frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 28 April 2021
doi: 10.3389/fnhum.2021.636191

Decoding Multiple Sound-Categories in the Auditory Cortex by Neural Networks: An fNIRS Study

So-Hyeon Yoo[1], *Hendrik Santosa*[2], *Chang-Seok Kim*[3] *and Keum-Shik Hong*[1]*

[1] *School of Mechanical Engineering, Pusan National University, Busan, South Korea,* [2] *Department of Radiology, University of Pittsburgh, Pittsburgh, PA, United States,* [3] *Department of Cogno-Mechatronics Engineering, Pusan National University, Busan, South Korea*

OPEN ACCESS

Edited by:
Andreas K. Maier,
University of Erlangen Nuremberg,
Germany

Reviewed by:
Evangelos Paraskevopoulos,
Aristotle University of Thessaloniki,
Greece
Giancarlo Valente,
Maastricht University, Netherlands

***Correspondence:**
Keum-Shik Hong
kshong@pusan.ac.kr

Specialty section:
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *01 December 2020*
Accepted: *31 March 2021*
Published: *28 April 2021*

Citation:
Yoo S-H, Santosa H, Kim C-S and Hong K-S (2021) Decoding Multiple Sound-Categories in the Auditory Cortex by Neural Networks: An fNIRS Study.
Front. Hum. Neurosci. 15:636191.
doi: 10.3389/fnhum.2021.636191

This study aims to decode the hemodynamic responses (HRs) evoked by multiple sound-categories using functional near-infrared spectroscopy (fNIRS). The six different sounds were given as stimuli (English, non-English, annoying, nature, music, and gunshot). The oxy-hemoglobin (HbO) concentration changes are measured in both hemispheres of the auditory cortex while 18 healthy subjects listen to 10-s blocks of six sound-categories. Long short-term memory (LSTM) networks were used as a classifier. The classification accuracy was 20.38 ± 4.63% with six class classification. Though LSTM networks' performance was a little higher than chance levels, it is noteworthy that we could classify the data subject-wise without feature selections.

Keywords: functional near-infrared spectroscopy (fNIRS), long short-term memories (LSTMs), auditory cortex, decoding, deep learning

INTRODUCTION

Recognizing sound is one of the important senses in everyday life. People are always exposed to a variety of sounds, and they can know what sound it is without being conscious. This ability allows people to avoid various dangers and facilitates communication with others. The auditory stimulus that enters through the outer ear is transmitted to the auditory cortex through the auditory nerve. It is clear that the temporal cortex is activated differently by different sounds. Neural responses in the auditory cortices have been studied using diverse modalities like electroencephalography (EEG; Wong et al., 2007; Hill and Scholkopf, 2012; Liu et al., 2015), magnetoencephalography (MEG; Hyvarinen et al., 2015), eletrocorticogram (ECoG; Pasley et al., 2012; Herff et al., 2015), functional magnetic resonance imaging (fMRI; Wong et al., 2008; Gao et al., 2015; Zhang et al., 2016), functional near-infrared spectroscopy (fNIRS; Plichta et al., 2011; Kovelman et al., 2012; Dewey and Hartley, 2015), and multimodal imaging (i.e., concurrent fNIRS, fMRI, and/or MEG; Kovelman et al., 2015; Corsi et al., 2019) to identify this process. In these studies, the complexities of brain responses evoked by the perception of sounds have been investigated to improve the quality of life.

Griffiths and Warren (2004) argued that analyzing auditory objects in the two-dimensional space (frequency and time) rather than one-dimensional space (frequency or time) is more meaningful, and thus acoustic experiences that produce two-dimensional images need to be investigated. But, they did not provide specific sound categories in their work. Theunissen and Elie (2014) showed that natural sounds facilitate the characterizations of the stimulus-response functions for neurons than white noise or simple synthetic sounds. Salvari et al. (2019) demonstrated significant activation and interconnection differences between natural sounds and human-made object sounds (music and artificial sounds) in the prefrontal areas using MEG. However, there were no significant differences between music and artificial sounds. Liu et al. (2019) demonstrated that predictions of tuning properties of putative feature-selective neurons match data from the marmoset's primary auditory cortex. Also, they showed that the exact algorithm of marmoset's call classification could successfully be applied in call classification in other species.

Identifying the sound that a person hears using a brain-computer interface (BCI) enables us to know what the person is hearing. The more diverse sounds a BCI device can discern, the more variant conditions are identified. For those who have lost vision, sound may be an alternative tool to communicate with other people in a non-contact way. If it is possible to classify more sounds, we can increase the control commands for an external device. Zhang et al. (2015) have researched decoding brain activation from multiple sound categories in the human temporal cortex. Seven different sound categories (English, non-English, vocal, Animal, mechanical, music, and nature) were used for classification in their fMRI work. They reported sound-category-selective brain maps showing distributed patterns of brain activity in the superior temporal gyrus and the middle temporal gyrus. However, analyses of such responses were hampered by the machine noise produced during fMRI experiments (Scarff et al., 2004; Fuchino et al., 2006).

fNIRS is a non-invasive brain imaging method that uses near-infrared light (700–900 nm) to penetrate the head and records oxygenation changes in the cerebral blood flow. fNIRS is a promising method for analyzing sound and speech processing. Compared to fMRI, fNIRS measurement is not noisy, and such measurements can be made in an environment more conducive to infant studies. Owing to these advantages, fNIRS shows significant potential for real-time brain monitoring while the subject is moving. According to fNIRS analyses, newborns consistently exhibit a strong hemodynamic response to universally preferred syllables, which suggests that the early acquisition and perception of language can be detected using categorical linguistic sounds (Gomez et al., 2014). The applications of this technology have the potential to provide feedback for speech therapy or in the tuning of hearing aid devices (e.g., cochlear implants) at an early stage of development based on brain recordings (Mushtaq et al., 2020). Several groups have demonstrated fNIRS use for measuring brain responses in deaf children with cochlear implants (Sevy et al., 2010; Pollonini et al., 2014).

In fNIRS applications, classification has been used in lie detection (Bhutta et al., 2015), drowsiness detection (Khan and Hong, 2015), mental workload detection (Herff et al., 2014), brain disease (Yoo et al., 2020), and the fNIRS-EEG-based hybrid BCI (Yuan et al., 2019; Lin et al., 2020). In fMRI applications, classification has also been used to decode the brain responses evoked by sight (Kohler et al., 2013; Smith, 2013) and sound (Staeren et al., 2009; Zhang et al., 2015). Lotte et al. (2007) and Pereira et al. (2009) reviewed the classification algorithms for EEG and fMRI data, respectively.

Recently, numerous studies have focused on improving classification accuracy by applying deep learning technology to brain signal classification, in addition to artificial neural networks (ANNs; Badai et al., 2020; Flynn et al., 2020). Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are representative forms of ANNs. The CNN is robust in processing large image data sets (Lee et al., 2020). It has been widely implemented in brain signal processing, including fMRI (Erturk et al., 2020), deep brain stimulation (Kakusa et al., 2020), and EEG (Lun et al., 2020). Besides, CNNs have been used to diagnose brain diseases in the fNIRS domain (Xu et al., 2019a; Yang et al., 2019, 2020). RNNs are capable of predicting and classifying sequential data. They have been widely applied in robotics for various purposes and systems such as obstacle avoidance control (Xu et al., 2019b; Zheng et al., 2019; Zhao et al., 2020), self-organizing robot control (Smith et al., 2020), collision-free compliance control (Zhou et al., 2019), dynamic neural robots (Tekulve et al., 2019), and self-driving system (Chen et al., 2019). Recently, RNNs have achieved impressive results in detecting seizures (Sirpal et al., 2019), brain injuries (Ieong et al., 2019), and pain (Hu et al., 2019b), as well as in discriminating attention-deficit hyperactivity disorder (Dubreuil-Vall et al., 2020).

Long short-term memory (LSTM) is a type of RNN incorporating a progressive model (Hochreiter and Schmidhuber, 1997). Compared to RNN, LSTM networks possess a "gate" to reduce the vanishing gradient problem and allow the algorithm to more precisely control the information that needs to be retained in memory and the information that must be removed. LSTM is also considered superior to RNNs when handling large sequences of data. Additionally, compared to CNN, it exhibits better performance in classifying highly dynamic nonlinear time-series data such as EEG data (Tsiouris et al., 2018; Li et al., 2020).

This study aims to develop a communication method for the completely paralyzed with no vision. We identify the sound that a person hears by measuring task-evoked hemodynamic responses from the auditory cortex: In our early work (Hong and Santosa, 2016), four sound categories were classified. When remotely communicating with people without vision, visual or motor cortex-based BCIs may not be applicable. Sound will be a vital tool to communicate. In this article, we increased the number of sound categories from four to six. The more diverse sounds are classified, the more variant conditions are identified. Eventually, we can diversify the control commands to operate an external device. Sound-based BCI using audio stimuli is promising because we can use such audio signals in our daily lives (i.e., a passive BCI is possible). In this article, auditory-evoked HRs are measured using fNIRS, and subsequently, LSTM is applied to analyze fNIRS' ability to distinguish individual sounds out of six classes.

MATERIALS AND METHODS

Subjects

A total of 18 subjects participated in the experiment (age: 26.89 ± 3.49 years; seven females, two left-handed). All subjects had normal hearing and no history of any neurological disorder. All subjects were informed about the nature and purpose of the respective experiments before obtaining their written consent. For the experiment, each subject lay down on a bed. All subjects were asked to remain relaxed, close their eyes, and avoid significant body movements during the experiment. The subjects were asked to listen attentively to various audio stimuli and guess the category of each stimulus. After the experiment, all participants were asked to explain verbally whether they could precisely distinguish what they heard. The fNIRS experimentation was done on healthy subjects and the entire experimental procedure was carried out in accordance with the Declaration of Helsinki and guidelines approved by the Ethics Committee of the Institutional Review Board of Pusan National University.

Audio Stimuli

The audio stimuli consisted of six different sound categories selected from a popular website (http://www.youtube.com). As shown in **Table 1**, the first and second categories entailed speech in English and other languages (non-English). The subjects were Indonesian, Korean, Chinese, Vietnamese, and Pakistani. Each participant had a common recognition of English but failed to recognize the other languages. The third and fourth categories were annoying sounds and nature sounds. The fifth category was a segment of classical music (Canon in D by Pachelbel). The sixth category was gunshot sounds at a frequency of 1 Hz. Each category consisted of six different sounds (except the gunshots, which had the same repeated sound). Each subject was exposed to 36 trials (i.e., six sound categories × six trials). The audio stimuli were presented in a pseudo-randomized order. Each stimulus consisted of 10 s of the sound followed by 20 s of silence.

Additionally, pre- and post-trials of classical music were added (to avoid sudden hearing), neither of which was included in the data processing. Accordingly, the entire fNIRS recording lasted for approximately 19 min. All audio stimuli were digitally mixed using the Adobe Audition software (MP3-format file: 16-bit quantification, 44.1 kHz sampling, stereo channel) and normalized to the same intensity level. Active noise-cancellation earbuds (Sony MDR-NC100D) were utilized for acoustic stimulation of all subjects with the same sound-level setting. After each fNIRS recording session, all subjects reported that they could accurately distinguish the sound among the sound categories for all trials.

fNIRS Measurements

Figure 1 shows the continuous-wave fNIRS system's optode configuration (DYNOT: DYnamic Near-infrared Optical Tomography; NIRx Medical Technologies, Brooklyn, NY, USA) for bilateral imaging of the auditory cortex in both hemispheres. The emitter–detector distance was 23 mm, while the sampling rate was set to 1.81 Hz at two wavelengths (760 and 830 nm). The optode configuration consisted of 3 × 5 arrays (eight emitters and seven detectors) with 22 channels for each hemisphere. The two 22-channel sets were placed on the scalp, covering the left (Chs. 1–22) and right (Chs. 23–44) temporal lobes. According to the International 10-20 System, Chs. 16 and 38 were placed at T2 and T4, respectively (Santosa et al., 2014). In the left hemisphere, both Broca's area and Wernicke's area were covered by this configuration. Finally, the lights in the room were switched off to minimize signal contamination from ambient light sources during the experiments.

Preprocessing

The optical data of two wavelengths were converted into relative oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) concentration changes using the modified Beer-Lambert law (Hiraoka et al., 1993) using MATLAB^{TM} (2020b, MathWorks, USA). Owing to the uniform emitter-detector distance, constant values of differential path-length factors were used for all channels (i.e., d = 7.15 for λ = 760 nm and d = 5.98 for λ = 830 nm). In previous studies, HbO data were activated significantly higher than HbR for given stimuli. Therefore, only HbO data were processed in this study. The HbO data were filtered to remove physiological and artificial noises using the fifth-order Butterworth band-pass filter with cutoff frequencies of 0.01 Hz and 0.1 Hz. The filtered data was chopped for each trial.

TABLE 1 | Audio categories (M: male, F: female).

Trial	Human vocal hearing		Nonvocal hearing			
	English	Non-English	Annoying sound	Nature sound	Music	Gunshot
1	M	Russian (F)	Baby cry	River	Canon in D	10 times
2	F	German (F)	Car alarm	Forest (day time)	Canon in D	10 times
3	M	French (F)	Police siren	Rain	Canon in D	10 times
4	MF*	Bulgarian (MF*)	Horror sound	Jungle	Canon in D	10 times
5	F	Italian (MF)	Male scream	Ocean waves	Canon in D	10 times
6	F	Japanese (F)	Nuclear alarm siren	Waterfall	Canon in D	10 times

MF denotes male–female conversation.

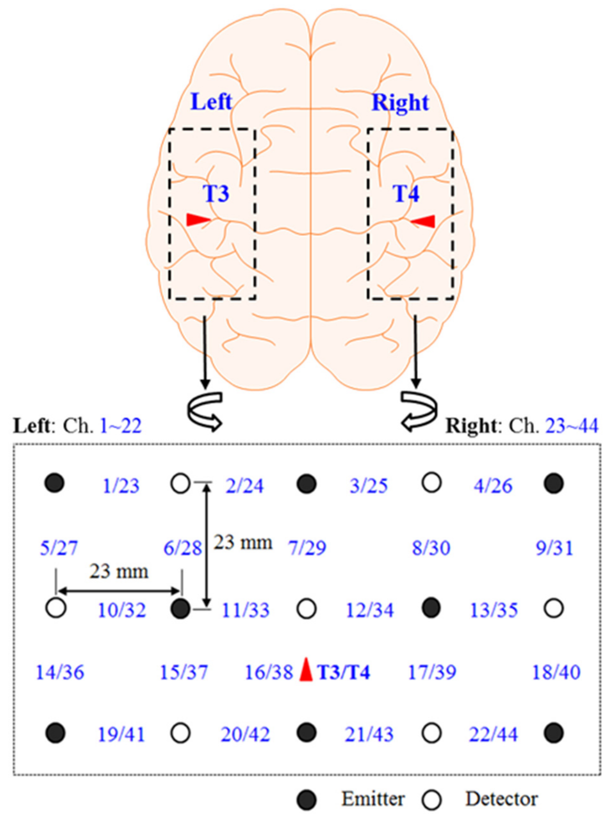


FIGURE 1 | Optode configuration: the numbers represent the measurement channels, where Chs. 16 and 38 coincide with T3 and T4 locations in the International 10–20 system (Hong and Santosa, 2016).

Feature Extraction for Support Vector Machine

The mean, slope, kurtosis, and skewness values of HbO signals were used as support vector machine (SVM) features. SVM classification was performed twice: One for "within-subject" and the other for "across-subject." Within-subject classification is a standard classification method for the fNIRS study. Considering the total number of trials for one subject was 36, 6-fold cross-validation was performed for each subject. For the across-subject classification, we used the entire data set. In this case, the total number of trials was 648 (i.e., multiplication of the number of subjects and the number of trials. Ten-fold and leave-one-out cross-validation were performed for the across-subject classification. Training and testing sets were divided randomly by MATLAB^{TM} function *cvpartition* for cross-validation. The same data partitions were used for SVM and LSTM.

LSTM

A recurrent neural network (RNN) is a type of artificial neural network wherein hidden nodes are connected with directional edges as a directed cycle. It is well known as an effective tool to process sequential data such as voice and handwriting. The RNN has the following structure (Hochreiter and Schmidhuber, 1997):

$$y_t = W_{hy}\, h_t + b_y, \tag{1}$$
$$h_t = \tanh\left(W_{hh}\, h_{t-1} + W_{xh}\, x_t + b_h\right), \tag{2}$$

where y_t indicates the output of the present state; subscript t is the discrete time step; W_{hy}, W_{hh}, and W_{xh} are the parameters from layer to layer; b_y is the bias of the output y; h_t is the hidden state vector; x_t is the input vector; and b_h is the bias of the hidden state vector h.

The LSTM is a special kind of recurrent neural network, compensating for the vanishing gradient problem. It has a structure of cell-states in the hidden state of RNN. The basic formulas for LSTM are as follows.

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_f), \tag{3}$$
$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + b_o), \tag{4}$$
$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o), \tag{5}$$
$$g_t = \tanh(W_{xg} x_t + W_{hg} h_{t-1} + b_g), \tag{6}$$
$$c_t = f_t \circ c_{t-1} + i_t \circ g_t, \tag{7}$$
$$h_t = o_t \circ \tanh(c_t), \tag{8}$$

where f_t is the activation vector of forgetting gate to forget past information; i_t is the activation vector of input gate to memorize the current information; o_t is the activation vector of output gate; g_t is the activation vector of the cell input; c_t is the cell state vector; W_{xf}, W_{hf}, W_{xi}, W_{hi}, W_{xo}, W_{ho}, W_{xg}, and W_{hg} are the weight matrices of the input and recurrent connections; b_f, b_i, b_o, and b_g are the parametric bias vectors; and $\circ$ is the Hadamard product. In the LSTM networks, cell state and hidden state are calculated recursively.

In this article, LSTM networks are applied in two ways, like the two cases (within-subject, across-subject) in the SVM classification. **Figure 2** represents the LSTM networks used in this article. First, for within-subject classification, a bi-LSTM layer of eight hidden layers was used with two maximum epochs and three mini-batch sizes (Kang et al., 2020). Second, the bi-LSTM layer of 16 hidden layers was used for across-subject classification with three maximum epochs and three mini-batch sizes, see **Table 2**. The number of hidden layers, maximum epoch, and mini-batch size were selected to avoid overfitting (Sualeh and Kim, 2019). Additionally, a bi-LSTM layer of 256 hidden layers was examined for across-subject classification (to compare with 16 hidden layers). Also, 6- and 10-fold and leave-one-out cross-validations were performed in the same way as SVM.

RESULTS

In the experiment, a total of 18 subjects listened to six repetitions of each of the six categories of sound stimuli (36 total trials). The six categories were English (E), non-English (nE), nature sounds (NS), music (M), annoying sounds (AS), and gunshot (GS). The within-subject classification accuracies were 21.35 ± 6.71% for SVM and 19.14 ± 9.16% for LSTM, respectively; see **Figure 3**.

When the cross-validations of SVM and LSTM were performed separately, the accuracies of the 10-fold across-subject

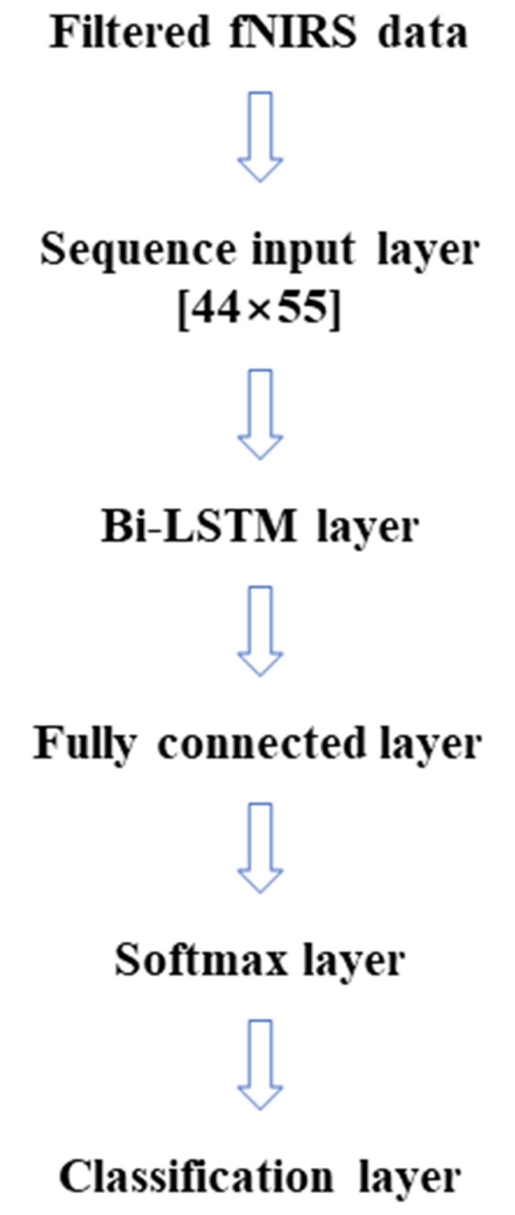


FIGURE 2 | Simple bi-LSTM model for classification.

TABLE 2 | A single bi-LSTM network.

	bi-LSTM structure		
	Across-subject		**Within-subject**
	10-fold	**Leave-one-out**	
Input size	44 × 55		44 × 55
Training data set	584	612	30
Testing data set	64	36	6
The number of hidden layers	16		8

classification with 16 hidden layers were 16.83 ± 3.90% for SVM and 20.38 ± 4.63% for LSTM, respectively. **Figure 4** shows the confusion matrices for training and testing for the 10-fold across-subject classification. The hypergeometric *p*-values were calculated using the confusion matrix in **Figure 4B**. The *p*-values were 0.3745 (E), 0.3123 (nE), 0.0232 (NS), 0.0946 (M), 0.0129 (AS), and 0.3701 (GS). For a fair comparison between SVM and LSTM, we repeated the 10-fold cross-validation using the same data partitioning using the same code. In this case, the results were 15.73 ± 3.00% for SVM and 21.44 ± 4.57% for LSTM. Henceforth, we could not find a significant difference between the two cases regarding the data partitioning method.

Using the same data partitioning for SVM and LSTM, we also performed the leave-one-out across-subject classification. The results were 16.83 ± 3.90% for SVM and 20.52 ± 6.15% for LSTM. **Figure 5** shows the confusion matrices for training and testing in the leave-one-out case. The hypergeometric *p*-values were calculated using the confusion matrix in **Figure 5B**. The *p*-values were 0.4274 (E), 0.5488 (nE), 0.0129 (NS), 0.0036 (M), 0.1078 (AS), and 0.3769 (GS). The low *p*-values indicate that the classifier could successfully classify the sounds.

LSTM showed better accuracies in the across-subject classification but worse accuracies in the within-subject classification: This result was somewhat unexpected in comparison to the four-sound case (Hong and Santosa, 2016). It suggests that, in the six categories case, the subjects heard too many sound-categories, and they had difficulty in distinguishing them. Overall, when there are many hidden layers in the classifier, training becomes better than when there are few, but overfitting to the training data occurs. It is noted that the across-subject classification accuracies of LSTM with 256 hidden layers were 99.9% for training and 23.15% for testing, which is an overfitting case.

DISCUSSIONS

The previous studies in the literature have shown that various sound categories were processed differently in the brain. Staeren et al. (2009) showed that different sound categories evoked significant BOLD responses in a large expanse of the auditory cortex, including bilaterally the Heschl's gyrus, the superior temporal gyrus, and the upper bank of the superior temporal sulcus. Zhang et al. (2015) revealed that sound category-selective brain maps demonstrated distributed brain activity patterns in the superior temporal gyrus and the middle temporal gyrus. Plichta et al. (2011) reported that pleasant and unpleasant sounds increased auditory cortex activation than neutral sounds in their fNIRS research. Our results showed that nature sounds, music, and annoying sounds were classified better than other categories. Nature sounds and music are considered pleasant sounds, and annoying sounds are unpleasant sounds (Plichta et al., 2011). Classifying emotionally-neutral sounds (i.e., E, nE, and GS) "individually" from other categories is considered difficult.

Deep learning algorithms have been developed to increase classification accuracy and stability (Shan et al., 2019; Park and Jung, 2020; Sung et al., 2020). RNNs have been developed to improve their performance likewise; memristor-based RNNs (Yang et al., 2021), chaotic delayed RNNs with unknown parameters and stochastic noise (Yan et al., 2019), reformed recurrent Hermite polynomial neural network (Lin and Ting,

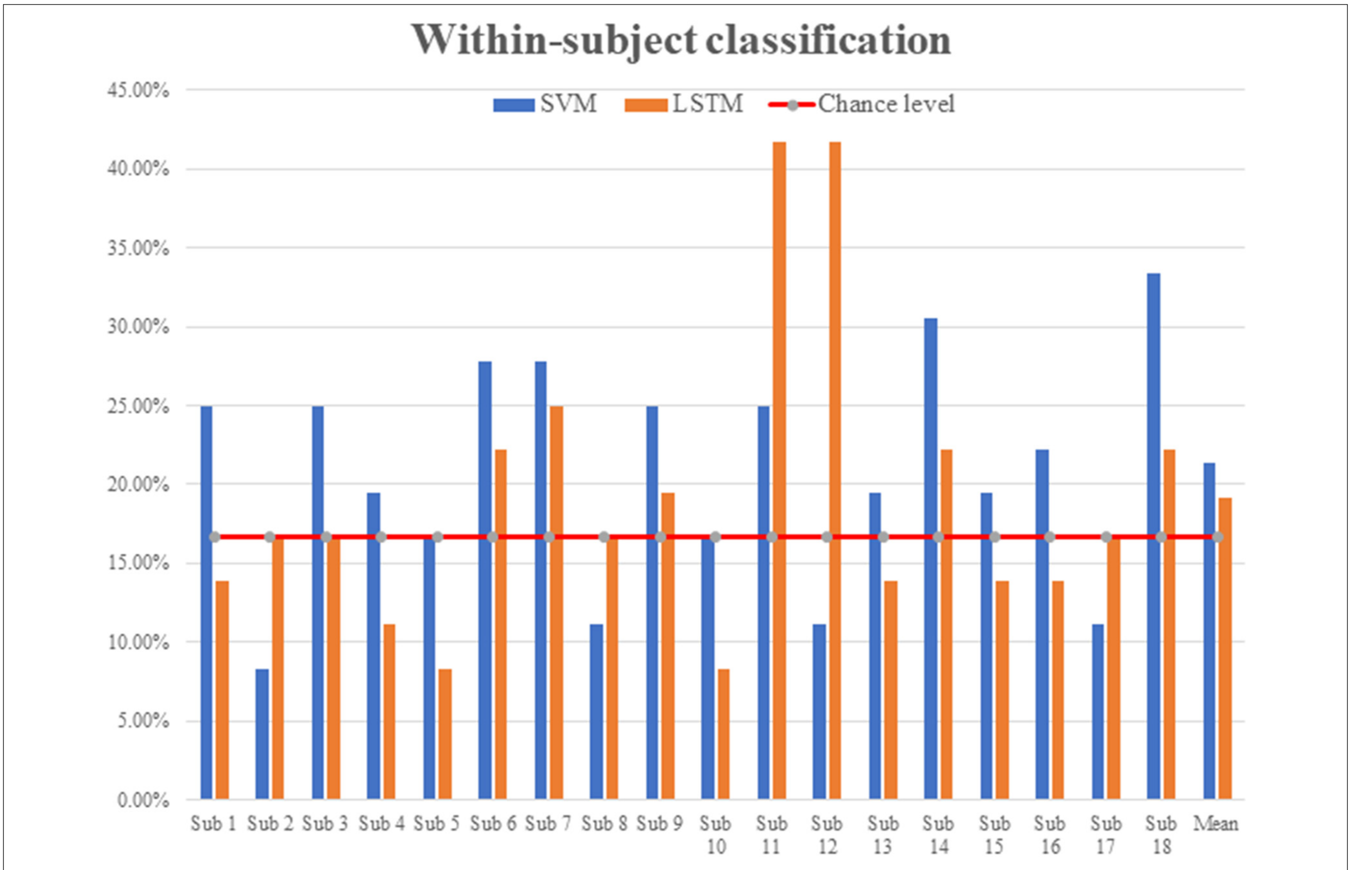
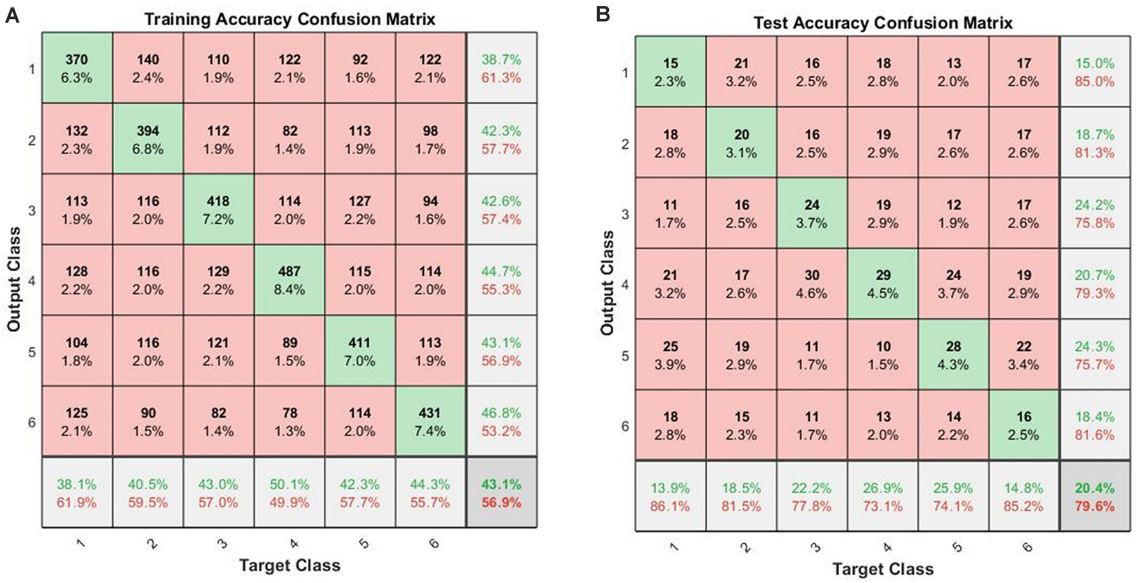


FIGURE 3 | Within-subject classification accuracies (SVM vs. LSTM).

2019). The developed RNNs have been applied in various brain imaging techniques (Hu et al., 2019a; Plakias and Boutalis, 2019). Wang et al. (2019) achieved 98.50% of classification accuracy by convolutional RNNs for individual recognition based on resting-state fMRI data. Qiao et al. (2019) proposed the application of bi-LSTM to decoding visual stimuli based on fMRI images from the visual cortex. The classification accuracies were 60.83 ± 1.17% and 42.50 ± 0.74% for each subject in five categories. The number of training samples and validation samples were 1,750 and 120 for each subject. Compared with the existing research, a limitation of this research is the small size of the data set.

The conventional classification technique is the process of distinguishing data from a set of categories based on a training data set on the category classes of which are known (Klein and Kranczioch, 2019; Pinti et al., 2019). The individual observations are analyzed into a set of features selected and executed by a classifier (Khan et al., 2014). In a more detailed process, a classifier is a function that takes the values of various features. For example, the mean, slope, skewness, and kurtosis values of HbO and HbR signals from individual trials can be used as the feature set (Tai and Chau, 2009). In our previous research (Hong and Santosa, 2016), decoding four-class sounds categories using fNIRS showed the 46.17 ± 6.25% (left) and 40.28 ± 6.00% (right) accuracies using LDA, while showing 38.35 ± 5.39% (left) and 36.99 ± 4.23% (right) using SVM. In the previous study, the classification was performed with the following steps: filtering, selecting region of interest, feature extraction from the region of interest, and classification. In the current study, to compare with the proposed method, the conventional classification technique was applied with the following steps: filtering, feature extraction from all channels, and classification. For LSTM networks, only filtering has been applied before classification.

The LSTM network may indirectly extract unstructured features from the data, and the network's weighting factors are optimized during the training session. The network can be trained only after simple filtering. The results showed that SVM is better for within-subject classification and LSTM is better for across-subject classification. It seems that fNIRS data involve different physiological data per subject, but this physiological difference is not removed with simple filtering. Also, the sample sizes of within-subject classification were 30 for training and six for validation. The sample sizes of across-subject classification were 584 for training and 64 for validation. Additionally, there are no significant differences between 10-fold and leave-one-out validations. The dataset of 10-fold validation might use the same subject's data in either training or testing. The size of



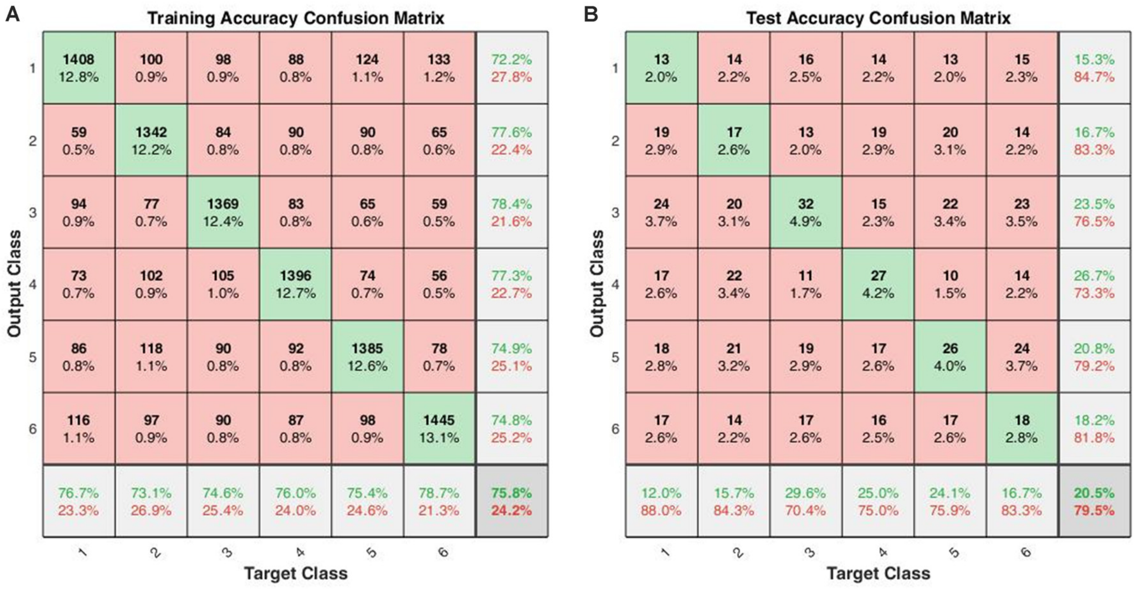
A Training Accuracy Confusion Matrix

	1	2	3	4	5	6	
1	370 6.3%	140 2.4%	110 1.9%	122 2.1%	92 1.6%	122 2.1%	38.7% 61.3%
2	132 2.3%	394 6.8%	112 1.9%	82 1.4%	113 1.9%	98 1.7%	42.3% 57.7%
3	113 1.9%	116 2.0%	418 7.2%	114 2.0%	127 2.2%	94 1.6%	42.6% 57.4%
4	128 2.2%	116 2.0%	129 2.2%	487 8.4%	115 2.0%	114 2.0%	44.7% 55.3%
5	104 1.8%	116 2.0%	121 2.1%	89 1.5%	411 7.0%	113 1.9%	43.1% 56.9%
6	125 2.1%	90 1.5%	82 1.4%	78 1.3%	114 2.0%	431 7.4%	46.8% 53.2%
	38.1% 61.9%	40.5% 59.5%	43.0% 57.0%	50.1% 49.9%	42.3% 57.7%	44.3% 55.7%	43.1% 56.9%

B Test Accuracy Confusion Matrix

	1	2	3	4	5	6	
1	15 2.3%	21 3.2%	16 2.5%	18 2.8%	13 2.0%	17 2.6%	15.0% 85.0%
2	18 2.8%	20 3.1%	16 2.5%	19 2.9%	17 2.6%	17 2.6%	18.7% 81.3%
3	11 1.7%	16 2.5%	24 3.7%	19 2.9%	12 1.9%	17 2.6%	24.2% 75.8%
4	21 3.2%	17 2.6%	30 4.6%	29 4.5%	24 3.7%	19 2.9%	20.7% 79.3%
5	25 3.9%	19 2.9%	11 1.7%	10 1.5%	28 4.3%	22 3.4%	24.3% 75.7%
6	18 2.8%	15 2.3%	11 1.7%	13 2.0%	14 2.2%	16 2.5%	18.4% 81.6%
	13.9% 86.1%	18.5% 81.5%	22.2% 77.8%	26.9% 73.1%	25.9% 74.1%	14.8% 85.2%	20.4% 79.6%

FIGURE 4 | Across-subject classification accuracies (LSTM): **(A)** confusion matrix for training, **(B)** confusion matrix for testing (Class 1: English; Class 2: non-English; Class 3: Nature sound; Class 4: Music; Class 5: Annoying sound; Class 6: Gunshot).



A Training Accuracy Confusion Matrix

	1	2	3	4	5	6	
1	1408 12.8%	100 0.9%	98 0.9%	88 0.8%	124 1.1%	133 1.2%	72.2% 27.8%
2	59 0.5%	1342 12.2%	84 0.8%	90 0.8%	90 0.8%	65 0.6%	77.6% 22.4%
3	94 0.9%	77 0.7%	1369 12.4%	83 0.8%	65 0.6%	59 0.5%	78.4% 21.6%
4	73 0.7%	102 0.9%	105 1.0%	1396 12.7%	74 0.7%	56 0.5%	77.3% 22.7%
5	86 0.8%	118 1.1%	90 0.8%	92 0.8%	1385 12.6%	78 0.7%	74.9% 25.1%
6	116 1.1%	97 0.9%	90 0.8%	87 0.8%	98 0.9%	1445 13.1%	74.8% 25.2%
	76.7% 23.3%	73.1% 26.9%	74.6% 25.4%	76.0% 24.0%	75.4% 24.6%	78.7% 21.3%	75.8% 24.2%

B Test Accuracy Confusion Matrix

	1	2	3	4	5	6	
1	13 2.0%	14 2.2%	16 2.5%	14 2.2%	13 2.0%	15 2.3%	15.3% 84.7%
2	19 2.9%	17 2.6%	13 2.0%	19 2.9%	20 3.1%	14 2.2%	16.7% 83.3%
3	24 3.7%	20 3.1%	32 4.9%	15 2.3%	22 3.4%	23 3.5%	23.5% 76.5%
4	17 2.6%	22 3.4%	11 1.7%	27 4.2%	10 1.5%	14 2.2%	26.7% 73.3%
5	18 2.8%	21 3.2%	19 2.9%	17 2.6%	26 4.0%	24 3.7%	20.8% 79.2%
6	17 2.6%	14 2.2%	17 2.6%	16 2.5%	17 2.6%	18 2.8%	18.2% 81.8%
	12.0% 88.0%	15.7% 84.3%	29.6% 70.4%	25.0% 75.0%	24.1% 75.9%	16.7% 83.3%	20.5% 79.5%

FIGURE 5 | Across-subject classification accuracies (leave-one-out validation of the LSTM with 16 hidden layers): **(A)** confusion matrix for training, **(B)** confusion matrix for testing (Class 1: English; Class 2: non-English; Class 3: Nature sound; Class 4: Music; Class 5: Annoying sound; Class 6: Gunshot).

the training data set was bigger in the leave-one-out validation. According to this result, if the sample size increases, the LSTM network would show better performance than the conventional method (SVM). Also, if we have enough training data, it would be enough for ignoring the subjects' physiological differences in the LSTM network. The LSTM network with 256 hidden layers showed slightly better performance than others, but the network overfitted with the training data set rapidly. Simplifying the data classification time can contribute to the commercialization of future diagnostics using fNIRS or BCI technology, given that it can reduce the classification time. Although the results in this study are not outstanding, it is worthwhile to show the potential of deep learning-based fNIRS signal classification technology.

CONCLUSION

This article aimed to identify hearing sounds using the HRs from the auditory cortices. The proposed audio-signal-based BCI is to be used for completely paralyzed people, for whom visual or motor cortex-based BCI may not be suitable. In this study, we used fNIRS signals evoked by audio-stimuli from multiple sound-categories. Compared with the conventional method, the LSTM-based approach could decode the brain activities without heavy pre-processing of the data, such as regression, feature selection, and feature extraction. Though the LSTM network's performance was a little higher than the chance level, it is noteworthy that we could classify the six sounds virtually without defining the region of interest and feature extraction. The approach using audio stimuli is promising for a passive-type BCI using ordinary sounds in our daily lives. This study has a limitation on the number of data, which needs to be improved in the future.

DATA AVAILABILITY STATEMENT

The datasets generated in this article are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Institutional Review Board of Pusan National University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

S-HY conducted the literature review and wrote the first draft of the manuscript. HS obtained the experimental data and initiated the work. C-SK participated in the revision of the manuscript. K-SH conceived the idea, corrected the manuscript, and finalized the work. All the authors have approved the final manuscript.

FUNDING

This work was supported by the National Research Foundation (NRF) of Korea under the auspices of the Ministry of Science and ICT, Korea (grant no. NRF-2020R1A2B5B03096000).

REFERENCES

Badai, J., Bu, Q., and Zhang, L. (2020). Review of artificial intelligence applications and algorithms for brain organoid research. *Interdiscip. Sci.* 12, 383–394. doi: 10.1007/s12539-020-00386-4

Bhutta, M. R., Hong, M. J., Kim, Y. H., and Hong, K. S. (2015). Single-trial lie detection using a combined fNIRS-polygraph system. *Front. Psychol.* 6:709. doi: 10.3389/fpsyg.2015.00709

Chen, J. N., Chen, J. Y., Zhang, R. M., and Hu, X. B. (2019). Toward a brain-inspired system: deep recurrent reinforcement learning for a simulated self-driving agent. *Front. Neurorobot.* 13:40. doi: 10.3389/fnbot.2019.00040

Corsi, M. C., Chavez, M., Schwartz, D., Hugueville, L., Khambhati, A. N., Bassett, D. S., et al. (2019). Integrating EEG and MEG signals to improve motor imagery classification in brain-computer interface. *Int. J. Neural Syst.* 29:1850014. doi: 10.1142/S0129065718500144

Dewey, R. S., and Hartley, D. E. H. (2015). Cortical cross-modal plasticity following deafness measured using functional near-infrared spectroscopy. *Hear. Res.* 325, 55–63. doi: 10.1016/j.heares.2015.03.007

Dubreuil-Vall, L., Ruffini, G., and Camprodon, J. A. (2020). Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. *Front. Neurosci.* 14:251. doi: 10.3389/fnins.2020.00251

Erturk, M. A., Panken, E., Conroy, M. J., Edmonson, J., Kramer, J., Chatterton, J., et al. (2020). Predicting *in vivo* MRI gradient-field induced voltage levels on implanted deep brain stimulation systems using neural networks. *Front. Hum. Neurosci.* 14:34. doi: 10.3389/fnhum.2020.00034

Flynn, M., Effraimidis, D., Angelopoulou, A., Kapetanios, E., Williams, D., Hemanth, J., et al. (2020). Assessing the effectiveness of automated emotion recognition in adults and children for clinical investigation. *Front. Hum. Neurosci.* 14:70. doi: 10.3389/fnhum.2020.00070

Fuchino, Y., Sato, H., Maki, A., Yamamoto, Y., Katura, T., Obata, A., et al. (2006). Effect of fMRI acoustic noise on sensorimotor activation examined using optical topography. *NeuroImage* 32, 771–777. doi: 10.1016/j.neuroimage.2006.04.197

Gao, P. P., Zhang, J. W., Fan, S. J., Sanes, D. H., and Wu, E. X. (2015). Auditory midbrain processing is differentially modulated by auditory and visual cortices: an auditory fMRI study. *NeuroImage* 123, 22–32. doi: 10.1016/j.neuroimage.2015.08.040

Gomez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A. H., Cattarossi, L., Nespor, M., et al. (2014). Language universals at birth. *Proc. Natl. Acad. Sci. U S A* 111, 5837–5841. doi: 10.1073/pnas.1318261111

Griffiths, T., and Warren, J. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892. doi: 10.1038/nrn1538

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., et al. (2015). Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front. Neurosci.* 9:217. doi: 10.3389/fnins.2015.00217

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2014). Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* 7:935. doi: 10.3389/fnhum.2013.00935

Hill, N. J., and Scholkopf, B. (2012). An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. *J. Neural Eng.* 9:026011. doi: 10.1088/1741-2560/9/2/026011

Hiraoka, M., Firbank, M., Essenpreis, M., Cope, M., Arridge, S. R., Vanderzee, P., et al. (1993). A monte-carlo investigation of optical pathlength in inhomogeneous tissue and its application to near-infrared spectroscopy. *Phys. Med. Biol.* 38, 1859–1876. doi: 10.1088/0031-9155/38/12/011

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780.

Hong, K. S., and Santosa, H. (2016). Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. *Hear. Res.* 333, 157–166. doi: 10.1016/j.heares.2016.01.009

Hu, R. H., Huang, Q. J., Wang, H., He, J., and Chang, S. (2019a). Monitor-based spiking recurrent network for the representation of complex dynamic patterns. *Int. J. Neural Syst.* 29:1950006. doi: 10.1142/S0129065719500060

Hu, X. S., Nascimento, T. D., Bender, M. C., Hall, T., Petty, S., O'Malley, S., et al. (2019b). Feasibility of a real-time clinical augmented reality and artificial intelligence framework for pain detection and localization from the brain. *J. Med. Internet Res.* 21:e13594. doi: 10.2196/13594

Hyvarinen, P., Yrttiaho, S., Lehtimaki, J., Ilmoniemi, R. J., Makitie, A., Ylikoski, J., et al. (2015). Transcutaneous vagus nerve stimulation modulates tinnitus-related beta- and gamma-band activity. *Ear Hear.* 36, E76–E85. doi: 10.1097/AUD.0000000000000123

Ieong, H. F. H., Gao, F., and Yuan, Z. (2019). Machine learning: assessing neurovascular signals in the prefrontal cortex with non-invasive bimodal electro-optical neuroimaging in opiate addiction. *Sci. Rep.* 9:18262. doi: 10.1038/s41598-019-54316-6

Kakusa, B., Saluja, S., Dadey, D. Y. A., Barbosa, D. A. N., Gattas, S., Miller, K. J., et al. (2020). Electrophysiology and structural connectivity of the posterior hypothalamic region: much to learn from a rare indication of deep brain stimulation. *Front. Hum. Neurosci.* 14:164. doi: 10.3389/fnhum.2020.00164

Kang, H., Yang, S., Huang, J., and Oh, J. (2020). Time series prediction of wastewater flow rate by bidirectional LSTM deep learning. *Int. J. Control Autom. Syst.* 18, 3023–3030. doi: 10.1007/s12555-019-0984-6

Khan, M. J., and Hong, K. S. (2015). Passive BCI based on drowsiness detection: an fNIRS study. *Biomed. Opt. Express* 6, 4063–4078. doi: 10.1364/BOE.6.004063

Khan, M. J., Hong, M. J. Y., and Hong, K. S. (2014). Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface. *Front. Hum. Neurosci.* 8:244. doi: 10.3389/fnhum.2014.00244

Klein, F., and Kranczioch, C. (2019). Signal processing in fNIRS: a case for the removal of systemic activity for single trial data. *Front. Hum. Neurosci.* 13:331. doi: 10.3389/fnhum.2019.00331

Kohler, P. J., Fogelson, S. V., Reavis, E. A., Meng, M., Guntupalli, J. S., Hanke, M., et al. (2013). Pattern classification precedes region-average hemodynamic response in early visual cortex. *NeuroImage* 78, 249–260. doi: 10.1016/j.neuroimage.2013.04.019

Kovelman, I., Mascho, K., Millott, L., Mastic, A., Moiseff, B., and Shalinsky, M. H. (2012). At the rhythm of language: brain bases of language-related frequency perception in children. *NeuroImage* 60, 673–682. doi: 10.1016/j.neuroimage.2011.12.066

Kovelman, I., Wagley, N., Hay, J. S. F., Ugolini, M., Bowyer, S. M., Lajiness-O'Neill, R., et al. (2015). Multimodal imaging of temporal processing in typical and atypical language development. *Ann. N. Y. Acad. Sci.*. 1337, 7–15. doi: 10.1111/nyas.12688

Lee, S. J., Choi, H., and Hwang, S. S. (2020). Real-time depth estimation using recurrent CNN with sparse depth cues for SLAM system. *Int. J. Control Autom. Syst.* 18, 206–216. doi: 10.1007/s12555-019-0350-8

Li, R. L., Wu, Q., Liu, J., Li, C., and Zhao, Q. B. (2020). Monitoring depth of anesthesia based on hybrid features and recurrent neural network. *Front. Neurosci.* 14:26. doi: 10.3389/fnins.2020.00026

Lin, C. T., King, J. T., Chuang, C. H., Ding, W. P., Chuang, W. Y., Liao, L. D., et al. (2020). Exploring the brain responses to driving fatigue through simultaneous EEG and fNIRS measurements. *Int. J. Neural Syst.* 30:1950018. doi: 10.1142/S0129065719500187

Lin, C. H., and Ting, J. C. (2019). Novel nonlinear backstepping control of synchronous reluctance motor drive system for position tracking of periodic reference inputs with torque ripple consideration. *Int. J. Control Autom. Syst.* 17, 1–17. doi: 10.1007/s12555-017-0703-0

Liu, F., Maggu, A. R., Lau, J. C. Y., and Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Front. Hum. Neurosci.* 8:1029. doi: 10.3389/fnhum.2014.01029

Liu, S. T., Montes-Lourido, P., Wang, X., and Sadagopan, S. (2019). Optimal features for auditory categorization. *Nat. Commun.* 10:1302. doi: 10.1038/s41467-019-09115-y

Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* 4, R1–R13. doi: 10.1088/1741-2560/4/2/R01

Lun, X. M., Yu, Z. L., Chen, T., Wang, F., and Hou, Y. M. (2020). A simplified CNN classification method for MI-EEG *via* the electrode pairs signals. *Front. Hum. Neurosci.* 14:338. doi: 10.3389/fnhum.2020.00338

Mushtaq, F., Wiggins, I. M., Kitterick, P. T., Anderson, C. A., and Hartley, D. E. H. (2020). The benefit of cross-modal reorganization on speech perception in pediatric cochlear implant recipients revealed using functional near-infrared spectroscopy. *Front. Hum. Neurosci.* 14:308. doi: 10.3389/fnhum.2020.00308

Park, J., and Jung, D. J. (2020). Deep convolutional neural network architectures for tonal frequency identification in a lofargram. *Int. J. Control Autom. Syst.* 19, 1103–1112. doi: 10.1007/s12555-019-1014-4

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol.* 10:e1001251. doi: 10.1371/journal.pbio.1001251

Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45, S199–S209. doi: 10.1016/j.neuroimage.2008.11.007

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., and Tachtsidis, I. (2019). Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of diverse signal filtering methods within a general linear model framework. *Front. Hum. Neurosci.* 12:505. doi: 10.3389/fnhum.2018.00505

Plakias, S., and Boutalis, Y. S. (2019). Lyapunov theory-based fusion neural networks for the identification of dynamic nonlinear systems. *Int. J. Neural Syst.* 29:1950015. doi: 10.1142/S0129065719500151

Plichta, M. M., Gerdes, A. B. M., Alpers, G. W., Harnisch, W., Brill, S., Wieser, M. J., et al. (2011). Auditory cortex activation is modulated by emotion: a functional near-infrared spectroscopy (fNIRS) study. *NeuroImage* 55, 1200–1207. doi: 10.1016/j.neuroimage.2011.01.011

Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M. S., and Oghalai, J. S. (2014). Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hear. Res.* 309, 84–93. doi: 10.1016/j.heares.2013.11.007

Qiao, K., Chen, J., Wang, L. Y., Zhang, C., Zeng, L., Tong, L., et al. (2019). Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices. *Front. Neurosci.* 13:692. doi: 10.3389/fnins.2019.00692

Salvari, V., Paraskevopoulos, E., Chalas, N., Muller, K., Wollbrink, A., Dobel, C., et al. (2019). Auditory categorization of man-made sounds versus natural sounds by means of MEG functional brain connectivity. *Front. Hum. Neurosci.* 13:1052. doi: 10.3389/fnins.2019.01052

Santosa, H., Hong, M. J., and Hong, K. S. (2014). Lateralization of music processing auditory cortex: an fNIRS study. *Front. Behav. Neurosci.* 8:418. doi: 10.3389/fnbeh.2014.00418

Scarff, C. J., Dort, J. C., Eggermont, J. J., and Goodyear, B. G. (2004). The effect of MR scanner noise on auditory cortex activity using fMRI. *Hum. Brain Mapp.* 22, 341–349. doi: 10.1002/hbm.20043

Sevy, A. B. G., Bortfeld, H., Huppert, T. J., Beauchamp, M. S., Tonini, R. E., and Oghalai, J. S. (2010). Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. *Hear. Res.* 270, 39–47. doi: 10.1016/j.heares.2010.09.010

Shan, C. H., Guo, X. R., and Ou, J. (2019). Deep leaky single-peaked triangle neural networks. *Int. J. Control Autom. Syst.* 17, 2693–2701. doi: 10.1007/s12555-018-0796-0

Sirpal, P., Kassab, A., Pouliot, P., Nguyen, D. K., and Lesage, F. (2019). fNIRS improves seizure detection in multimodal EEG-fNIRS recordings. *J. Biomed. Opt.* 24, 1–9. doi: 10.1117/1.JBO.24.5.051408

Smith, K. (2013). Reading minds. *Nature* 502, 428–430. doi: 10.1038/502428a

Smith, S. C., Dharmadi, R., Imrie, C., Si, B. L., and Herrmann, J. M. (2020). The DIAMOND model: deep recurrent neural networks for self-organizing robot control. *Front. Neurorobot.* 14:62. doi: 10.3389/fnbot.2020.00062

Staeren, N., Renvall, H., De Martino, F., Goebel, R., and Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Curr. Biol.* 19, 498–502. doi: 10.1016/j.cub.2009.01.066

Sualeh, M., and Kim, G. W. (2019). Simultaneous localization and mapping in the epoch of semantics: a survey. *Int. J. Control Autom. Syst.* 17, 729–742. doi: 10.1007/s12555-018-0130-x

Sung, H. J., Park, M. K., and Choi, J. W. (2020). Automatic grader for flatfishes using machine vision. *Int. J. Control Autom. Syst.* 18, 3073–3082. doi: 10.1007/s12555-020-0007-7

Tai, K., and Chau, T. (2009). Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface. *J. Neuroeng. Rehabil.* 6:39. doi: 10.1186/1743-0003-6-39

Tekulve, J., Fois, A., Sandamirskaya, Y., and Schoner, G. (2019). Autonomous sequence generation for a neural dynamic robot: scene perception, serial order and object-oriented movement. *Front. Neurorobot.* 13:95. doi: 10.3389/fnbot.2019.00095

Theunissen, F., and Elie, J. (2014). Neural processing of natural sounds. *Nat. Rev. Neurosci.* 15, 355–366. doi: 10.1038/nrn3731

Tsiouris, K. M., Pezoulas, V. C., Zervakis, M., Konitsiotis, S., Koutsouris, D. D., and Fotiadis, D. I. (2018). A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput. Biol. Med.* 99, 24–37. doi: 10.1016/j.compbiomed.2018.05.019

Wang, L. B., Li, K. M., Chen, X., and Hu, X. P. P. (2019). Application of convolutional recurrent neural network for individual recognition based on resting state fMRI data. *Front. Neurosci.* 13:434. doi: 10.3389/fnins.2019.00434

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872

Wong, P. C. M., Uppunda, A. K., Parrish, T. B., and Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *J. Speech Lang. Hear. Res.* 51, 1026–1041. doi: 10.1044/1092-4388(2008/075)

Xu, L. Y., Geng, X. L., He, X. Y., Li, J., and Yu, J. (2019a). Prediction in autism by deep learning short-time spontaneous hemodynamic fluctuations. *Front. Neurosci.* 13:1120. doi: 10.3389/fnins.2019.01120

Xu, Z. H., Zhou, X. F., and Li, S. (2019b). Deep recurrent neural networks based obstacle avoidance control for redundant manipulators. *Front. Neurorobot.* 13:47. doi: 10.3389/fnbot.2019.00047

Yan, Z. L., Liu, Y. M., Huang, X., Zhou, J. P., and Shen, H. (2019). Mixed script capital H-infinity and script capital L-2—script capital L-infinity anti-synchronization control for chaotic delayed recurrent neural networks. *Int. J. Control Autom. Syst.* 17, 3158–3169. doi: 10.1007/s12555-019-0263-6

Yang, C., Liu, Y. C., Li, F. M., and Li, Y. F. (2021). Finite-time synchronization of a class of coupled memristor-based recurrent neural networks: static state control and dynamic control approach. *Int. J. Control Autom. Syst.* 19, 426–438. doi: 10.1007/s12555-019-0616-1

Yang, D., Hong, K.-S., Yoo, S.-H., and Kim, C.-S. (2019). Evaluation of neural degeneration biomarkers in the prefrontal cortex for early identification of patients with mild cognitive impairment: an fNIRS study. *Front. Hum. Neurosci.* 13:317. doi: 10.3389/fnhum.2019.00317

Yang, D., Huang, R., Yoo, S.-H., Shin, M.-J., Yoon, J. A., Shin, Y.-I., et al. (2020). Detection of mild cognitive impairment using convolutional neural network: temporal-feature maps of functional near-infrared spectroscopy. *Front. Aging Neurosci.* 12:141. doi: 10.3389/fnagi.2020.00141

Yoo, S.-H., Woo, S.-W., Shin, M.-J., Yoon, J. A., Shin, Y.-I., and Hong, K.-S. (2020). Diagnosis of mild cognitive impairment using cognitive tasks: a functional near-infrared spectroscopy study. *Curr. Alzeimer Res.* 17, 1145–1160. doi: 10.2174/1567205018666210212154941

Yuan, Z., Zhang, X., and Ding, M. Z. (2019). Editorial: techniques advances and clinical applications in fused EEG-fNIRS. *Front. Hum. Neurosci.* 13:408. doi: 10.3389/fnhum.2019.00408

Zhang, C. C., Pugh, K. R., Mencl, W. E., Molfese, P. J., Frost, S. J., Magnuson, J. S., et al. (2016). Functionally integrated neural processing of linguistic and talker information: an event-related fMRI and ERP study. *NeuroImage* 124, 536–549. doi: 10.1016/j.neuroimage.2015.08.064

Zhang, F. Q., Wang, J. P., Kim, J., Parrish, T., and Wong, P. C. M. (2015). Decoding multiple sound categories in the human temporal cortex using high resolution fMRI. *PLoS One* 10:e0117303. doi: 10.1371/journal.pone.0117303

Zhao, W. F., Li, X. X., Chen, X., Su, X., and Tang, G. R. (2020). Bi-criteria acceleration level obstacle avoidance of redundant manipulator. *Front. Neurorobot.* 14:54. doi: 10.3389/fnbot.2020.00054

Zheng, W., Wang, H. B., Zhang, Z. M., Li, N., and Yin, P. H. (2019). Multi-layer feed-forward neural network deep learning control with hybrid position and virtual-force algorithm for mobile robot obstacle avoidance. *Int. J. Control Autom. Syst.* 17, 1007–1018. doi: 10.1007/s12555-018-0140-8

Zhou, X. F., Xu, Z. H., and Lie, S. (2019). Collision-free compliance control for redundant manipulators: an optimization case. *Front. Neurorobot.* 13:50. doi: 10.3389/fnbot.2019.00050

Conflict of Interest: The authors declare that they have no conflict of interest. This research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.