generally precluded.

This gap can be at least partially filled by the methods of near-infrared spectroscopy (NIRS) and imaging (NIRI). NIRS and NIRI are sensitive to changes in brain oxygenation[2] while at the same time they (1) are nonconfining, (2) can be made relatively insensitive to motion, (3) are relatively low cost, (4) can be portable and unobtrusive, and (5) are noninvasive, using non-ionizing radiation. Moreover, diffuse optical methods can be adapted to low power consumption (for use in truly ambulatory or remote monitoring) and can quantify multiple biologically and clinically relevant compounds. Thus, NIRS-based techniques promise to enable a much wider range of clinical and research questions that can be asked and answered using these popular imaging modalities. The main drawbacks to NIRS and NIRI lie in their modest spatial resolution and limited depth penetration.

NIRS and NIRI

Technique summary. A number of detailed reviews have appeared in the literature describing the principles of NIRS,[2,3] so we only briefly describe the basic concepts here. NIRS is based on the same principles as pulse oximetry: shine red or near-infrared light onto the scalp and detect the remitted light some distance away. Measurement of 2 or more colors of introduced light enables the calculation of oxyhemoglobin (O_2Hb), deoxyhemoglobin (HHb), and total hemoglobin concentrations within the probed tissue. Regional changes in oxygenation, in turn, reflect regional changes in brain activity. For brain measurements, source-detector separations typically range from 25 to 50mm, with the resulting measurement being sensitive to, at most, the outermost 10mm of cortex in an adult human. Individual measurements can be made in 1 to 10ms, providing excellent temporal resolution of the hemodynamic responses to brain activity. NIRS-derived hemodynamic measurements closely parallel fMRI findings,[4] including the timing and oxygenation changes expected from a prototypical hemodynamic response function.[5]

From the Division of Psychiatric Neuroscience, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA (Strangman, Rauch); Division of Behavioral and Mental Health, Spaulding Rehabilitation Hospital, Harvard Medical School, Boston, MA (Strangman, Rauch); and Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Harvard Medical School, Boston, MA (Goldstein, Stein).

Supported by the National Institute of Neurological Disorders and Stroke, National Institutes of Health (grant no. K25-NS046554).

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.

Reprint requests to Gary Strangman, Neural Systems Group, Massachusetts General Hospital, 149 13th St, Psychiatry, Rm 2651, Charlestown, MA 02129, e-mail: *strang@nmr.mgh.harvard.edu.*

0003-9993/06/8712S-10982$32.00/0
doi:10.1016/j.apmr.2006.07.269

NIRI (also known as diffuse optical tomography or optical topography) is a straightforward extension of NIRS: multiple, overlapping, NIRS-type measurements are made and then combined (via either topography[6] or tomography[7]) to generate images of spatial variations of absorption and scattering properties of brain tissue. These images can provide information about changes in both hemodynamic responses to brain activity. The diffusion of light through the tissue inherently limits the spatial resolution of the technique to approximately 5 to 10mm in adult human brain imaging.

NIRS in monitoring of gait. The portability and relative motion robustness of NIRS enables monitoring in experimental contexts not previously possible. One such domain is the study of gait, on which 2 NIRS studies have already been conducted with healthy subjects.[8,9]

Miyai et al[8] used diffuse optical topography (ie, NIRI without tomographic reconstruction) to map the activation patterns in healthy and patient populations during treadmill walking. This work showed the possibility of overcoming several obstacles in ambulatory brain imaging by using NIRI. First, it was an existence proof that reliable and artifact-free signals can be recorded from subjects while walking. Second, the results showed evidence of regional activation associated with treadmill walking, suggesting that systemic changes in blood flow and blood pressure do not necessarily overpower signals associated with functional brain modulations.[10] Third, the observed brain activity showed regional specificity to various aspects of walking.

Suzuki et al[9] then investigated the effect of walking and running speed on brain activity patterns. Nine healthy subjects were investigated while walking at 2 rates of ambulation (3km/h, 5km/h) and while running (9km/h), starting with 30 seconds of rest, then 90 seconds of ambulation, and finally another 30 seconds of rest (with ≈10- to 18-s transition periods). They observed hemodynamic changes in bilateral prefrontal and premotor cortex associated with higher locomotor speeds. Medial prefrontal cortex activity (reflected by O_2Hb increases during the task) was most prominent during running, suggesting a role for this region in the control of running relative to walking. Together, these studies show a unique role for NIRS in gait monitoring.

Neuroimaging Applications in Stroke Populations

Non-NIRS in stroke rehabilitation. Although non-NIRS brain imaging studies of stroke rehabilitation are still modest in number and have their challenges (restrictions in terms of scan time, achievable task, and/or patient characteristics), they have made tremendous strides in producing a coherent picture of cortical changes underlying rehabilitative changes.[1,11] Most such studies examine finger tapping, hand tapping, sequential finger opposition, or somatosensory stimulation. Generally speaking, in the early stages of rehabilitation, hemiparetic patients tend to activate relevant neuronal networks more bilaterally (ie, in both the affected and nonaffected hemisphere) when using the paretic limb, either compared with the nonparetic limb or compared with healthy control subjects.[12-16] Later stages of rehabilitation tend to be associated with the more typical contralateral activation pattern when using the affected limb.[17-21]

NIRS in stroke rehabilitation. To date, only 5 studies using NIRS-based techniques have investigated various aspects of recovery or rehabilitation from stroke. One of these investigated a hand-grasp task,[22] 2 were gait studies,[23,24] and 2 were implemented in conjunction with a rehabilitation program: a case study investigating constraint-induced movement therapy[25] and a case series investigating a wide range of rehabilitation tasks including passive, cognitive, and active (including gait) tasks.[26]

NIRS in stroke rehabilitation: hand-grasp task. The hand-grasp study by Kato et al[22] was conducted on 6 chronic stroke patients with minimal hemiparesis and 5 similarly aged control subjects (all right handed). Both NIRI (optical topography) and fMRI (in a separate session) data were collected, and the researchers compared results from the 2 groups and the 2 neuroimaging modalities. Findings were similar to some previous neuroimaging studies of finger tapping in hemiparetic subjects, showing more neural activity in the side of the brain ipsilateral to the moving hand.[11] Interestingly, the NIRS measurement seemed to be more sensitive (or generate more robust activations) on the ipsilateral side relative to fMRI, an effect that has previously been observed elsewhere as well. The implications of this ipsilateral enhancement are currently unknown.

NIRS in stroke rehabilitation: cross-sectional rehabilitation investigations. The first cross-sectional study using NIRS in a rehabilitation setting, and one of the first investigations of NIRS in any sort of rehabilitation context, was by Saitou et al,[26] who investigated 44 hemiplegic patients (>3mo poststroke) and 24 control subjects. Each subject performed a subset of 13 rehabilitation tasks, including head-up tilt, mental calculation, cycle ergometer, reading words, listening to classical music, passive wrist and finger extension, passive movement of affected upper limbs with a pulley, reciprocal extension of knee joints, isotonic extension of the nonparalyzed knee joint, supine hip elevation, facilitated flexion-extension movement of paralyzed leg, repeated stand-up and sit down (stand-up), and free gait. The researchers found significant percentages of subjects exhibiting changes in cerebral blood volume and cerebral oxygen volume over the prefrontal cortex in nearly every task. The results require some interpretive caution, given that the probe was located near the sagittal sinus in controls and not in patients and given that some substantial changes in heart rate and/or blood pressure were observed that can influence NIRS signals. Nevertheless, the work clearly showed the suitability, applicability, and sensitivity of NIRS in a clinical rehabilitation setting.

The second study in a rehabilitation context was a gait-based rehabilitation study by Miyai et al[23] that followed their investigation of gait in healthy controls discussed earlier. Six nonambulatory, severe stroke patients were monitored with NIRS during hemiplegic treadmill gait approximately 3 months poststroke (for 3 subjects, this was their first poststroke gait experience). Subjects alternated between walking at 0.2km/h for 30 seconds and 30-second periods of rest, all under partial body support, either with mechanical assistance in swinging the paretic leg or with a therapist-mediated facilitation of paretic leg swing. The researchers found similar activations as in their previous study of control subjects but enhanced activity of premotor and presupplementary motor cortices, primarily in the unaffected hemisphere. They proposed that these regions play an important role in restoration of gait in patients with severe stroke. The group also found that the therapist-facilitated swing induced stronger regional brain activation than mechanical assistance. Thus, if it can be established that increased activity levels are associated with more rapid improvement of functional capabilities, the NIRS-based technique might eventually be used in conjunction with traditional therapist-based rehabilitation to help guide therapy and gauge efficacy.

Longitudinal NIRS studies. Finally, 2 studies have used NIRS in longitudinal studies of rehabilitation-related tasks. As such, these studies go beyond mapping the product of

rehabilitation and begin to address questions about neurophysiologic changes associated with the process of recovery and rehabilitation.

Miyai et al[24] described another follow-up on their previously described gait studies. In this study, a total of 8 stroke patients were investigated, 4 of whom came from the previous study plus 4 new subjects. All underwent the above-described procedure of NIRS monitoring during gait, but at 2 sessions: approximately 3 months poststroke and again approximately 2 months later. Once again, they found medial primary sensorimotor cortex activity associated with walking, with greater amplitude in the unaffected hemisphere, as well as premotor and supplementary motor activity. On the second session, subjects exhibited an "improvement" in the laterality of activation; that is, activity in the unaffected hemisphere decreased and that in the affected hemisphere increased. This improvement mirrors findings in recovery of hand motor function using serial fMRI scans,[27,28] thereby generalizing the previous hand-function findings to recovery of gait.

Finally and most recently, Park et al[25] describe a longitudinal case study using NIRI serially during 10 consecutive days of constraint-induced movement therapy. The subject was a 73-year-old, right-hemiparetic patient with a subcortical stroke 4 months before the study. Daily NIRI images were obtained while the poststroke subject and a healthy control participant performed a functional key-turning task. In the poststroke subject, again there was a trend toward increasing laterality (ie, increasing activity in the contralateral hemisphere as compared with the ipsilateral hemisphere). In contrast, the control participant exhibited no such changes. At the same time, for the poststroke subject, O_2Hb exhibited a decreasing trend in activation magnitude across the sessions. These findings were accompanied by marked improvements in various functional clinical scores (Wolf Motor Function Test, Motor Activity Log, functional key grip test).

Although preliminary in nature, these studies show the applicability of NIRS in a rehabilitation context, and the case study show the ability of NIRS-based techniques to monitor brain activity associated with functional tasks throughout a rehabilitation program.

NIRS Test-Retest Reliability

Clearly, NIRS is just beginning to be used to investigate neural processes associated with (re)learning to perform a task. Several important and unanswered questions are fundamental to this application, however. One issue concerns test-retest reliability. A number of test-retest fMRI studies have shown that the spatial extents of hemodynamic responses to a given task are reproducible across sessions, days, and even months.[29-34] The spatial reproducibility is not perfect, however, even in the simple motor tasks used in investigations of stroke rehabilitation.[34]

Test-retest reliability of the hemodynamic response is clearly important for interpreting learning and rehabilitation-related changes in the hemodynamic response. This is illustrated in figure 1, where it is clear that the reproducibility of amplitude modulations is particularly important. (Note that learning-related amplitude modulations may appear as linearly decreasing, linearly increasing, or nonlinear in nature.) Two previous studies have reported test-retest data from functional brain activation experiments using NIRS. The first described between-session test-retest reliability of amplitude modulations.[35] Results suggested reasonable repeatability (R range, .42−.87), although the retest intervals were widely varied and the sample size was sufficiently small (N=5) so as to preclude population inferences. A more recent NIRS study examined the between-session reliability of the Student t scores computed from the NIRS data in an event-related design visual checkerboard task.[36] This study found good channel quantity and channel location reliability between subjects (intraclass correlation coefficient, ≈0.8), with weaker reliability on the individual subject level. Neither of these studies, however, examined reliability *within* a scanning session. As an important initial step toward careful application of NIRS and NIRI in rehabilitation contexts, we therefore sought to evaluate the test-retest reliability of NIRS-derived O_2Hb and HHb concentration modulations on a motor control task typical of those used in neuroimaging studies of stroke rehabilitation (complex, sequential finger opposition), where progressive changes in brain activation during a session would be important. Specifically, we evaluated the test-retest reliability of single blocks of a motor task and then investigated the effect of lag between retests, as well as the effect of averaging on NIRS test-retest reliability.

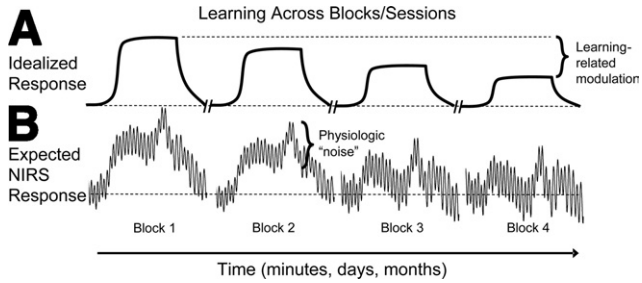

Fig 1. Schematic of neurophysiologic changes expected from a longitudinal rehabilitation and learning study. (A) Idealized response from a region exhibiting linear decreases in activity with over time, assuming a block-design experimental paradigm. Depending on the task or brain region, linear increases or nonlinear (including nonmonotonic) changes may be found. (B) Depiction of an example NIRS response (in arbitrary optical density units) reflecting the idealized response in (A) plus typical physiologic fluctuation observed in NIRS signals (fast oscillations equals cardiac response; slower oscillations equals respiration and/or Mayer waves[48]). Note that the learning-related modulation is smaller than the overall task-related modulation and hence will be harder to detect.

METHODS

Participants

Nineteen healthy subjects (6 men, 13 women; mean age, 32±10y; range, 22−55y), all right handed,[37] underwent near-infrared optical recording. The study was approved by the institutional review board at the Massachusetts General Hospital, where the experiments were performed, and all subjects gave their prior, written informed consent.

NIRI Recording Procedure

We performed near-infrared recordings using a custom-built, continuous-wave NIRS instrument[a] with 16 dual-color laser sources and 32 detectors coupled to 2.7-mm silicon fiber bundles.[b] The bandwidth for each source-detector pair was 3Hz, sampled at 10Hz. The flexible fiber-holder probe was positioned centered over the Cz location in the International 10−20 system[38] and reached laterally from 2 to 4cm lateral to C3 and C4, depending on the subject's head size (fig 2A). We recorded pulse oximetry,[c] respirometry, and pressure-pulse,[d] as well as stimulus trigger signals synchronized with the NIRS data.

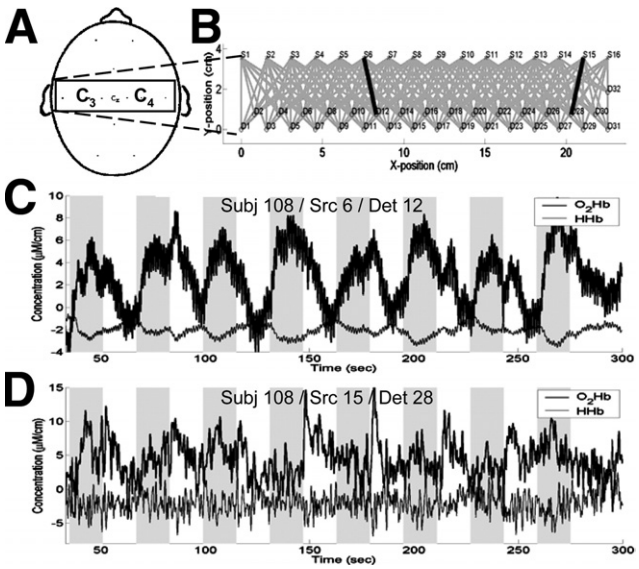

Fig 2. (A) NIRI probe placement and (B) and geometry, where S is the source position (16 total) and D is the detector position (32 total). (C) Example O_2Hb (heavy line) and HHb (light line) responses to 1 run of task performance (shaded regions are finger opposition) for source 6 and detector 12 (left black line in [B], near C3). Notice the substantial task-related modulation. (D) A similar measurement for source 15 and detector 29 (right black line in [B], lateral to C4). Notice the substantial lack of task-related modulation in this region. Abbreviations: Det, (optical) detector; Src, (laser) source; Subj, subject.

Task

We asked each subject to perform a complex finger-to-thumb opposition task (little-index-ring-middle, repeat, paced at 2.5Hz by a blinking asterisk). A block-design was used: 35 seconds of rest (fixation cross), then seven 16-second periods of activity alternating with eight 16-second periods of rest, ending with an extra 25-second period of rest (total, 300s). This task is similar to that used in various fMRI studies of stroke and stroke recovery.[17,18,39,40] Each subject performed 2 such runs during a single scanning session, with an unrelated, 6-minute event-related motor task occurring in between, spanning approximately 15 minutes—a time period appropriate for evaluating easily performed or learned tasks. After the blocked and event-related motor control tasks, subjects also performed resting baseline measurements. We discuss only the block-design results here.

Data Analysis

We converted optical data from individual source-detector pairs preprocessed (high-pass filter, .017Hz; low-pass filter, 3Hz) and wavelength-pairs to O_2Hb and HHb concentrations using the modified Beer-Lambert law, a standard approach.[41,42] Instead of estimating the differential path-length factor (required by the modified Beer-Lambert law), we computed concentrations in units of μM/cm.

For each subject, we pruned source-detector pairs (out of a possible 1024) to remove low- or poor-quality signals (signal-to-noise ratio [SNR], <5; or if $>0.5\%$ of the O_2Hb or HHb values exceeded 100μM/cm—ie, not physiologic). To be consistent with previous fMRI test-retest studies, we also pruned channels from a given run based on a statistical comparison of task time points (overbars in fig 3) versus rest time points (underbars in fig 3). Channels with no significant modulation between task and rest were also excluded (Student t test, $P>.05$; Bonferroni-adjusted for the number of tests per subject). We did not prune any full or partial runs or blocks suspected of motion or physiologic artifact because (1) such a procedure can be highly subjective, and (2) no pruning provides a meaningful lower bound on test-retest reliability.

Next, we computed amplitude modulations in [O_2Hb] and [HHb] by subtracting the average signal level (O_2Hb or HHb) during the last half of a task period (overbars) from the average signal recorded during the last half of the immediately following rest period (underbars; see fig 3). The resulting 8-second lag from task onset–offset accounted for delays in the hemodynamic response to task changes.

We performed test-retest evaluations via Pearson correlation coefficients as recommended by Rousson et al.[43] Pearson r values were computed across all surviving measurements for a given subject, block by block. Using a linear mixed effects model,[44] we then modeled the effect of lag on these test-retest correlations using 3 different lags: adjacent (comparing amplitude-modulations from adjacent blocks; 1 vs 2, 2 vs 3; 3 vs 4,

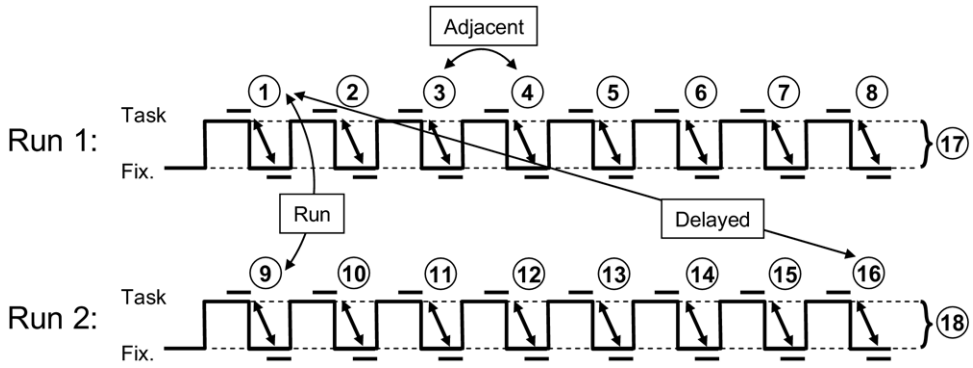

Fig 3. The experimental paradigm and illustration of test-retest comparisons. Subjects performed two 5-minute–long runs, each including 8 blocks of finger-opposition task alternated with fixation rest periods. Block-by-block modulation amplitudes were computed by averaging O_2Hb signals (and, separately, HHb signals) for the last 8 seconds of each 16-second task or rest period (short black overbars and underbars, respectively) and subtracting these 2 values (task-rest; thick angled arrows). Test-retest evaluations were performed for 3 lags: adjacent block-pairs (eg, 1 vs 2, 2 vs 3, etc; lag=1), for corresponding blocks across runs (eg, 1 vs 9, 2 vs 10, etc; lag=8), and delayed (block 1 vs 16; lag=15). We also evaluated the effect of averaging the modulation amplitudes across pairs (eg, 1+2 vs 3+4), quadruples (eg, 1+2+3+4 vs 5+6+7+8), and entire runs (17 vs 18).

etc; lag, 1 block); run (1 vs 9, 2 vs 10, etc; lag, 8 blocks); and delayed (1 vs 16; lag, 15 blocks) (see fig 3). Lag was modeled as a fixed effect, but each subject was allowed to have his or her own intercept. Conceptually, we wanted to evaluate how test-retest correlations changed over time, expecting such mean correlations to differ across subjects and to diminish somewhat with longer lags. All mixed-effects model parameters were computed via restricted maximum likelihood in the R statistical environment.[e]

Because the lag analysis was based only on single-block data, which we assumed would be noisy, we followed up the lag analysis with 2 analyses examining the effect of averaging on test-retest reliability. To do so, we averaged pairs, quadruples, and entire runs' worth (octuples) of amplitude modulations. We again modeled the effect of averaging via linear mixed-model analysis, separately for the adjacent and run lags described above.

RESULTS

Despite the general motion robustness of NIRS, data from 1 subject, who had substantial difficulty sitting still during the experiment, had to be excluded because of several motion-related artifacts (>10% change in raw signal in <2s). Data from 3 additional subjects had 2 or fewer channels that survived the SNR, amplitude, and statistical pruning process. Thus, we analyzed 15 subjects, for each of whom we had 2 full runs of more than 3 channels. From these, we obtained a mean ± standard deviation of 52±30 O_2Hb measurements per subject (total, 785; range, 12−116), and 36±26 HHb measurements per subject (total, 536; range, 4−91). Pearson correlation coefficients (r) were first computed across channels within-subject at multiple lags. A histogram of all block-pair–by–block-pair correlation coefficients (combining O_2Hb and HHb at all lags) appears in figure 4. Note the bulk of the distribution lies around +0.5, indicating substantial block-to-block reproducibility. Negative correlations arise from, for example, block-to-block physiologic variability, motion artifacts, or other unknown sources.

In the lag analysis for [O_2Hb], we found a significant R intercept of .33 ($F_{1,329}$=49.8, P<.001) and a significant slope of −.012 ($F_{1,329}$=49.8, P=.019). Thus, while the test-retest reliability significantly decreased with temporal separation of the blocks, this effect was small. The [HHb] findings were nearly identical: intercept of .33 ($F_{1,329}$=49.1, P<.001) and slope of −.017 ($F_{1,329}$=10.5, P=.013). Average Pearson r values at each lag appear in figure 5. Individual subjects had significantly different intercepts from one another, emphasizing significant individual variability in our test-retest evaluation (variance, .12; 95% confidence interval, .07−.21; residual variance, .39).

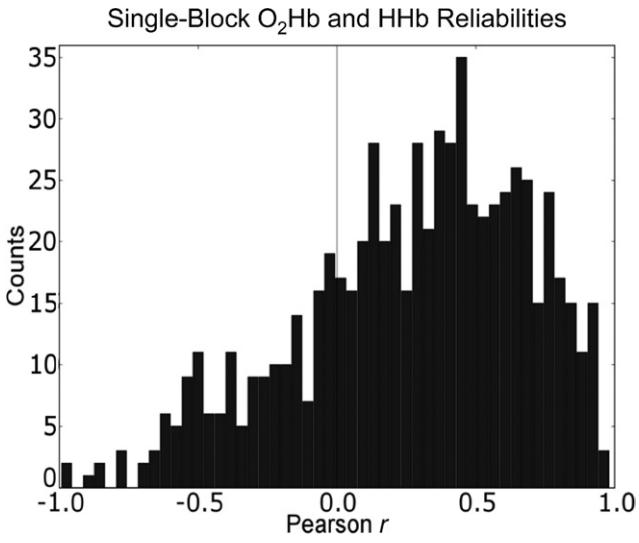

Fig 4. Histogram of single-block test-retest reliability coefficients (Pearson r values) for adjacent (lag=1) pairs, combined across O_2Hb and HHb.

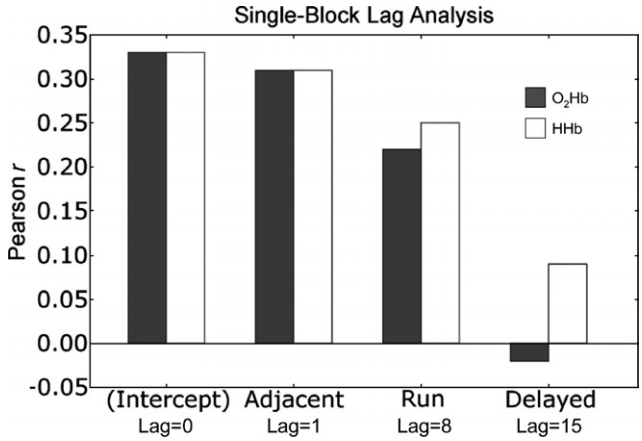

Fig 5. Effect of lag on Pearson product-moments for single-block test-retest reliability. Mean O_2Hb and HHb Pearson r coefficients are shown for test-retest repetitions of adjacent (lag=1 block; N=252), run (lag=8 blocks; N=144), and delayed (lag=15 blocks, N=18) (see text). Note the estimates for longer lags are based on progressively fewer data points and hence include substantially more variability.

The above findings place a rough lower bound on single-block test-retest reliability of approximately .32 (range, −1 to 1). We next investigated the effect of block-averaging on both the adjacent test-retest comparisons, as well as the run-to-run comparisons. For adjacent block-averages, the [O_2Hb] Pearson r intercept was .22 ($F_{1,314}$=60.0, P<.001) with a significant slope of .11 ($F_{1,314}$=21.2, P<.001); for [HHb], the intercept was .21 ($F_{1,314}$=108.4, P<.001) and slope was .13 ($F_{1,314}$=29.1, P<.001). The effect of averaging, therefore, was quite substantial: each block included in the average boosted the correlation coefficient by an average of about .12.

When comparing block-averaged data across runs instead of within runs (our run averaging analysis), the Pearson r intercept for [O_2Hb] was .23 ($F_{1,329}$=35.6, P<.001) with a significant slope of .044 ($F_{1,329}$=16.8, P=.000); for [HHb], the intercept was .20 ($F_{1,329}$=33.8, P<.001) and slope was .058 ($F_{1,329}$=28.2, P<.001). Thus, averaging significantly improved test-retest reliability across runs as well, providing approximately half the benefit found within run. Average Pearson r values for each averaging level (plus the model intercepts) are shown graphically in figure 6. Note that adjacent and run slopes were positive and, respectively, about 10 and about 4 times larger than the (negative) slope for lag, suggesting that averaging could more than compensate for reliability degradation associated with lag.

DISCUSSION

We examined test-retest reliability of NIRS-derived [O_2Hb] and [HHb] modulations observed within a single scanning session while subjects performed a motor control task typical for stroke recovery neuroimaging studies. Even with single-

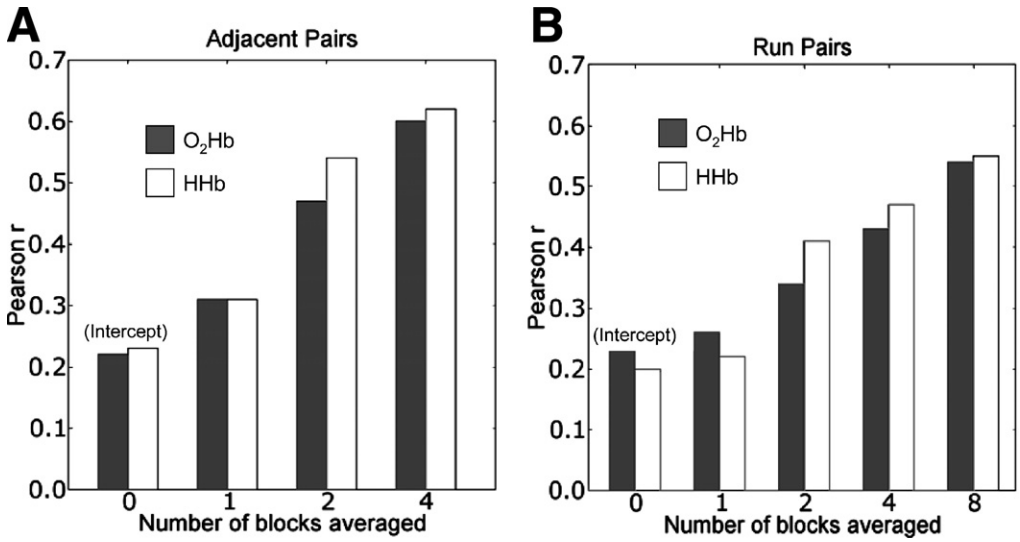

Fig 6. Effect of averaging on Pearson product-moments for NIRS test-retest reliability. Mean O_2Hb and HHb Pearson *r* coefficients are shown for (A) adjacent comparisons with 1-, 2-, and 4-block averages (along with the mixed-effect model intercept) and (B) run-to-run comparisons, with 1-, 2-, 4-, and 8-block averages (along with the intercept).

block data (16-s worth), we found significant reproducibility ($r \approx 0.3$) across all measures, with notable consistency across [O_2Hb] and [HHb]. This reliability was dramatically increased with block-averaging.

It is important to carefully compare the present findings with previous neuroimaging test-retest studies. Prior studies[30,32,34,45] of motor test-retest reliability in fMRI have observed good reliability in neuronal activation (generally ≈0.5 to 0.6) when performing identical tasks within-subject and within-day. Although these numbers are considerably larger than our *r* equal to about 0.3, direct comparison of these numbers is misleading. First and most saliently, the smallest unit of measure in the previous studies was an entire experimental run (data range, 140−512s). In contrast, our minimal unit consisted of 16 seconds of NIRS data. We elected to use such a minimal unit of measure in part because the stamina of stroke subjects is often reduced, which may limit the number of blocks that may be obtained in a single session. An appropriate direct comparison between our results and the fMRI studies would be our 8-block average test-retest results (ie, 64s of data). In this context, the NIRS results are quite comparable with those found with fMRI (≈0.5–0.6) (see fig 6).

Second, the primary test-retest measure in the previous fMRI studies was the consistency of spatial activity patterns, rather than the brain activity modulation amplitudes investigated here (although 1 study[45] also examined activation amplitudes). We focused on amplitudes instead of spatial distributions because (1) amplitude modulations are a critical measurement for investigating rehabilitation-related changes in brain activity (see fig 1) and (2) spatial resolution for NIRS is limited and still undercharacterized because of considerable, ongoing research on tomographic reconstruction. There is, in fact, no particular reason to assume that spatial reliability and amplitude reliability would necessarily match quantitatively, nor necessarily track one another qualitatively. Nevertheless, when comparing like amounts of data, our amplitude-based reliability measures do appear to match fMRI spatial-based reliability measures even quantitatively.

Finally, we noted significant interindividual variability in all analyses. This has been observed in a previous NIRS study of intersession reliability[36] as well as other hemodynamics-based neuroimaging modalities, with many possible causes including differences in baseline physiology, task performance, attention, motivation, and difficulty of the task for the subject. Additional work to illuminate the nature and source of such individual variability would certainly be useful to future clinical investigations of rehabilitation, where the individual subject is of primary importance.

Study Limitations

The present study examined test-retest reliability only *within* a scanning session, over a span of approximately 15 minutes. Although task-learning can be achieved in this time frame for a limited set of tasks, many rehabilitation-related changes in performance occur over the much longer time scales of days, weeks, months, and even years. Monitoring neural changes over longer time scales requires removal and replacement of the NIRS probe, which will likely affect the test-retest reliability of the technique. Investigation of between-session test-retest reliability in NIRS monitoring, therefore, remains an important area of exploration.

Our statistical models only allowed subject intercepts to be a random effect, not subject slopes. In addition, we did not use any other predictors of individual subject intercepts (age, task performance accuracy, physiologic variation). Given the significant variability across subjects, these variables should be investigated further to better understand and model the source(s) of individual subject variability. Moreover, we did not systematically investigate the effect of our statistical thresholding procedure. Although the selected procedure was comparable with the thresholding applied in fMRI test-retest investigations, one should consider the effect of different thresholds (and indeed different SNR pruning) on test-retest reliability estimates.

Finally, we were not able to provide an explanation for the pattern of change in correlations across lags. Intuitively, one might presume that physiologic states are quasi-stationary and hence would influence neighboring blocks more similarly than blocks at longer temporal separations. However, such a hypothesis remains to be investigated, preferably with many runs per session and with sessions spanning multiple days.

CONCLUSIONS

The results reported here strongly suggest that NIRS-based techniques exhibit test-retest reliabilities comparable with

fMRI when repeatedly examining simple motor tasks. Other fMRI work[32,46,47] has suggested that these test-retest reliability figures generalize to other, nonmotor, domains as well. Thus, insofar as the activated and modulated brain region lies in superficial cortex (ie, accessible to NIRS measurements), within-session NIRS-based measurement reliability—even without physiologic or motion-artifact filtering—should provide retest reliabilities comparable with fMRI. This is particularly encouraging given that a standard step in fMRI data analysis is to try to compensate for motion artifacts. Because neurophysiologic responses to task demands show within- and particularly between-subject variability, longitudinal rehabilitation investigations using NIRS (or other techniques) will continue to require interpretive caution. However, we find that NIRS-derived amplitude modulations associated with task performance are as reliable as spatial maps using fMRI and sufficiently so for longitudinal investigations of stroke rehabilitation and recovery.

In sum, NIRS techniques are still quite new in the domain of stroke (or in fact any) rehabilitation research, but they have already begun to reveal patterns of neural activity associated with gait and various in situ rehabilitation tasks, something that would have been difficult or impossible to acquire with conventional neuroimaging methods. As with any new methodology, caution must be exercised both in applying the technique and (per these test-retest findings) in the data interpretation, particularly with regard to longitudinal data. Given appropriate attention to such cautions, however, NIRS techniques appear well positioned to assess neurophysiologic responses in a wide range of new tasks in rehabilitation research. By also providing novel information (details about blood oxygenation and volume), such techniques should nicely complement existing neuroimaging techniques for investigating the brain-based consequences of poststroke rehabilitation and recovery.

Acknowledgments: We thank Michelle Rotman, Stephen Kosslyn, PhD, and Quan Zhang, PhD, for their respective assistance and support in this work.

References

1. Calautti C, Baron JC. Functional neuroimaging studies of motor recovery after stroke in adults: a review. Stroke 2003;34:1553-66.
2. Strangman G, Boas DA, Sutton JP. Noninvasive brain imaging using near infrared light. Biological Psychiatry 2002;52:679-93.
3. Villringer A, Chance B. Non-invasive optical spectroscopy and imaging of human brain function. Trends Neurosci 1997;20: 435-42.
4. Strangman G, Culver JC, Thompson JH, Boas DB. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. Neuroimage 2002;17:719-31.
5. Jasdzewski G, Strangman G, Wagner J, Kwong KK, Poldrack RA, Boas DA. Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. Neuroimage 2003;20:479-88.
6. Kawaguchi F, Ichikawa N, Fujiwara N, Yamashita Y, Kawasaki S. Clinically available optical topography system. Hitachi Rev 2001;50:18-22.
7. Arridge SR. Optical tomography in medical imaging. Inverse Problems 1999;15:R41-93.
8. Miyai I, Tanabe HC, Sase I, et al. Cortical mapping of gait in humans: a near-infrared spectroscopic topography study. Neuroimage 2001;14:1186-92.
9. Suzuki M, Miyai I, Ono T, et al. Prefrontal and premotor cortices are involved in adapting walking and running speed on the treadmill: an optical imaging study. Neuroimage 2004;23:1020-6.
10. Firbank M, Okada E, Delpy DT. A theoretical study of the signal contribution of regions of the adult head to near-infrared spectroscopy studies of visual evoked responses. Neuroimage 1998;8:69-78.
11. Cramer SC, Bastings EP. Mapping clinically relevant plasticity after stroke. Neuropharmacology 2000;39:842-51.
12. Silvestrini M, Cupini LM, Placidi F, Diomedi M, Bernardi G. Bilateral hemispheric activation in the early recovery of motor function after stroke. Stroke 1998;29:1305-10.
13. Cuadrado ML, Egido JA, Gonzalez-Gutierrez JL, Varela-De-Seijas E. Bihemispheric contribution to motor recovery after stroke: A longitudinal study with transcranial Doppler ultrasonography. Cerebrovasc Dis 1999;9:337-44.
14. Caramia MD, Palmieri MG, Giacomini P, Iani C, Dally L, Silvestrini M. Ipsilateral activation of the unaffected motor cortex in patients with hemiparetic stroke. Clin Neurophysiol 2000;111: 1990-6.
15. Nelles G, Spiekermann G, Jueptner M, et al. Reorganization of sensory and motor systems in hemiplegic stroke patients. A positron emission tomography study. Stroke 1999;30:1510-6.
16. Traversa R, Cicinelli P, Bassi A, Rossini PM, Bernardi G. Mapping of motor cortical reorganization after stroke. A brain stimulation study with focal magnetic pulses. Stroke 1997;28:110-7.
17. Weiller C, Ramsay SC, Wise RJ, Friston KJ, Frackowiak RS. Individual patterns of functional reorganization in the human cerebral cortex after capsular infarction. Ann Neurol 1993;33: 181-9.
18. Chollet F, DiPiero V, Wise RJ, Brooks DJ, Dolan RJ, Frackowiak RS. The functional anatomy of motor recovery after stroke in humans: a study with positron emission tomography. Ann Neurol 1991;29:63-71.
19. Weiller C, Chollet F, Friston KJ, Wise RJ, Frackowiak RS. Functional reorganization of the brain in recovery from striatocapsular infarction in man. Ann Neurol 1992;31:463-72.
20. Honda M, Nagamine T, Fukuyama H, Yonekura Y, Kimura J, Shibasaki H. Movement-related cortical potentials and regional cerebral blood flow change in patients with stroke after motor recovery. J Neurol Sci 1997;146:117-26.
21. Nelles G, Spiekramann G, Jueptner M, et al. Evolution of functional reorganization in hemiplegic stroke: a serial positron emission tomographic activation study. Ann Neurol 1999;46:901-9.
22. Kato H, Izumiyama M, Koizumi H, Takahashi A, Itoyama Y. Near-infrared spectroscopic topography as a tool to monitor motor reorganization after hemiparetic stroke: a comparison with functional MRI. Stroke 2002;33:2032-6.
23. Miyai I, Yagura H, Oda I, et al. Premotor cortex is involved in restoration of gait in stroke. Ann Neurol 2002;52:188-94.
24. Miyai I, Yagura H, Hatakenaka M, Oda I, Konishi I, Kubota K. Longitudinal optical imaging study for locomotor recovery after stroke. Stroke 2003;34:2866-70.
25. Park SW, Butler AJ, Cavalheiro V, Alberts JL, Wolf SL. Changes in serial optical topography and TMS during task performance after constraint-induced movement therapy in stroke: a case study. Neurorehabil Neural Repair 2004;18:95-105.
26. Saitou H, Yanagi H, Hara S, Tsuchiya S, Tomura S. Cerebral blood volume and oxygenation among poststroke hemiplegic patients: effects of 13 rehabilitation tasks measured by near-infrared spectroscopy. Arch Phys Med Rehabil 2000;81:1348-56.
27. Marshall RS, Perera GM, Lazar RM, Krakauer JW, Constantine RC, DeLaPaz RL. Evolution of cortical activation during recovery from corticospinal tract infarction. Stroke 2000;31:656-61.
28. Carey JR, Kimberley TJ, Lewis SM, et al. Analysis of fMRI and finger tracking training in subjects with chronic stroke. Brain 2002;125(Pt 4):773-88.
29. Genovese CR, Noll DC, Eddy WF. Estimating test-retest reliability in functional MR imaging. I: Statistical methodology. Magn Reson Med 1997;38:497-507.

30. Maitra R, Roys SR, Gullapalli RP. Test-retest reliability estimation of functional MRI data. Magn Reson Med 2002;48:62-70.
31. Manoach DS, Halpern EF, Kramer TS, et al. Test-retest reliability of a functional MRI working memory paradigm in normal and schizophrenic subjects. Am J Psychiatry 2001;158:955-8.
32. Noll DC, Genovese CR, Nystrom LE, et al. Estimating test-retest reliability in functional MR imaging. II: Application to motor and cognitive activation studies. Magn Reson Med 1997;38:508-17.
33. Nybakken GE, Quigley MA, Moritz CH, Cordes D, Haughton VM, Meyerand ME. Test-retest precision of functional magnetic resonance imaging processed with independent component analysis. Neuroradiology 2002;44:403-6.
34. Yetkin FZ, McAuliffe TL, Cox R, Haughton VM. Test-retest precision of functional MR in sensory and motor task activation. AJNR Am J Neuroradiol 1996;17:95-8.
35. Watanabe A, Matsuo K, Kato N, Kato T. Cerebrovascular response to cognitive tasks and hyperventilation measured by multichannel near-infrared spectroscopy. J Neuropsychiatry Clin Neurosci 2003;15:442-9.
36. Plichta MM, Herrmann MJ, Baehne CG, et al. Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable? Neuroimage 2006;31:116-24.
37. Oldfield R. The assessment and analysis of handedness. The Edinburgh inventory. Neuropsychologia 1971;9:97-113.
38. Homan RW, Herman J, Purdy P. Cerebral location of international 10-20 system electrode placement. Electroencephalogr Clin Neurophysiol 1987;66:376-82.
39. Cao Y, D'Olhaberriague L, Vikingstad EM, Levine SR, Welch KM. Pilot study of functional MRI to assess cerebral activation of motor function after poststroke hemiparesis. Stroke 1998; 29:112-22.
40. Seitz RJ, Hoflich P, Binkofski F, Tellmann L, Herzog H, Freund HJ. Role of the premotor cortex in recovery from middle cerebral artery infarction. Arch Neurol 1998;55:1081-8.
41. Cope M, Delpy DT, Reynolds EO, Wray S, Wyatt J, Van der Zee P. Methods of quantitating cerebral near infrared spectroscopy data. Adv Exp Med Biol 1987;222:183-9.
42. Cope M, van der Zee P, Essenpreis M, Arridge SR, Delpy DT. Data analysis methods for near infrared spectroscopy of tissue: problems in determining the relative cytochrome aa3 concentration. Proc SPIE 1991;1431:251-62.
43. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. Stat Med 2002;21:3431-46.
44. Pinheiro JC, Bates DM. Mixed-effects models in S and S-Plus. New York: Springer; 2000.
45. Marshall I, Simonotto E, Deary IJ, et al. Repeatability of motor and working-memory tasks in healthy older volunteers: assessment at functional MR imaging. Radiology 2004;233:868-77.
46. Kiehl KA, Liddle PF. Reproducibility of the hemodynamic response to auditory oddball stimuli: a six-week test-retest study. Hum Brain Mapp 2003;18:42-52.
47. Nybakken GE, Quigley MA, Moritz CH, Cordes D, Haughton VM, Meyerand ME. Test-retest precision of functional magnetic resonance imaging processed with independent component analysis. Neuroradiology 2002;44:403-6.
48. Obrig H, Neufang M, Wenzel R, et al. Spontaneous low frequency oscillations of cerebral hemodynamics and metabolism in human adults. Neuroimage 2000;12:623-39.

Suppliers

a. CW5; TechEn Inc, 115 Cedar St, Milford, MA 01757.
b. Prototype; Fiberoptics Technology, 1 Quassett Rd, Pomfret, CT 06258.
c. Nonin 8600FO; Nonin Medical, 2605 Fernbrook Ln N, Plymouth, MN 55447-4755.
d. ADInstruments, 2205 Executive Cir, Colorado Springs, CO 80906.
e. R statistical software, version 2.1.0. Available at: http://www.r-project.org Accessed August 8, 2006.