



# Using a Data-Driven Approach to Estimate Second-Language Proficiency From Brain Activation: A Functional Near-Infrared Spectroscopy Study

Miaomei Lei<sup>1\*</sup>, Toshinori Miyoshi<sup>1</sup>, Ippeita Dan<sup>2</sup> and Hiroki Sato<sup>3</sup>

<sup>1</sup> Research & Development Group, Hitachi, Ltd., Tokyo, Japan, <sup>2</sup> Research and Development Initiatives, Applied Cognitive Neuroscience Laboratory, Chuo University, Tokyo, Japan, <sup>3</sup> Department of Bioscience and Engineering, College of Systems Engineering and Science, Shibaura Institute of Technology, Saitama, Japan

## OPEN ACCESS

### Edited by:

Ning Liu,  
Stanford University, United States

### Reviewed by:

Hendrik Santosa,  
University of Pittsburgh, United States  
Keum-Shik Hong,  
Pusan National University,  
South Korea

### \*Correspondence:

Miaomei Lei  
miaomei.lei.jt@hitachi.com

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 21 October 2019

**Accepted:** 08 June 2020

**Published:** 10 July 2020

### Citation:

Lei M, Miyoshi T, Dan I and  
Sato H (2020) Using a Data-Driven  
Approach to Estimate  
Second-Language Proficiency From  
Brain Activation: A Functional  
Near-Infrared Spectroscopy Study.  
*Front. Neurosci.* 14:694.  
doi: 10.3389/fnins.2020.00694

While non-invasive brain imaging has made substantial contributions to advance human brain science, estimation of individual state is becoming important to realize its applications in society. Brain activations were used to classify second-language proficiencies. Participants in functional near-infrared spectroscopy (fNIRS) experiment were 20/20 native Japanese speakers with high/low English abilities and 19/19 native English speakers with high/low Japanese abilities. Their cortical activities were measured by functional near-infrared spectroscopy while they were conducting Japanese/English listening comprehension tests. The data-driven method achieved classification accuracy of 77.5% in the case of Japanese speakers and 81.9% in the case of English speakers. The informative features predominantly originated from regions associated with language function. These results bring an insight of fNIRS neuroscience and its applications in society.

**Keywords:** machine learning, language proficiency, brain activation, feature selection, native Japanese speakers, native English speakers

## INTRODUCTION

Language, which differentiates human beings from other living species, plays an important role in our daily lives. The neural basis of language has been investigated with various techniques for functional neuroimaging (Price, 2012; Quresima et al., 2012). Functional near-infrared spectroscopy (fNIRS) is an optical neuroimaging technique that measures brain activity by monitoring the hemodynamic changes in cerebral cortex response of brain activation. Its main advantages are relatively low cost, portability, safety, low acoustic noise (compared to functional magnetic resonance imaging), and easiness to operate (Scholkmann et al., 2014; Hong and Yaqub, 2019). In the context of fNIRS community, hemodynamic changes (which represent brain activation) have been used as a useful indicator to demonstrate speech perception in infants (Pena et al., 2003; Bortfeld et al., 2009; Sato et al., 2012) and language comprehension in adults

(Sato et al., 1999; Schecklmann et al., 2008; Lei et al., 2018). Since conventional analysis of fNIRS data has focused on human brain activity at the group level, these studies have traditionally drawn a population-level conclusion about general patterns across a large number of participants. Knowledge from these studies has important implications for advancing our understanding of how the human brain processes language. To further translate this knowledge into practical applications in society, individual estimation or classification of language ability (e.g., speech-comprehension level and second-language proficiency) on the basis of neuroimaging data across participants is a topic of interest.

Research interest in estimating the state of an individual by applying machine learning using fNIRS data has been increasing. fNIRS data with corresponding labels/classes are used to train a machine-learning classifier/model. The trained classifier is applied to the unknown data to estimate the labels. For example, in previous studies, mental arithmetic and music imagery (Power et al., 2010) motor imagery (Naseer and Hong, 2013), and subjective preference (Luu and Chau, 2008) were estimated. These studies showed the feasibility of establishing a predictive machine-learning model based on the state of individuals and their underlying brain activity. However, they focused on categorical discrimination to estimate the state of a participant using data of the participant. It is still a challenge to estimate the state of individuals on the basis of neuroimaging data of others, namely, estimation of state across participants. Clinical studies focusing on classification of diseases or disorders have made progresses in such estimation across participants (Hosseini et al., 2018; Sutoko et al., 2019) however, estimation of, for example, language ability, which may subtly differ across groups, remains unstudied.

The major difficulty concerning estimation across participants based on neuroimaging data is the relatively small data sample with individual differences. Individual differences refer to the variations across participants even though they have the same label in the same population group, for example, patients who have the same disease. It has been recognized that the brain structure and its corresponding function show high individual variability even among a healthy population group (Raz et al., 2005; Qin et al., 2014; Finn et al., 2015). From the viewpoint of machine learning, if input patterns have high individual differences, data in the feature space will be almost impossible to separate according to the label. It is thus difficult to construct a machine-learning classifier and model for the estimation, that is, separating feature vectors based on the information of the label. On the other hand, estimation with a small number of data samples is also a challenge; that is, the data dimensionality is usually much higher than the number of samples available for classifier training (Fan et al., 2007; An et al., 2017). This typical problem is known in machine-learning literature as the “curse of dimensionality” (Bellman, 2015). It may make the model unstable or cause the problem of overfitting (Guyon and Elisseeff, 2003), which is the condition that model fits accurately to the training data (including inherent noise) but fits poorly to unknown test data. In addition, in a practical situation, increasing the number of data samples is not always possible, for example, when the

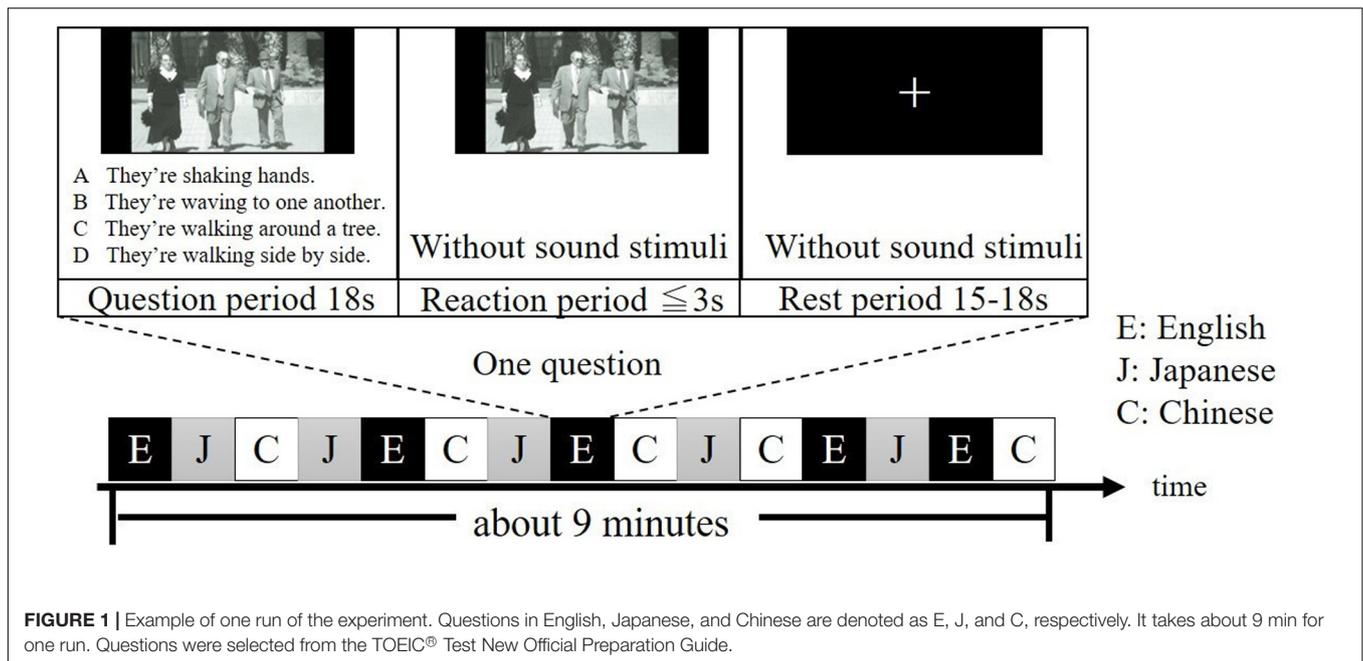
number of patients with a particular disease is limited. To solve this unbalance between number of features and sample size with the aim of increasing classification accuracy, various methods of feature selection (to extract a subset of most-informative features) have been proposed (Saeys et al., 2007; Pereira et al., 2009; Hu et al., 2013; Mwangi et al., 2014; Hong et al., 2018). Common methods of feature selection include using *t*-test (De Martino et al., 2008) ANOVA (analysis of variance) (Akama et al., 2014; Lei et al., 2014) Pearson correlation coefficient (Fan et al., 2007) and prior knowledge (Chu et al., 2012). Recently, sparse-feature selection has become one of the choices for data-driven feature selection (Tibshirani, 1996). Sparse techniques combine both machine learning and feature-reduction steps by enlisting a L1-norm regularization, resulting in a reduced subset of relevant features (Zou and Hastie, 2005).

In our previous study, we reported significant differences between brain-activity patterns in regard to correct responses and incorrect responses of a second language at group level (Lei et al., 2018). A reliable quantitative tool for evaluating second-language proficiency based on brain-activation patterns may help people to learn a second language more efficiently. In the present study, we aimed to estimate second-language proficiency using functional brain activity provided by fNIRS data applying machine learning methods. fNIRS data were collected from native Japanese speakers with high/low second-language (English) proficiency and native English speakers with high/low second-language (Japanese) proficiency. Brain activities were recorded by fNIRS when the subjects (speakers) were doing listening-comprehension tasks in English, Japanese, and an unknown language (Chinese). High second-language proficiency and low second-language proficiency is estimated cross participants. To overcome difficulties with estimation across participants and further improve classification performance, the informative features were extracted by using a method of sparse-feature selection. The generalization capability of the machine-learning methods was confirmed by analyzing two independent-validation population groups of native Japanese speakers and native English speakers. In addition, using the label of second-language proficiency classification of the first language and an incomprehensible unknown language was also conducted.

## MATERIALS AND METHODS

### Participants

All participants in the present study, categorized as native Japanese speakers and native English speakers, were right-handed. The native Japanese speakers were 65 healthy adults (mean age  $\pm$  SD:  $28.5 \pm 2.8$ ; range: 24–33; 35 males and 30 females). Based on their TOEIC® Listening & Reading scores, two groups with different English proficiency were recruited. The high-proficiency group contained 32 participants (mean age  $\pm$  SD:  $27.8 \pm 2.6$ ; range: 24–32; 18 males and 14 females) with TOEIC® Listening & Reading scores above 700. The low-proficiency group contained 33 participants (mean age  $\pm$  SD:  $29.2 \pm 2.8$ ; range: 24–33; 17 males and 16 females) with TOEIC® Listening & Reading scores under 500.



**FIGURE 1** | Example of one run of the experiment. Questions in English, Japanese, and Chinese are denoted as E, J, and C, respectively. It takes about 9 min for one run. Questions were selected from the TOEIC® Test New Official Preparation Guide.

The native English speakers were 66 healthy adults (mean age  $\pm$  SD:  $28.7 \pm 2.9$ ; range: 24–33; 34 males and 32 females). They included nationals from Australia, Canada, New Zealand, the United Kingdom, and the United States, who were temporarily staying in Japan for periods ranging from 0.2 to 12 years (mean years  $\pm$  SD:  $3.4 \pm 2.6$ ). On the basis of their self-assessments of Japanese proficiency, namely, whether they can speak Japanese or not, the participants were categorized into the high-proficiency group or the low-proficiency group. The high-proficiency group was composed of 31 people (mean age  $\pm$  SD:  $29.3 \pm 3.1$ ; range: 24–33; 15 males and 16 females), and low-proficiency group was composed of 35 people (mean age  $\pm$  SD:  $28.1 \pm 2.7$ ; range: 24–33; 19 males and 16 females). In addition, all participants did not have experience of learning Chinese.

Data were obtained according to the standards of the internal review board of Research & Development Group, Hitachi, Ltd. Data from volunteers were obtained according to the standards of internal review board on Research & Development Group, Hitachi, Ltd., following receipt of written informed consent.

### Auditory Stimuli and Task Design

Listening comprehension questions from “TOEIC® Listening Test Part 1: Photographs”<sup>1</sup> were used (ETS, 2005, 2007, 2008, 2012). Each question relates to a photograph with four short explanations. The explanation that most accurately describes the photograph is to be chosen. These listening-comprehension

<sup>1</sup>Written pledge was submitted to copyright owner IIBC (The Institute for International Business Communication) and permission was obtained from the copyright owner. Questions were selected from the TOEIC® Test New Official Preparation Guide vol. 1 (2005), vol. 2 (2007), vol. 3 (2008), and vol. 5 (2012) published by IIBC, and copyrighted by Educational Testing Service (ETS). TOEIC is a registered trademark of ETS. This publication is not endorsed or approved by ETS.

questions were respectively translated into Japanese and Chinese by the native speakers. Sound stimuli were created from the recorded voice of a professional female announcer who is bilingual in Japanese and English, and has learned Chinese as a third language.

All participants were given two runs, each of which contained 15 different questions. During each run, questions in Japanese, English and Chinese were presented five times, respectively, in a pseudo-randomized order (Figure 1). Note that questions in the same language were not given continuously. For each question, a period of 18 s was for presenting the question (question period), a period of less than or equal to 3 s was for answering the question (reaction period) and an arbitrary period between 15 to 18 s was for resting (rest period). The experimental session was conducted in a quiet, dimly lit room. Participants were instructed to look at the photograph on the screen and listen to the four explanations in the question period, answer the question in the reaction period and look at a fixation cross on the screen in the rest period. Specifically, after listening the four explanations of the question, participants were asked to press the button as quick as possible during the reaction period. After pressing the button, the photograph will disappear, and a fixation cross will be shown. Finally, the participants were instructed to silently fix their eyes on the cross and no response was required during the rest period, when they were also asked to think nothing as possible as they can. To ensure that each participant clearly understood the experiment procedure during the on-line tasks, the participants did practice tasks similar to experimental tasks in advance.

### fNIRS Measurement

An optical-topography system (ETG-4000; Hitachi Medical, Japan) was used to measure change in concentration of cerebral

cortical hemoglobin. Absorption of near-infrared light at two wavelengths (695 and 830 nm) was measured with a sampling rate of 10 Hz. Two 3 × 5 optode probe sets were placed over the bilateral frontal and temporal areas by referring to the international 10–20 system of electrode placement. Each optode-probe set consists of eight emitters and seven detectors, resulting in 22 measurement channels. The source-detector distance was fixed at 3 cm. For spatial registration, virtual registration (Tsuzuki et al., 2007) was used to register the channel positions in relation to the Montreal Neurological Institute (MNI) standard brain space (Collins et al., 1994; Brett et al., 2002). The anatomical estimation is based on LBPA40 (Shattuck et al., 2008) and Brodmann’s atlas (Rorden and Brett, 2000). The channel positions include regions related to auditory language processing (Friederici et al., 2000; Obrig et al., 2010; Price, 2012; Hall et al., 2013).

### fNIRS Data Preprocessing

For analyzing the fNIRS data, Mathematica (Version 10.0, Wolfram Research, Inc., IL, United States) and Matlab (Version 2017a, Mathworks Inc., Natick, MA, United States) were used. Based on the modified Lambert–Beer law, concentration changes in oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb) on each measurement channel were obtained (Maki et al., 1995). A band-pass filter (0.01–0.8 Hz) was then applied for noise reduction as in a previous study (Sasai et al., 2012; Santosa et al., 2013; Tak and Ye, 2014). The time-continuous data were divided into 33-s language blocks, which consisted of the 18-s question period, the reaction period (less than or equal to 3 s) and the rest period (between 15 and 18 s). After all language blocks were extracted, the baseline was corrected by using linear fitting to a mean signal over 5 s before the task and over the last 5 s of the task. Since optical measurements correspond to the hemodynamic signals, which are an indirect measure of neuronal activity. The hemodynamic signals (representing blood flow) are delayed in relation to the actual neuronal activity (Dehaene-Lambertz et al., 2002). Therefore, in consideration of the delay, the most-informative part of the comprehension during the task period (that is, the amplitude between 5 and 18 s averaged over each extracted language block on each measurement channel) was used to calculate brain activation.

The activation indicator used as an input feature is based on the significant differences between the oxy-Hb and deoxy-Hb signals (Cui et al., 2010). Since fNIRS simultaneously measures the concentration changes of oxy-Hb and deoxy-Hb, this indicator reflects activation strength. For each measurement channel of each participant, the activation indicator is defined as

$$\text{Activation indicator} = \frac{\overline{oxy} - \overline{deoxy}}{\sqrt{\frac{S_{oxy}^2}{n} + \frac{S_{deoxy}^2}{m}}} \quad (1)$$

where  $\overline{oxy}$  and  $\overline{deoxy}$  are sample means,  $S_{oxy}$  and  $S_{deoxy}$  are sample standard deviations, and  $n$  and  $m$  are sample sizes.

Since there are 44 measurement channels, the number of activation indicators for one participant is 44, and the input

**TABLE 1** | Details of participants after participant selection.

	Native Japanese speakers		Native English speakers	
	High-proficiency group	Low-proficiency group	High-proficiency group	Low-proficiency group
<i>N</i>	20	20	19	19
Female/male	8/12	10/10	6/13	8/11
Age (mean ± SD)	28.1 ± 2.6	29.4 ± 2.9	29.2 ± 2.9	28.5 ± 2.4

feature can be represented as a vector,  $A = (a_1, \dots, a_{44})$ . The number of dimensions of the original input feature is 44.

The label of the participants was re-examined. As a matter of fact, the participants in the high-proficiency group showed a low rate of correct answers, and the participants in the low-proficiency group showed a high rate of correct answers; that is, label proficiency group and label rate of correct answers contradict. To remove ambiguous data, participants whose measurement data did not contradict were further selected from both the native Japanese speakers and native English speakers. After those participants were selected, as for the native Japanese speakers, 20 participants were left in the high-proficiency group, and 20 participants were left in the low-proficiency group; and as for native English speakers, 19 participants were left in the high-proficiency group, and 19 participants were left in the low-proficiency group. The details about number, sex and age are shown in **Table 1**. As for both the native Japanese speakers and native English speakers, the high- and low-language proficiency groups were age-gender matching groups.

### Algorithm Evaluation

The following conventional methods, which were shown to be promising by various classification studies, were used to classify the language proficiency into the high or low group.

- Support Vector Machine (SVM)
- Sparse Logistic Regression (Yamashita, 2009) (SLR)
- K-Nearest-Neighbors based on Euclidean distance of original input features (KNN,  $K = 5$ ).

Using brain-activation vectors [for example,  $A = (a_1, \dots, a_{44})$ ] for classifying each participant into the high or low groups was evaluated. Concretely, a support-vector machine (SVM) is considered to be a promising and popular algorithm among those used in classification studies, and it has been used in a variety of fNIRS studies (Li et al., 2016; Hosseini et al., 2018). Moreover, a SVM has already been used to examine the diagnostic potential of neuroimaging for a range of psychiatric disorders (Nieuwenhuis et al., 2012; Orrù et al., 2012). A SVM with a linear kernel was adopted for the binary-classification problem. The algorithm known as sparse logistic regression (SLR) (Yamashita, 2009) is an extension of logistic regression to automatically select features related to a label. Logistic regression is extended to a Bayesian framework by using a technique known as automatic relevance determination (ARD) from the neural-network literature. By combining logistic regression with ARD, SLR is obtained. SLR

is effective for removing irrelevant features, such that their associated weights are automatically set to zero, leading to a sparse weight vector for classification. In the implementation of this study, default values in the SLR toolbox were used to do the classification. K-nearest neighbor(s) (KNN) using majority voting (Duda et al., 2012) was used for classification. In this study, K was fixed to 5. K-nearest neighbors defines the label of test data by looking at the K-closest training data in the feature space. And it is sensitive to the local structure of the data.

Leave-one-out cross validation (LOOCV) was applied for cross validation. In detail, the data are divided into N folds (N = 40 for the native Japanese speakers; N = 38 for the native English speakers). In each leave-one-out cross-validation fold, all except one participant (N-1) were used as training data; the one participant left out was used as test data to determine which group the participant came from. This process is repeated once for each participant.

The classification accuracy of second language proficiency was computed to verify the estimation performance of the algorithm. A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such a system is commonly evaluated using the data in the matrix.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

### Informative-Feature Selection

Given the small number of data sets and the high dimensions of the data, to further improve classification accuracy, feature selection or feature extraction is necessary (Guyon and Elisseeff, 2003; Akama et al., 2014). By selecting informative features, the machine-learning algorithm can give stable results and the physical interpretations of selected features are also important and worth discussing by means of neuroscience.

In this study, sparse canonical correlation analysis (SCCA) is applied to select the informative features. SCCA identifies sparse linear combinations of two sets of highly correlated variables (Witten et al., 2009). It has been shown to be useful in the analysis of high-dimensional neuroimaging data, namely, when two sets of variables are available for the same set of samples (Yahata et al., 2016).

Specifically, N observations (participants) of paired variables  $X \in \mathbb{R}^{d1}$  and  $Y \in \mathbb{R}^{d2}$  are given, X is an N × d1 matrix comprising the first set of variables, and Y is an N × d2 matrix comprising the second set of variables. L1-norm SCCA can be formulated as

$$\begin{aligned} \max_{v_X, v_Y} v_X^T X^T Y v_Y \text{ subject to } \|v_X\|_1 \leq \lambda_X, \|v_Y\|_1 \leq \lambda_Y, \\ \|v_X\|_2 \leq 1, \|v_Y\|_2 \leq 1, \end{aligned} \tag{3}$$

where hyperparameters  $\lambda_X$  and  $\lambda_Y$  indicate the sparseness of projection vectors  $v_X$  and  $v_Y$ , respectively. The projection matrices are  $v_X \in \mathbb{R}^{d1 \times m}$  and  $v_Y \in \mathbb{R}^{d2 \times m}$ , where m = min

(d1, d2). In this study, X is the input feature, and Y is the label for language proficiency. Before input to SCCA, the training set is centered to have zero mean and scaled to have unit variance.

To select the common informative feature, the data sets for the second language were applied, and the LOOCV described above was used for cross validation. A feature selected from more than 95% of participants during the leave-one-out procedure was defined as a common informative feature. That is, a feature was selected when it was shown more than 38 times by the Japanese speakers (N = 40) and more than 36 times by the English speakers (N = 38). After the common informative features were selected, machine-learning methods were used to classify language proficiency to confirm that classification accuracy has been improved.

## RESULTS

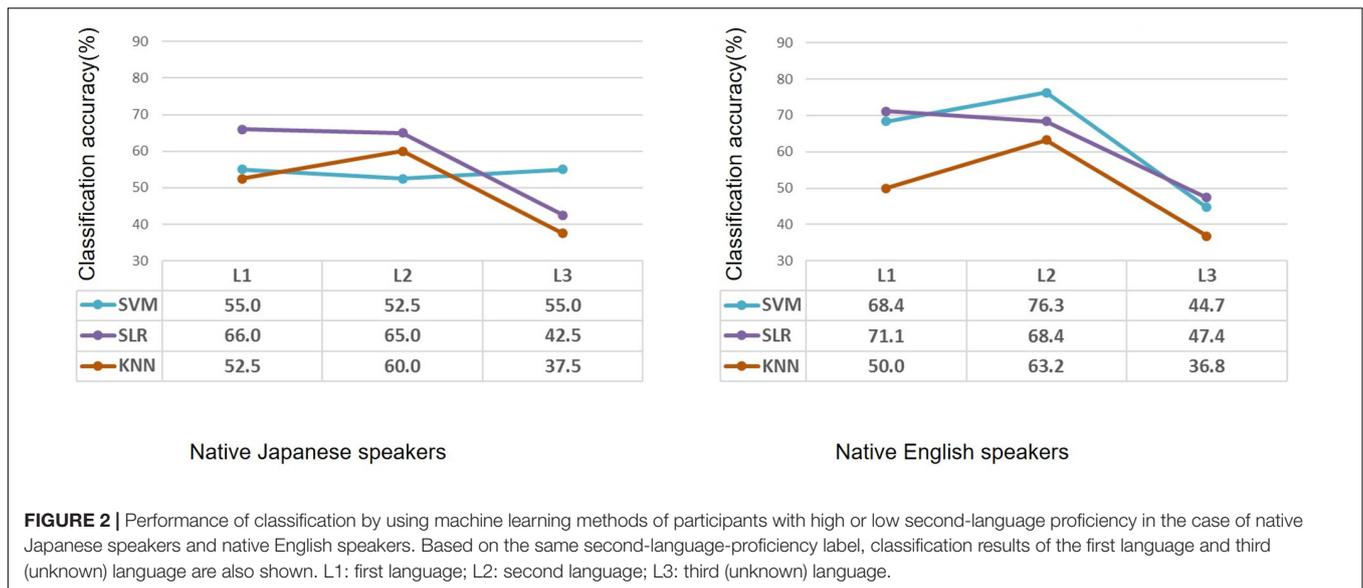
### Classification Performance

Classification performance for the native Japanese speakers and the native English speakers is shown in Figure 2. Classification accuracies of the SVM for the native Japanese speakers were 55.0%, 52.5%, and 55.0% in terms of first language (L1), second language (L2) and third (unknown) language (L3), respectively, which show low classification accuracy. When SLR was used, classification accuracies were 66.0%, 65.0%, and 42.5% for the three languages, respectively; namely, classification of L1 and L2 showed higher accuracy than that for L3. When KNN was used, classification accuracies were 52.5%, 60.0%, and 37.5%; that is, only classification of L2 showed higher accuracy. For the native English speakers, classification accuracies of SVM were 68.4%, 76.3%, and 44.7% for L1, L2, and L3, respectively; that is, classification of L2 showed higher accuracy. When SLR was used, classification accuracies were 71.1%, 68.4%, and 47.4% for the three languages, respectively. When KNN was used, classification accuracies were 50.0%, 63.2%, and 36.8%, respectively. SLR showed the highest classification accuracy for the second language in the case of the Japanese speakers, and SVM showed the highest classification accuracy for the second language in the case of the English speakers.

The SVM showed unstable classification results: it cannot classify language proficiency in the case of the native Japanese speakers. SLR showed unexpected high classification accuracy when classifying L1. Using the label of second-language proficiency, SLR showed the highest classification accuracy for L1 in the case of both the native Japanese speakers and native English speakers. K-nearest neighbor showed reasonable classification results; namely, classification accuracy for L2 is higher, and that for L1 and L3 is near to chance level.

### Informative Features Shared Between High- and Low-Language-Proficiency Groups

To extract informative features that can improve classification accuracy, sparse canonical correlation analysis was used. The same leave-one-out cross validation (LOOCV) procedure was



**FIGURE 2 |** Performance of classification by using machine learning methods of participants with high or low second-language proficiency in the case of native Japanese speakers and native English speakers. Based on the same second-language-proficiency label, classification results of the first language and third (unknown) language are also shown. L1: first language; L2: second language; L3: third (unknown) language.

adopted to select the features. Common informative features were defined as features selected from more than 95% of the participants. Spatial distribution of informative features is shown in **Figure 3**, and anatomical information about the features is listed in **Table 2**. As for the native Japanese speakers, the selected common informative features correspond to channel 1, channel 6, and channel 22 on the left hemisphere and channel 16 on the right hemisphere. The anatomical information about these features indicate the pars opercularis, part of Broca’s area, left precentral gyrus, left inferior temporal gyrus and right superior temporal gyrus. As for the native English speakers, the selected common informative features correspond to channel 2, channel 9, channel 15, and channel 17 on the left hemisphere and channel 19 on the right hemisphere. The anatomical information about these features indicate the left postcentral gyrus, left angular gyrus, part of Wernicke’s area, left superior and middle temporal gyrus and right middle temporal gyrus.

After feature selection, the informative features were used to classify each participant into the high-proficiency group or the low-proficiency group. Classification accuracy for L1, L2, and L3 in the case of the native Japanese speakers and the native English speakers is shown in **Figure 4**. As for the native Japanese speakers, when SVM was used, classification accuracy for L2 was the highest, i.e., 75%. When SLR was used, classification accuracies were 70.0%, 75.0%, and 55.0% for L1, L2, and L3, respectively; similarly, the accuracy was highest for L2. When KNN was used, classification accuracies were 55.0%, 77.5%, and 65.0%. As for the native English speakers, when SVM was used, classification accuracies for L1, L2, and L3 were 76.3%, 81.9%, and 57.9%, respectively. When SLR was used, classification accuracies were 68.4%, 79.0%, and 63.2% for L1, L2, and L3, respectively. When KNN was used, classification accuracies were 63.2%, 73.7%, and 63.2%; namely, classification accuracy was highest for L2.

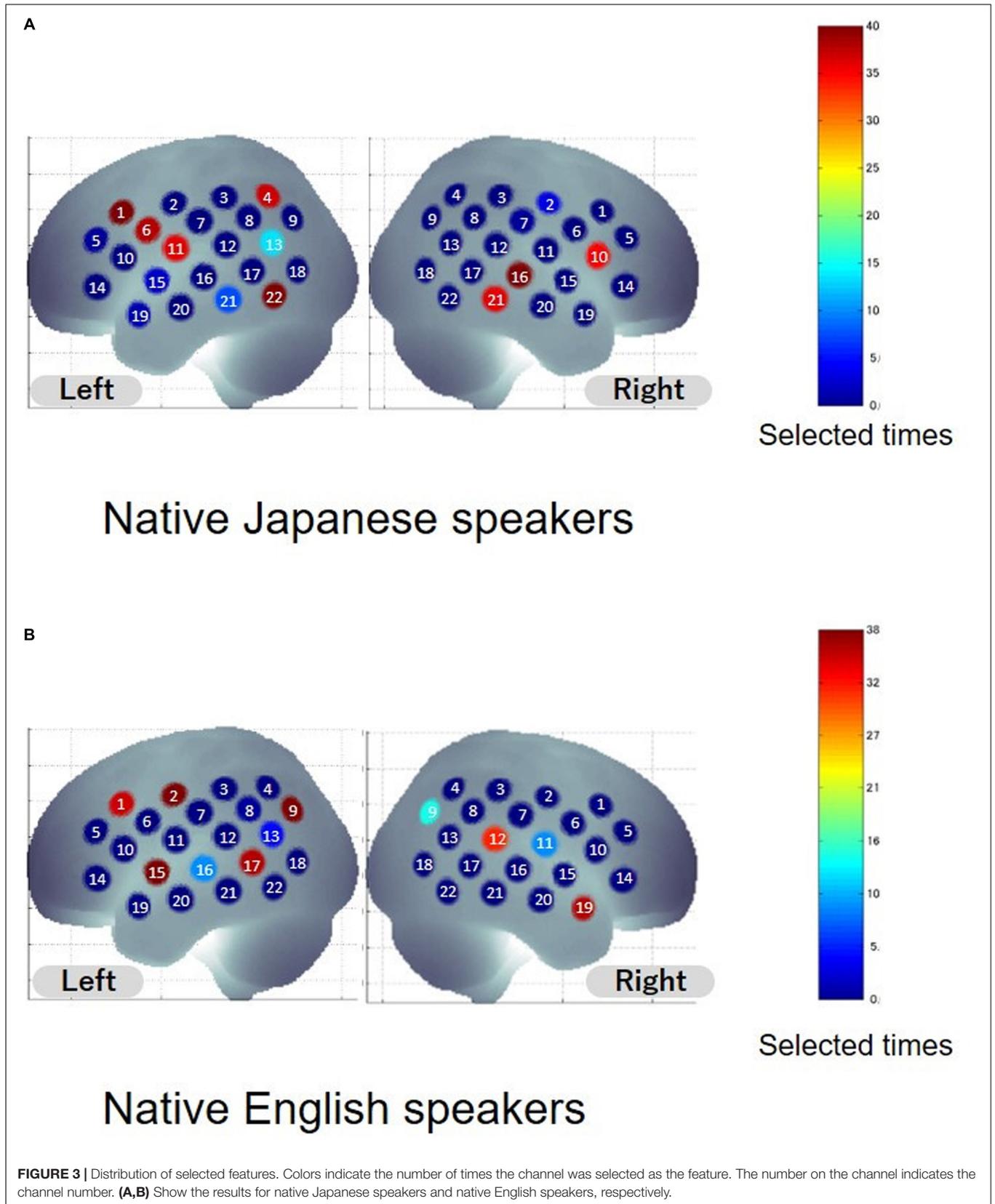
After feature selection, as expected, classification accuracy for the second language was improved; meanwhile, classification accuracies for the first and unknown languages tend to be at

the chance level. On the other hand, the SVM showed higher accuracy for L3 in the case of the native Japanese speakers and for L1 in the case of the native English speakers. SLR also showed higher accuracies for L1 in both cases. When K-nearest neighbor was used, classification accuracy tended to be reasonable; that is, it showed higher classification accuracy for L2. After feature selection, all the algorithms showed higher classification accuracy for L2. These results suggest that the informative features, which are related to second-language proficiency, are important for improving classification accuracy.

## DISCUSSION

In this study, machine-learning methods—using activation patterns in fNIRS data— were used to classify individuals with high second-language proficiency or low second-language proficiency, in the case of both native Japanese speakers and native English speakers. After feature selection, all methods showed higher classification accuracy for the second language, suggesting that the validity of feature selection. Also, the activation patterns of frontal-temporal region are important indicators to estimate individual language proficiency.

In the field of neuroimaging studies applying machine learning, it is believed that to achieve better classification accuracy, informative features must be extracted (Norman et al., 2006; De Martino et al., 2008; Pereira et al., 2009; An et al., 2017). The higher classification performance demonstrated in this study indicates the validity of feature selection. Optimal feature extraction avoids over-fitting and eliminates the effects of noisy variables that are irrelevant to the classification problem. How to extract intrinsic features is an important research focus. SCCA was used as one of the methods for extracting informative features for individual estimation (Yahata et al., 2016). CCA can derive projection vectors that have maximum correlation with desired labels (e.g., a label for



**TABLE 2 |** Spatial and anatomical information of selected features.

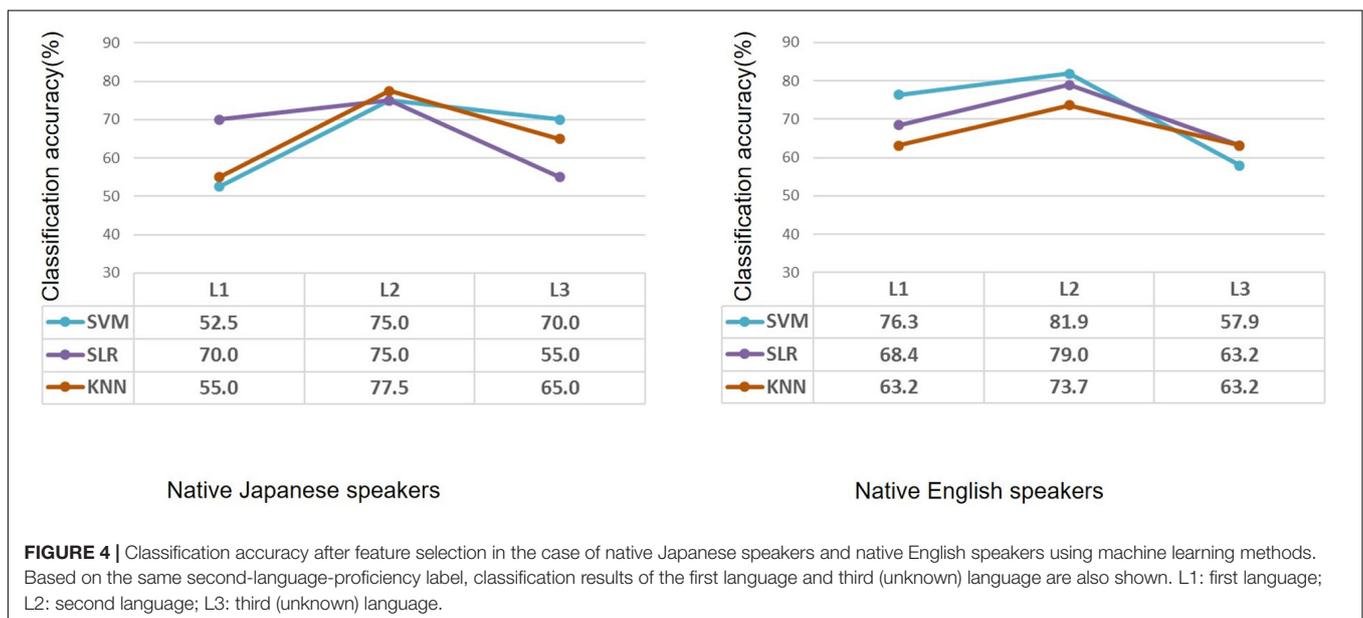
Channels	MNI coordinates		Anatomical information
	x,y,z		
Japanese speakers	Left Ch 1	-50, 22, 38	Left middle frontal gyrus (BA 44)
	Ch 6	-61, 7, 28	Left precentral gyrus (BA 6)
	Ch 22	-61, -62, -8	Left inferior temporal gyrus (BA 37)
English speakers	Right Ch 16	71, -25, 4	Right superior temporal gyrus (BA 21)
	Left Ch 2	-60, -7, 43	Left postcentral gyrus (BA 4)
	Ch 9	-52, -72, 35	Left angular gyrus (BA 39)
	Ch 15	-62, 2, 1	Left superior temporal gyrus (BA 48)
	Ch 17	-67, -50, 6	Left middle temporal gyrus (BA 21)
Right Ch 19	58, 10, -18	Right middle temporal gyrus (BA 21)	

language proficiency). Using L1-norm regularization will lead to sparse solutions. As a result, features only related to desired labels can be extracted, so label-unrelated variables can be eliminated. Conventional methods of feature selection need careful engineering and considerable domain expertise to design a feature extractor that transforms raw data into an appropriate feature vector. SCCA allows an input to be composited from raw data; thus, it makes it possible to automatically extract the informative features required for the classification task.

Analyzing the most-discriminative features shared between high- and low-language-proficiency groups revealed that native Japanese speakers and native English speakers utilize different specific brain regions, but they show the same tendency, that is, Broca’s area, Wernicke’s area and the temporal cortex. The reason for activation of different specific brain regions may be due to the differences between brain shapes of native Japanese speakers and native English speakers; consequently, specific brain regions may deviate during spatial registration

of measurement channels (Chee et al., 2011). Previous studies have found evidence that the two languages extensively overlap in regard to the classical language areas, namely Broca’s area and Wernicke’s area. Specifically, a variety of regions, including the left frontal region (Price et al., 1999; Lehtonen et al., 2005; Abutalebi and Green, 2008) and the bilateral supramarginal gyri (Price et al., 1999) have been observed to be involved in bilingual language comprehension and processing. Those studies also suggested that no single region is responsible for language comprehension and processing. Moreover, multiple studies have suggested that the bilateral temporal-frontal network is involved in processing during auditory language comprehension (Friederici, 2002; Price, 2012; Fengler et al., 2016). Concretely, syntactic and semantic information are processed predominately by the left hemisphere, while processing of prosodic information occurs predominately in the right hemisphere (Friederici et al., 2000; Friederici, 2002). Studies on sentence-comprehension tasks also reported left laterality plays a primary and significant role in language comprehension (Harrington et al., 2006; Sanjuán et al., 2010; Niskanen et al., 2012). In the present study, the brain region of selected features are consistent with the previous findings; namely, multiple cortical regions in a temporal-frontal network were observed to be related to language comprehension irrespective of native language, and the informative features in these brain regions play an important role in improving classification accuracy.

Using the label of second-language proficiency, classification of L1 and L3 results in a higher classification accuracy than the chance level. One possible explanation of this result is the relation between the second-language ability and the native-language ability (Brevik et al., 2016; Guo, 2018). In addition, during the fNIRS-measurement experiment, the psychological stressors of the high-language-proficiency group and the low-language-proficiency group may differ. Since all the languages



were randomly presented, the tests for L1 and L3 may be affected by the different psychological stressors.

## CONCLUSION

Machine-learning methods were used for distinguishing second-language proficiency individually for both native Japanese speakers and native English speakers. By extracting informative features, the machine-learning methods showed higher classification accuracy for the second language. The informative features showed the effectiveness of feature selection in improving classification accuracy. Moreover, brain-activation patterns measured by fNIRS have the potential to serve as biomarkers for identifying language proficiency. Finally, the same approach could potentially be used with other biological data with similar characteristic to those of fNIRS data.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## REFERENCES

- Abutalebi, J., and Green, D. W. (2008). Control mechanisms in bilingual language production: neural evidence from language switching studies. *Lang. Cogn. Process.* 23, 557–582. doi: 10.1080/01690960801920602
- Akama, H., Murphy, B., Lei, M. M., and Poesio, M. (2014). Cross-participant modelling based on joint or disjoint feature selection: an fMRI conceptual decoding study. *Appl. Inform.* 1:1.
- An, L., Adeli, E., Liu, M., Zhang, J., Lee, S.-W., and Shen, D. (2017). A hierarchical feature and sample selection framework and its application for Alzheimer's disease diagnosis. *Sci. Rep.* 7:45269.
- Bellman, R. E. (2015). *Adaptive Control Processes: a Guided Tour*. Princeton, NJ: Princeton university press.
- Bortfeld, H., Fava, E., and Boas, D. A. (2009). Identifying cortical lateralization of speech processing in infants using near-infrared spectroscopy. *Dev. Neuropsychol.* 34, 52–65. doi: 10.1080/87565640802564481
- Brett, M., Johnsrude, I. S., and Owen, A. M. (2002). Opinion: the problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3, 243–249. doi: 10.1038/nrn756
- Brevik, L. M., Olsen, R. V., and Hellekjær, G. O. (2016). The complexity of second language reading: investigating the L1-L2 relationship. *Read. Foreign Lang.* 28, 161–182.
- Chee, M. W. L., Zheng, H., Goh, J. O. S., Park, D., and Sutton, B. P. (2011). Brain structure in young and old East Asians and Westerners: comparisons of structural volume and cortical thickness. *J. Cogn. Neurosci.* 23, 1065–1079. doi: 10.1162/jocn.2010.21513
- Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., and Initiative, A. S. D. N. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 60, 59–70. doi: 10.1016/j.neuroimage.2011.11.066
- Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205. doi: 10.1097/00004728-199403000-00005
- Cui, X., Bray, S., and Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49, 3039–3046. doi: 10.1016/j.neuroimage.2009.11.050

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by internal review board on Research & Development Group, Hitachi, Ltd. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

HS, ID, and TM conceived and designed the study. HS, TM, ID, and ML conducted the experiment. ML, TM, and HS carried out the analysis of the data. ML wrote the manuscript. All the authors reviewed the manuscript.

## ACKNOWLEDGMENTS

We thank Yoshiki Niwa, Yukiko Hirabayashi, Tsukasa Funane, Stephanie Sutoko, and Takashi Numata for their technical assistance in the fNIRS measurement experiment.

- De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., and Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* 43, 44–58. doi: 10.1016/j.neuroimage.2008.06.037
- Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015. doi: 10.1126/science.1077066
- Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern Classification*. Hoboken, NJ: John Wiley & Sons.
- ETS (2005). *TOEIC® Test New Official Preparation Guide*, Vol. 1. Tokyo: The Institute for International Business Communication.
- ETS (2007). *TOEIC® Test New Official Preparation Guide*, Vol. 2. Tokyo: The Institute for International Business Communication.
- ETS (2008). *TOEIC® Test New Official Preparation Guide*, Vol. 3. Tokyo: The Institute for International Business Communication.
- ETS (2012). *TOEIC® Test New Official Preparation Guide*, Vol. 5. Tokyo: The Institute for International Business Communication.
- Fan, Y., Rao, H., Hurt, H., Giannetta, J., Korczykowski, M., Shera, D., et al. (2007). Multivariate examination of brain abnormality using both structural and functional MRI. *Neuroimage* 36, 1189–1199. doi: 10.1016/j.neuroimage.2007.04.009
- Fengler, A., Meyer, L., and Friederici, A. D. (2016). How the brain attunes to sentence processing: relating behavior, structure, and function. *Neuroimage* 129, 268–278. doi: 10.1016/j.neuroimage.2016.01.012
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., et al. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671. doi: 10.1038/nn.4135
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/s1364-6613(00)01839-8
- Friederici, A. D., Meyer, M., and Von Cramon, D. Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain Lang.* 74, 289–300. doi: 10.1006/brln.2000.2313
- Guo, L. (2018). Modeling the relationship of metacognitive knowledge, L1 reading ability, L2 language proficiency and L2 reading. *Read. Foreign Lang.* 30, 209–231.
- Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.

- Hall, M., Chaudhary, U., Rey, G., and Godavarty, A. (2013). Fronto-temporal mapping and connectivity using NIRS for language-related paradigms. *J. Neurolinguistics* 26, 178–194. doi: 10.1016/j.jneuroling.2012.06.002
- Harrington, G. S., Buonocore, M., and Farias, S. T. (2006). Intrasubject reproducibility of functional MR imaging activation in language tasks. *Am. J. Neuroradiol.* 27, 938–944.
- Hong, K.-S., Khan, M. J., and Hong, M. J. (2018). Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces. *Front. Hum. Neurosci.* 12:246. doi: 10.3389/fnhum.2018.00246
- Hong, K.-S., and Yaqub, M. A. (2019). Application of functional near-infrared spectroscopy in the healthcare industry: a review. *J. Innov. Opt. Health Sci.* 12:1930012. doi: 10.1142/s179354581930012x
- Hosseini, R., Walsh, B., Tian, F., and Wang, S. (2018). An fNIRS-based feature learning and classification framework to distinguish hemodynamic patterns in children who stutter. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1254–1263. doi: 10.1109/tnsre.2018.2829083
- Hu, X.-S., Hong, K.-S., and Ge, S. S. (2013). Reduction of trial-to-trial variability in functional near-infrared spectroscopy signals by accounting for resting-state functional connectivity. *J. Biomed. Opt.* 18:017003. doi: 10.1117/1.jbo.18.1.017003
- Lehtonen, M. H., Laine, M., Niemi, J., Thomsen, T., Vorobyev, V. A., and Hugdahl, K. (2005). Brain correlates of sentence translation in Finnish-Norwegian bilinguals. *Neuroreport* 16, 607–610. doi: 10.1097/00001756-200504250-00018
- Lei, M., Akama, H., and Murphy, B. (2014). Neural basis of language switching in the brain: fMRI evidence from Korean-Chinese early bilinguals. *Brain Lang.* 138, 12–18. doi: 10.1016/j.bandl.2014.08.009
- Lei, M., Miyoshi, T., Niwa, Y., Dan, I., and Sato, H. (2018). Comprehension-dependent cortical activation during speech comprehension tasks with multiple languages: functional near-infrared spectroscopy study. *Jpn. Psychol. Res.* 60, 300–310. doi: 10.1111/jpr.12218
- Li, J., Qiu, L., Xu, L., Pedapati, E. V., Erickson, C. A., and Sunar, U. (2016). Characterization of autism spectrum disorder with spontaneous hemodynamic activity. *Biomed. Opt. Express* 7, 3871–3881.
- Luu, S., and Chau, T. (2008). Decoding subjective preference from single-trial near-infrared spectroscopy signals. *J. Neural Eng.* 6:016003. doi: 10.1088/1741-2560/6/1/016003
- Maki, A., Yamashita, Y., Ito, Y., Watanabe, E., Mayanagi, Y., and Koizumi, H. (1995). Spatial and temporal analysis of human motor activity using noninvasive NIR topography. *Med. Phys.* 22, 1997–2005. doi: 10.1118/1.597496
- Mwangi, B., Tian, T. S., and Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics* 12, 229–244. doi: 10.1007/s12021-013-9204-3
- Naseer, N., and Hong, K.-S. (2013). Classification of functional near-infrared spectroscopy signals corresponding to the right-and left-wrist motor imagery for development of a brain-computer interface. *Neurosci. Lett.* 553, 84–89. doi: 10.1016/j.neulet.2013.08.021
- Nieuwenhuis, M., Van Haren, N. E., Pol, H. E. H., Cahn, W., Kahn, R. S., and Schnack, H. G. (2012). Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *Neuroimage* 61, 606–612. doi: 10.1016/j.neuroimage.2012.03.079
- Niskanen, E., Könönen, M., Villberg, V., Nissi, M., Ranta-Aho, P., Säisänen, L., et al. (2012). The effect of fMRI task combinations on determining the hemispheric dominance of language functions. *Neuroradiology* 54, 393–405. doi: 10.1007/s00234-011-0959-7
- Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005
- Obrig, H., Rossi, S., Telkemeyer, S., and Wartenburger, I. (2010). From acoustic segmentation to language processing: evidence from optical imaging. *Front. Neuroenergetics* 2:13. doi: 10.3389/fnene.2010.00013
- Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., and Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* 36, 1140–1152. doi: 10.1016/j.neubiorev.2012.01.004
- Pena, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11702–11705. doi: 10.1073/pnas.1934290100
- Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45, S199–S209.
- Power, S. D., Falk, T. H., and Chau, T. (2010). Classification of prefrontal activity due to mental arithmetic and music imagery using hidden Markov models and frequency domain near-infrared spectroscopy. *J. Neural Eng.* 7:26002.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847. doi: 10.1016/j.neuroimage.2012.04.062
- Price, C. J., Green, D. W., and Von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain* 122(Pt 12), 2221–2235. doi: 10.1093/brain/122.12.2221
- Qin, S., Young, C. B., Duan, X., Chen, T., Supekar, K., and Menon, V. (2014). Amygdala subregional structure and intrinsic functional connectivity predicts individual differences in anxiety during early childhood. *Biol. Psychiatry* 75, 892–900. doi: 10.1016/j.biopsych.2013.10.006
- Quaresima, V., Bisconti, S., and Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain Lang.* 121, 79–89. doi: 10.1016/j.bandl.2011.03.009
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., et al. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cereb. Cortex* 15, 1676–1689. doi: 10.1093/cercor/bhi044
- Rorden, C., and Brett, M. (2000). Stereotaxic display of brain lesions. *Behav. Neurol.* 12, 191–200. doi: 10.1155/2000/421719
- Saeyes, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics* 23, 2507–2517. doi: 10.1093/bioinformatics/btm344
- Sanjuán, A., Bustamante, J.-C., Forn, C., Ventura-Campos, N., Barrós-Loscertales, A., Martínez, J.-C., et al. (2010). Comparison of two fMRI tasks for the evaluation of the expressive language function. *Neuroradiology* 52, 407–415. doi: 10.1007/s00234-010-0667-8
- Santosa, H., Jiyoun Hong, M., Kim, S.-P., and Hong, K.-S. (2013). Noise reduction in functional near-infrared spectroscopy signals by independent component analysis. *Rev. Sci. Instrum.* 84:073106. doi: 10.1063/1.4812785
- Sasai, S., Homae, F., Watanabe, H., Sasaki, A. T., Tanabe, H. C., Sadato, N., et al. (2012). A NIRS-fMRI study of resting state network. *Neuroimage* 63, 179–193. doi: 10.1016/j.neuroimage.2012.06.011
- Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., et al. (2012). Cerebral hemodynamics in newborn infants exposed to speech sounds: a whole-head optical topography study. *Hum. Brain Mapp.* 33, 2092–2103. doi: 10.1002/hbm.21350
- Sato, H., Takeuchi, T., and Sakai, K. L. (1999). Temporal cortex activation during speech recognition: an optical topography study. *Cognition* 73, B55–B66.
- Schecklmann, M., Ehlis, A. C., Plichta, M. M., and Fallgatter, A. J. (2008). Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency. *Neuroimage* 43, 147–155. doi: 10.1016/j.neuroimage.2008.06.032
- Scholkman, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., et al. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27. doi: 10.1016/j.neuroimage.2013.05.004
- Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39, 1064–1080. doi: 10.1016/j.neuroimage.2007.09.031
- Sutoko, S., Monden, Y., Tokuda, T., Ikeda, T., Nagashima, M., Kiguchi, M., et al. (2019). Distinct methylphenidate-evoked response measured using functional near-infrared spectroscopy during Go/No-Go task as a supporting differential diagnostic tool between attention-deficit/hyperactivity disorder and autism spectrum disorder comorbid children. *Front. Hum. Neurosci.* 13:7. doi: 10.3389/fnhum.2019.00007
- Tak, S., and Ye, J. C. (2014). Statistical analysis of fNIRS data: a comprehensive review. *Neuroimage* 85, 72–91. doi: 10.1016/j.neuroimage.2013.06.016
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

- Tsuzuki, D., Jurcak, V., Singh, A. K., Okamoto, M., Watanabe, E., and Dan, I. (2007). Virtual spatial registration of stand-alone fNIRS data to MNI space. *Neuroimage* 34, 1506–1518. doi: 10.1016/j.neuroimage.2006.10.043
- Witten, D. M., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10, 515–534. doi: 10.1093/biostatistics/kxp008
- Yahata, N., Morimoto, J., Hashimoto, R., Lisi, G., Shibata, K., Kawakubo, Y., et al. (2016). A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat. Commun.* 7:11254.
- Yamashita, O. (2009). *Sparse Logistic Regression Toolbox*. Available online at: [https://bicr.atr.jp/~oyamashi/SLR\\_WEB.html](https://bicr.atr.jp/~oyamashi/SLR_WEB.html) (accessed June 17, 2020).
- Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

**Conflict of Interest:** ML and TM was employed by the company Hitachi, Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lei, Miyoshi, Dan and Sato. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.