NeuroImage 128 (2016) 373–384

Contents lists available at ScienceDirect

NeuroImage

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

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.

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).

http://dx.doi.org/10.1016/j.neuroimage.2016.01.015
1053-8119/

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 6**th** grade, we report both raw scores and norm scores (age norms for children, 6**th** 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)
Reading fluency[a]			
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)
WISC/WAIS[c]			
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.