PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Investigating the cortical representation of speech melody using near-infrared spectroscopy

Kurt STEINMETZGER[1]; Martin ANDERMANN[1]; Esther MEGBEL[1]; Zhengzheng SHEN[2]; Mark PRAETORIUS[2]; André RUPP[1]

[1] Section of Biomagnetism, Department of Neurology, Heidelberg University Hospital, Germany

[2] Section of Otology and Neuro-otology, ENT Clinic, Heidelberg University Hospital, Germany

ABSTRACT

Objective measures that reliably quantify the listening success with cochlear implants (CIs) are hardly available, since CIs are incompatible with several common neuroscientific methods. However, especially for pre-lingually implanted children, such measures are urgently needed to better monitor the development of their hearing. A promising method that does not interfere with the function of CIs is near-infrared spectroscopy (NIRS). We are assuming that cortical activation patterns which resemble those of normal-hearing control subjects indicate that an implantation has been successful. In a first step, we thus presented normal-hearing subjects with speech-like sounds and vowels that differed regarding their prosodic properties, to evaluate the precision and usability of NIRS. The stimuli either had a more or less clear pitch, or varied regarding their pitch contours. Same as additionally obtained magnetoencephalography (MEG) data, which allow for a clear spatial differentiation of the cortical areas activated by the different stimuli, the NIRS data show different spatial activation patterns, despite the lower spatial resolution of this method. The present results thus suggest that NIRS is suitable for investigating cortical responses to stimuli that vary regarding specific acoustic properties and can therefore be used for the neurophysiological assessment of CI-based hearing.

Keywords: human auditory cortex, speech melody, near-infrared spectroscopy

1. INTRODUCTION

To provide the basis for later studies investigating pitch perception in cochlear implant (CI) users, functional near-infrared spectroscopy (NIRS) data were obtained from normal-hearing participants listening to vowel sequences with different pitch contours. The current experiment thus aimed to investigate whether and how changes in the cortical representation of speech sounds with different prosodic contours become apparent in NIRS measurements. The theoretical motivation for this project is that the listening success of CI users generally varies widely across subjects and that, especially for pre-lingually implanted young children, objective measures that help to explain these performance differences and are urgently needed. NIRS is currently the most promising neuroscientific method to investigate CI-based hearing, as the data are unaffected by the electromagnetic signals of the CI processor (1, 2). A crucial limitation when listening through a CI is the restricted access to pitch information, which is largely confined to exploiting the temporal pitch cues transmitted by the signal envelopes (3). However, individual differences in the ability to use these pitch cues, as reflected by cortical activation patterns that resemble those of normal-hearing listeners to a greater or lesser degree, may serve as an objective marker of CI-based listening success.

NIRS measurements are based on the relatively low absorption of infrared light by biological tissue, resulting in a so-called optical window into the brain (4, 5). Sources optodes placed directly on the scalp emit infrared light towards the brain, while detector optodes positioned at scalp sites nearby record the amount of light that has passed the cortical area in between the two sensors. The more active a given cortical region is, the more light will be absorbed by it, since brain activity causes an inflow of oxygenated blood containing higher concentrations of red oxyhemoglobin (HbO). This increase in HbO concentration over time, and the concurrent smaller decrease of the concentration of

[1] kurt.steinmetzger@uni-heidelberg.de

deoxygenated hemoglobin (HbR) not considered in the current paper, can be measured using NIRS. One methodological difficulty that has received much attention in recent years (6) is that NIRS signals are strongly affected by superficial blood flow changes, which may obscure the cortical effects of interest. An increasingly more common way to separate these two signal components, which has also been used here, is the use of additional measurement channels with a short source detector spacing. Contrary to the longer standard channels, the light travelling from source to detector optode will not reach the cortex in this case and the recorded signals can thus be used to estimate and subtract the contribution of the superficial component from the data.

The stimuli used in this experiment were German vowels concatenated into continuous sequences. The pitch contour within each block was either the same for each individual vowel (*same pitch*) or varied between vowels (*different pitch*). Hence, sequences with the same pitch throughout sounded rather monotonous, whereas the ones with different pitch contours had a somewhat melody-like quality. Previous neuroimaging work has shown that while stimuli with fixed pitch contours mainly activate Heschl's gyrus and the planum temporale, melodies lead to additional activity slightly further anterior to those two structures (7). Furthermore, the neural activity evoked by transitions between different vowel types that had the same prosodic contours could be localised to the planum temporale (8). Consequently, we expected the vowel sequences with *different pitch* contours to activate a slightly larger cortical area than the *same pitch* equivalents. Since the findings reported in those two studies were based on methods with a higher spatial resolution (functional magnetic resonance imaging – fMRI – and MEG, respectively), the current experiment thus also served as a general test of whether NIRS is suitable for precise investigations of auditory perception.

2. METHODS

2.1 Participants

NIRS data obtained from 16 normal-hearing listeners (7 female, 9 male) were analysed. Their ages ranged from 21 to 56 years, with a mean of 27.3 years. All participants were native speakers of German and had audiometric thresholds of less than 20 dB hearing level (HL) at octave frequencies between 125 and 8000 Hz. All subjects gave written consent and the study was approved by the local research ethics committee (Medical Faculty, University of Heidelberg).

2.2 Stimuli

The stimulus materials used in this study were recordings of the German vowels /a/, /e/, /i/, /o/ and /u/ spoken by an adult male German talker. The recordings were made in an anechoic room and digitised with 24-bit resolution and a 48-kHz sampling rate, using a condenser microphone (Brüel & Kjær, type 4193) and an RME Babyface audio interface. Each vowel was cut at zero-crossings right before vowel onset, limited to a length of 800 ms using a 50-ms Hann-windowed offset ramp, and high-pass filtered at 50 Hz (zero-phase-shift third-order Butterworth).

Subsequently, the F0 contours of the vowels were manipulated with the STRAIGHT vocoder software (9) implemented in MATLAB, which allows to alter the prosodic properties of the stimulus materials without affecting their spectral filter. Each vowel was re-synthesised with two different mean F0s (80 Hz and 120 Hz) and five different prosodic contours (flat, rising straight, falling straight, rising curved and falling curved), resulting in a final set of 50 stimuli. For the non-flat contours, the F0 increased or decreased by a total of a perfect fifth relative to the mean F0. The mean F0 was always specified to be the mid-point of the non-flat contours, such that the maximum and minimum F0 values were equally far above and below the mean.

Finally, the stimuli were low-pass filtered (zero-phase-shift first-order Butterworth) at 3.5 kHz – to match the frequency response of the Etymotic Research ER3 headphones used in a corresponding MEG experiment – and normalised to a common root-mean-square (RMS) level. Examples of the stimuli are shown in Fig. 1.

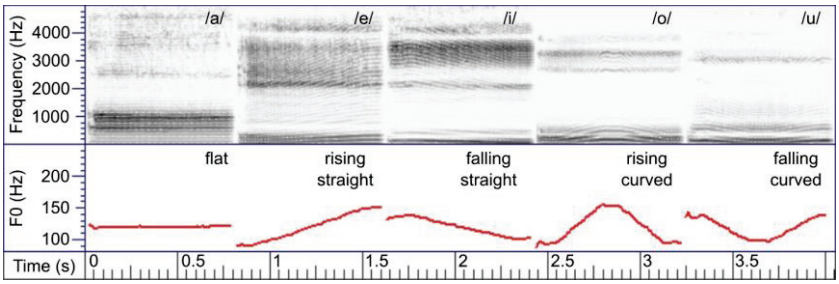

Figure 1 – Stimuli. Narrow-band spectrograms (upper panel) of the 5 German vowels used in the current experiment, each synthesised with one of the 5 different F0 contours (lower panel). The vowels were subsequently concatenated into blocks of 20 stimuli, in which the pitch contours were either the same throughout (same pitch) or varied between vowels (different pitch).

2.3 Procedure

As hemoglobin concentration changes evolve over the course of several seconds, a block design with long stimulation periods and equally long pauses was used to maximise the size of the experimental effects. The individual vowels were thus concatenated into blocks of 20 stimuli with a total duration of 16 s. These stimulus blocks were followed by pauses with random durations ranging from 16–20 s. The vowels were grouped into ten experimental conditions, according to their prosodic properties. The first five conditions (*same pitch*) each comprised only one of the five different prosodic contours. Hence, the prosodic contours could not vary between the individual vowels in these blocks. Five more conditions (*different pitch*), in which the contours varied within each block, were formed of vowels with rising, falling, straight, and curved contours, or a mixture of all five contour types. Each participant was presented with 10 blocks in each experimental condition. There was one block for every combination of vowel type and mean F0 in each condition, to ensure that they would occur equally often across conditions. The order of the blocks as well as the order of the contours within each block were both randomised without any constraints. The experiment thus consisted of 100 stimulus blocks framed by 101 pauses, amounting to a total duration of about 57 mins. For simplicity, the data in each of the five *same* and *different pitch* conditions were pooled together for analysis.

The experiment took place in a sound-attenuating and electrically shielded booth, with the participant sitting in a comfortable reclining chair during data acquisition. To ensure that the same experimental design can also be used with young children in the future, there was no behavioural task. Since the presentation of a silent movie was found to negatively affect the recorded data in pilot tests, the experiment also did not contain a visual distraction. Short pauses were inserted roughly every 10 mins to ensure the vigilance of the subjects. The stimuli were converted with 24-bit resolution and a sampling rate of 48 kHz using an RME ADI-8 DS sound card and presented via Etymotic Research ER2 headphones attached to a Tucker-Davis HB7 headphone buffer. The presentation level was set to 70 dB SPL, using an artificial ear (Brüel & Kjær, type 4157) connected to a Brüel & Kjær measurement amplifier (type 2610). Simultaneously recorded 64-channel EEG data are available for each subject but were not considered in the current paper.

2.4 NIRS recording and analysis

NIRS signals were recorded with a continuous-wave NIRx NIRScout 16x16 system at a sampling rate of 7.8125 Hz. Eight light sources and eight photodetector optodes were symmetrically placed over both hemispheres using a standard EEG cap. The optode positions for each individual subject were digitised with a Polhemus 3SPACE ISOTRACKII system before the experiment. The source optodes emitted light pulses with wavelengths of 760 and 850 nanometres. The same sequential illumination pattern was simultaneously used for both hemispheres to avoid interference between adjacent sources. The chosen optode layout was devised to optimally cover the auditory cortex and resulted in 22 measurement channels per hemisphere, 20 of which had a standard source-to-detector distance of 30 mm, while the other 2 had a shorter 15-mm spacing (Fig. 2).

The data were pre-processed using the HomER2 toolbox, version 2.8 (10). The raw light intensity signals were first converted to optical density values and then corrected for motion artefacts. A kurtosis-based wavelet algorithm with a threshold value of 3.3 (11) was used to identify and correct motion artefacts by rejecting spectral components of the signal rather than time segments. Measurement channels with poor signal quality were then excluded from further analysis based on their scalp coupling index (12). This processing step was carried out by band-pass filtering the optical density signals of both wavelengths between 0.5–2.5 Hz (third-order low-pass and fifth-order high-pass zero-phase-shift Butterworth filters) to emphasise the heart-beat related signal fluctuations and correlating the two signals. Channels with values below 0.5 were excluded, since a good contact between optode and scalp will result in high correlations. On average, 1.6 channels per subject were excluded (26 in total, max. 10 per subject). Next, the motion-corrected signals of the remaining channels were band-pass filtered between 0.01–0.2 Hz (same filter settings as above) to isolate the task-related neural activity and converted to HbO concentration values based on the modified Beer-Lambert law (4). The differential path length factors required for the conversion were determined based on the wavelength and the age of the subject (13). The pre-processed waveforms in each experimental condition were then block averaged from -2–32 s around stimulus onset and baseline corrected by subtracting the mean of the signal in the pre-stimulus window from each sample point.

Secondly, the pre-processed data were statistically evaluated and visualised with SPM-fNIRS, version r3 (14). Based on the principles of the general linear model (GLM), the SPM framework (15) tests how well the measured data can be predicted with an expected hemodynamic response function (HRF, Fig. 4). Within SPM-fNIRS, the optode positions of each subject were first probabilistically rendered onto a standard brain surface in the MNI coordinate system (16). The signals were then temporally smoothed using the shape of the expected HRF (‘pre-colouring’) to avoid autocorrelations issues when estimating the model (17). The data of each individual subject were statistically modelled by convolving the continuous HbO signals with separate regressors for each of the ten experimental conditions. The expected HRFs were based on the standard canonical HRF implemented in SPM and the stimulus duration was set to the length of the blocks (16 s). To allow the shape of the measured HRFs to vary slightly, the temporal and spatial derivatives of the canonical HRF were included in the GLM too (18). Furthermore, the first component of a principal component analysis of the signals recorded with the four short channels was used as an additional nuisance regressor, as this serves to estimate and remove the so-called global scalp-hemodynamic component (6), i.e. the superficial signal component. After estimating the GLM for each subject, contrast vectors were defined to evaluate the effects of the main regressors in each experimental condition. The additional regressors representing the derivatives and the global scalp component were set to zero in all contrasts to statistically control for their effects. Group-level statistics were computed by testing whether a main regressor, or combinations thereof, had beta weights that were significantly different from zero using one-sample t-tests. A customised version of the SPM-fNIRS plotting routine was devised to visualise the optode and channel positions (Fig. 2) as well as the functional activations (Fig. 3).

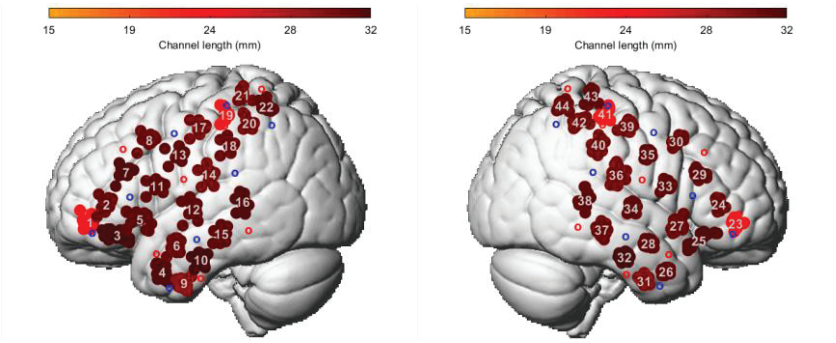

Figure 2 – NIRS optode and channel positions. The average positions of the source and detector optodes are indicated by the red and blue circles, respectively. The average positions of the resulting measurement channels are indicated by the grey numbers. The locations of the red dots show the channel positions for each individual subject and the hue of the dots represents the distances between the sources and detectors.

3. RESULTS

The cortical activation patterns in response to vowel sequences with different prosodic properties are summarised in Fig. 3. The left column of the figure shows the cortical activity that was evoked by blocks consisting of vowels that had the same pitch contours throughout. A significant increase of neural activity was observed near the posterior part of the right auditory cortex (channel 36), while there was no significant activation of the left auditory cortex. In contrast, vowel sequences with different pitch contours within the individual stimulus blocks evoked activity in the right auditory cortex that was both stronger and had a wider spatial distribution (channels 34, 36, and 38; Fig. 3, middle column). However, there was again no significant activation of the left auditory cortex. Accordingly, when comparing the two stimulus classes directly, a cluster of channels near the right auditory cortex exhibited a stronger activation in response to vowels sequences with different pitch contours (Fig. 3, right column). This channel cluster (numbers 25, 27, 34, 36 and 38) was distributed along the right superior temporal gyrus. For the left auditory cortex, on the other hand, no significant differences were observed.

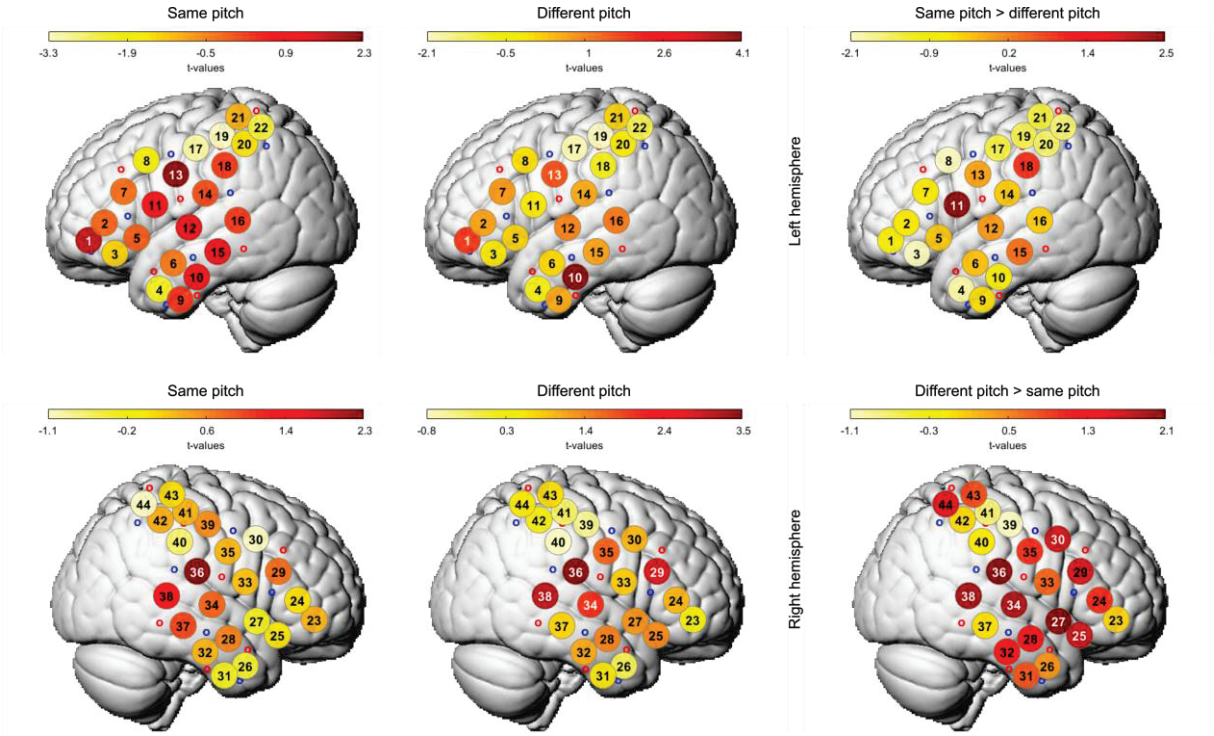

Figure 3 – Neural activation patterns in response to vowel sequences with the same (left column) or different pitch contours (middle column), as well as the differences between both stimulus classes (right column). NIRS channels showing significant positive activations ($p < .05$) are indicated by white numbers.

As shown in the lower part of Fig. 3, vowel sequences with same and different pitch contours, as well as the comparison of these two stimulus classes, all resulted in significant effects at channel 36. To further investigate this finding, Fig. 4 shows the average time courses of the HbO concentration changes at this location for the two stimulus conditions. Both hemodynamic response functions were found to reach their maximum around 5 s after stimulus onset. However, the HbO concentration level for sequences with the same pitch quickly returned to baseline level after this initial peak, while the HRF for vowels with different pitch contours showed a sustained increase throughout the stimulation period. In line with this observation, the predicted HRF that was used to statistically model these data resembles the measured HRF of the vowel sequences with different pitch to a much greater degree, after weighting it with the subject-specific regressions weights.

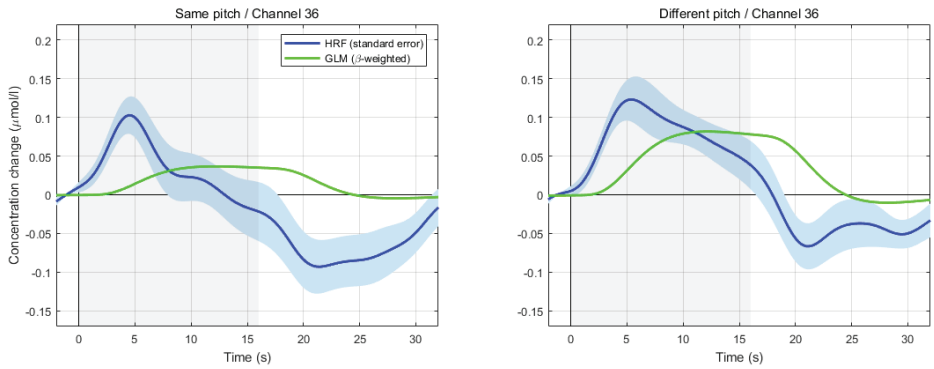

Figure 4 – Hemodynamic response functions. Measured (blue) and predicted (green) concentration changes of oxyhemoglobin (HbO) evoked by vowel sequences with same and different pitch contours. The 16-s stimulation period is indicated by the grey shading and the predicted response functions are shown after multiplication with the estimated regression weights.

4. DISCUSSION

In the present NIRS experiment, right-lateralised activations of the auditory cortex were found in response to vowel sequences with same or different pitch contours. The observed activations showed a wider spatial distribution as well as a longer duration for sequences with different pitch contours.

With respect to the spatial distribution, the results are in line with earlier findings reporting additional neural activity anterior to primary auditory cortex for stimulus sequences with melodic rather than monotonous pitch contours (7). However, although this fMRI study also found strongly right-lateralised activity when comparing these two stimulus types, bilateral activity in auditory cortex was observed in response to stimuli with the same pitch. The lateralised activation pattern obtained in the current study has, however, also been found in several MEG studies using simple tonal stimuli and passive experimental paradigms. One explanation for this may be that the increased cortical folding of the left auditory cortex results in MEG signal cancellations (19). For NIRS, in turn, the increased folding might have the consequence that this area is located slightly further away from scalp and cannot be fully reached by the infrared light. The penetration depth of NIRS is assumed to be only about half the source detector distance (5) – i.e. ~15 mm for the long channels – and hence even small structural difference across hemispheres could greatly affect the results.

The time courses of the HbO concentration changes revealed a more sustained effect for vowel sequences with different pitch contours, as opposed to a shorter response following stimulus onset for sequences with the same pitch. This finding agrees with results showing that the hemodynamic response measured with fMRI also decreases after an initial peak when stimuli with a consistent pitch are used (20). Contrary to these fMRI data, there was however no indication of an offset response in the present data, despite the markedly higher temporal resolution of NIRS signals.

5. CONCLUSIONS

The current results have demonstrated, firstly, that NIRS has a sufficiently high spatial resolution to distinguish between speech sounds with different prosodic properties. Secondly, the time courses of the hemodynamic responses recorded with NIRS also showed clear differences depending on the stimulus type. Taken together, this suggests that NIRS is a suitable method for precise neurophysiological investigations of auditory perception, in addition to its many practical advantages.

ACKNOWLEDGEMENTS

We are grateful to the Dietmar Hopp Stiftung (Grant No. 2301 1239) and MED-EL for supporting our research for a period of three years. We would also like to thank Helmut Riedel for technical support and Alexander Gutschalk for helpful advice regarding the interpretation of the results.

REFERENCES

1. Sevy AB, Bortfeld H, Huppert TJ, Beauchamp MS, Tonini RE, Oghalai JS. Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. Hearing Res. 2010;270(1):39–47.
2. Anderson CA, Wiggins IM, Kitterick PT, Hartley DE. Adaptive benefit of cross-modal plasticity following cochlear implantation in deaf adults. Proc Natl Acad Sci USA. 2017;114(38):10256–61.
3. Steinmetzger K, Rosen S. The role of envelope periodicity in the perception of masked speech with simulated and real cochlear implants. J Acoust Soc Am. 2018;144(2):885–96.
4. Scholkmann F, Kleiser S, Metz AJ, Zimmermann R, Pavia JM, Wolf U, et al. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. Neuroimage. 2014;85:6–27.
5. Pinti P, Tachtsidis I, Hamilton A, Hirsch J, Aichelburg C, Gilbert S, et al. The present and future use of functional near‐infrared spectroscopy (fNIRS) for cognitive neuroscience. Ann N Y Acad Sci. 2018.
6. Sato T, Nambu I, Takeda K, Aihara T, Yamashita O, Isogaya Y, et al. Reduction of global interference of scalp-hemodynamics in functional near-infrared spectroscopy using short distance probes. Neuroimage. 2016;141:120–32.
7. Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. Neuron. 2002;36:767–76.
8. Andermann M, Patterson RD, Vogt C, Winterstetter L, Rupp A. Neuromagnetic correlates of voice pitch, vowel type, and speaker size in auditory cortex. Neuroimage. 2017;158:79–89.
9. Kawahara H, Irino T. Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In: Divenyi P, editor. Speech separation by humans and machines. Boston, MA: Springer; 2005. p. 167–80.
10. Huppert TJ, Diamond SG, Franceschini MA, Boas DA. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl Opt. 2009;48(10):D280–D98.
11. Chiarelli AM, Maclin EL, Fabiani M, Gratton G. A kurtosis-based wavelet algorithm for motion artifact correction of fNIRS data. Neuroimage. 2015;112:128–37.
12. Pollonini L, Olds C, Abaya H, Bortfeld H, Beauchamp MS, Oghalai JS. Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. Hearing Res. 2014;309:84–93.
13. Scholkmann F, Wolf M. General equation for the differential pathlength factor of the frontal human head depending on wavelength and age. J Biomed Opt. 2013;18(10):105004.
14. Tak S, Uga M, Flandin G, Dan I, Penny W. Sensor space group analysis for fNIRS data. J Neurosci Methods. 2016;264:103–12.
15. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp. 1994;2(4):189–210.
16. Singh AK, Okamoto M, Dan H, Jurcak V, Dan I. Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. Neuroimage. 2005;27(4):842–51.
17. Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited—again. Neuroimage. 1995;2(3):173–81.
18. Plichta MM, Heinzel S, Ehlis A-C, Pauli P, Fallgatter AJ. Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: a parametric validation study. Neuroimage. 2007;35(2):625–34.
19. Shaw ME, Hämäläinen MS, Gutschalk A. How anatomical asymmetry of human auditory cortex can lead to a rightward bias in auditory evoked fields. Neuroimage. 2013;74:22–9.
20. Steinmann I, Gutschalk A. Sustained BOLD and theta activity in auditory cortex are related to slow stimulus fluctuations rather than to pitch. J Neurophysiol. 2012;107(12):3458–67.