NeuroImage 133 (2016) 144–150

Contents lists available at ScienceDirect

NeuroImage

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

Full Length Articles

The neural correlates of processing scale-invariant environmental sounds at birth

Judit Gervain [a,b,*], Janet F. Werker [c], Alexis Black [c], Maria N. Geffen [d]

[a] Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France
[b] Laboratoire Psychologie de la Perception, CNRS, Paris, France
[c] Department of Psychology, University of British Columbia, Vancouver, Canada
[d] Department of Otorhinolaryngology – Head and Neck Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

ARTICLE INFO

Article history:
Received 25 August 2015
Accepted 1 March 2016
Available online 5 March 2016

Keywords:
Auditory perception
Scale-invariance
Efficient neural coding
Newborns
Near-infrared spectroscopy

ABSTRACT

Sensory systems are thought to have evolved to efficiently represent the full range of sensory stimuli encountered in the natural world. The statistics of natural environmental sounds are characterized by scale-invariance: the property of exhibiting similar patterns at different levels of observation. The statistical structure of scale-invariant sounds remains constant at different spectro-temporal scales. Scale-invariance plays a fundamental role in how efficiently animals and human adults perceive acoustic signals. However, the developmental origins and brain correlates of the neural encoding of scale-invariant environmental sounds remain unexplored. Here, we investigate whether the human brain extracts the statistical property of scale-invariance. Synthetic sounds generated by a mathematical model to respect scale-invariance or violate it were presented to newborns. In alternating blocks, the two sound types were presented together in an alternating fashion, whereas in non-alternating blocks, only one type of sound was presented. Newborns' brain responses were measured using near-infrared spectroscopy. We found that scale-invariant and variable-scale sounds were discriminated by the newborn brain, as suggested by differential activation in the left frontal and temporal areas to alternating vs. non-alternating blocks. These results indicate that newborns already detect and encode scale-invariance as a characteristic feature of acoustic stimuli. This suggests that the mathematical principle of efficient coding of information guides the auditory neural code from the beginning of human development, a finding that may help explain how evolution has prepared the brain for perceiving the natural world.

Introduction

To what extent is the human brain prepared at birth to process the natural world? Natural signals such as natural environmental sounds possess characteristic statistical regularities in their structure. Accurately representing and encoding these regularities is an essential function of our perceptual systems. One such important statistical regularity of environmental sounds is scale-invariant spectro-temporal structure, i.e. the property of exhibiting similar structures or patterns at different levels of observation.

The statistical structure of scale-invariant sounds remains constant at multiple spectro-temporal scales. This feature has been identified not only for environmental sounds but also for music and speech (Voss and Clarke, 1975) and is thought to be fundamental to how the mammalian auditory system perceives acoustic signals (Pallier et al., 1998; Smith and Lewicki, 2006). At the level of neuronal responses measured in the adult animal brain, it has been shown that the auditory system encodes sounds that possess scale-invariant features more efficiently than sounds that lack this structure (Escabi et al., 2003; Rieke et al., 1995; Woolley et al., 2005). Water sounds exhibit scale-invariance, which humans perceive as an attribute of natural sounds. We constructed a generative mathematical model to create artificial sounds that either obeyed scale-invariant spectro-temporal structure or lacked this relationship. We found that human adults perceived the artificial sounds that obeyed scale-invariant statistical structure as natural, but judged those that lack it as unnatural (Geffen et al., 2011). These results point to the importance of scale-invariant spectro-temporal properties in the perception of a sound as natural. We were therefore interested in understanding whether and how sensitivity to the scale-invariant spectro-temporal structure of sounds emerges throughout development.

Young infants have sophisticated auditory abilities that support detection, processing, categorization, and learning of complex sounds (Moore, 2002; Saffran et al., 2006; Werner et al., 2012). Sounds of ecological importance, such as human speech, are processed with particular efficiency (Gervain and Mehler, 2010; Jusczyk, 1981; Werker and Curtin, 2005). Infants' auditory perception extends beyond speech to other sounds such as music or human action sounds (e.g. Trehub and Hannon, 2006; Geangu et al., 2015), and, as we recently demonstrated

* Corresponding author at: Laboratoire Psychologie de la Perception (UMR 8158), CNRS-Université Paris Descartes, 45 rue des Saints-Pères, 75006, Paris, France.
E-mail address: judit.gervain@parisdescartes.fr (J. Gervain).

http://dx.doi.org/10.1016/j.neuroimage.2016.03.001
1053-8119/

in a behavioral study, to natural sounds obeying scale-invariant spectro-temporal structure. We found that infants perceived scale-invariant water sounds generated by a mathematical model (Geffen et al., 2011) differentially as compared to variable-scale sounds, and grouped scale-invariant, but not variable-scale sounds, into one perceptual category (Gervain et al., 2014). This discrimination ability was not simply driven by increased familiarity with scale-invariant sounds, as in a control study using two categories of scale-invariant sounds (less and more familiar, as rated by adults), infants showed no preference for the more familiar sounds. This study provided the first evidence that by 5 months of age, infants can distinguish between sounds that possess and those that lack scale-invariant spectro-temporal structure at the behavioral level. Whether this ability is present even earlier in development, and if so, what is its neural substrate, is as yet unexplored. Here, we hypothesized that the differential sensitivity to scale-invariant and variable-scale sounds is driven by differences in neuronal correlates involved in processing these two classes of sounds in the infant brain from birth on.

Materials and methods

To test our hypothesis, we performed an optical brain imaging study with newborns, who were presented with scale-invariant and variable-scale water sounds. Previously, we developed a 3-parameter mathematical model (Eq. (1)) that generates sounds that are classified as "water" by adult observers if they are scale-invariant, and as a variety of other, non-natural sounds (e.g. robot sounds, grease sizzling, hissing, etc.) if they are variable-scale (Geffen et al., 2011). These sounds consisted of a population of randomly spaced gamma tone chirps from a wide range of frequencies (Geffen et al., 2011; and Fig. 1A). The gammatone transform is widely used to approximate the transformation of a sound into spectral bands at the cochlear stage (Goblick and Pfeiffer, 1969; Depireux et al., 2001). Each chirp was characterized by its frequency, amplitude, and cycle constant of decay (Eq. 1).

$$G_n(t) = \int_{\tau=0}^{\infty} (t-\tau)e^{-f_n\tau/Q}\sin(2\tau f_n\tau)y(t-\tau)d\tau \quad (1)$$

where $y(t)$ is the signal, G_n is the gammatone transform in frequency band n, f is the center frequency, τ is the delay time, and Q is the bandwidth or cycle constant of decay.

This allowed us to directly manipulate whether the water sounds generated by the model respected scale-invariance across spectral bands or not. Scale-invariant sounds were generated when the temporal structure of the chirps scaled relative to their center frequency. For variable-scale sounds, the chirps in different spectral bands varied in their temporal structure relative to their center frequency (Fig. 1). We used near-infrared spectroscopy to test whether the newborn brain is already able to discriminate between these artificially generated scale-invariant and variable-scale water sounds, similarly to adults (Geffen et al., 2011) and 5-month-old infants (Gervain et al., 2014).

Participants

Twenty-two healthy, full-term neonates (9 females; mean age 1.73 days, range 0–3 days; Apgar score ≥ 8) born in the Vancouver area participated. Data from 8 additional infants were collected, but excluded from the data analysis as they (i) failed to complete the experiment due to fussiness and crying (5), or (ii) provided poor quality data due to large motion artifacts or thick hair (3). All parents gave informed consent prior to participation. The Ethics Boards of the University of British Columbia and BC Women's Hospital, where the experiments took place, granted permission.

Material

Stimuli in the scale-invariant and variable-scale categories were generated using our computational model of water sounds (Geffen et al., 2011; and Fig. 1A). Sounds in both categories were matched for the frequency range, sound pressure level (amplitude root mean square), amplitude and timing parameters of the chirps. Chirps were gamma tone functions with parameters amplitude, frequency f, onset time, and cycle constant of decay Q drawn randomly from distinct probability distributions.

The only difference between the sounds in the two categories was the relation between Q and f. For both types of sounds, the distribution

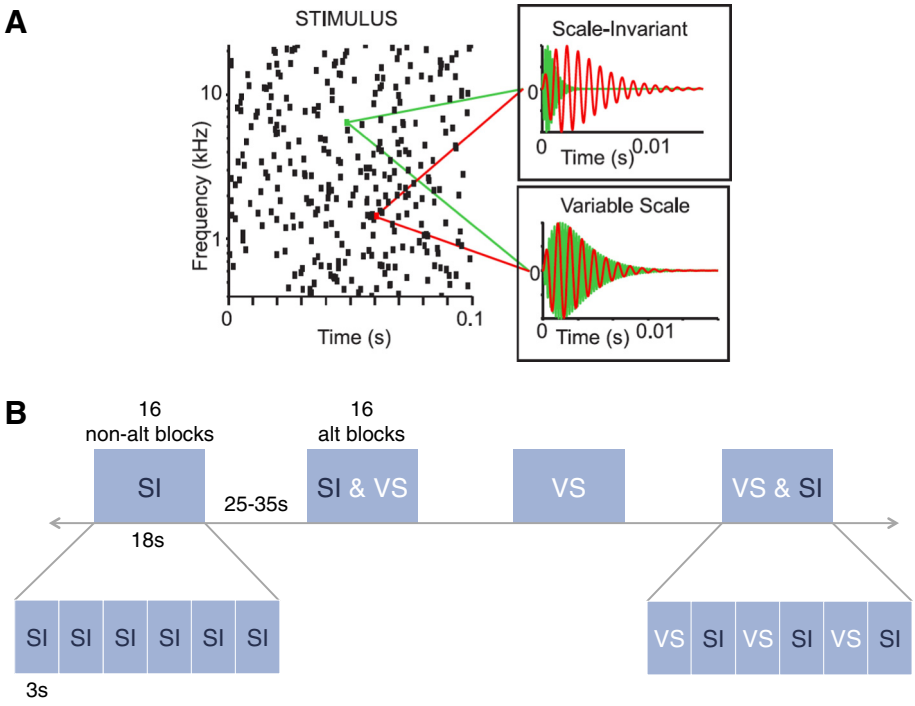

Fig. 1. A. The generative model used to synthesize the sound stimuli presented in the current study. B. The experimental design. SI: scale-invariant; VS: variable-scale.

of f was uniform random in log-frequency space between 400 and 20,000 Hz. The number of chirps per second was determined by the rate r, which varied between 53, 530, and 5340 chirp/octave/second. The distribution for the onset of the chirps was uniform random, while the amplitude was drawn from an inverse-uniform distribution. For scale-invariant sounds, Q varied randomly between 1, 2, 3.1, and 4. By contrast, for variable-scale sounds, Q scaled proportionally to f ($Q = 0.1 \times f$, $Q = 0.01 \times f$, $Q = 0.001 \times f$ or $Q = 0.005 \times f$). The parameter values for both categories were based on Geffen et al. (2011) and Gervain et al. (2014).

The above parameters were used to generate 3 s-long sound chunks, each with its own Q and r, for both categories. Sequences of 6 chunks were then continuously concatenated to produce 18 s-long sound stimulation blocks. Two types of blocks were created (Fig. 1B). Non-alternating or homogeneous blocks contained 6 chunks from the same sound category, whereas alternating or mixed blocks contained 3 chunks from the scale-invariant and 3 from the variable-scale category, concatenated in strict alternation. The alternating/non-alternating design is commonly used in behavioral studies with infants to uncover fine-grained discrimination abilities (Best and Jones, 1998; Maye et al., 2002). The rationale of this design is the following. If participants cannot discriminate the two stimulus categories, then alternating and non-alternating blocks do not sound different to them. If, by contrast, they can tell the stimulus categories apart, the two types of blocks sound different. This design has now been successfully employed in fNIRS studies with infants to probe fine-grained discrimination abilities (e.g. Gervain et al., 2012; Sato et al., 2010). If the NIRS responses to alternating and non-alternating blocks are different, then it can be concluded that the brain discriminates between the two stimulus categories.

Alternating and non-alternating blocks were presented in alternation. The experiment comprised 16 non-alternating blocks, half scale-invariant, half variable-scale, and 16 alternating blocks, half of which started with a scale-invariant chunk, the other half with a variable-scale chunk. Altogether, 80 different alternating and 80 different non-alternating blocks were created. The type of the first block (non-alternating scale-invariant, non-alternating variable-scale, alternating scale-invariant, or alternating variable-scale first) was randomized and counterbalanced across infants. Blocks were separated by pauses varying between 25 and 35 s in duration to avoid inducing phase-locked brain responses.

The mean and standard deviation over time of the sound pressure level for the alternating and non-alternating stimuli were verified with a handheld sound pressure level meter (Extech 407,735, set to A-type frequency weighting) positioned approximately at the level of the infant ear. Ten alternating blocks, twelve non-alternating scale-invariant, and twelve non-alternating variable-scale blocks were measured at 24 time intervals within the duration of each block. There was no difference between the mean in sound pressure level for the alternating (mean 69.08 dB sound pressure level relative to 20μPa (SPL), standard deviation 0.66 dB) and non-alternating (mean 68.89 dB SPL, standard deviation 0.71 dB) blocks ($t(32) = -0.722$, $p = 0.476$). Nor was there a difference between the two types of non-alternating blocks (SI: 68.99 dB SPL, SD = 0.57; VS: 68.79 dB, SD = 0.84; $t(22) = 0.672$, $p = 0.508$). These objective measures of intensity make it likely, but do not guarantee that the two types of stimuli were subjectively perceived by infants as having the same loudness.

Procedure

Infants were tested with a HITACHI ETG-4000 NIRS machine (source-detector separation: 3 cm; two continuous wavelengths of 695 nm and 830 nm) in a dimly lit sound-attenuated room at BC Women's Hospital, lying in their cribs throughout the 25-min-long test session. At least one parent was present at all times. Babies were tested while in a state of quiet rest or sleep. The NIRS optical probes were placed on infants' heads bilaterally (12 channels per hemisphere; Fig. 2) using the tragus, the vertex, and the ears as surface landmarks (Gervain et al., 2012, 2008; Peña et al., 2003).

Sound stimuli were administered through two loudspeakers positioned at a distance of 1.5 m from the baby's head, at an angle of 30°, elevated to the same height as the infant's crib. A Macintosh computer played the stimuli and operated the NIRS machine, running PsyScope experimental software. The NIRS machine used 0.7 mW laser power.

Data processing and analysis

Changes in the concentration of oxygenated hemoglobin (oxyHb) and deoxygenated hemoglobin (deoxyHb) were calculated from the absorption of near-infrared light as metabolic indicators of neural activity. OxyHb and deoxyHb were entered into the data analysis.

To eliminate high-frequency noises (heartbeat etc.) and overall trends, the data were band pass filtered between 0.01 and 0.7 Hz. Movement artifacts, defined as concentration changes larger than 0.1 mmol*mm over 0.2 s, were removed by rejecting block-channel pairs where artifacts occurred. For the non-rejected blocks, a baseline was linearly fitted between the means of the 5 s preceding the onset of the block and the 5 s starting 33 s after the onset of the block (18 s of stimulation plus 15 s of resting period). The entire data set of an infant was removed from the analysis if the infant contributed less than 30% non-rejected blocks (3 infants for the alternating vs. non-alternating statistical comparison and 1 additional infant for the scale-invariant vs. variable-scale comparison, see below). We averaged the concentration change of oxyHb or deoxyHb across all blocks of each condition in each channel for each participant over a 25 s time window starting at the onset of stimulation to ensure that we capture that peak response, which may have quite long latencies in young infants (Gervain et al., 2011).

We analyzed the data statistically by conducting channel-by-channel t-tests comparing each block type to baseline as well as the two types of blocks to each other. The results were corrected for multiple comparisons using the False Discovery Rate (Benjamini and Hochberg, 1995) method. We report both uncorrected and corrected results, if at least the former is significant (reported as $p < 0.05$). If the uncorrected result is significant, but the corrected one is marginal, we report exact p values for the marginal result. To further confirm the results obtained by the t-tests, we also performed analyses of variance (ANOVAs) over specific regions of interest (ROIs). Two ANOVAs were conducted: one comparing the Alternating and non-alternating blocks, and one comparing the scale-invariant and variable-scale blocks within the non-alternating block types. The first ANOVA allowed us to address the main question of the study, i.e. whether the newborn brain distinguishes between scale-invariant and variable-scale environmental sounds. The second ANOVA was conducted to further probe into how the two sound categories were processed, directly comparing the localization and amplitude of the responses to the two types of sounds. As this latter comparison includes half as many blocks as the former, i.e. only the non-alternating ones, 1 neonate from the 22 participants included in the former analysis did not contribute sufficient data toward this latter analysis and was therefore not included in it. Importantly, as NIRS measures metabolic correlates of neural activity, i.e. changes in the concentration of oxyHb and deoxyHb, and not neural activity directly, not finding a significant difference in the direct comparison of the non-alternating blocks for the two conditions does not mean that the brain cannot distinguish the two conditions. Rather, it suggests that the two conditions require the same amount of metabolic resources in the same brain areas to be processed, independently of whether they can be discriminated or not.

The ROIs (Fig. 2A) were defined on the basis of previous studies using similar probe placement, experimental design, and setup (Peña et al., 2003; Gervain et al., 2012), and were confirmed by the channel-by-channel t-test results. Accordingly, the channels of interest were the anterior channels, i.e. channels 1–6 in the LH and channels 13–17 and 19 in the RH, following Gervain et al. (2008, 2012). Furthemore,

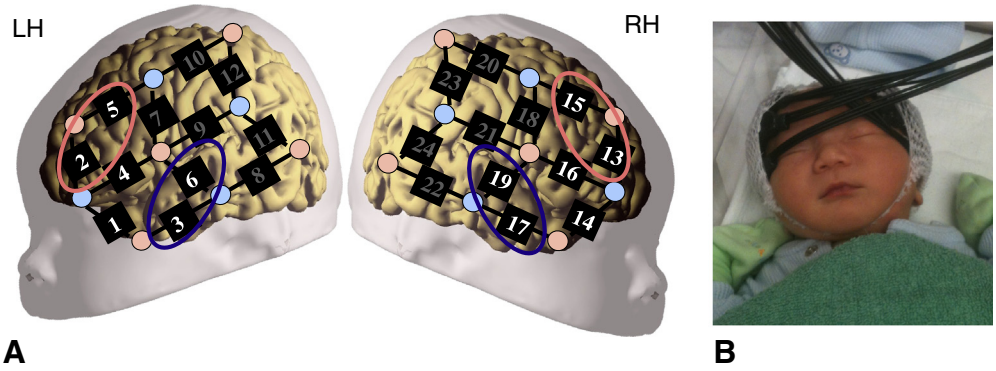

Fig. 2. Probe placement. A. The schematic illustration of the NIRS channels placed bilaterally on the participants' heads. Red dots indicate light sources, blue dots light detectors. Channels are coupled source-detector pairs, indicated by the numbered black squares (channels of interest in white, other channels in grey). The red ovals indicate the two frontal ROIs, the blue ovals the two temporal ROIs. B. Newborn participant wearing the NIRS headgear.

for the ANOVA comparing alternating vs. non-alternating blocks, channels 3 and 6 (left hemisphere, LH) and channels 17 and 19 (RH) comprised the ROIs covering the bilateral temporal areas, known to be responsible for auditory processing in adults and infants (Dehaene-Lambertz et al., 2002, 2008; Friederici et al., 2002; Peña et al., 2003). Channels 2 and 5 (LH) and channels 13 and 15 (RH) made up the bilateral frontal ROIs, documented to be involved in higher order sequence learning and, in particular, discrimination in alternating/non-alternating paradigms (Gervain et al., 2012). For the ANOVA comparing the scale-invariant and variable-scale non-alternating blocks, the ROIs were somewhat enlarged to compensate for the greater variability in the data due to the smaller number of trials that could be included. Thus, channels 1, 3, 6, and 8 were included in the left temporal ROI, channels 2, 4, 5, and 7 were included in the left frontal ROI, channels 14, 17, 19, and 22 were included in the right temporal ROI and channels 13, 15, 16, and 18 were included in the right frontal ROI. The analyses were also performed with the more restricted ROIs used in the first ANOVA, and the pattern of results were similar, but more babies had missing data points.

All statistical analyses were conducted on oxyHb as well as on deoxyHb. When unreported, the analyses for deoxyHb were conducted, but did not reach significance.

Results

Alternating vs. non-alternating blocks

The oxyHb and deoxyHb concentrations were averaged across blocks separately for the alternating and non-alternating conditions. The resulting grand average across all infants is shown in Fig. 3.

As a first analysis, we conducted channel-by-channel t-tests comparing the concentration changes of oxyHb and deoxyHb against baseline for each block type separately as well as comparing the two types of blocks to each other. As compared to baseline (Fig. 4), we observed significant activation in channel 3 with oxyHb ($p < 0.01$ uncorrected, $p = 0.1$ after False Discovery Rate (Benjamini and Hochberg, 1995) correction for multiple comparisons) and channels 3 and 13 with deoxyHb for the alternating blocks ($p < 0.05$, uncorrected, n.s. after FDR correction) and in channels 2, 3, 5 ($p < 0.05$ uncorrected as well as after FDR correction) and 17 ($p < 0.05$ uncorrected, n.s. after FDR correction) with oxyHb and channels 1, 2, 3, 5, 13, 16, 17, and 19 with deoxyHb for the non-alternating blocks ($p < 0.05$ uncorrected as well as after FDR correction). When directly comparing the two types of blocks (Fig. 3), we obtained significantly higher activation in channel 2 for the non-alternating than for the alternating blocks with oxyHb ($t(21) = 3.061$, $p = 0.006$ uncorrected and $p = 0.071$ after FDR correction).

In a second analysis, we performed a repeated-measures ANOVA with factors block type (Alt/Non-Alt), hemisphere (LH/RH), and ROI (temporal/frontal) using oxyHb as the dependent variable. We obtained a main effect of Hemisphere ($F(1,21) = 6.19$, $p = 0.014$) due to greater overall activation in the left than in the right hemisphere, as well as a main effect of ROI ($F(1,21) = 9.64$, $p = 0.002$) due to greater overall activation in the temporal than in the frontal ROIs. Furthermore, the interaction block type × hemisphere × ROI was significant ($F(1,21) = 4.87$, $p = 0.029$). Scheffe's post hoc tests revealed that this interaction was carried by greater activation in the left temporal than in the left frontal ROI for the alternating blocks ($p = 0.002$), by greater activation in the left frontal ROI for the non-alternating than for the alternating blocks ($p = 0.001$), by greater activation for the non-alternating blocks in the left than in the right frontal ROI ($p = 0.002$), and by greater activation in the temporal than in the frontal ROI in the RH for non-alternating blocks. A similar ANOVA with deoxyHb yielded a main effect of block type due to greater overall activation, i.e. greater decrease in deoxyHb concentration change, for the non-alternating blocks ($F(1,21) = 8.08$, $p = 0.005$).

Scale-invariant vs. variable-scale non-alternating blocks

The oxyHb and deoxyHb concentrations were averaged across blocks separately for the scale-invariant and variable-scale blocks within the non-alternating condition. The resulting grand average across all infants is shown in Fig. 5.

As a first analysis, we performed channel-by-channel t-tests against baseline separately for the two conditions as well as comparing the two conditions. The t-tests against baseline (Fig. 6) yielded significant activation in channels 2 and 3 with oxyHb ($p < 0.05$ uncorrected and $p = 0.059$ after FDR correction) and in channels 1, 2, 3, 4, 13, 14, 16, 17, and 19 with deoxyHb for the scale-invariant blocks ($p < 0.05$ uncorrected as well as after FDR correction); and in channel 3 with oxyHb ($p < 0.05$ uncorrected, n.s. after FDR correction) and in channels 5 and 13 with deoxyHb for the variable-scale blocks ($p < 0.05$ uncorrected, n.s. after FDR correction). We obtained no channelwise differences in activation when directly comparing the two types of blocks.

In a second analysis, we performed a repeated-measures ANOVA with factors block type (scale-invariant/variable-scale), Hemisphere (LH/RH), and ROI (temporal/frontal) using oxyHb as the dependent variable. We obtained a marginally significant main effect of Hemisphere due to the greater involvement of the LH ($F(1,20) = 2.93$, $p = 0.089$). A similar ANOVA with deoxyHb yielded no significant main effects or interactions.

Discussion

In this study, we tested how the newborn brain processes scale-invariance in complex sounds that replicate the statistical structure of

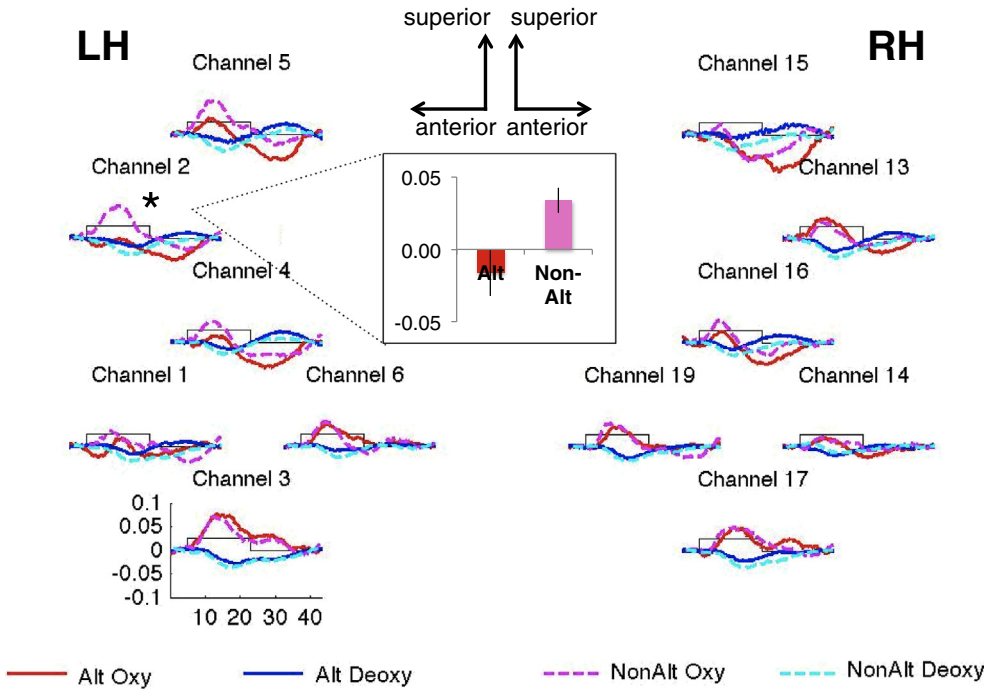

Fig. 3. The grand average hemodynamic response obtained in the experiment for the alternating/non-alternating blocks. Channels are plotted following the probe placement illustrated in Fig. 2A. The x-axis represents time in seconds; the y-axis shows concentration in mmol × mm. The rectangle along the x-axis indicates the time of stimulation. OxyHb and deoxyHb concentration changes in response to the alternating blocks are shown in red and blue, respectively (continuous line). OxyHb and deoxyHb concentration changes in response to the non-alternating blocks are shown in magenta and cyan, respectively (dashed line). The asterisk indicates the channel that showed significantly greater activation to non-alternating as compared to alternating blocks. The inset represents the oxyHb concentration change averaged over the analysis time window in channel 2 in the two conditions with error bars showing the standard error of the mean.

a class of natural sounds. Using NIRS, we measured newborns' brain responses to alternating and non-alternating presentation blocks of synthesized water sounds that either respect scale-invariance or not, as defined by the relation between the temporal and frequency parameters of chirps in a mathematical model that generated the sounds.

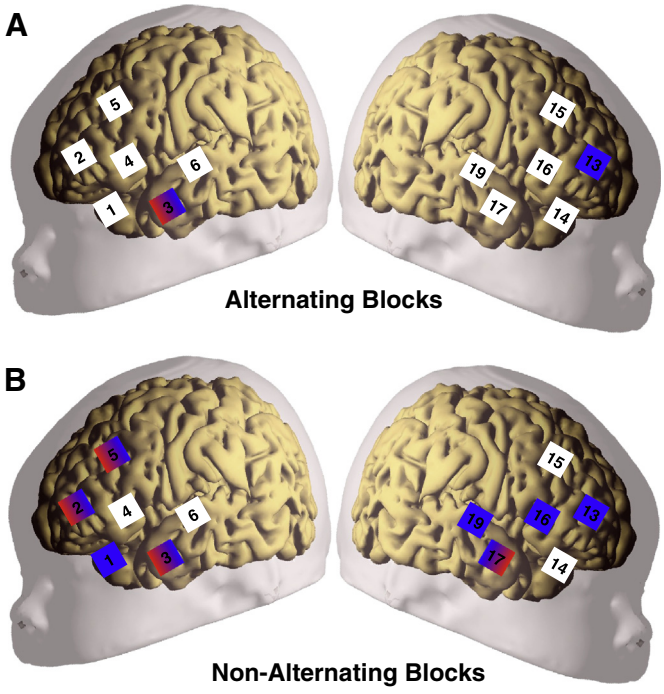

Fig. 4. T-test comparison against baseline for the alternating (A) and non-alternating (B) blocks. Squares represent channels. Channels in red show statistically significant activation with respect to baseline in oxyHb, those in blue for deoxyHb and those with a red-blue color gradient for both oxyHb and deoxyHb.

We found that the left frontal and temporal areas of the newborn brain were differentially activated by alternating and non-alternating blocks of scale-invariant and variable-scale water sounds. It needs to be noted that while the SI and VS stimuli were generated in the same way and have similar sound pressure levels, it cannot be excluded given our experimental design that newborns might still have perceived the two sound categories as having different subjective loudness. Nevertheless, the difference between scale-invariant and variable-scale water sounds was observed both in channel-by-channel comparisons and in an overall ANOVA. These results strengthen the possibility that the newborn brain encodes the statistical property of scale-invariance in natural stimuli, potentially allowing for categorization of scale-invariant sounds that we observed behaviorally in older infants.

The observed localization of the responses to the left hemisphere is consistent with adult studies which have demonstrated that fast changing or temporally modulated auditory events preferentially engage the left temporal areas (Hickok and Poeppel, 2007; Zatorre and Belin, 2001). Indeed, our stimuli contained prominent fast modulation components. More importantly for the present study, we found greater differential activation in left frontal regions, possibly involving Broca's area, to the non-alternating blocks. The responses in this area indicate discrimination between scale-invariant and variable-scale water sounds. The direction of this differential response, i.e. a larger response to the non-alternating as opposed to the alternating blocks, is consistent with previous results. Using the same study design and NIRS setup in newborns but using language sounds (Gervain et al., 2012), we previously found greater activation for the block type from which a clear perceptual category or regularity can be extracted, i.e. the non-alternating blocks.

Additionally, the direct comparison of scale-invariant and variable-scale non-alternating blocks indicates that the two types of sounds are generally processed in a similar fashion and in the same brain areas,

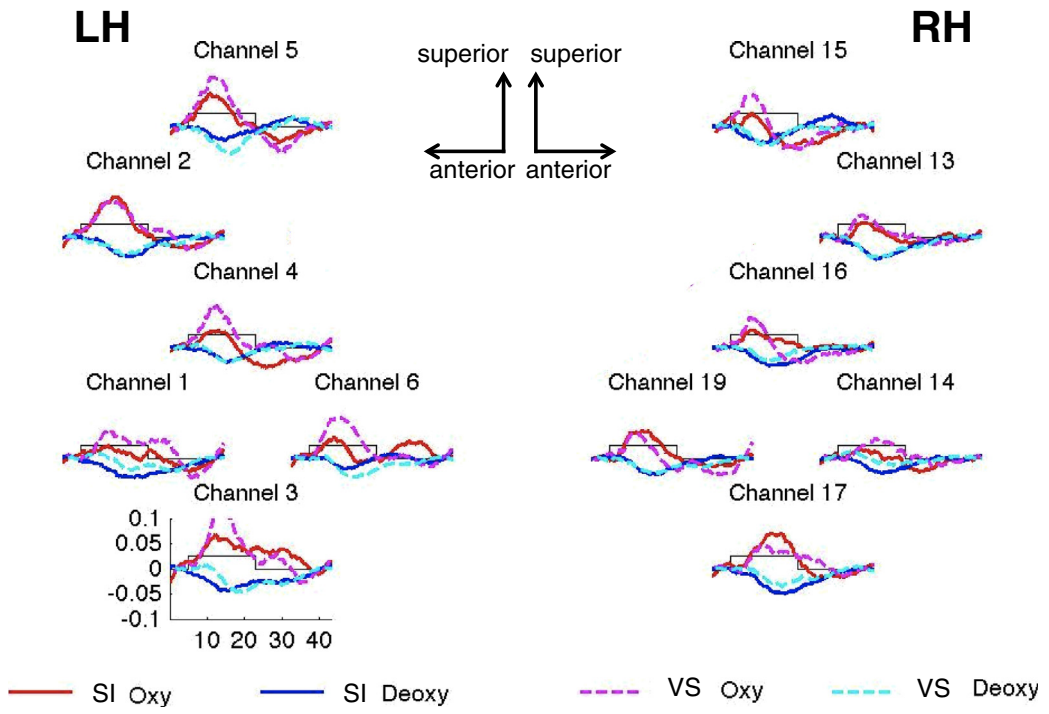

Fig. 5. The grand average hemodynamic response obtained in the experiment for the scale-invariant and variable-scale non-alternating blocks. Plotting conventions are the same as for Fig. 3. SI: scale-invariant, VS: variable-scale.

as no significant differences emerged in the direct channel-by-channel comparisons and the overall ANOVA. This latter analysis, however, confirms the left hemisphere advantage found in the alternating/non-alternating comparison, in line with the proposal of the left hemisphere's specialization for fast temporally modulated sounds. Additionally, the greater number of channels that showed significant activation as compared to baseline for the scale-invariant as compared to the variable-scale sounds points in the direction of the perceptual asymmetry found in our previous behavioral study with 5-month-olds (Gervain et al., 2014), whereby infants readily discriminated scale-invariant and variable-scale water sounds when they were habituated to the natural category, i.e. to scale-invariant water sounds, but not if they were habituated to variable-scale sounds, which do not serve as an appropriate basis to anchor the natural perceptual category of water sounds.

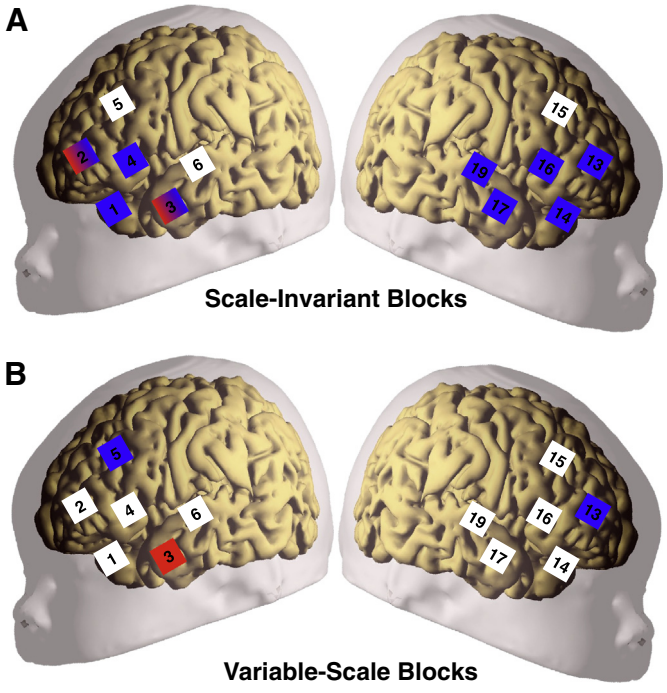

Fig. 6. T-test comparison against baseline for the scale-invariant (A) and variable-scale (B) non-alternating blocks. Plotting conventions are the same as for Fig. 4.

The efficient neural coding hypothesis argues that the mammalian sensory systems have evolved to encode environmental stimuli at an information theoretical optimum, capturing the greatest amount of information at the lowest possible cost (Attneave, 1954; Barlow, 1961; Simoncelli and Olshausen, 2001). To achieve this, the sensory systems need to extract the statistical structure of environmental stimuli—an information theoretical prediction that has been confirmed for the visual system (Olshausen and Field, 2004; Simoncelli and Olshausen, 2001) and more recently for the auditory system (Lewicki, 2002; Ming and Holt, 2009; Smith and Lewicki, 2006). Electrophysiologically measured responses from the visual system and auditory nerve fibers of animal models have been shown to closely match the statistical properties of natural signals, efficiently encoding scale-invariant stimuli (Olshausen and Field, 2004; Smith and Lewicki, 2006). The drive to encode natural stimuli with high efficiency might have given rise to the observed sensitivity for scale-invariance in the infant brain.

For several classes of natural acoustic signals, our auditory system displays invariance to temporal stretching or compression. Adults and children are able to adapt to and comprehend accelerated speech in their native language as well as in rhythmically similar languages (Banai and Lavner, 2012; Guiraud et al., 2013; Orchik and Oelschlaeger, 1977; Pallier et al., 1998; Sebastián-Gallés et al., 2000). Adults also readily categorize accelerated or decelerated recordings of water sounds as being natural (Geffen et al., 2011). Because water sounds are characterized by a scale-invariant temporal structure, their temporal distortions do not take them outside the natural statistical boundaries. It is plausible that the same principle might underlie invariance in perception to temporally stretched or compressed speech: despite large changes in temporal structure, the distorted speech still falls into the statistical range to which our brain is sensitive. This suggests that the property of

acoustically invariant structure in the temporal domain might be shared between speech and other natural sounds.

In summary, we showed in this study that the newborn brain detects scale-invariance in natural sounds. These results extend previous research indicating that the mammalian brain has evolved to process naturally occurring sounds that have this scale-invariant property, by demonstrating that such encoding is seen in the human brain, and from the first days of postnatal life. These results thus advance theories of evolution by showing that the human brain is prepared, early in life, to process natural sounds as special signals. The fact that scale-invariant natural sounds activate the same neural regions that are activated in response to speech in the neonate raises important questions for further study regarding the characterization of speech, and the relation between the processing of natural environmental sounds to the processing of communicative sounds.

Acknowledgments

This research was supported by a HFSP Young Researcher Grant nr. RGY0073/2014 to MNG and JG, an Emergence(s) grant from the City of Paris, an ANR grant nr. ANR-15-CE37-0009-01, and a French National Research Agency (ANR) "Investissements d'Avenir" grant nr. ANR-10-LABX-0083 to JG, Natural Sciences and Engineering Research Council DG-81103, Canadian Foundation for Innovation, and the Canadian Institute for Advanced Research to JFW, The Burroughs Wellcome Award at the Scientific Interface, the Klingenstein Award in the Neurosciences, NIH R01DC014479 award, Pennsylvania Lions Club Hearing Research Fellowship and the NIH R03DC013660 award to MNG. The authors declare no competing financial interests.

References

Attneave, F., 1954. Some informational aspects of visual perception. Psychol. Rev. 61 (3), 183.
Banai, K., Lavner, Y., 2012. Perceptual learning of time-compressed speech: more than rapid adaptation. PLoS One 7 (10), e47099.
Barlow, H.B., 1961. Possible principles underlying the transformation of sensory messages. Sens. Commun. 217–234.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57 (1), 289–300.
Best, C., Jones, C., 1998. Stimulus-alternation preference procedure to test infant speech discrimination. Infant Behav. Dev. 21, 295-295.
Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L., 2002. Functional neuroimaging of speech perception in infants. Science 298 (5600), 2013–2015.
Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Dehaene, S., 2008. How does early brain organization promote language acquisition in humans? Eur. Rev. 16 (04), 399–411.
Depireux, D.A., Simon, J.Z., Klein, D.J., Shamma, S.A., 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J. Neurophysiol. 85 (3), 1220–1234.
Escabi, M.A., Miller, L.M., Read, H.L., Schreiner, C.E., 2003. Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. J. Neurosci. 23 (37), 11489–11504.
Friederici, A.D., Steinhauer, K., Pfeifer, E., 2002. Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. Proc. Natl. Acad. Sci. 99 (1), 529–534.
Geangu, E., Quadrelli, E., Lewis, J.W., Cassia, V.M., Turati, C., 2015. By the sound of it: an ERP investigation of human action sound processing in 7-month-old infants. Dev. Cogn. Neurosci. 12, 134–144.
Geffen, M.N., Gervain, J., Werker, J.F., Magnasco, M.O., 2011. Auditory perception of self-similarity in water sounds. Front. Integr. Neurosci. 5.
Gervain, J., Mehler, J., 2010. Speech perception and language acquisition in the first year of life. Annu. Rev. Psychol. 61, 191–218.
Gervain, J., Berent, I., Werker, J.F., 2012. Binding at birth: the newborn brain detects identity relations and sequential position in speech. J. Cogn. Neurosci. 24 (3), 564–574.
Gervain, J., Macagno, F., Cogoi, S., Peña, M., Mehler, J., 2008. The neonate brain detects speech structure. Proc. Natl. Acad. Sci. 105 (37), 14222–14227.
Gervain, J., Werker, J.F., Geffen, M.N., 2014. Category-specific processing of scale-invariant sounds in infancy. PLoS One 9 (5), e96278.
Gervain, J., Mehler, J., Werker, J.F., Nelson, C.A., Csibra, G., Lloyd-Fox, S., Shukla, M., Aslin, R.N., 2011. "Near-Infrared Spectroscopy: A report from the McDonnell infant methodology consortium". Developmental Cognitive Neuroscience 1 (1), 22–46.
Goblick Jr., T.J., Pfeiffer, R.R., 1969. Time-domain measurements of cochlear nonlinearities using combination click stimuli. J. Acoust. Soc. Am. 46 (4B), 924–938.
Guiraud, H., Ferragne, E., Bedoin, N., Boulenger, V., 2013. Adaptation to natural fast speech and time-compressed speech in children. Proceedings from INTERSPEECH 2013: *the 14th Annual Conference of the International Speech Communication Association*, pp. 1370–1374.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402.
Jusczyk, P.W., 1981. The processing of speech and nonspeech sounds by infants: Some implications. In: Aslin, R., Alberts, J., Petersen, M. (Eds.), Development of perception: Psychobiological perspectives vol. 1. Academic Press, New York, pp. 191–216.
Lewicki, M.S., 2002. Efficient coding of natural sounds. Nat. Neurosci. 5 (4), 356–363.
Maye, J., Werker, J.F., Gerken, L.A., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82, B101–B111.
Ming, V.L., Holt, L.L., 2009. Efficient coding in human auditory perception. J. Acoust. Soc. Am. 126 (3), 1312–1320.
Moore, D.R., 2002. Auditory development and the role of experience. Br. Med. Bull. 63 (1), 171–181.
Olshausen, B.A., Field, D.J., 2004. Sparse coding of sensory inputs. Curr. Opin. Neurobiol. 14 (4), 481–487.
Orchik, D.J., Oelschlaeger, M.L., 1977. Time-compressed speech discrimination in children and its relationship to articulation. Ear Hear. 3 (1), 37–41.
Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., Mehler, J., 1998. Perceptual adjustment to time-compressed speech: A cross-linguistic study. Mem. Cogn. 26 (4), 844–851.
Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., Mehler, J., 2003. Sounds and silence: An optical topography study of language recognition at birth. Proc. Natl. Acad. Sci. U. S. A. 100 (20), 11702–11705.
Rieke, F., Bodnar, D.A., Bialek, W., 1995. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proc. R. Soc. Lond. Ser. B Biol. Sci. 262 (1365), 259–265.
Saffran, J.R., Werker, J.F., Werner, L.A., 2006. The infant's auditory world: hearing, speech, and the beginnings of language. In: Siegler, R., Kuhn, D. (Eds.), Handbook of Child Development vol. 2. John Wiley & Sons Inc., Hoboken, NJ, US, pp. 58–108.
Sato, Y., Sogabe, Y., Mazuka, R., 2010. Development of hemispheric specialization for lexical pitch-accent in Japanese infants. J. Cogn. Neurosci. 22 (11), 2503–2513.
Sebastián-Gallés, N., Dupoux, E., Costa, A., Mehler, J., 2000. Adaptation to time-compressed speech: phonological determinants. Percept. Psychophys. 62 (4), 834–842.
Simoncelli, E.P., Olshausen, B.A., 2001. Natural image statistics and neural representation. Annu. Rev. Neurosci. 24 (1), 1193–1216.
Smith, E.C., Lewicki, M.S., 2006. Efficient auditory coding. Nature 439 (7079), 978–982.
Trehub, S., Hannon, E., 2006. Infant music perception: domain-general or domain-specific mechanisms? Cognition 100, 73–99.
Voss, R.F., Clarke, J., 1975. '1/*f*noise' in music and speech. Nature 258, 317–318.
Werker, J.F., Curtin, S., 2005. PRIMIR: a developmental framework of infant speech processing. Lang. Learn. Dev. 1 (2), 197–234.
Werner, L., Fay, R.R., Popper, A.N., 2012. Human auditory development vol. 42. Springer New York.
Woolley, S.M., Fremouw, T.E., Hsu, A., Theunissen, F.E., 2005. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat. Neurosci. 8 (10), 1371–1379.
Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11 (10), 946–953.