Hearing Research 420 (2022) 108483

Contents lists available at ScienceDirect

Hearing Research

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

# Cortical activity evoked by voice pitch changes: A combined fNIRS and EEG study

Kurt Steinmetzger[a,*], Esther Megbel[a], Zhengzheng Shen[b], Martin Andermann[a], André Rupp[a]

[a] Section of Biomagnetism, Department of Neurology, Heidelberg University Hospital, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany
[b] Section of Otology and Neurootology, ENT Clinic, Heidelberg University Hospital, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany

ARTICLE INFO

*Article history:*
Received 8 June 2021
Revised 2 March 2022
Accepted 10 March 2022
Available online 11 March 2022

*Keywords:*
Speech
Hearing
Pitch perception
Event-related potentials
ERP source reconstructions
Adaptation

ABSTRACT

Functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG) data were simultaneously obtained from normal-hearing listeners presented with continuous natural vowel sequences to study the interrelation of the haemodynamic and electrophysiological cortical responses evoked by voice pitch changes. fNIRS topographies and distributed ERP source reconstructions both indicated additional activity in the right superior temporal cortex if the prosodic contours varied between successive vowels, rather than being the same throughout the sequences. The source-level ERPs furthermore revealed two temporally and spatially separable adaptation processes in superior temporal cortex: Firstly, the early P1 component was bilaterally attenuated when vowels with the same prosodic contours were presented repeatedly, reflecting sensory adaptation. Secondly, the later P2 and sustained potential components were smaller in the right hemisphere during sequences without prosodic changes, which is taken to represent an attention-based adaptation effect. The present results demonstrate the convergence of both methods and demonstrate which ERPs underlie the right-lateralised activity in superior temporal cortex evoked in response to pitch changes that has been observed in many neuroimaging studies.



## 1. Introduction

Pitch changes in auditory signals transmit crucial information, such as prosodic variations in speech and melody in music. Pitch information also promotes the formation of separate auditory streams (Oxenham, 2008), which is important for understanding speech in noise (Cullington et al., 2008; Steinmetzger et al., 2015). Consequently, the restricted access to pitch information is a critical limitation when listening through a cochlear implant (CI; Macherey et al., 2014; Oxenham, 2008; Steinmetzger et al., 2018).

Neuroimaging and lesion studies have shown that secondary auditory regions in the right hemisphere are critically involved in the processing of pitch changes (Johnsrude et al., 2000; Meyer et al., 2002; Patterson et al., 2002; Zatorre et al., 2001). For example, a functional magnetic resonance imaging (fMRI) study by Patterson et al. (2002) showed that while tone sequences with a fixed pitch mainly evoke activity in primary auditory cortex, pitch changes elicit additional activity anterior to it. Specifically, stronger activity in response to pitch changes was observed in the right planum polare and superior temporal sulcus (STS), both located in the superior temporal cortex (STC). In agreement with these findings, fMRI evidence demonstrated that vowels with a fixed pitch elicit prominent bilateral activity in superior temporal areas (Uppenkamp et al., 2006), while vowels with prosodic modulations evoke additional activity in right superior temporal areas (Obleser et al., 2006). Accordingly, theoretical models of auditory cortical processing (Poeppel, 2003; Zatorre et al., 2002) posit that pitch changes and slow spectral modulations should engage right superior temporal areas, in particular their anterior part.

In contrast, the electrophysiological activity evoked by the processing of pitch changes in the right STC is poorly understood. In a recent magnetoencephalography (MEG) study, Andermann et al. (2021) found that the transient auditory responses (i.e., P1, N1, and P2) were strongly reduced in response to stimulus sequences with a fixed pitch compared to melodic sequences with discrete, stepwise pitch changes. These results are in line with earlier MEG data (Patterson et al., 2016; Rupp et al., 2005) showing reduced N1 and P2 amplitudes in response to melodies compared to tone sequences without pitch changes. Furthermore, the dipole sources in Rupp et al. (2005) indicated that the N1 and P2 evoked by pitch changes might originate from a slightly more anterior location than pitch-onset responses, and that they were lateralised to the right hemisphere. However, these

* Corresponding author.
*E-mail address:* kurt.steinmetzger@uni-heidelberg.de (K. Steinmetzger).

https://doi.org/10.1016/j.heares.2022.108483
0378-5955/

dipole-based MEG studies were concerned with activity generated in primary auditory cortex only and it is also unclear whether these findings similarly apply to continuous pitch glides typically found in natural speech.

In the current study, the cortical processing of pitch changes in sequences of natural vowels was investigated using simultaneously recorded functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG) data. In contrast to the literature referred to above, the combination of both methods enabled the concurrent assessment of the haemodynamic and electrophysiological cortical activity evoked by the stimuli. Focussing on the event-related potentials (ERPs) ensured the comparability of the results obtained with both methods, as the synchronised postsynaptic activity underlying the ERPs has been shown to underlie the blood-oxygen-level-dependent (BOLD) responses measured using fNIRS (Logothetis, 2003; Logothetis et al., 2001). By computing distributed source reconstructions of the ERPs, the combination of both methods also allowed to assess the similarity of the resulting cortical activity maps. This approach was intended to determine whether the cortical activity evoked by pitch changes can be accurately detected using fNIRS, and thereby validate the suitability of this method.

The materials in the present study were a set of natural German vowels synthesised with a range of prosodic contours. These vowels were concatenated into continuous sequences in which the pitch contours were either the same for each individual vowel (FIXED PROSODY) or varied between vowels (VARIABLE PROSODY), while any other acoustic differences were controlled for. Hence, there were gliding pitch changes in both conditions, but they were repetitive in the FIXED PROSODY condition, while they varied across vowels in the VARIABLE PROSODY condition. Both conditions were expected to elicit right-lateralised activity in the STC, but the non-repetitive pitch contours in the VARIABLE PROSODY condition were assumed to evoke additional activity in this area, particularly its anterior portion. As the pitch contours of the vowels had durations of several hundred ms, both the transient ERP components and the subsequent sustained potential in the right STC were hypothesised to have larger amplitudes in the VARIABLE PROSODY condition.

The current study is furthermore intended to lay the foundation for subsequent work with CI users. By testing a sample of young normal-hearing listeners, we aimed to obtain a standard model of how these stimuli are processed. In contrast to EEG data, which are contaminated by electrical artefacts when testing CI users, fNIRS is ideally suited to study CI-based hearing, as the devices do not interfere with the measurements. To ensure that the experimental paradigm is also suitable for testing infant CI users, there was no behavioural task. The differences between the prosodic contours used here were designed to be readily apparent with normal hearing. However, as they primarily have to rely on temporal rather than spectral pitch cues, this linguistically relevant contrast is more difficult to perceive for CI users and individual difference are large (Chatterjee et al., 2008; Everhardt et al., 2020; Steinmetzger et al., 2018). It is thus assumed that if differences between the FIXED and VARIABLE PROSODY conditions are evident at the cortical level, this should go along with better CI-based performance and higher speech intelligibility scores. Hence, the current paradigm might potentially be useful for an objective evaluation of CI-based hearing, particularly in young children where reliable behavioural measures are hard to obtain.

## 2. METHODS

### 2.1. Participants

Twenty subjects (10 females, 10 males; mean age 26.2 years, standard deviation 8 years) participated in this experiment. All subjects were German native speakers and had audiometric thresholds of less than 20 dB HL (hearing level) at octave frequencies between 125 and 8000 Hz. All participants reported that they are right-handed, have no history of neurological or psychiatric illnesses, and were not taking any medications at the time of testing. Written consent was obtained prior to the experiment. The study was approved by the local research ethics committee (Medical Faculty, University of Heidelberg) and was conducted in accordance with the Declaration of Helsinki.

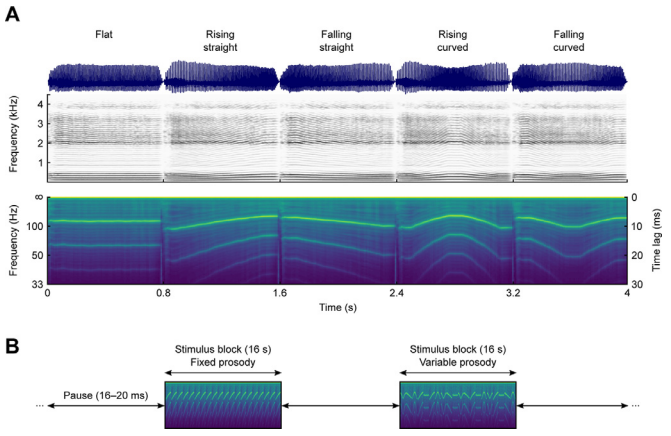


**Fig. 1.** Example stimuli and experimental design. A) Waveforms, narrow-band spectrograms, and summary autocorrelation function spectrograms of the vowel /e/ processed to have five different prosodic contours. B) The individual vowels were concatenated into blocks of 20 vowels lasting 16 s and alternated with pauses lasting 16–20 s. The prosodic contours within these blocks were either the same throughout (FIXED PROSODY) or varied between the individual vowels (VARIABLE PROSODY).

### 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, Nærum, Denmark) and an RME Babyface audio interface (Haimhausen, Germany). 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 3$^{rd}$-order Butterworth).

Subsequently, the original *F*0 contours of the vowels were manipulated with the STRAIGHT vocoder software (Kawahara et al., 2005) implemented in MATLAB (MathWorks, Natick, MA, USA), which allows to alter the prosodic properties of the stimulus materials without affecting their spectral envelope. Each of the five vowels was re-synthesised with two different mean fundamental frequencies (*F*0s; 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 *F*0 increased or decreased by a perfect fifth relative to the mean *F*0. The mean *F*0 was always specified to be the mid-point of the non-flat contours, such that the maximum and minimum *F*0 values were equally far above and below the mean. Finally, the stimuli were low-pass filtered (zero-phase-shift 1$^{st}$-order Butterworth) at 3.5 kHz – to match the frequency response of the Etymotic Research ER3 headphones (Elk Grove Village, IL, USA) used in a corresponding MEG experiment – and normalised to a common root-mean-square level.

Examples of the stimuli are shown in Fig. 1A. The narrow-band spectrograms depicted in the upper half of the figure demonstrate that the spectrum of the example vowel (/e/, mean *F*0 = 120 Hz)

is indeed very similar across the five prosodic versions, while the waveforms differ markedly. To visualise the different prosodic contours, the lower half of the plot shows spectrographic representations of summary autocorrelation functions (SACFs; Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). The individual autocorrelation functions were calculated for the low-pass filtered ($2^{nd}$-order Butterworth, cut-off 1 kHz) outputs of 22 gammatone filters with centre frequencies ranging from 0.2–4 kHz and summed together into SACFs. This procedure was applied with a step size of 1 ms and a Hann-window size of 5 ms to yield spectrographic representations of the SACFs across the duration of the stimuli. The time lag of the first peak in the SACF spectrograms represents the *F*0 of the stimuli.

### 2.3. Experimental design and procedure

To maximise the BOLD responses, a block design with long stimulation periods and equally long pauses was used (Fig. 1B). The individual vowels were thus concatenated into blocks of 20 stimuli with a total duration of 16 s, which were followed by pauses with random durations ranging from 16–20 s (Henson, 2007). Furthermore, the silent gaps between the individual vowels in each block were minimised to ensure that the ERPs were reflecting pitch-change responses rather than energy-onset responses (Krumbholz et al., 2003).

The experiment comprised two conditions: In the FIXED PROSODY condition, all vowels within a block had the same prosodic contour (*flat, rising straight, falling straight, rising curved*, or *falling curved*). In the VARIABLE PROSODY condition, the contours varied between the vowels within each block (*rising, falling, straight, curved*, or *a mixture of all five contour types*). Each participant was presented with 50 blocks in each condition, i.e., 10 blocks of each contour type or combination, respectively. In both conditions, the same vowel with the same mean *F*0 was presented throughout a given block and hence there was one block for each of the ten combinations of vowel type and mean *F*0. The order of the blocks as well as the order of the prosodic contours within each block of the VARIABLE PROSODY condition 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. As the EEG data were analysed relative to the onset of the individual vowels in each block, this design resulted in 2000 trials (2 conditions x 50 blocks x 20 vowels). The range of prosodic contours, vowels, and *F*0s was included to ensure the generalisability of the results and to make the experiment less monotonous. Although these factors could in principle be analysed too, the current paper solely focussed on the differences between the FIXED and VARIABLE PROSODY conditions.

The experiment took place in a sound-attenuating and electrically shielded room, with the participant sitting in a comfortable reclining chair during data acquisition. Throughout the experiment, the light was dimmed to the lowest level to minimise ambient light from interfering with the fNIRS recordings. There was no behavioural task, but pauses were inserted about every 15 mins to ensure the vigilance of the subjects. Although this procedure enabled no control of the attentional state of the subjects, it provides the advantage that it can be used in infant studies without modification. The stimuli were converted with 24-bit resolution at a sampling rate of 48 kHz using an RME ADI-8 DS sound card and presented over Etymotic Research ER2 earphones attached to a Tucker-Davis Technologies HB7 headphone buffer (Alachua, FL, USA). The presentation level was set to 70 dB SPL, using an artificial ear (Brüel & Kjær, type 4157) connected to a corresponding measurement amplifier (Brüel & Kjær, type 2610).

The fNIRS and EEG methodology described in the next two sections largely resembles that of a previous study (Steinmetzger et al., 2020). Stimulus materials, the fNIRS and EEG data, as well as the code used to process these data are all available at: https://osf.io/ps47b.

### 2.4. fNIRS recording and analysis

fNIRS signals were recorded with a continuous-wave NIRScout 16 × 16 system (NIRx Medizintechnik, Berlin, Germany) at a sampling rate of 7.8125 Hz. Eight source optodes and eight detector optodes were placed symmetrically over each hemisphere by mounting them on an EEG cap (EasyCap, Herrsching, Germany). The source optodes emitted infrared light with wavelengths of 760 and 850 nm. To avoid interference between adjacent sources, only a single source optode per hemisphere was illuminated at a given time. The chosen optode layout was devised to optimally cover the auditory cortex and associated areas, resulting in 22 measurement channels per hemisphere, of which 20 had a standard source-to-detector distance (∼30 mm), while the remaining 2 had a shorter spacing (∼15 mm). Note that the optimal short-channel spacing of 8.4 mm (Brigadoi et al., 2015) could not be realised with the current fNIRS system and the signals obtained via the short channels therefore contained a small proportion of cortical activity. The optode and reference positions for each individual subject were digitised with a Polhemus 3SPACE ISOTRAK II system (Colchester, VT, USA) before the experiment.

The data were pre-processed using the HOMER2 toolbox (version 2.8; Huppert et al., 2009) and custom MATLAB code. 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 (Chiarelli et al., 2015) 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 (SCI; Pollonini et al., 2014). SCIs were computed by filtering the optical density signals of both wavelengths between 0.5–2.5 Hz ($3^{rd}$-order low-pass and $5^{th}$-order high-pass Butterworth filters, applied forwards and backwards), to emphasise the heart-beat related signal fluctuations, and correlating the filtered signals. Channels with correlation coefficients below 0.5 were excluded (mean = 1.7 channels/subject, max. 10 per subject, no excluded short channels), as this indicates a poor contact between optodes and scalp. As we did not pre-select the subjects, the sample included several participants with long, dark hair. Thus, a lower SCI threshold than the one suggested by Pollonini and colleagues (0.75) was used to limit data loss. Next, the motion-corrected signals of the remaining channels were band-pass filtered between 0.01–0.5 Hz (same filter types as for the SCI above), to isolate the task-related neural activity, and subsequently converted to concentration values based on the modified Beer-Lambert law (Scholkmann et al., 2014). The differential path length factors required for the conversion were determined based on the wavelengths of the light and the age of the subject (Scholkmann et al., 2013).

Secondly, the pre-processed data were statistically evaluated and topographically visualised with SPM-fNIRS (version r3; Tak et al., 2016). Based on the principles of the general linear model (GLM), the SPM framework tests how closely the stimulus-evoked signal changes resemble a canonical haemodynamic response function (HRF). In SPM-fNIRS, the optode positions of each subject were first transformed from subject space to MNI space, after which they were probabilistically rendered onto the ICBM-152 cortical template surface. This was achieved by applying a least-squares approach in which a set of digitised reference points distributed across the whole head (4 external fiducials and 17 positions of the extended international 10-20 system) was aligned to template values (Singh et al., 2005).

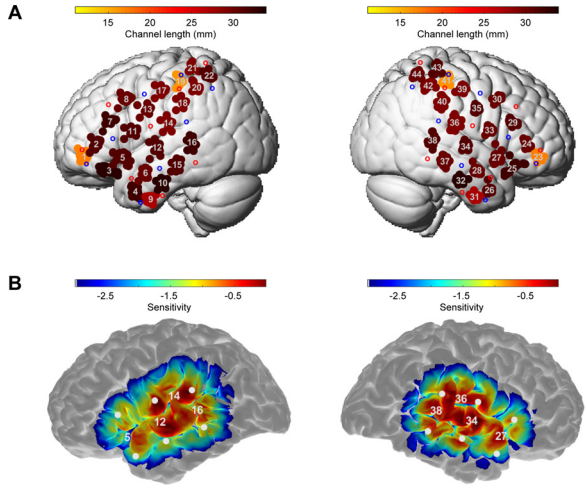


**Fig. 2.** fNIRS measurement layout. A) Group-averaged positions of the source and detector optodes (red and blue circles, respectively) and the resulting measurement channels (grey numbers), rendered onto the ICBM-152 cortical template surface. The point cloud around each channel number indicates the positions of the individual subjects and the colour of the cloud corresponds to the mean distance between the respective source-detector pair. B) Sensitivity profiles of the measurement channels forming the two auditory regions of interest (ROIs) covering the superior temporal cortices, based on the digitisation of the optode positions and a structural MRI scan of a representative subject.

The signals were then temporally smoothed using the shape of the canonical HRF waveform ('pre-colouring') to avoid autocorrelation issues when estimating the model (Worsley et al., 1995). The data of the individual subjects were statistically modelled by convolving the continuous signals obtained from each long channel with separate regressor functions representing the FIXED and VARIABLE PROSODY conditions. The standard SPM double-gamma function was used as canonical HRF and convolved with 16-s boxcar functions following the onsets of the stimulus blocks. The HbO data were modelled with positive HRFs, while the concentration changes were assumed to be negative for the HbR analysis. To allow the time course of the measured concentration changes to vary slightly, the temporal and spatial derivatives of the canonical HRF were included as additional regressors (Plichta et al., 2007). Furthermore, the first component of a principal component analysis of the pre-processed signals of the four short channels was used as an additional nuisance regressor. This procedure serves to estimate and remove the so-called global scalp-haemodynamic component (Sato et al., 2016), i.e., the superficial signal component.

After estimating the HbO and HbR GLMs for each subject, contrast vectors were defined to assess the functional activations. To evaluate the activity in the FIXED or VARIABLE PROSODY conditions, the respective regressors were set to 1, while all other regressors including the derivatives and the global scalp component were set to 0 to statistically control for their effects. When comparing the two conditions, the regressor representing the VARIABLE PROSODY condition was set to 1, while the regressor of the FIXED PROSODY condition was set to -1, as we expected stronger activity in the former condition. As before, all other regressors were set to 0 to control for their effects.

A customised version of the SPM-fNIRS plotting routine was devised to topographically visualise the measurement layout as well as the functional activations. The optode locations as well as the resulting channel positions and distances are shown in Fig. 2A. The analyses presented in this paper focussed on responses recorded from the auditory cortices in both hemispheres. Therefore, two auditory regions of interest (ROIs) were defined a priori, each comprising four channels positioned along the superior temporal cortices (left ROI: ch# 5, 12, 14, and 16; right ROI: ch# 27, 34, 36, and 38). To verify the coverage of the auditory cortices, the sensitivity profiles of these channels were computed using AtlasViewer (version 2.2, default settings; Aasted et al., 2015) and are provided in Fig. 2B. In the plots showing the functional activations (Fig. 3A & 4A), the short channels were omitted as they are assumed to not reflect any cortical responses. The waveforms in the HRF plots (Fig. 3B & 4B), were averaged from -2–32 s around block onset and baseline corrected by subtracting the mean amplitude in the pre-stimulus window from each sample point. The HRFs are shown both after the pre-processing ('Total HRF') and again after regressing out the contribution of the short channels and

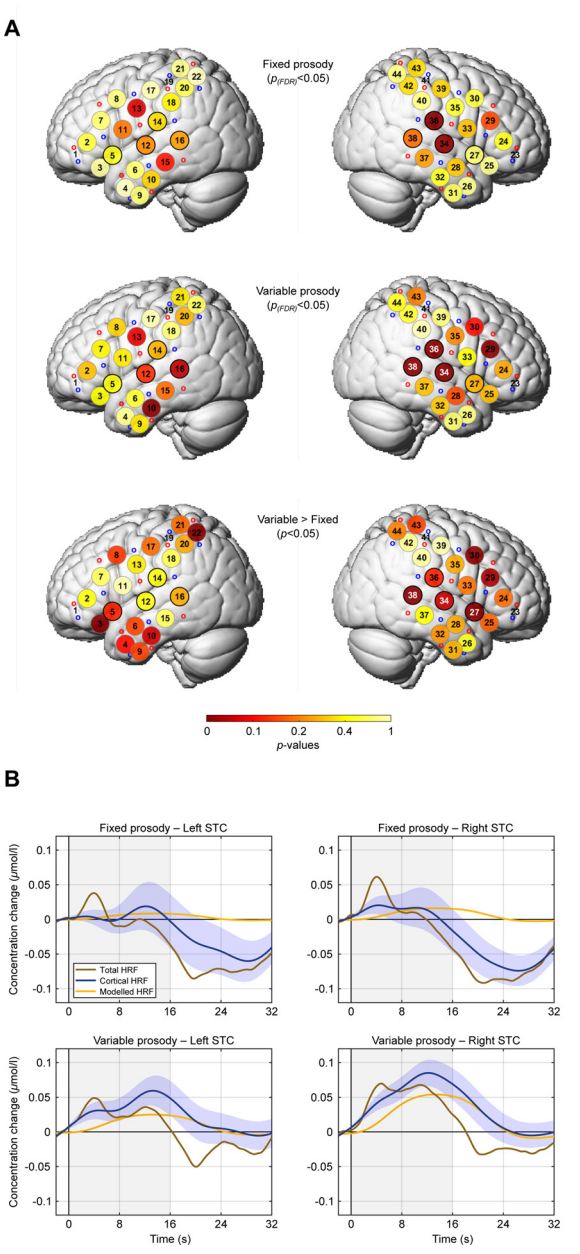


**Fig. 3.** fNIRS oxygenated haemoglobin (HbO) results. A) Group-level topographies. The upper two rows show the activation maps for the FIXED and VARIABLE PROSODY conditions, respectively. The lower row shows where the activity in the VARIABLE PROSODY condition exceeded that in the FIXED PROSODY condition. Channels within the two auditory regions of interest (ROIs) are outlined by black circles and white channel numbers indicate significant ROI channels. B) Group-level haemodynamic response functions (HRFs) for both conditions and ROIs. The shading around the cortical HRFs indicates the standard error of the mean.

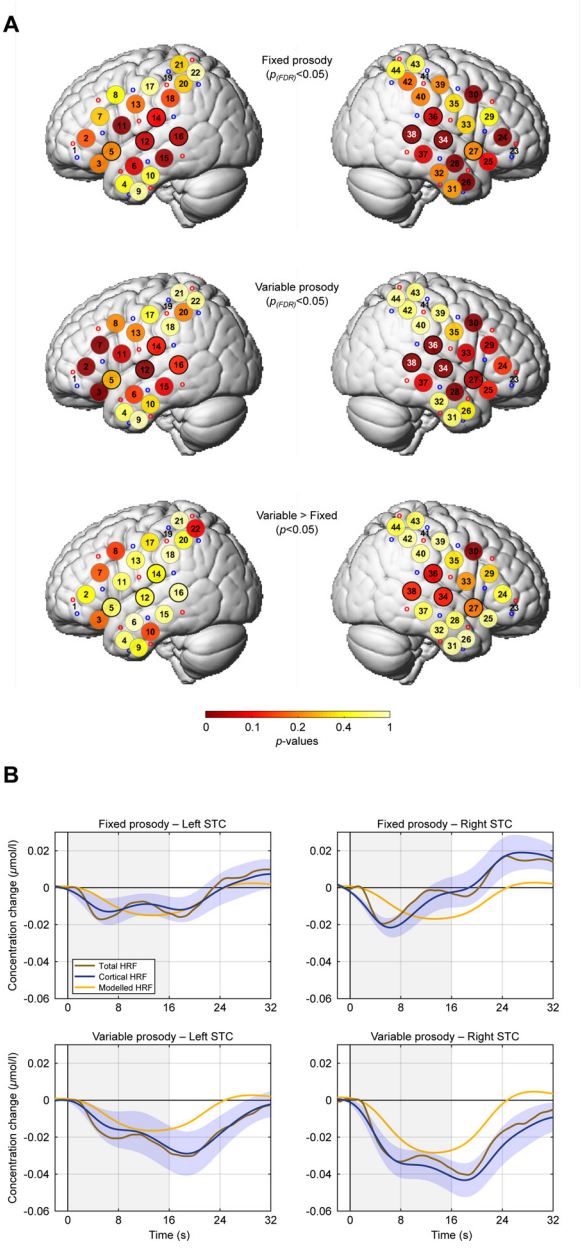


**Fig. 4.** fNIRS de-oxygenated haemoglobin (HbR) results. A) Group-level topographies. B) Group-level HRFs. All plotting details apart from the scaling are the same as for the HbO results shown in Fig. 3.

pre-colouring the signals ('Cortical HRF'), to illustrate the effect of removing the non-cortical signal component. Additionally, the $\beta$-weighted canonical HRFs ('Model HRF') are shown to demonstrate how well the cortical HRFs can be explained by the HRF models.

### 2.5. EEG recording and analysis

Continuous EEG signals were recorded using a 64-channel BrainVision actiCHamp system (Brain Products, Gilching, Germany). CI-compatible custom EEG caps with holes for the transmitter coils at electrode positions P7 and P8 were used. Apart from this omission, the 60 scalp electrodes were arranged according to the extended international 10-20 system. Four additional electrodes were placed around the eyes to record vertical and horizontal eye movements. The EEG data were recorded with an initial sampling rate of 500 Hz, an online anti-aliasing low-pass filter with a cut-off frequency of 140 Hz, and were referenced to the right mastoid. As with the fNIRS optodes, the electrode positions of each subject were digitized with a Polhemus 3SPACE ISOTRAK II system before the experiment.

The data were pre-processed offline using FieldTrip (version 20180924; Oostenveld et al., 2011) and custom MATLAB code. The continuous waveforms were first segmented into epochs ranging from -0.3–0.9 s relative to vowel onset. Next, linear trends and the DC component were removed by subtracting a 1$^{st}$-order polynomial and the epochs were low-pass filtered at 15 Hz (4$^{th}$-order Butterworth, applied forwards and backwards). The low cut-off frequency was chosen to improve the signal-to-noise ratio of the data and thus enable reliable source localisations. Very similar event-related potentials (ERPs) were obtained with a 30-Hz cut-off (Fig. S1). The epochs were then re-referenced to the mean of both mastoids and down-sampled to 250 Hz. After visually identifying and excluding bad channels (total = 9, max. 3 per subject), the data were decomposed into 20 principal components to detect and eliminate eye artefacts. After the four eye electrodes were removed from the data, epochs in which the amplitudes between -0.2–0.8 s around stimulus onset exceeded ±80 µV or the $z$-transformed amplitudes differed by more than 17.5 standard deviations from the mean of all channels were excluded from further processing. On average, 88% of the trials (1765/2000 per subject) passed the rejection procedure. In two subjects, EEG battery issues led to data loss, so that only 60% (1204) and 24% (479) of the trials remained after the rejection procedure. Bad channels were then interpolated using the weighted average of the neighbouring channels. Lastly, the data were re-referenced to the average of all 60 channels and each individual epoch was baseline corrected by subtracting the mean amplitude from -0.1–0 s before stimulus onset.

Distributed source reconstructions of the ERPs were computed using the MNE-dSPM approach implemented in Brainstorm (version 3.200625; Dale et al., 2000; Tadel et al., 2011). The electrode positions of each subject were first co-registered to the ICBM152 MRI template by aligning three external fiducial points (LPA, RPA, and Nz) and subsequently projecting the electrodes to the scalp of the template MRI. A Boundary Element Method (BEM) volume conduction model based on the ICBM152 template and the corresponding cortical surface (down-sampled to 15.000 vertices) were used as head and source models. The BEM head model was computed using OpenMEEG (version 2.4.1; Gramfort et al., 2010) and comprised three layers (scalp, outer skull, and inner skull) with 1082, 642, and 642 vertices, respectively. To validate the accuracy of the template-based source reconstructions, structural MRIs of four subjects were also obtained and used to devise individual head and source models. Linear MNE-dSPM solutions with dipole orientations constrained to be normal to the cortex were estimated after pre-whitening the forward model with the averaged noise covariance matrix calculated from the individual trials in a time window from -0.2–0 s before vowel onset. The default parameter settings for the depth weighting (order = 0.5, max. amount = 10), noise covariance regularisation (regularise noise covariance = 0.1), and regularisation parameter (SNR = 3) were used throughout. The individual source reconstructions were then converted to absolute values and spatially smoothed (Gaussian kernel, full width at half maximum = 3 mm).

## 3. Results

### 3.1. fNIRS data

#### 3.1.1. Oxygenated haemoglobin (HbO) results

The group-level fNIRS HbO topographies are shown in Fig. 3A. The upper two rows depict the functional activations for the FIXED and VARIABLE PROSODY conditions, whereas the lower row shows

where the activity in the VARIABLE PROSODY condition exceeded that in the FIXED PROSODY condition. To statistically evaluate these three contrasts, it was separately assessed for each channel in the left (ch# 5, 12, 14, and 16) and right (ch# 27, 34, 36, and 38) auditory ROIs whether the beta weights were significantly greater than zero. As Shapiro-Wilk tests indicated at least one ROI channel with a non-normal distribution of the beta weights in each contrast ($p < 0.05$), non-parametric tests were used throughout. The false discovery rate (FDR) across the eight ROI channels was controlled using the Benjamini-Hochberg procedure. After applying this correction, no significant activity was evident in the FIXED PROSODY condition, whereas three channels in the right ROI (ch# 34, 36, and 38) reached significance level in the VARIABLE PROSODY condition (Wilcoxon signed-rank tests, $p_{(FDR)} < 0.05$). Furthermore, when not correcting for multiple comparisons among the ROI channels, three channels along the right STC (ch# 27, 34, and 38) showed significantly greater activity in the VARIABLE PROSODY condition compared to the FIXED PROSODY condition (Wilcoxon rank-sum tests, $p < 0.05$). None of the ROI channels showed an effect in the opposite direction (FIXED > VARIABLE, $p \geq 0.572$).

To test for general effects across conditions and hemispheres, the beta weights were then averaged across the four channels in each ROI. After doing so, Shapiro-Wilk tests indicated no violations of normality anymore and the data were analysed using a two-way repeated measures ANOVA. The results indicated a significant main effect of condition ($F_{(1,19)} = 4.743$, $p = 0.042$), reflecting the overall greater responses in the VARIABLE PROSODY condition, but neither a main effect of hemisphere nor an interaction of the two factors ($p \geq 0.204$).

To complement the topographical results, the grand-averaged time courses of the HbO concentration changes in both conditions are shown in Fig. 3B, separately for both ROIs. In the FIXED PROSODY condition, the cortical haemodynamic response functions (HRFs) of the left as well as the right STC exhibited hardly any increase in HbO concentration during the stimulation period, followed by a marked decrease after sounds offset. In contrast, clear increases in HbO concentration are evident in the VARIABLE PROSODY condition, particularly in the right STC.

#### 3.1.2. De-oxygenated haemoglobin (HbR) results

The group-level fNIRS HbR topographies are shown in Fig. 4A and were plotted and statistically evaluated in the same way as the HbO data. In contrast to the HbO data, the activity of channels in the right ROI reached significance ($p_{(FDR)} < 0.05$) in the FIXED (ch# 34 and 38) as well as the VARIABLE PROSODY condition (ch# 34, 36, and 38). However, despite a strong trend in the right ROI ($p = 0.061$ for ch# 36, $p = 0.098$ across all four ROI channels), the activity in the VARIABLE PROSODY condition did not significantly exceed that in the FIXED PROSODY condition. As for the HbO data, none of the channels in the two ROIs showed an effect in the opposite direction (FIXED > VARIABLE, $p \geq 0.122$). After averaging the beta weights across the four channels in each ROI, a two-way repeated measures ANOVA furthermore revealed neither main effects of condition or hemisphere nor an interaction ($p \geq 0.233$). Nevertheless, the time courses of the HbR concentration changes shown in Fig. 4B vary distinctly between the two conditions, with distinctly more pronounced responses in the VARIABLE PROSODY condition in both conditions.

### 3.2. EEG data

#### 3.2.1. Sensor-level results

The grand-average ERPs relative to the onset of the individual vowels contained in the stimulation blocks are shown in Fig. 5A. On the left, the sensor waveforms in the FIXED and VARIABLE PROSODY conditions are depicted after averaging them over 16 electrodes in the fronto-central scalp region. The selected electrodes are highlighted in the inset scalp map in Fig. 5A and were chosen as auditory ERPs usually have the largest amplitudes in this scalp region. In both conditions, prominent P1 and P2 components are evident, while the intermediate N1 was less pronounced. To statistically evaluate the data, the time windows of the transient auditory ERP components were determined after averaging across conditions and taking the RMS across all scalp channels (P1 = 50–110 ms, N1 = 110–170 ms, and P2 = 170–300 ms).

After averaging the ERP amplitudes across all ROI channels and the respective time windows, permutation-based (n = 10.000) paired *t*-tests across the two conditions indicated significantly larger positive amplitudes for the P1 ($t_{(19)} = 5.02$, $p < 0.001$), N1 ($t_{(19)} = 2.83$, $p = 0.006$), and P2 ($t_{(19)} = 1.93$, $p = 0.034$) in the VARIABLE PROSODY condition. The corresponding scalp maps are shown on the right of Fig. 5A.

As the fNIRS results showed that the haemodynamic responses in the VARIABLE PROSODY condition peaked later than in the FIXED PROSODY condition, suggesting more sustained neural activity throughout the stimulation blocks, the ERP amplitudes and latencies in the first and second half of the blocks were subsequently analysed separately to test for a similar effect. During the first half of the blocks, only the P1 amplitude ($t_{(19)} = 1.90$, $p = 0.032$) was found to be larger in the VARIABLE PROSODY condition. No significant differences were evident for the N1 ($t_{(19)} = -0.07$, $p = 0.533$) and P2 ($t_{(19)} = -0.97$, $p = 0.827$). During the second half of the blocks, however, the P1 ($t_{(19)} = 4.24$, $p < 0.001$), N1 ($t_{(19)} = 3.34$, $p = 0.002$), and P2 ($t_{(19)} = 2.60$, $p = 0.008$) all had significantly more positive amplitudes in the VARIABLE PROSODY condition.

#### 3.2.2. Source-level results

dSPM-based distributed source reconstructions of the ERPs averaged across all vowels contained in the stimulation blocks and the entire duration of the individual vowels (0–800 ms) are shown in Fig. 5C. The results indicate right-lateralised bilateral activity in the primary auditory cortex in both conditions. Additional activity was located to the right STS, particularly in the VARIABLE PROSODY condition. Furthermore, the activity in the right primary auditory cortex extended further anterior into planum polare in the VARIABLE PROSODY condition.

To evaluate the accuracy of the source reconstructions, dSPM solutions of four subjects for which structural MRI scans were available were compared to solutions based on the ICBM152 template MRI. As shown in Fig. S2, activity was generally larger when the subject-specific MRIs were used, while the overall right-lateralised pattern of activity remained unchanged. Similarly, the cortical activity was largely confined to the cortical sulci for the MRI- and template-based dSPM solutions. The individual MRIs thus enabled to localise a greater amount of the energy of the surface-recorded signals to the auditory cortex and adjacent regions, but the results were not qualitatively different from the template-based solutions.

To statistically analyse the source distributions, the time-averaged source activity in the STC was extracted for both hemispheres, based on the Desikan-Killiany atlas implemented in Brainstorm. A two-way repeated measures ANOVA revealed no main effects for the factors condition and hemisphere ($p \geq 0.183$), but a significant interaction ($F_{(1,19)} = 5.029$, $p = 0.037$). A one-sided post-hoc *t*-test comparing the difference between conditions across hemispheres confirmed that this effect was driven by the greater responses in the VARIABLE PROSODY condition in the right hemisphere ($t_{(19)} = 2.243$, $p = 0.016$).

Next, the source waveforms extracted from the right and left STC were analysed (Fig. 5D), based on the same component time windows as in the sensor-level analyses. For each component, a

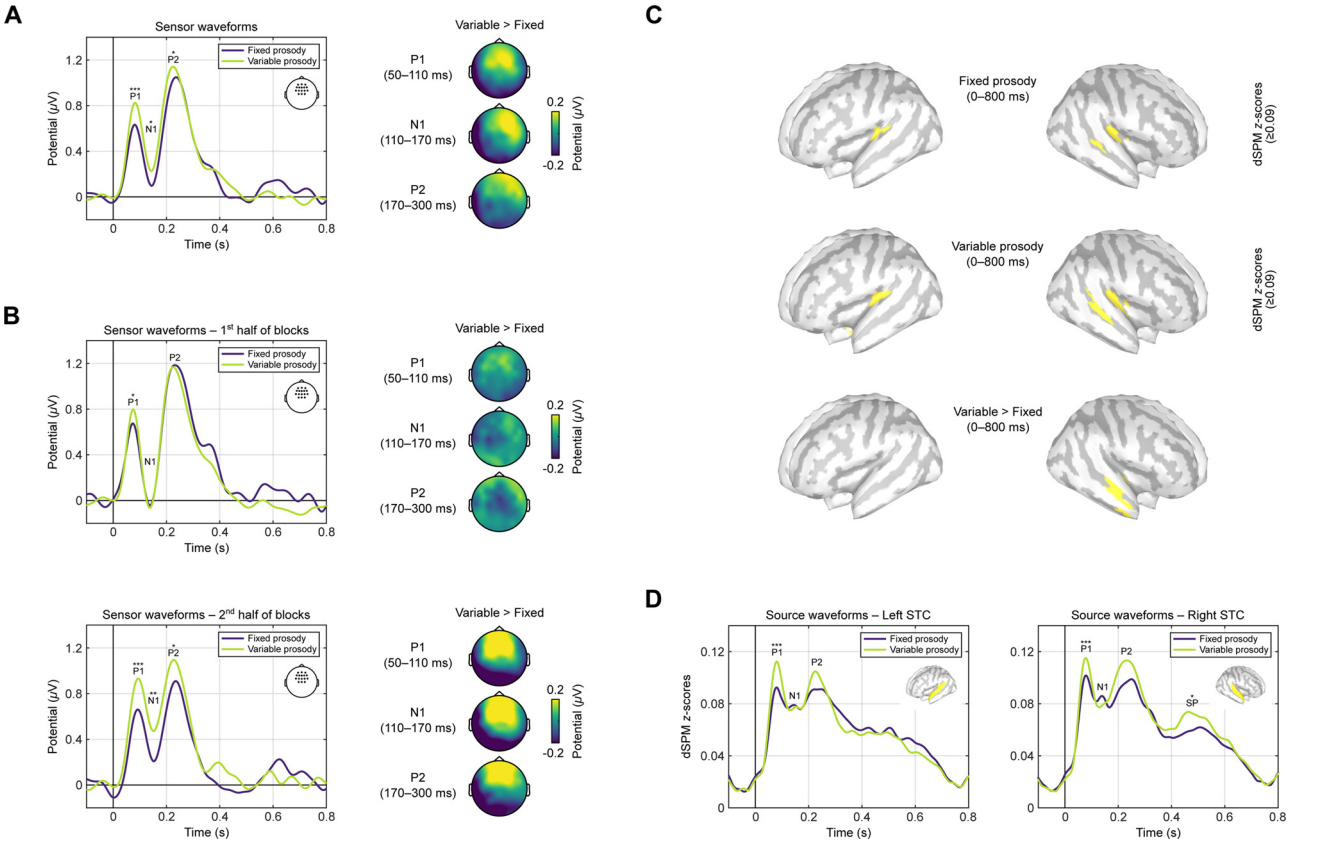


**Fig.5.** EEG results. A) Sensor-level event-related potentials (ERPs) and scalp maps across all vowels in the stimulus blocks. The significance level of the ERP components is indicated by the black asterisks. The scalp maps on the right show differences between conditions after time-averaging the voltages across the specified time windows. B) Sensor-level ERPs and scalp maps for the first and second half of the stimulus blocks. C) dSPM source reconstructions of the ERPs averaged across all vowels in the blocks and the entire duration of the individual vowels. The upper two rows show the cortical activity for the FIXED and VARIABLE PROSODY conditions, respectively. In the lower row, the difference between the two conditions is shown. D) Source-level ERPs across all vowels in the stimulus blocks, extracted from the left and right superior temporal cortices (STCs). The STCs are highlighted in the inserted cortical surface plots. SP = sustained potential.

two-way repeated measures ANOVAs with the factors condition and hemisphere was computed. For the P1, a highly significant main effect of condition was observed ($F_{(1,19)} = 17.584$, $p < 0.001$), reflecting the greater amplitudes in the VARIABLE PROSODY condition, but no effect of hemisphere or interaction ($p \geq 0.352$). In contrast to the sensor level results, no effects were observed for the N1 ($p \geq 0.676$). Despite of a trend for a main effect of condition, there were also no significant effects for the P2 ($p \geq 0.101$). Nevertheless, the condition comparison showed that the P2 in the VARIABLE PROSODY condition was larger in the right hemisphere ($t_{(19)} = 2.180$, $p = 0.021$). Finally, the source waveforms also revealed a late effect (300–800 ms) that was not apparent at the sensor level. No main effects ($p \geq 0.124$) but a significant interaction was found for this sustained potential ($F_{(1,19)} = 5.981$, $p = 0.024$). Here, a one-sided post-hoc *t*-test comparing the difference between conditions across hemispheres confirmed that this effect was driven by the greater responses in the VARIABLE PROSODY condition in the right hemisphere too ($t_{(19)} = 2.446$, $p = 0.012$).

## 4. DISCUSSION

### *4.1. Prosodic variability elicits additional activity in right superior temporal cortex*

The present fNIRS and EEG results both indicated right-lateralised STC activity in response to continuous vowel sequences, regardless of whether the pitch contours of the individual vowels were the same throughout the blocks (FIXED PROSODY) or varied between successive vowels (VARIABLE PROSODY). Furthermore, both methods also revealed additional activity in the right STC in the VARIABLE PROSODY condition. In the fNIRS HbO topographies, this effect is reflected in stronger activity along the right STC, although it did not result in a significant condition*hemisphere interaction. The ERP source reconstructions, in contrast, revealed this type of interaction and stronger activity in the VARIABLE PROSODY condition was particularly evident in the anterior portion of the right STC. The current results thus agree with studies that have found right-lateralised responses to melodies and speech signals with prosodic variations (Dimitrijevic et al., 2013; Meyer et al., 2002; Norman-Haignere et al., 2019). More specifically, they are in line with the notion that pitch changes are primarily processed in the anterior part of the right STC (Johnsrude et al., 2000; Patterson et al., 2002; Zatorre et al., 2002).

Adding to this, the current results have shown that right-hemispheric activity in the superior temporal cortex not only reflects the processing of pitch changes *per se*, but also the variability of the pitch changes contained in the stimulus materials. Furthermore, the EEG data revealed that the late ERP components (P2 and sustained potential) are driving the BOLD effects in right superior temporal areas that reflect the processing of pitch changes in the right STC.

Although the fNIRS and EEG results were largely coherent, the lack of a significant condition*hemisphere interaction in the fNIRS data suggests that this method has a lower sensitivity for detect-

ing auditory activity. As the ERP source reconstructions localised the activity elicited by the stimuli primarily to the cortical sulci (cf Fig. 5.C), this may be explained with the limited depth resolution of the fNIRS measurements. With a standard source-detector distance of 30 mm, the majority of the infrared light is assumed to reach no deeper than about 15 mm (Patil et al., 2011; Pinti et al., 2018). This points at a general limitation of fNIRS in auditory research and raises the possibility that experimental effects may be missed when solely relying on this method, particularly with regard to primary auditory cortex activity emanating from deep in the lateral sulcus.

### 4.2. The ERPs reflect early and late adaption effects

The fNIRS waveforms recorded from the right superior temporal cortex not only exhibited higher amplitudes in the VARIABLE PROSODY condition but were also found to peak significantly later compared to the FIXED PROSODY condition. While the time course of the responses in the VARIABLE PROSODY condition matched the shape of the canonical HRF model well, the waveforms in the FIXED PROSODY condition receded back to baseline level markedly earlier. The pattern in the latter condition is consistent with fMRI data showing monotonically increasing adaptation effects in response to uniform stimulus sequences, an effect also referred to as repetition suppression (Grill-Spector et al., 2006; Steinmann et al., 2012).

In contrast, the ERP data revealed a more complex picture of the underlying neural processes. The P1 amplitudes were larger in the VARIABLE PROSODY condition and this early effect was observed in the STC of both hemispheres. However, larger P2 and sustained potential amplitudes in the VARIABLE PROSODY condition were only found in the right STC. Both of these effects may be explained with greater stimulus-specific adaptation, i.e., the electrophysiological equivalent of repetition suppression (Briley et al., 2013; Fishman et al., 2012; Ulanovsky et al., 2003).

Diminished auditory P1 amplitudes in response to repeated stimuli have been referred to as P50 suppression, a sensory adaptation effect that appears to be unaffected by attention (Jerger et al., 1992; Todorovic et al., 2012). On the other hand, the larger P2 and sustained potential amplitudes in the VARIABLE PROSODY condition may represent an attention-related adaptation effect. This interpretation is in line with the observation that only the auditory ERP components following the P1 are modulated by attention (Picton et al., 1974). The repetitiveness of the FIXED PROSODY condition likely caused the participants to pay less attention to these predictable stimulus sequences after the first few vowels, leading to the observed adaptation effect during the second half of the stimulus blocks.

The current ERP results furthermore agree with the MEG data of Todorovic et al. (2012), who reported similar early and late adaptation effects in an active listening task. In that study, repetition suppression effects for tone pairs with a fixed compared to a variable pitch were also observed for the P1 as well as during a later time window including P2 and sustained response, but not for the intermediate N1. However, in contrast to the sensor-level MEG data of Todorovic et al. (2012), the ERP source analysis in the current study also demonstrated a differential contribution of the right and left auditory cortex to the early and late adaptation effects. The ERP source waveforms showed that the later effect was lateralised to the right STC, whereas the earlier effect was evident in the STC of both hemispheres. More generally, the finding that both the early and late ERP components differed between the FIXED and VARIABLE PROSODY conditions is consistent with the extended duration of the prosodic contours, which the listeners had to track over the course of several hundred ms in order to identify them.

### 4.3. Comparison of the oxygenated (HbO) and de-oxygenated (HbR) fNIRS data

In accordance with the literature on the physiological bases of fNIRS signals (Kirilina et al., 2012; Tachtsidis et al., 2016), the HbR data were less affected by non-cortical artifacts, as indicated by the greater similarity of the total and cortical HRFs. However, the smaller response amplitudes and the correspondingly smaller signal-to-noise ratio compared to the HbO data outweighed this advantage. Overall, the HbR data showed greater activity than the HbO data for the individual conditions, whereas the condition comparison instead returned more pronounced differences for the HbO data. Furthermore, in contrast to the HbO data, the HbR time courses contained separate sound onset and offset peaks. This is in line with fMRI data recorded in response to continuous auditory stimuli (Harms et al., 2002; Steinmann et al., 2012) and thus provides further evidence for the similarity of HbR and BOLD signal changes (Cui et al., 2011; Strangman et al., 2002).

In contrast to a previous fNIRS-EEG study in which we tested normal-hearing listeners with a comparable passive auditory paradigm (Steinmetzger et al., 2020), the HbO and HbR results were largely coherent in the current experiment. In particular, our previous data showed prominent negative HbO responses in fronto-temporal areas, whereas the HbR data showed the expected activity in auditory areas. A possible explanation for this might be that HbO signals are affected by activity of the sympathetic branch of the autonomic nervous system (ANS; Tachtsidis and Scholkmann, 2016). The absence of a similar effect in the current data thus suggests that no substantial changes in ANS activity were elicited. Furthermore, the congruence of the HbO and HbR results suggests that the current results have not been affected by temporalis muscle artefacts, which have been shown to result in concurrent positive HbO and HbR amplitude deflections in the anterior temporal scalp region (Zimeo Morais et al., 2017).

### 4.4. Differences between the sensor- and source-level ERPs

In contrast to the P1 and P2 amplitudes, which were larger in the VARIABLE PROSODY condition at both the sensor and source levels, larger N1 amplitudes were only evident in the sensor waveforms. However, previous MEG studies (Andermann et al., 2021; Patterson et al., 2016; Rupp et al., 2005) have consistently found the N1 to be enhanced in response to melodies as compared to fixed pitch sequences, making an effect in the opposite direction unlikely. Since this effect was not present at the source level, it seems to have been caused by the superposition of the N1 by the neighbouring P1 and P2 components at the scalp level.

Secondly, while auditory MEG studies have consistently reported a sustained field after the initial transient responses (e.g., Andermann et al., 2021; Eulitz et al., 1995; Gutschalk et al., 2002), no sustained potential was evident at the scalp level here (Fig. 5A). In contrast, the source waveforms extracted from the right STC showed a significantly larger amplitude in the VARIABLE PROSODY condition during this period (300–800 ms, Fig. 5D). However, as shown in Fig. S3, after re-referencing the sensor-level data to the vertex electrode (Cz), a sustained potential difference was also evident at the scalp level, particularly at right fronto-temporal electrode sites. This demonstrates that the average reference required for computing the source reconstructions merely attenuated this slow potential, implying that the mean across all scalp channels contained a similar deflection.

## 5. Conclusions

The concurrently obtained fNIRS and EEG data analysed in the current study showed that vowel sequences in which the prosodic

contours vary between successive vowels elicit additional activity in the right superior temporal cortex, compared to sequences in which the prosodic contours are the same throughout. These results demonstrate that cortical activity in right secondary auditory areas not only reflects the processing of prosodic pitch changes as such, but also the variability of these pitch changes.

Moreover, the ERP source analysis enabled the temporal and spatial dissociation of two separate adaptation effects in the superior temporal cortex. Early after vowel onset, differences between the two conditions were confined to the P1, which was larger when the prosodic contours varied between vowels. In addition to this early bilateral effect taken to reflect sensory adaption, smaller P2 and sustained potential amplitudes were observed during sequences without prosodic changes. This later effect was localised to the right hemisphere and is interpreted as an attention-based adaptation effect caused by the repetitiveness of these sequences.

In summary, the current results have shown which electrophysiological processes underlie the right-lateralised activity in secondary auditory regions that many neuroimaging studies have reported when using materials with pitch changes and slow spectral modulations. From a methodological point of view, the results show that the distribution of cortical activity measured with fNIRS and EEG was largely coherent, although the sensitivity of the EEG data was higher than that of the fNIRS data.

## Declaration of Competing Interest

None of the authors has any competing interests to declare.

## CRediT authorship contribution statement

**Kurt Steinmetzger:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Esther Megbel:** Investigation, Data curation. **Zhengzheng Shen:** Investigation. **Martin Andermann:** Conceptualization, Funding acquisition, Resources. **André Rupp:** Conceptualization, Supervision, Funding acquisition.

## Acknowledgments

We are grateful to the Dietmar Hopp Stiftung (grant number 2301 1239) for supporting our research. Parts of this research have been presented at the *23rd International Congress on Acoustics*: https://doi.org/10.18154/RWTH-CONV-239373.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.heares.2022.108483.

## References

Aasted, C.M., Yücel, M.A., Cooper, R.J., Dubb, J., Tsuzuki, D., Becerra, L., Petkov, M.P., Borsook, D., Dan, I., Boas, D.A, 2015. Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. Neurophotonics 2, 020801. doi:10.1117/1.NPh.2.2.020801.

Andermann, M., Günther, M., Patterson, R.D., Rupp, A., 2021. Early cortical processing of pitch height and the role of adaptation and musicality. Neuroimage 225, 117501. doi:10.1016/j.neuroimage.2020.117501.

Brigadoi, S., Cooper, R.J, 2015. How short is short? Optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy. Neurophotonics 2, 025005. doi:10.1117/1.NPh.2.2.025005.

Briley, P.M., Krumbholz, K., 2013. The specificity of stimulus-specific adaptation in human auditory cortex increases with repeated exposure to the adapting stimulus. J. Neurophysiol. 110, 2679–2688. doi:10.1152/jn.00547.2012.

Chatterjee, M., Peng, S.-C, 2008. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. Hearing Res 235, 143–156. doi:10.1016/j.heares.2007.11.004.

Chiarelli, A.M., Maclin, E.L., Fabiani, M., Gratton, G., 2015. A kurtosis-based wavelet algorithm for motion artifact correction of fNIRS data. Neuroimage 112, 128–137. doi:10.1016/j.neuroimage.2015.02.057.

Cui, X., Bray, S., Bryant, D.M., Glover, G.H., Reiss, A.L, 2011. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. Neuroimage 54, 2808–2821. doi:10.1016/j.neuroimage.2010.10.069.

Cullington, H.E., Zeng, F.-G., 2008. Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. J. Acoust. Soc. Am. 123, 450–461. doi:10.1121/1.2805617.

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55–67. doi:10.1016/S0896-6273(00)81138-1.

Dimitrijevic, A., Pratt, H., Starr, A., 2013. Auditory cortical activity in normal hearing subjects to consonant vowels presented in quiet and in noise. Clin. Neurophysiol. 124, 1204–1215. doi:10.1016/j.clinph.2012.11.014.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., Elbert, T., 1995. Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. J. Neurosci. 15, 2748–2755. doi:10.1523/JNEUROSCI.15-04-02748.1995.

Everhardt, M.K., Sarampalis, A., Coler, M., Baskent, D., Lowie, W., 2020. Meta-analysis on the identification of linguistic and emotional prosody in cochlear implant users and vocoder simulations. Ear Hear 41, 1092–1102. doi:10.1097/AUD.0000000000000863.

Fishman, Y.I., Steinschneider, M., 2012. Searching for the mismatch negativity in primary auditory cortex of the awake monkey: deviance detection or stimulus specific adaptation? J. Neurosci. 32, 15747–15758. doi:10.1523/JNEUROSCI.2835-12.2012.

Gramfort, A., Papadopoulo, T., Olivi, E., Clerc, M., 2010. OpenMEEG: opensource software for quasistatic bioelectromagnetics. BioMedical Eng. OnLine 9, 45. doi:10.1186/1475-925X-9-45.

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. 10, 14–23. doi:10.1016/j.tics.2005.11.006.

Gutschalk, A., Patterson, R.D., Rupp, A., Uppenkamp, S., Scherg, M., 2002. Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage 15, 207–216. doi:10.1006/nimg.2001.0949.

Harms, M.P., Melcher, J.R, 2002. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. J. Neurophysiol. 88, 1433–1450. doi:10.1152/jn.2002.88.3.1433.

Henson, R., 2007. Efficient experimental design for fMRI. In: Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.), Statistical parametric mapping: The analysis of functional brain images. Academic Press, London, pp. 193–210. doi:10.1016/B978-0-12-372560-8.X5000-1.

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. doi:10.1364/AO.48.00D280.

Jerger, K., Biggins, C., Fein, G., 1992. P50 suppression is not affected by attentional manipulations. Biol. Psychiatry 31, 365–377. doi:10.1016/0006-3223(92)90230-W.

Johnsrude, I.S., Penhune, V.B., Zatorre, R.J, 2000. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123, 155–163. doi:10.1093/brain/123.1.155.

Kawahara, H., Irino, T., 2005. Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In: Divenyi, P. (Ed.), Speech separation by humans and machines. Springer, Boston, MA, pp. 167–180. doi:10.1007/0-387-22794-6_11.

Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., Ittermann, B., Jacobs, A.M., Tachtsidis, I., 2012. The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. Neuroimage 61, 70–81. doi:10.1016/j.neuroimage.2012.02.074.

Krumbholz, K., Patterson, R., Seither-Preisler, A., Lammertmann, C., Lütkenhöner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb. Cortex 13, 765–772. doi:10.1093/cercor/13.7.765.

Logothetis, N.K., 2003. The underpinnings of the BOLD functional magnetic resonance imaging signal. J. Neurosci. 23, 3963–3971. doi:10.1523/JNEUROSCI.23-10-03963.2003.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157. doi:10.1038/35084005.

Macherey, O., Carlyon, R.P., 2014. Cochlear implants. Curr. Biol. 24, R878–R884. doi:10.1016/j.cub.2014.06.053.

Meddis, R., Hewitt, M.J, 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. J. Acoust. Soc. Am. 89, 2866–2882. doi:10.1121/1.400725.

Meddis, R., O'Mard, L., 1997. A unitary model of pitch perception. J. Acoust. Soc. Am. 102, 1811–1820. doi:10.1121/1.420088.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum. Brain Mapp. 17, 73–88. doi:10.1002/hbm.10042.

Norman-Haignere, S.V., Kanwisher, N., McDermott, J.H., Conway, B.R, 2019. Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. Nat. Neurosci. 22, 1057–1060. doi:10.1038/s41593-019-0410-7.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., Eulitz, C., 2006. Vowel sound extraction in anterior superior temporal cortex. Hum. Brain Mapp. Rauschecker, J.P. 27, 562–571. doi:10.1002/hbm.20201.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 156869. doi:10.1155/2011/156869.

Oxenham, A.J., 2008. Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. Trends Amplif 12, 316–331. doi:10.1177/1084713808325881.

Patil, A.V., Safaie, J., Moghaddam, H.A., Wallois, F., Grebe, R., 2011. Experimental investigation of NIRS spatial sensitivity. Biomedical Optics Express 2, 1478–1493. doi:10.1364/BOE.2.001478.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D, 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776. doi:10.1016/S0896-6273(02)01060-7.

Patterson, R.D., Andermann, M., Uppenkamp, S., Rupp, A., 2016. Locating Melody Processing Activity in Auditory Cortex with Magnetoencephalography. In: van Dijk, P. (Ed.), Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing, pp. 363–369. doi:10.1007/978-3-319-25474-6_38.

Picton, T.W., Hillyard, S.A, 1974. Human auditory evoked potentials. II: Effects of attention. Electroencephalogr. Clin. Neurophysiol. 36, 191–200. doi:10.1016/0013-4694(74)90156-4.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., Burgess, P.W, 2018. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. Ann. N. Y. Acad. Sci. 1464, 5–29. doi:10.1111/nyas.13948.

Plichta, M.M., Heinzel, S., Ehlis, A.-C., Pauli, P., Fallgatter, A.J, 2007. Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: a parametric validation study. Neuroimage 35, 625–634. doi:10.1016/j.neuroimage.2006.11.028.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time. Speech communication 41, 245–255. doi:10.1016/S0167-6393(02)00107-3.

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. Hearing Res 309, 84–93. doi:10.1016/j.heares.2013.11.007.

Rupp, A., Uppenkamp, S., 2005. Neuromagnetic representation of short melodies in the auditory cortex. Fortschritte der Akustik 31, 473–474.

Sato, T., Nambu, I., Takeda, K., Aihara, T., Yamashita, O., Isogaya, Y., Inoue, Y., Otaka, Y., Wada, Y., Kawato, M., 2016. Reduction of global interference of scalp-hemodynamics in functional near-infrared spectroscopy using short distance probes. Neuroimage 141, 120–132. doi:10.1016/j.neuroimage.2016.06.054.

Scholkmann, F., Wolf, M., 2013. General equation for the differential pathlength factor of the frontal human head depending on wavelength and age. Journal of Biomedical Optics 18, 105004. doi:10.1117/1.JBO.18.10.105004.

Scholkmann, F., Kleiser, S., Metz, A.J., Zimmermann, R., Pavia, J.M., Wolf, U., Wolf, M., 2014. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. Neuroimage 85, 6–27. doi:10.1016/j.neuroimage.2013.05.004.

Singh, A.K., Okamoto, M., Dan, H., Jurcak, V., Dan, I., 2005. Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. Neuroimage 27, 842–851. doi:10.1016/j.neuroimage.2005.05.019.

Steinmann, I., Gutschalk, A., 2012. Sustained BOLD and theta activity in auditory cortex are related to slow stimulus fluctuations rather than to pitch. J. Neurophysiol. 107, 3458–3467. doi:10.1152/jn.01105.2011.

Steinmetzger, K., Rosen, S., 2015. The role of periodicity in perceiving speech in quiet and in background noise. J. Acoust. Soc. Am. 138, 3586–3599. doi:10.1121/1.4936945.

Steinmetzger, K., Rosen, S., 2018. The role of envelope periodicity in the perception of masked speech with simulated and real cochlear implants. J. Acoust. Soc. Am. 144, 885–896. doi:10.1121/1.5049584.

Steinmetzger, K., Shen, Z., Riedel, H., Rupp, A., 2020. Auditory cortex activity measured using functional near-infrared spectroscopy (fNIRS) appears to be susceptible to masking by cortical blood stealing. Hearing Res., 108069 doi:10.1016/j.heares.2020.108069.

Strangman, G., Culver, J.P., Thompson, J.H., Boas, D.A, 2002. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. Neuroimage 17, 719–731. doi:10.1006/nimg.2002.1227.

Tachtsidis, I., Scholkmann, F., 2016. False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. Neurophotonics 3, 031405. doi:10.1117/1.NPh.3.3.031405.

Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M, 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 8. doi:10.1155/2011/879716.

Tak, S., Uga, M., Flandin, G., Dan, I., Penny, W., 2016. Sensor space group analysis for fNIRS data. J. Neurosci. Methods 264, 103–112. doi:10.1016/j.jneumeth.2016.03.003.

Todorovic, A., de Lange, F.P., 2012. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. J. Neurosci. 32, 13389–13395. doi:10.1523/JNEUROSCI.2227-12.2012.

Ulanovsky, N., Las, L., Nelken, I., 2003. Processing of low-probability sounds by cortical neurons. Nat. Neurosci. 6, 391–398. doi:10.1038/nn1032.

Uppenkamp, S., Johnsrude, I.S., Norris, D., Marslen-Wilson, W., Patterson, R.D, 2006. Locating the initial stages of speech–sound processing in human temporal cortex. Neuroimage 31, 1284–1296. doi:10.1016/j.neuroimage.2006.01.004.

Worsley, K.J., Friston, K.J, 1995. Analysis of fMRI time-series revisited—again. Neuroimage 2, 173–181. doi:10.1006/nimg.1995.1023.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11, 946–953. doi:10.1093/cercor/11.10.946.

Zatorre, R.J., Belin, P., Penhune, V.B, 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46. doi:10.1016/S1364-6613(00)01816-7.

Zimeo Morais, G., Scholkmann, F., Balardin, J., Furucho, R., de Paula, R.C.V., Biazoli, C., Sato, J., 2017. Non-neuronal evoked and spontaneous hemodynamic changes in the anterior temporal region of the human head may lead to misinterpretations of functional near-infrared spectroscopy signals. Neurophotonics 5, 011002. doi:10.1117/1.NPh.5.1.011002.