Hearing Research 333 (2016) 157–166

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy

Keum-Shik Hong [a, b, *], Hendrik Santosa [a]

[a] Department of Cogno-Mechatronics Engineering, Pusan National University, 2 Busandaehak-ro, Geumjeong-gu, Busan 46241, Republic of Korea
[b] School of Mechanical Engineering, Pusan National University, 2 Busandaehak-ro, Geumjeong-gu, Busan 46241, Republic of Korea

ARTICLE INFO

Article history:
Received 14 June 2015
Received in revised form
15 January 2016
Accepted 18 January 2016
Available online 29 January 2016

Keywords:
Functional near-infrared spectroscopy (fNIRS)
Auditory cortex
Classification
Multiple sound categories

ABSTRACT

The ability of the auditory cortex in the brain to distinguish different sounds is important in daily life. This study investigated whether activations in the auditory cortex caused by different sounds can be distinguished using functional near-infrared spectroscopy (fNIRS). The hemodynamic responses (HRs) in both hemispheres using fNIRS were measured in 18 subjects while exposing them to four sound categories (English-speech, non-English-speech, annoying sounds, and nature sounds). As features for classifying the different signals, the mean, slope, and skewness of the oxy-hemoglobin (HbO) signal were used. With regard to the language-related stimuli, the HRs evoked by understandable speech (English) were observed in a broader brain region than were those evoked by non-English speech. Also, the magnitudes of the HbO signals evoked by English-speech were higher than those of non-English speech. The ratio of the peak values of non-English and English speech was 72.5%. Also, the brain region evoked by annoying sounds was wider than that by nature sounds. However, the signal strength for nature sounds was stronger than that for annoying sounds. Finally, for brain-computer interface (BCI) purposes, the linear discriminant analysis (LDA) and support vector machine (SVM) classifiers were applied to the four sound categories. The overall classification performance for the left hemisphere was higher than that for the right hemisphere. Therefore, for decoding of auditory commands, the left hemisphere is recommended. Also, in two-class classification, the annoying vs. nature sounds comparison provides a higher classification accuracy than the English vs. non-English speech comparison. Finally, LDA performs better than SVM.

1. Introduction

The aim of this study is to identify the cortical brain regions associated with particular sounds in everyday life. Four different sound-categories (English-speech, non-English-speech, annoying sounds, and nature sounds) are investigated. As a means of neuronal activity detection, the hemodynamic responses (HRs) (Cope et al., 1988) upon various sounds measured by functional near-infrared spectroscopy (fNIRS) are utilized. As features of HR signals (in distinguishing different sounds), the mean, slope and skewness of the oxy-hemoglobin (HbO) signal are used in the classification process. Two classification techniques, namely linear discriminant analysis (LDA) and support vector machine (SVM), are applied.

Scientists have investigated auditory responses using various modalities such as electroencephalography (EEG) (Herrmann et al., 2013; Kong et al., 2014; Liu et al., 2015), functional magnetic resonance imaging (fMRI) (Scarff et al., 2004b; Wong et al., 2008; Olulade et al., 2011; Alho et al., 2014; Talavage et al., 2014; Weichenberger et al., 2015; Butler et al., 2015; Hall and Lomber, 2015), and fNIRS (Sevy et al., 2010; Pollonini et al., 2014; Dewey and Hartley, 2015; Murata et al., 2015). These studies investigated the complexities in the human auditory processing that is involved for various sound categories. An fMRI study revealed that there is a selectivity category for specific sounds within the auditory cortex (Sharda and Singh, 2012). Staeren et al. (2009) studied two-class classification using SVM for four sound categories (i.e., cats, female singers, acoustic guitars, and tones). The average classification

Abbreviations: fNIRS, functional near-infrared spectroscopy; HbO, oxy-hemoglobin; HbR, deoxy-hemoglobin; HR, hemodynamic response; MHR, modeled HR; LDA, linear discriminant analysis; SVM, support vector machine

* Corresponding author. Department of Cogno-Mechatronics Engineering and School of Mechanical Engineering, Pusan National University, 2 Busandaehak-ro, Geumjeong-gu, Busan 46241, Republic of Korea.
E-mail address: kshong@pusan.ac.kr (K.-S. Hong).

http://dx.doi.org/10.1016/j.heares.2016.01.009
0378-5955/

accuracies in their two-class classification problems using fMRI for female singers vs. acoustic guitars, female singers vs. cat sounds, and acoustic guitars vs. cat sounds were 69, 69, and 70%, respectively. Zhang et al. (2015) applied four methods (three versions of SVM, one of LDA) to decode the brain activities evoked by the audio-stimuli of seven sound categories: In their fMRI study, an average classification accuracy of 40% was achieved using the multi-class support vector machine-recursive feature elimination method (see Fig. 6 in their paper).

The mechanical noise produced in fMRI experiments, however, is problematic for auditory cortex studies, as it can cause interference when measuring brain activities evoked by sounds (Scarff et al., 2004a; Fuchino et al., 2006). For this reason, fNIRS has several advantages over fMRI in identifying cortical areas associated with specific sounds. Beside the key advantage (the silence of the machine), others include portability, real-time applicability, and inexpensiveness. Additionally, fNIRS offers a good trade-off between spatial and temporal resolution compared with both EEG and fMRI (Hu et al., 2013). Given these advantages, fNIRS has been applied to various applications such as neurology for stroke recovery (Liebert et al., 2005), psychiatry (Ernst et al., 2012), experimental psychology for language studies (Kovelman et al., 2012), and brain-computer interface (BCI) (Hu et al., 2012; Kamran and Hong, 2014; Naseer and Hong, 2015a, 2015b; Hong et al., 2015). Recently, Putze et al. (2014) studied a two-class classification problem (silent movie vs. audio book) in the visual and auditory cortices using LDA and hybrid EEG-fNIRS equipment. Their average classification accuracies were 75.8% (using HbO) and 70.9% (using HbR), respectively. Besides the auditory cortex, Herff et al. (2012) investigated classification of two-class problems in the Broca's and Wernicke's areas using fNIRS and LDA. In their work, three cases, that is, audible-speech vs. silence, audible-speech vs. imagery-speech, and silence vs. imagery-speech, were investigated resulting in 68, 65, and 54% classification accuracies, respectively.

fNIRS nomally uses two wavelengths to distinguish oxy-from deoxy hemoglobin (HbR) (or additional wavelengths to distinguish other chromophores such as water, lipid, etc). The measurement depth of NIR light penetration is almost half the emitter-detector distance (Stothers et al., 2008; Bhutta et al., 2015). The optimal emitter-detector distance depends on the NIR light intensity and the cortical region to be investigated (Ferrari and Quaresima, 2012; Naseer et al., 2014). Although the maximum light penetration depth of fNIRS is only about 4 cm, it has an acceptable spatial resolution for HR monitoring in the auditory cortex, considering the benefits of the absence of mechanical noises (Kovelman et al., 2009).

Classification is a process for distinguishing data classes (Duda et al., 2001) and involves selection of features and execution of a classifier (Liu and Yu, 2005). As features for distinguishing different stimuli, the mean, slope, and skewness (Tai and Chau, 2009; Naseer and Hong, 2013; Khan et al., 2014) values of HbO signals for individual trials over all channels are adopted. As classification algorithms, the LDA and SVM techniques have been widely utilized in various fNIRS applications involving the detection of drowsiness (Khan and Hong, 2015), mental workload (Herff et al., 2014) and speaking modes (i.e., audible-, silent-, and imagery-speech) (Herff et al., 2012) as well as in a hybrid BCI application using fNIRS and EEG (Putze et al., 2014). Whereas LDA separates the data into two or more classes (Fukunaga, 1990), SVM maximizes the margins of the selected hyperplanes (Burges, 1998). The usability of these algorithms has been well established in the literature; see, for example, relevant review papers (Lotte et al., 2007; Pereira et al., 2009) and multi-class problems (Garrett et al., 2003; Schlogl et al., 2005).

In this study, fNIRS is utilized to investigate the HRs evoked in hearing the audio-stimuli of four different sound-categories, namely English-speech, non-English-speech, annoying sounds and nature sounds, as presented in a pseudo-randomized order to 18 healthy subjects. Pre-processing techniques of noise removal and statistical analysis are used to enhance classification accuracy. Then, the HbO signals are decoded via the multi-class classifiers LDA and SVM using the mean, slope, and skewness values as features to distinguish the different sounds.

2. Materials and methods

2.1. Subjects

A total of 18 subjects (age: 28.11 ± 4.32 years; 6 females; 3 left-handed) participated in the experiment. All of them had normal hearing and no previous history of any neurological disorder. All were informed about the purpose of the experiment before providing their written informed consent. They were asked to avoid body motion and to remain relaxed with their eyes closed during the experiment. As selective attention is influential to the activation pattern of the auditory cortex (Jancke et al., 1999), the subjects were asked to listen to the audio-stimuli attentively and to guess, for each stimulus, the category. After the experiment, all of the subjects reported whether they were able to distinguish the individual audio-stimuli accurately or not. The experimental procedure was conducted in accordance with the ethical standards encoded in the latest Declaration of Helsinki and the guidelines approved by the Institutional Review Board of Pusan National University.

2.2. Audio-stimuli

The stimuli consist of 4 different sound categories including two languages (English, non-English) and two types of sound. The speech samples were selected based on a language proficiency test, and the sound categories (annoying sounds, nature sounds) selected from the youtube website (http://www.youtube.com). Each category consists of 6 different sounds. In one experiment, the participants were exposed to 24 audio-stimuli (10 s stimulus followed by 20 s resting; see Table 1) presented diotically in a pseudo-randomized order. Besides the 24 stimuli, one pre- and one post-trial (music, Canon in D by Pachelbel) were added though not included in the analysis. The entire fNIRS recording duration was 13 min. Regarding the speech hearing, none of the subjects recognized any language except English. These audio-stimuli were digitally mixed using Adobe Audition software (MP3-format file: 16-bit quantification, 44.1 kHz sampling, stereo channel) and normalized to the same intensity level. An active noise-cancellation earbud (Sony MDR-NC100D) was used, at the same sound-level setting, for all of the subjects.

To evaluate the overall sound quality (i.e., accuracy, enjoyability,

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

Trial	Speech hearing		Sound hearing	
	English	Non-English	Annoying sound	Nature sound
1	M	Russian (F)	Baby cry	River
2	F	German (F)	Car alarm	Forest (day time)
3	M	French (F)	Police siren	Rain
4	MF[a]	Bulgarian (MF*)	Horror sound	Jungle
5	F	Italian (MF)	Male scream	Ocean waves
6	F	Japanese (F)	Nuclear alarm siren	Waterfall
SQ[b]	8.6	7.0	7.8	8.1

[a] MF denotes male-female conversation.
[b] SQ stands for the subjective sound quality of each category evaluated by 7 participants (1 worst, 10 best).

fidelity, etc.) in advance, a pre-questionnaire on audio-stimuli was given to 7 persons who did not participate in the acquisition of fNIRS data. They were asked to listen to the audio-stimuli of four sound-categories and to score them on a scale of 1 (the worst) to 10 (the best). Additionally, for each audio-stimulus, they were instructed to guess the specific sound as well as its category. The last row in Table 1 shows that the average score of each category fell within the 7.0–8.6 range, indicating the suitability of the employed audio-stimuli for the testing purposes. Furthermore, all of the subjects reported an ability to distinguish each sound accurately, in every trial, for all sound-categories.

2.3. fNIRS data

All HR data from the auditory cortex were acquired with the continuous-wave fNIRS system (DYNOT: DYnamic Near-infrared Optical Tomography; NIRx Medical Technologies, Brooklyn, NY). The emitter-detector distance was 23 mm (see Fig. 1), and the sampling rate was set to 1.81 Hz. The data were measured simultaneously at two wavelengths (760 and 830 nm) for two 22-channel sets, that is, one set (i.e., channels 1–22) for the left hemisphere and another (i.e., channels 23–44) for the right hemisphere, including the auditory cortex, respectively. The optode configuration in Fig. 1 shows two sets of 3 × 5 arrays (8 emitters and 7 detectors), one in each hemisphere. The optodes in the left hemisphere were positioned to cover the Broca's area and Wernicke's area as well as the auditory cortex. Accordingly, Chs. 16 and 38 corresponded to the T3 and T4 locations, respectively, in the international 10–20 system. Finally, during the experiment, all of the lights in the room were switched off to minimize signal contamination.

The measured intensity data of the two wavelengths were converted to relative HbO and HbR concentration changes using the modified Beer–Lambert law (Cope et al., 1988; Kamran and Hong, 2013; Hong and Nguyen, 2014). Subsequently, the open-source software NIRS-SPM (Ye et al., 2009) was utilized in the authors' own Matlab™ (Math-works, Natick, MA) code. In this study, constant values of differential path-length factor (i.e., d = 7.15 for λ = 760 nm and d = 5.98 for λ = 830 nm) were used for all of the channels. Since HbO signals are more direct to the given stimuli than HbR signals (i.e., the signal-to-noise ratio of HbO is higher than that of HbR) (Wolf et al., 2002; Schecklmann et al., 2008; Holper et al., 2009), the mean, slope, and skewness only of the HbO concentration changes were used in the subsequent analysis.

2.4. Pre-processing

Pre-processing of fNIRS data is an important step, the purpose of which is to remove physiological noises and to minimize signal variations. It can be split into two stages: spectral filtering and scatter correction (Rinnan et al., 2009). Spectral filtering removes both additive and multiplicative effects in the spectra, after which scatter correction (e.g., normalization or rescaling) reduces the variability among samples due to scatter. In this study, filtering, detrending, and rescaling were applied in the pre-processing process. Since an auditory stimulus of 10 s was followed by a 20 s resting period, the stimulation frequency was approximately 0.033 Hz (i.e., 1/30). First, the physiological noises of respiratory (about 0.3 Hz) and cardiac signals (about 1 Hz) contained in the HRs were removed using a 0.15 Hz low-pass filter (Santosa et al., 2013). Second, the trend of the signal (e.g., a low-frequency drift) was removed from the time-series data by the detrending technique (Tanabe et al., 2002), according to which the *detrend* function in Matlab™ subtracts the best fit line (or mean) from the data. In the third step, due to the wide variation in the classification stage, the rescaling method (Naseer and Hong, 2015a, 2015b) by which the *Mat2gray* function in Matlab™ re-scales the data within the 0–1 range was applied.

2.5. Activation maps

Verification of cortical activation is the most important step in fNIRS data analysis. Hu et al. (2010) showed that activation can be estimated by fitting the measured HR to a regression model using the recursive least squares algorithm. The modeled (or ideal) HR is that which is expected for a given stimulus: In Fig. 2, it is represented by the dotted (red) line computed by convolving the

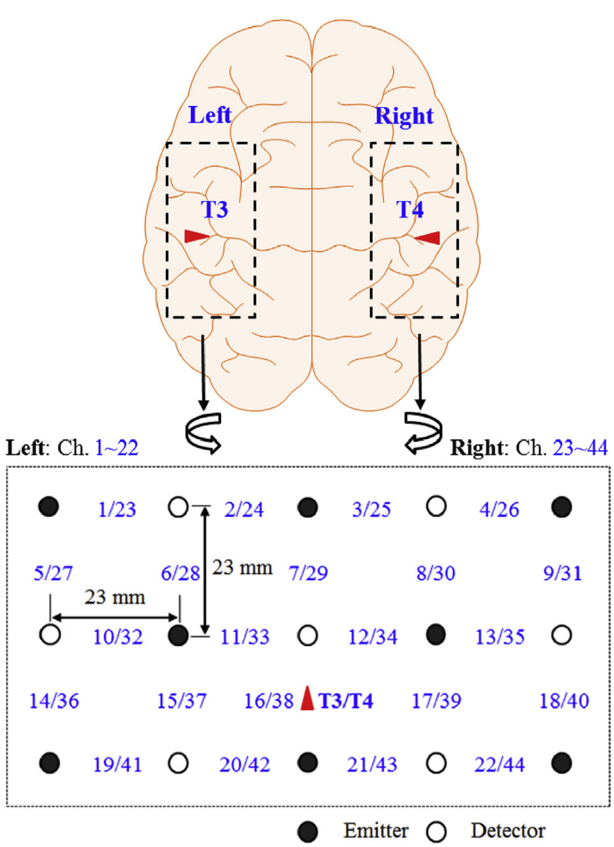

Fig. 1. Optodes configuration: Numbers represent the measurement channels. The channel number 16 (and 38) coincide with the T3 (and T4) location in the International 10–20 System (Santosa et al., 2014).

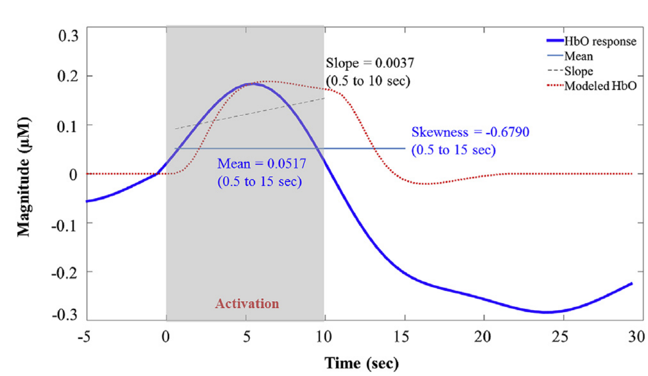

Fig. 2. Features defined for classification: Mean, slope, and skewness values of the HbO signal (solid blue curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

stimulus pattern (i.e., 10 s activation and 20 s rest) and the canonical HR function (HRF) available in the SPM 8 software (Wellcome Trust Center for Neuroimaging, London, UK) (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). The profile of the canonical HRF used in this study is as follows. Delay of response from onset: 6 s, delay of undershoot: 16 s, dispersion of response: 1 s, dispersion of undershoot: 1 s, ratio of peak to undershoot: 6, length of kernel: 21 s, see (Hong and Nguyen, 2014) for the impulse HRF.

The t-value is defined as the ratio of the weighting coefficient to the modeled HR (in the process of fitting the measured stimulus-evoked HR to the modeled one), and its standard error (Friston et al., 2008). A high t-value indicates that the signal is highly correlated with the modeled HR. In this study, the t-values were calculated for individual trials using the *robustfit* function available in Matlab™, which has been described in detail in (Santosa et al., 2014; see Eqs. (3)–(5)). Specifically, the HbO signal in an experiment (i.e., 1320 data points for 24 trials) was segmented to 55 data points per trial. As seen in Figs. 3A and 4A, the obtained t-values were displayed as a map in order to illustrate the activation in the covered brain region; the intermediate values were interpolated with the Matlab function *interp2* using the 22 t-values from 22 channels. On these t-maps, the numbers, the color in a pixel, and the color bar in the lower-right corner indicate the channel numbers, signal intensity, and color scale of the t-value of that pixel, respectively. Fig. 3 is an activation map of Sub. 11 (chosen because he had no hair)'s HbO data for four audio-stimuli in the left and right hemispheres, each t-value being the average of six trials for each sound category. Fig. 4 shows the activation map averaged over the 18 subjects, thus demonstrating the overall trends.

2.6. Regions of interest

The regions-of-interest (ROIs) of four audio-stimuli are investigated: The ROIs denote the brain areas where the t-values are higher than the critical t-value (t_{crt}). In this study, t_{crt} was set as 1.6736, as computed from the degree of freedom of the data ($N - 1 = 54$) and the statistical significance level ($\alpha = 0.05$ for one-tailed test). The ROIs were found to differ for the respective sound categories. The signals from each ROI were used for further analysis in feature selection and classification, because using the signals only from the associated ROI (rather than from all of the channels) had improved the classification accuracy. In the case of Subject 11, Chs. 1, 2, 5, 6, 8 (in the left hemisphere) and 24, 27, 28, 29, 33, 36, 37 (in the right hemisphere) for English-speech hearing, and Chs. 9, 10, 22 (in the left hemisphere) and 27, 41 (in the right hemisphere) for non-English-speech hearing, showed that $t > t_{crt}$. These channels are marked as crosses, plus-signs, triangles and circles, respectively, in Fig. 3B. To observe the trends over 18 subjects, the respective signals were averaged (see Fig. 4A), the ROIs from which are identified in Fig. 4B. If the number of channels showing $t > t_{crt}$ is less than two (i.e., the number of channels in an ROI is too small), more than two channels with high t-values can be included. Fortunately, such a case was not encountered in our experiment.

2.7. Feature selection and classification

The signals from the ROI of a given sound stimulus were averaged first. Then, the mean, slope, and skewness values of the averaged HbO signal were used as features for classification. The mean (for 0.5–15 s) differentiates the occurrence of an activation from the resting state; the slope (during 0.5–10 s) indicates the speediness of the occurrence of the response; the skewness is a measure of the asymmetry of a signal in terms of the probability distribution around its mean relative to a normal distribution, and therefore can differentiate the shapes of HbO signals. For example, from the solid (blue) curve in Fig. 2 (which is the averaged HbO of the 18 subjects' English-speech hearing in the left hemisphere), the mean, the slope, and the skewness values are 0.0517, 0.0037, and −0.6790, respectively. In this study, the *mean*, *polyfit*, and *skewness* functions available in Matlab™ were used.

For classification, we used the LDA and SVM classifiers to verify the individual HbO signals. The two classifiers were compared in order to ensure that the HR can be classified from different sound-categories. The *classify* and *multisvm* functions available in

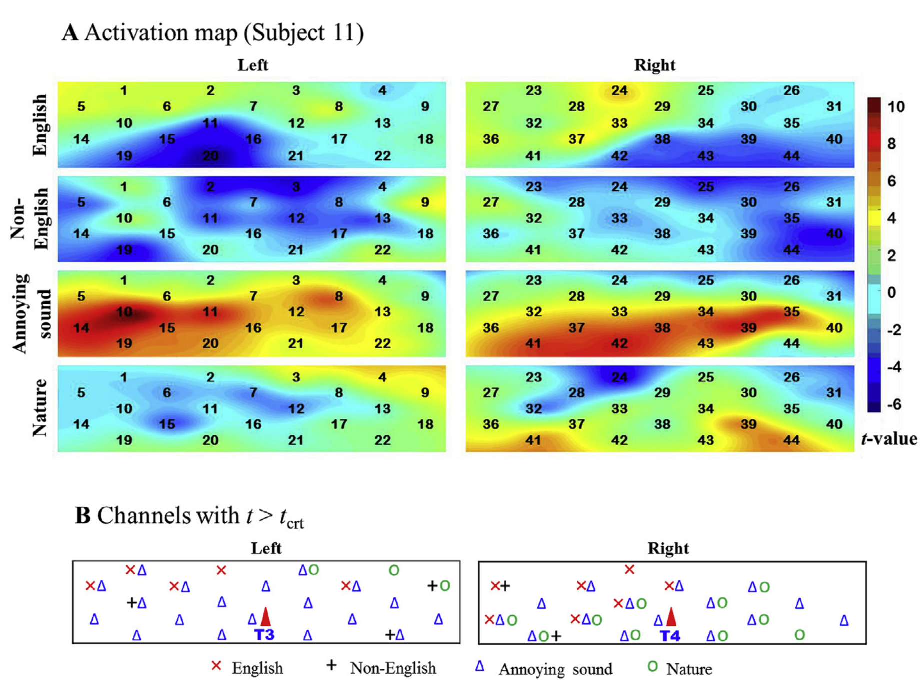

Fig. 3. HbO in the left and right auditory cortices evoked by four different sound-categories (Subject 11): Active channels appear differently upon different auditory stimuli.

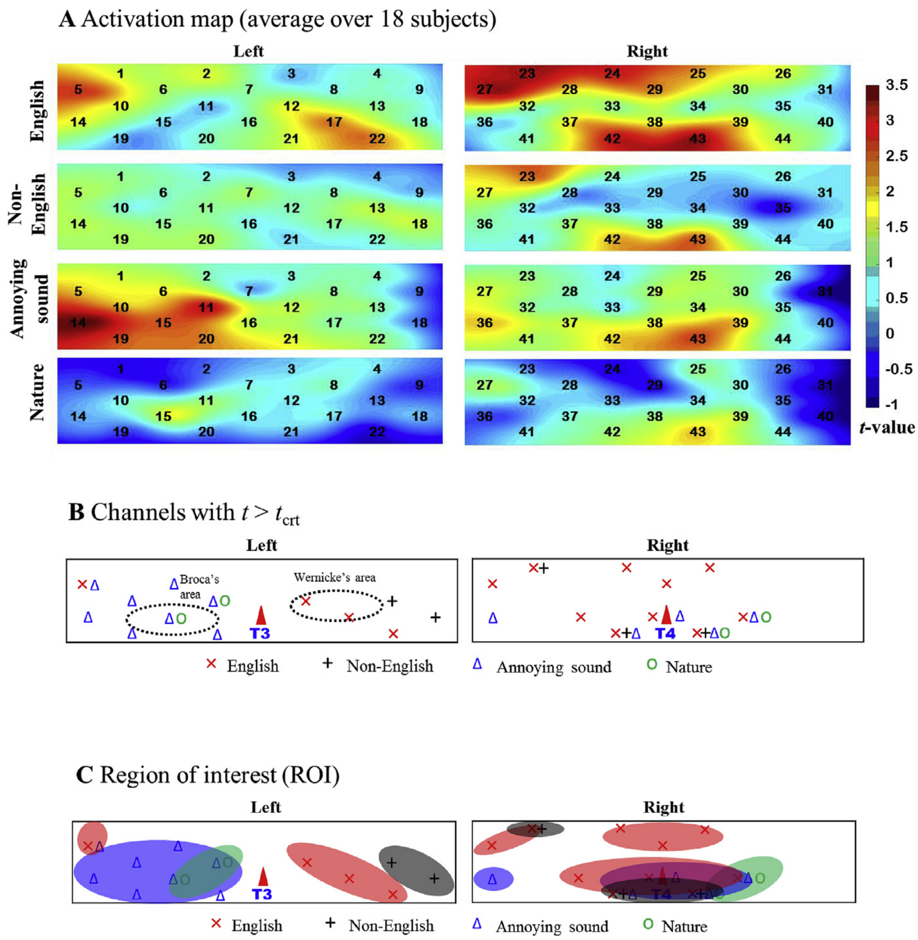

Fig. 4. The averaged HbO (over 18 subjects) in the left and right auditory cortices upon four different sound-categories: The ROIs for individual categories are specified.

Matlab™ were used as the LDA and SVM classifiers, respectively. To determine the classification accuracy, we used 4-runs of 4-fold cross-validation. This method entails the following steps: i) randomly break the data into 4 sets (i.e., 24/4 = 6); ii) train on 3 datasets and test on 1; iii) repeat 4 times and take the mean accuracy. We ran the cross-validation two times to obtain the average and standard deviation for every subject.

3. Results

Fig. 5 presents Subject 11's three-dimensional (3D) plot of the mean, slope, and skewness values of HbO (*x*-, *y*-, and *z*-axes, respectively) from the left and right hemispheres. It displays data for a total of 24 trials (six trials for each of four categories), which are marked by crosses (×, red) for English, plus-signs (+, black) for non-English, upward-pointing triangles (Δ, blue) for annoying sounds, and circles (○, green) for nature sounds. It can be seen that the sets of nature sounds and annoying sounds are clearly distinguishable due to the difference in their mean values.

Fig. 6 compares the average HbO signals and the standard deviations over the 18 subjects for English and non-English hearing in the left and right hemispheres, while Fig. 7 compares those of the annoying and nature sounds. The averaging was performed on 108 data points (i.e., 18 subjects × 6 trials) for each category. The shaded areas along the mean values represent their standard errors. The numbers inside Figs. 6 and 7 indicate the peak values of the individual HbO responses. For example, 0.1890 and 0.0888 in Fig. 7 are the peak values for nature and annoying sounds in the left hemisphere, respectively.

To determine whether the classification accuracies were consistent across the subjects, a cross-subject analysis was performed. Fig. 8 shows the individual subjects' classification accuracies for the four audio categories using LDA and SVM in the left and right hemispheres. More specifically, it compares the means of the classification accuracies for the individual subjects with their standard deviations, while Fig. 9 depicts those of the two-class classification problems. In Figs. 8 and 9, the first bar (wide downward diagonal), second bar (dashed horizontal), third bar (light horizontal), and fourth bar (wide upward diagonal) indicate the classification accuracies using LDA (green color) and SVM (red color) for the left and right hemispheres, respectively, where the chance level is indicated by the horizontal (blue) line. As for the results variance across the subjects, Subject 5 showed the lowest accuracies. The last bars (Avrg) in both figures are the averages over the 18 subjects: using LDA, the accuracies for four-class classification were 46.17 ± 6.25% (left) and 40.28 ± 6.00 (right), and using SVM they were 38.35 ± 5.39% (left) and 36.99 ± 4.23% (right). It was found that the LDA accuracies were higher than those of SVM, and that the accuracies in the left hemisphere were higher in general than those in the right hemisphere.

Next, to investigate language-related classification capability, two-class classification problems were performed. Fig. 9A and B plot the classification results for speech hearing (English vs. Non-English) and sound hearing (annoying sounds vs. natural sounds), respectively. In both cases, as can be seen, the classification performance was significantly above the chance (i.e., 50%) level. As

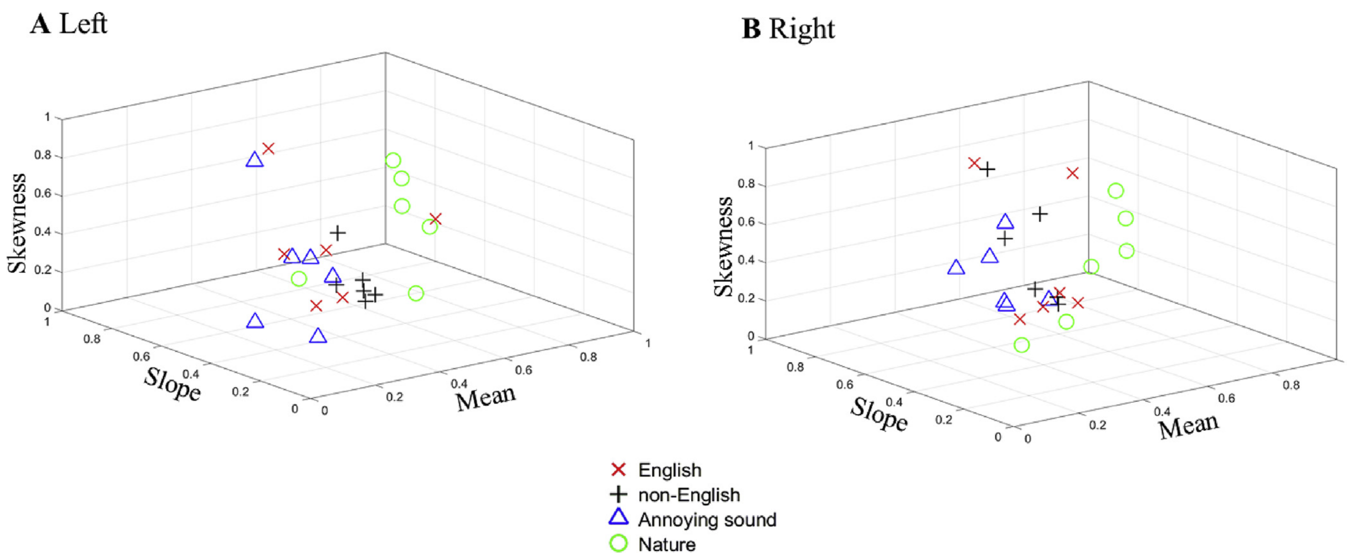

Fig. 5. Example of 3D scatter plot of the mean, slope, and skewness values of the HbOs upon 24 trials (Subject 11).

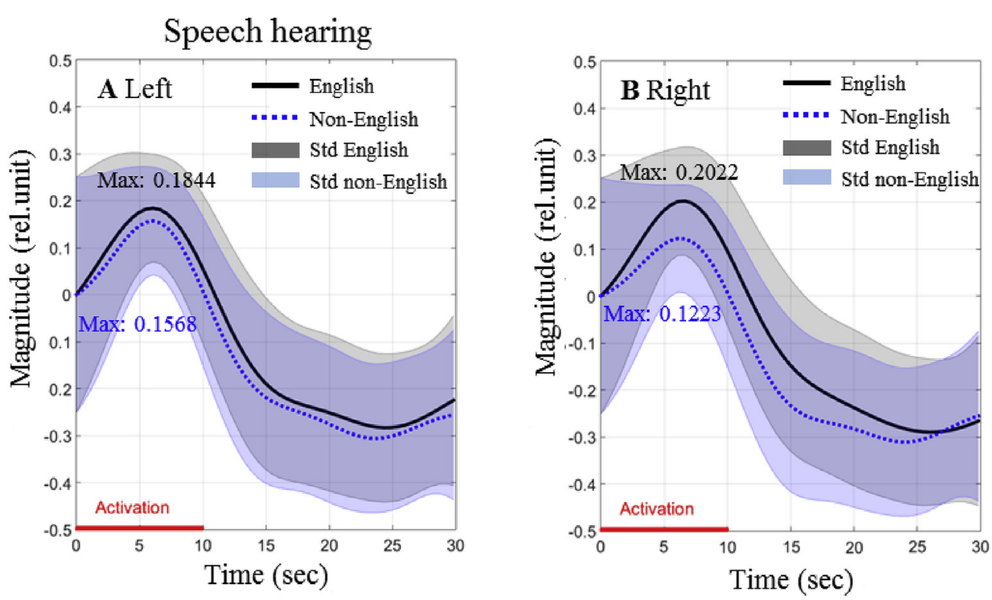

Fig. 6. The averaged HbOs (over 18 subjects) and their standard deviations for English and non-English speech.

shown in Fig. 9A (speech hearing), the average classification accuracies using LDA were 71.03 ± 8.72% (left) and 70.03 ± 8.97% (right) and those by SVM, 68.18 ± 8.30% (left) and 68.07 ± 7.59% (right). As shown in Fig. 9B (sound hearing), the average classification accuracies using LDA were 74.97 ± 11.74% (left) and 71.80 ± 9.89% (right), and those by SVM, 72.34 ± 9.72% (left) and 72.15 ± 9.77% (right), respectively. The overall-averaged classification accuracies were 70.53 ± 8.79% (LDA) and 68.11 ± 7.90% (SVM) for speech hearing and 73.39 ± 10.82% (LDA) and 72.24 ± 9.61% (SVM) for sound hearing.

4. Discussion

The authors used the fNIRS technique because of its crucial advantages for analysis of sound-evoked brain activation: non-invasiveness and silence. Along with this, fNIRS has a high potential as a neuroimaging tool, because it can demonstrate real-time imaging in everyday life. To compare brain activities, two types of information (spatial and temporal) were examined. The *t*-map (see Figs. 3 and 4) depicts the spatial distribution of activation in terms of the correlation level of the measured HR with the expected HR for a given stimulus. However, it cannot reveal the signal strength in the given location. Therefore, to determine the intensity of brain activity, the temporal magnitude of the HR (Figs. 6 and 7) should be examined together with the activation map. In this study, four audio categories (English-speech, non-English-speech, annoying sounds, and nature sounds) were investigated. The detection area in the left hemisphere (see Fig. 1) includes the Wernicke's and Broca's areas, both of which are related to language processing (see Fig. 4B). The obtained results will be discussed in two aspects: asymmetry and classification.

Asymmetry in the functional responses in the left and right hemispheres was observed (Toga and Thompson, 2003). The left hemisphere is known to be responsible for language, math, and

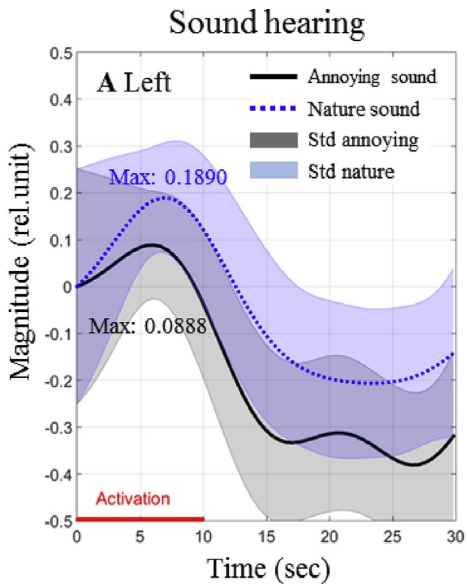

Fig. 7. The averaged HbOs (over 18 subjects) and their standard deviations for annoying and nature sounds.

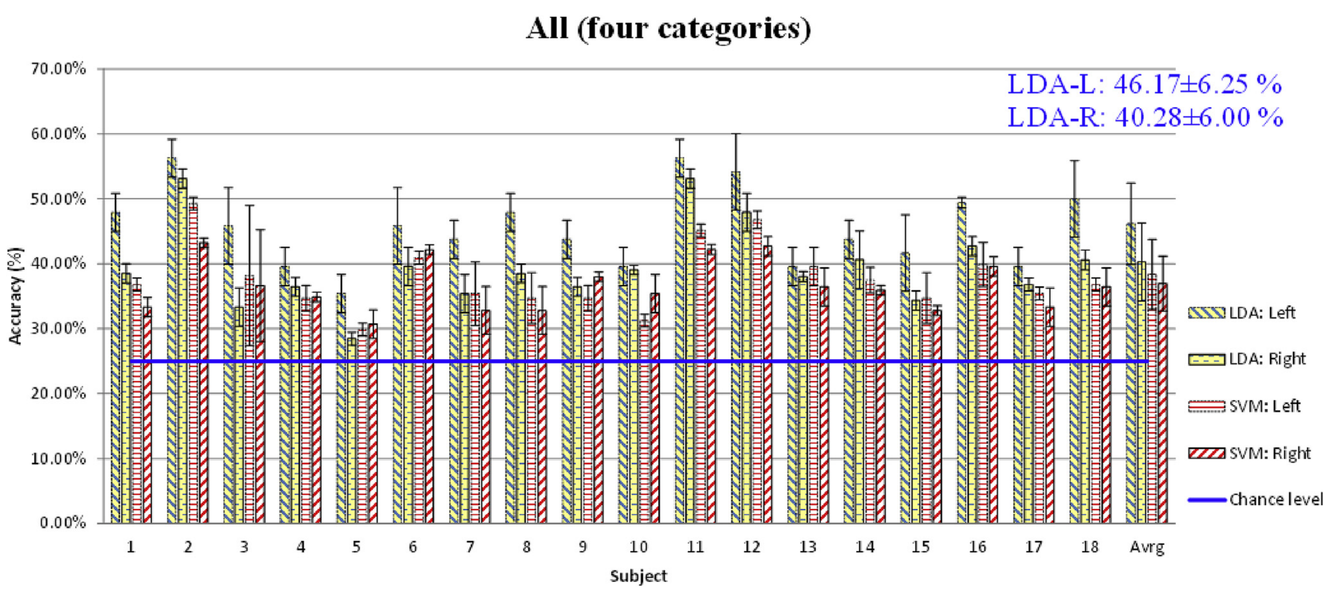

Fig. 8. Classification accuracies of 4 sound-categories (from the left and right hemispheres) and the overall average over 18 subjects: LDA vs. SVM (the bar on top represents the standard deviation of the obtained mean value).

logic, while the right hemisphere is responsible for spatial abilities, visual imagery, music, etc. Particularly, the auditory cortex in the left hemisphere is known to be dominant in the hearing of phonemic (~12–50 Hz) transitions in speech, while right lateralization occurs in hearing the syllabic (~3–7 Hz) transition in speech (Poelmans et al., 2012).

Comparing the first and second rows in Fig. 4A, it can be seen that the brain is more activated when hearing understandable language (English) than non-understandable language (non-English). The same phenomenon was found in Fig. 6, in which the peak values of the HR when hearing English-speech are higher than those when hearing non-English speech in both hemispheres. If the peak magnitudes are compared, the percentile ratio of the peak value of non-English hearing to that of English hearing is 85% (i.e., 01568/0.1844) in the left hemisphere and is 60% (i.e., 0.1223/0.2022) in the right hemisphere.

In the presence of left lateralization in language processing, such a phenomenon did not occur in our case (i.e., in English-speech hearing). As seen in the first row of Fig. 4A, the color in the right side is more red (i.e., highly correlated) when hearing English-speech. This means that in this case, right lateralization exists. This is seen also in Fig. 6: when hearing English-speech, the peak value on the right side (i.e., 0.2022) was higher than that on the left side (i.e., 0.1844). The authors believe that this was due to the fact that segmented speech-stimuli data were used in the experiment (i.e., 10 s speech and 20 s rest; 0.033 Hz). Also the dominant frequency in the speech stimuli was close to 4 Hz, which belongs to the frequency band of syllables in speech. This might be congruent with Abrams et al. (2008) in that right lateralization occurs for slow and syllabic-rate modulation in the auditory cortex.

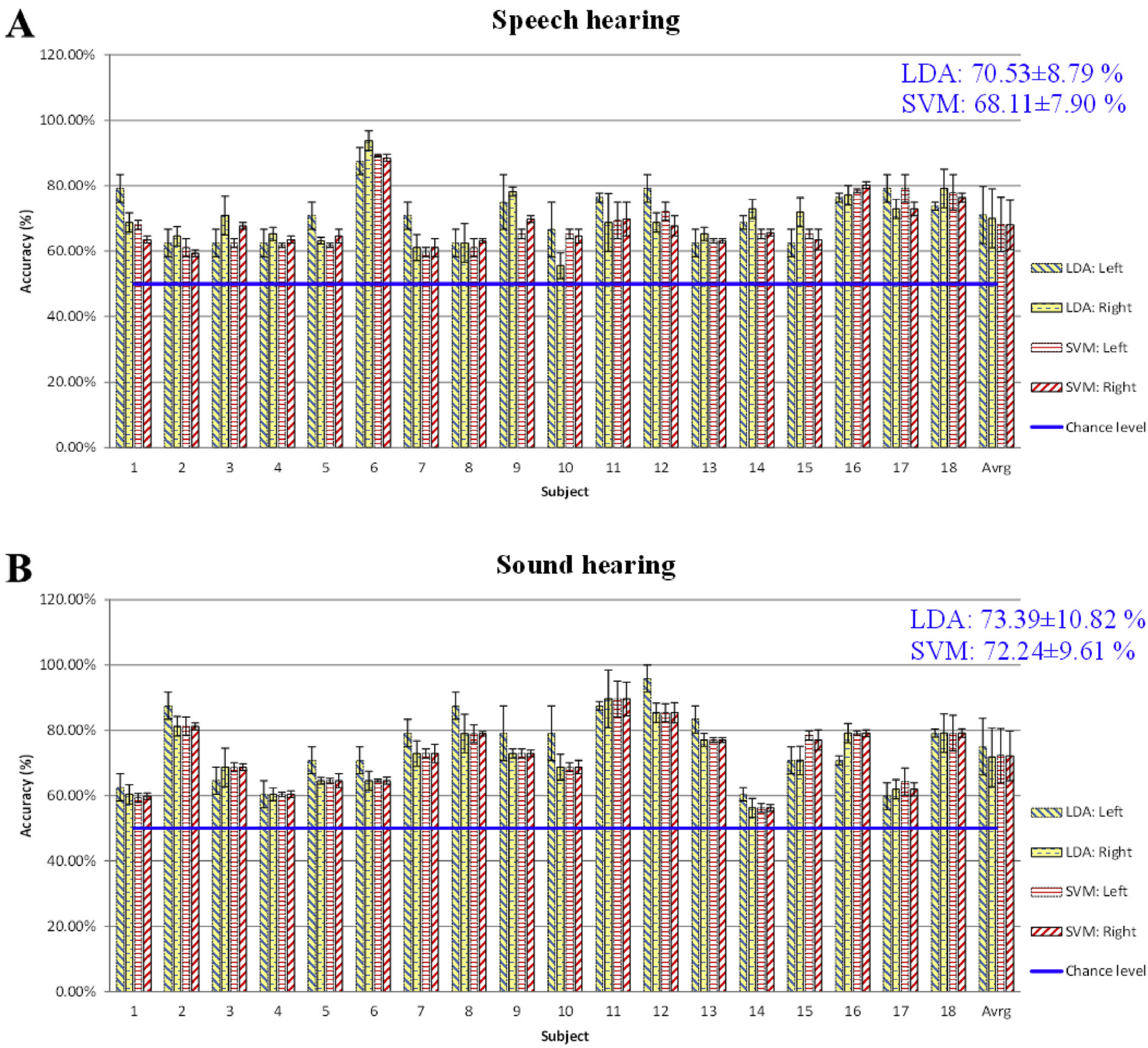

Fig. 9. Classification accuracies of two sound-categories: (a) Speech hearing (English vs. non-English), (b) sound hearing (annoying vs. natural).

Asymmetry in hearing non-language sounds was observed as well. First, as seen in Fig. 4A–C, annoying sounds prevailed in a broader brain region than did nature sounds. Comparing the third and fourth rows in Fig. 4, the color of the *t*-map for annoying sounds is more yellow throughout the region than that of nature sounds. Additionally, since the signal strength cannot be seen in this *t*-map, the temporal HR data of both sound categories were examined as well: In Fig. 7, the peak values of the HRs for annoying sounds are 0.0888 (left) and 0.1006 (right), whereas those of nature sounds are 0.1890 (left) and 0.1702 (right). This reflects the fact that humans respond more profoundly to nature sounds than to annoying sounds. This result is congruent with Plichta et al. (2011) in that the HR to pleasant sounds increases as compared with that to unpleasant sounds.

The average classification accuracies using the HbO signals from the left hemisphere were higher than those from the right hemisphere (see Table 2). In the two-class classification problems, the average performance in the left hemisphere was 73.0 ± 10.23%, while that in the right hemisphere was 70.9 ± 11.05%. In the four-class cases, the accuracy in the left hemisphere was 46.17 ± 6.25%, while that in the right hemisphere was 40.28 ± 6.0%. Therefore, if only one side of the brain is to be chosen for BCI purposes, the left hemisphere is recommended.

Regarding the classifiers' performances (see Fig. 9), the overall two-class classification accuracies of LDA were 70.53 ± 8.79% (speech hearing) and 73.39 ± 10.82% (sound hearing), whereas those of SVM were 68.11 ± 7.90% (speech hearing) and 72.24 ± 9.61% (sound hearing), respectively. Therefore, it was concluded that LDA performs better than SVM in classifying the HR signals evoked by audio stimuli.

Variations in signal strength and, therefore, classification accuracy are due to several factors. i) Skull and scalp thicknesses (Lynnerup et al., 2005) and hair darkness. In the present study for example, the subject with the best experimental conditions was Subject 11, who had no hair; the data for Subject 11, therefore, were used in illustrating the activation map and scatter plot in Figs. 3 and

Table 2
Comparison of classification accuracies.

		Left (%)	Right (%)	Average (%)
A) LDA				
Two-class classification	Speech	71.03 ± 8.72	70.03 ± 8.97	70.53 ± 8.79
	Sound	74.97 ± 11.74	71.80 ± 13.13	73.39 ± 10.82
Four-class classification		46.17 ± 6.25	40.28 ± 6.00	43.22 ± 6.72
B) SVM				
Two-class classification	Speech	68.16 ± 8.30	68.07 ± 7.59	68.11 ± 7.90
	Sound	72.34 ± 9.72	72.15 ± 9.77	72.24 ± 9.61
Four-class classification		38.35 ± 5.39	36.99 ± 4.23	37.67 ± 4.82

5, respectively. ii) A subject's concentration level. In the current experiment, some of the participants reported difficulty in focusing due to fatigue; minimizing such subject-wise variation, then, is another important factor to consider. iii) The size of a data set. A larger amount of data for cross-validation would yield a better result, but it would also consume more experimental time, thus leading to subject fatigue. It should be noted too, that to make the HR return to the baseline, at least 20 s is needed. Finally, a possible reduction of variation in classification accuracy can be accomplished by using the optimum activation period to increase the number of trials with good environmental conditions.

5. Conclusions

This study investigated the use of functional near-infrared spectroscopy to decode the hemodynamic responses evoked by audio-stimuli from four different sound categories (English-speech, non-English-speech, annoying sounds, and nature sounds). To account for and handle the large variations of data in the multi-class offline classification problem, three data processing steps including pre-processing, feature-selection, and classifier selection were examined. As features for classification, the mean, slope, and skewness values of HbO were used. Interestingly, the classification accuracies were higher in the left hemisphere than in the right hemisphere. Further, we demonstrated that the HR differs from those different sound-categories, which fact reflects the reported hemispheric lateralization in the auditory cortex areas. Finally, we concluded that the fNIRS signals of the HR, as evoked by audio-stimuli representing the four different sound-categories, are distinguishable. The overall results suggest that it is possible to decode the responses to different sound-categories in the auditory cortex areas.

Conflicts of interest

There are no conflicts of interest.

Acknowledgments

This work was supported by the National Research Foundation of Korea under the auspices of the Ministry of Science, ICT and Future Planning, Korea (grant no. NRF-2014-R1A2A1A10049727). The authors are grateful for the anonymous reviewers and Melissa J. Hong at FIRST 5 Santa Clara County, 4000 Moorpark Ave., San Jose, CA 95117, USA, for their constructive comments on the manuscript.

References

Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J. Neurosci. 28, 3958–3965.
Alho, K., Rinne, T., Herron, T.J., Woods, D.L., 2014. Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. Hear. Res. 307, 29–41.
Bhutta, M.R., Hong, M.J., Kim, Y.-H., Hong, K.-S., 2015. Single-trial lie detection using a combined fNIRS-polygraph system. Front. Psychol. 6, 709.
Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. 2, 121–167.
Butler, B.E., Hall, A.J., Lomber, S.G., 2015. High-field functional imaging of pitch processing in auditory cortex of the cat. PLoS One 10, e0134362.
Cope, M., Delpy, D.T., Reynolds, E.O., Wray, S., Wyatt, J., van der Zee, P., 1988. Methods of quantitating cerebral near infrared spectroscopy data. Adv. Exp. Med. Biol. 215, 183–189.
Dewey, R.S., Hartley, D.E.H., 2015. Cortical cross-modal plasticity following deafness measured using functional near-infrared spectroscopy. Hear. Res. 325, 55–63.
Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, second ed. Wiley-Interscience, New York.
Ernst, L.H., Schneider, S., Ehlis, A.C., Fallgatter, A.J., 2012. Functional near infrared spectroscopy in psychiatry: a critical review. J. Near Infrared Spec. 20, 93–105.
Ferrari, M., Quaresima, V., 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. NeuroImage 63, 921–935.
Friston, K.J., Ashburner, J.T., Kiebel, K.E., Nichols, T.E., Penny, W.D., 2008. Statistical parametric mapping: the analysis of functional brain images. Academic Press, San Diego.
Fuchino, Y., Sato, H., Maki, A., Yamamoto, Y., Katura, T., Obata, A., Koizumi, H., Yoro, T., 2006. Effect of fMRI acoustic noise on sensorimotor activation examined using optical topography. NeuroImage 32, 771–777.
Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition, second ed. Academic Press, New York.
Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H., 2003. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans. Neur. Sys. Rehab. 11, 141–144.
Hall, A.J., Lomber, S.G., 2015. High-field fMRI reveals tonotopically-organized and core auditory cortex in the cat. Hear. Res. 325, 1–11.
Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., Schultz, T., 2014. Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. Front. Hum. Neurosci. 7, 935.
Herff, C., Heger, D., Putze, F., Guan, C.T., Schultz, T., 2012. Cross-subject classification of speaking modes using fNIRS. Neural Inf. Process. 7664, 417–424.
Herrmann, B., Henry, M.J., Scharinger, M., Obleser, J., 2013. Auditory filter width affects response magnitude but not frequency specificity in auditory cortex. Hear. Res. 304, 128–136.
Holper, L., Biallas, M., Wolf, M., 2009. Task complexity relates to activation of cortical motor areas during uni- and bimanual performance: a functional NIRS study. NeuroImage 46, 1105–1113.
Hong, K.-S., Naseer, N., Kim, Y.-H., 2015. Classification of prefrontal and motor cortex signals for three-class fNIRS-BCI. Neurosci. Lett. 587, 87–92.
Hong, K.-S., Nguyen, H.-D., 2014. State-space models of impulse hemodynamic responses over motor, somatosensory, and visual cortices. Biomed. Opt. Express 5, 1778–1798.
Hu, X.-S., Hong, K.-S., Ge, S.S., Jeong, M.-Y., 2010. Kalman estimator- and general linear model-based on-line brain activation mapping by near-infrared spectroscopy. Biomed. Eng. Online 9, 82.
Hu, X.-S., Hong, K.-S., Ge, S.S., 2012. fNIRS-based online deception decoding. J. Neural Eng. 9, 026012.
Hu, X.-S., Hong, K.-S., 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.
Jancke, L., Mirzazade, S., Shah, N.J., 1999. Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. Neurosci. Lett. 266, 125–128.
Kamran, M.A., Hong, K.-S., 2013. Linear parameter-varying model and adaptive filtering technique for detecting neuronal activities: an fNIRS study. J. Neural Eng. 10, 056002.
Kamran, M.A., Hong, K.-S., 2014. Reduction of physiological effects in fNIRS waveforms for efficient brain-state decoding. Neurosci. Lett. 580, 130–136.
Khan, M.J., Hong, K.-S., 2015. Passive BCI based on drowsiness detection: an fNIRS study. Biomed. Opt. Express 6, 4063–4078.
Khan, M.J., Hong, M.J., Hong, K.-S., 2014. Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface. Front. Hum. Neurosci. 8, 244.

Kong, Y.Y., Mullangi, A., Ding, N., 2014. Differential modulation of auditory responses to attended and unattended speech in different listening conditions. Hear. Res. 316, 73–81.
Kovelman, I., Mascho, K., Millott, L., Mastic, A., Moiseff, B., Shalinsky, M.H., 2012. At the rhythm of language: brain bases of language-related frequency perception in children. NeuroImage 60, 673–682.
Kovelman, I., Shalinsky, M.H., White, K.S., Schmitt, S.N., Berens, M.S., Paymer, N., Petitto, L.A., 2009. Dual language use in sign-speech bimodal bilinguals: fNIRS brain-imaging evidence. Brain Lang. 109, 112–123.
Liebert, A., Wabnitz, H., Steinbrink, J., Moller, M., Macdonald, R., Rinneberg, H., Villringer, A., Obrig, H., 2005. Bed-side assessment of cerebral perfusion in stroke patients based on optical monitoring of a dye bolus by time-resolved diffuse reflectance. NeuroImage 24, 426–435.
Liu, F., Maggu, A.R., Lau, J.C.Y., Wong, P.C.M., 2015. Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. Front. Hum. Neurosci. 8, 1029.
Liu, H., Yu, L., 2005. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowl. Data Eng. 17, 491–502.
Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., Arnaldi, B., 2007. A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural Eng. 4, R1–R13.
Lynnerup, N., Astrup, J.G., Sejrsen, B., 2005. Thickness of the human cranial diploe in relation to age, sex and general body build. Head. Face. Med. 1, 13.
Murata, A., Park, J., Kovelman, I., Hu, X.-S., 2015. Culturally non-preferred cognitive tasks require compensatory attention: a functional near infrared spectroscopy (fNIRS) investigation. Cult. Brain 3, 53–67.
Naseer, N., Hong, K.-S., 2015a. Decoding answers to four-choice questions using functional near-infrared spectroscopy. J. Near Infrared Spec. 23, 23–31.
Naseer, N., 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.
Naseer, N., Hong, K.S., 2015b. fNIRS-based brain-computer interfaces: a review. Front. Hum. Neurosci. 9, 3.
Naseer, N., Hong, M.J., Hong, K.-S., 2014. Online binary decision decoding using functional near-infrared spectroscopy for the development of brain-computer interface. Exp. Brain Res. 232, 555–564.
Olulade, O., Hu, S., Gonzalez-Castillo, J., Tamer, G.G., Luh, W.M., Ulmer, J.L., Talavage, T.M., 2011. Assessment of temporal state-dependent interactions between auditory fMRI responses to desired and undesired acoustic sources. Hear. Res. 277, 67–77.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45, S199–S209.
Plichta, M.M., Gerdes, A.B.M., Alpers, G.W., Harnisch, W., Brill, S., Wieser, M.J., Fallgatter, A.J., 2011. Auditory cortex activation is modulated by emotion: a functional near-infrared spectroscopy (fNIRS) study. NeuroImage 55, 1200–1207.
Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M.S., 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.
Poelmans, H., Luts, H., Vandermosten, M., Ghesquiere, P., Wouters, J., 2012. Hemispheric asymmetry of auditory steady-state responses to monaural and diotic stimulation. Jaro J. Assoc. Res. Oto. 13, 867–876.
Putze, F., Hesslinger, S., Tse, C.-Y., Huang, Y.Y., Herff, C., Guan, C.T., Schultz, T., 2014. Hybrid fNIRS-EEG based classification of auditory and visual perception processes. Front. Hum. Neurosci. 8, 373.
Rinnan, A., van den Berg, F., Engelsen, S.B., 2009. Review of the most common preprocessing techniques for near-infrared spectra. Trac-Trend Anal. Chem. 28, 1201–1222.
Santosa, H., Hong, M.J., Hong, K.-S., 2014. Lateralization of music processing auditory cortex: an fNIRS study. Front. Behav. Neurosci. 8, 418.
Santosa, H., Hong, M.J., Kim, S.P., Hong, K.-S., 2013. Noise reduction in functional near-infrared spectroscopy signals by independent component analysis. Rev. Sci. Instrum. 84, 073106.
Scarff, C.J., Dort, J.C., Eggermont, J.J., Goodyear, B.G., 2004a. The effect of MR scanner noise on auditory cortex activity using fMRI. Hum. Brain Mapp. 22, 341–349.
Scarff, C.J., Reynolds, A., Goodyear, B.G., Ponton, C.W., Dort, J.C., Eggermont, J.J., 2004b. Simultaneous 3-T fMRI and high-density recording of human auditory evoked potentials. NeuroImage 23, 1129–1142.
Schecklmann, M., Ehlis, A.C., Plichta, M.M., Fallgatter, A.J., 2008. Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency. NeuroImage 43, 147–155.
Schlogl, A., Lee, F., Bischof, H., Pfurtscheller, G., 2005. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. J. Neural Eng. 2, L14–L22.
Sevy, A.B.G., Bortfeld, H., Huppert, T.J., Beauchamp, M.S., Tonini, R.E., 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.
Sharda, M., Singh, N.C., 2012. Auditory perception of natural sound categories - an fMRI study. Neuroscience 214, 49–58.
Staeren, N., Renvall, H., De Martino, F., Goebel, R., Formisano, E., 2009. Sound categories are represented as distributed patterns in the human auditory cortex. Curr. Biol. 19, 498–502.
Stothers, L., Shadgan, B., Macnab, A., 2008. Urological applications of near infrared spectroscopy. Can. J. Urol. 15, 4399–4409.
Tai, K., Chau, T., 2009. Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface. J. Neuroeng. Rehabil. 6, 39.
Talavage, T.M., Gonzalez-Castillo, J., Scott, S.K., 2014. Auditory neuroimaging with fMRI and PET. Hear. Res. 307, 4–15.
Tanabe, J., Miller, D., Tregellas, J., Freedman, R., Meyer, F.G., 2002. Comparison of detrending methods for optimal fMRI preprocessing. NeuroImage 15, 902–907.
Toga, A.W., Thompson, P.M., 2003. Mapping brain asymmetry. Nat. Rev. Neurosci. 4, 37–48.
Weichenberger, M., Kuhler, R., Bauer, M., Hensel, J., Bruhl, R., Ihlenfeld, A., Ittermann, B., Gallinant, J., Koch, C., Sander, T., Kuhn, S., 2015. Brief bursts of infrasound may improve cognitive function – an fMRI study. Hear. Res. 328, 87–93.
Wolf, M., Wolf, U., Toronov, V., Michalos, A., Paunescu, L.A., Choi, J.H., Gratton, E., 2002. Different time evolution of oxyhemoglobin and deoxyhemoglobin concentration changes in the visual and motor cortices during functional stimulation: a near-infrared spectroscopy study. NeuroImage 16, 704–712.
Wong, P.C.M., Uppunda, A.K., Parrish, T.B., Dhar, S., 2008. Cortical mechanisms of speech perception in noise. J. Speech Lang. Hear. Res. 51, 1026–1041.
Ye, J.C., Tak, S., Jang, K.E., Jung, J., Jang, J., 2009. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. NeuroImage 44, 428–447.
Zhang, F.Q., Wang, J.P., Kim, J., Parrish, T., Wong, P.C.M., 2015. Decoding multiple sound categories in the human temporal cortex using high resolution fMRI. PLoS One 10, e0117303.