*THIS IS A PRE-PUBLICATION VERSION – PLEASE EMAIL FOR ARTICLE*

# Title: The emergence of cerebral specialisation for the human voice over the first months of life

**Short title: Cerebral specialisation for human voice**

**To: Social Neuroscience**

**Lloyd-Fox, S.[1*], Blasi, A.[2], Mercure, E.[3], Elwell, C.E.[4] & Johnson, M.H.[1]**

[1] Centre for Brain and Cognitive Development, Birkbeck, University of London.

[2] Department of Forensic and Neurodevelopmental Science, Institute of Psychiatry, Kings College London, UK.

[3]Institute of Cognitive Neuroscience, University College London, UK.

[4] Department of Medical Physics and Bioengineering, University College London, UK.

* Corresponding author: Sarah Lloyd-Fox, s.fox@bbk.ac.uk. Tel: 0207 631 6327. Fax: 0207 631 6587.

**Keywords**

Functional Near Infrared Spectroscopy (fNIRS), infancy, voice processing, development, STS region,

**Abstract**

How specialized is the infant brain for processing voice within our environment? Research in adults suggests that portions of the temporal lobe play an important role in differentiating vocalizations from other environmental sounds (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin & Grosbras, 2010), however very little is known about this process in infancy. Recent research in infants has revealed discrepancies in the cortical location of voice selective activation, as well as the age of onset of this response (Grossman et al., 2010; Blasi et al, 2011). The current study used functional near infrared spectroscopy (fNIRS) to further investigate voice processing in awake four to seven month old infants. While listening to voice and non-voice sounds, there was robust and widespread activation in bilateral temporal cortex. Further, voice-selective regions of the bilateral anterior temporal cortex evidenced a steady increase in voice selective activation (voice > non voice activation) over four to seven months of age. These findings support a growing body of evidence that suggests that the emergence of cerebral specialisation for human voice sounds evolves over the first six months of age.

## Introduction

The ability to identify a human voice within our rich and cluttered auditory environment is of clear importance as a basis for social interactions. The voice not only conveys information through speech, but also provides cues about a person's gender, age, emotional state and wellbeing (Latinus & Belin, 2011). Functional neuroimaging studies have revealed a network of brain areas specialized for processing information about other human beings. Within this network, areas of the temporal lobes demonstrate stronger activation when adults listen to human vocalizations (including speech, laughter, crying, coughing, etc) over non-voice environmental sounds and acoustically matched stimuli (Belin et al., 2000; Belin & Grosbras, 2010). Furthermore, research on voice processing in adults suggests that voice sensitive regions of the temporal cortex, particularly within the superior temporal sulcus (STS) region (the term 'STS region' includes regions of the superior temporal gyrus, sulcus and middle temporal gyrus) appear to be more prevalent in the right than the left hemisphere (Assal, Aubert & Buttet, 1981; Belin et al., 2000; Van Lancker & Canter, 1982; Van Lancker, Kreiman & Cummings, 1989). Interestingly, and perhaps related to these functional findings, the anatomy of the left superior temporal sulcus (STS) appears to develop more slowly than the right over the first four months of life (Leroy et al., 2011), and there is continued asymmetry in the anatomical structure of the STS region into adulthood (Ochiai et al., 2004; van Essen et al., 2005). The STS region has also been associated with the perception of speech-like mouth movements (infants - Lloyd-Fox et al., 2011; adults - Pelphrey et al., 2005) and silent lip reading (Calvert et al., 1997). Furthermore, this anterior STS region activation does not arise in response to the perception of non-speech like mouth movements (Calvert et al., 1997) such as facial gurning (i.e. twitching or closed-mouth gestures), suggesting that this region of the temporal cortex is particularly sensitive to voice associated visual cues.

Voice processing abilities are present very early in life. Behavioural studies measuring changes in heart rate in neonates have found that they are able to discriminate their parent's voice from others postnatally (Ockleford et al., 1988), and even prenatally (Kisilevsky et al., 2003). Furthermore, recent findings comparing French and German newborn crying patterns suggest that from the first days of life infants display an accent

within their cry melody, which is related to their parent's language (Mampe, Friederici, Christophe & Wermke, 2009). These results suggest that infants are able to identify, and learn from, voices in their surroundings from a very early age. In contrast, speech processing is not fully developed until much later in an infant's life (Friederici, 2005). Despite these findings, comparatively little is known about the cerebral processing of voice sounds (i.e. laughter, crying) in infancy, and how the specialised 'voice sensitive' regions develop over the first year of life. This question is of crucial importance, not only to better our understanding of typical development, but also to investigate the early origins of disorders which influence social cognition such as autism and schizophrenia.

The investigation of speech processing in very young infants suggests that the temporal cortex may not show the same degree of sensitivity to such auditory stimuli as it does in adulthood. Functional Magnetic Resonance Imaging (fMRI) research with two-to-three month old infants has reported activation in a network of left lateralized brain areas to the perception of speech, particularly within the temporal lobe (Dehaene-Lambertz et al., 2002). However, in contrast to adults, no difference was observed within the temporal lobe between the activation to forward and backward speech (Dehaene-Lambertz et al., 2002), between mother's and stranger's voice, or between speech and music (Dehaene-Lambertz et al., 2010).

A recent study used functional Near Infrared Spectroscopy (fNIRS) to investigate cerebral voice processing in four and seven month old infants (Grossman, Oberecker, Koch & Frederici, 2010). Seven-month-old infants revealed a greater response to the voice sounds relative to the non-voice sounds, as evidenced by an increase in oxygenated haemoglobin ($HbO_2$) in a posterior part of the temporal lobe, bilaterally. In line with previous findings in adults, this response was greater in the right than the left hemisphere. In contrast, no such voice-sensitive response was evident in the four-month-olds. Instead, the younger infants showed a significantly greater response to the non-voice sounds relative to the voice sounds in a right temporal region of the cortex. Why did this differential response to non-voice and voice stimuli vary from four to seven months of age? The voice stimuli included both non-speech (i.e. crying, laughing) and speech (words and non-words) sounds, and so it is difficult to ascertain whether this differential response across the two age groups was related to speech or to the processing of other

types of vocalizations. Further, the non-voice stimuli included animal vocal sounds that may share features with those emitted by humans; therefore it is unknown whether this had an impact on the strength of the contrast between the voice and non-voice conditions. Finally, the non-voice stimuli included a mixture of sounds, which would be both familiar and unfamiliar to infants. Thus, it is undetermined whether degree of stimulus familiarity had an effect on the results.

Another recent study has used fMRI to examine the neural basis of voice processing in three-to-seven month old infants (Blasi et al., 2011). This study found defined areas of the right anterior middle and superior temporal gyri that evidenced a greater response to human non-speech voice sounds compared with non-voice environmental sounds (Blasi et al., 2011), in line with the right lateralized voice-sensitive region found in adults (Belin et al., 2000). A further bilateral area of the medial frontal gyrus also revealed a significantly stronger response. In contrast, the non-voice sounds elicited a significantly greater response than the voice sounds in an area of the left posterior superior temporal gyrus. Therefore, in line with the fNIRS study (Grossman et al., 2010), the fMRI study finds selective responses to the voice and non-voice sounds in young infants. However, conclusions from this study are limited, given that it was undertaken with sleeping infants and the sleeping state has been shown to reduce contrast effects in studies with adults (Csizch et al., 2002). Further, the application of fNIRS rather than fMRI may promote the elucidation of developmental changes as it allows the measurement of changes in both oxy-haemoglobin ($HbO_2$) and deoxy-haemoglobin (HHb), and can be applied to the study of awake older infants.

In the present study we used fNIRS to further explore the development of voice processing in infancy by presenting non-speech *voice* sounds (i.e. crying, laughing, coughing, yawning) and familiar *non-voice* sounds (i.e. toys rattling, water running) to four-to-seven month old infants. fNIRS is inherently a very similar technique to fMRI in the regard that it measures haemodynamic response to neuronal activation. Although the fNIRS technique provides data with less spatial resolution than fMRI, research from adults has shown a high degree of correlation between simultaneous recordings of haemodynamic responses with fNIRS and fMRI (Steinbrink et al., 2006). Given the ease of use of this method with infants (Lloyd-Fox, Blasi & Elwell, 2010), and its capacity to

accommodate a good degree of movement from the infants, this technique has now been successfully used in over fifty functional neuroimaging infant studies. To ensure that infants remain relatively still and attentive, the majority of fNIRS auditory studies with awake infants are required to use concurrent presentation of visual stimuli (Lloyd-Fox et al., 2010). In the current experiment the visual stimuli employed were from a previous study with this age group (Lloyd-Fox et al., 2009), where the haemodynamic responses are known. To account for multi-modal effects from this concurrent visual stimulation, the response during a third 'no sound' condition (i.e. visual only) was subtracted from the responses during the auditory experimental conditions. Predictions for the active areas of the fNIRS arrays were informed by the spatial resolution of fNIRS (light transport models; Fukui et al., 2003; structural infant MRI; Salamon et al., 1990) and from the findings of previous work on voice processing in infants and adults (i.e. Belin et al., 2000; Blasi et al., 2011; Grossman et al., 2010). According to these previous findings, we predicted human voice specific activation in the current study would be localised to an anterior portion of the STS region, approximately at T3/T4 of the 10/20 system, and that this selective activation would be greater in the right hemisphere. Furthermore, we examined whether the pattern of responses to the non-voice and voice stimuli differs as a function of age across the group of infants.

## Methods

### *Participants*

Thirty-three healthy four to seven month old infants (15 female; $\tilde{x}$ - 161.22 days; range, 115-209 days) participated in this study. All infants were born full term (37-42 weeks gestation) and with normal birth weight (> 2,500g). There were no known hearing deficits in this population according to a pre-test parent report. Furthermore, the infants had undergone a UK National Health Service screening hearing test in their first week of life. A further seventeen infants participated but were excluded from the study as either they failed to attend for the minimum four trials per experimental condition (n = 15); there was a technical problem with data collection (n = 1); or they had excessive hair which prevented data collection (n = 1). This attrition rate is within the standard range for infant fNIRS studies (see review by Lloyd-Fox et al., 2010). Some infants also

participated in an fMRI study with similar stimuli (Blasi et al, 2011), but never on the same day. All parents gave informed consent before the study and the ethics committee at Birkbeck, University of London, approved the study design.

*Procedure*

Infants wore custom-built NIRS headgear consisting of three source – detector arrays (see Fig. 1) containing a total of 38 channels (source - detector separations; 26 at 2cm, 12 at 4.5cm) and were tested with the UCL topography system (NTS; Everdell et al, 2005).

********** INSERT FIGURE 1 APPROXIMATELY HERE **********

This system used two continuous wavelengths of source light at 770 and 850nm. The different channel separations allowed the measurement of activation at different depths into the cortex. Based on an understanding of light transport and given that the cortex is approximately 0.5cm from the skin surface in this age group (measure taken from structural MRIs; Salamon et al., 1990), the channel separations used in the current study were predicted to penetrate up to a depth of approximately 2cm from the skin surface, potentially allowing measurement of both the gyri and parts of the sulci near to the surface of the cortex. Before the infants' began the study, measurements of their head circumference, and distance between glabella, ears and inion were taken and the location of the channels and arrays relative to these anatomical landmarks were recorded. The distance from the midpoint of the headband over the forehead (the glabella) to the midpoint of the temporal arrays (channel 9; left hemisphere and 28; right hemisphere) is fixed at 11cm and is aligned approximately with T3 and T4 of the 10-20 system on an average five month old infant head (43cm; unpublished observation from the 200+ infants of this age range tested at Birkbeck). Measurements from this group of infants showed that the average head circumference was 43.06cm, and the average distance from the glabella to the ear (T3/T4) was 11.09cm (s.d. 0.64) (in addition measurements were taken in relation to the 10-20 system – nasion to inion and ear to ear via the vertex). Therefore across the majority of the infants the position of the channels varied, relative to T3/T4 by no more than 1cm. With the use of age appropriate infant structural MRIs,

anatomical scalp landmarks and the 10-20 system we can therefore approximate the location of underlying cortical regions for the infants, and draw tentative comparisons of general regional activation with findings from adult populations.

The infants sat on their parent's lap in a dimly lit and sound attenuated room. The parent was instructed to refrain from interacting with the infant during the stimuli presentation unless the infant became fussy or sought their attention. The experiment ended when the infants became bored or fussy as judged by the experimenter who was monitoring their behaviour. The sequence of stimulus presentation is illustrated in Fig 1. The session began with a rest period (30 secs) to familiarise them with the general experimental setup. Following this the trials alternated one after the other, beginning with a 10 second baseline trial followed by a 10 second experimental trial. The three types of auditory experimental trials (*voice, non-voice and silence*) were presented pseudo-randomly to prevent anticipatory effects, and to ensure the infant was presented with an equal number of trials per condition after every 12 trials. Within the *voice* condition, infants were presented with non-speech adult vocalizations, which included coughing, yawning, throat clearing, laughing and crying. Within the *non-voice* trials the infants were presented with naturalistic environmental sounds (that are not human or animal produced, but are likely to be familiar to infants of that age, which included water running, and toys such as rattles, squeaky toy, spinning balls). Voice and non-voice stimuli were chosen from the Montreal Affective Voices (for more detail see Belin, Fillion-Bilodeau, & Gosselin, 2008) and the stimuli of the voice functional localizer (http://vnl.psy.gla.ac.uk/resources_main.php). The voice stimuli were produced from a range of male and female speakers. Additional non-voice stimuli were also recorded by the authors (toy sounds). Each stimulus sequence lasted 8 seconds and consisted of four different sounds (of voice or non-voice stimuli) presented for 0.37 – 2.92 seconds each, interleaved by short silence periods (between 0.16 and 0.24 seconds). The stimulus categories were equivalent in terms of average sound intensity and duration ($p > 0.65$). Each of the four sounds within a stimulus sequence (trial) differed – i.e. laughing, crying, yawning, coughing - but over the course of the session these sounds were presented repeatedly as they were taken from a range of 16 different voice/non-voice stimuli.

During the presentation of the acoustic stimuli, the infants were encouraged to remain attentive and relatively still with the presentation of visual stimuli, displayed on a 117cm plasma screen with a viewing distance of approximately 100cm. The visual stimuli consisted of static images of transport during the baseline trials and dynamic videos of social stimuli during the experimental trials. These visual trials were presented for varying duration between 8-10 seconds, to avoid inducing anticipatory brain activity. Given that the responses to such visual stimuli are known in this age group from previous work (Lloyd-Fox et al., 2009), this visual effect could then be subtracted from the auditory activation with the use of the data from the silent experimental trials. A minimum of four valid trials per experimental condition was required to include an infant in the study.

*Data Processing and Analysis*

Within each optical array, light reaching the detectors will have travelled from the sources through the skin, skull and underlying brain tissue. The NIRS system measured the absorption of this light, from which the changes in oxy-haemoglobin ($HbO_2$) and deoxy-haemoglobin (HHb) concentration (μMol) were calculated and used as haemodynamic indicators of neural activity (Obrig & Villringer, 2003). Initially, the recorded NIR attenuation measurements for each infant were analysed and trials or channels were rejected from further analysis based on the quality of the signals using artifact detection algorithms (Lloyd-Fox et al., 2009, 2010). For each infant, the channels that survived these rejection criteria were entered into further group analyses. Inclusion criteria required each channel to contain valid data in all three experimental conditions. A minimum of four valid trials per condition were set as a threshold for inclusion within infants, and the maximum number of rejected channels could not exceed 10 (of the 26 2cm channels). Given the lower level of light at the 4.5cm channels, the signal – noise ratio was expected to be lower and therefore the threshold was set higher with no minimum limit imposed. Three of the 12 4.5cm channels (12, 19 and 31) were excluded from further group analyses, as they did not yield any valid data across the group of infants.

For each infant, the attenuation signal (from the reflected near-infrared light) was low-pass filtered, using a cut off frequency of 1.8Hz. The data was then divided into blocks consisting of 4 sec of the baseline trial prior to the onset of the 10 sec experimental trial, plus the following 10 sec baseline trial. The attenuation data was detrended with a linear fit between the first and last 4 sec of each 24 sec block. The data was then converted into changes in concentration (μMol) in $HbO_2$ and HHb using the modified Beer Lambert law (Delpy et al, 1988) and assuming a differential pathlength factor for infants (5.13; based on Duncan et al., 1995). The maximum haemodynamic changes in both $HbO_2$ and HHb concentration were analysed. Either a significant increase in $HbO_2$ concentration, or a significant decrease in HHb, is commonly accepted as an indicator of cortical activation in infant work (Lloyd-Fox et al, 2010). If $HbO_2$ and HHb were to either increase or decrease significantly in unison, the signal was considered unreliable and not included in the data set.

Following this, valid experimental stimulus trials for each condition were averaged together within channels for each infant, and a time course of the mean concentration change in $HbO_2$ and HHb was compiled for each channel. These average time courses for each infant were then compiled into grand averaged time response curves of the haemodynamic responses (across all infants) for each channel. A time window was selected between 8 and 16 seconds post experimental stimulus onset. This period of time was selected to include the range of maximum concentration changes observed across infants for $HbO_2$ and HHb. Statistical comparisons of the response to experimental - baseline trials across all infants were made using the valid data for each channel. Paired sample channel-by-channel t-tests were performed on the group data during the specified experimental trial time window, firstly to compare the maximum change (or amplitude) in $HbO_2$ and HHb during the *voice* and *non-voice* conditions relative to the silent condition, and secondly in relation to each other. Finally, post hoc comparisons were conducted on the voice-selective channels to investigate the correlation between age of participant and the amplitude of individual voice-selective haemodynamic responses.

**Results**

In an initial analysis, the grand averaged haemodynamic responses (μMol) of all 33 infants were assessed. T-tests (two-tailed, with a threshold of $p<0.05$) compared the stimulus-related changes in $HbO_2$ and HHb concentration (during the time window of activation described in the methods) evoked by the *voice and non-voice* conditions (experimental conditions 1 and 2) relative to the *silent* visual-only condition (experimental condition 3) in each channel. The significant results are illustrated in Fig. 2 and Table 1.

********** INSERT FIGURE 2 APPROXIMATELY HERE **********

As illustrated in Fig. 2, the haemodynamic responses to the auditory experimental conditions relative to the silent experimental condition were largely localised to the middle/posterior area of each lateral array, centred approximately over the bilateral STS region of the cortex. While the response to the *voice* condition was largely localised to bilateral cortex in a more anterior portion of the STS region (7 of the 9 significant channels; 8, 9, 10, 13, 29, 32, 33), the bilateral response to the *non-voice* condition was localised over a wide area of the arrays, which approximated to the inferior frontal region (4 of the 12 significant channels; 1, 2, 5, 23), the anterior STS region (5 of the 12 significant channels; 9, 13, 28, 29, 32) and the posterior STS region (3 of the 12 significant channels; 29, 33, 34) of the cortex. In a secondary analysis, comparisons were made between the two auditory experimental conditions (*non-voice compared with voice*), within those channels which had been found to be voice, or non-voice sensitive, illustrated in Fig. 2 and Table 1. This analysis revealed that the posterior region of the lateral arrays displayed greater haemodynamic responses to the *non-voice* condition relative to the *voice* condition. These responses were centred approximately over the bilateral posterior STS region. In contrast, the analysis revealed a greater haemodynamic response to the *voice* condition relative to the *non-voice* condition in a left middle area of the array, centred over a more anterior portion of the STS region. There were also two

further channels that revealed a trend for significance in the right hemisphere (channel 21; t = 1.99, p = 0.056; and channel 38; t = 2.19; p = 0.059). To show an example of the time course of the haemodynamic responses, Fig. 3 illustrates the response to the voice and non-voice stimuli in channel 8 (significant voice selective response in HHb) and channel 13 (significant non-voice selective response in $HbO_2$) located in the left hemisphere.

********** INSERT FIGURE 3 APPROXIMATELY HERE **********

In secondary analyses, the haemodynamic responses were investigated further by assessing whether there were changes in voice-selective activation with age. Over the group of infants, Pearson product moment correlation coefficients (r) were calculated in channels 8 and 10 (those channels which displayed a significant effect in the group analyses). This analysis evaluated the strength of linear dependence between age of participant and the amplitude of the voice-selective response – as measured by subtracting the maximum decrease in HHb during the non-voice condition from the maximum in the voice selective condition for each infant. For channel 8 the correlation between the strength of the haemodynamic response (*to voice relative to non-voice*) and the age of participant was borderline significant (see Fig. 4: r = 0.36, p = 0.05). There was no significant correlation in channel 10. Given the results in the left hemisphere – and our hypothesis concerning the predicted right lateralized voice selective response - post-hoc comparisons were conducted on the homologue channels (relative to channel 8 and 10) in the right hemisphere, Channel 27 (the homologue of channel 8) revealed a significant correlation between the strength of the voice-selective haemodynamic response (HHb) and age of participant (see Fig. 4: r = 0.38, p = 0.04). In line with the findings from channel 10, there was no significant correlation in the homologous channel on the right (channel 29). Furthermore, there were no significant correlations between the strength of the non-voice selective haemodynamic responses and age of participant in the channels that exhibited a significant non-voice selective effect in the group analyses.

********** INSERT FIGURE 4 APPROXIMATELY HERE **********

## Discussion

*Voice selective processing*

In this study we identified a voice selective region of the temporal cortex in four-to-seven month old infants. Even accepting certain limitations of fNIRS localisation (see methods) there are striking similarities between our findings and previous fMRI research in infants (Blasi et al., 2011) and adults (Belin et al., 2000), as well as with recent fNIRS findings (Grossman et al., 2010). To visualise the significant results across the infant voice processing studies (current findings; Blasi et al., 2011; Grossman et al., 2010) and to relate these with the adult responses (Belin et al., 2000), Fig. 5 illustrates the relative locations of the areas of maximum activation in each study.

********** INSERT FIGURE 5 APPROXIMATELY HERE **********

In the current study, in addition to activation to the visual stimuli in the current study there was robust and widespread activation to both the voice and non-voice sounds. Further, the voice-selective region of the temporal cortex (voice > non-voice activation) relates well with findings from the aforementioned studies. Moreover, the current findings evidence a steady increase in voice selective activation over four to seven months of age, providing further evidence that cerebral specialisation for human voice sounds emerges sometime over the first six months of age.

While previous research reports bilateral STS region activation, with a greater selective response in the right hemisphere (Belin et al., 2000; Blasi et al., 2011; Grossman et al., 2010; though see review by Petkov, Logothetis & Obleser, 2009), the current findings were stronger in the left hemisphere. Nevertheless, although the overall voice selective group effects were confined to the left hemisphere, we also found some evidence of a selective response in the homologous region in the right hemisphere, supporting previous findings of a bilateral response. In the right hemisphere, near significant group effects were evidenced, and correlation analyses revealed a stronger

voice-selective response as the age of participant increased from four to seven months. Indeed, these findings are in line with previous fNIRS research (Grossman et al., 2010), which found evidence of voice-selective activation in the right hemisphere at seven, but not four months of age. It is possible that voice-selective specialisation could vary considerably across young infants - with the response becoming more robust by the second half of the first year of life – particularly in the right hemisphere. Indeed, there is preliminary evidence of hemispheric differences in myelination, showing significantly more rapid myelination in the temporal, occipital and parietal lobes of the left hemisphere, compared with the right in infants from three – seven months of age (Deoni et al., 2011). This data was derived from a subset of the same infants that took part in the current study. Furthermore, visual inspection of data from a subset of individuals from the present study (9 infants), who also completed the fMRI infant study (Blasi et al., 2011), suggest that the majority of the infants (approximately 70%) who evidenced temporal cortex activation to the voice stimuli (compared with the silent condition) in the fNIRS study also evidenced activation to the neutral voice stimuli (relative to silence) in the temporal lobe in the fMRI study. Currently, work is underway to analyse a larger dataset; to assess the different stimulus-specific activation in further detail; and to more accurately localize these responses within individuals using fiducials and structural MRIs to co-register the individual infants' fNIRS and fMRI results.

Continued investigation of these factors will help elucidate the relative contributions of stimulus specific effects, age of participant and anatomical development (i.e. degree of myelination) on voice selective processing in infancy.

These latest developmental findings on voice selective cortical responses in humans (current research, Blasi et al., 2011; Grossman et al., 2010) are in line with research on voice processing in macaque monkeys, highlighting the role of this region in the processing of auditory communication signals. In a recent neuroimaging study of macaques using high-resolution fMRI, Petkov and colleagues (2008) were able to identify an auditory region in the anterior superior temporal plane that prefers species-specific vocalizations over other vocalizations and sounds. As nonhuman primates lack a number of linguistic capabilities evident in humans (Hauser, Chomsky & Fitch, 2002), these findings would suggest that the processing of the human voice does not rely on

linguistic processing. Indeed we know that speech and voice discrimination in humans dissociate, suggesting the two are supported by different systems within the anterior and middle temporal lobes (Scott, 2008; Rauschecker & Scott, 2009). Furthermore, when participants are asked to focus on the voice of the speakers rather than the linguistic content of the same sentences the activation in the anterior ST region is significantly increased (Von Kriegstein, Eger, Kleinschmidt & Giraud, 2003).

Fullerton and Pandya (2007) have shown that the human superior temporal cortex has differentiated into multiple areas in humans relative to macaques, which could have arisen in response to the evolution of speech and language. These comparative studies suggest that the developmental progression of voice processing in infancy may not depend on language/speech processing, but rather that voice processing operates as a dissociable system for identifying the initial sensory components of auditory communicative signals. A recent fNIRS study with four-month-old infants sheds further light on the development of voice processing in the temporal cortex (Minagawa-Kawai et al., 2011). They presented five forms of auditory stimuli: native speech, non-native speech, human voice (laughter/crying), monkey voice (postive/negative calls) and scrambled sounds (of all four conditions). The results revealed left lateralized temporal responses to the speech conditions, a restricted right lateralized response to the human voice, and a widespread bilateral response to the monkey voice. Interestingly the scrambled sounds (and non-native speech) elicited a restricted response in the left hemisphere only, suggesting that the bilateral response to the monkey voice was not based on novelty alone. It is possible that early in development voice processing is not yet restricted to species-specific stimuli, causing both human and other animal voices to activate these temporal regions. Comparative work with infant macaques would further elucidate whether this developmental plasticity for voice processing is specific to humans or extends to other non-human primates.

*Non-voice selective processing*

We found evidence of robust and widespread bilateral activation to the non-voice sounds, relative to the silent condition, extending from the inferior frontal cortex to an extensive portion of the STS region. In addition, the current study also found evidence of regions

that show a greater response to the non-voice stimuli relative to the voice stimuli. This activation was localized to a bilateral posterior portion of the STS region. Strikingly, this finding is similar to the infant fMRI findings (Blasi et al., 2011), which also evidenced a greater selective response to the non-voice stimuli in this cortical region, though in the fMRI findings it was left-lateralized (see Fig. 5). In both the current results and the results from the fMRI study, the voice-selective regions were in a more anterior location along the STS region, while the non-voice selective regions were localized in a more posterior location. Furthermore, in contrast to voice-selective cortex, there was no evidence of age related changes in non-voice selectivity in the current findings, or in the recent fMRI findings (Blasi et al., 2011). This contrasts with a previous fNIRS study (Grossman et al., 2010), which reported non-voice selective activation in four month olds, but not in seven month olds. However, comparisons with the previous fNIRS study are limited, as the position of their NIRS measurement probes were in a more anterior location than ours (Grossman et al., 2010), and thus they would have been unable to measure the location of non-voice selective activation evident in recent infant work (current findings; Blasi et al., 2011).

The findings from the current study indicate that four-to-seven month olds are able to discriminate between these types of auditory stimuli - with defined regions of the temporal cortex selectively responding to both voice and non-voice sounds. However, it would appear to involve a different, and possibly immature mechanism, compared to that seen in adults. While all three of the studies in infants reveal some degree of significantly greater activation to non-voice relative to voice stimuli, the studies in adults do not find such effects (Belin et al., 2000). How may we reconcile these differences?

As suggested previously (Belin & Grosbras, 2010), this non-voice selective activation could reflect an immature response in infancy. Interestingly, however, some adult research on the comparison of language and environmental sounds found that these two types of stimuli recruit highly overlapping cortical regions (i.e. Dick et al., 2007). These findings are similar to the current study, which also evidenced overlapping regions of activation to non-speech voice sounds and environmental sounds. Furthermore, while adult studies do report some differences in activation to language and environmental sounds (i.e. Girard & Price, 2001; Humphries et al., 2001; Specht & Reul, 2003, Thierry,

Giraud & Price, 2003), there is considerable divergence in the reported extent and location of the responses.

These cross-study differences in activation in the infant studies could depend in part upon the exact stimuli used, methodological details, and task constraints. For example, while the infant research using familiar age-appropriate non-voice environmental sounds (i.e. toys rattling, water running) reported greater responses to the non-voice compared with the voice sounds in a posterior STS region of the cortex (current findings; Blasi et al., 2011), the research using a combination of relatively unfamiliar environmental sounds (i.e. orchestra music, planes) and non-human voice stimuli (i.e. animal cries) only found a small level of non-voice selective activation in infants (Grossman et al., 2010). Indeed, a recent fNIRS study with four-month-olds (Minagawa-Kawai et al., 2011) evidenced a widespread response to monkey voice in the temporal cortex - in a similar location to the response to speech and human voice stimuli – suggesting that the inclusion of animal cries within non-voice stimuli (Grossman et al., 2010) could affect the sensitivity of the voice/non-voice contrasts.

*Methodological considerations*

All stimuli where chosen to represent sounds that young infants are likely to be familiar with, but we acknowledge that comparing between-category familiarity level is extremely difficult to assess. To provide more infant-appropriate stimuli, the non-voice stimuli were intentionally modified from previous adult work (Belin et al., 2000), to include sounds that would be more likely to be familiar to infants (i.e. toys, water, bells). Though the voice sounds were also intended to be familiar to the infants, it is possible that the infants were not highly familiar with the adult crying included within the voice stimuli. Yet, as the voice stimulus trials were a mixed sample of four types of sounds (i.e. crying, laughing, yawning, coughing), it is unlikely that the adult crying stimuli would have driven an overall greater response in this condition. Indeed, it is equally plausible that a proportion of the non-voice stimuli could have been less familiar or salient, but it was intended that overall, the selection of stimuli would be comparable between conditions. Furthermore, we acknowledge that the voice and non-voice stimuli could have differed in terms of their affective value. Yet this is a limitation which is probably true of all studies

on voice processing, as there are few stimuli which holds an equivalent, or higher, affective value than the human voice.

A further consideration is whether brain activation was enhanced by the multi-modal stimulus presentation. Though we attempted to account for multi-modal effects by subtracting the effects of the visual condition from the auditory conditions, it is possible that the functional responses could have been enhanced by the presence of multi-modal stimuli, relative to activation that would result from the visual or auditory presentation alone. However it should be noted that this potential confound exists for many studies with awake infants as either additional visual/auditory stimuli are used to maintain attention, or, when awake, the infants would be looking at whatever is within their environment during stimulation. Indeed without the use of a controlled visual stimulus, there would have been a substantial increase in the level of artifact in the signal, resulting from increased body movement from the infants, which could have compromised the reliability of the results. The previous fNIRS voice processing study (Grossman et al., 2010) also presented visual stimuli alongside the vocal and non-vocal sounds. Therefore the different types of multimodal presentation (i.e. visual human versus non-human stimuli) may potentially have contributed to the differences in observed activation across the two studies, particularly in the non-voice condition. Yet, in support of the current findings (see Fig. 5), one can see that the locations of our non-voice and voice selective responses are similar to the infant MRI findings (Blasi et al., 2011) - a study which did not have concurrent visual stimuli as the infants were asleep - suggesting that our results were unlikely to have been primarily driven by multimodal stimulation.

*Measuring $HbO_2$ and HHb changes*

Finally, the current study evidenced significant haemodynamic effects in both $HbO_2$ and HHb concentration changes, a finding with important implications for fNIRS infant research. Traditionally, a significant increase in $HbO_2$ and/or a significant decrease in HHb concentration are considered indicators of cortical activation (Villringer and Chance, 1997). However, while adult fNIRS studies show $HbO_2$ and HHb responses consistent with BOLD fMRI signal changes (i.e. a decrease in HHb and parallel increase in $HbO_2$), infant data often reveals a less consistent pattern of HHb changes in response

to neural activation (Baird et al, 2002; Sakatani, et al, 1999; Meek et al, 1998). It is uncertain whether in infants the standard decrease in HHb should be expected, and it remains a matter of much controversy in both fNIRS and fMRI developmental research (Marcar et al., 2004). Indeed, as HHb results are often inconsistent in infants, they tend to be presented less in published papers. A recent review (Lloyd-Fox et al., 2010) reported that of the 36 infant fNIRS studies published by early 2010, 44% failed to report both $HbO_2$ and HHb. It is paramount that simultaneous fMRI and fNIRS research is undertaken to investigate the infant haemodynamic response further.

The current findings highlight the importance of reporting both $HbO_2$ and HHb concentration changes, as it is an important consideration here that the HHb effects alter the significance and degree of activation. Therefore, we reiterate our recommendation to report both $HbO_2$ and HHb results, particularly with a view for drawing comparisons with other fNIRS and fMRI research.

*Conclusions*

The voice-sensitive temporal brain responses identified in the current infant fNIRS research, add to converging evidence from infant and adult fMRI research. Further, these regions of the temporal cortex evidenced a steady increase in voice selective activation (voice > non voice activation) over four to seven months of age, suggesting that this cerebral specialisation emerges, and matures over the first months of life. These findings have important implications for neurodevelopmental disorders such as autism, as recent research has found atypical functioning of the temporal lobe in this population (Gervais et al 2004; Raznahan et al., 2010; Kaiser et al., 2010). In addition to confirming and building on the previous research of cerebral specialisation to voice processing with infants. The current findings also demonstrate the potential of the fNIRS method, including the promise of elucidating the relationship between age of onset of cognitive skills, individual differences and their relation to other developmental milestones.

Acknowledgements

We would like to thank all of the parents and infants who took part in this study, and the reviewers for their informative feedback. This work was supported by the UK Medical Research Council Programme Grant G0701484.

References

Assal, G., Aubert, C., & Buttet, J. (1981). Asymétrie cérébrale et reconnaissance de la voix. *Revue Neruology*, *137*, 255-268.

Baird, A. A., Kagan, J., Gaudette, T., Walz, K. A., Hershlag, N., & Boas, D. A. (2002). Frontal Lobe Activation during Object Permanence: Data from Near-Infrared Spectroscopy. *NeuroImage*, *16*(4), 1120-1126. doi:10.1006/nimg.2002.1170

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531.

Belin, P., & Grosbras, M. H. (2010). Before Speech: Cerebral Voice Processing in Infants. *Neuron*, *65*(6), 733–735.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312.

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, A., Deeley, Q., Barker, G., et al. (2011). Early specialization in the infant brain for processing the human voice and its emotions: an fMRI Study. *Current Biology, 21*, 1220-1224.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593.

Czisch, M., Wetter, T.C., Kaufmann, C., Pollmächer, T., Holsboer, F., & Auer, D.P. (2002). Altered processing of acoustic stimuli during sleep: reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. *NeuroImage, 16*, 251-258.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*(5600), 2013-2015. doi:10.1126/science.1077066

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, *114*(2), 53-65. doi:10.1016/j.bandl.2009.09.003

Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S., & Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, *33*(12), 1433-1442.

Deoni, S., Mercure, E., Blasi, A., Gasston, D., Thomson, A., Johnson, M. H., Williams, S., et al. (2011). Mapping Infant Brain Myelination with Magnetic Resonance Imaging. *Journal of Neuroscience*, *31*(2), 784-791.

Dick, F., Saygin, A. P., Galati, G., Pitzalis, P., & Bentrovato, S. (2007). What is involved and what is necessary for complex linguistic and non-linguistic auditory processing: Evidence from fMRI and lesion data. *Journal of Cognitive Neuroscience*, *19*(5), 799–816.

Duncan, A., Meek, J. H., Clemence, M., Elwell, C. E., Tyszczuk, L., Cope, M., & Delpy, D. (1995). Optical pathlength measurements on adult head, calf and forearm and the head of the newborn infant using phase resolved optical spectroscopy. *Physics in Medicine and Biology*, *40*, 295.

Everdell, N. L., Gibson, A. P., Tullis, I. D. C., Vaithianathan, T., Hebden, J. C., & Delpy, D. T. (2005). A frequency multiplexed near-infrared topography system for imaging

functional activation in the brain. *Review of Scientific Instruments*, *76*(9), 093705. doi:10.1063/1.2038567

Friederici, A. D. (2006). The neural basis of language development and its impairment. *Neuron*, *52*(6), 941–952.

Fukui, Y., Ajichi, Y., & Okada, E. (2003). Monte Carlo prediction of near-infrared light propagation in realistic adult and neonatal head models. *Applied Optics*, *42*(16), 2881-2887.

Fullerton, B.C. & Pandya, D.N. (2007) Architectonic analysis of the auditory related areas of the superior temporal region in human brain. *Journal of Comparative Neurology, 504,* 470-498.

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthélémy, C., et al. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, *7*(8), 801-802. doi:10.1038/nn1291

Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience*, *13*(6), 754–765.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, *65*(6), 852-858. doi:10.1016/j.neuron.2010.03.001

Hauser, M.D., Chomsky, N. & Fitch, W.T. (2002) The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport*, *12*(8), 1749.

Kaiser, M. D., Hudac, C. M., Shultz, S., Lee, S. M., Cheung, C., Berken, A. M., Deen, B., et al. (2010). Neural signatures of autism. *Proceedings of the National Academy of Sciences*, *107*(49), 21223 -21228. doi:10.1073/pnas.1010412107

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K., et al. (2003). Effects of Experience on Fetal Voice Recognition. *Psychological Science*, *14*(3), 220 -224. doi:10.1111/1467-9280.02435

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145.

Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J.-F., & Dehaene-Lambertz, G. (2011). Early Maturation of the Linguistic Dorsal Pathway in Human Infants. *J. Neurosci.*, *31*(4), 1500-1506. doi:10.1523/JNEUROSCI.4141-10.2011

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience and Biobehavioral Reviews*, *34*(3), 269-284. doi:10.1016/j.neubiorev.2009.07.008

Lloyd-Fox, S., Blasi, A., Everdell, N., Elwell, C. E., & Johnson, M. H. (2011). Selective Cortical Mapping of Biological Motion Processing in Young Infants. *Journal of Cognitive Neuroscience*, 23(9), 2521-2532.

Lloyd-Fox, S., Blasi, A., Volein, A., Everdell, N., Elwell, C. E., & Johnson, M. H. (2009). Social perception in infancy: a near infrared spectroscopy study. *Child Development*, *80*(4), 986-999. doi:10.1111/j.1467-8624.2009.01312.x

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994–1997.

Marcar, V. L., Strassle, A. E., Loenekker, T., Schwarz, U., & Martin, E. (2004). The influence of cortical maturation on the BOLD response: an fMRI study of visual cortex in children. *Pediatric research*, *56*(6), 967.

Meek, J. H., Firbank, M., Elwell, C. E., Atkinson, J., Braddick, O., & Wyatt, J. S. (1998). Regional hemodynamic responses to visual stimulation in awake infants. *Pediatric Research*, *43*(6), 840-843.

Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R. & Dupoux, E. (2011) Optical brain imaging reveals general auditory and language specific processing in early infant development, *Cerebral Cortex, 21,* 254-261.

Obrig, H., & Villringer, A. (2003). Beyond the visible—imaging the human brain with light. *Journal of Cerebral Blood Flow & Metabolism*, *23*(1), 1–18.

Ochiai, T., Grimault, S., Scavarda, D., Roch, G., Hori, T., Rivière, D., Mangin, J. F., et al. (2004). Sulcal pattern and morphology of the superior temporal sulcus. *NeuroImage*, *22*(2), 706-719. doi:10.1016/j.neuroimage.2004.01.023

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early human development*, *18*(1), 27–36.

Pelphrey, K. A., Morris, J., Michelich, C., Allison, T., & McCarthy, G. (2005). Functional Anatomy of Biological Motion Perception in Posterior Temporal Cortex: An fMRI Study of Eye, Mouth and Hand Movements. *Cerebral Cortex*, *15*(12), 1866 -1876. doi:10.1093/cercor/bhi064

Petkov, C., Kayser, C., Steudel, T., Whittingstall, K., Augath, M. & Logothetis, N.K. (2008) A voice region in the monkey brain. *Nature Neuroscience, 11,* 167-174.

Rauschecker, J.P. & Scott, S.K. (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*, 718-725.

Raznahan, A., Toro, R., Daly, E., Robertson, D., Murphy, C., Deeley, Q., Bolton, P. F., et al. (2010). Cortical Anatomy in Autism Spectrum Disorder: An In Vivo MRI Study on the Effect of Age. *Cerebral Cortex*, *20*(6), 1332 -1340. doi:10.1093/cercor/bhp198

Salamon, G., Raynaud, C., Regis, J., & Rumeau, C. (1990). *Magnetic resonance imaging of the pediatric brain*. New York, NY: Raven Press.

Sakatani, K., Chen, S., Lichty, W., Zuo, H., & Wang, Y. P. (1999). Cerebral blood oxygenation changes induced by auditory stimulation in newborn infants measured by near infrared spectroscopy. *Early Human Development*, *55*(3), 229-236.

Scott, S. (2008) Voice processing in monkey and human brains, *Trends in Cognitive Sciences, 12,* 323-325.

Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *Neuroimage*, *20*(4), 1944–1954.

Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., & Obrig, H. (2006). Illuminating the BOLD signal: combined fMRI-fNIRS studies. *Magnetic Resonance Imaging*, *24*(4), 495-505. doi:10.1016/j.mri.2005.12.034

Thierry, G., Giraud, A. L., & Price, C. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron*, *38*(3), 499–506.

Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *NeuroImage*, *28*(3), 635-662. doi:10.1016/j.neuroimage.2005.06.058

Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, *1*(2), 185–195.

Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, *11*(5), 665–674.

Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, *20*(10), 435-442. doi:10.1016/S0166-2236(97)01132-6.

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. *Cognitive Brain Research, 17*(1), 48-55.

Table 1: The results from the t-test channel-by-channel analysis across the two auditory experimental conditions For each contrast the results for the significant increase in $HbO_2$ and/or decrease in HHb concentration are displayed. Please note that the degrees of freedom may vary across channels as individual infants may have invalid data in some channels (this is especially the case for the larger channel separations – 4.5cm – where the signal is weaker in some individuals). *ROI - approximate cortical region of each channel (Ch): IFr – inferior frontal cortex, aST – anterior portion of the STS region, and pST – posterior portion of the STS region – (see Fig.1 for location of channels on a schematic infant head).*

| *$HbO_2$ (uM)* | | | | *HHb (uM)* | | | | *$HbO_2$ (uM)* | | | | *HHb (uM)* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ch** | *t* | *p* | *ROI* | **Ch** | *t* | *p* | *ROI* | **Ch** | *t* | *p* | *ROI* | **Ch** | *t* | *p* | *ROI* |
| **Voice Condition > Silence Condition** | | | | | | | | | | | | | | | |
| Left Hemisphere | | | | | | | | Right Hemisphere | | | | | | | |
| **9** | 4.59 | 0.00007 | aST | **8** | 4.56 | 0.00009 | aST | **28** | 2.85 | 0.008 | aST | **21** | 2.44 | 0.02 | IFr |
| | | | | **9** | 2.22 | 0.034 | aST | **32** | 4.29 | 0.0002 | aST | **28** | 5.66 | < 0.00001 | aST |
| | | | | **10** | 2.37 | 0.024 | aST | **33** | 2.09 | 0.045 | pST | **29** | 3.32 | 0.002 | aST |
| | | | | **13** | 3.2 | 0.003 | aST | **38** | 2.58 | 0.032 | pST | **33** | 3.75 | 0.0008 | pST |
| **Non-Voice Condition > Silence Condition** | | | | | | | | | | | | | | | |
| Left Hemisphere | | | | | | | | Right Hemisphere | | | | | | | |
| **2** | 2.92 | 0.006 | IFr | **1** | 2.72 | 0.016 | IFr | **28** | 3.21 | 0.003 | aST | **23** | 2.05 | 0.048 | IFr |
| **9** | 4.05 | 0.0004 | aST | **9** | 2.57 | 0.01 | aST | **32** | 6.65 | < 0.00001 | aST | **28** | 6.15 | < 0.00001 | aST |
| **11** | 11.8 | 0.001 | IFr | **13** | 6.37 | < 0.00001 | aST | **33** | 3.52 | 0.001 | pST | **29** | 2.69 | 0.011 | aST |
| **13** | 2.85 | 0.008 | aST | **15** | 4.31 | 0.0002 | pST | **34** | 2.86 | 0.007 | pST | **32** | 3.84 | 0.0005 | aST |
| | | | | | | | | **38** | 2.58 | 0.037 | pST | **33** | 4.61 | 0.00006 | pST |
| | | | | | | | | | | | | **38** | 3.01 | 0.02 | pST |
| **Voice Condition > Non-Voice Condition** | | | | | | | | | | | | | | | |
| Left Hemisphere | | | | | | | | Right Hemisphere | | | | | | | |
| | | | | **8** | 2.08 | 0.047 | aST | | | | | | | | |
| | | | | **10** | 2.28 | 0.03 | aST | | | | | | | | |
| **Non-Voice Condition > Voice Condition** | | | | | | | | | | | | | | | |
| Left Hemisphere | | | | | | | | Right Hemisphere | | | | | | | |
| | | | | **13** | 2.24 | 0.032 | aST | **32** | 2.42 | 0.021 | aST | | | | |
| | | | | **15** | 4.15 | 0.0002 | pST | **34** | 2.43 | 0.02 | pST | | | | |

**Figures**

Figure 1: Illustrations of the procedure used in this experiment. Upper panel: The experimental design showing the order and timing of stimulus presentation. Lower panel: A schematic infant head showing the approximate location of the two arrays and channels there within. Though the exact underlying anatomical brain structures may vary across individuals, this positioning allowed recording from inferior frontal cortex (IFr), anterior portions of the superior temporal region (aST) and posterior portions of the superior temporal region (pST), which are labelled on the figure. Note that in addition to the 2cm channels shown on this schematic there are 4.5cm channels located at a lower depth underneath this set of channels at the following locations; under channel 5 are channels 3 and 11; under channel 9 are channels 7 and 16; under channel 14 are channels 12 and 16; under channel 24 are channels 22 and 30; under channel 28 are channels 26 and 35; and under channel 33 are channels 31 and 38.

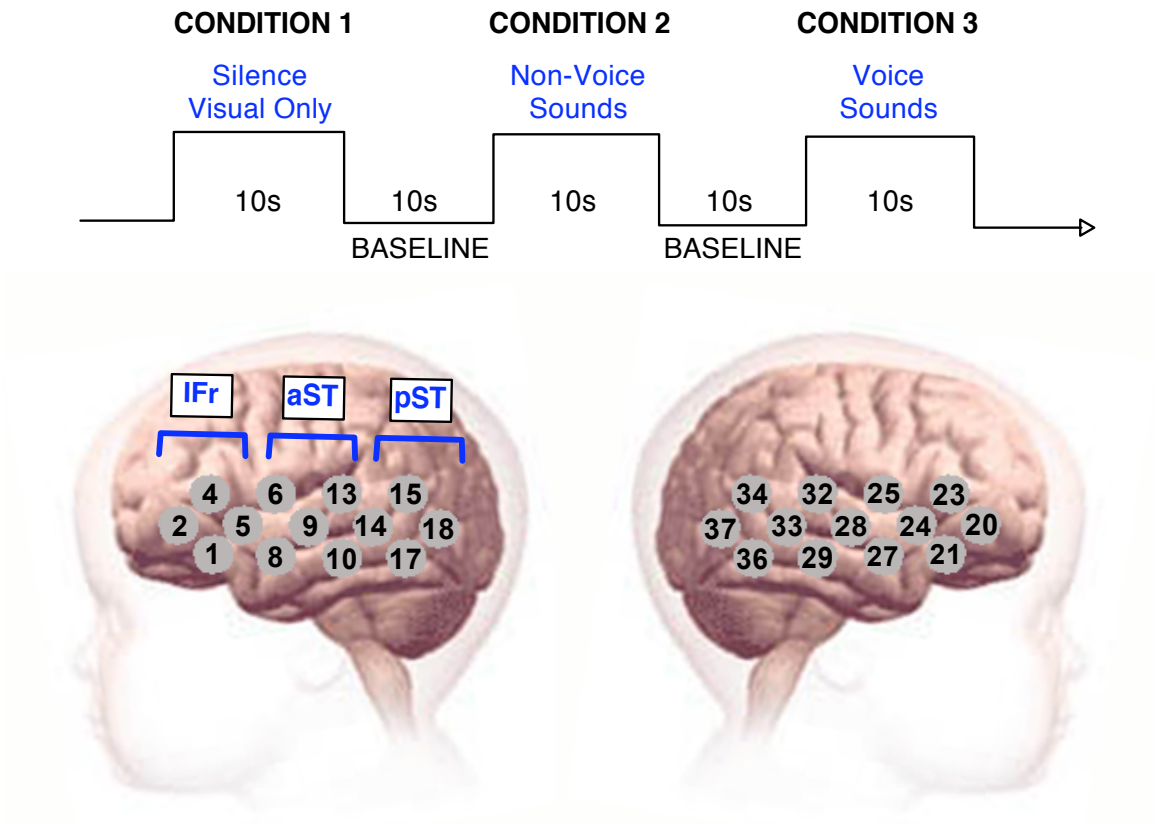


Figure 2: A schematic of the infant head showing the statistically significant effects ($p<0.05$) in the channel-by-channel t-test analysis for: upper panel - voice condition > silence condition; middle panel - non-voice condition > silence condition; and lower panel - voice condition versus non-voice condition. In the upper two panels, the channels that revealed a significant response during the specified time window of activation are plotted in red (increase in $HbO_2$ concentration) and blue (decrease in HHb concentration). In the lower panel, the channels that revealed a significantly greater response (either $HbO_2$ or HHb) to one condition relative to the other are plotted in *orange* for the voice > non-voice contrast and in *green* for the non-voice > voice contrast. Channels are plotted following the same layout as in Fig. 1. Note that the schematic is used for illustrative purposes only and does not imply mapping of channels onto brain anatomy.

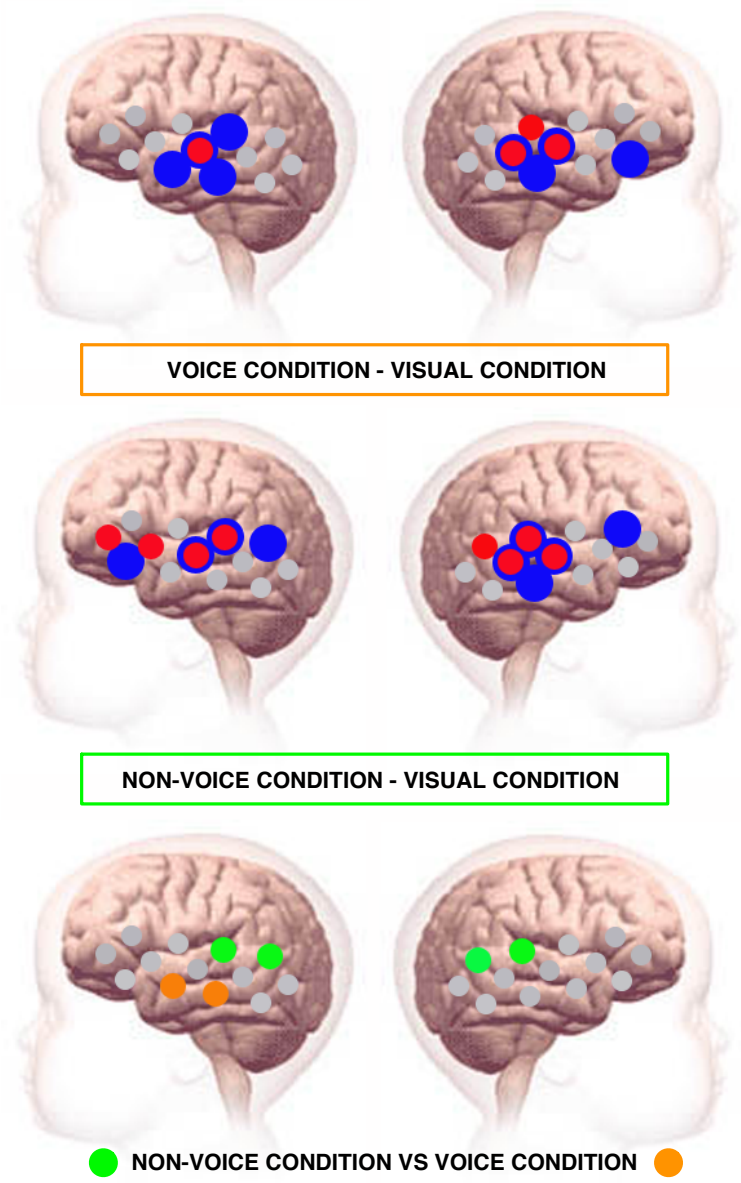
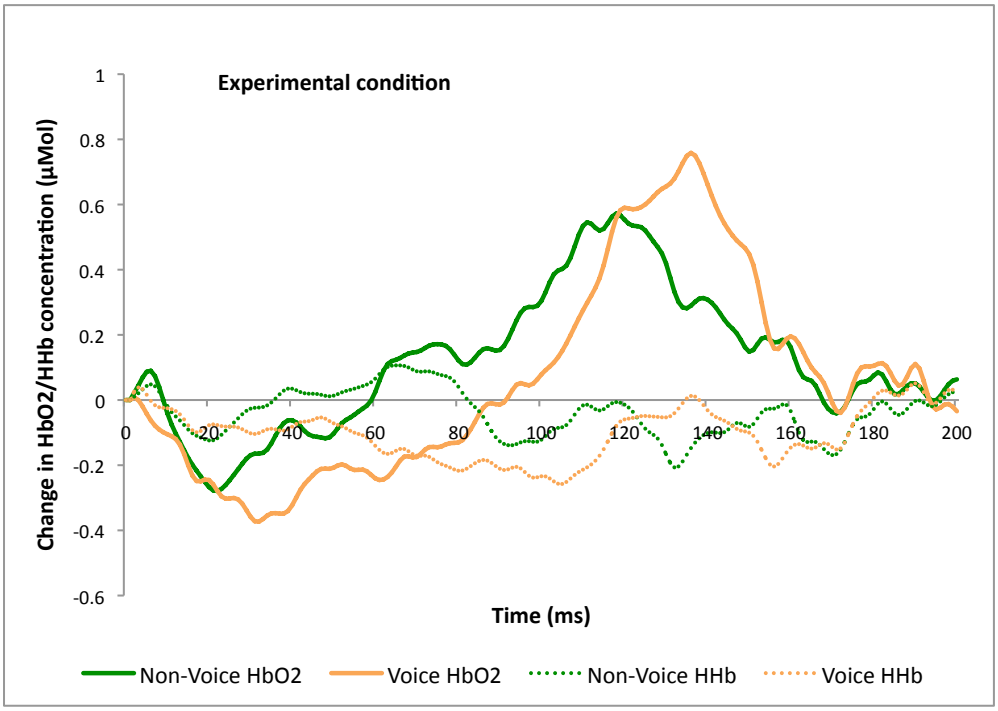


Figure 3: An example of the grand averaged haemodynamic time courses of the group results. The plots show the response to the voice and non-voice stimuli in channel 8 and channel 13 in the left hemisphere.



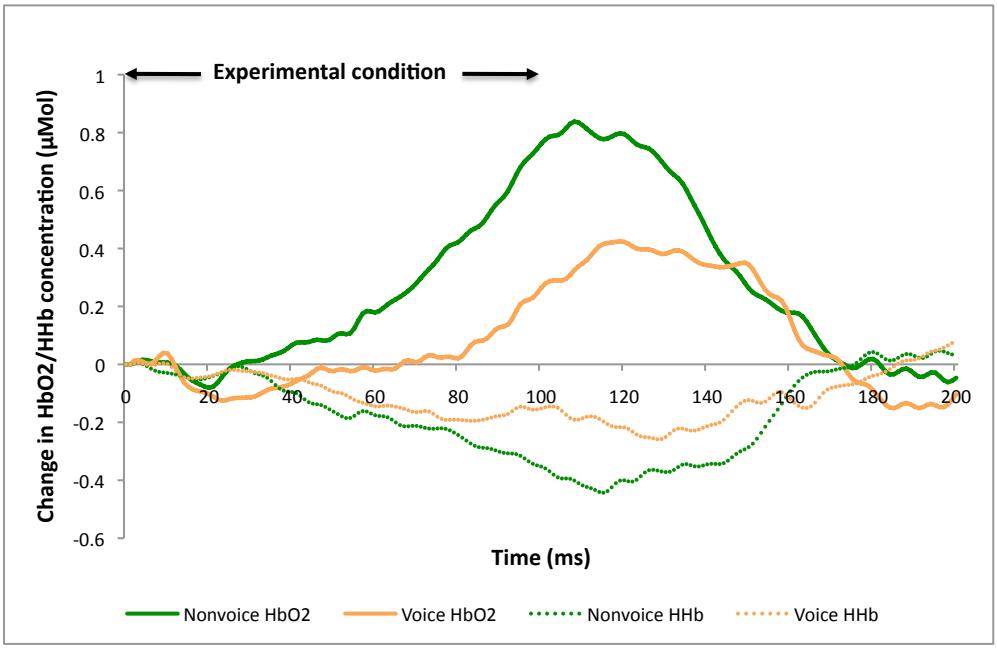



Figure 4: The upper (channel 8) and lower (channel 27) panels on the right show plots of individual infants' voice selective amplitude in HHb (uMol) – as measured by the subtraction of the *non-voice* peak haemodynamic response from the *voice* peak haemodynamic response – compared with their age. The schematics on the left show the position of this voice selective response within the lateral arrays.

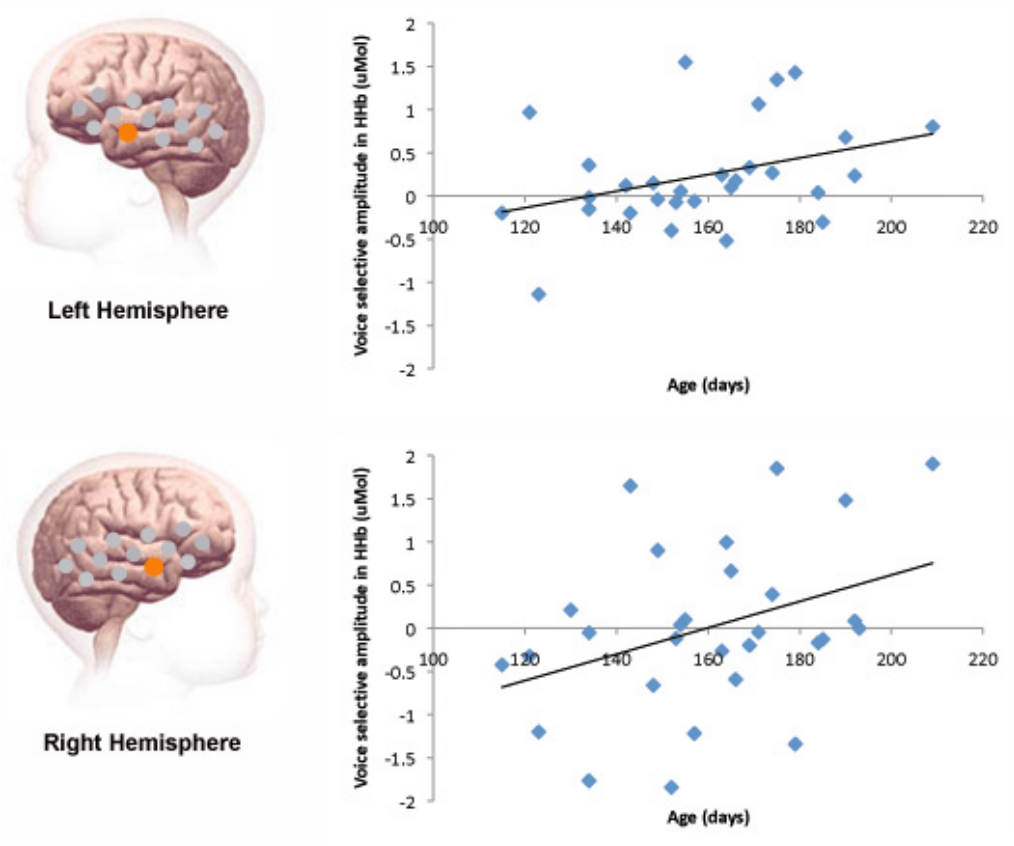


Figure 5: This schematic illustrates the approximate location of the areas of maximum *voice selective* (**orange**) and *non-voice selective* (**green**) activation across four voice-processing studies (current findings (*circle*); Belin et al., 2000 (*square*); Blasi et al., 2011(*triangle*); Grossman et al., 2010 (*diamond*)). The shapes containing a smaller white triangle highlight age-related changes in voice-selective activation. For the MRI studies, this was achieved by using a T2T converter *(wwwneuro03.uni-muenster.de/ger/t2tconv/),* which can project the coordinates of the activation from MRI to the surface of a schematic brain. To overlay the approximate location of the results from the NIRS studies we used the 10-20 system component (yellow grid) of the T2T converter. The location of the NIRS channels in relation to the 10-20 coordinates were calculated using the head dimensions of an average 4 – 7 mth old infant (see *Methods* for head measurements from the current participants in relation to the location of our NIRS channels). The position of the channels showing significant *voice and non-voice selective* activation were then superimposed onto the schematic in relation to the 10-20 coordinates.