Hearing Research 309 (2014) 84–93

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy

Luca Pollonini [a,*], Cristen Olds [b], Homer Abaya [b], Heather Bortfeld [c], Michael S. Beauchamp [d], John S. Oghalai [b]

[a] Abramson Center for the Future of Health and Department of Engineering Technology, University of Houston, 300 Technology Building, Suite 123, Houston, TX 77204, USA
[b] Department of Otolaryngology – Head and Neck Surgery, Stanford University, 801 Welch Road, Stanford, CA 94305-5739, USA
[c] Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020, USA
[d] Department of Neurobiology and Anatomy, University of Texas Health Science Center at Houston, 6431 Fannin St., Suite MSB 7.046, Houston, TX 77030, USA

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
22 November 2013
Accepted 25 November 2013
Available online 14 December 2013

ABSTRACT

The primary goal of most cochlear implant procedures is to improve a patient's ability to discriminate speech. To accomplish this, cochlear implants are programmed so as to maximize speech understanding. However, programming a cochlear implant can be an iterative, labor-intensive process that takes place over months. In this study, we sought to determine whether functional near-infrared spectroscopy (fNIRS), a non-invasive neuroimaging method which is safe to use repeatedly and for extended periods of time, can provide an objective measure of whether a subject is hearing normal speech or distorted speech. We used a 140 channel fNIRS system to measure activation within the auditory cortex in 19 normal hearing subjects while they listed to speech with different levels of intelligibility. Custom software was developed to analyze the data and compute topographic maps from the measured changes in oxyhemoglobin and deoxyhemoglobin concentration. Normal speech reliably evoked the strongest responses within the auditory cortex. Distorted speech produced less region-specific cortical activation. Environmental sounds were used as a control, and they produced the least cortical activation. These data collected using fNIRS are consistent with the fMRI literature and thus demonstrate the feasibility of using this technique to objectively detect differences in cortical responses to speech of different intelligibility.

1. Introduction

Hearing loss is the fourth most common developmental disorder in the United States and deafness is the most common sensory disorder (Bhasin et al., 2006). The ability to hear during the early years of life is critical for the development of speech, language, and cognition. Early identification and intervention can prevent severe psychosocial, educational, and linguistic repercussions (Robinshaw, 1995; Yoshinaga-Itano et al., 1998). Hence, universal newborn hearing screening programs have been implemented throughout the United States. Similarly, cochlear implantation is encouraged for deaf children. Once implanted, the device needs to be programmed, and this process is repeated iteratively over several months to best fit the child's acoustical needs. The frequency spectrum is divided up and the energy within each specific frequency band is routed to each channel. While the program map needs to be gradually adjusted over time to account for individual variations and needs (Leake et al., 2008), this frequency-to-channel mapping is based on the tonotopic organization of the cochlea (Sridhar et al., 2006).

Cochlear implantation (CI) in young children has been an extraordinary success, however there are individual patients that do poorly (Miyamoto et al., 1994). Possible causes of poor performances are several and diverse, including less-than-optimal programming of the cochlear implant and compromised cochlear innervation. An objective test of how well speech information is received by the brain would be a helpful tool in CI research. For instance, it could provide a quantitative feedback measure in pediatric CI users who are unable to deliver an interpretable behavioral response. In older populations, it could allow an assessment between cortical and behavioral responses to speech, and potentially estimate the pattern of cochlear innervation.

* Corresponding author. Tel.: +1 713 743 4159.
E-mail addresses: lpollonini@uh.edu (L. Pollonini), ceo@stanford.edu (C. Olds), habaya@ohns.stanford.edu (H. Abaya), heather.bortfeld@uconn.edu (H. Bortfeld), Michael.S.Beauchamp@uth.tmc.edu (M.S. Beauchamp), joghalai@ohns.stanford.edu (J.S. Oghalai).

0378-5955/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2013.11.007

fMRI studies have shown that the receptive language region of the auditory cortex (the superior temporal gyrus) demonstrates more activation in response to clear speech than to scrambled speech or environmental sounds in normal-hearing subjects (Belin et al., 2000). However, fMRI is difficult to use in children hearing through a cochlear implant because the device is ferromagnetic and because children typically require general anesthesia to remain still during the test. To overcome these obstacles, we have instead been studying the feasibility of functional near-infrared spectroscopy (fNIRS) to image brain activity in response to speech. This approach is proven to be safe and appropriate for use in assessing language development in children (Gallagher et al., 2012; Quaresima et al., 2012; Rossi et al., 2012; Wilcox et al., 2008, 2009). Previously, we used a 4 channel fNIRS system and demonstrated its ability to detect cortical activation in response to speech (Sevy et al., 2010). Herein, we sought to determine whether fNIRS is sensitive enough to detect differences in cortical activation evoked by different quality levels of speech. To improve our sensitivity to more subtle differences between stimuli, we used a 140 channel fNIRS system and developed custom software to analyze the large amount of data in a meaningful fashion. Consistent with the fMRI literature, fNIRS also demonstrated that the more distorted the speech was, the less region-specific brain activation was observed.

2. Materials and methods

2.1. Participants and experimental protocol

The experimental protocol was approved by the Institutional Review Board of Stanford University, and all subjects signed an informed consent form before participating to the test. Twenty-one adults (age range: 19–63; 18 right-handed, 3 left-handed) were recruited for this study. All passed a hearing screening test and had auditory thresholds of <30 dB HL at 500, 1000 and 2000 Hz. There was no selection bias for hair or skin color during the recruiting process. During the experiment, each subject was asked to sit comfortably in a lit room and to look at a screen showing animated images unrelated to the content or timing of the auditory stimuli. This was done to reduce head movements and to help maintain attention.

We designed four different auditory stimuli. The first stimulus was composed of clips of normal speech, taken from auditory books checked out from the Palo Alto Library. There were three different stories available ("Animal farm", "Great expectations" and "War of the worlds") and the subject was permitted to select which one he/she wanted to listen to. This was done in an effort to increase subject compliance and attentiveness to the auditory stimuli. All of the narrators for the different stories were men and there were no appreciable differences between their voice characteristics except that two spoke with a British accent. Each clip followed the story line sequentially from the previous clip and thus was designed to keep the subject interested in the story line.

The second stimulus was termed channelized speech, and its purpose was to simulate a well-programmed cochlear implant. Thus, the speech clips were re-created using only 12 band-passed frequency "channels" to convey the information. Like the processor of a cochlear implant, we first measured the energy content from the speech clips within the range of 250–8000 Hz within 20 logarithmically-equal frequency bands (or channels). This was done using intervals of 40 ms. An ideal cochlear implant would then direct the energy contained within each frequency band to the channels that stimulate the auditory nerves tuned to those

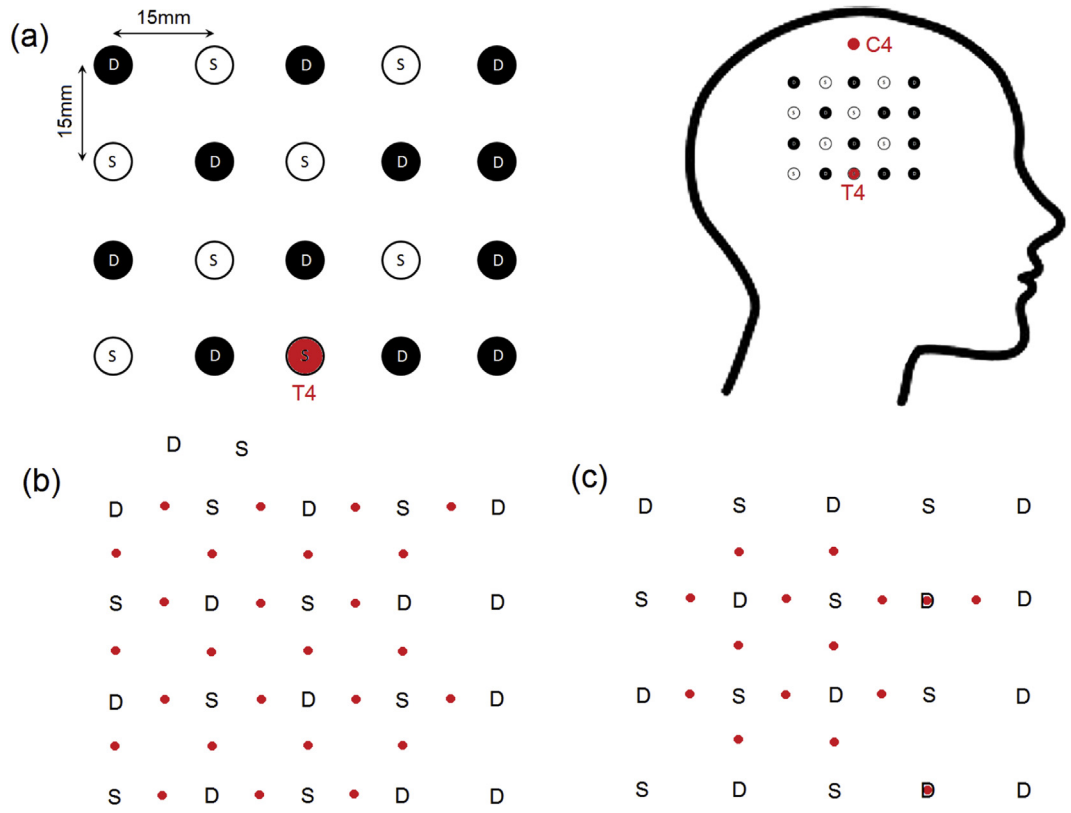

Fig. 1. Headset layout. (a) Layout of optical sources (empty circles) and photodetectors (filled circles) of the right hemisphere probe. The optical source located in the middle of the bottom horizontal line of the probe grid was aligned with T4 of the 10–20 international system. The left hemisphere had an identical layout. (b) Layout of shallow and (c) deep channels determined as mid-point between each source-detector pair.

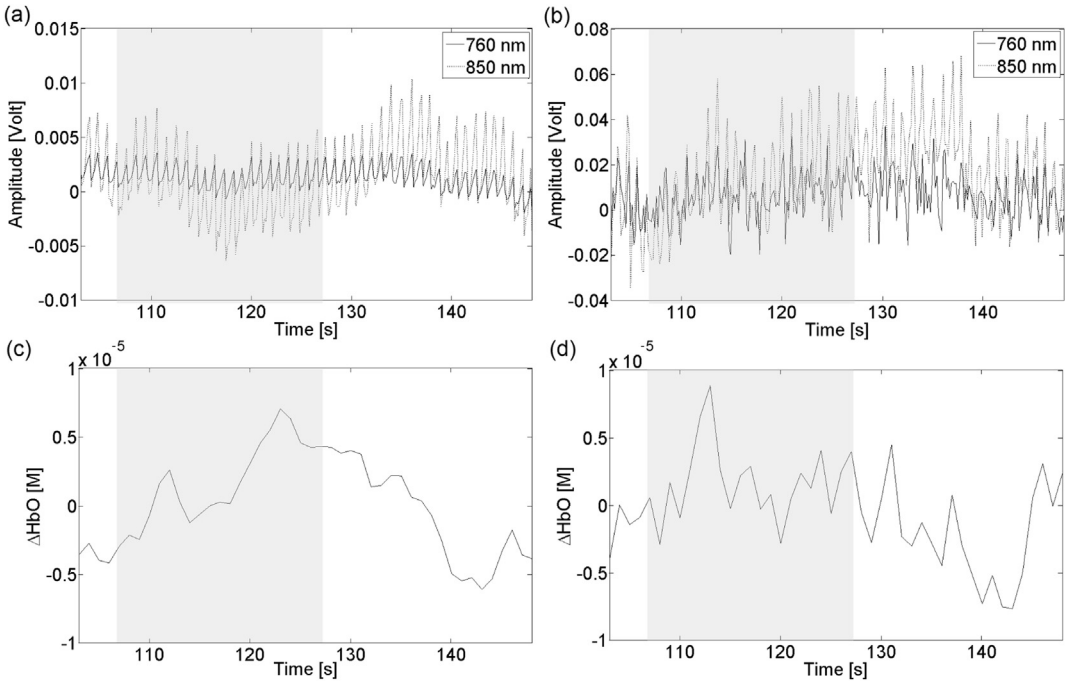

Fig. 2. Representative data from two channels with different scalp coupling indices. (a, b) Raw signals for both wavelengths of transmitted light. (c, d) The calculated HbO signals. The plots on the left (a, c) come from a channel with good scalp contact; the plots on the right (b, d) come from a channel with poor scalp contact. The grayed region indicates the time during which the normal speech stimulus was presented.

frequencies. However because these experiments were done on normal-hearing patients, we simulated this process and re-created sound by summing white noise that was band-pass filtered to each channels' frequency range and amplitude-modulated. In general, channelized speech could be understood by a normal-hearing listener, but it certainly was not as clear as normal speech.

The third stimulus was termed scrambled speech. Initially, we processed the sound identically to channelized speech. However, we simulated an idealized, worst-case scenario of a poorly-programmed cochlear implant in which the sound energy processed by each channel is stimulating auditory nerves not tuned to those frequencies. Thus for example, low-frequency sounds might stimulate high-frequency nerves and vice versa. We accomplished this by using the energy within each stimulus frequency band to amplitude-modulate the noise within channels not in the assigned order, but instead using a random order. In general, this allowed the listener to appreciate the timing of a sentence but not to understand any of the words or meaning. Since the randomized frequency-scrambling process adopted in this study represents an exaggeration of the effect of the current spread occurring in a real-world scenario, the actual hearing experience of individuals with poorly-programmed cochlear implant is likely to be better than the simulation performed in this study.

The fourth stimulus was created by juxtaposing short audio clips (3 s) of environmental sounds such as hardware (e.g. hammering, drilling) or animal noises. This stimulus represented a non-speech control task that does not require language processing.

Auditory stimuli were presented using a pair of desktop loudspeakers (Altec Lansing, New York NY) at a normal listening volume (60 dB HL) for 20 s, followed by 20 s of silence. Each block consisted of four different auditory stimuli presented in random order. Blocks were repeated 5 times with the random order varied each time. Thus, the total length of time required to perform an experiment was 13 min and 20 s (40 s total per stimulus/silence *4 stimuli *5 blocks). Exemplary audio clips for each story and condition are provided as Supplementary Material. Prior to the experiment, the subjects were asked to try their best to pay attention to the narrated words regardless of the speech degradation.

2.2. Instrumentation

Functional near-infrared spectroscopy data was collected with a multi-channel system with 16 points of illumination and 24 points of detection (NIRScout, NIRx Medical Technologies LLC., Glen Head, NY). Each illumination channel consists of two co-located LEDs with emission wavelengths of 760 nm and 850 nm and a miniature collimating lens. Light collection was achieved with optical fiber bundles coupled to silicon photodetectors. Sources and detectors were embedded into plastic snap-on enclosures for mounting. To investigate the response of the primary auditory cortex in both hemispheres, we arranged the FNIRS optodes (i.e. optical sources and detectors) in two symmetrical probes, each with 8 illuminating and 12 collecting channels (Fig. 1a).

Each probe layout contained a total of 70 usable source-detector pairings (or channels) with source-detector distances of 15 mm (26 channels), 21 mm (3 channels), 30 mm (2 channels), 33 mm (28 channels), 42 mm (2 channels), and 45 mm (9 channels). Channels

Table 1
Summary of source-detector distance for all channels with SCI>0.75 in one representative subject.

Source-detector distance [mm]	Left hemisphere (# channels)	Right hemisphere (# channels)
15	21	26
21	3	2
30	2	1
33	7	7
42	0	1
45	1	1
Total	34	38

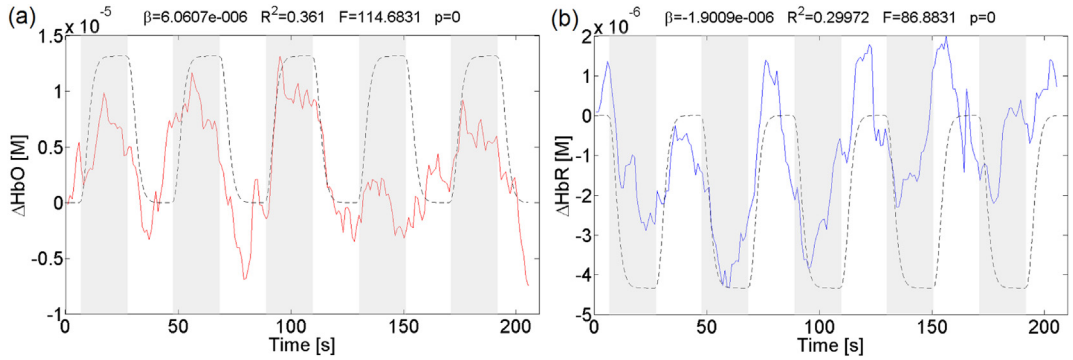

Fig. 3. HbO and HbR data from one channel of a representative subject. The time course of the response for (a) HbO and (b) HbR to five repetitions of the normal speech stimulus. The dashed lines are the predicted hemodynamic responses. The grayed regions indicate the time periods when the stimulus was presented.

with source-detector distances greater than 45 mm were not included in the analysis due to negligible intensity of the photo-detected signal (26 additional potential channels). Light sources were time-multiplexed to remove the potential for artifacts from the other wavelength at that same site or either wavelength from the other sources. However, fNIRS signals were acquired simultaneously from both hemispheres, as the light emitted on one hemisphere could not be detected by the photodiodes of the contralateral probe. Thus, an overall sampling rate of 6.25 Hz for each hemisphere was achieved.

After the participant was comfortably seated in the experiment room, the fNIRS headgear was placed onto the subject's head and secured in place with Velcro straps. The probes were positioned in a repeatable location on each subject's head by aligning the optical source in the middle of the bottom row at the T3 and T4 locations of the 10–20 international system in the left and right hemisphere, respectively. The tightness of the headgear was adjusted to guarantee secure contact between the optodes and the scalp while still offering a level of comfort acceptable to the subject. Prior to the auditory stimulation, the experimenter verified that the hair layer did not obstruct the light passage between sources and detectors by visualizing the interface between them and the skin. Intervening hair was moved out of the way with a cotton-tipped applicator.

2.3. Signal filtering and channels selection

Noise in functional NIRS measurements arises from gross motion of the subject's head as well as physiological noise sources, including cardiac, respiratory and vasomotor (Mayer waves) (Fekete et al., 2011b). There is an active literature on noise reduction in fNIRS, although a consensus on optimal strategies has not been reached (Ayaz et al., 2010; Cui et al., 2010; Fekete et al., 2011a;

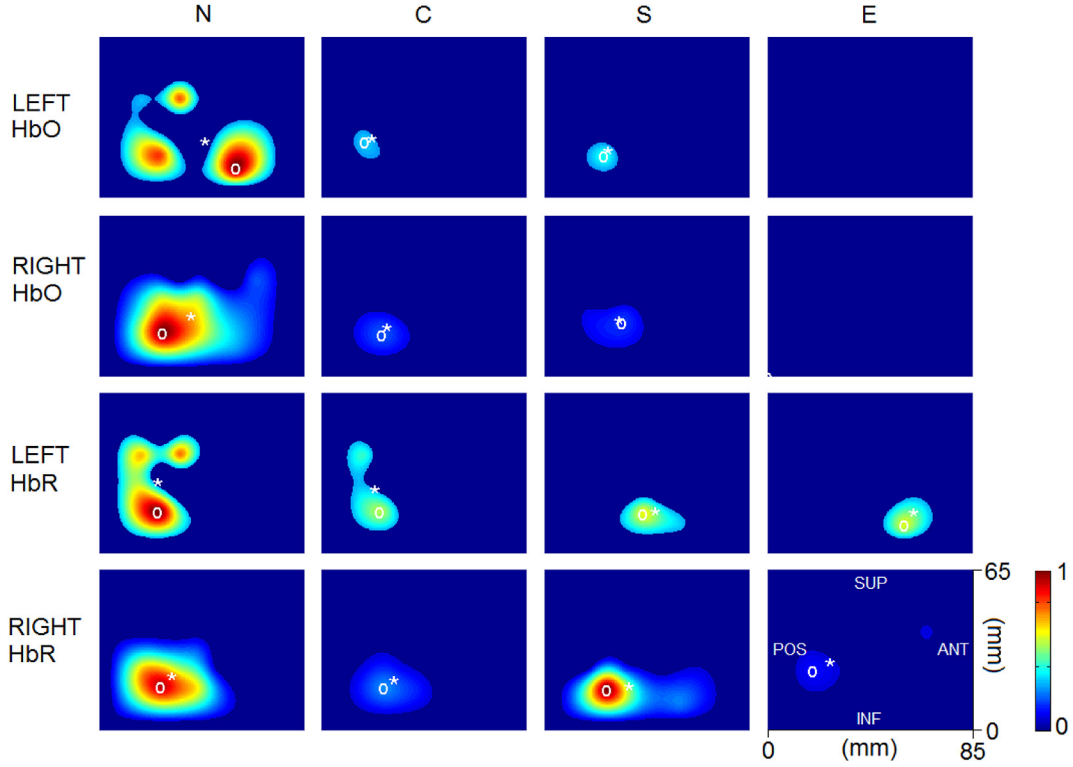

Fig. 4. Statistical parametric maps of a representative subject. The color-coded maps show the responses to normal speech (N), channelized speech (C), scrambled speech (S), and environmental sounds (E). Each map reports the location of center of mass (*) and peak of activity (○) of active areas. The scalp orientation of each plot is shown in the bottom right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

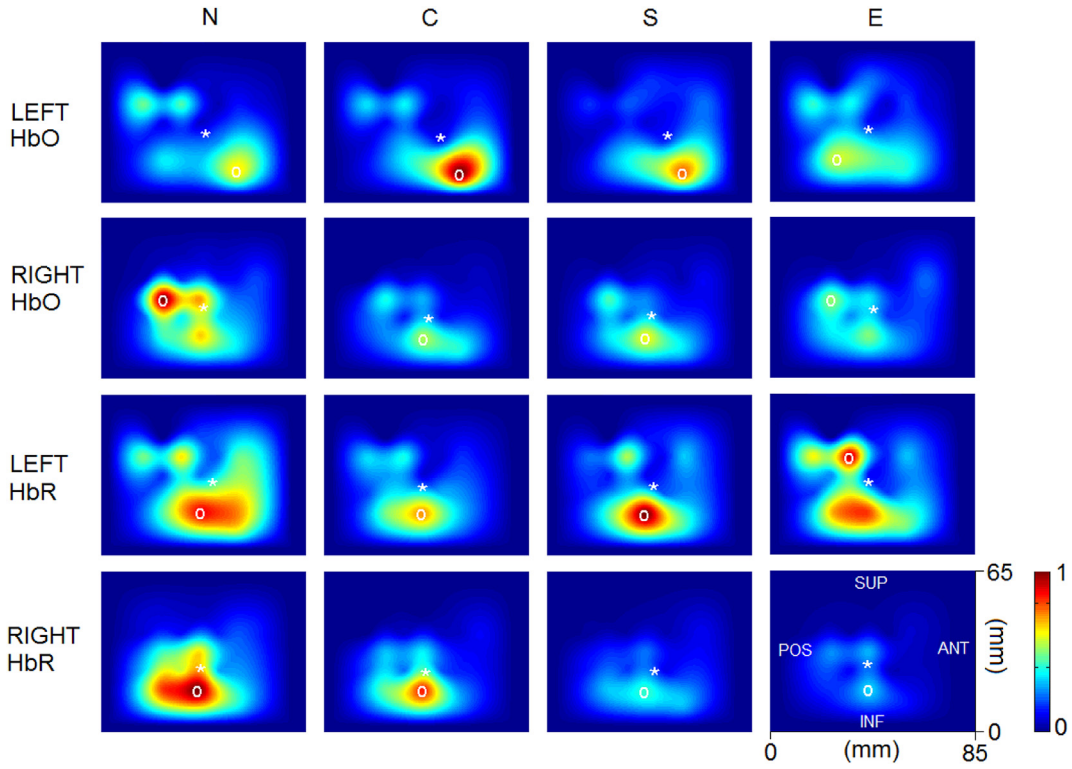

Fig. 5. Grand average of statistical parametric maps. The color-coded maps show the group average of the responses to normal speech (N), channelized speech (C), scrambled speech (S), and environmental sounds (E). The scalp orientation of each plot is shown in the bottom right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Huppert et al., 2009; Izzetoglu et al., 2005, 2010; Pei et al., 2007; Ye et al., 2009). Substantial differences between the optical probes and FNIRS instrumentation used at different sites further complicate methods development.

Although a strong cardiac oscillation in fNIRS raw signals is undesirable for measuring evoked cortical hemodynamic responses, its presence is related to intracranial physiological parameters and thus indicates a good contact between the optical probe and the scalp (Themelis et al., 2004, 2007). Therefore, the first step of our data processing was to identify optical channels showing a strong cardiac signal. Since the LED sources at 760 nm and 850 nm were co-located, an optical channel in good contact with the scalp exhibited a prominent synchronous cardiac pulsation in both photodetected signals. This observation was independent of the amplitude of the output voltage of the photodetector, which in turn depends on the inter-distance between sources and detector. For each channel, we filtered both photodetected signals between 0.5 and 2.5 Hz to preserve only the cardiac component and normalized the resulting signals to balance any difference between their amplitude. Then, we computed the cross-correlation and we extracted the value at a time lag of 0 to quantify the similarity between the filtered signals. In-phase and counter-phase identical waveforms yielded a zero-lag cross-correlation value of 1 and −1 respectively, whereas a null value derived from totally uncorrelated signals. Therefore, the zero-lag cross-correlation between photodetected signals of the same channel was used as a quantitative measure of the signal-to-noise ratio of the channel. We termed this value the *scalp coupling index* (SCI).

Fig. 2a and b illustrate examples of raw signals of channels with good and poor contact with the scalp, respectively. Both channels were selected from a single fNIRS scan performed on a volunteer. The resulting scalp coupling indices (SCIs) were 0.97 and 0.48, respectively. Fig. 2c and d show the resulting change in oxyhemoglobin (HbO) signals in response to normal speech (grayed interval). The optical channel with a good scalp contact revealed a clean task-related hemodynamic change, whereas the channel with poor scalp contact resulted in a noisy hemodynamic signal. Based on this, we only used channels with an SCI-value of +0.75 or greater when analyzing the data in this study; channels with a lower SCI value were considered to have poor scalp coupling and were not considered for further processing.

The raw signals from the selected channels were then bandpass-filtered between 0.016 Hz and 0.8 Hz to remove cardiac and respiratory noise sources. The signals were normalized to their temporal mean value and de-trended using a 10th degree polynomial function to eliminate slow-trended temporal drifts (Pei et al., 2007). Finally, the modified Beer–Lambert law (Cope et al., 1988; Kocsis et al., 2006) was applied to calculate the changes in oxygenated and deoxygenated hemoglobin concentration for each channel and time point. Since continuous-wave FNIRS does not allow estimating the differential path length factor in tissues, the hemoglobin concentration measurements was computed per unit path length.

2.4. Topographic mapping

Hemodynamic signals obtained from both hemispheres were linearly correlated with the expected hemodynamic response according to general linear model theory (GLM). The expected response was modeled with a gamma-variate function (Cox, 1996; Ward, 1998) convolved with the event stimuli, in a manner similar to the analysis of fMRI data (Beauchamp et al., 2004, 2008). For oxyhemoglobin (HbO), we expected an increase of concentration due to the augmented blood supply required by the neuronal activation; conversely, deoxyhemoglobin (HbR) is predicted to

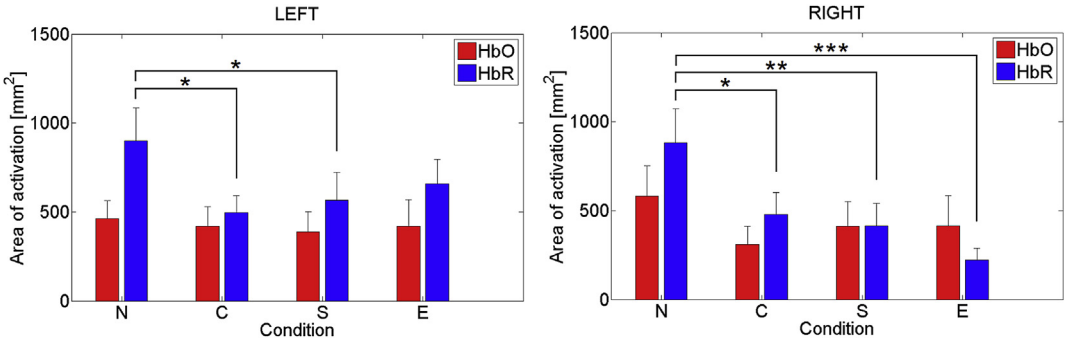

Fig. 6. Group average area of activation. Group means ($N = 19$) and standard error of active areas derived from deep maps. Normal speech (N), channelized speech (C), scrambled speech (S), and environmental sounds (E). (*) $p < 0.05$, (**) $p < 0.01$, (**) $p < 0.001$.

decrease due to increased oxygen consumption occurring at the cortical site of activation. For each channel and hemoglobin type (i.e. HbO and HbR), linear modeling (performed using the *regress* function in MATLAB) yielded the weight of the regression (β-value), goodness of fit (R^2-value), *F*-statistics (*F*-value), and probability of occurrence of the null-hypothesis (*p*-value) for each of the four stimulus categories.

A statistical parametric map for each hemisphere and speech condition was reconstructed using the *F*-values of each channel. We created two different maps, a shallow map and a deep map, for each hemisphere. We assumed that a source-detector pair that is close together will primarily detect hemodynamic changes that occur in a relatively shallow region (i.e. the scalp). In contrast, we assumed that a source-detector pair that is farther apart will be able to detect hemodynamic changes that are deeper (the outer layer of the brain (Sevy et al., 2010)). Channels with source-detector distances of 15–21 mm were used to generate the shallow map and channels with source-detector distances greater than 30 mm were used to generate the deep map.

The location of each channel was determined as the mid-point between each source-detector pair, which corresponded to the surface projection of the point of maximum depth reached by the photons migrating from the source to the detector. The resulting layout of the shallow and deep channels is illustrated in Fig. 1b and c. A color-coded activation map was created by interpolating the *F*-values of the channels surviving the SCI-thresholding using a bidimensional polyharmonic spline. The pixel size was set at 0.25 mm × 0.25 mm. For this study, in which we wanted to only study brain activation, we eliminated the shallow channels from all analyses and only studied the deep channels.

SCI-based channel selection and topographic imaging were implemented in MATLAB, whereas bandpass filtering and hemoglobin concentrations changes were calculated with the software provided by the manufacturer (Pei et al., 2007).

2.5. Image analysis and statistics

Topographic maps for each hemoglobin state (HbO and HbR), each hemisphere (left and right), and each stimulus (normal speech, channelized speech, scrambled speech and environmental sounds) were computed for each individual participating in the study. A threshold value was arbitrarily set on the *F*-statistics maps to identify the areas where the cerebral hemodynamic response was significant. After thresholding, the surface area of the significant regions was calculated.

At group level, we used the analysis of variance (ANOVA) to test if the area activated by the four auditory stimuli was significantly different. The ANOVA significance level was set at $p = 0.1$. Prior to applying ANOVA, we tested if the sphericity assumption was verified. Where indicated by the ANOVA test, we then evaluated for pairwise differences between stimuli using the paired Student's *t*-test with the significance level set at $p = 0.05$.

2.6. fMRI imaging

One subject (a 32 year old normal-hearing female who did not take part in the fNIRS testing) was tested with blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI). Scanning was performed using a Siemens 3 T MRI scanner at the Baylor College of Medicine Center for Advanced MRI. First, two repetitions of a T1-weighted MP-RAGE sequences were collected and used to create a cortical surface model using FreeSurfer (Fischl et al., 1999) and visualized using SUMA (Saad and Reynolds, 2012). Next, five scan series of gradient-echo EPI were collected (33 3.0 mm axial slices, 64 × 64 matrix, in-plane dimension 3.0 mm by 3.0 mm). Each scan series contained 160 brain volumes with TR = 2 s. The same stimulus set as used in the fNIRS experiment was presented using a block design. There were 8 blocks in each scan series. Each block contained one stimulus type (20 s) followed by no-stimulus baseline (20 s). There were 2 repetitions of each type of stimulus block in each scan series, in pseudo-random order (10 blocks of each stimulus type across all scan series). EPI volumes were concatenated, aligned to the high-resolution MP-RAGE, resampled to 2.5 mm isotropic, and analyzed using the generalized linear model in AFNI (Cox, 1996) with head motion estimates used as regressors of no interest (for command lines, see (Nath and Beauchamp, 2011). There were four regressors of interest, one for each stimulus category. Activation maps were created separately for each stimulus category, using the *t*-statistic for that category's regressor of interest with a threshold of $t > 3.7$. This threshold was selected because it provided a false discovery rate corrected statistical significance (q) of $q < 0.01$. The volume of active cortex for each stimulus category was estimated using a clustering algorithm that found all neighboring voxels exceeding the significance threshold. For each stimulus category, there were only two clusters with volume >400 voxels, consisting of auditory cortex (and neighboring regions) in left and right hemisphere.

3. Results

3.1. Functional NIRS – representative analysis from one subject

fNIRS signal processing was performed in a representative subject (32 year old male, right-handed). Initially, we selected the cleanest channels based on the value of the scalp contact index (SCI) in each channel. With a SCI threshold of 0.75, 34 and 38 channels survived the selection in the left and right hemisphere, respectively. Table 1 summarizes the source-detector separation of

the selected channels. The channels were then processed and analyzed. The HbO signal increased following the onset of speech stimuli, whereas the HbR signal decreased for most of the auditory trials (Fig. 3). Consistent with cortical activation, the HbO response increased and the HbR response decreased during five stimulus trials. The *F*-values resulting from the GLM regression were used to construct statistical parametric maps for all hemodynamic signals and speech conditions. Subsequently, we imposed an *F*-value threshold of 15 to determine the significant areas of activation (Fig. 4). The largest area of cortical activation was found for normal speech, and the responses were reduced with channelized and scrambled speech. Environmental sounds produced the least area of cortical activation.

We also determined the point of highest activation as the pixel with the highest *F*-value. We calculated the center of mass by averaging the *F*-values for all significant pixels. These values shifted with the different stimuli, but did not demonstrate a clear pattern.

3.2. Group analysis

Out of the twenty-one adults recruited for this study, two subjects had gross and excessive movement artifact in their recorded data tracings as visualized by eye and were excluded from further analysis. All 19 of the remaining subjects (90% of the total population) had measurable fNIRS responses, defined as a significant cortical response to at least one of the auditory stimuli in at least one channel. We then calculated the areas of cortical activation imposing a common statistical threshold of $F > 15$ and compared the effect of the different stimuli. First, we computed the group average of the statistical parametric maps. After group averaging, the maps were normalized to the maximum *F*-value across the four auditory conditions (Fig. 5). As a result, all maps were color-coded between magnitudes 0 (deep blue) and 1 (dark red). The channelized speech evoked the strongest response in the HbO map in the left hemisphere, whereas normal speech evoked the strongest HbO response in the right hemisphere. The HbR maps demonstrated more activation with normal speech than with the other stimuli in both hemispheres, although this was particularly obvious in the right hemisphere.

We then quantified these data by calculating the area of activation for each stimulus in all subjects (Fig. 6). Each significant pixel was given equal weighting for this calculation. Changes in the area of activation based on the HbO response were not significantly different across conditions for either hemisphere (one-way ANOVA, LH: $p = 0.96$, RH: $p = 0.49$). However, the area of activation based on the HbR response varied between the four stimuli in both hemispheres (one-way ANOVA, LH: $p = 0.08$, RH: $p = 0.007$). Since the area of activation evoked by normal speech was clearly larger than the area of activation evoked the other stimuli, we then studied this effect in a more detailed fashion. In both hemispheres, the area activated by normal speech was larger than by channelized speech (paired *t*-test, LH: $p = 0.010$, RH: $p = 0.034$) and scrambled speech (LH: $p = 0.049$, RH: $p = 0.005$). Normal speech activated a larger area than environmental sounds only in the right hemisphere (LH: $p = 0.056$, RH: $p = 0.0007$). We found no differences in the area of activation evoked by channelized speech or scrambled speech conditions in either hemisphere for either HbO or HbR (paired *t*-test, $p > 0.30$ for all comparisons).

To assess whether there were shifts in the location of cortical activation associated with the intelligibility of the speech, we calculated the peak of activation and the center of mass of the active area for each subject for each stimulus. The distance between active areas of normal speech and the other three stimuli as well as between channelized and scrambled speech was calculated pairwise (N–C, N–S, N–E, C–S). Specifically, we calculated the vector difference between the peaks of activity (and centers of mass) of pairwise combinations for all subjects and applied the Hotelling's multivariate *T*-squared test on its orthogonal-axis components. We found a significant shift between the peaks of activity of normal and channelized speeches for HbR in the right hemisphere

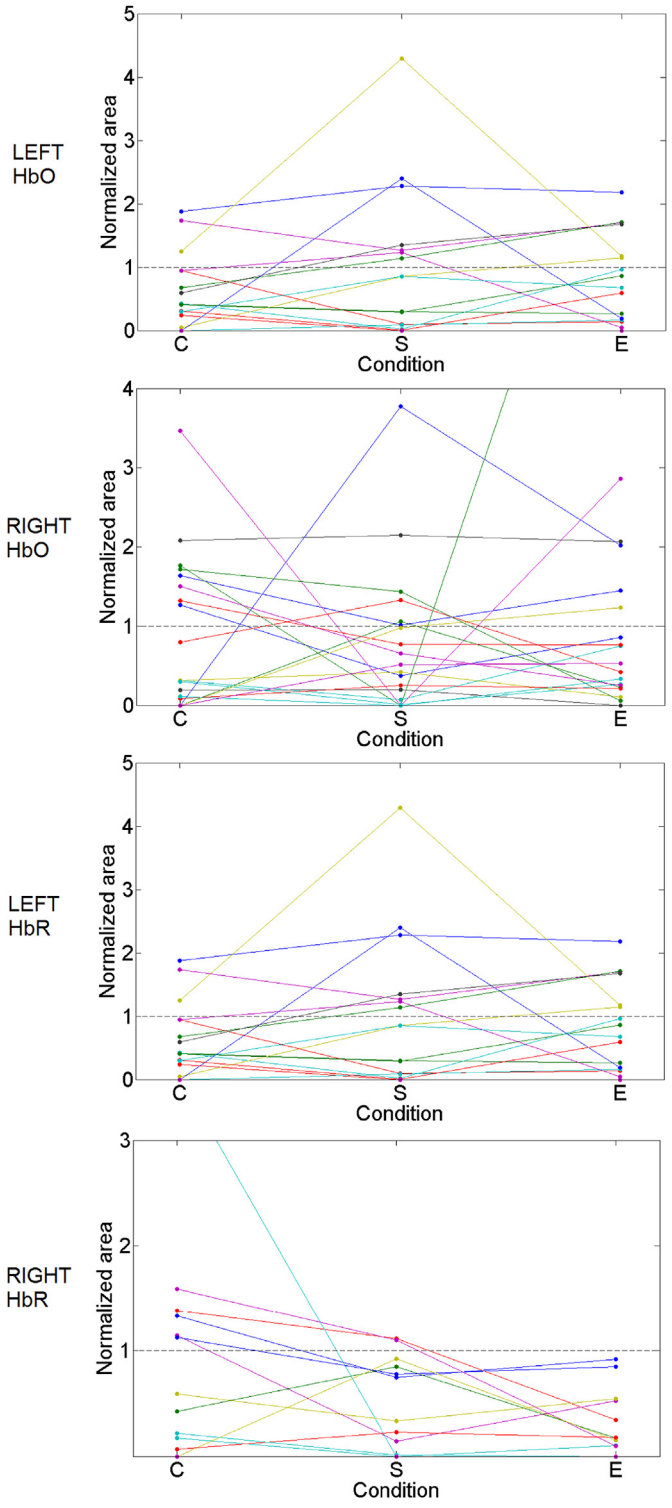

Fig. 7. Individual responses normalized to normal speech. Activation areas for HbO and HbR for the left and right hemispheres. Areas are normalized to the response area measured with normal speech. Channelized speech (C), scrambled speech (S), and environmental sounds (E).

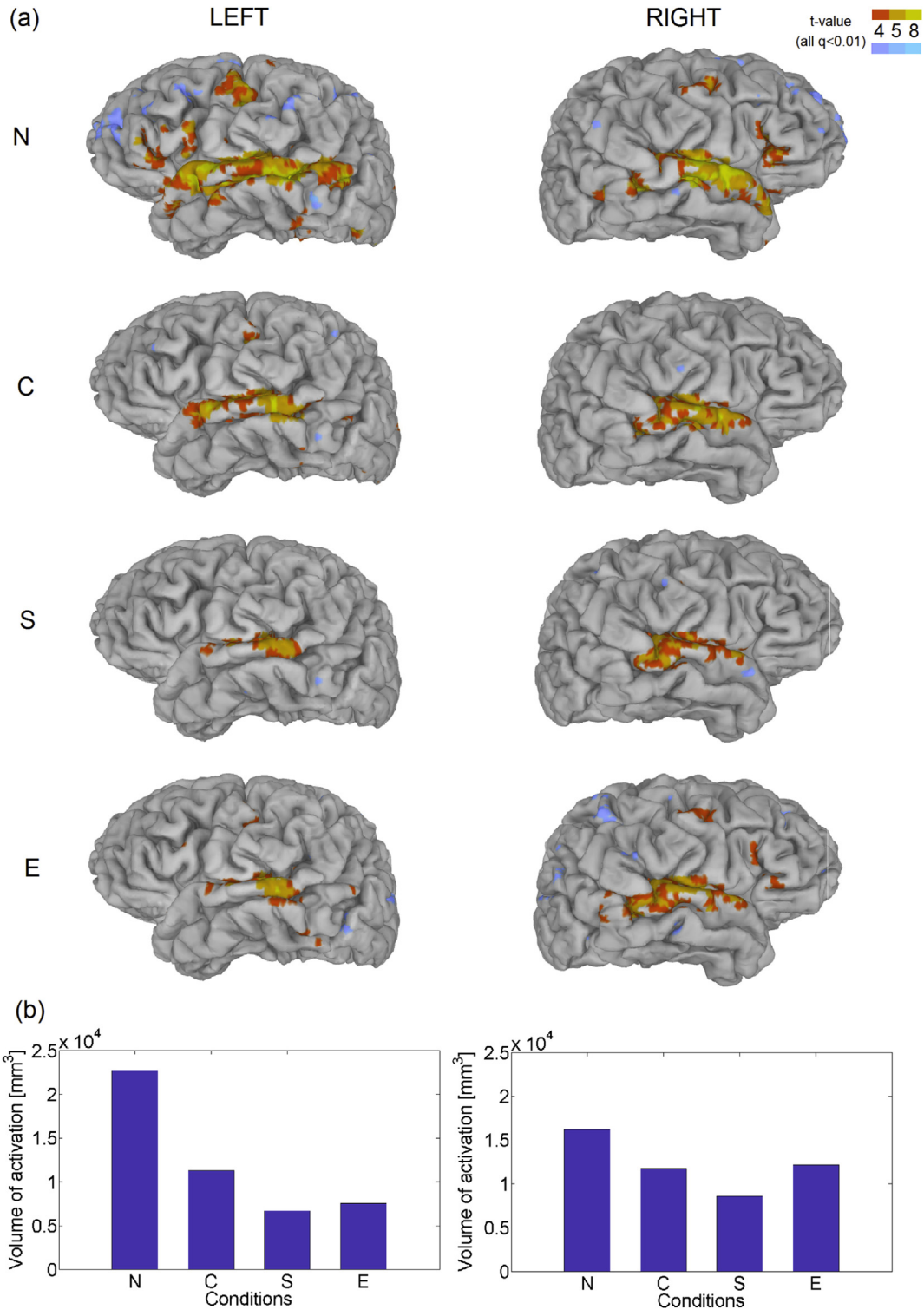

Fig. 8. fMRI activations of a representative subject. (a) BOLD fMRI responses to sound stimuli. Orange-to-yellow color scale shows voxels that were significant activated to sound stimuli (sound > resting baseline); blue color scale shows voxels that were significantly deactivated to sound stimuli (sound < resting baseline). A lateral view of a reconstruction of the cortical surface model is shown for the left hemisphere (left column) and right hemisphere (right column) for normal speech, channelized speech, scrambled speech, environmental sounds. (b) Volume of active cortex for each sound stimulus in left hemisphere (left plot) and right hemisphere (right plot). The volume of voxels exceeding the significance threshold ($t > 3.7, q < 0.01$) for each stimulus category after a cluster analysis; results show only the largest cluster in each hemisphere, consisting of left and right auditory cortex. Normal speech (N), channelized speech (C), scrambled speech (S), and environmental sounds (E). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

($p = 0.04$). However, no other statistically-significant shifts were found. Importantly, while the averaged HbO signal in the right hemisphere demonstrates obvious differences between normal speech and the other three conditions (see the second row in Fig. 5), this effect was due to one subject having a large difference, whereas the rest of the subjects did not. Therefore, since the calculation of shifts for statistical testing (Hotelling's test) was performed at the individual level, this was not a statistically significant effect.

3.3. Single-subject analysis

Finally, we compared single-subject responses to the group averages in order to evaluate the feasibility of using fNIRS to assess cortical responses on individual subjects. To achieve this, we plotted each subjects' change in response, normalized to his/her response to normal speech (Fig. 7). No clear differences emerged in the left hemisphere for both HbO and HbR, and in the right hemisphere for HbO. In the right hemisphere, the majority of the subjects demonstrated less HbR activation as the stimuli became less speech-like, and only few subjects had a response for channelized speech greater than normal speech. However, the differences between conditions at individual level were not as evident as in group averages. This finding is aligned with previous fNIRS studies in which group-level results were found to be sufficiently repeatable, whereas single-subject analyses were less reproducible (Plichta et al., 2006; Schecklmann et al., 2008; Zhang et al., 2011).

3.4. Functional MRI

As a control, we also performed fMRI to verify the stimulus paradigm on one subject (Fig. 8A). All four stimulus conditions strongly activated auditory and auditory-related cortex regions. To quantify the differences between the stimulus conditions, we measured the volume of active cortex in the auditory cortex of the left and right hemispheres. In both hemispheres, the largest volume of active cortex was evoked by normal speech stimuli (Fig. 8B). This is consistent with our findings with fNIRS (compare with Fig. 5), particularly within the right hemisphere for HbR, that is the paramagnetic molecule which metabolic changes are measured by fMRI.

4. Discussion

This study shows that the auditory cortex of normal-hearing individuals responds differently to variations in speech intelligibility, and that these differences are detectable using fNIRS. The use of large channel arrays is feasible, enhances the imaging resolution, and yields quantitative data that are able to be more accurately compared. An important caveat is that these results may not describe the receptive language ability of the auditory cortex that translates behaviorally into speech understanding. Nevertheless, fNIRS is a useful technique to investigate the cortical response to speech stimulation in groups of human subjects. Further work is needed to develop an experimental protocol and data interpretation technique that permit reliable single-subject analysis.

The largest area of activation was found when subjects listened to normal speech; degrading the speech decreased the area of activation and, in limited circumstances, shifted the position of the peak of activity. This is consistent with previous fMRI studies demonstrating that regions of human auditory cortex selectively respond to speech (Belin, 2006; Belin et al., 2002, 2000).

Our data indicate that fNIRS is most responsive to changes in activation within the right hemisphere, which is surprising because language processing has been predominantly associated to cortical activity in the left hemisphere (Frost et al., 1999). This may be because these stimuli evoke different responses in more superficial areas of the right cortex than in the left cortex, making them more easily detected by fNIRS.

In our previous study (Sevy et al., 2010), we used a system that offered only two functional NIRS channels per hemisphere, whereas the system we used in this study had 70 functional channels per hemisphere. The benefit of having more channels is that we could develop topographic maps and measure the area of activation and the center of mass, rather than just the activation amplitude and time course. This makes the data analysis more similar to that used with fMRI. As well, including more channels allows the identification of channels with a poor signal-to-noise ratio. These channels could then be excluded from the analysis process. Therefore, comparisons between stimuli for any given subject were performed using only those subjects' channels which had good sensitivity to hemodynamic changes.

One somewhat surprising result of this study is our finding of larger effects in the HbR responses than in the HbO responses. This was seen in our previous fNIRS study as well (Sevy et al., 2010). However, most fNIRS studies only present HbO data (Ferrari and Quaresima, 2012) because it has a lower noise level, and thus more obvious responses. With our system, we found that the HbR and HbO noise levels were comparable, possibly because we removed channels with low signal-to-noise levels prior to analysis. In any case, the BOLD response from the fMRI data, which comes from changes in HbR, confirmed our observations.

In this study, we only enrolled normal-hearing adults in order to best define the capabilities of fNIRS in this most straightforward situation. However, a long-term goal of this research is to use fNIRS to objectively assess how well speech is relayed to the auditory cortex in deaf subjects after cochlear implantation. Of course, our simulations of what a patient hears through a cochlear implant may not be representative, although we did base these simulations on the principles of vocoded speech (Shannon et al., 1995). In this study, we presumed that when a deaf patient hears normal speech through a cochlear implant, it sounds like channelized speech in the best situation (no loss of auditory nerves, well-programmed cochlear implant, etc.). We simulated the worst situation using scrambled speech, which is probably worse than would typically be found in human subjects. The channels would still stimulate the cochlea in a semi-tonotopic fashion, although the coiled nature of the cochlea and the fact that low frequency neurons travel through the modiolus near the basal turn channels could lead to unusual frequency transposition effects. In addition, the brain plasticity that occurs after cochlear implantation may elicit fNIRS responses that could be erroneously interpreted. Therefore, further research is needed to assess whether fNIRS provides a valid objective measure of speech perception in human subjects hearing through a cochlear implant.

Acknowledgments

This research was supported by NIH R56DC010164 and R01DC010075.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.heares.2013.11.007

References

Ayaz, H., Izzetoglu, M., Shewokis, P.A., Onaral, B., 2010. Sliding-window motion artifact rejection for Functional Near-Infrared Spectroscopy. In: Conf Proc IEEE Eng Med Biol Soc 2010, pp. 6567–6570.

Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41, 809–823.
Beauchamp, M.S., Yasar, N.E., Frye, R.E., Ro, T., 2008. Touch, sound and vision in human superior temporal sulcus. Neuroimage 41, 1011–1020.
Belin, P., 2006. Voice processing in human and non-human primates. Philos. Trans. R Soc. Lond. B Biol. Sci. 361, 2091–2107.
Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Brain Res. Cogn. Brain Res. 13, 17–26.
Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.
Bhasin, T.K., Brocksen, S., Avchen, R.N., Van Naarden Braun, K., 2006. Prevalence of four developmental disabilities among children aged 8 years–Metropolitan Atlanta Developmental Disabilities Surveillance Program, 1996 and 2000. MMWR Surveill. Summ. 55, 1–9.
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. 222, 183–189.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Cui, X., Bray, S., Reiss, A.L., 2010. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. Neuroimage 49, 3039–3046.
Fekete, T., Rubin, D., Carlson, J.M., Mujica-Parodi, L.R., 2011a. A stand-alone method for anatomical localization of NIRS measurements. Neuroimage 56, 2080–2088.
Fekete, T., Rubin, D., Carlson, J.M., Mujica-Parodi, L.R., 2011b. The NIRS Analysis Package: noise reduction and statistical inference. PLoS One 6, e24322.
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.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.
Frost, J.A., Binder, J.R., Springer, J.A., Hammeke, T.A., Bellgowan, P.S.F., Rao, S.M., Cox, R.W., 1999. Language processing is strongly left lateralized in both sexes: evidence from functional MRI. Brain 122, 199–208.
Gallagher, A., Beland, R., Lassonde, M., 2012. The contribution of functional near-infrared spectroscopy (fNIRS) to the presurgical assessment of language function in children. Brain Lang. 121, 124–129.
Huppert, T.J., Diamond, S.G., Franceschini, M.A., Boas, D.A., 2009. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl. Opt. 48, D280–D298.
Izzetoglu, M., Devaraj, A., Bunce, S., Onaral, B., 2005. Motion artifact cancellation in NIR spectroscopy using Wiener filtering. IEEE Trans Biomed. Eng. 52, 934–938.
Izzetoglu, M., Chitrapu, P., Bunce, S., Onaral, B., 2010. Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. Biomed. Eng. Online 9, 16.
Kocsis, L., Herman, P., Eke, A., 2006. The modified Beer-Lambert law revisited. Phys. Med. Biol. 51, N91–N98.
Leake, P.A., Stakhovskaya, O., Hradek, G.T., Hetherington, A.M., 2008. Factors influencing neurotrophic effects of electrical stimulation in the deafened developing auditory system. Hear Res. 242, 86–99.
Miyamoto, R.T., Osberger, M.J., Todd, S.L., Robbins, A.M., Stroer, B.S., Zimmerman-Phillips, S., Carney, A.E., 1994. Variables affecting implant performance in children. Laryngoscope 104, 1120–1124.
Nath, A.R., Beauchamp, M.S., 2011. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J. Neurosci. 31, 1704–1714.
Pei, Y., Xu, Y., Barbour, R.L., 2007. NAVI-SciPort solution: a problem solving environment (PSE) for NIRS data analysis. In: Human Brain Mapping 2007, Chicago, IL.
Plichta, M.M., Herrmann, M.J., Baehne, C.G., Ehlis, A.C., Richter, M.M., Pauli, P., Fallgatter, A.J., 2006. Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable? Neuroimage 31, 116–124.
Quaresima, V., Bisconti, S., Ferrari, M., 2012. A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. Brain Lang. 121, 79–89.
Robinshaw, H.M., 1995. Early intervention for hearing impairment: differences in the timing of communicative and linguistic development. Br. J. Audiol. 29, 315–334.
Rossi, S., Telkemeyer, S., Wartenburger, I., Obrig, H., 2012. Shedding light on words and sentences: near-infrared spectroscopy in language research. Brain Lang. 121, 152–163.
Saad, Z.S., Reynolds, R.C., 2012. Suma. Neuroimage 62, 768–773.
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.
Sevy, A.B., 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.
Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.
Sridhar, D., Stakhovskaya, O., Leake, P.A., 2006. A frequency-position function for the human cochlear spiral ganglion. Audiol. Neuro-Otol. 11 (Suppl 1), 16–20.
Themelis, G., Selb, J., Thaker, S., Stott, J.J., Custo, A., Boas, D.A., Franceschini, M.-A., 2004. Depth of Arterial Oscillation Resolved with NIRS Time and Frequency Domain. Optical Society of America, p. WF2.
Themelis, G., D'Arceuil, H., Diamond, S.G., Thaker, S., Huppert, T.J., Boas, D.A., Franceschini, M.A., 2007. Near-infrared spectroscopy measurement of the pulsatile component of cerebral blood flow and volume from arterial oscillations. J. Biomed. Optics 12, 014033.
Ward, B.D., 1998. Deconvolution Analysis of FMRI Time Series Data. NIMH NIH – AFNI documentation.
Wilcox, T., Bortfeld, H., Woods, R., Wruck, E., Boas, D.A., 2008. Hemodynamic response to featural changes in the occipital and inferior temporal cortex in infants: a preliminary methodological exploration. Dev Sci 11, 361–370.
Wilcox, T., Bortfeld, H., Woods, R., Wruck, E., Armstrong, J., Boas, D., 2009. Hemodynamic changes in the infant cortex during the processing of featural and spatiotemporal information. Neuropsychologia 47, 657–662.
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.
Yoshinaga-Itano, C., Sedey, A.L., Coulter, D.K., Mehl, A.L., 1998. Language of early- and later-identified children with hearing loss. Pediatrics 102, 1161–1171.
Zhang, H., Duan, L., Zhang, Y.-J., Lu, C.-M., Liu, H., Zhu, C.-Z., 2011. Test–retest assessment of independent component analysis-derived resting-state functional connectivity based on functional near-infrared spectroscopy. Neuroimage 55, 607–615.