

Developmental refinement of cortical systems for speech and voice processing

Citation for published version (APA):

Bonte, M., Ley, A., Scharke, W., & Formisano, E. (2016). Developmental refinement of cortical systems for speech and voice processing. *Neuroimage*, 128, 373-384.
<https://doi.org/10.1016/j.neuroimage.2016.01.015>

Document status and date:

Published: 01/03/2016

DOI:

[10.1016/j.neuroimage.2016.01.015](https://doi.org/10.1016/j.neuroimage.2016.01.015)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

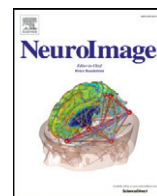
www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Developmental refinement of cortical systems for speech and voice processing



Milene Bonte *, Anke Ley, Wolfgang Scharke, Elia Formisano

Department of Cognitive Neuroscience and Maastricht Brain Imaging Center, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands

ARTICLE INFO

Article history:

Received 15 July 2015

Accepted 6 January 2016

Available online 14 January 2016

Keywords:

Auditory cortex

Brain development

fMRI

Language

Phonological skills

ABSTRACT

Development typically leads to optimized and adaptive neural mechanisms for the processing of voice and speech. In this fMRI study we investigated how this adaptive processing reaches its mature efficiency by examining the effects of task, age and phonological skills on cortical responses to voice and speech in children (8–9 years), adolescents (14–15 years) and adults. Participants listened to vowels (/a/, /i/, /u/) spoken by different speakers (boy, girl, man) and performed delayed-match-to-sample tasks on vowel and speaker identity. Across age groups, similar behavioral accuracy and comparable sound evoked auditory cortical fMRI responses were observed. Analysis of task-related modulations indicated a developmental enhancement of responses in the (right) superior temporal cortex during the processing of speaker information. This effect was most evident through an analysis based on individually determined voice sensitive regions. Analysis of age effects indicated that the recruitment of regions in the temporal–parietal cortex and posterior cingulate/cingulate gyrus decreased with development. Beyond age-related changes, the strength of speech-evoked activity in left posterior and right middle superior temporal regions significantly scaled with individual differences in phonological skills. Together, these findings suggest a prolonged development of the cortical functional network for speech and voice processing. This development includes a progressive refinement of the neural mechanisms for the selection and analysis of auditory information relevant to the ongoing behavioral task.

© 2016 Elsevier Inc. All rights reserved.

Introduction

Learning to recognize human voices and speech is a crucial skill for a healthy cognitive and social development. Efficient processing of speech is pivotal to language acquisition and verbal communication (Jusczyk and Luce, 2002). The analysis of voice itself is fundamental for social interaction as it allows identifying a speaker and her/his emotional state (Belin et al., 2004). Because spoken language contains both types of information, recognition of voices and speech content requires the formation of distinct representations of the same input signal while ignoring the irrelevant dimension and possible interference of background noise. In adults, this selective processing has been shown to rely on (auditory) cortical mechanisms that enable flexible representations of the same sound depending on the current behavioral goal (Bonte et al., 2009; Bonte et al., 2014; von Kriegstein et al., 2003; Schall et al., 2015). The developmental trajectory that leads to these efficient neural representations remains largely unknown and is the focus of the present study. In particular, we aim to trace task-induced changes in the cortical analysis of voice and speech by measuring functional magnetic resonance imaging (fMRI) responses while children, adolescents and adults selectively attend to either speaker or speech sound identity.

Basic brain functions of voice and speech perception emerge during the first year of life. Enhanced processing of the mother's voice in newborns is one of the earliest neuro-functional markers of auditory learning (Beauchemin et al., 2011; Webb et al., 2015). In 3 to 7 months-old infants, listening to voices as compared to non-vocal sounds recruits regions in the right superior temporal cortex that are also part of voice sensitive regions in adults (Belin et al., 2000; Blasi et al., 2011; Grossmann et al., 2010). A similarly early specialization for speech has been found to rely on left-dominant posterior superior temporal cortical regions (Dehaene-Lambertz et al., 2002, 2010) and involves a transition from generic language-universal to language-specific responses emphasizing speech sound categories of the native language (Cheour et al., 1998; Kuhl, 2000; Kuhl and Rivera-Gaxiola, 2008).

While the basic cortical network for speech and voice perception is in place early on, more fine-grained morphological and functional characteristics continue to change throughout childhood and adolescence (Bonte and Blomert, 2004; Bonte et al., 2013; Giedd et al., 1999; Gogtay et al., 2004; Pang and Taylor, 2000; Sharma et al., 1997; Sowell et al., 2002). Such an extended developmental time course may allow a prolonged process of refinement during which experience and learning contribute to the shaping and fine tuning of relevant brain circuitry (Johnson, 2001, 2011). Thus, similar to a gradual developmental progression of cortical stimulus selectivity for faces (Johnson et al., 2009), voice evoked superior temporal cortical responses were recently

* Corresponding author. Fax: +31 43 3884125.

E-mail address: m.bonte@maastrichtuniversity.nl (M. Bonte).

shown to be less selective and more spatially diffuse in 8–9 year old children changing to more selective and focal responses in adults (Bonte et al., 2013).

Functional refinement may also show as an increasingly selective neural response to specific task demands. Studies examining higher-order language functions indeed suggest that children's brain responses may be less adaptive to different task contexts than those of adults. For example, in fMRI studies comparing patterns of cortical activation during reading and oral language tasks, adults showed selective activation of extrastriate visual areas during reading (Booth et al., 2001; Church et al., 2008), and of auditory superior temporal gyrus areas during the processing of spoken word forms (Booth et al., 2001), whereas in children both visual and auditory regions were significantly active in both task contexts. A similar developmental increase in task-selective activity has been observed within the visual face perception network by comparing the processing of different face aspects (identity, expression, gaze) across children, adolescents and adults (Cohen-Kadosh et al., 2013). Next to a gradual sharpening of task selectivity within commonly activated regions, children may also recruit additional brain regions as compared to adults (or vice versa) to perform the same task, even in the absence of behavioral differences (Brown et al., 2005; Church et al., 2008; Cohen-Kadosh et al., 2013). To our knowledge, so far no study has investigated these types of task and age dependent changes within the (auditory) cortical systems for speech and voice perception.

Beyond age and task effects, the brain's responses to speech may also depend on inter-individual variability in language and reading skills, and in particular, on an individual's proficiency in handling the sound structure of spoken language (phonological skills). In children and adults with developmental dyslexia, who often exhibit phonological processing deficits, the posterior superior temporal cortex shows reduced responsiveness to speech (Blau et al., 2009; Blau et al., 2010; Monzalvo et al., 2012). Furthermore, in typically reading children, the strength of left superior temporal gyrus (STG) activity during phonological task performance has been found to correlate with reading (Brennan et al., 2013) and/or phoneme categorization skills (Conant et al., 2014). Thus an additional aim of this study is to investigate whether speech evoked cortical responses in children, adolescents and adults relate to phonological and reading skills.

The present study investigates the effects of age, task demands, phonological and reading skills on neural processing in the developing cortical speech/voice processing network. We measure fMRI responses while children, adolescents and adults listen to vowel sounds spoken by a boy, girl and male speaker and perform delayed-match-to-sample tasks on vowel and speaker identity. The speech stimuli and experimental tasks are designed with the aim to obtain comparable behavioral accuracy across age groups. In both tasks decision pictures follow speech sounds with a delay of ~5 s, which allows focusing our analysis on the first part of each trial and model task-dependent (auditory) cortical extraction of speaker/vowel information prior to the presentation of the decision picture and the subsequent motor response. Because voice selective superior temporal regions show spatial variability across individuals (Pernet et al., 2015), and this variability may change with development (Bonte et al., 2013), in addition to whole-brain analyses we examine this task-dependency within individually determined voice sensitive regions of interest. Finally, participants' performance on offline language tasks is used to investigate whether individual differences in phonological and/or reading skills explain additional variance in speech evoked fMRI responses as compared to age related changes.

Materials and methods

Participants

(f)MRI measurements were performed in thirty-seven Dutch-speaking participants (13 children, 14 adolescents, 10 adults). Analysis

was performed in thirty-three participants: 10 children (8f, mean (SD) age 9.1 ± 0.7 yrs), 13 adolescents (11f, mean age 14.1 ± 0.5 yrs) and 10 adults (6f, mean age 24.1 ± 2.4 yrs). Data of 3 children and 1 adolescent were discarded: 2 children did not complete the measurement, 1 child moved too much during the measurements (>3 mm in 6 out of 8 functional runs), and data of 1 adolescent were discarded due to technical problems during the measurement. Data of the adult participants were also used in a previous multivariate decoding study including a larger number of single trial responses per participant (Bonte et al., 2014). In the current study we used a subset of these data matched in number of trials to those acquired in children and adolescents (see section **Experimental design and procedure**). Adults and adolescents received a monetary reward for participation (€5 per hour), children could select a toy (e.g. car, ball, bracelet, book) after both sessions. Informed consent was obtained from all adult and adolescent participants and from the parents of adolescents and children, according to the approval by the Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University.

All participants were native speakers of Dutch with normal hearing. Auditory detection thresholds were assessed with a pure tone audiogram for frequencies between 250 and 8000 Hz and yielded thresholds at 0–20 decibels HL. With the exception of 1 left-handed participant per age group, all participants were right-handed. Handedness was assessed with an adapted version of Annett's questionnaire (1979), in which four adult-oriented items (striking a match, using thread, using a broom, using a shovel) were replaced with child-oriented items (drawing, using a spoon, using a hair comb, turning a page).

Children, adolescents and adults performed a series of psychometric tests, which showed normal language and cognitive abilities in all participants (Table 1). Cognitive abilities were assessed with the block design, similarities and digit span sub-tests of the WISC (Wechsler et al., 2000) or WAIS (Kort et al., 2005). Reading fluency was assessed by measuring the number of correctly read items within 1.5 min (Blomert and Vaessen, 2009), including high (30 s) and low (30 s) frequency words and pseudowords (30 s). Phonological skills were assessed with a phoneme deletion task consisting of 23 pseudowords with a CVC or CCVCC structure (Blomert and Vaessen, 2009). In this task, the participant is asked to omit as fast as possible a consonant at either the beginning or end of the pseudoword or within a consonant cluster. As the reading and phoneme deletion tests only provide age-appropriate norms until the end of 6th grade, we report both raw scores and norm scores (age norms for children, 6th grade norms for adolescents and adults). Normal reading and phonological skills are indicated by the age-appropriate scores in children and scores of at least 1 SD above the

Table 1
Psychometric test results across age groups.

	Children (n = 10)	Adolescents (n = 13)	Adults (n = 10)
<i>Reading fluency^a</i>			
High freq words	48.9 (7.2)	61.2 (7.5)	67.7 (4.7)
Low freq words	42.5 (8.1)	57.8 (6.0)	64.5 (5.1)
Pseudowords	26.4 (4.7)	42.2 (6.0)	46.2 (5.4)
t-score fluency ^b	52.9 (10)	63.0 (11)	72.0 (5.4)
Phoneme deletion ^a	16.9 (5.2)	22.2 (0.9)	21.6 (2.0)
t-Score ^b	51.0 (10.8)	61.9 (3.8)	59.4 (7.5)
<i>WISC/WAIS^c</i>			
Similarities	10.4 (3.0)	12.1 (3.0)	13.3 (2.8)
Block design	10.5 (2.8)	10.8 (3.0)	11.5 (3.4)
Digit span	11.1 (2.5)	12.3 (2.7)	11.9 (3.2)

^a Raw scores, number of correct items per 30 s (reading fluency), or of 23 items in total (phoneme deletion).

^b t-Scores, mean = 50, SD = 10, children: age-appropriate norms; adolescents/adults: norms end of 6th grade.

^c Age-appropriate norm scores, mean = 10, SD = 3.

mean of 6th grade norms in adolescents and adults. Accordingly, none of the participants (or their parents) reported any history of reading problems. ANOVA analyses showed no significant group differences in age-normed scores on the WISC/WAIS subtests. As expected, age groups did differ in their raw scores on the reading and phonological tests. Reading fluency scores led to significant age group effects for high ($F(2,30) = 20.5$; $p < 0.001$) and low ($F(2,30) = 30.4$; $p < 0.001$) frequency words and pseudowords ($F(2,30) = 37.7$; $p < 0.001$). Except for a non-significant difference between adolescents and adults in the number of correctly read pseudowords ($p = 0.094$), post-hoc t-tests showed significant pair-wise group differences on all raw reading scores. As for phoneme deletion, results showed a main effect of age group ($F(2,30) = 9.5$; $p = 0.001$) and significant post-hoc t-test difference between children and adolescents ($p < 0.001$) and children and adults ($p < 0.005$), but not between adolescents and adults ($p = 0.630$).

Stimuli

Stimuli and experimental paradigm were the same as in a previous multivariate fMRI decoding study in the adult participants (Bonte et al., 2014). Speech stimuli consisted of three natural Dutch vowels (/a/, /i/, and /u/) spoken by three native Dutch speakers (sp1: 9-year-old boy, sp2: 9-year-old girl, and sp3: adult male). We included two utterances of each vowel for each speaker, to introduce acoustic variability typical of natural speech (Fig. 1A). We used children voices in addition to an adult voice in order to make our paradigm more children-friendly. Unlike adult voices, children voices are not readily distinguished based on F0 and their identification additionally relies on formant frequencies (Bennett and Weinberg, 1979; Perry et al., 2001). We expected that children's daily exposure to other children's voices increased the likelihood of obtaining similar behavioral

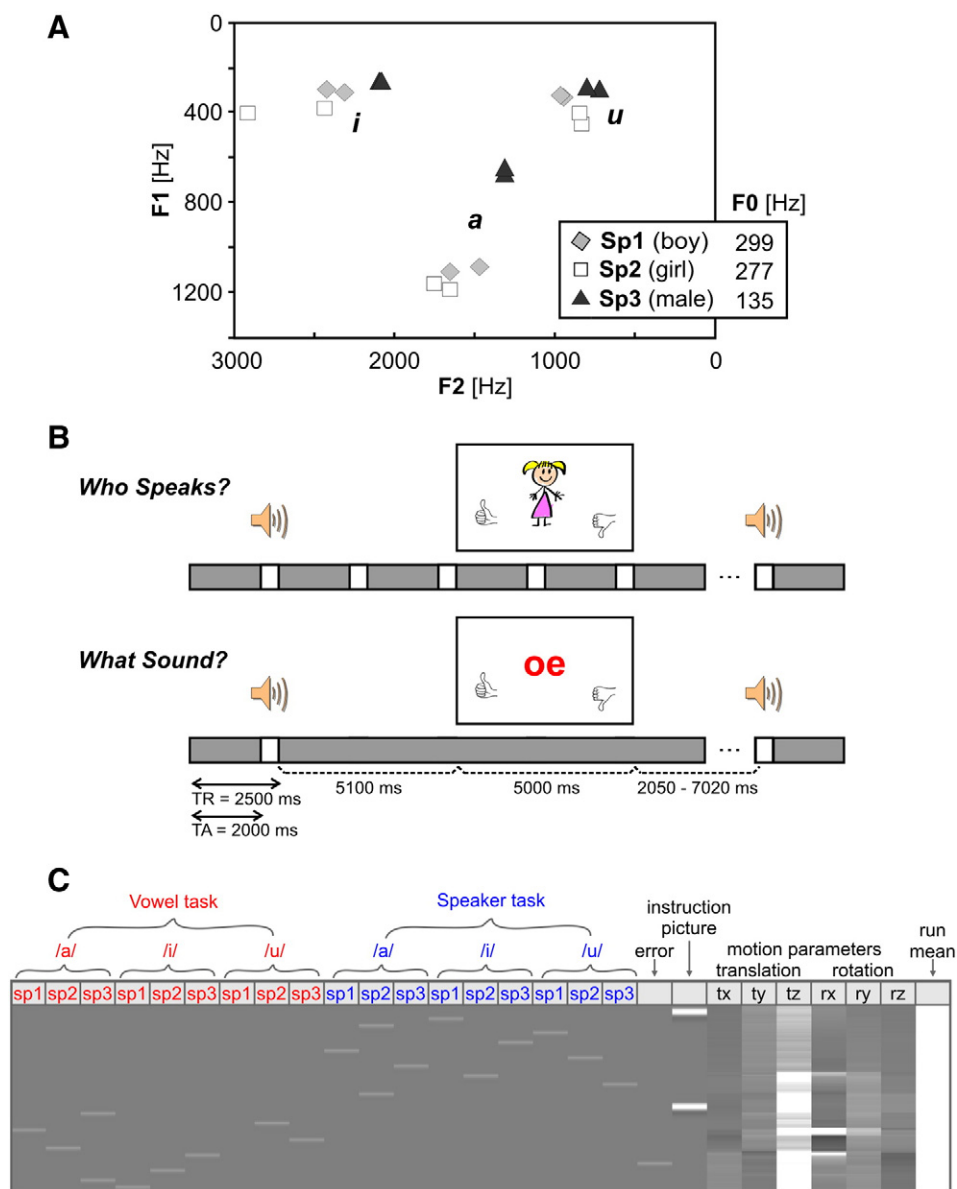


Fig. 1. Stimuli and design. A. F1/F2 formant values for each of the 9 speech stimuli (2 utterances per vowel for each speaker) and mean \pm SD fundamental frequency (F0) values for each of the three speakers. B. Schematic overview of the fMRI stimulation protocol and example experimental trials for the speaker and the vowel task. C. Experimental design illustrated for one experimental run of one child. Note that the 9 speech stimuli per task are displayed as separate predictors for illustrative purposes only; in our first-level design matrix they were collapsed into one speech sound predictor per task. Additional predictors included error trials, instruction pictures, estimated motion correction parameters, and the mean fMRI signal per run per subject.

performance across age groups. The spoken vowels were digitized (44.1 kHz sampling rate), D/A converted (16 bit resolution), band pass filtered (80 Hz–10.5 kHz), and down-sampled to 22.05 kHz. The vowels were edited with PRAAT (Boersma and Weenink, 2002), including equalization of stimulus duration to 350 ms (original range 258–364 ms) using PSOLA (100–400 Hz as extrema for the F0 contour). Length equalization did not lead to any detectable changes in stimulus F0. Sound intensity level was numerically equalized by matching peak amplitudes across all stimuli. Finally, stimuli were faded with 100 ms exponential onset and offset ramps to avoid acoustic transients (clicks) created by a sharp cut-off.

Experimental design and procedure

Task-dependent processing of the speech stimuli was investigated by comparing brain responses during the performance of delayed-match-to-sample tasks on speaker and vowel identity (Fig. 1B; see also Bonte et al., 2014). Both tasks consisted of (1) the presentation of one of the speech stimuli, followed by (2) a decision picture presented at the center of the screen, 5.1 s after speech sound offset, followed by (3) a match/mismatch response of the participant, indicated by pressing a response button with the right index or middle finger respectively. In the speaker task, decision pictures showed cartoons of a girl (see Fig. 1B), a boy or a man. In the vowel task, decision pictures showed the letter combinations 'aa', 'ie' or 'oe' (see Fig. 1B), corresponding to the pronunciation of the 3 Dutch spoken vowels. Decision pictures remained on screen until the button press, or for a maximum time of 5 s. Speech stimuli were presented in pseudo-randomized order in a mixed blocked/slow event-related design with alternations of 3-min "speaker" and "vowel" task blocks (Fig. 1C), presented in a counterbalanced order (see below). An instruction picture signaled the beginning of the next task block to the subject. Each of the 3-min task blocks contained 10 or 11 trials with an inter trial interval of on average 15 s (6 TRs, range 5–7 TRs or 12.5–17.5 s). Each trial consisted of the presentation of one of the 9 speech stimuli (e.g. /a/spoken by the boy). Within a block a speech stimulus was repeated maximally two times, and this repetition never occurred in two subsequent trials. Furthermore, as we had two utterances of each of our 9 speech stimuli, we made sure that the repetition always involved the other utterance of the same stimulus, so that acoustically identical sounds were never repeated within the same task block. Half of the trials involved matching and the other half mismatching pictures, presented in pseudo-randomized order, balanced per task, across experimental runs, and for each of the 9 speech stimuli.

All participants performed two fMRI sessions with a break of 1 to 14 days between sessions. Prior to the first fMRI session, participants were familiarized with the speech stimuli and performed practice trials to make sure both speaker and vowel tasks were understood and the three speakers and vowels were recognized correctly. At the start of the second session, the practice trials were repeated. In children and adolescents, both fMRI sessions consisted of four 6-min experimental runs, each run consisting of 2 alternations of 3-min speaker and vowel task blocks (run 1, 3, 5 and 7: speaker – vowel task; run 2, 4, 6 and 8: vowel – speaker task). In adults both fMRI sessions consisted of three 12-min experimental runs, each run consisting of 4 alternations of 3-min speaker and vowel task blocks (run 1, 3 and 5: speaker – vowel – vowel – speaker task; run 2, 4 and 6: vowel – speaker – speaker – vowel task). In total, we presented 8 (children/adolescents) or 12 (adults) different sequences of speech stimuli, each of them occurring once for the speaker and once for the vowel task, across different fMRI sessions. Each 3-min vowel/speaker task block included 10 or 11 trials during which each of the 9 speech stimuli was presented once and one or two of the stimuli were repeated using a different utterance. Across both fMRI sessions, each of the 9 speech stimuli was presented 14 (adults) or 9 (children, adolescents) times per task. To equalize the number of trials for the 9 speech conditions across groups, the analysis reported here only includes runs 1, 2, 5 and 6 of the adults.

fMRI measurements

Brain imaging was performed with a Siemens Allegra 3 Tesla scanner (head setup) at the Maastricht Brain Imaging Center. Children and adolescents were acquainted to the scanning environment and trained to minimize head movement using a simulation scanner. Movement training consisted of two 6-min blocks during which children/adolescents watched a cartoon movie that was shortly interrupted whenever head movement would exceed 3 mm (see also Bonte et al., 2013). During the fMRI sessions functional runs ($3 \times 3 \times 3 \text{ mm}^3$) were collected using a standard echoplanar-imaging (EPI) sequence (repetition time [TR] = 2500 ms, acquisition time [TA] = 2000 ms, field of view [FOV] = 192 mm \times 192 mm, matrix size = 64 \times 64, echo time [TE] = 32 ms). Each volume consisted of 33 slices (distance factor 10%), covering the whole brain, except the most superior part of the posterior parietal cortex in some participants. The speech stimuli were presented binaurally at a comfortable listening level via MR compatible headphones, in the 500-ms silent gap between two volume acquisitions (Fig. 1B). We used a slow event-related design such that the average inter-trial-interval between two speech stimuli was 15 s (range 12.5 to 17.5 s). Speaker/vowel decision pictures were presented 5.1 s after the offset of the speech stimuli to allow a clear estimation of the auditory activation before the onset of visual and response-related activity. During both scanning sessions a high-resolution structural scan ($1 \times 1 \times 1 \text{ mm}^3$) was collected using a T1-weighted three-dimensional ADNI sequence ([TR] = 2050 ms, [TE] = 2.6 ms, 192 sagittal slices).

At the end of the second session, we additionally acquired an 8-min functional voice localizer run. Voice localizer data of 2 children were not further analyzed due to excessive movement during the measurement. In the present study, the localizer data are used to investigate task-dependent processing of voices/speech sounds in individually determined regions of interest (voice ROIs). See Bonte et al. (2013) for a detailed analysis of developmental changes in voice selective activity and auditory cortical morphology across the same participant groups. Localizer data were collected using the same EPI sequence and slice positioning as the main experiment, but with a TR of 3 s, leaving 1 s of silence for sound presentation. The localizer consisted of 12 stimulation blocks (18 s/6 volumes) alternated with 12 s (4 volumes) rest. In adults we collected 24 stimulation blocks, of which only the first 12 were used to determine the voice ROIs. During the stimulation blocks, participants listened to either (1) vocal sounds (7 non-speech sounds and 5 meaningless speech sounds), (2) other natural categories (musical instruments, environmental and animal sounds) or (3) tones (amplitude modulated (8 Hz) tones ranging from 0.3–3 kHz) (Belin et al., 2000; Bonte et al., 2013).

fMRI pre-processing

Functional MRI data were subjected to conventional pre-processing in BrainVoyager QX 2.8 (Brain Innovation). In particular, slice scan-time correction was performed with respect to the first slice of each volume using sinc interpolation based on information about the TA (2000 ms) and the order of slice scanning (ascending, interleaved). Data were high-pass temporal filtered to remove nonlinear drifts of five or less cycles per time course. Three-dimensional motion correction was performed by spatial alignment of all volumes of a subject to the first volume of the third functional run of each session by rigid body transformations. The third functional run directly followed the structural scan that was acquired midway each fMRI session. Preprocessed functional data were co-registered to each individual subject's structural images and anatomical and functional data were normalized to Talairach space (Goebel et al., 2006). The functional data, acquired with a voxel size of $3 \times 3 \times 3 \text{ mm}^3$, were spatially smoothed with a Gaussian kernel of $4 \times 4 \times 4 \text{ mm}^3$ FWHM. For all participants, functional runs were only included if the estimated head movements were within one voxel

(3 mm) in any direction. In adults all data were included, in adolescents on average 7.2 (SD 1.2) out of 8 runs were included and in children 6.5 (SD 1.4) out of 8 runs. Based on the high-resolution anatomical scans, individual cortical surfaces were reconstructed from gray–white matter segmentations. An anatomically aligned group-average cortical surface representation was obtained by aligning all 33 individual cortical surfaces using a moving target-group average approach based on curvature information (cortex-based alignment, Goebel et al., 2006). The alignment procedure was performed on spherical cortical surface reconstructions of each individual subject and started with a rigid alignment to a spherical curvature map of a randomly chosen subject to minimize differences in rotation and orientation of the reconstructions. This was followed by a non-linear morphing alignment to a dynamic group average curvature map using a coarse to fine (multi-scale) approach. The employed coarse-to-fine scheme ensures that small local structures that may differ across brains (e.g. small gyri and notches) are placed in the same relative position in each brain, i.e. the alignment approach tolerates local curvature differences within established more global correspondence (e.g. large sulci and gyri) and creates improved co-registration of functional data (Goebel et al., 2006; see also Bonte et al., 2013).

Whole brain fMRI analysis

fMRI signal time courses were mapped from volume space to cortical surface space by sampling values located between the grey/white

matter boundary and up to 4 mm into grey matter towards the pial surface using trilinear interpolation, and averaging these values to obtain a single value for each vertex of a cortex mesh. Random effects (RFX) general linear model (GLM) analyses were performed on time course data sampled on individual cortical surface meshes, aligned to the cortical group surface mesh using cortex-based alignment. Our first-level GLM model included one predictor for all speech sounds per task for each trial with a correct response (convolved with a double gamma hemodynamic response function with standard values). The model additionally included separate predictors to model error trials and instruction pictures at the onset of each vowel and speaker task block, as well as confound predictors including each participant's motion correction parameters and the mean BOLD signal for each functional run per subject (see Fig. 1C).

Functional maps were calculated using second level contrast analyses. Standard contrast maps were calculated to assess the overall quality of sound-evoked fMRI responses during the speaker and vowel tasks (speaker task > baseline; vowel task > baseline) in each of the three age groups (Fig. 2) and to assess the overall effects of task (speaker task > vowel task, Fig. 3). Additional 'weighted' contrast analyses were performed to test the effects of age and offline behavioral scores (phonological skills, reading fluency) on overall and task specific responses. First, subjects' age (in months) was demeaned and z-scored and was entered as a contrast vector to 'weight' each subject's response (beta). We then tested for effects of age on overall responses (speaker task + vowel task, Fig. 4) and age \times task interaction (speaker task – vowel task).

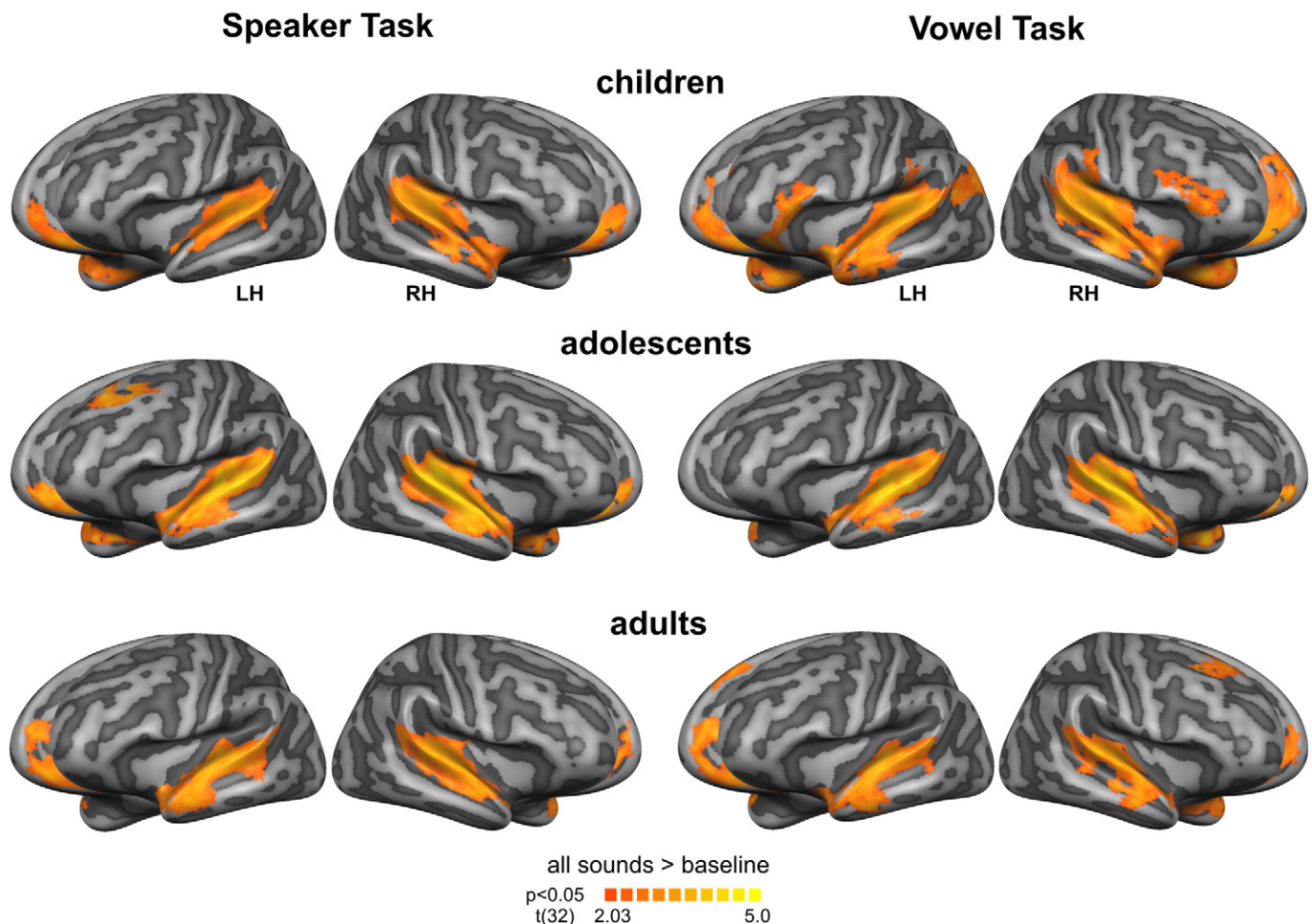


Fig. 2. Speech-evoked BOLD responses. Functional maps illustrating similar speech evoked activity across age groups during the speaker and vowel tasks. Maps are based on random effects contrasts (t-statistics) and visualized on inflated representations of the left (LH) and right (RH) hemispheres (light gray: gyri and dark gray: sulci), resulting from the realignment of the cortices of all 33 participants. Maps are corrected for multiple comparisons using a cluster size correction at $p < 0.05$.

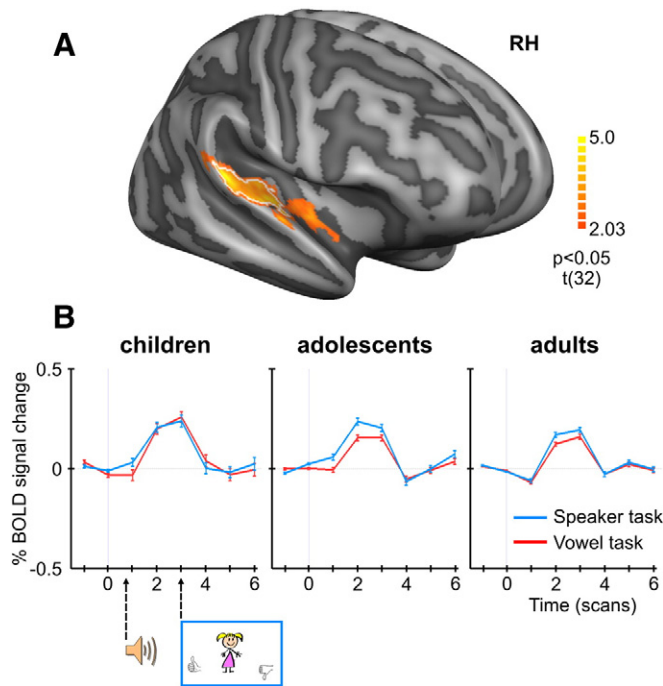


Fig. 3. Main effect of speaker task. Functional map showing a stronger speech evoked right superior temporal cortical response during speaker as compared to vowel task trials. The map is based on random effects contrast analyses [(speaker > vowel task) & (speaker task + vowel task > baseline)] and cluster size corrected at $p < 0.05$ with an initial vertex-level threshold of $p_{\text{vertex}} = 0.05$. The white outline indicates the right posterior PT/STG region that survived cluster-size correction ($p < 0.05$) with a primary threshold of $p_{\text{vertex}} = 0.01$. The corresponding time-course of task-related activity is illustrated for all age groups by plotting BOLD responses (percentage signal change vs volume acquisitions, TR resolution). The arrows indicate the onset of the speech sound and of the response picture. Note that the baseline and response periods in these event-related average plots are chosen for illustrative purposes only; fMRI activity was modeled using a GLM predictor at speech sound onset.

Similar ‘weighted’ contrast analyses were performed to test the effects of Phonological and Reading fluency scores (see Fig. 6 and Results). As these offline behavioral scores correlated with age, before entering the contrast analysis, each score was separately orthogonalized with respect to age using Gram–Schmidt orthogonalization (Wilf, 1962).

In all contrast analyses, significance was assessed using vertex-wise t -statistics and corrected for multiple comparisons by applying a surface-based cluster-size threshold correction. For the latter, an initial vertex-level threshold of $p_{\text{vertex}} = 0.05$ was selected and maps were submitted to a whole-brain correction criterion ($p < 0.05$) based on Monte Carlo simulations (5000 iterations) which also accounted for the estimated map’s spatial smoothness (Forman et al., 1995; Goebel et al., 2006). Additionally, to ensure that the resulting corrections did not critically depend on the selection of the primary threshold, we repeated the same procedure with an initial vertex-level threshold of $p_{\text{vertex}} = 0.01$. Significant regions ($p < 0.05$) under this criterion were marked with white outlines in Figs. 3, 4 and 6.

Finally, we quantified the strength of individual subject’s brain activity in regions with significant effects of age (Fig. 4) and phonological score (Fig. 6) by calculating percent signal change values. To this end, the subjects’ betas corresponding to the speaker or to the vowel task were divided by the subjects’ estimated baseline (b_0), multiplied by 100 and scaled by the trial amplitude in the task predictors.

ROI based analysis

Next to the whole-brain analyses we investigated developmental changes in task-dependent speaker/vowel effects within individually determined regions of interest based on the voice localizer (voice

ROIs; Belin et al., 2000; Bonte et al., 2013). The voice ROIs were defined on each individual participant’s cortical surface mesh and included regions showing significantly stronger activity to voices as compared to both other sound categories and tones (voices > (other + tones) / 2). To prevent large between-subjects differences in the size of the ROIs, the exact statistical threshold was set on an individual basis (see also Bonte et al., 2013; Bonte et al., 2014; Frost and Goebel, 2012). As illustrated by the regions of high inter-subject consistency (Fig. 5A—probabilistic group maps), in most participants the voice ROI included bilateral clusters on the posterior superior temporal gyrus (STG) / superior temporal sulcus (STS) and on the middle STG/STS (at the lateral extremity of Heschl’s sulcus). In 1 adult, 2 adolescents and 3 children, the posterior cluster only reached significance in the right hemisphere and in 2 adolescents the middle cluster only reached significance in one hemisphere (1 left and 1 right). In addition, in 9 of 10 adults, 10 of 13 adolescents and 7 of 8 children the voice ROI also included an anterior STG/STS cluster (at the lateral adjacency of Heschl’s gyrus/first transverse sulcus). The anterior cluster was bilateral/right-lateralized in respectively 5/4 adults, 3/7 adolescents and 5/2 children. In the two children for whom we did not have reliable voice localizer data, we used the average voice ROI of the 8 remaining children. ROI based RFX GLM analysis was performed in the left and right voice ROIs including one predictor per speech sound condition (correct trials only), as well as separate predictors to model error trials and instruction pictures at the onset of each vowel and speaker task block, and confound predictors including participant’s motion correction parameters. The obtained beta estimates were analyzed using repeated measures GLM with task (speaker vs. vowel) and hemisphere (left vs. right) as within subjects factors and age group (children, adolescents, adults) as between subjects factor.

Results

Behavioral results

All participants performed well-above chance level (50%) on the delayed-match-to-sample speaker and vowel tasks (Table 2). Accuracy of speaker identification was comparable across age groups (Group: $F(2,30) = 0.37$; n.s.), but girl/boy voices were more difficult to recognize than the adult voice (Stimulus: $F(2,60) = 18.0$; $p = 0.000$; mean (SD) % correct: boy 89.9 (6.9); girl 82.9 (12.2); man 96.7 (5.3)). Although accuracy of vowel identification was consistently high, results led to a significant difference between children and adolescents/adults (Group: $F(2,30) = 10.0$; $p = 0.000$; mean (SD) % correct: children 95.4 (3.6); adolescents 99.1 (1.5); adults 99.3 (1.0)), without significant stimulus differences. Because decision pictures were presented 5.1 s after speech sound offset, we focused on accuracy, rather than reaction time measures as a reliable indicator of correct recognition of vowels/speakers.

Speech-evoked fMRI activity

Overall, speech evoked BOLD responses showed similar statistical significance levels across age groups indicating comparable data quality (Fig. 2). Across groups, these responses showed activity in a wide expanse of bilateral superior temporal cortex and in medial frontal cortex (Fig. 2). Especially in children, during both tasks, the superior temporal activity extended posteriorly towards the temporoparietal junction, whereas in adults the activity extended towards the right STS especially during the speaker task (for a further analysis of these age and task effects see section Task and age dependent fMRI responses to speech). The activity maps also showed significant activation clusters in additional regions including the bilateral inferior frontal cortex and the left posterior STS/middle temporal gyrus (MTG) during the vowel task in children, the left superior frontal gyrus during the speaker task in

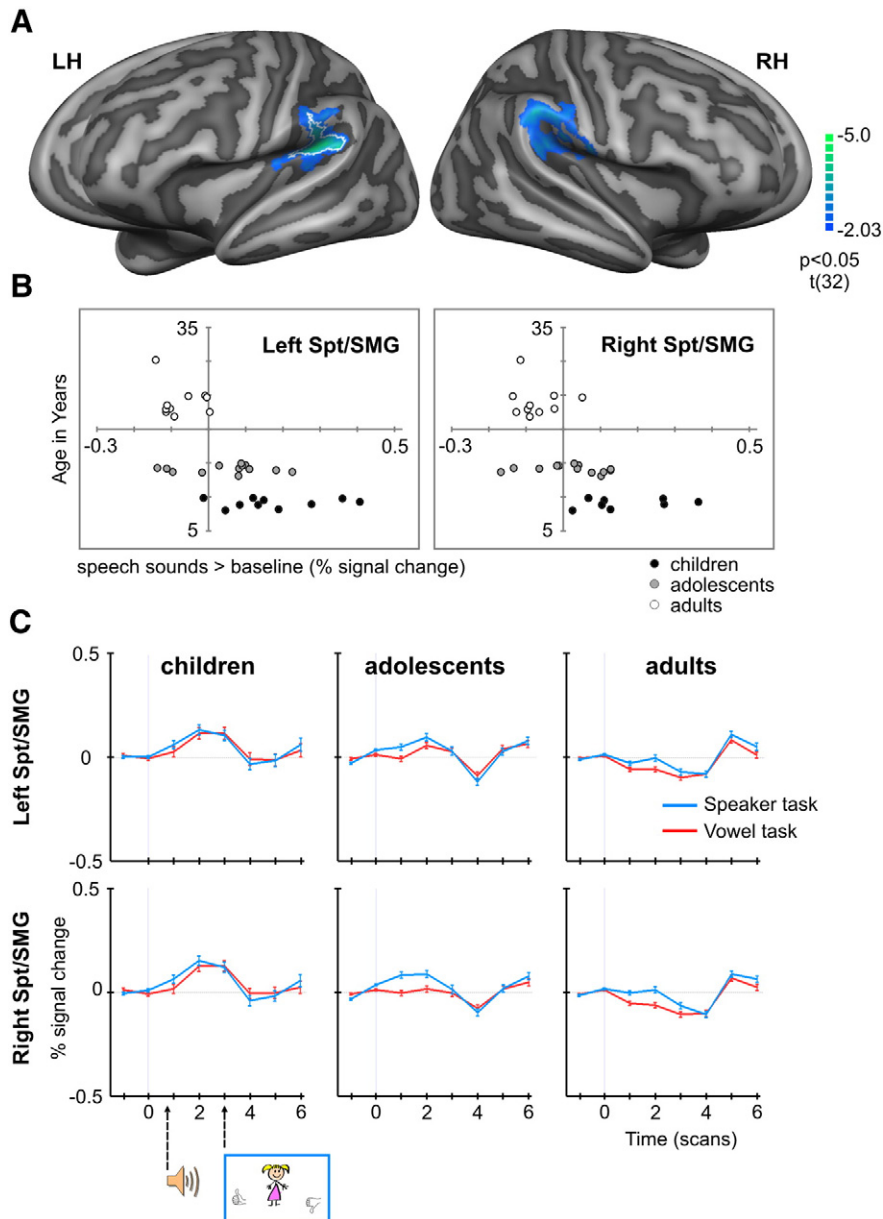


Fig. 4. Main effect of age. A. Functional maps showing regions where speech evoked activity scaled with age. Maps are based on age-weighted random effects contrast analyses (see Methods) and cluster size corrected at $p < 0.05$ with an initial voxel-level threshold of $p_{\text{vertex}} = 0.05$. The white outline indicates the left sylvian parietal temporal (Spt)/supramarginal gyrus (SMG) region that survived cluster-size correction ($p < 0.05$) with a primary threshold of $p_{\text{vertex}} = 0.01$. B. The relation between age and the strength of speech evoked responses (percent BOLD signal change values, averaged across speaker and vowel tasks), in the left and right Spt/SMG across age groups. C. The time-course of task-related activity in the left and right Spt/SMG across age groups. BOLD percentage signal change is plotted with respect to volume acquisitions (TR resolution). The arrows indicate the onset of the speech sound and of the response picture.

adolescents and the right superior frontal gyrus during the vowel task in adults.

Task and age dependent fMRI responses to speech

We assessed the effects of task, age and age \times task interaction in several contrast analyses (see Methods). Firstly, our analysis yielded a main effect of task with stronger responses during the speaker as compared to the vowel task in the right superior temporal cortex encompassing the planum temporale (PT), middle to posterior STG and Heschl's gyrus (primary vertex-level threshold of $p = 0.01$, white outline Fig. 3A), extending towards the first transverse sulcus (primary vertex-level threshold of $p = 0.05$, Fig. 3A).

Secondly, our analysis showed a significant age-related decrease in the recruitment of bilateral temporal–parietal clusters encompassing sylvian parietal temporal (Spt) extending to the supramarginal gyrus (SMG, Fig. 4A, B), and of the left posterior cingulate and right cingulate gyrus (Table 3). These regions showed a clear sound-evoked response during both tasks in children, a smaller response – especially during the vowel task – in adolescents, and no response in adults (Fig. 4C). This age-related change was most clearly present in the left Spt/SMG (white outline in Fig. 4A) and left posterior cingulate (Table 3) that both survived cluster-size multiple comparison correction using a primary vertex-level threshold of $p = 0.01$.

Thirdly, the age \times task interaction contrast did not lead to significant effects in the whole-brain analysis. However, inspection of the time courses of sound-evoked responses in regions with significant task

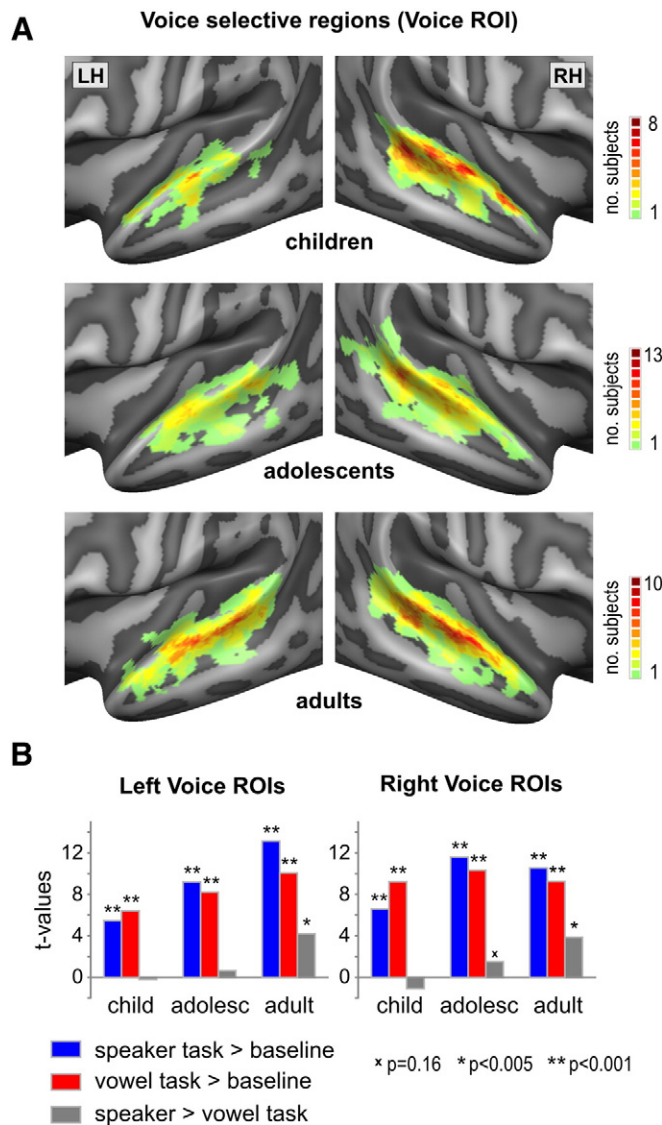


Fig. 5. Speaker task effect in the voice ROIs. A. Probabilistic maps illustrating the spatial overlap of the individually determined voice selective regions (voices > (other + tones) / 2) across age groups. The maps are projected on inflated and aligned group-averaged representations left (LH) and right (RH) superior temporal cortices. B. Analysis of speaker/vowel task effects (t-values) within the left and right voice ROIs. Task effects were assessed by applying a random effects GLM analysis in the individually determined voice selective regions.

effects suggested that it was due to enhanced activity during speaker as compared to vowel task performance in adolescents and adults, whereas this difference was not apparent in children (Fig. 3B, see also ROI-based analysis).

Speaker task effects in voice ROIs

Due to individual variation in the exact location of voice-selective auditory cortical regions (Bonte et al., 2013), the whole brain GLM analysis may be not sensitive enough to detect task-related group differences. Thus, we also investigated task-dependent speaker/vowel processing in individually determined voice ROIs (Fig. 5A). In each of the age groups, the left and right voice ROIs robustly responded to speech sounds during both speaker and vowel task trials (Fig. 5B). Statistical analysis of these responses (beta estimates) demonstrated a developmental change in task dependent activity (task \times group interaction ($F(2,30) = 3.6$; $p < 0.05$, no significant interaction with

Table 2

Mean(se) identification accuracy of vowel and speaker task performance across age groups.

	Children (n = 10)	Adolescents (n = 13)	Adults (n = 10)
<i>Vowel task</i>			
/a/	94.1 (2.1)	99.7 (0.3)	99.7 (0.3)
/i/	98.0 (0.9)	99.1 (0.7)	99.6 (0.4)
/u/	94.0 (2.0)	98.5 (0.8)	98.6 (0.7)
<i>Speaker task</i>			
Boy	91.2 (1.8)	89.4 (2.4)	89.3 (2.4)
Girl	82.3 (4.5)	81.7 (4.3)	84.9 (2.7)
Man	94.7 (2.3)	97.0 (1.4)	98.3 (1.1)

hemisphere). Post-hoc t-tests showed significant speaker task effects in adults ($p < 0.005$ in both left and right voice ROIs), whereas adolescents only showed a tendency in the right Voice ROI ($p = 0.16$) and children showed no significant task effects. Consistent with the fact that these ROIs were selected with a “voice” localizer, there was an overall tendency for the speaker task to elicit stronger activity (main task effect $F(1,30) = 3.4$; $p = 0.07$).

Phonological skill dependent fMRI responses to speech

Because the brain's response to speech may vary with phonological and/or reading skills (Blau et al., 2009; Blau et al., 2010; Monzalvo et al., 2012), we investigated its relation with individual differences in offline phoneme deletion and reading scores. Because groups significantly differed in performance on these tests, it was crucial to orthogonalize the phonological/reading scores with respect to age and consider variance orthogonal to age alone. Whereas individual differences in reading scores did not explain additional variance as compared to age, phonological skills significantly scaled with the strength of speech evoked activity in the left posterior PT towards STG and on the right middle STG/STS (Fig. 6A). Because age-orthogonalized phonological scores similarly scaled with speech evoked activity during both speaker and vowel task trials (Fig. 6B), and showed no significant interactions with task, the illustrated functional contrast map includes both speaker and vowel task activity (Fig. 6A). This behaviorally-weighted contrast was most significant in the left PT (white outline in Fig. 6A). Furthermore, visual inspection of individual subject data (Fig. 6B) suggested that it was mainly driven by the children and by some of the adults with lower age-orthogonalized phonological scores, whereas the relation saturated for participants with highest scores.

Discussion

In this fMRI study we investigated developmental changes in the selective cortical processing of speaker and speech sound information. We specifically investigated the effects on this processing of task, age and phonological skills across children, adolescents and adults. Our findings speak for a gradual progression towards selective and efficient (auditory) cortical processing of speech and voice

Table 3

Brain regions showing overall age effects.

Brain region	Side	Talairach (center of gravity)		
		x	y	z
Spt/SMG **	Left	−49	−31	22
Posterior cingulate **	Left	−17	−55	9
Spt/SMG *	Right	49	−28	27
Cingulate gyrus *	Right	14	−33	39

* Surviving cluster-size correction ($p < 0.05$) with a primary threshold of $p_{\text{vertex}} = 0.05$.

** Surviving cluster-size correction ($p < 0.05$) with a primary threshold of $p_{\text{vertex}} = 0.01$.

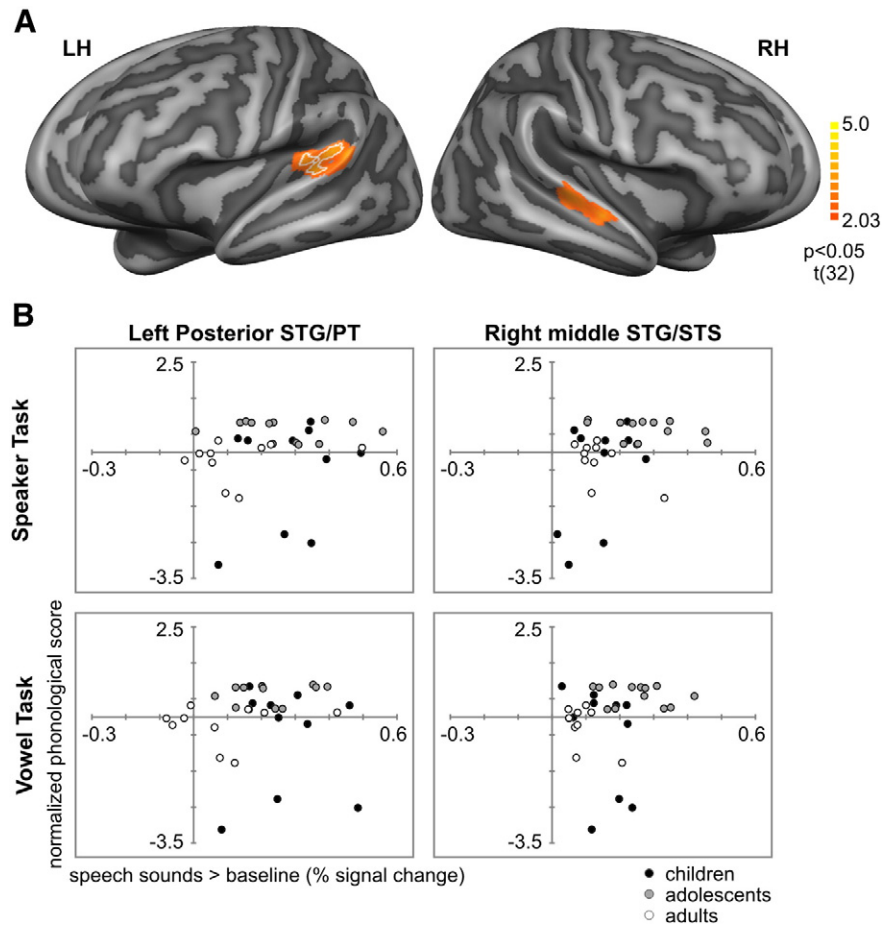


Fig. 6. Speech-evoked responses scale with phonological skills. A. Functional maps showing regions where speech evoked activity scaled with phonological skills. Maps are based on a weighted contrast derived from age-orthogonalized phonological scores (see Methods) and cluster size corrected at $p < 0.05$ with an initial vertex-level threshold of $p_{\text{vertex}} = 0.05$. The white outline indicates the left PT cluster that survived cluster-size correction ($p < 0.05$) with a primary threshold of $p_{\text{vertex}} = 0.01$. B. Plots illustrating the relation between age-orthogonalized phonological scores and the strength of speech evoked responses (percent BOLD signal change values) in the left PT/STG and right STG/STS for the speaker (above) and vowel (below) tasks across age groups.

throughout childhood and adolescence. Furthermore, they suggest that this processing is shaped by individual differences in phonological skills.

To study the selective processing of voice/speech information, we used delayed match-to-sample tasks involving the extraction and maintenance of speaker or vowel identity – and suppression of the task-irrelevant dimension – until the presentation of a decision picture several seconds later. Because we modelled BOLD responses locked to the presentation of the speech sounds, our results primarily reflect the auditory perceptual phase rather than the maintenance phase. By using a children-friendly task design and stimuli spoken by a boy, girl and adult male voice, we obtained similar in-scanner behavioral accuracy across age groups as well as a comparable overall pattern of speech evoked fMRI responses. In both task contexts, speech sounds evoked widespread and bilateral superior temporal cortex activity indicating sensory/perceptual processing of the speech sounds (Binder et al., 2000; Scott et al., 2000). The activation of medial frontal regions most likely relates to more cognitive aspects of the tasks including short-term memory and/or activation of task-relevant stimulus–response mappings (Duncan and Owen, 2000; Euston et al., 2012). Some of the overall activity maps (Fig. 2) also indicated the involvement of additional regions such as, for example, the bilateral inferior frontal cortex and the left posterior temporal cortex during vowel task performance in children. Due to the presence of smaller, below threshold activity clusters in the same regions across tasks, age groups and hemispheres, none of these additional regions yielded significant effects of age and/

or task. The involvement of (left) inferior frontal cortex concurs with its known role in speech perception (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009), especially during effortful lexical-semantic analysis of noisy speech signals (Davis and Johnsruide, 2003; Eisner et al., 2010; Obleser and Kotz, 2010). The left posterior STS/MTG cluster overlaps with a larger region in the inferior parietal lobe recruited during the processing of learned audio-visual associations (Killian-Hütten et al., 2011; Naumer et al., 2009). This activation may thus relate to matching of vowel sounds to letters during vowel task performance. The right hemispheric homologue of this region showed a significant vowel task effect in our previous analysis in the adults (Bonte et al., 2014). In the current data this task specificity was still present in adults and did show a non-significant tendency to decrease with age (data not shown). In future studies, it would thus be interesting to examine possible developmental changes in the posterior temporal STS/MTG and its relation with reading experience, e.g. by specifically targeting the processing of audiovisual associations of letters and speech sounds.

Developmental changes in task-dependent processing of speech/voice sounds were mostly related to the selective processing of speaker information. Speaker task performance led to enhanced activity especially along the right middle to posterior STG, PT and Heschl's gyrus – extending to the first transverse sulcus at a more lenient statistical threshold – confirming the role of these regions in the analysis of human vocal sounds during active or passive tasks (e.g. Belin et al., 2000; Formisano et al., 2008; Bonte et al., 2013; Bonte et al., 2014;

Moerel et al., 2012; von Kriegstein et al., 2003). Most interestingly, this speaker task modulation showed a gradual increase with age, especially when analyzed across individually determined voice selective regions. In these same regions and across the same age groups, we have previously reported an age-related focalization of responses to voices compared to other sound categories (Bonte et al., 2013). Together, these results demonstrate that the superior temporal “voice” network undergoes a continued functional refinement throughout childhood and adolescence. Interestingly, our findings parallel those of investigations in the visual cortex, where a developmental increase in task selectivity has been observed within the cortical face processing network (Cohen-Kadosh et al., 2013).

The enhancement of responses in superior temporal (voice) regions during our speaker task compared to the vowel task with identical stimuli confirms that the processing of vocal sounds in these regions is speech-independent (Bonte et al., 2014; Schall et al., 2015). Furthermore, the age-dependency of this effect suggests that this processing refines with development. Our data and analyses, however, do not allow drawing conclusions on how development affects the cortical representation of voices at the level of individual speaker identity. Addressing such a fine-grained level of representation requires a statistical assessment of the distinctiveness of responses to individual speakers and how this is influenced by the tasks. In other words, it would require a significant task * speaker identity interaction, which was not present in our univariate analysis, or performing a multivoxel pattern analysis (MVPA). In adult participants – using the full data set – we were able to successfully perform MVPA and show task-modulations of individual speaker and vowel representations (see Bonte et al., 2014). Here, the reduced number of trials available in children and adolescents for practical constraints prevented us to perform a similar analysis. Studying developmental changes at the level of speaker identity representations remains a challenge for future studies that may benefit from signal to noise improvements (e.g. through optimized MR coils, or using prospective motion correction, Maclaren et al., 2013).

Overall age effects were characterized by a developmental decrease in speech evoked activity at the temporal–parietal border encompassing Spt and SMG, and in the left posterior cingulate and right cingulate gyrus, with strongest effects in the left lateralized regions. This decrease pertains to neural responses prior to the presentation of the decision picture, suggesting that children, and to a lesser extent adolescents, relied on additional brain regions to extract and/or rehearse task-relevant speaker/vowel information, despite comparable behavioral accuracies across groups. An additional recruitment of left fronto-parietal regions including the Spt in children as compared to adults has been reported recently during performance of a phoneme discrimination task with synthetic consonant vowel syllables along a /ba-/da/continuum (Conant et al., 2014; Liebenthal et al., 2005). Moreover, during higher-order language tasks requiring overt reading, repetition or word generation, children have been shown to more strongly activate left temporal–parietal (SMG) regions (Brown et al., 2005; Church et al., 2008, but see Krishnan et al., 2014). The (left) Spt and SMG constitute adjacent sensorimotor regions within the dorsal auditory pathway that maps heard speech sounds onto vocal-tract related motor actions (Buchsbaum et al., 2005; Hickok and Poeppel, 2007). In adults, the Spt and/or SMG support the repetition of spoken words (Hickok et al., 2009), verbal rehearsal in working memory (Buchsbaum et al., 2011; Fegen et al., 2015; Jacquemot and Scott, 2006) and subvocal rehearsal when speech perception is linguistically or perceptually complicated (Price, 2010). The observed developmental decrease in Spt/SMG activity may thus indicate a relatively strong reliance on these sensorimotor and/or verbal working memory functions during the performance of the vowel and speaker tasks in school-aged children. In particular, it is possible that adults forewent subvocal rehearsal due to the low working memory load of one sound, whereas especially children but possibly also adolescents, adopted a different strategy and did rely on subvocal

rehearsal to reach the same behavioral accuracy. Less specific changes, however, may also play a role. In particular, we also observed an age-related activity decrease in the left posterior cingulate and, at a more lenient threshold, the right cingulate gyrus (see also Brown et al., 2005; Church et al., 2008), possibly related to selective attention to task-relevant information and/or suppression of distracting information (Corbetta et al., 2008; Luna et al., 2001). To further understand the age-related decrease in (left) Spt/SMG recruitment, it would be interesting to study its relation to continued anatomical changes in the posterior superior temporal cortex (Sowell et al., 2002; Gogtay et al., 2004; Giedd et al., 1999) as well as its relation to developmental changes in functional connectivity with other speech and/or task related regions. This could be done, for example, using a task involving a parametric increase in sensorimotor and/or working memory demands (e.g. Fegen et al., 2015). Together, both the developmental tuning of superior temporal cortical regions to speaker task demands and the age-related reduction in the number of recruited regions is consistent with predictions of increasing task selectivity and functional network reorganization as postulated by the interactive specialization framework of functional brain development (Johnson, 2001, 2011).

Besides age and task effects, the strength of speech evoked activity in the left posterior PT/STG and, at a more lenient threshold, the right middle STG/STS, scaled with inter-individual variability in phonological skills. For the left PT/STG, this finding extends previous reports that related reduced speech responsiveness in this region to phonological deficits in developmental dyslexia (Blau et al., 2009; Blau et al., 2010; Monzalvo et al., 2012) and that related left PT/STG activity during phonological task performance to reading and/or phoneme categorization skills in normal readers (Brennan et al., 2013; Conant et al., 2014). Because of its role in voice processing (Belin et al., 2000; Bonte et al., 2013; Moerel et al., 2012), the additional involvement of the right middle STG/STS cluster most likely reflects the prominence of voice analysis in the current experimental design and stimuli. Visual inspection of individual subject's brain activation estimates within the left PT/STG and right middle STG/STS clusters showed that the relation with phonological scores was mainly driven by the children and by some of the adults with lower age-orthogonalized phonological scores. Whereas our GLM analysis tested a linear contrast, the relation seems to follow a sigmoidal function with saturation of the scores/relation in adolescents and most of the adults. Further investigation across participant groups with larger variations in phonological and/or reading skills is thus essential. In particular, to understand what drives these overall differences in response strength, future studies could use more fine-grained multivariate methods to investigate developmental changes in superior temporal speech representations in larger groups of children during normal and dyslexic reading development.

In conclusion, the present study indicates a prolonged period of functional refinement of cortical systems for speech/voice processing during childhood and adolescence. This development involves a shift from less (task) selective brain responses in children to selective and more focal modulation of activity in adults as well as an overall effect of individual differences in phonological skills.

Acknowledgments

This work was supported by The Netherlands Organization for Scientific Research (NWO VENI-grant no. 451-07-002 to MB). We thank Jeannette Boschma and Annemarie Graus for assistance in data acquisition and Giancarlo Valente for useful discussion on data analysis. Special thanks go to the parents and children for their participation and the St. Aloysius school, OBS de Perroen, Bernard Lievegoed School and het Porta Mosana College in Maastricht for their support. Data are available upon request for researchers who meet the criteria for access to confidential data.

References

- Annett, M., 1979. Family handedness in three generations predicted by the right shift theory. *Ann. Hum. Genet.* 42, 479–491.
- Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez-Montes, E., Belin, P., Beland, R., Francoeur, D., Carceller, A.M., Wallois, F., Lassonde, M., 2011. Mother and stranger: an electrophysiological study of voice processing in newborns. *Cereb. Cortex* 21, 1705–1711.
- Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–135.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.
- Bennett, S., Weinberg, B., 1979. Acoustic correlates of perceived sexual identity in preadolescent children's voices. *J. Acoust. Soc. Am.* 66, 989–1000.
- Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.
- Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G.J., Renvall, V., Deoni, S., Gasston, D., Williams, S.C., Johnson, M.H., Simmons, A., Murphy, D.G., 2011. Early specialization for voice and emotion processing in the infant brain. *Curr. Biol.* 21, 1220–1224.
- Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., Blomert, L., 2010. Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133, 868–879.
- Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., Blomert, L., 2009. Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr. Biol.* 19, 503–508.
- Blomert, L., Vaessen, A., 2009. *Differentiaal Diagnostiek van Dyslexie: Cognitieve Analyse van Lezen en Spellen [Dyslexia Differential Diagnosis: Cognitive Analysis of Reading and Spelling]*. Boom Test Publishers, Amsterdam.
- Boersma, P., Weenink, D., 2002. Praat 4.0: A System for Doing Phonetics with the Computer [Computer Software]. Universiteit van Amsterdam, Amsterdam, The Netherlands.
- Bonte, M., Blomert, L., 2004. Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clin. Neurophysiol.* 115, 409–423.
- Bonte, M., Frost, M.A., Rutten, S., Ley, A., Formisano, E., Goebel, R., 2013. Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *NeuroImage* 83C, 739–750.
- Bonte, M., Hausfeld, L., Scharke, W., Valente, G., Formisano, E., 2014. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J. Neurosci.* 34, 4548–4557.
- Bonte, M., Valente, G., Formisano, E., 2009. Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *J. Neurosci.* 29, 1699–1706.
- Booth, J.R., Burman, D.D., Van Santen, F.W., Harasaki, Y., Gitelman, D.R., Parrish, T.B., Marsel Mesulam, M.M., 2001. The development of specialized brain systems in reading and oral-language. *Child Neuropsychol.* 7, 119–141.
- Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, J., Booth, J.R., 2013. Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Hum. Brain Mapp.* 34, 3354–3368.
- Brown, T.T., Lugar, H.M., Coalson, R.S., Miezin, F.M., Petersen, S.E., Schlaggar, B.L., 2005. Developmental changes in human cerebral functional organization for word generation. *Cereb. Cortex* 15, 275–290.
- Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M., Hickok, G., 2011. Conduction aphasia, sensory-motor integration, and phonological short-term memory—an aggregate analysis of lesion and fMRI data. *Brain Lang.* 119, 119–128.
- Buchsbaum, B.R., Olsen, R.K., Koch, P., Berman, K.F., 2005. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48, 687–697.
- Cheour, M., Cepaniene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Naatanen, R., 1998. Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.* 1, 351–353.
- Church, J.A., Coalson, R.S., Lugar, H.M., Petersen, S.E., Schlaggar, B.L., 2008. A developmental fMRI study of reading and repetition reveals changes in phonological and visual mechanisms over age. *Cereb. Cortex* 18, 2054–2065.
- Cohen Kadosh, K., Johnson, M.H., Dick, F., Cohen Kadosh, R., Blakemore, S.J., 2013. Effects of age, task performance, and structural brain development on face processing. *Cereb. Cortex* 23, 1630–1642.
- Conant, L.L., Liebenthal, E., Desai, A., Binder, J.R., 2014. fMRI of phonemic perception and its relationship to reading development in elementary- to middle-school-age children. *NeuroImage* 89, 192–202.
- Corbetta, M., Patel, G., Shulman, G.L., 2008. The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324.
- Davis, M.H., Johnsruide, I.S., 2003. Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431.
- Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L., 2002. Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015.
- Dehaene-Lambertz, G., Montavont, A., Jobert, A., Alliol, L., Dubois, J., Hertz-Pannier, L., Dehaene, S., 2010. Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain Lang.* 114, 53–65.
- Duncan, J., Owen, A.M., 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483.
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., Scott, S.K., 2010. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J. Neurosci.* 30, 7179–7186.
- Euston, D.R., Gruber, A.J., McNaughton, B.L., 2012. The role of medial prefrontal cortex in memory and decision making. *Neuron* 76, 1057–1070.
- Fegen, D., Buchsbaum, B.R., D'Esposito, M., 2015. The effect of rehearsal rate and memory load on verbal working memory. *NeuroImage* 105, 120–131.
- Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647.
- Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322, 970–973.
- Frost, M.A., Goebel, R., 2012. Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage* 59, 1369–1381.
- Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci.* 2, 861–863.
- Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27, 392–401.
- Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent 3rd, T.F., Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L., Thompson, P.M., 2004. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8174–8179.
- Grossmann, T., Oberecker, R., Koch, S.P., Friederici, A.D., 2010. The developmental origins of voice processing in the human brain. *Neuron* 65, 852–858.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hickok, G., Okada, K., Serences, J.T., 2009. Area STp in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* 101, 2725–2732.
- Jacquemot, C., Scott, S.K., 2006. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* 10, 480–486.
- Johnson, M.H., 2001. Functional brain development in humans. *Nat. Rev. Neurosci.* 2, 475–483.
- Johnson, M.H., 2011. Interactive specialization: a domain-general framework for human functional brain development? *Dev. Cogn. Neurosci.* 1, 7–21.
- Johnson, M.H., Grossmann, T., Cohen, K.K., 2009. Mapping functional brain development: building a social brain through interactive specialization. *Dev. Psychol.* 45, 151–159.
- Jusczyk, P.W., Luce, P.A., 2002. Speech perception and spoken word recognition: past and present. *Eur. Hear.* 23, 2–40.
- Kilian-Hutten, N., Valente, G., Vroomen, J., Formisano, E., 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J. Neurosci.* 31, 1715–1720.
- Kort, W., Schittekatte, M., Dekker, P.H., Verhaeghe, P., Compagn, E.L., Bosmans, M., Vermeir, G., 2005. *WISC-III NL Harcourt Test Publishers*, London.
- Krishnan, S., Leech, R., Mercure, E., Lloyd-Fox, S., Dick, F., 2014. Convergent and divergent fMRI responses in children and adults to increasing language production demands. *Cereb. Cortex* <http://dx.doi.org/10.1093/cercor/bhu120>.
- Kuhl, P.K., 2000. A new view of language acquisition. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11850–11857.
- Kuhl, P., Rivera-Gaxiola, M., 2008. Neural substrates of language acquisition. *Annu. Rev. Neurosci.* 31, 511–534.
- Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621–1631.
- Luna, B., Thulborn, K.R., Munoz, D.P., Merriam, E.P., Garver, K.E., Minshew, N.J., Keshavan, M.S., Genovese, C.R., Eddy, W.F., Sweeney, J.A., 2001. Maturation of widely distributed brain function subserves cognitive development. *NeuroImage* 13, 786–793.
- Maclaren, J., Herbst, M., Speck, O., Zaitsev, M., 2013. Prospective motion correction in brain imaging: a review. *Magn. Reson. Med.* 69, 621–636.
- Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J. Neurosci.* 32, 14205–14216.
- Monzalvo, K., Fluss, J., Billard, C., Dehaene, S., Dehaene-Lambertz, G., 2012. Cortical networks for vision and language in dyslexic and normal children of variable socioeconomic status. *NeuroImage* 61, 258–274.
- Naumer, M.J., Doehrmann, O., Muller, N.G., Muckli, L., Kaiser, J., Hein, G., 2009. Cortical plasticity of audio-visual object representations. *Cereb. Cortex* 19, 1641–1653.
- Obleser, J., Kotz, S.A., 2010. Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex* 20, 633–640.
- Pang, E.W., Taylor, M.J., 2000. Tracking the development of the N1 from age 3 to adulthood: an examination of speech and non-speech stimuli. *Clin. Neurophysiol.* 111, 388–397.
- Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage* 119, 164–174.
- Perry, T.L., Ohde, R.N., Ashmead, D.H., 2001. The acoustic bases for gender identification from children's voices. *J. Acoust. Soc. Am.* 109, 2988–2998.
- Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191, 62–88.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Schall, S., Kiebel, S.J., Maess, B., von Kriegstein, K., 2015. Voice identity recognition: functional division of the right STS and its behavioral relevance. *J. Cogn. Neurosci.* 27, 280–291.
- Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 (Pt 12), 2400–2406.

- Sharma, A., Kraus, N., McGee, T.J., Nicol, T.G., 1997. Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalogr. Clin. Neurophysiol.* 104, 540–545.
- Sowell, E.R., Thompson, P.M., Rex, D., Kornsand, D., Tessner, K.D., Jernigan, T.L., Toga, A.W., 2002. Mapping sulcal pattern asymmetry and local cortical surface gray matter distribution in vivo: maturation in perisylvian cortices. *Cereb. Cortex* 12, 17–26.
- von Kriegstein, K., Eger, E., Kleinschmidt, A., Giraud, A.L., 2003. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res. Cogn. Brain Res.* 17, 48–55.
- Webb, A.R., Heller, H.T., Benson, C.B., Lahav, A., 2015. Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3152–3157.
- Wechsler, D., Van der Steene, G., Vertommen, H., Bleichrodt, N., Uiterwijk, J.M., 2000. *WAIS-III NL*. Swets & Zeitlinger, Lisse.
- Wilf, H.S., 1962. *Mathematics for the Physical Sciences*. Dover, New York.