Towards the next generation of near-infrared spectroscopy

BY YOKO HOSHI*

Integrated Neuroscience Research Project, Tokyo Metropolitan Institute of Medical Science, 2-1-6 Kamikitazawa, Setagaya-ku, Tokyo 156-8506, Japan

Although near-infrared spectroscopy (NIRS) was originally designed for clinical monitoring of tissue oxygenation, it has also been developing into a useful tool for neuroimaging studies (functional NIRS). Over the past 30 years, technology has developed and NIRS has found a wide range of applications. However, the accuracy and reliability of NIRS have not yet been widely accepted, mainly because of the difficulties in selective and quantitative detection of signals arising in cerebral tissue, which subject the use of NIRS to a number of practical restrictions. This review summarizes the strengths and advantages of NIRS over other neuroimaging modalities and demonstrates specific examples. The issues of selective quantitative measurement of cerebral haemoglobin during brain activation are also discussed, together with the problems of applying the methods of functional magnetic resonance imaging data analysis to NIRS data analysis. Finally, near-infrared optical tomography—the next generation of NIRS—is described as a potential technique to overcome the limitations of NIRS.

Keywords: near-infrared spectroscopy; continuous wave; time-resolved spectroscopy; diffuse optical tomography

1. Introduction

Near-infrared spectroscopy (NIRS) can measure changes in the oxygenation states of haemoglobin (Hb) and myoglobin (Mb) and the oxidation–reduction state of cytochrome c oxidase (cyt. ox.) in living tissue [1]. Since the oxygenation–deoxygenation states of Hb and Mb and the redox state of cyt. ox. depend on the oxygen concentration in blood, muscle and mitochondria, respectively [2], NIRS was originally designed for clinical monitoring of tissue oxygenation [3,4]. The redox state of cyt. ox. has attracted attention as a new tissue oxygenation indicator; however, the specificity and accuracy of its measurements are still not generally established [5–8].

Since the early 1990s, NIRS has been in development as a useful tool for neuroimaging studies, the so-called functional NIRS (fNIRS) [9–12]. Over the past
30 years, a wide range of NIRS instruments have been developed. Among these, the instruments for continuous wave (CW) measurements based on the modified Lambert–Beer law (MLB) [13] (CW-type instruments) are the most readily available commercially. Instruments of this type allow observation of dynamic changes in regional cerebral blood flows (rCBFs) in real time by monitoring the concentration changes in cerebral Hb. The recent advent of multi-channel CW-type instruments has greatly increased the use of NIRS in a variety of neuroimaging studies [14–17].

However, the central issue of NIRS remains to be explored. That is, it is difficult to quantitatively measure Hb concentrations in the cerebral tissue separately from those in the extracerebral tissue. These problems have limited the use of NIRS, and the limitations have not been well understood by many NIRS users. Various approaches have been tried, but it has been indicated that the approach using the MLB is difficult to solve the quantitative issues of NIRS. Here, diffuse optical tomography (DOT) offers a potential technique for the quantitative detection of focal concentration changes in cerebral Hb, and DOT is expected to be able to overcome the limitations of NIRS.

This paper will first focus on CW measurements, with specific examples mainly from our own studies of the strengths and advantages of NIRS measurements over other neuroimaging modalities. Then, it summarizes the issues of selective and quantitative measurements of cerebral Hb concentration changes and the solutions that are available. Finally, the future prospects of NIRS will be discussed.

2. Functional near-infrared spectroscopy

The strengths and advantages of CW-type instruments are as follows: (i) temporal resolution is high (less than 1 s) and non-invasive, which allows long-time continuous measurements in real time and repeating of measurements at short intervals, (ii) measurements can be performed with fewer motion restrictions and in natural environments, and (iii) measurements simultaneous with other techniques for neuroimaging and/or other physiological measurements are possible. These strengths and advantages of NIRS not only enable neuroimaging studies to be performed on subjects who have not been fully examined until now, such as children, the elderly and patients with psychoneurological problems, but also allow new neuroimaging studies that are different from those performed with positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). Some examples of NIRS measurements that make the most of its strengths and advantages are described next.

(a) Event-related NIRS

Because of the high temporal resolution, NIRS can detect brief, event-related haemodynamic responses [18,19]. Different from the data analysis in event-related fMRI, which often uses haemodynamic response functions (HRFs), there are no standard HRFs for NIRS data analysis; however, event-related NIRS data can be analysed without using a HRF. For example, we have employed event-related NIRS to examine rCBF changes directly related to emotions but not to cognitive operations in the anterior frontal region conducted while the subject
viewed pleasant, unpleasant or neutral pictures [20]. Each picture was presented for 6s, followed by the presentation of a white cross-hair for 14s. Changes in oxygenated Hb (oxy-Hb), an indicator of changes in rCBF [21], from the value at the stimulus onset during the 6s stimulation period, which were directly related to emotion generation, were statistically analysed using a two-way ANOVA and the two-tailed Dunnett’s method as a post hoc test. It was found that very unpleasant emotion was accompanied by an increase in oxy-Hb in the bilateral ventrolateral prefrontal cortices (VLPFCs), while very pleasant emotion was accompanied by a decrease in oxy-Hb in the left dorsolateral prefrontal cortex. These findings suggest that the bilateral VLPFCs are involved in the processing of externally induced unpleasant emotions.

(b) Free-motion neuroimaging studies

The NIRS measurements can be performed with fewer motion restrictions, in a daily life environment, which opens new dimensions in neuroimaging studies. In particular, a portable NIRS instrument combined with wireless telemetry (the wearable NIRS system) allows subjects to move during measurements with portable ECG or electroencephalogram (EEG) instruments [22]. Employing a single- or two-channel NIRS instrument, we succeeded in detecting haemodynamic changes associated with emotion and cognitive function in infants and preschool children [23]. These studies demonstrate that NIRS enables a neuroscientific approach to developmental psychology. By using the wearable NIRS system, we also monitored haemodynamic responses to movements such as walking on a balance beam in preschool and school children [24]. The wearable NIRS system has a variety of possible applications, such as investigations of cerebral haemodynamic changes during gravity-induced loss of consciousness [25] and during high-altitude training. To increase the usefulness of the free-motion neuroimaging studies with NIRS, we are now developing a 24-channel wearable NIRS system (light-emitting diode based frequency-division-multiplex Fourier-transform NIRS system) (figure 1).

(c) Multimodal measurements

There are a number of different functional neuroimaging modalities, such as fMRI, PET and magnetoencephalography (MEG), available to detect different aspects of brain functions and each has its different advantages and drawbacks. Combined measurements with several of these modalities can be expected to provide complementary and synergistic results. As NIRS can simply be combined with any of the other modalities, a number of studies combined with NIRS and other modalities have been reported [26–28].

In recent years, the brain–computer interface (BCI) has attracted attention as it provides users with an alternative output channel rather than the physiological output path of the brain, i.e. the efferent nervous system and muscles [29]. With a BCI, patients with disabilities can be provided with an alternative mode of communication and control, available for those with amyotrophic lateral sclerosis and ‘locked-in’ patients. Most of the current BCI devices rely on the electrical activity of the brain that produces scalp EEG signals; however, the scalp EEG signals are inherently noisy, nonlinear and non-localized. Some research groups
have proposed NIRS as an alternative to electrical signals [30], and it is expected that simultaneous measurements with EEG and NIRS will improve the accuracy of decoding brain signals.

3. Problems of near-infrared spectroscopy

Now that NIRS has found a variety of applications, a number of methodological and technical issues still remain to be explored, in addition to the practical difficulties, such as positioning sources and detectors. This review will describe basic and intractable problems in the use of NIRS.

(a) Quantitative and selective detection of near-infrared spectroscopy signals

(i) Quantification of near-infrared spectroscopy signals

A central issue for NIRS is that concentration changes in cerebral Hb cannot be quantified, and this has hindered widespread use of NIRS in clinical medicine and research. When Hb concentration changes are global within tissue, quantification is possible with time-resolved spectroscopy (TRS) and frequency domain spectroscopy (FDS), which can determine optical path lengths. Where Hb concentration changes are localized, such as in functional brain activation, however, the concentration changes cannot be quantified accurately. The optical path lengths determined by TRS and FDS are the mean total path length (t-PL) but not the mean partial path length (p-PL) in the cerebral tissue. As the t-PL is much longer than the p-PL [31,32], the Hb concentration changes are underestimated when the t-PL is substituted for the MLB (partial volume reduction). However, since an accurate measurement of p-PL is not feasible, quantification of focal changes in Hb concentrations is not possible.
To quantitatively analyse NIRS data obtained from CW-type instruments without measuring the p-PL, assumptions such as that the ratio of the source–detector spacing to the t-PL and the ratio of the p-PL to the t-PL are constant have often been made. Arranging the source–detector spacing for pairs of values at equal distances, in which the t-PL and p-PL can be considered constant if the assumption is correct, multi-channel CW-type instruments generate topographic images of relative concentration changes in Hb. However, the assumptions on which this is based are not valid. As is shown in figure 2, the ratio of the source–detector spacing to the t-PL (the differential path length factor (DPF)) \cite{13} varies with each position. Further, p-PL is negatively correlated to the t-PL at a fixed source–detector spacing, and the ratio of the p-PL to the t-PL varies with wavelength and measurement position \cite{33}. These findings mean that the substitution of t-PL for the MLB results in underestimated as well as inaccurate results. Therefore, when Hb concentration changes occur equally within the measured area, a topographic image generated by a multi-channel CW-type instrument does not show a distribution of the Hb concentration changes but displays a distribution of the p-PLs. Further, the amplitudes of NIRS signals vary with the source–detector position \cite{34,35}. This makes direct comparisons of amplitudes across regions an invalid comparison of regional differences in Hb concentration changes.

(ii) Influence of the extracerebral tissue

The light detected on the scalp contains information about both the cerebral and the extracerebral tissue, and changes in the scalp blood flow can influence NIRS signals. Thus, it is necessary to separate the NIRS signals originating in the cerebral tissue from those coming from the extracerebral tissue. To enable this, a multi-detector system consisting of CW-type instruments was developed \cite{36,37}; here, however, separation of NIRS signals is still incomplete. Recently, a new method with a multi-distance probe arrangement based on simulation results

<table>
<thead>
<tr>
<th>CH</th>
<th>PL (cm)</th>
<th>DPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.45</td>
<td>6.82</td>
</tr>
<tr>
<td>2</td>
<td>18.32</td>
<td>6.11</td>
</tr>
<tr>
<td>3</td>
<td>15.94</td>
<td>5.31</td>
</tr>
<tr>
<td>4</td>
<td>11.81</td>
<td>3.94</td>
</tr>
<tr>
<td>5</td>
<td>12.24</td>
<td>4.08</td>
</tr>
<tr>
<td>6</td>
<td>13.40</td>
<td>4.47</td>
</tr>
<tr>
<td>7</td>
<td>16.22</td>
<td>5.41</td>
</tr>
<tr>
<td>8</td>
<td>19.08</td>
<td>6.36</td>
</tr>
<tr>
<td>9</td>
<td>16.09</td>
<td>5.36</td>
</tr>
<tr>
<td>10</td>
<td>15.11</td>
<td>5.02</td>
</tr>
</tbody>
</table>

**Figure 2.** Mean total pathlengths (PLs) and differential path length factors (DPFs) measured on the scalp adjacent to the left lateral prefrontal cortex in a healthy adult. Ten regions between each pair of illuminators and detectors, which were termed as channels (CHs), were measured by TRS. The wavelength is 759 nm. The illuminator (red circles) to detector (black circles) spacing is 3 cm. (Online version in colour.)
has been proposed [38]. The multi-distance FDS [39] and TRS [40], in which diffusion theory is used, have been applied to determine the optical properties of the cerebral tissue in the two-layer head model. Monitoring of oxygen saturation of the cerebral tissue has been performed with spatially resolved spectroscopy [41]. In addition to refining the measurements, mathematical methods, such as principal component analysis and independent component analysis, have also been used to remove extracerebral interference from NIRS signals [42–44]. In some cases, these methods have been successful in eliminating artefacts arising from a subject’s physical and/or physiological activities, but further investigation is required to enable routine application to NIRS measurements.

Functional connectivity in the resting brain is one of the current research interests in neuroimaging studies [45,46], and NIRS has also been employed in investigating functional connectivity [47,48]. However, it must be borne in mind that the cerebral blood flow as well as the scalp blood flow fluctuate spontaneously during rest. Figure 3 shows oxy-Hb changes in the left frontal region in the resting state in a subject measured by TRS. The TRS data can be analysed by two methods: one is based on the MLB, which is analogous to CW measurements (MLB method), and the other is based on diffusion theory (DE-fit method) (see §4a). In figure 3, both the methods produced different results: the upper and lower traces were oxy-Hb changes calculated by the DE-fit and the MLB method, respectively. Such a discrepancy is occasionally observed in resting-state TRS measurements. It has been reported that the DE-fit method shows higher sensitivity to the deeper regions of the head, while the MLB method is more sensitive to absorbance changes of the superficial layers [49,50]. Thus, it is assumed that the lower trace (the MLB method) mainly reflects the skin blood flow in the case of figure 3. This indicates that NIRS signals in the resting state measured by CW-type instruments can be influenced by skin blood flow, as well as in the activation state. Thus, NIRS instruments that can selectively measure cerebral Hb are recommended for investigations of functional connectivity.

*Phil. Trans. R. Soc. A* (2011)
Towards the next generation

(b) Near-infrared spectroscopy data analysis

Unlike PET and fMRI, there are no standard methods of NIRS data analysis, and a number of methods of analysis have been proposed and tried. This is not in itself a problem, and it is potentially beneficial to perform non-standard analyses of NIRS data. Nevertheless, new NIRS users often experience difficulties in establishing an appropriate method of data analysis and it is also time-consuming to perform the analysis, especially on multi-channel NIRS data. Thus, some statistical analysis tool sets for NIRS, which are based on the methods used in fMRI data analysis, have been developed [51,52]. However, these methods do not necessarily consider the situation in which there are several distinct differences in the nature of signals generated and analysed by fMRI and NIRS; for example, the fMRI signals reflect haemodynamic changes occurring in the cerebral tissue, whereas the NIRS signals do not correspond simply to Hb concentration changes in the cerebral tissue as described above. The special characteristics of NIRS signals must be carefully considered when developing standard statistical analysis methods, which must be valid and reliable.

Model-based, event-related and analyses combing these two, which are in general incorporated into the statistical analysis tool sets mentioned above, have also been tried [53,54]. As these approaches do not require quantification of Hb concentration changes, they are useful in the analysis of NIRS data. Model-based analysis is widely used for data analysis in fMRI and PET studies. However, this method cannot universally be applied to NIRS data, because the pattern of Hb changes often varies with measurements, although high reproducibility has been reported by one research group [55]. In addition, it is difficult to generate accurate haemodynamic models for NIRS measurements, as the duration of stimuli is generally longer than that in fMRI and PET.

Although grouped data analysis is crucial for fMRI and PET studies, it is very difficult to perform this kind of analysis on NIRS data. To combine NIRS data for multiple subjects, the same brain regions must be measured. However, inter-subject variation in the measured areas is unavoidable, because of differences in head size and shape, and as anatomical structures of the brain are different from one subject to another, while the size of a source–detector light guide holder is fixed. In addition, the holder can accidentally become dislocated from the ideal position when it is attached to the head. It appears to be a commonly held misconception in the cognitive neuroscience community that grouped data analysis is more reliable than the analysis of data for individuals. Focusing on measures of central tendency is not only insufficient to understand complex physiological phenomena in the brain but can also lead to misunderstanding [56,57]. The significance of the analysis of data for discrete individuals needs to be more appreciated. Similar to MEG data analysis, analytical results for individuals can be organized into grouped data for further statistical analysis of NIRS data [16,20].

4. Future prospects

To overcome the limitations of NIRS, it will be necessary to develop methods which are not based on the MLB to be able to selectively and quantitatively measure cerebral Hb. Today, DOT is the most promising technique to achieve
this aim. In this section, a general account of TRS, which is one of the basic measurement methods for DOT, is initially given, and then DOT is described.

(a) Time-resolved spectroscopy

In TRS, ultrashort (picosecond order) laser pulses are irradiated on the tissue, and the intensity of the emerging light is detected as a function of time (the temporal point spread function; TPSF) with picosecond resolution [58,59]. Because time-resolved instrumentation was developed as laboratory-based devices, it has been difficult to use in the clinical environment. However, recent advances in technology have made the development of a compact 64-channel time-resolved optical imaging system possible [60]. Now, single-/two-channel TRS instruments are commercially available [61].

The mean transit time of the scattered photons is calculated with the TPSF. The mean t-PL is determined by multiplying the light speed in the media by the mean transit time of the scattered photons. However, it must be noted that the mean t-PL determined by TRS measurements of the head is a summation of the p-PLs within the cerebral and extracerebral tissues.

The TPSF contains information about the optical properties of the media, that is, absorption ($\mu_a$) and reduced scattering ($\mu'_s$) coefficients. Propagation of photons in living tissue, which is assumed to be a homogeneous medium, can be described by the diffusion equation (equation (4.1)) [62],

$$\frac{1}{c} \frac{\partial \varphi(r, t)}{\partial t} = \nabla [D(r) \nabla \varphi(r, t)] - \mu_a(r) \varphi(r, t) + S(r, t), \quad (4.1)$$

where $\varphi(r, t)$ is the diffuse photon fluence rate at position $r$ and the time $t$, $c$ is the speed of light in the tissue, $D = 1/3\mu'_s(r)$ is the optical diffusion coefficient and $S(r, t)$ is the photon source. For a semi-infinite half-space geometry, the solution of equation (4.1) for the impulse input is expressed by equation (4.2) [63],

$$R(l, t) = \frac{z_0 \exp(-\mu_a ct)}{t (4\pi Dct)^{3/2}} \exp\left(-\frac{l^2 + z_0^2}{4Dct}\right), \quad (4.2)$$

where $R(l, t)$ is the intensity of the reflected light, $l$ and $t$ are the distance and time from the impulse input, respectively, and $z_0 = 1/\mu'_s$. By fitting equation (4.2) to the observed TPSF, the values of $\mu_a$ and $\mu'_s$ are determined.

It is still not conclusively established whether the human head can be considered as a homogeneous and/or a simple slab model. Thus, different mathematical approaches, such as deriving the solution of the diffusion equation for multi-layered turbid media [64,65] and analysing moments of the TPSF [40], have been tried. Nevertheless, it has been reported that the estimated $\mu_a$ is nearly equal to the deeper layer $\mu_a$ under the condition where the $\mu_a$ of the upper layer is larger than that of the lower layer in two-layered slab models [49]. We have also confirmed that the curve-fitting method based on the diffusion equation (DE-fit method) is more sensitive to $\mu_a$ changes in the cerebral tissue by measuring the adult head with TRS during carotid endarterectomy [50]. This is explained by the fact that the later time (falling) part of the TPSF is critical for the
Figure 4. Changes in oxy-Hb in the left forehead of a healthy adult during hyperventilation. Dotted line indicates oxy-Hb changes calculated by the MLB method (arbitrary units, arb. units). Solid line denotes oxy-Hb changes calculated by the DE-fit method (molar concentration).

determination of the $\mu_a$ in the fitting process. In contrast, the time-independent calculation of TRS data based on the MLB (MLB method), which is analogous to CW measurements, is quite sensitive to $\mu_a$ changes in extracerebral tissue. Figure 4 shows an example: changes in oxy-Hb in the left forehead measured by TRS during hyperventilation. The oxy-Hb calculated by the MLB method increased during hyperventilation, which probably reflected increases in the scalp blood flow owing to elevated sympathetic nerve activity. In contrast, the oxy-Hb calculated in the DE-fit method decreased, reflecting a reduction in the cerebral blood flow caused by hypocapnia.

(b) *Time-resolved domain diffuse optical tomography*

Although changes in cerebral Hb concentrations can be more selectively measured by the DE-fit method, the selective and quantitative accuracy are insufficient. While several different approaches with single-channel TRS have been proposed [66–68], further investigation must be continued to apply these methods to human head measurement. Unlike this, DOT is a technique for reconstructing images of Hb concentration changes using multiple light sources and detectors [69], and this technique is thought to be the most promising for the quantitative detection of focal changes in cerebral haemodynamics. DOT has been under development since the early 1990s, and it can be performed with TRS [70], FDS [71] and CW-type instruments [72]. Recently, CW high-density DOT with high spatial resolution has been developed and applied to functional haemodynamic maps of the adult human visual cortex [73]. However, this CW-DOT provides only qualitative images because it is based on linear image reconstruction, and nonlinear iterative reconstruction schemes are required.
to obtain quantitative images. Since TRS can provide the richest information about the distribution of optical properties in biological tissue [74], we have been developing a time-resolved domain DOT (TR-DOT).

In the 1990s, we developed the 64-channel TR-DOT system in a research and development project for optical tomographic imaging systems for the New Energy and Industrial Technology Development Organization and Ministry of International Trade and Industry of the Japanese Government [60]. This system has three 100ps pulsed LDs (wavelengths of 761, 797 and 834nm) and 64 fibre bundles that are connected to 64 corresponding time-resolved detection devices. Each of the detecting devices consists of an optical attenuator, a high-speed, high-sensitivity photomultiplier tube and a time-correlated single photon counting circuit that contains a miniaturized constant fraction discriminator/time-to-amplitude converter module and a signal acquisition unit with an A/D converter. To reduce the number of optical fibres, a special coaxial bundle was developed: the centre part is for illumination, and the annular part for detection.

Optical technology advanced further in the 2000s, and now more high-quality light sources and detector devices are available. However, TR-DOT is still not ready for clinical use, mainly because of difficulties in describing photon migration in biological tissue (the forward model) and developing image reconstruction algorithms based on nonlinear inversion of an appropriate forward model (the inverse model). The diffusion equation is often used as the forward model, whereas the diffusion approximation is not valid for tissue subjecting photons to little scattering, such as cerebrospinal fluid and small source–detector spacings. In contrast, the radiative transport equation (RTE) accurately describes the photon propagation in tissue. Although this integro-differential equation is not easily solved even by numerical methods, the RTE is more reliable as the forward model [75,76],

Various image reconstruction algorithms, which are essentially based on inverse problem techniques, have been proposed [77–79] and applied to human measurements [80–82]. However, the image quality is inherently low, because the problem is usually nonlinear, ill-posed and underdetermined because of the diffusive nature of the photon migration. In addition to improving the forward and inverse models, recently, combining DOT with a high-resolution structure-oriented imaging modality, such as MRI or ultrasound, has been reported to successfully improve image resolution and accuracy [83–85]. Although the development of DOT is extremely difficult, recent technological and methodological advances are encouraging. TR-DOT can be extended to fluorescence TR-DOT [86], which enables in vivo molecular imaging. Overall, these techniques will enable non-invasive investigation of brain functions at all levels, from the molecule/cell level to the system/individual level.

5. Conclusions

Over the last 30 years, NIRS has been making steady progress. Its strengths and advantages have provided new directions for neuroimaging studies. On the other hand, the central issues of NIRS remain to be resolved, and before that is achieved there will be limitations in the areas of application of NIRS. With DOT,
the next generation of NIRS, there is great promise to overcome the present-day limitations of NIRS. The DOT technique offers the most exciting future prospects for biomedical optics.

References

Y. Hoshi


*Phil. Trans. R. Soc. A* (2011)


