When Do Visual Category Representations Emerge in Infants' Brains?

Xiaoqian Yan[1,2,3*], Sarah Tung[1+], Bella Fascendini[1+], Yulan Diana Chen[1,2], Anthony M. Norcia[1,2#], Kalanit Grill-Spector[1,2,4#]

1. Department of Psychology, Stanford University, Stanford, CA 94305, USA
2. Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA 94305, USA
3. Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China
4. Neurosciences Program, Stanford University, Stanford, CA 94305, USA

* corresponding author

\# co-senior-author

\+ equal contribution

Key words: Infant, object categorization, EEG

Abstract

Organizing the continuous stream of visual input into categories like places or faces is important for everyday function and social interactions. However, it is unknown when neural representations of these and other visual categories emerge. Here we used steady state evoked potential electroencephalography to measure cortical responses in infants at 3-4 months, 4-6 months, 6-8 months, and 12-15 months, when they viewed controlled, gray-level images of faces, limbs, corridors, characters, and cars. We found that distinct responses to these categories emerge at different ages. Reliable brain responses to faces emerge first, at 4-6 months, followed by limbs and places around 6-8 months. Between 6-15 months response patterns become more distinct, such that a classifier can decode what an infant is looking at from their brain responses. These findings, have important implications for assessing typical and atypical cortical development as they not only suggest that category representations are learned, but also that representations of categories that may have innate substrates emerge at different times during infancy.

Introduction

Visual categorization is important for everyday activities and is amazingly rapid: adults categorize the visual input in about one-tenth of a second[1,2]. In adults and school-age children, this key behavior is supported by both clustered and distributed responses to visual categories in high-level visual cortex in ventral and lateral temporal cortex (VTC and LOTC, respectively)[3,4]. A visual category consists of items that share common visual features and configurations[2,5–8]; e.g., corridors share features of floors, walls, and ceilings, with a typical spatial relationship. Clustered regions in VTC and LOTC[4,9,10] respond more strongly to items of ecologically-relevant categories (faces, bodies, places, words) than other stimuli[5,11–16] and distributed neural responses across VTC and LOTC[4,9,10] are reliable across items of a category but distinct across items of different categories. However, it is unknown when these visual category representations emerge in infants' brains.

Behaviorally, infants can perform some level of visual categorization within the first year of life. Measurements of infants' looking preferences and looking times suggest that visual saliency impacts young infants' viewing patterns[17] : between 4-10 months of age, infants can behaviorally distinguish between faces and objects[17,18] and between different animals like cats and dogs[19,20]. Later on, between 10-19 months, infants behaviorally distinguish broader-level animate versus inanimate categories[17]. Neurally, electroencephalographic (EEG) studies have found stronger responses to images of faces vs. objects or textures in 4-12-month-olds[21–23] and that stimulus category can be decoded from distributed responses slightly but significantly above chance in 6-15-month-olds[24,25]. Functional magnetic resonance imaging (fMRI) studies have found stronger responses to videos of faces[26,27], bodies[27], and places[26,27] vs. objects in clustered regions in VTC and LOTC of 2-10-month-olds. However, because prior studies used different types of stimuli and age ranges, it is unknown when representations to various categories emerge during the first year of life. To address this key open question, we examined when neural representations to different visual categories emerge during infancy using EEG in infants of 4 age groups spanning 3-15 months of age.

We considered two main hypotheses regarding the developmental trajectories of category representations. One possibility is that representations to multiple categories emerge together because infants need to organize the barrage of visual input to understand what they see. Supporting this hypothesis are findings of (i) selective responses to faces, places, and body parts in VTC and LOTC of 2-10-month-olds[27], and (ii) above chance classification of distributed EEG responses to toys, bodies, faces, houses in 6-8 month-olds[25] as well as animals and body parts in 12-15-month-olds[24].

Another possibility is that representations of different categories may emerge at different times during infancy. This may be due to two reasons. First, representations of ecologically-relevant categories like faces, body parts, and places may be innate because of their evolutionary importance[28–31], whereas representations for other categories, may develop later only with learning[5,10,32]. Supporting this hypothesis are findings that newborns and young infants tend to orient to faces[33] and face-like stimuli[34], as well as have cortical responses to face-like stimuli[35], but word representations only emerge in childhood with the onset of reading instruction[5,10,36]. Second, even if visual experience is necessary for the development category representations (including faces[37–40]) categories that are seen more frequently earlier in infancy may develop before others. Measurements using head mounted cameras suggest that infants' visual diet (composition of visual input) varies across categories and age: The visual diet of 0-3-month-olds contains ~25% faces and <10% hands, that of 12-15-month-olds contains ~20% faces and ~20% hands [41,42], and that of 24-month-olds contains ~10% faces, and ~25% hands. Thus, looking behavior in infants predicts that representations of faces may emerge before that of limbs.

<insert Fig 1 here>

Results

45 infants from four age groups: 3-4 months (n = 17, 7 females), 4-6 months (n = 14, 7 females), 6-8 months (n = 15, 6 females), and 12-15 months (n = 15, 4 females) participated in EEG experiments. Twelve participants were part of an ongoing longitudinal study and came for several sessions spanning at least 3 months apart. Infants viewed gray-scale images from five visual categories present in infants' environments (faces, limbs, corridors, characters, and cars) while EEG was recorded. Different from prior infant studies[21,23–27], we used images that have been widely used in fMRI studies[5,7,43–45] and are largely controlled for low-level properties such as luminance, contrast, similarity, and spatial frequency (**Fig 1B** and supplementary **Table S3**). We use a Steady-State Visual Evoked Potential[23,46–48] (SSVEP) paradigm: In each 70-s sequence, images from five categories were shown every 0.233-s; one of the categories was the target, so different images from that category appeared every 1.167-s, and the rest of the images were drawn from the other four categories in a random order (**Fig 1A**). Images of all categories appeared at equal probability and no images were repeated[8]. Infants participants in 5 conditions, which varied by the target category. We used the EEG-SSVEP approach because: (i) it affords a high signal-to-noise ratio with short acquisitions making it effective for infants [23,46], (ii) it has been successfully used to study responses to faces in infants[23,46,49], and (iii) it enables measuring both general visual response to images by examining responses at the image presentation frequency (4.286 Hz), as well as category-selective responses by examining responses at the category frequency (0.857 Hz, **Fig 1A**).

As the EEG-SSVEP paradigm is novel and we are restricted in the amount of data we can obtain in infants, we first tested if we can use this paradigm and a similar amount of data to detect category-selective responses in adults. Results show that in adults we can (i) reliably measure category-selective responses (Supplementary **Figs S1-S2**) and (ii) decode category information from distributed spatiotemporal response patterns (Supplementary **Figs S3**) with the same amount of data that we obtained in infants. This validates the SSVEP paradigm for measuring category-selectivity. As infants have lower cortical visual acuity, we also tested if the stimuli are distinguishable to infants. Thus, we simulated how they may look to infants by filtering the images to match the cortical acuity of 3-month-olds Supplementary **Fig S4**). Despite being blurry, images of different categories are readily distinguishable by adults (Supplementary **Movies 1-5**), suggesting that there is sufficient visual information in the lower spatial frequencies of the stimuli for infants to distinguish visual categories.

Robust visual responses in occipital regions to visual stimuli in all infant age groups

We first tested if there are significant visual responses to our stimuli in infants' brains by evaluating the amplitude of responses at the image presentation frequency (4.286 Hz) and its first three harmonics. We found that in all age groups, visual responses were concentrated spatially over occipital electrodes (**Fig 2A-D, left panel, Figure S1A**). Quantification of the mean visual response amplitude over a region of interest (ROI) spanning 9 electrodes over early visual cortex revealed significant responses in all infant age groups at the image frequency and its first 3 harmonics (response amplitudes significantly above zero with FDR corrected at 4 levels; except for the first harmonic at 8.571 Hz in 6-8-month-olds; **Fig 2A-D, right panel**). We also tested if experimental noise varied across age groups. Noise level was estimated in the occipital electrodes by measuring the amplitude of response in frequencies up to 8.571 Hz excluding image presentation frequencies (4.286 Hz and harmonics) and category frequencies (0.857 Hz and harmonics) as this frequency range includes the relevant main harmonics. We found no significant difference in noise across age groups (**Fig 2E**). These analyses indicate that infants were looking at the stimuli as there are significant visual responses even in the youngest 3-4 months-old infants' and there are no significant differences in noise levels across infants of different ages.

<insert Fig 2 here>

Prior EEG data[21,50] suggest that the timing and waveform of visual responses may vary across development. To complement the frequency domain analysis, we transformed the responses at image frequency and its harmonics to the time domain using an inverse Fourier transformation for two reasons. First, the time-domain provides access to information about response timing and waveform that is not

directly accessible from an analysis of responses of individual harmonics. Second the total visual response is better reflected in the time-domain as the individual harmonic amplitudes can sum constructively.

We observed that during the 233-ms image presentation, temporal waveforms had two deflections in 3-4-month-olds (one negativity and one positivity, **Fig 2F**) and four deflections for infants older than 4 months (two minima and two maxima, **Fig 2F**). To evaluate developmental effects, we examined the latency and amplitude of the peak visual response during two time windows related to the first deflection (60-90 ms), and second deflection (90-160 ms for 3-4-month-olds, and 90-110 ms for other age groups). In general, we find that the latency of the peak deflection decreased from 3 to 15 months (**Fig 2G, H**). As data includes both cross-sectional and longitudinal measurements and we observed larger development earlier than later in infancy, we used a linear mixed model (LMM) to model peak latency as a function of the logarithm of age (see Methods). Results reveal that the latency of the peak deflection significantly and differentially decreased with age in the two time-windows ($\beta_{\text{age x time window}}$ = -45.78, 95% CI: -58.39 – -33.17, $t_{(118)}$ = -7.19, $p = 6.39 \times 10^{-11}$; LMM with age and time window as fixed effects, and participant as a random effect, all stats in **Supplementary Table S4-S5**). There were larger decreases in the peak latency in the second than first time window (**Fig 2G, H**, first: β_{age} = -7.44, 95% CI: -13.82 – -1.06, $t_{(118)}$ = -2.33, $p_{FDR} < .05$; second: β_{age} = -46.91, 95% CI: -56.56 – -37.27, $t_{(59)}$ = -9.73, $p_{FDR} < .001$). Peak amplitude also differentially develops across the two windows ($\beta_{\text{age x time window}}$ = -4.90, 95% CI: -8.66 – -1.14, $t_{(118)}$ = -2.58, p = .01, Supplementary **Table S6-S7**). The decrease in peak amplitude with age was significant only for the second deflection (β_{age} = -3.59, 95% CI: -6.38 – -0.81, $t_{(59)}$ = -2.58, p = .01, LMM). These data suggest that the temporal dynamics of visual responses over occipital cortex develop from 3 to 15 months of age.

What is the nature of category-selective responses in infants?

We next examined if in addition to visual responses to the rapid image stream, there are also category-selective responses in infants, by evaluating the amplitude of responses at the category frequency (0.857 Hz) and its harmonics. This is a selective response as it reflects the relative response to images of category above the general visual response. **Fig 3** shows the spatial distribution and amplitude of the mean category response for faces and its harmonics in each age group. Mean category-selective responses to limbs, cars, corridors, and words are shown in Supplementary **Figs S5-S8**. We analyzed mean responses over two ROIs spanning 7 electrodes each over left and right occipitotemporal cortex where high-level visual regions are located[51].

<insert Fig 3 here>

We found significant group-level category responses to some but not all categories and a differential development of category-selective responses during infancy. The largest and earliest developing category-selective responses were to faces. In contrast to visual responses, which were centered over occipital electrodes (**Fig 2A-D,** left panel), significant categorical responses to faces (at 0.857 Hz and its first harmonic, 1.714 Hz) were observed over lateral occipitotemporal electrodes (**Fig 3A-D,** left panel). Notably, there were significant responses to faces over bilateral occipitotemporal electrodes in 4-6-month-olds at 0.857 Hz (**Fig 3B,** response amplitudes significantly above zero with Hotelling's T2 statistic, $p_{FDR} < .05$, FDR corrected over 2 levels: the category frequency and its first harmonic), as well as 6-8-month-olds and 12-15-month-olds at the category frequency and its first harmonic (**Fig 3C,D,** both $p_{FDR} < .05$). However, there were no significant responses to faces in 3-4-month-olds at either the category frequency or its harmonics (**Fig 3A, right panel**). These data suggest that face-selective responses start to reliably emerge over lateral occipitotemporal cortex between 4 and 6 months of age.

We did not find significant group-level category-selective responses that survived FDR correction to any of the other categories before 6 months of age (Supplementary **Fig S8,** except for a weak but statistically significant response for cars in the right occipitotemporal ROI in 3-4-month-olds). Instead, we found significant category-selective responses that survived FDR correction for (i) limbs in 6-8-month-olds in the right occipitotemporal ROI (Supplementary **Fig S5**), (ii) corridors in 6-8-month-olds and 12-15 months-old in the left occipitotemporal ROI (Supplementary **Fig S6**), and (iii) characters in 6-8-month-olds in the right occipitotemporal ROI, and in 12-15-month-olds in bilateral occipitotemporal ROI (Supplementary **Fig S7**).

We evaluated the temporal dynamics of category-selective waveforms by transforming the data at the category frequency and its harmonics to the time domain. This analysis was done separately for each of the left and right occipitotemporal ROIs for each category and age group. Consistent with frequency domain analyses, average temporal waveforms over lateral-occipital ROIs show significant responses to faces that emerge at ~4 months of age (**Fig 4A,** significant responses relative to zero, cluster-based nonparametric permutation 10,000 times, two-tailed t-test, at $p < .05$). The temporal waveforms of responses to faces in infants show an initial positive deflection peaking ~500 ms after stimulus onset followed by a negative deflection peaking at ~900 ms. Notably, mean waveforms associated with limbs, corridors, and characters in lateral occipital-ROIs, are different from faces: there is only a single negative defection which peaks at ~500 ms after stimulus onset, which is significant only in 6-8 and 12-15-month-olds (**Fig 4B-D**). There was no significant category response to cars in infants except for a late (~1000 ms) positive response in 4-6-month-olds (**Fig 4E**). These results show that both the timing and waveform differ

across categories, which suggests that there might be additional category information in the distributed spatiotemporal response.

<insert Fig 4 here>

We next examined the development of the peak response and latency of the category waveforms separately for the right and left lateral occipitotemporal ROIs. We found significant development of the peak response in the right lateral occipitotemporal ROI which varied by category ($\beta_{category\ x\ age}$ = -1.09, 95% CI – -2.00 – -0.14, $t_{(301)}$ = -2.26, p_{FDR} < .05, LMM as a function of log (age) and category; participant: random effect). Post-hoc analyses revealed that the peak response for faces significantly increased from 3 to 15 months (**Fig 4A-right,** β_{age} = 7.27, 95% CI: 4.03 – 10.51, $t_{(59)}$ = 4.50, p_{FDR} < .05, LMM as a function of log(age); participant: random effect) and the peak response for limbs significantly decreased (**Fig 4B-right,** β_{age} = -2.90, 95% CI: -5.41 – -0.38, $t_{(59)}$ = -2.31, p = .02, not significant after FDR correction over 5 category levels). There were no other significant developments of peak amplitude (Supplementary **Tables S8-S9**).

Additionally, for all categories, the latency of the peak response in the right occipitotemporal ROI significantly decreased from 3 to 15 months of age (β_{age} = -173.17, 95% CI: -284.73 – -61.61, $t_{(301)}$ = -3.05, p = .002, LMM as a function of log(age) and category; participant: random effect). We found no significant development of peak latency in the left occipitotemporal ROI (Supplementary **Tables S8-9**).

Are spatiotemporal patterns of responses to visual categories consistent across infants?

As we observed different mean waveforms over the lateral occipital ROIs for the five categories (**Fig 4**), we asked whether the distributed spatiotemporal patterns of brain responses evoked by each category are unique and reliable across infants. We reasoned that if different categories generated consistent distributed spatiotemporal responses, an independent classifier would be able to predict the category an infant was viewing from their distributed spatiotemporal pattern of response. Thus, we used a leave-one-out-cross-validation (LOOCV) approach (**Fig 5A**) and tested if a classifier can decode the category a left-out infant viewed based on the similarity of their distributed spatiotemporal response to the mean response to each of the categories in the remaining N-1 infants. We calculated for each infant the mean category-waveform (same as **Fig 4**) across the occipital and lateral occipitotemporal ROIs and concatenated the waveforms across the three ROIs to generate the distributed spatiotemporal response to a category (**Fig 5A**). The classifier was trained and tested separately for each age group.

Results reveal two main findings. First, the LOOCV classifier decoded category information from brain responses significantly above the 20% chance level in infants aged 6 months and older but not in younger

infants (**Fig 5B**, 6-8-month-olds, significant above chance: $t_{(14)} = 4.1$, $p_{FDR} < .01$, one-tailed, FDR corrected over 4 age groups; 12-15-month-olds, $t_{(14)} = 3.4$, $p_{FDR} < .01$). This suggests that spatiotemporal patterns of responses to different categories become reliable across infants after 6 months of age. Second, examination of classification by category, shows that the LOOCV classifier successfully determined from spatiotemporal responses when infants were viewing faces in 64% of 4-6-month-olds, in 93% of 6-8-month-olds, and 87% of 12-15-month-olds (**Fig 5C**). In contrast, classification performance was numerically lower for the other categories (successful classification in less than 40% of the infants). This suggests that a reliable spatiotemporal response to faces that is consistent across infants develops after 4 months of age and dominates classification performance.

<insert Fig 5 here>

What is the nature of categorical spatiotemporal patterns in individual infants?

While the prior analyses leverage the power of averaging across electrodes and infants, this averaging does not provide insight to fine-grained neural representations within individual infants. To examine the finer-grain representation of category information within each infant's brain, we examined the distributed spatiotemporal responses to each category across the 23 electrodes spanning the left and right occipitotemporal cortex in each infant. We tested: (i) if categorical representations in an infant's brain are reliable across different images of a category, and (ii) if category representations become more distinct during the first year of life. We predicted that if representations become more similar across items of a category and more dissimilar between items of different categories then category distinctiveness (defined as the difference between mean within and between category similarity) would increase from 3 to 15 months of age.

To examine the representational structure, we calculated representation similarity matrices across odd/even split-halves of the data in each infant. Each cell in the RSM quantifies the similarity between two spatiotemporal patterns: On-diagonal cells of the RSM quantify the similarity of distributed spatiotemporal responses to different images from the same category and off-diagonal cells quantify the similarity of spatiotemporal responses to images from different categories. Categorical structure will manifest in RSMs as positive on diagonal values indicating reliable within-category spatiotemporal responses which are higher than off-diagonal between category similarity (**Fig 6** and Supplementary **Fig S3B**).

Examination of mean RSMs in each age group reveals no reliable category information in individuals at 3-4-month-olds or 4-6-month-olds, as within-category similarity is not significantly above zero (**Fig 6A**,

3-4-month-olds: on-diagonal, -0.03 $\pm$ 0.06, p = .96, one-tailed; 4-6-month-olds: on diagonal: 0.009 $\pm$ 0.11, p = .38). However, starting around 6 months some category structure emerges in the RSMs. In particular, distributed responses to faces become reliable as within category similarity for faces is significantly above zero in 6-8-month-olds (**Fig 6A,** 0.31 $\pm$ 0.24, $t_{(14)}$ = 5.1, p_{FDR} < .05, FDR corrected over 5 category levels), and stays reliable in 12-15-month-olds (**Fig 6A,** 0.26 $\pm$ 0.24, $t_{(14)}$ = 4.18, p_{FDR} < .05). Distributed responses to limbs become reliable later on as within category similarity for limbs is significantly above zero in 12-15-months-olds (**Fig 6A,** 0.11 $\pm$ 0.21, $t_{(14)}$ = 1.98, p = 0.03, but not surviving FDR correction at 5 levels).

<insert Fig 6 here>

Next, we evaluated the development of category distinctiveness, which was calculated for each category and infant. Individual infants' category distinctiveness is shown in **Fig 6B** (infants ordered by age) and in the scatterplots in **Fig 6C**. In infants younger than 4 months (120 days) category distinctiveness is largely close to zero or even negative, suggesting no differences between spatiotemporal responses to one category vs another. Category distinctiveness increases with age and becomes more positive from 84 to 445 days of age (**Fig 6B,C**). The biggest increase is for faces where after ~6 months of age (194 days) face distinctiveness is consistently positive in individual infants (13/15 infants aged 194-364 days and 12/15 infants aged 365-445 days). The increase in distinctiveness is more modest for other categories and appears later in development. For example, positive distinctiveness for limbs and cars in individual infants is consistently observed after 12 months of age (**Fig 6B,C**; limbs: 9/15 infants aged 365-445 days vs 5/15 infants aged 194-364 days**;** cars: 12/15 365-445 days vs 7/15 194-364 days).

Using LMMs we determined if distinctiveness significantly changed with age (log transformed) and category (participant, random factor). Results indicate that category distinctiveness significantly increased from 3 to 15 months of age (β_{age} = 0.77, 95% CI: 0.54 –1.00, $t_{(301)}$ = 6.62, $p = 1.67\times10^{-10}$), and further that development significantly varies across categories ($\beta_{age\ x\ category}$ = -0.13, 95% CI: -0.2 – -0.06, $t_{(301)}$ = -3.61, $p = 3.5\times10^{-4}$; main effect of category, $\beta_{category}$ = 0.27, 95% CI: 0.11 – 0.43, $t_{(301)}$ = 3.38, $p = 8.2\times10^{-4}$). Post-hoc analyses for each category (**Fig 6C**) reveal that distinctiveness significantly increased with age for faces (β_{age} = 0.9, 95% CI: 0.6 – 1.1, $t_{(59)}$ = 6.8, p_{FDR} < .001), limbs (β_{age} = 0.4, 95% CI: 0.2 – 0.6, $t_{(59)}$ = 5.0, p_{FDR} < .001), characters (β_{age} = 0.2, 95% CI: 0.02 – 0.3, $t_{(59)}$ = 2.2, p_{FDR} < .05), and cars (β_{age} = 0.4, 95% CI: 0.2 – 0.5, $t_{(59)}$ = 3.7, p_{FDR} < .001). Post-hoc t-tests show that for faces, the category distinctiveness is significantly above zero after 6 months of age (6-8-month-olds, $t_{(14)}$ = 6.73, p_{FDR} < .05; 12-15-month-olds, $t_{(14)}$ = 5.30, p_{FDR} < .05) and for limbs and cars at 12-15 months of age (limbs: $t_{(14)}$ = 2.19, p_{FDR} < .05; cars: $t_{(14)}$ = 4.53, p_{FDR} < .05).

This suggests that category distinctiveness slowly emerges in the visual cortex of infants from 3 to 15 months of age, with the largest and fastest development for faces.

Discussion

We find that both selective responses to items of a category over others across lateral occipital ROIs and the distinctiveness of distributed visual category representations progressively develop from 3 to 15 months of age. Notably, we find a differential development of category-selective responses (**Fig 7**), whereby responses to faces emerge the earliest, at 4-6 months of age and continue to develop through the first year of life. Category-selective responses to limbs, corridors, and characters follow, emerging at 6-8 months of age. Our analysis of the distinctiveness of the distributed spatiotemporal patterns to each category also find that distributed representations to faces become more robust in 6-8-month-olds and remain robust in 12-15-month-olds. While the distinctiveness of distributed patterns to limbs and cars only become reliable at 12-15 months of age. Together these data suggest a rethinking of the development of category representations during infancy as they not only suggest that category representations are learned, but also that representations of categories that may have innate substrates such as faces, bodies, and places emerge at different times during infancy.

<insert Fig 7 here>

Reliable category representations start to emerge at 4 months of age

While 3-4 months old infants have significant and reliable evoked visual responses over early visual cortex, we find no reliable category representations of faces, limbs, corridors, or characters in these young infants. Both analyses of average responses across lateral occipital ROIs and analyses of distributed spatiotemporal responses across visual cortex find no reliable category representations in 3-4-month-olds, either when examining mean response across an ROI or in distributed spatiotemporal patterns across visual cortex. The earliest categorical responses we find are for faces, and they emerge at 4-6 months of age.

Is it possible that there are some category representations in 3-4-month-olds, but we lack the sensitivity to measure them? We believe this is unlikely, because (i) we can measure significant visual responses from the same 3-4-month-olds, (ii) with the same amount of data, we can measure category selective responses and decode category information from distributed spatiotemporal responses in infants older than 4 months and in adults.

317 Ours findings together with a recent fMRI study in 2-10-month-olds[27] provide accumulating evidence
318 for multiple visual categories representations in infants' brains before the age of one. However, there are
319 also differences across studies. The earliest we could find reliable group-level category-selective
320 responses for faces was 4-6-month-olds and for limbs and corridors only after 6 months of age. In contrast,
321 Kosakowski and colleagues[27] report category-selective responses to faces, bodies, and scenes in example
322 4-5-months-olds. Group average data in their study found significant face- and place-selective responses
323 in infants' ventral temporal cortex (VTC) but not in lateral occipital cortex (LOTC), and significant body-
324 selective responses in LOTC, but not VTC. Because Kosakowski et al.[27] report group-averaged data across
325 infants spanning 8 months, their study does not provide insights to the time course of this development.
326 We note that, the studies differ in both measurement modalities (EEG/fMRI) and in the types of stimuli
327 infants viewed. In[27] infants viewed isolated, colored, and moving stimuli, but in our study, infants viewed
328 still, gray-level images on phase-scrambled backgrounds, which were controlled for several low level
329 properties. Thus, future research is necessary to determine whether differences between findings are due
330 to differences in measurement modalities or differences in stimulus format.
331
332 **Face representations emerge around 4-6 months of age**
333 Recognizing faces (e.g., a caregiver's face) is crucial for infant's daily lives. Converging evidence from
334 many studies suggest that infants have significant and reliable face-selective neural responses at 4-6
335 months of age[22,23,26,52,53]. While some studies report responses to face-like (high contrast paddle-like)
336 stimuli in newborns[34,35,54] and significant visual evoked responses to faces in 3-month-olds[55–58], these
337 studies have largely compared responses to an isolated faces vs. another isolated object. In contrast, we
338 do not find reliable face-selective responses (**Fig 3 & 4**) or reliable distributed representations (**Fig 5 & 6**)
339 to faces in 3-4-month-olds when responses to faces are contrasted to many other items and when stimuli
340 are shown on a background rather than in isolation. Our findings are consistent with longitudinal research
341 in macaques showing that robust cortical selectivity to faces takes several months to emerge[39] and
342 support the hypothesis that experience with faces is necessary for the development of cortical face
343 selectivity[37,39,40].
344 Our data also reveal that face-selective responses and distributed representations to faces become
345 more robust in 6-8-month-olds and remain robust in 12-15-month-olds. For example, successful decoding
346 of faces in the group level was observed in 80% of individual infants based on several minutes of EEG data.
347 Reliable distributed spatiotemporal responses to different images of faces become significantly different
348 from responses to images from different categories. This robust decoding has important clinical

ramifications as it may serve as an early biomarker for cortical face processing, which is important for early detection of social and cognitive developmental disorders such as Autism[59,60] and Williams Syndrome[61]. Future research is necessary for elucidating the relationship between the development of brain responses to faces to infant behavior. For example, it is interesting that at 6 months of age, when we find robust face representations, infants also start to exhibit recognition of familiar faces (like parents) and stranger anxiety[62].

One fascinating aspect of the development of cortical face selectivity is that among the categories we tested, selectivity to faces seems to emerge the earliest at around 4 months of age, yet the development of selectivity and distributed representations to faces is protracted compared to objects and places[14,63,15]. Indeed, in both our data and prior work, face-selective responses and distributed representations to faces in infants are immature compared to adults[21,25], and a large body of work has shown that face selectivity[5,14,15,50,63–65] and distributed representations to faces [10] continue to develop during childhood and adolescence. This suggests that not only experience during infancy but also life-long experience with faces, sculpts cortical face selectivity. We speculate that the extended cortical plasticity for faces may be due to both the expansion of social circles (family, friends and acquaintances) across the lifespans and also the changing statistics of faces we socialize with (e.g., child and adult faces have different appearance).

New insight about development: different category representations emerge at different times

To our knowledge, our study is the first to examine the development of both ROI level and spatiotemporal distributed responses in infants across the first year of life. We note that both analyses find that category information to faces develops before other categories. However, there are also some differences across analyses (**Fig 7**). For example, for limbs and corridors we find significant category-selective responses at the ROI level in lateral occipitotemporal ROIs starting at 6-8 months but no reliable distinct distributed responses across visual cortex at this age. In contrast, for cars, we find an opposite pattern where there is a distinct spatiotemporal pattern in 12-15 month-olds even as there is no significant car-selective response in the ROI level. As these approaches have different senstivities, they reveal insights to the nature of the underlying representations. For example, as visible in **Fig 4**, limbs and corridor have a clear category selective waveform in both in 6-8 and 12-15-months-olds, but the waveform of limbs and its spatial distribution is not that different from that to corridors, which may explain why distinctiveness of spatitemporal patterns for limbs is low in 6-8-moths old (**Fig 6**). Likewise, even as there is no signifincat response for cars (**Fig 4e**), its spatiotemporal pattern is consistently different than for other categories giving rise to a distinctive spatiotemporal response by 12 months (**Fig 6**).

In sum, the key finding from our study is that the development of category selectivity during infancy is non-uniform: face-selective responses and representations of distributed patterns develop before representations to limbs and other categories. We hypothesize that this differential development of visual category representations may be due to differential visual experience with these categories during infancy. This hypothesis is consistent with behavioral research using head-mounted cameras that revealed that the visual input during early infancy is dense with faces, while hands become more prevalent in the visual input later in development and especially when in contact with objects[41,42]. Additionally, a large body of research has suggested that young infants preferentially look at faces and face-like stimuli[17,18,33,34], as well as look longer at faces than other objects[41], indicating that not only the prevalence of faces in babies' environments but also longer looking times may drive the early development of face representations. Future studies that examine the visual diet, looking behavior, and brain development using additional behaviorally relevant categories such as food[66–68] can test how environmental and experiential differences may influence infants' developmental trajectories

Together our findings not only suggest that visual experience is necessary for the development of visual category representations, including faces, but also necessitate a rethinking of how visual category representations develop in infancy. Moreover, this differential development during infancy is evident even for categories that have evolutionary importance and may have innate substrates such as faces, bodies, and places[28–31]. Finally, our findings have important ramifications for theoretical and computational models of visual development as well as for the assessment of atypical infant development.

Methods

Participants

Ethical permission for the study was obtained from the Institutional Review Board of Stanford University. Parents of the infant participants provided written informed consent prior to their first visit and also prior to each session if they came for multiple sessions. Participants were paid 20$/hour for participation. Participants were recruited via ads on social media (Facebook and Instagram).

Sixty-two full-term, typically developing infants were recruited. Twelve participants were part of an ongoing longitudinal study and came for several sessions spanning ~3 months apart (seven 3-4-month-olds, three 4-6-month-olds, eight 6-8-month-olds, and twelve 12-15-month-olds). Data from nineteen infants (nine 3-4-month-olds, six 4-6-month-olds, and eight 6-8-month-olds; among whom seven were longitudinal) were acquired in two visits within a two-week span to obtain a sufficient number of valid data epochs. Supplementary **Table S1** contains participants' demographic information (sex and race). The

youngest infants were 3 months old, as the EEG setup requires the infants to be able to hold their head and look at the screen in front of them. 23 adults (14 females) also participated in the study. All participants had normal/corrected-to-normal vision and provided written informed consent.

Data e*xclusion criteria:* We excluded participants who had less than 20 valid epochs (1.1667-s/epoch) per category, had noise/muscular artifacts during the EEG recordings, couldn't record data, or had no visual responses over the occipital electrodes. As such, we excluded (1) five infants due to an insufficient number of epochs, (2) two infants who had no visual responses, (3) ten infants due to technical issues during data collection, and (4) three adults due to excessive noise/muscular artifacts during EEG. In total, we report data from 45 infants (Supplementary **Table S1**) and 20 adults (13 females, 19-38 years) that met inclusion criteria.

Visual stimuli

Natural grayscale images of adult faces, limbs, corridors, characters, and cars are used as stimuli, with 144 images per category from the fLOC image database (https://github.com/VPNL/fLoc) [8]. The size, view, and retinal position of the items varied, and the items were overlaid on phase-scrambled backgrounds that were generated from a randomly drawn image in the stimulus set. The images were also controlled for multiple low-level differences between stimuli of different categories including their luminance, contrast, similarity and spatial frequency power distributions using the SHINE toolbox [69]. As only five of ten categories from [8] were used, we evaluated the stimuli used in our experiments to test if they differed in (i) contrast, (ii) luminance, (iii) similarity, and (iv) spatial frequency. Results show that categories were largely matched on most metrics (**Fig 1B** and Supplementary Materials). The stimuli were presented on a gamma-corrected OLED monitor screen (SONY PVM-2451; SONY Corporation, Tokyo Japan) at a screen resolution of 1920 × 1080 pixels and a monitor refresh rate of 60 Hz. When viewed from 70 cm away, each image extended a visual angle of approximately 12°.

EEG Protocol

The experiments were conducted in a calm, dimly illuminated lab room. Stimuli were presented using custom stimulus presentation software with millisecond timing precision. During testing, infant participants were seated on their parent's laps in front of the screen at a distance of 70 cm. One experimenter stood behind the presentation screen to monitor where the infant was looking. The visual presentation was paused if the infant looked away from the screen and was continued when the infant looked again at the center of the screen. To motivate infants to fixate and look at the screen, we presented at the center of the screen small (~1°) colored cartoon images such as butterflies, flowers, ladybugs. For adults, we used a fixation cross of the same size instead of the cartoons and asked the participants to

445 fixate and indicate when the fixation's color changed by pressing a space bar key on a keyboard. EEG
446 measurements for infant participants continued until the infant no longer attended the screen and we
447 obtained between 2-12 different 70-s sequences per individual. For adult participants, we acquired 12
448 sequences per individual.

449 A frequency-tagging paradigm [23,46] was used to measure brain responses. In the experiment,
450 randomly selected images from 5 categories were presented sequentially at a rate of 4.286 Hz (~233 ms
451 per image) with no inter stimulus interval during each 70-s sequence. For each condition, one category
452 was determined as the target category; for this category random selected images from that category were
453 presented first and followed by four images randomly drawn from the other four categories with no
454 regular order (**Fig 1A)**. The target images are therefore presented periodically at 0.857 Hz (i.e. 4.286 Hz/5),
455 but the intervals between sequential presentations of images from the other 4 categories was not periodic.
456 The probability of image occurrences across categories was equal at 20%. The experiment had five
457 conditions one for each of the following target categories: faces, limbs, corridors, characters, and cars.
458 Each 70-s experimental sequence was composed of five 14-s long conditions which included a 1.1667-s
459 stimulus fade-in and a 1.1667-s stimulus fade-out. The order of the category conditions was randomized
460 within each 70-s sequence. No image was repeated within a sequence. Two presentation frequencies
461 were embedded in the experiment: (i) the image frequency (4.286 Hz), which is predicted to elicit visual
462 responses to all stimuli over occipital visual cortex, and, (ii) the category frequency (0.857 Hz), which is
463 predicted to elicit a category-selective response over lateral occipital-temporal electrodes.

464 **EEG Acquisition**

465 EEG data were recorded at 500 Hz from a 128-channel EGI High-Density Geodesic Sensor Net. For
466 infants, the net was connected to a NetAmps 300 amplifier (Electrical Geodesics, Inc., Eugene, OR, USA).
467 For the adults, the net was connected to a NetAmps400 amplifier. The EEG recording was referenced
468 online to a single vertex (electrode Cz) and the channel impedance was kept below 50 KΩ.

469 **Pre-processing**

470 EEG recordings were down-sampled to 420 Hz and were filtered using a 0.03-50 Hz bandpass filter
471 with custom signal processing software. For infants, channels with more than 20% of samples exceeding
472 a 100-150 µV amplitude threshold were replaced with the average amplitude of its six nearest-neighbor
473 channels. The continuous EEG signals were then re-referenced to the common average of all channels and
474 segmented into 1166.7-ms epochs (i.e., duration of five stimuli starting with a target category image
475 followed with four images drawn from the rest four categories). Epochs with more than 15% of time
476 samples exceeding threshold (150 - 200 µV) were excluded further on a channel-by-channel basis [70]. For

477 adults, the two-step artifact rejection was performed with different criteria as EEG response amplitudes
478 are lower in adults than infants [70]. EEG channels with more than 15% of samples exceeding a 30 µV
479 amplitude threshold were replaced by the average value of their neighboring channels. Then the EEG
480 signals were re-referenced to the common average of all channels and segmented into 1.1667-s epochs.
481 Epochs with more than 10% of time samples exceeding threshold (30 - 80 µV) were excluded on a channel-
482 by-channel basis [71].

483 **Supplementary Table S2** shows the number of epochs (1.1667-s each) we acquired before and after
484 data pre-processing summing across all five categories. We used data after pre-processing for further
485 analyses. There was no significant difference in the number of pre-processed epochs across infant age-
486 groups ($F_{(3,57)} = 1.5$, $p = .2$).

487 **Univariate EEG analyses**

488 Both image update and categorical EEG visual responses are reported in the frequency and time
489 domain over three regions-of-interest (ROIs): two occipito-temporal ROIs (left occipitotemporal (LOT):
490 channels 57, 58, 59, 63, 64, 65 and 68; right occipitotemporal (ROT) channels: 90, 91, 94, 95, 96, 99, and
491 100) and one occipital ROI (channels 69, 70, 71, 74, 75, 76, 82, 83 and 89). These ROIs were selected a
492 priori based on a previously published study [51]. We further removed several channels in these ROIs for
493 two reasons: (1) Three outer rim channels (i.e., 73, 81, and 88) were not included in the occipital ROI for
494 further data analysis for both infant and adult participants because they were consistently noisy. (2) Three
495 channels (66, 72, and 84) in the occipital ROI, one channel (50) in the LOT ROI, and one channel (101) in
496 the ROT ROI were removed because they did do not show substantial responses in the group-level
497 analyses.

498 ***Frequency domain analysis.*** Individual participant's preprocessed EEG signals for each stimulus
499 condition were averaged over two consecutive epochs (2.3334-s). The averaged time courses for each
500 participant were then converted to the frequency domain at a frequency resolution of 0.4286 Hz via a
501 Discrete Fourier Transform (DFT). The frequency bins of interest are at exactly every other bin in the
502 frequency spectrum. The real and imaginary Fourier coefficients for each of the categorical and image
503 update responses for each condition were averaged across participants (vector averaging) to obtain a
504 group-level estimates. The amplitudes of response were then computed from the coherently averaged
505 vector. Hotelling's T2 statistic [72] was used to test whether response amplitudes were significantly different
506 from zero. We used Benjamini's & Hochberg's false discovery rate (FDR) procedure to correct for multiple
507 comparisons.

Image-update visual responses (image frequency): The amplitude and phase of the evoked response at the image presentation rate and its first 3 harmonics (4.286 Hz, 8.571 Hz, 12.857 Hz, and 17.143 Hz).

Categorical responses (category frequency): The amplitude and phase of the response at the category repetition frequency and its second harmonic (0.857 Hz, 1.714 Hz) for each category condition.

Time domain analyses. Preprocessed time domain EEG signals of each participant were low-passed filtered with a 30Hz cut-off. The raw EEG signals contain many components including categorical responses 0.857 Hz and harmonics), general visual responses (4.286 Hz and harmonics) and noise. To separate out the temporal waveforms of these two responses, we first transformed the epoch-averaged (2.3334-s) time series of each condition into frequency domain using a DFT. Then, we used an inverse DFT to transform back to the time domain keeping only the responses at the category frequency and its harmonics, zeroing the other frequencies. The same approach was used to separate the visual responses by using an inverse DFT transform of the responses at 4.286 Hz and harmonics.

Categorical responses. We kept responses at frequencies of interest (0.857 Hz and its harmonics up to 30 Hz, excluding the harmonic frequencies that overlapped with the image frequency and its harmonics) and zeroed responses in other frequencies. Then we applied an inverse Fourier transform to transform the data to the time domain. We further segmented the time series into 1.1667-s epochs and averaged across these epochs for each condition and individual. The mean and standard error across participants were computed for each condition at each time-point.

Image update visual responses. A similar procedure was performed except that frequencies of interest are 4.286 Hz and its harmonics, and the rest were zeroed. As temporal waveforms for image-update responses were similar across different category conditions, we averaged waveforms across all five conditions and report the mean response (**Fig 2**).

Statistical analyses. To determine time windows in which amplitudes were significantly different from zero for each condition, we used a cluster-based nonparametric permutation t-test, 10000 permutations, with a threshold of $p < 0.05$, two-tailed) on the post-stimulus onset time points (0-1167 ms) [73,74]. The null hypothesis is that the evoked waveforms are not different from zero at any time point. For each permutation, we assigned random signs to the data of individual participants and computed the group-level difference against zero using a t-test. We then calculated the cluster-level statistic as the sum of t-values in the consecutive time points with p-values less than 0.05 [75]. We calculated the maximum cluster-level statistic for each permutation to generate a non-parametric reference distribution of cluster-level statistics. We rejected the null hypothesis if the cluster-level statistic for any consecutive time points in the original data was larger than 97.5% or smaller than 2.5% of the values in the null distribution.

540 **Decoding analyses**

541 ***Group-level.*** We used a leave-one-out-cross-validation (LOOCV) classifier to test if spatiotemporal
542 responses patterns to each of the five categories were reliable across participants. The classifier was
543 trained on averaged data from N-1 participants and tested on how well it predicted the category the left-
544 out participant was viewing from their brain activations. This procedure was repeated for each left-out
545 participant. We calculated the averaged category temporal waveform for each category across channels
546 of our three ROIs: 7 LOT, 9 occipital, and 7 ROT, as the exact location of the channels varies across
547 individuals. Then we concatenated the waveform from the three ROIs to form a spatiotemporal response
548 vector. At each iteration, the LOOCV classifier computed the correlation between each of the five category
549 vectors from the left-out participant (test data, for an unknown stimulus) and each of the mean
550 spatiotemporal vectors across the N-1 participants (training data, labeled data). The winner-take-all (WTA)
551 classifier classifies the test vector to the category of the training vector that yields the highest correlation
552 with the test vector. We computed group mean decoding performance across all N iterations for each
553 category, and the group mean decoding accuracies across five categories.

554 ***Individual level.*** Similar to group level with 2 differences: (i) All analyses were done within an
555 individual using a split-half approach. That is, the classifier was trained on one half of the data (i.e., odd
556 or even trials) and tested on the other half of the data. (ii) Spatiotemporal patterns for each category used
557 the concatenated waveforms across 23 channels spanning the occipital and bilateral occipitotemporal
558 ROIs.

559 ***Category distinctiveness*** [10]***.*** Category distinctiveness is defined as the difference between the
560 similarity (correlation coefficient) of spatiotemporal responses within a category across odd and even
561 splits using different image and the average between-category similarity of spatiotemporal responses
562 across odd and even splits [10]. Distinctiveness is higher when the within category similarity is positive and
563 the between category similarity is negative and varies from -2 (no category information) to 2 (maximal
564 category information). We computed category distinctiveness for each of the 5 categories as in each infant
565 and determined if it varied from 3 to 15 months of age.

566 **Statistical Analyses of Developmental Effects**

567 To quantify developmental effects, we used linear mixed models (LMMs, [76] with the 'fitlme' function
568 in MATLAB version 2021b (MathWorks, Inc.). LMMs allow explicit modeling of both within-subject effects
569 (e.g., longitudinal measurements) and between-subject effects (e.g., cross-sectional data) with unequal
570 number of points per participants, as well as examine main and interactive effects of both continuous
571 (age) and categorical (e.g., stimulus category) variables. We used random-intercept models that allow the

intercept to vary across participants (term: 1|participant). In all LMMs, we measured development as a function of log 10(age in days) as development is typically faster earlier on. Indirect evidence comes from neuroimaging and post-mortem studies showing that the structural development of infants' brain is nonlinear, with development in the first two years being rapid, especially the first year [77].

We report slope (rate of development), interaction effects, and their significance. Table 1 summarizes LMMs used in this study.

Table 1. Linear mixed models (LMMs)

Variable	LMM formula	Results
Peak latency: latency of the peak waveform in each time window; window 1: 60-90 ms or window 2: 90-160 ms for 3-4-month-olds, and 90-110 ms for other age groups	Peak latency ~ 1 + log10(age in days) × time window + (1|participant)	Fig 2G. Supplementary Table S4.
	Peak latency ~ 1+ log10(age in days) + (1|participant)	Fig 2H. Supplementary Table S5.
Peak amplitude: peak response in a 400-700ms time window	Peak amplitude ~ 1+ log10(age in days) × category + (1|participant)	Supplementary Table S7.
	Peak amplitude ~ 1+ log10(age in days) + (1|participant)	Fig 4. Supplementary Table S8.
Category distinctiveness of spatiotemporal responses for each of the five categories	Category distinctiveness ~ log10(age in days) × category + (1|participant)	
	Category distinctiveness ~ log10(age in days) + (1|participant)	Fig 6B.

Analysis of Noise

To test whether EEG noise levels vary with age, for example whether noise in the EEG data is larger in the younger infants than older ones, we quantified the response amplitudes in the occipital ROI in the frequency domain, at frequency bins next to the category and image frequency bins (white bars in **Fig 2A-D**, right panel). The noise level was quantified as the amplitude of response up to 8.571 Hz excluding image presentation frequencies (4.286 Hz and harmonics) and category frequencies (0.857 Hz and harmonics) as this frequency range includes the relevant main harmonics (**Fig 2E**). We used a LMM to test if noise varies with age, with participant as a random effect:

Noise amplitude ~ 1 + log10(age in days) + (1|participant)

We found no significant differences in noise amplitude across infant age groups (**Fig 2E**, mean amplitude across the first five noise bins: β_{age} = -.005, 95% CI: -.12 – .11, $t_{(59)}$ = -.09, p = .93; mean noise across the first 10 bins: β_{age} = .04, 95% CI: -.03 – .12, $t_{(59)}$ = 1.16, p = .25, linear mixed models, LMMs, with age (log transformed) as fixed effects and participant as a random effect).

Data availability

Individual EEG data, and code for all analyses are available on Github: https://github.com/VPNL/InfantObjectCategorization.

Acknowledgements

This work was supported by grants from the Wu Tsai Neurosciences Institute of Stanford University, the Human Centered Artificial Intelligence Institute of Stanford University to KGS and AMN.

Author contributions

XY designed the experiment, coded the analysis pipeline, collected the data, analyzed data, and wrote the paper. AMN designed the experimental protocol and provided funding and instrumentation. SF and BF recruited the participants. ST, BF, and YC collected the data. ST and BF equally contributed to the study. KGS designed the experiment, oversaw the data collection and analyses, provided funding, and wrote the paper. All authors read the manuscript, provided critical revisions and approved the final paper for submission.

Competing interests

The authors declare no competing interests.

613 **References**

614 1. Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381**, 520–
615 522 (1996).
616 2. Grill-Spector, K. & Kanwisher, N. Visual Recognition: As Soon as You Know It Is there, You Know
617 What It Is. *Psychological Science* **16**, (2005).
618 3. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its
619 role in categorization. *Nat Rev Neurosci* **15**, 536–548 (2014).
620 4. Bugatus, L., Weiner, K. S. & Grill-Spector, K. Task alters category representations in prefrontal but
621 not high-level visual cortex. *Neuroimage* **155**, 437–449 (2017).
622 5. Nordt, M. *et al.* Cortical recycling in high-level visual cortex during childhood development. *Nat Hum*
623 *Behav* **5**, 1686–1697 (2021).
624 6. Margalit, E. *et al.* Ultra-high-resolution fMRI of Human Ventral Temporal Cortex Reveals Differential
625 Representation of Categories and Domains. *J. Neurosci.* **40**, 3008–3024 (2020).
626 7. Gomez, J. *et al.* Microstructural proliferation in human cortex is coupled with the development of
627 face processing. **355**, (2017).
628 8. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Temporal Processing Capacity in High-Level Visual
629 Cortex Is Domain Specific. *Journal of Neuroscience* **35**, 12412–12424 (2015).
630 9. Haxby, J. V. *et al.* Distributed and Overlapping Representations of Faces and Objects in Ventral
631 Temporal Cortex. *Science* **293**, 2425–2430 (2001).
632 10. Nordt, M. *et al.* Longitudinal development of category representations in ventral temporal cortex
633 predicts word and face recognition. *Nat Commun* **14**, 8010 (2023).
634 11. Kanwisher, N., McDermott, J. & Chun, M. M. The Fusiform Face Area: A Module in Human
635 Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience* **17**, 4302–4311
636 (1997).

637 12. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**,
638 598–601 (1998).
639 13. Downing, P. E., Jiang, Y., Shuman, M. & Kanwisher, N. A Cortical Area Selective for Visual Processing
640 of the Human Body. *Science* **293**, 2470–2473 (2001).
641 14. Golarai, G. *et al.* Differential development of high-level visual cortex correlates with category-
642 specific recognition memory. *Nature Neuroscience* **10**, 512–522 (2007).
643 15. Scherf, K. S., Behrmann, M., Humphreys, K. & Luna, B. Visual category-selectivity for faces, places
644 and objects emerges along different developmental trajectories. *Developmental Science* **10**, F15–
645 F30 (2007).
646 16. Dehaene-Lambertz, G., Monzalvo, K. & Dehaene, S. The emergence of the visual word form:
647 Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS*
648 *Biology* **16**, e2004103 (2018).
649 17. Spriet, C., Abassi, E., Hochmann, J.-R. & Papeo, L. Visual object categorization in infancy. *Proc. Natl.*
650 *Acad. Sci. U.S.A.* **119**, e2105866119 (2022).
651 18. Mondloch, C. J. *et al.* Face Perception During Early Infancy. *Psychol Sci* **10**, 419–422 (1999).
652 19. Quinn, P. C., Eimas, P. D. & Rosenkrantz, S. L. Evidence for Representations of Perceptually Similar
653 Natural Categories by 3-Month-Old and 4-Month-Old Infants. *Perception* **22**, 463–475 (1993).
654 20. Younger, B. A. & Fearing, D. D. Parsing Items into Separate Categories: Developmental Change in
655 Infant Categorization. *Child Development* **70**, 291–303 (1999).
656 21. Conte, S., Richards, J. E., Guy, M. W., Xie, W. & Roberts, J. E. Face-sensitive brain responses in the
657 first year of life. *NeuroImage* **211**, 116602 (2020).
658 22. Farzin, F., Hou, C. & Norcia, A. M. Piecing it together: Infants' neural responses to face and object
659 structure. *Journal of Vision* **12**, 6 (2012).

660 23. de Heering, A. & Rossion, B. Rapid categorization of natural face images in the infant right
661 hemisphere. *eLife* **4**, e06564 (2015).
662 24. Bayet, L. *et al.* Temporal dynamics of visual representations in the infant brain. *Developmental*
663 *Cognitive Neuroscience* **45**, 100860 (2020).
664 25. Xie, S. *et al.* Visual category representations in the infant brain. *Current Biology* **32**, 5422-5432.e6
665 (2022).
666 26. Deen, B. *et al.* Organization of high-level visual cortex in human infants. *Nat Commun* **8**, 13995
667 (2017).
668 27. Kosakowski, H. L. *et al.* Selective responses to faces, scenes, and bodies in the ventral visual
669 pathway of infants. *Current Biology* **32**, 265-274.e5 (2022).
670 28. Kanwisher, N. Functional specificity in the human brain: A window into the functional architecture
671 of the mind. *Proceedings of the National Academy of Sciences* **107**, 11163–11170 (2010).
672 29. Sugita, Y. Innate face processing. *Current Opinion in Neurobiology* **19**, 39–44 (2009).
673 30. Mahon, B. Z. & Caramazza, A. What drives the organization of object knowledge in the brain? *Trends*
674 *in Cognitive Sciences* **15**, 97–103 (2011).
675 31. Bi, Y., Wang, X. & Caramazza, A. Object Domain and Modality in the Ventral Visual Pathway. *Trends*
676 *in Cognitive Sciences* **20**, 282–290 (2016).
677 32. Behrmann, M. & Plaut, D. C. A vision of graded hemispheric specialization. *Ann N Y Acad Sci* **1359**,
678 30–46 (2015).
679 33. Pascalis, O., de Schonen, S., Morton, J., Deruelle, C. & Fabre-Grenet, M. Mother's face recognition
680 by neonates: A replication and an extension. *Infant Behavior & Development* **18**, 79–85 (1995).
681 34. Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. Newborns' preferential tracking of face-like
682 stimuli and its subsequent decline. *Cognition* **40**, 1–19 (1991).

683 35. Buiatti, M. *et al.* Cortical route for facelike pattern processing in human newborns. *Proceedings of*
684 *the National Academy of Sciences* **116**, 4625–4630 (2019).
685 36. Dehaene, S. *et al.* How Learning to Read Changes the Cortical Networks for Vision and Language.
686 *Science* **330**, 1359–1364 (2010).
687 37. Arcaro, M. J., Schade, P. F. & Livingstone, M. S. Universal Mechanisms and the Development of the
688 Face Network: What You See Is What You Get. *Annu Rev Vis Sci* **5**, 341–372 (2019).
689 38. Scott, L. S. & Arcaro, M. J. A domain-relevant framework for the development of face processing.
690 *Nat Rev Psychol* **2**, 183–195 (2023).
691 39. Livingstone, M. S. *et al.* Development of the macaque face-patch system. *Nat Commun* **8**, 14897
692 (2017).
693 40. Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R. & Livingstone, M. S. Seeing faces is necessary
694 for face-domain formation. *Nat Neurosci* **20**, 1404–1412 (2017).
695 41. Fausey, C. M., Jayaraman, S. & Smith, L. B. From faces to hands: Changing visual input in the first
696 two years. *Cognition* **152**, 101–107 (2016).
697 42. Jayaraman, S., Fausey, C. M. & Smith, L. B. Why are faces denser in the visual experiences of
698 younger than older infants? *Developmental Psychology* **53**, 38–49 (2017).
699 43. Allen, E. J. *et al.* A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial
700 intelligence. *Nat Neurosci* **25**, 116–126 (2022).
701 44. Lerma-Usabiaga, G., Carreiras, M. & Paz-Alonso, P. M. Converging evidence for functional and
702 structural segregation within the left ventral occipitotemporal cortex in reading. *Proceedings of the*
703 *National Academy of Sciences* **115**, E9981–E9990 (2018).
704 45. Jagadeesh, A. V. & Gardner, J. L. Texture-like representation of objects in human visual cortex.
705 *Proceedings of the National Academy of Sciences* **119**, e2115302119 (2022).

46. Farzin, F., Hou, C. & Norcia, A. M. Piecing it together: Infants' neural responses to face and object structure. *Journal of Vision* **12**, 6 (2012).

47. Heinrich, S. P., Mell, D. & Bach, M. Frequency-domain analysis of fast oddball responses to visual stimuli: A feasibility study. *International Journal of Psychophysiology* **73**, 287–293 (2009).

48. Liu-Shuang, J., Norcia, A. M. & Rossion, B. An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia* **52**, 57–72 (2014).

49. Rekow, D. *et al.* Odor-driven face-like categorization in the human infant brain. *Proceedings of the National Academy of Sciences* **118**, e2014979118 (2021).

50. Taylor, M. J., McCarthy, G., Saliba, E. & Degiovanni, E. ERP evidence of developmental changes in processing of faces. *Clinical Neurophysiology* **110**, 910–915 (1999).

51. Xie, W., McCormick, S. A., Westerlund, A., Bowman, L. C. & Nelson, C. A. Neural Correlates of Facial Emotion Processing in Infancy. *Dev Sci* **22**, e12758 (2019).

52. Guy, M. W., Zieber, N. & Richards, J. E. The Cortical Development of Specialized Face Processing in Infancy. *Child Dev* **87**, 1581–1600 (2016).

53. Halit, H., Csibra, G., Volein, Á. & Johnson, M. H. Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry* **45**, 1228–1234 (2004).

54. Goren, C. C., Sarty, M. & Wu, P. Y. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* **56**, 544–549 (1975).

55. Halit, H., de Haan, M. & Johnson, M. H. Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage* **19**, 1180–1193 (2003).

56. Macchi Cassia, V., Kuefner, D., Westerlund, A. & Nelson, C. A. A behavioural and ERP investigation of 3-month-olds' face preferences. *Neuropsychologia* **44**, 2113–2125 (2006).

730 57. Peykarjou, S. Frequency tagging with infants: The visual oddball paradigm. *Front. Psychol.* **13**,
731 1015611 (2022).
732 58. Tzourio-Mazoyer, N. *et al.* Neural Correlates of Woman Face Processing by 2-Month-Old Infants.
733 *NeuroImage* **15**, 454–461 (2002).
734 59. Rossion, B. Biomarkers of Face Perception in Autism Spectrum Disorder: Time to Shift to Fast
735 Periodic Visual Stimulation With Electroencephalography? *Biological Psychiatry: Cognitive*
736 *Neuroscience and Neuroimaging* **5**, 258–260 (2020).
737 60. Vettori, S. *et al.* Reduced neural sensitivity to rapid individual face discrimination in autism spectrum
738 disorder. *NeuroImage: Clinical* **21**, 101613 (2019).
739 61. Farran, E. K. *et al.* Characterizing the neural signature of face processing in Williams syndrome via
740 multivariate pattern analysis and event related potentials. *Neuropsychologia* **142**, 107440 (2020).
741 62. Kobayashi, M., Kakigi, R., Kanazawa, S. & Yamaguchi, M. K. Infants' recognition of their mothers'
742 faces in facial drawings. *Dev Psychobiol* **62**, 1011–1020 (2020).
743 63. Peelen, M. V., Fei-Fei, L. & Kastner, S. Neural mechanisms of rapid natural scene categorization in
744 human visual cortex. *Nature* **460**, 94–97 (2009).
745 64. Cantlon, J. F., Pinel, P., Dehaene, S. & Pelphrey, K. A. Cortical Representations of Symbols, Objects,
746 and Faces Are Pruned Back during Early Childhood. *Cerebral Cortex* **21**, 191–199 (2011).
747 65. Cohen, M. A. *et al.* Representational similarity precedes category selectivity in the developing
748 ventral visual pathway. *NeuroImage* **197**, 565–574 (2019).
749 66. Bannert, M. M. & Bartels, A. Visual cortex: Big data analysis uncovers food specificity. *Current*
750 *Biology* **32**, R1012–R1015 (2022).
751 67. Jain, N. Selectivity for food in human ventral visual cortex. *Communications Biology* **6**, 175 (2023).
752 68. Pennock, I. M. L. *et al.* Color-biased regions in the ventral visual pathway are food selective. *Current*
753 *Biology* **33**, 134-146.e4 (2023).

754 69. Willenbockel, V. *et al.* Controlling low-level image properties: The SHINE toolbox. *Behavior Research*
755 *Methods* **42**, 671–684 (2010).
756 70. Norcia, A. M., Pei, F. & Kohler, P. J. Evidence for long-range spatiotemporal interactions in infant and
757 adult visual cortex. *Journal of Vision* **17**, 12 (2017).
758 71. Kohler, P. J., Barzegaran, E., Norcia, A. M. & McCandliss, B. D. Parietal Contributions to Abstract
759 Numerosity Measured with Steady State Visual Evoked Potentials. 2020.08.06.239889 Preprint at
760 https://doi.org/10.1101/2020.08.06.239889 (2020).
761 72. Victor, J. D. & Mast, J. A new statistic for steady-state evoked potentials. *Electroencephalography*
762 *and Clinical Neurophysiology* **78**, 378–388 (1991).
763 73. Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W. & Norcia, A. M. Cue-Invariant Networks
764 for Figure and Background Processing in Human Visual Cortex. *J. Neurosci.* **26**, 11695–11708 (2006).
765 74. Blair, R. C. & Karniski, W. An alternative method for significance testing of waveform difference
766 potentials. *Psychophysiology* **30**, 518–524 (1993).
767 75. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of*
768 *Neuroscience Methods* **164**, 177–190 (2007).
769 76. Bosker, R. & Snijders, T. Multilevel Analysis : An Introduction to Basic and Advanced Multilevel
770 Modeling. 1–368 (2011).
771 77. Gilmore, J. H., Knickmeyer, R. C. & Gao, W. Imaging structural and functional brain development in
772 early childhood. *Nat Rev Neurosci* **19**, 123–137 (2018).
773

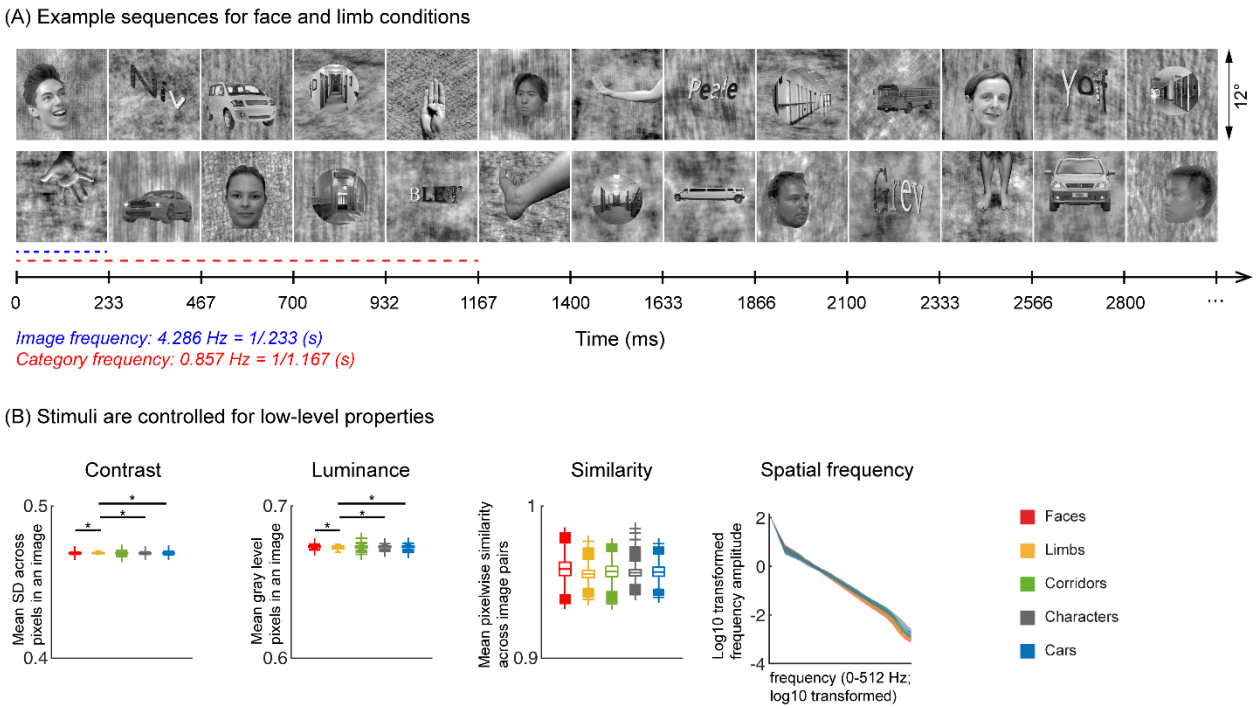


Figure 1. Experimental design and stimuli analysis. (A) Example segments of presentation sequences in which faces (top panel) and limbs (bottom panel) were the target category. Images spanning 12° containing gray-level images of items from different categories on a phase scrambled background appeared for 233 ms (frequency: 4.286 Hz). A different exemplar from a single category appeared every 5th image (frequency: 0.857 Hz). Between the target category images, randomly drawn images from the other four categories were presented. Sequences consisted of 20% images from each of the five categories and no images were repeated. Each category condition lasted for 14-s and contained 12 such cycles. Participants viewed in random order 5 category conditions: faces, limbs, corridors, characters, and cars forming a 70-s presentation sequence. (B) Images were controlled for several low-level properties using the SHINE toolbox as explained in [8]. Metrics are colored by category (see legend). *Contrast*: mean standard deviation of gray-level values in each image, averaged across 144 images of a category. *Luminance*: Mean gray-level of each image, averaged across 144 images of a category. *Similarity:* Mean pixel wise similarity between all pairs of images in a category. For all 3 metrics, boxplots indicate median, 25%, 75% percentiles, range, and outliers. Significant differences between categories are indicated by asterisks, for contrast and luminance (non-parametric permutation t-test $p < .05$, Bonferroni corrected); for image similarity, all categories are significantly different than others (non-parametric permutation testing, $p < 0.05$, Bonferroni corrected, except for corridors vs. cars, $p = .24$). *Spatial Frequency*: *Solid lines:* Distribution of spectral amplitude in each frequency averaged across 144 images in each category. *Shaded area*: standard deviation. Spatial frequency distributions are similar across categories.

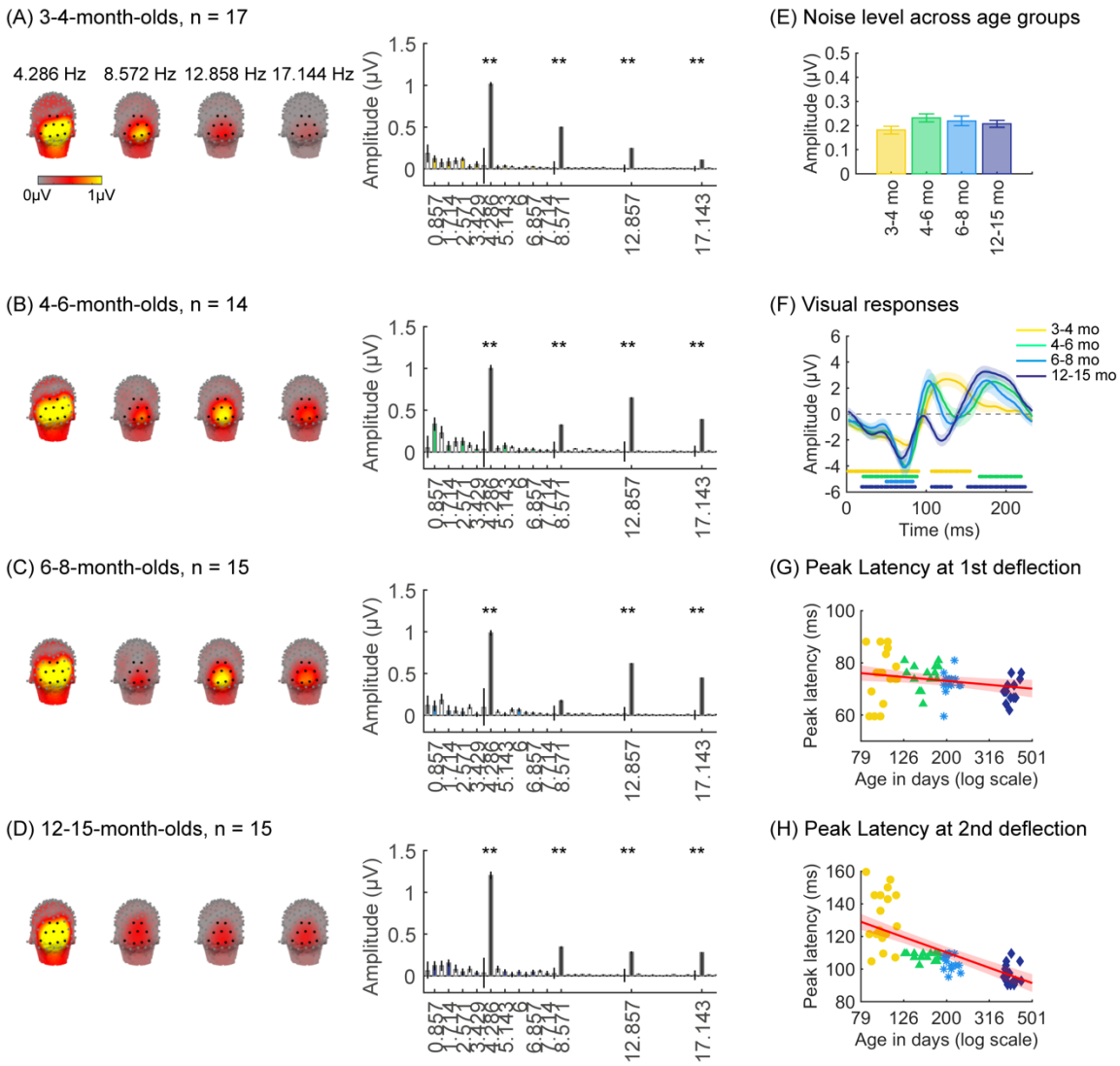


Figure 2. Strong visual responses over occipital cortex at the image-update frequency and its harmonics in all age groups. Each panel shows mean responses across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds, n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. *Left panels in each row:* spatial distribution of the visual response at the image update frequency and its first three harmonics. *Middle panels in each row:* mean Fourier Amplitude Spectrum across 9 occipital electrodes of the occipital ROI showing high activity at harmonics of the image update frequency marked out by thicker lines. Data are first averaged in each participant and each condition and then across participants. *Error bars:* standard error of the mean across participants. *Asterisks:* Response amplitudes significantly larger than zero, $p < .01$, FDR corrected. *Colored bars:* amplitude of response at category frequency and its harmonics. *White bars:* amplitude of response at noise frequencies. (E) Noise amplitudes in the frequency range up to 8.571 Hz (except for the visual response frequencies and visual category response frequencies) from the amplitude spectra in (A) for each age group (white bars on the spectra). *Error bars:* standard error of the mean across participants. (F) Mean image-update response over occipital electrodes for each age group. Waveforms are cycle averages over the period of the individual image presentation time (233-ms). *Lines:* mean response. *Shaded areas:* standard error of the mean across participants of each group. *Horizontal lines colored by age group:* significant responses vs. zero ($p < .05$ with a cluster-based analysis, see Methods). (G) Peak latency for the first peak in the 60-90 ms interval after stimulus onset. Each dot is a participant; Dots are colored by age group. *Line*: linear mixed model (LMM) estimate of peak latency as a function of log10(age). *Shaded area:* 95% confidence interval (CI). (H) Same as (G) but for the second peak in the 90-160 ms interval for 3-4-month-olds, and 90-110 ms for older infants.

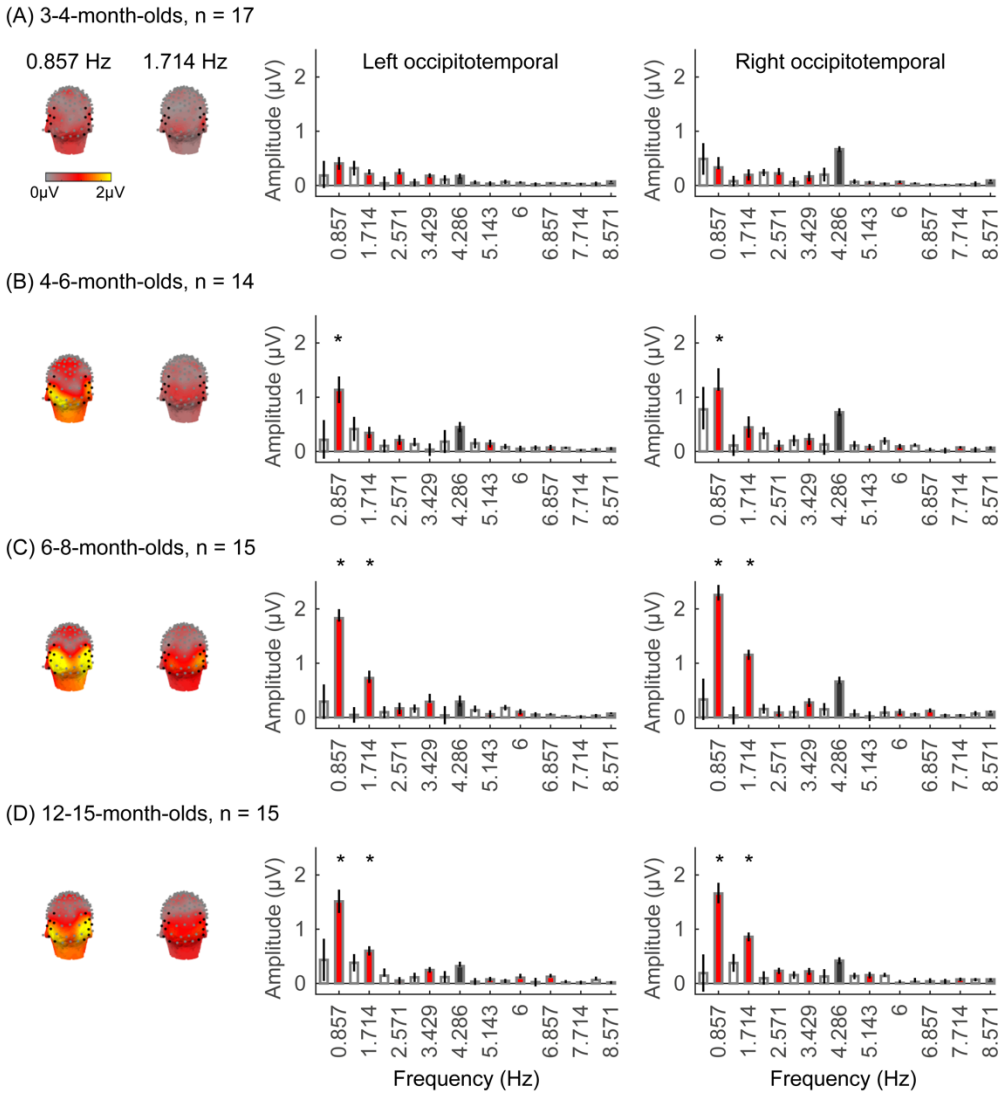


Figure 3. Face responses emerge over occipitotemporal electrodes after 4 months of age. Each panel shows mean responses at the category frequency and its harmonics across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds; n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. Left panels in each row: spatial distribution of response to category frequency at the 0.857 Hz and its first harmonic. Harmonic frequencies are indicted on the top. Right two panels in each row: mean Fourier amplitude spectrum across two ROIS: 7 left and 7 right occipitotemporal electrodes (shown in black on the left panel). Data are first averaged across electrodes in each participant and then across participants. Error bars: standard error of the mean across participants of an age group. Asterisks: significant amplitude vs. zero (p < .05, FDR corrected at 2 levels). Black bars: image frequency and harmonics; Colored bars: category frequency and harmonics. White bars: noise frequencies. Responses for the other categories (limbs, corridors, characters and cars) in Supplementary Figures S5-S8.

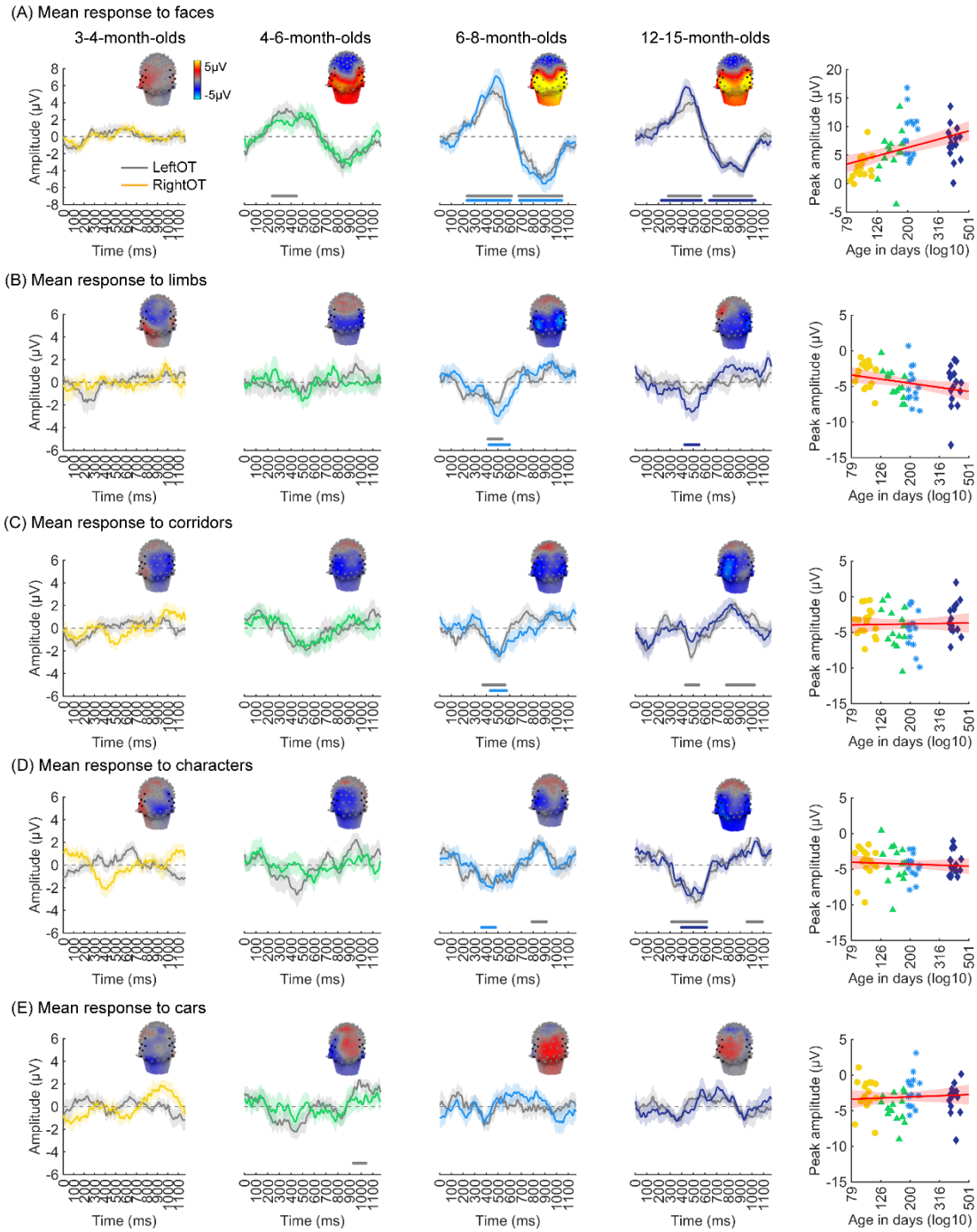


Figure 4. Temporal dynamics of category-selective responses as a function of age. Category-selective responses to (A) faces, (B) limbs, (C) corridors, (D) characters, and (E) cars over left and right occipitotemporal ROI. Data are averaged across electrodes of an ROI and across individuals. Left four panels in each row show the responses in the time domain for the four age groups. *Colored lines:* mean responses in the right occipitotemporal ROI. *Gray lines:* mean responses in the left occipitotemporal ROI. *Colored horizontal lines above x-axis:* significant responses relative to zero for the right OT ROI. *Gray horizontal lines above x-axis:* significant responses relative to zero for the left OT ROI. *Top*: 3-D topographies of the spatial distribution of the response to target category stimuli at a 483-500 ms time window after stimulus onset. *Right panel in each row:* amplitude of the peak deflection defined in a 400-700 ms time interval after stimulus onset. Each dot is a participant; Dots are colored by age group. *Red line:* linear mixed model (LMM) estimate of peak amplitude as a function of log10(age). *Shaded area:* 95% CI.

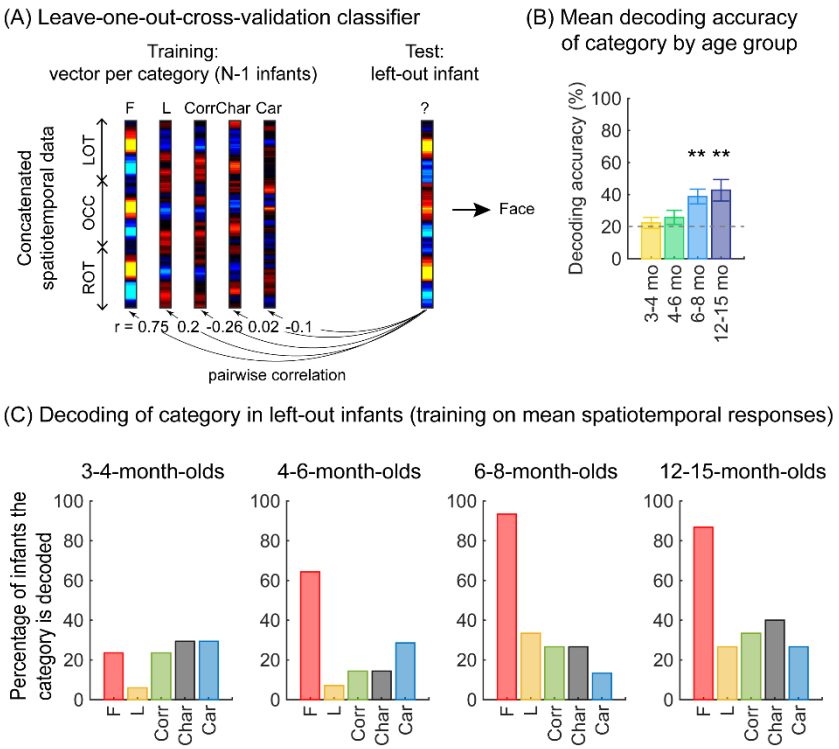


Figure 5. Successful decoding of faces from mean spatiotemporal responses starting from 4 months of age. (A) An illustration of winner-takes-all leave-one-out-cross validation (LOOCV) classifier from mean spatiotemporal response patterns of each category. Spatiotemporal patterns of response for each category are generated by concatenating the mean time courses from N-1 infants from three ROIs: left occipitotemporal (LOT), occipital (OCC), and right occipitotemporal (ROT). At each iteration, we train the classifier with the mean spatiotemporal patterns of each category from N-1 infants, and test how well it predicts the category the left-out infant is viewing from their spatiotemporal brain response. The winner-take-all (WTA) classifier determines the category based on the training vector that has highest pairwise correlation with the and test vectors. (B) Mean decoding accuracies across all five categories in each age group. *Asterisks:* significant decoding above chance level ($p < .01$, Bonferroni corrected, one-tailed). (C) Percentage of infants in each age group we could successfully decode for each category. *Dashed lines*: chance level.

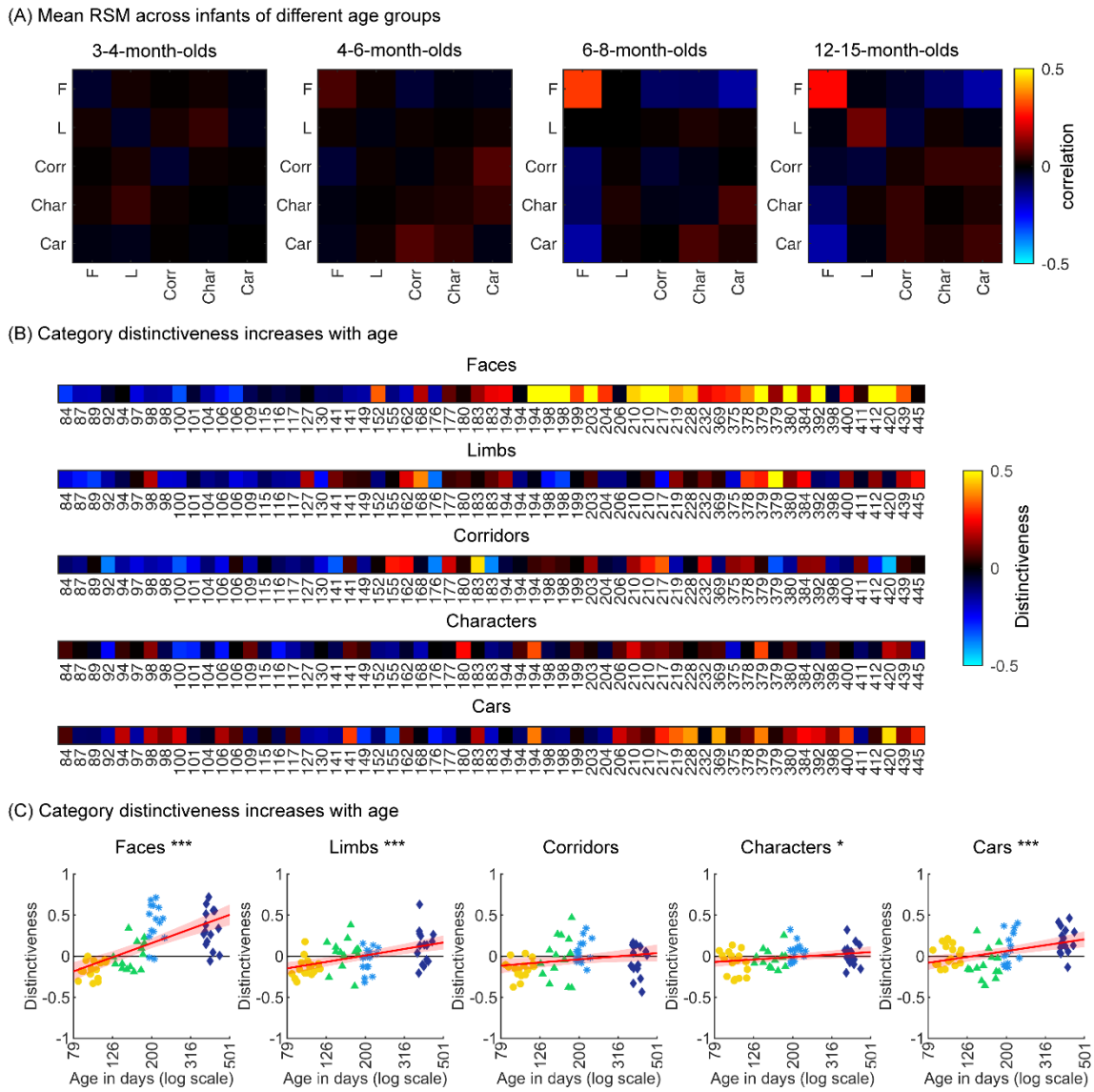


Figure 6. Individual split-half spatiotemporal pattern analyses reveal category information slowly emerges in the visual cortex after 6 months of age. (A) Representation similarity matrices (RSM) generated from odd/even split-halves of the spatiotemporal patterns of responses in individual infants. Spatiotemporal patterns for each category are generated by concatenating the mean time courses of each of 23 electrodes across left occipitotemporal (LOT), occipital (OCC), and right occipitotemporal (ROT). (B) *Top panel:* Category distinctiveness calculated for each infant and category by subtracting the mean between-category correlation values from the within-category correlation value. *Bottom panel:* Distinctiveness as a function of age; panels by category; Each dot is a participant. Dots are colored by age group. *Red line:* linear mixed model (LMM) estimates of distinctiveness as a function of log10(age). *Shaded area:* 95% CI.

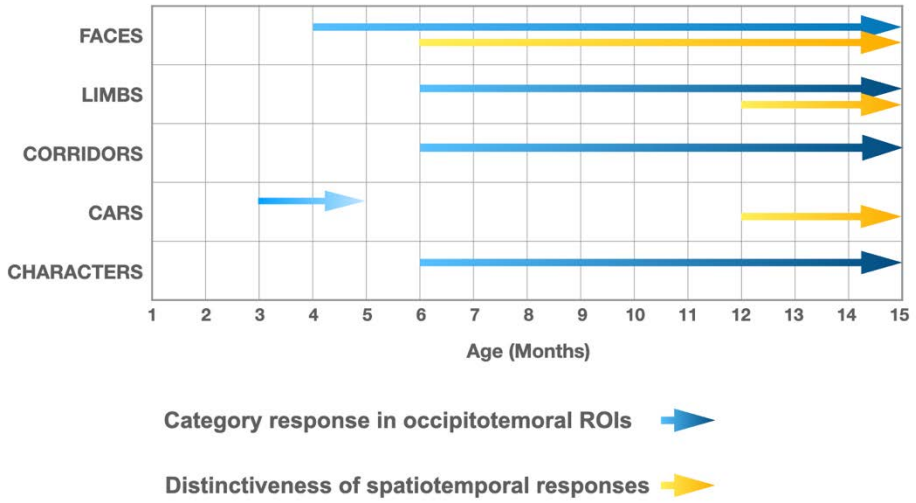


Figure 7. Selective responses to items of a category and distinctiveness in distributed patterns develop at different times during the first year of life. *Blue arrows*: presence of significant mean ROI category-selective responses in lateral occipital ROIs, combining results of analyses in the frequency and time domains. *Yellow arrows*: presence of significantly above zero distinctiveness in the distributed spatiotemporal response patterns across occipital and lateral occipital electrodes.

Supporting Information for
When Do Visual Category Representations Emerge in Infants' Brains?

Xiaoqian Yan[1,2,3], Sarah Tung[1+], Bella Fascendini[1+], Yulan Diana Chen[1,2], Anthony M. Norcia[1,2*], Kalanit Grill-Spector[1,2,4*]

5. Department of Psychology, Stanford University, Stanford, CA 94305, USA
6. Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA 94305, USA
7. Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China
8. Neurosciences Program, Stanford University, Stanford, CA 94305, USA

* co-senior-author

\+ equal contribution

Corresponding author: Xiaoqian Yan
Email: xqyan@fudan.edu.cn

This file includes:

- Supplementary Methods
- Supplementary Results
- Supplementary References
- Figures S1 to S8
- Tables S1 to S9
- Legends for Movies S1 to S5

Supplementary Materials

1. Demographic information

Supplementary Table 1. Demographic Information.

	3-4 MONTHS (N = 17)	**4-6 MONTHS (N = 14)**	**6-8 MONTHS (N = 15)**	**12-15 MONTHS (N = 15)**
Age at test (days)	84 - 117	127 - 183	194 - 232	369 - 445
Sex				
Female	7	7	6	4
Male	10	7	9	11
Race				
White	5	4	6	5
Black				
Asian		1	1	2
Mixed races	11	8	6	7
Unknown	1	1	2	1

Supplementary Table 2. Average number of valid epochs summed across all five categories for each age group before and after data pre-processing.

	3-4 MONTHS (N = 17)	4-6 MONTHS (N = 14)	6-8 MONTHS (N = 15)	12-15 MONTHS (N = 15)	ADULTS (N = 20)
Before pre-processing	281 (± 103)	270 (± 86)	346 (±111)	324 (±78)	600 (± 0)
After pre-processing	223 (± 89)	219 (± 73)	266 (± 91)	269 (± 77)	560 (± 37)
Ratio (after/before)	79%	81%	77%	83%	93%

897 **2. Low-level image properties analyses**

898 As images of items of visual categories vary both on low-level and high-level properties, and both
899 the EEG signal and babies' attention may be affected by low-level properties like contrast or luminance,
900 it is important to control for several low-level factors such as luminance, contrast, similarity spatial
901 frequency, as well as high-level properties such as familiarity across categories. Here we used a subset of
902 categories from[1], that did not contain familiar stimuli to our participants and were controlled for several
903 low-level factors using the SHINE toolbox [2]. To validate the effect of employing SHINE, we evaluated
904 several low-level metrics and tested if they differed across categories: (i) *Luminance:* mean gray level of
905 each image; (ii) *Contrast:* mean standard deviation of gray level values of each image; (iii) *Similarity:*
906 mean pixel-wise similarity between gray levels across pairs of images of a category, and (iv) *Spatial*
907 *frequency distribution*: we transformed each image to the Fourier Domain using FFT and calculated its
908 circular amplitude spectrum. We calculated these metrics for each image, then tested if contrast,
909 luminance, similarity metrics were significantly different across categories using pair-wise non-
910 parametric permutation t-tests (10,000 times, with Bonferroni correction). While it is difficult to
911 completely control multiple low-level metrics across categories[2], our analyses show that images are
912 largely controlled for several low-level metrics across categories, and that medians, means, and ranges
913 are matched across categories (**Fig 1B;** supplementary **Table 3**). With respect to differences in contrast
914 and luminance, there are no significant differences across categories except that limbs have slightly but
915 significantly lower luminance (*ps* < .05, Bonferroni corrected) and lower pairwise similarity (*ps* < .01) and
916 higher contrast (*ps* < .05, except for corridors) than other categories. For similarity, the medians and
917 ranges are similar across categories, but the outliers vary producing significant differences across
918 categories (permutation tests, *p* < .05, Bonferroni corrected). For spatial frequency distribution, the
919 means and standard deviations are similar across categories (Fig 1B-right). We further ran Kolmogrov-
920 Smirnov tests for each pair of categories with Bonferroni correction, our results revealed significantly
921 different distributions across category pairs (*ps* < .001, except for limbs and corridors).
922

923 **Supplementary Table 3. Mean (±SD) values of contrast, luminance and similarity metrices across**
924 **images within each five stimuli categories.**

	Faces	**Limbs**	**Corridors**	**Characters**	**Cars**
Contrast	0.469 (2.533e-4)	0.391 (3.91e-4)	0.404 (4.04e-4)	0.2856 (2.856e-4)	0.3269 (3.269e-4)
Luminance	0.6733 (0.0007)	0.6729 (0.0011)	0.6732 (0.0011)	0.6732 (0.0008)	0.6733 (0.0009)
Similarity	0.9855 (0.0063)	0.9555 (0.0039)	0.9569 (0.0051)	0.9563 (0.0033)	0.9570 (0.0044)

925
926

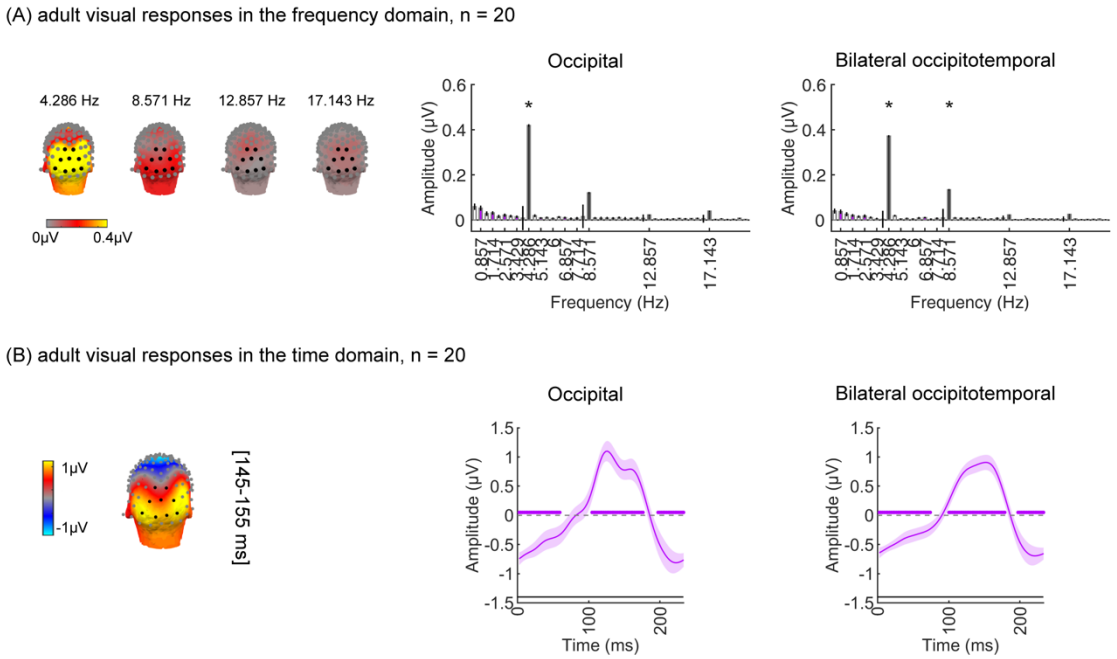
927 **3. Analyses of adult data**

928 In the study, one concern is that the amount of empirical data typically collected in infants is less
929 than in adults, which may compromise the experimental power to detect responses in infants as it may
930 affect the noise-bandwidth (sensitivity) of the frequency tagging analysis. Thus, we tested whether it is
931 possible to measure category-selective responses in 20 adults using the same experiment (**Fig 1**) and
932 same amount of data collected in infants using three types of analyses. By using the same number of
933 trials in adults as those obtained from our infant participants for data analyses, our goal was to test that
934 we had sufficient power to detect categorical responses from infants using the experimental paradigm.
935 We expect temporal and amplitude differences between adults and infants[3] as infants have immature
936 brains and skulls. For example, cortical gyrification, which determines the orientation of the electrical

fields generated in certain part of the brain region on the scalp, still undergoes development during the first two years of infants' lives[4]. Second, adults' skulls are thicker and have lower conductivity than infants' skulls, thus electrical signals on their scalp are lower than infants[5]. Nonetheless, we tested if we could in principle detect category information in adults with the same amount of data as infants. We reasoned that if category information can be detected in adults and signals are stronger in infants then we should have the power to detect category information in infants.

Visual Responses: We averaged the preprocessed data across two consecutive epochs (2.3334-s) and across conditions for each individual to measure visual responses to all images that were presented at 4.286 Hz and then transformed the data to the frequency domain. Supplementary **Fig S1A** shows the three-D topographies of group-averaged responses at 4.286 Hz and its first 3 harmonics (left panel) and the group-averaged Fourier amplitude spectrum over middle occipital and occipitotemporal ROIs (right panel). We used Hotelling's T2 statistic to test the amplitude significance and found significant visual responses in adults at 4.286 Hz over the occipital ROI containing 9 electrodes over early visual cortex ($p < .05$, FDR corrected at 4 levels) and significant visual responses at 4.286 Hz and its first two harmonics over the occipitotemporal ROI (14 electrodes) (*ps* < .05).

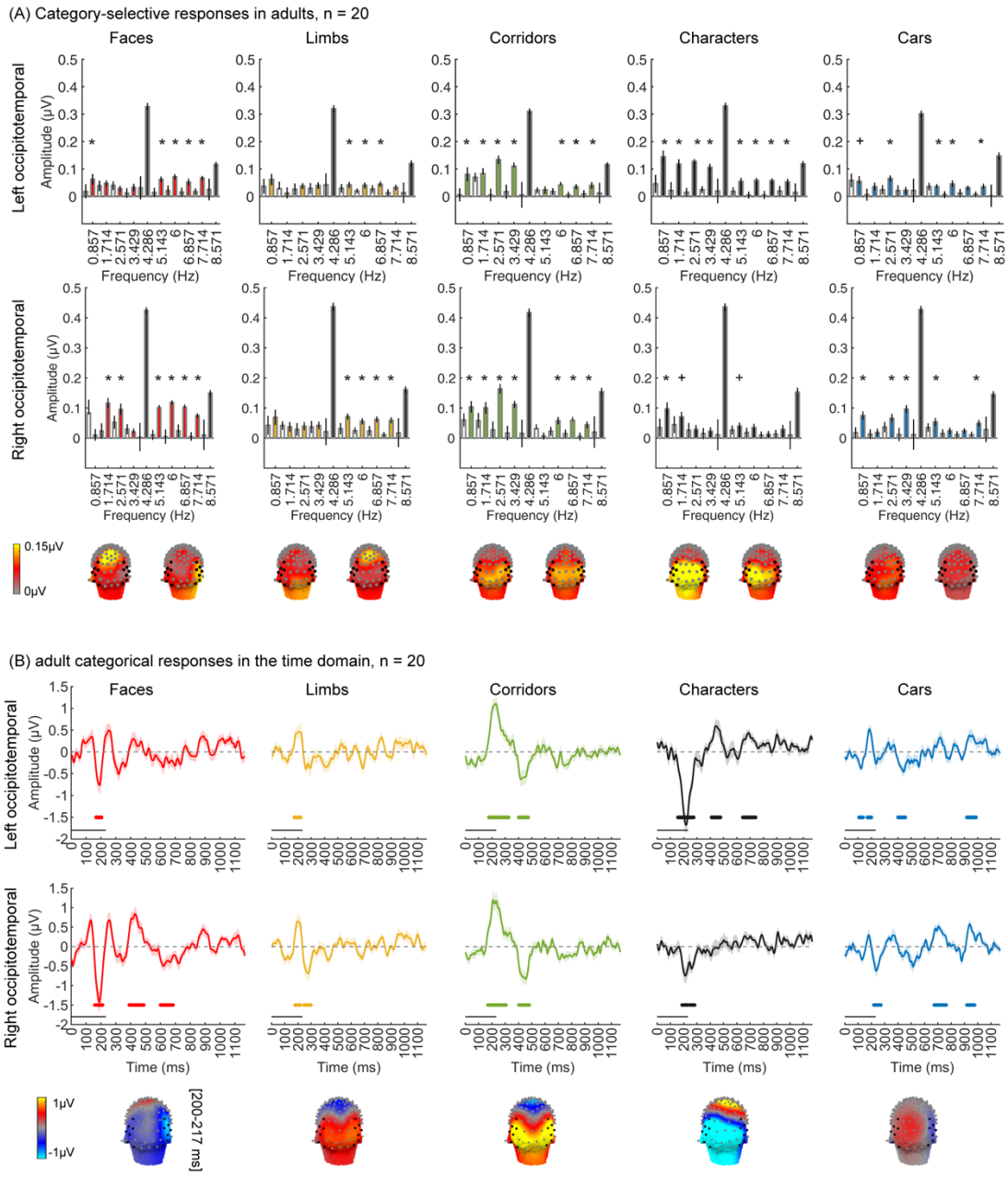
Examining the temporal waveform of the visual response (filtered at the image frequency and its harmonics), we found a relatively slow waveform with one peak, after 100 ms since stimulus onset (Supplementary **Fig S1B**).



Supplementary Figure 1. Robust visual and categorical responses recorded over occipitotemporal and occipital cortex in 20 adults. (A) *Left panel:* spatial distribution of visual response at 4.286 Hz and harmonic. Harmonic frequencies are indicted on the top. *Right panel:* mean Fourier amplitude spectrum across 14 electrodes in the occipitotemporal and 9 electrodes in the occipital ROIs. *Error bars:* standard error of the mean across participants. *Black bars:* image frequency and harmonics; Purple bars: category frequency and harmonics; Asterisks: significant response amplitude from zero at $p_{FDR} < .05$. (B) *Left:* spatial distribution of visual responses at time window 145-155 ms. Right: Mean visual responses over two ROIs in the time domain. Waveforms are shown for a time-window of 233-ms during which 1 image are shown. *Shaded area*: standard error of the mean across participants. Blank

line at around y = -1.5: stimulus onset duration. To define time windows in which amplitudes were significantly different from zero, we used a cluster-based nonparametric permutation t-test (1000 permutations, with a threshold of $p < 0.05$, two-tailed) on the post-stimulus onset time-points (0-1167 ms) [6,7].

Visual Category responses: As adults have mature category-selective regions, we next tested if using our SSVEP paradigm and the same amount of data as in infants, we identify significant category-selective responses in adults. First, we analyzed the mean response at the category frequency and its harmonics in lateral occipitotemporal regions of interest (ROIs). This is a selective response as it reflects the relative response to the target category (generalized across exemplars) above the general visual response to images of other categories. Despite lower response amplitudes in adults than infants, using the same amount of data as infants, adults show significant category-selective responses to each of these five categories (Supplementary **Fig S2A** for all categories). Supplementary **Fig S2A** shows the group-averaged Fourier amplitude spectrum over left and right occipitotemporal ROIs (top panels) and the three-D topographies of group-averaged responses at 0.857 Hz and its first harmonic (bottom panel). We used Hotelling's T2 statistic to test the amplitude significance and found significant above-zero category-selective responses in adults at 0.857Hz and its multiple harmonics over bilateral occipitotemporal ROIs containing 7 electrodes each (*ps* < .05, FDR corrected over 8 levels).



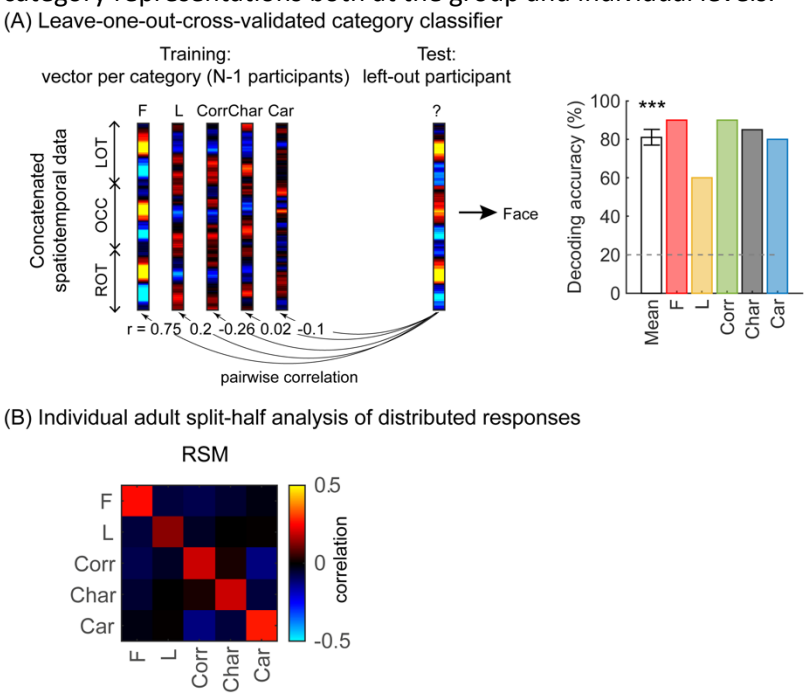
Supplementary Figure 2. Adult control: using the same amount of data as infants reveals strong category-selective responses in adults' occipitotemporal cortex. (A) Mean Fourier Amplitude Spectrum across 7 (left OT: 57, 58, 59, 63, 64, 65, 68; Right OT: 90, 91, 94, 95, 96, 99) electrodes in bilateral occipitotemporal ROIs. Data are first averaged in each participant and then across 20 participants. *Error bars*: standard error of the mean across participants. *Black bars*: visual response at image frequency and harmonics; *Colored bars:* categorical response at category frequency and harmonics. *Asterisks:* significant response amplitude from zero at $p_{FDR} < .05$ for category harmonics. Crosses: significant response amplitude from zero at $p < .05$ with no FDR correction. *Black dots:* ROI channels used in analysis. (B) Mean category-selective responses in the time domain. Data are averaged across electrodes of each of the left and right occipitotemporal ROI in each participant and then across participants. *Colored lines along x-axis at y = -1.5:* significant deflections against zero (calculated with a cluster-based method, see methods part). *Black line above x-axis:* stimulus onset duration. *Bottom panel:* spatial distribution of category-selective responses at time window 200-217 ms.

By transforming the filtered data at the category frequency and its harmonics to examine the temporal waveform of each category. Notably, each category generates a unique topography (bottom panels) at 200-217 ms after stimulus onset (Supplementary **Fig 2B**). For faces, we found an early negative deflection peaking ~200 ms after stimulus onset in both the left and right occipitotemporal ROIs, with the right occipitotemporal ROI showing a numerically larger mean response amplitude than the left (Supplementary **Fig S2B**-faces). Similarly, we found an early negative peak at around 200 ms for characters, and a left hemisphere dominance (at the 162-248 ms and 398-474 ms time windows (Supplementary **Fig S2B**-characters). For both limbs and corridors, there was an early positive waveform peaking at around 200 ms in both hemispheres with no hemispheric differences.

Second, we examined in the same 20 adults whether the spatiotemporal pattern of brain responses evoked by different visual categories are distinct from one another and are reliable across participants, using a leave-one-out-cross-validation (LOOCV) classifier approach with spatiotemporal timeseries concatenated with mean temporal waveforms from 3 ROIs: (Supplementary **Fig 3A-left**). Mean LOOCV classification performance was around 80% and significantly above chance in adults (Supplementary **Fig 3A-right**). This classification was associated with correct decoding of all 5 categories from distributed responses in the majority of participants.

Third, we examined in the same 20 adults if there is reliable category information in spatiotemporal distributed in each individual. Distributed spatiotemporal responses were measured in individual participants over 23 occipital and lateral occipital ROIs using split half of data. We computed the representations similarity matrix (RSM) across split halves and then calculated category distinctiveness by subtracting mean between category similarity from within category similarity for each category. Mean RSM across 20 adults shows that spatiotemporal patterns are more similar across items of a category than across items of different categories (Supplementary **Fig 3B-left**), and the category distinctiveness scores are significantly above zero for all five categories: faces, $t_{(19)} = 11.18$, $p_{FDR} < .05$; limbs, $t_{(19)} = 5.01$, $p_{FDR} < .05$; corridors, $t_{(19)} = 5.58$, $p_{FDR} < .05$; characters, $t_{(19)} = 6.56$, $p_{FDR} < .05$; and cars, $t_{(19)} = 10.59$, $p_{FDR} < .05$.

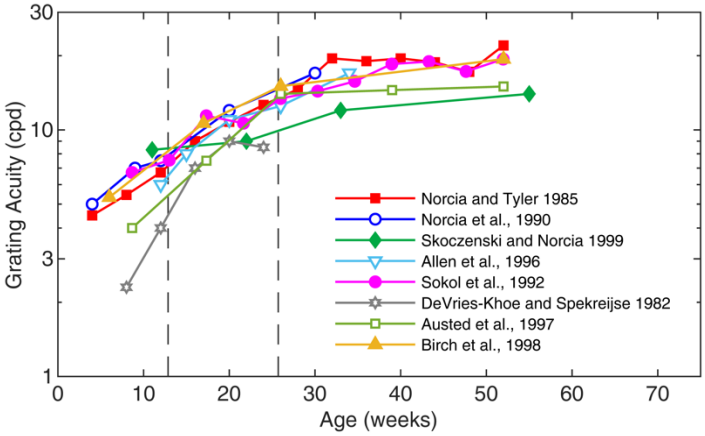
Together, these analyses suggest that the experimental paradigm has sufficient power to identify category representations both at the group and individual levels.



Supplementary Figure 3: Adult control: using the same amount of data as infants reveals distributed category-selective responses in adults' occipitotemporal cortex. (A) *Left:* An illustration of winner-takes-all leave-one-out-cross-validated classifier (LOOCV) using the spatiotemporal response patterns of each category. Spatiotemporal patterns of response for each category are generated by concatenating the mean time courses from each of the three ROIs: left occipitotemporal (LOT), occipital (OCC), and right occipitotemporal (ROT). At each iteration, we train the classifier with the mean spatiotemporal patterns of each category from N-1 participants, and test how well it predicts the category the left-out participant is viewing from their spatiotemporal brain response. This is a winner-take-all classifier which predicts the category based on the highest pairwise correlation between the training and testing vectors. *Right: White:* mean decoding accuracy across all five categories. In adults, this is significantly above chance level ($t_{(19)} = 15.4$, $p < .001$). *Colored:* decoding accuracy per category. (B) *Left:* Average adult representation similarity matrix (RSM) for odd/even splits of spatiotemporal patterns of categorical over 23 electrodes in the LOT, OCC, ROT. RSMs were generated in each participant and then averaged across all participants. *Diagonal:* correlation of distributed responses across different exemplars of the same category; *Off-diagonal:* correlations across different exemplars from different categories. *Acronyms:* F: faces; L: limbs; Corr: corridors; Char: characters; Car: Cars.

4. Validation of experimental paradigm

As visual acuity develops during the first year of life[8,9] (Supplementary **Fig S4**), one concern is that our controlled natural, gray-level stimuli may not be distinguishable to infants. Measurements of visual evoked potentials[8,9] suggest that visual acuity in 3-month-olds is around 5–8 cycles per degree (cpd) and in 6-month-olds around 10–16 cpd (Supplementary **Fig S4**). Thus, to simulate how our images may appear to infants, we filtered all images at 5 cpd. Despite being blurry, images of different categories are distinguishable and individual items retain their identity by visualization (Supplementary **Movies 1-5**).



Supplementary Figure 4. Grating acuity as a function of age measured with a swept spatial frequency technique combining EEG. (A) Acuity growth functions are similar across studies, with acuity increasing from 5-8 cycles per degree (cpd) in 3-month-olds to around 10-16 cpd in 6-month-olds. This figure is adapted from[10].

5. Linear mixed model (LMM) analyses of visual responses (associated with **Figure 2**)

Supplementary Table 4: Peak latency of visual responses by age and time-window (window 1: 60-90 ms; window 2: 90-160 ms for 3-4-month-olds, and 90-110 ms for other groups)

Formula: Peak latency ~ 1 + log10(Age) × time window + (1|Participant)

Significant effects are indicated by asterisks.

Parameter	*β*	*CI*	*df*	*t*	*p*
Intercept	-54.61	-100.82, -8.41	118	-2.34	.021*
Age	39.77	19.51, 60.02	118	3.89	.00017***
Window	141.68	112.92, 170.45	118	9.75	7.69e-17***
Age×Window	-45.78	-58.39, -33.17	118	-7.19	6.39e-11**

Supplementary Table 5: Peak latency of visual responses by age at each of the two time-windows

Formula: Peak latency ~ 1 + log10(Age) + (1|Participant)

Significant effects are indicated by asterisks.

Window	*Parameter*	*β*	*CI*	*df*	*t*	*p*
Window1	Intercept	90.19	75.66, 104.72	59	12.42	4.22e-18***
	Age	-7.44	-13.82, -1.06	59	-2.33	.02*
Window2	Intercept	218.09	196.05, 240.13	59	19.80	9.64e-28***
	Age	-46.91	-56.56, -37.27	59	-9.73	7.06e-14***

Supplementary Table 6: Analysis of peak amplitude of visual responses by age and time-window

Formula: Peak amplitude ~ 1 + log10(Age) × time window + (1|Participant)

Significant effects are indicated by an asterisk.

Parameter	*β*	*CI*	*df*	*t*	*p*
Intercept	-18.69	-32.35, -5.03	118	-2.71	.008**
Age	3.93	-2.05, 9.92	118	1.30	.20
Window	17.24	8.66, 25.82	118	3.98	.0001***
Age×Window	-4.90	-8.66, -1.14	118	-2.58	.011*

Supplementary Table 7: Peak amplitude of visual responses by age at each of the two time-windows

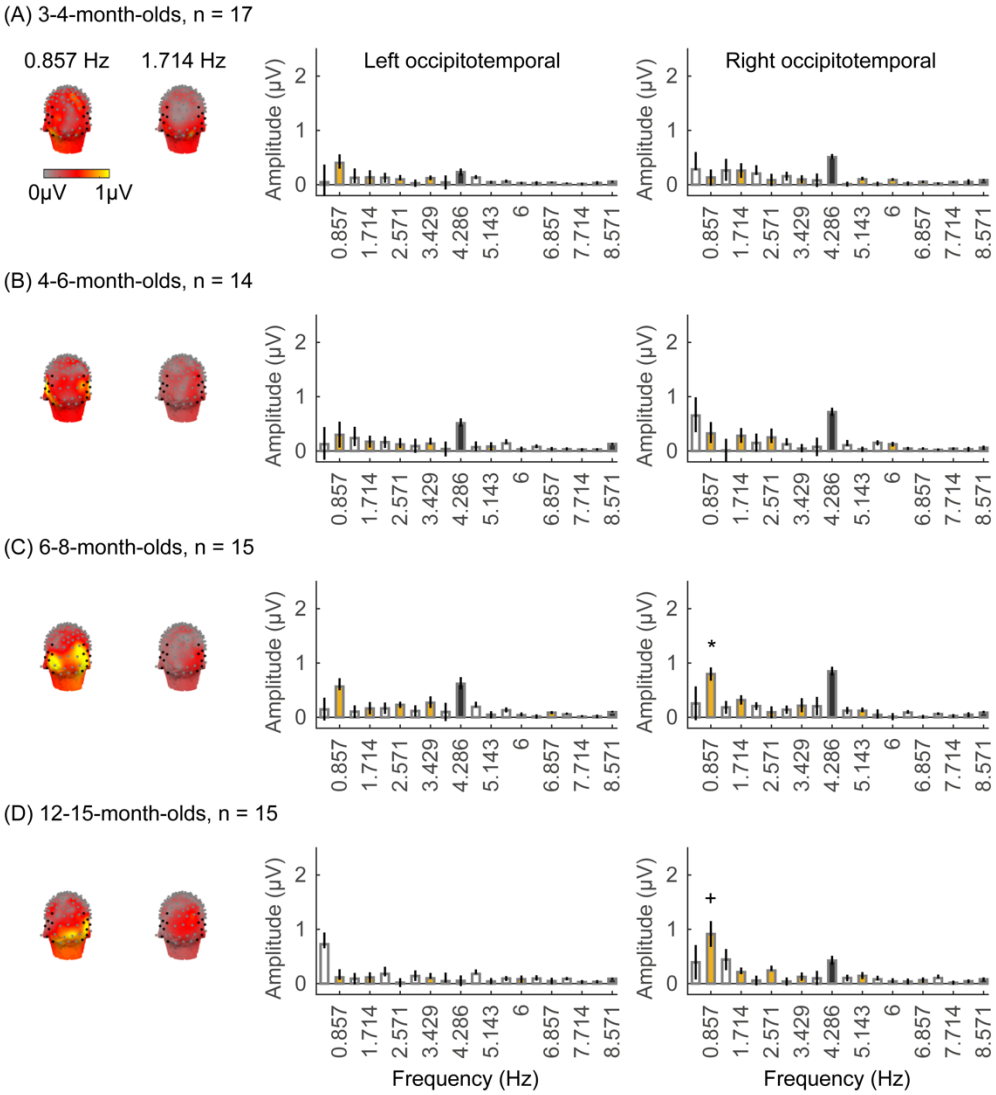
Formula: Peak amplitude ~ 1 + log10(Age) + (1|Participant)

Significant effects are indicated by asterisks.

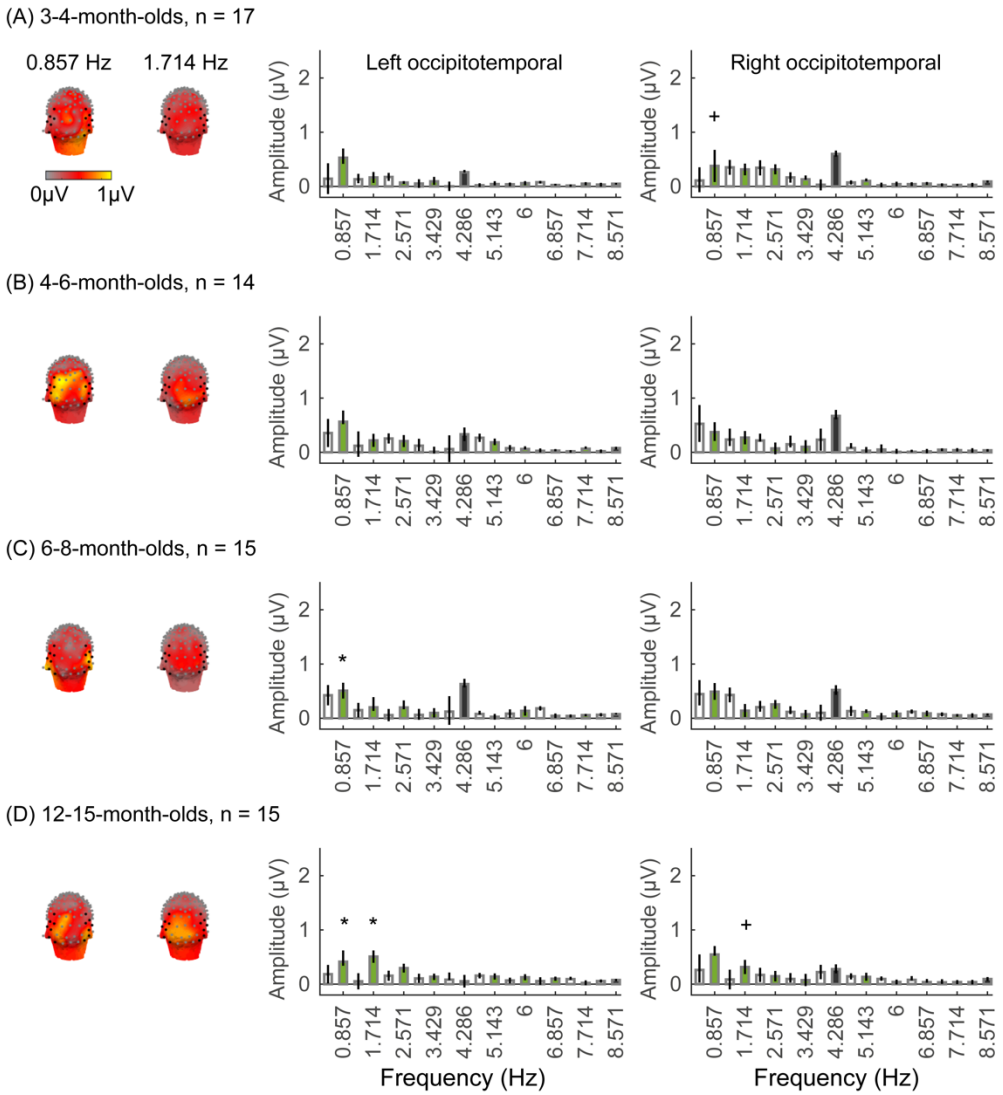
Window	*Parameter*	*β*	*CI*	*df*	*t*	*p*
Window1	Intercept	1.69	-3.42, 6.79	59	0.66	.51
	Age	0.91	-1.33, 3.15	59	0.82	.42
Window2	Intercept	11.16	4.80, 17.51	59	3.51	0.0009***
	Age	-3.59	-6.38, -0.81	59	-2.58	.012*

6. Frequency domain analyses of infants' categorical responses to limbs, corridors, characters and cars (associated with **Fig 3**)

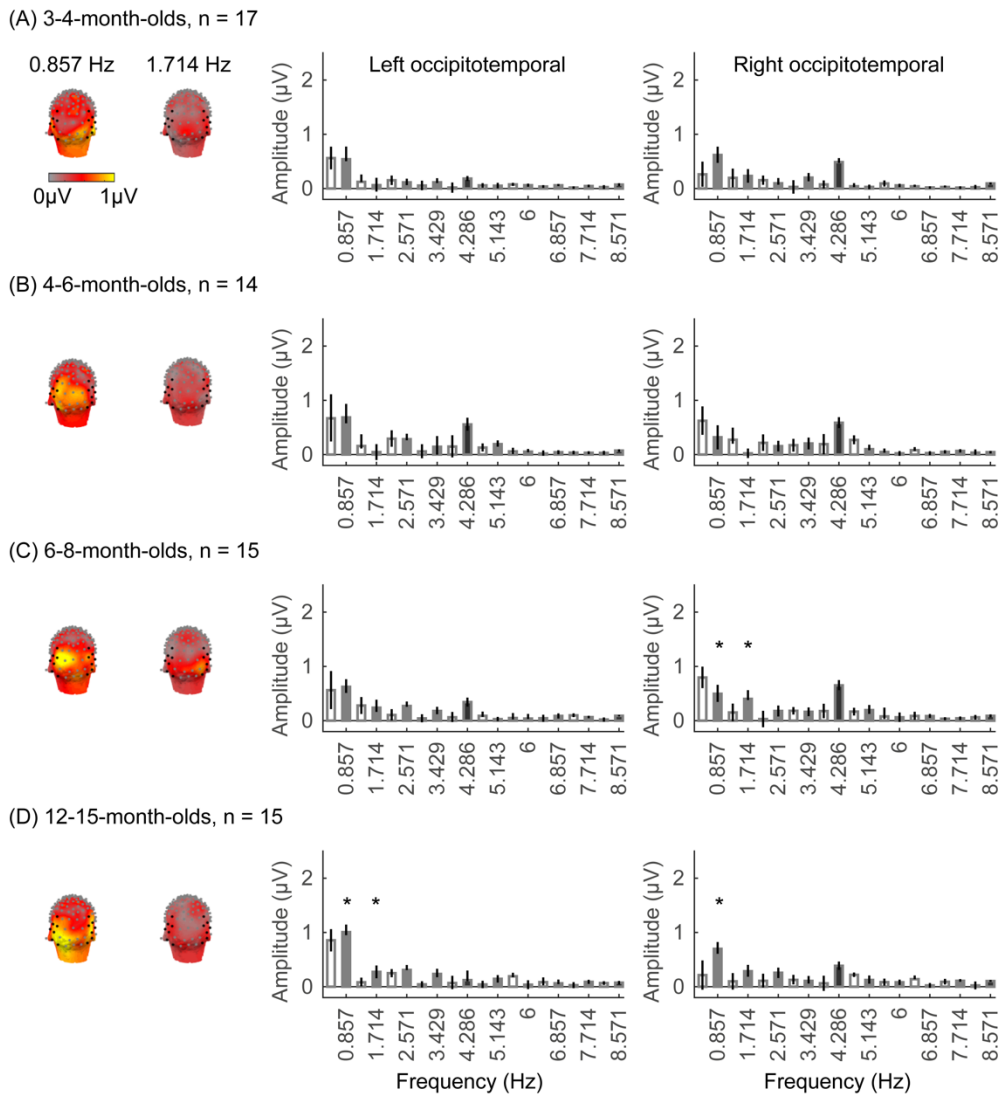
Fig 3 shows the group-averaged categorical response to faces. Supplementary **Figs S5-8** show group-averaged categorical responses to the rest four conditions other than faces in four age groups. We found significant responses in 6-8-month-olds for limbs, corridors, and characters, and in 12-15 months for corridors and characters.



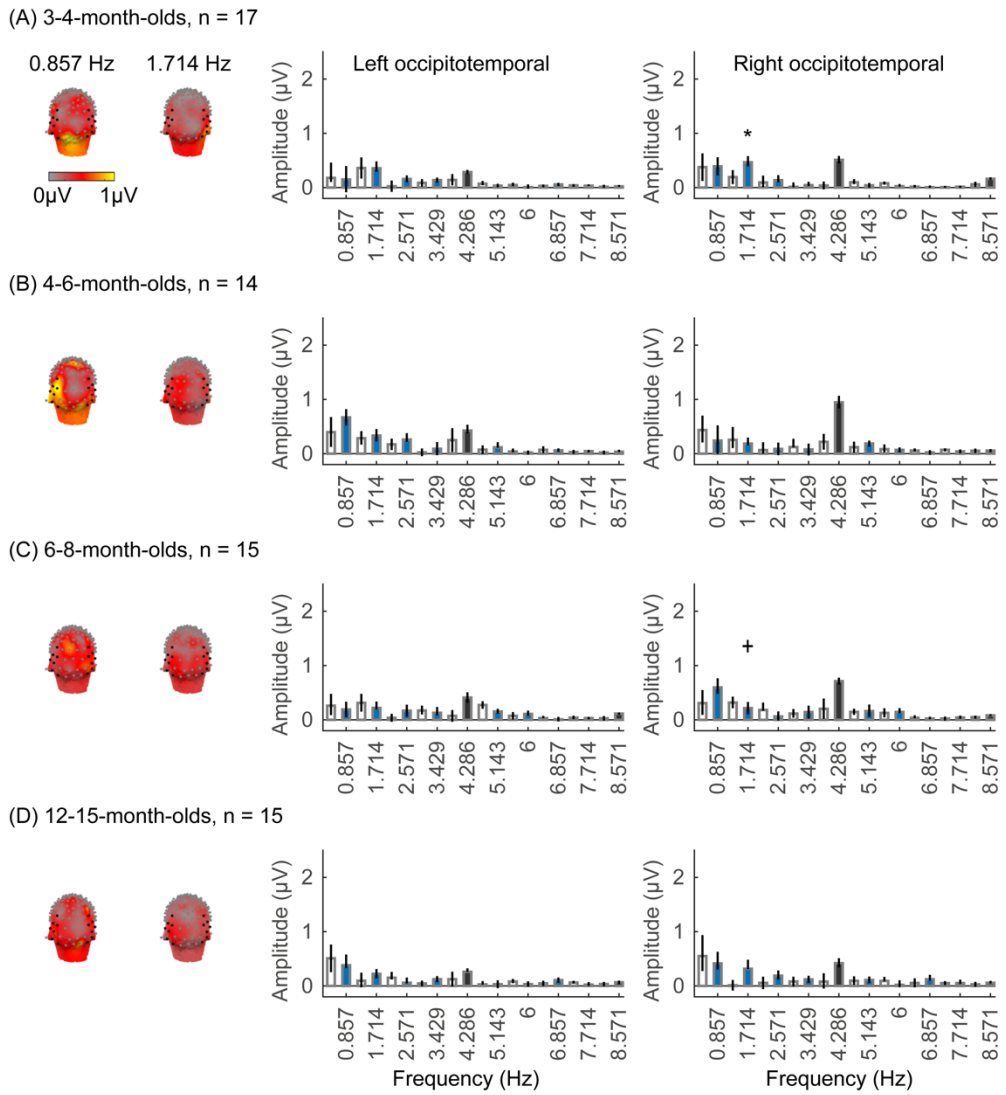
Supplementary Figure 5. Limb responses emerge over occipitotemporal electrodes after 6 months of age. Each panel shows mean responses at the category frequency (0.857 Hz) and its harmonics across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds; n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. *Left panels in each row:* spatial distribution of categorical response at 0.857 Hz and its first harmonic. Harmonic frequencies are indicted on the top. *Right two panels in each row*: mean Fourier Amplitude Spectrum across 7 left occipitotemporal electrodes and 7 right occipitotemporal (shown in black on the left panel). Data are first averaged in each participant and then across participants. *Error bars:* standard error of the mean across participants in an age group. *Asterisk:* significant amplitude vs. zero ($p < .05$, FDR corrected). Cross: significant amplitude vs. zero ($p < .05$, with no FDR correction). *Black bars:* image frequency and harmonics; *Colored bars:* category frequency and harmonics.



Supplementary Figure 6. Corridor responses emerge over occipitotemporal electrodes after 6 months of age. Each panel shows mean responses at the category frequency and its harmonics across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds; n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. *Left panels in each row:* spatial distribution of categorical response at 0.857 Hz and its first harmonic. Harmonic frequencies are indicted on the top. *Right two panels in each row*: mean Fourier Amplitude Spectrum across 7 left occipitotemporal electrodes and 7 right occipitotemporal (shown in black on the left panel). Data are first averaged in each participant and then across participants. *Error bars:* standard error of the mean across participants in an age group. *Asterisks:* significant amplitude vs. zero ($p < .05$, FDR corrected). Crosses: significant amplitude vs. zero ($p < .05$, with no FDR correction). *Black bars: image frequency and harmonics; Colored bars: category frequency and harmonics.*



Supplementary Figure 7. Significant characters responses found over occipitotemporal electrodes at 12-15 months of age. Each panel shows mean responses at the category frequency (0.857 Hz) and its harmonics across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds; n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. *Left panels in each row:* spatial distribution of categorical response at 0.857 Hz and its first harmonic. Harmonic frequencies are indicted on the top. *Right two panels in each row*: mean Fourier Amplitude Spectrum across 7 left occipitotemporal electrodes and 7 right occipitotemporal (shown in black on the left panel). Data are first averaged in each participant and then across participants. *Error bars:* standard error of the mean across participants in an age group. *Asterisks:* significant amplitude vs. zero ($p < .05$, FDR corrected). *Black bars:* image frequency and harmonics; *Colored bars:* category frequency and harmonics.



Supplementary Figure 8. Significant cars responses found over occipitotemporal electrodes at 3-4 months of age. Each panel shows mean responses at the category frequency and its harmonics across infants in an age group. (A) 3-4-month-olds, n=17; (B) 4-6-month-olds; n=14; (C) 6-8-month-olds, n=15; (D) 12-15-month-olds, n=15. *Left panels in each row:* spatial distribution of categorical response at 0.857 Hz and its first harmonic. Harmonic frequencies are indicted on the top. *Right two panels in each row*: mean Fourier Amplitude Spectrum across 7 left occipitotemporal electrodes and 7 right occipitotemporal (shown in black on the left panel). Data are first averaged in each participant and then across participants. *Error bars:* standard error of the mean across participants in an age group. *Asterisk:* significant amplitude vs. zero ($p < .05$, FDR corrected). Cross: significant amplitude vs. zero ($p < .05$, with no FDR correction). *Black bars:* image frequency and harmonics; *Colored bars:* category frequency and harmonics.

7. Linear mixed model (LMM) analyses of category-selective responses.

Supplementary Table 8: Analysis of peak amplitude of waveforms of category responses by age and category.

Separate LMMs were done separately for the left occipitotemporal (OT) and right OT ROIs.

Formula: Peak amplitude ~ 1 + log10(Age) × category + (1|Participant);

Peak latency ~ 1 + log10(Age) × category + (1|Participant).

Significant effects are indicated by an asterisk.

ROI/metric	Parameter	*β*	CI	df	*t*	*p*
Left OT/ amplitude	Intercept	-1.35	-7.36, 4.66	301	-0.44	.66
	Age	2.62	-0.02, 5.25	301	1.95	.052
	Category	0.18	-1.58, 1.93	301	0.20	.84
	Age×Catergory	-0.20	-0.97, 0.57	301	-0.51	.61
Left OT/ latency	Intercept	730.29	477.96, 982.61	301	5.70	2.9e-8***
	Age	-97.17	-207.76, 13.43	301	-1.73	.08
	Category	-43.35	-119.43, 32.73	301	-1.12	.26
	Age×Catergory	20.24	-13.11, 53.58	301	1.19	.23
Right OT/ amplitude	Intercept	-7.39	-14.44, -0.36	301	-2.07	.04*
	Age	5.53	2.45, 8.62	301	3.53	.0005***
	Category	2.19	0.06, 4.3	301	2.02	.04*
	Age×Category	-1.09	-2.00, -0.14	301	-2.26	.02*
Right OT/ latency	Intercept	922.47	667.95, 1177	301	7.13	7.38e-12***
	Age	-173.17	-284.73, -61.61	301	-3.05	.002**
	Category	-64.41	-141.15, 12.33	301	-1.65	.10
	Age×Catergory	28.49	-5.15, 62.12	301	1.67	.10

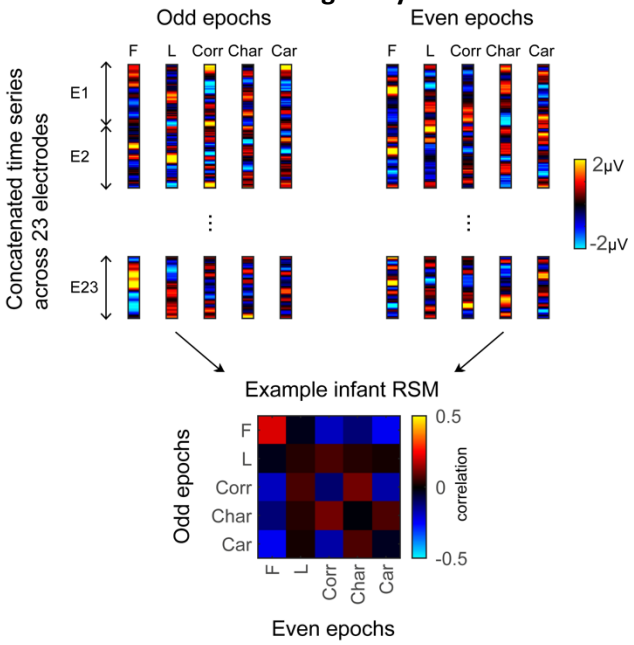
Supplementary Table 9: Analysis of peak amplitude of waveforms of category responses for each category in the right OT ROI.

Formula: Peak amplitude ~ age + (1|participant)

Significant effects are indicated by an asterisk.

Category	*Parameter*	*β*	*CI*	*df*	*t*	*p*
Faces	Intercept	-10.43	-17.81, -3.05	59	-2.83	.006**
	Age	7.27	4.03, 10.51	59	4.50	3.30e-5***
Limbs	Intercept	2.10	-3.62, 7.83	59	0.74	.46
	Age	-2.90	-5.41, -0.38	59	-2.31	.02*
Corridors	Intercept	-4.65	-11.09, 1.81	59	-1.44	.15
	Age	0.35	-2.47, 3.18	59	0.25	.80
Characters	Intercept	-2.79	-8.06, 2.49	59	-1.06	.3
	Age	-0.66	-2.97, 1.65	59	-0.57	.57
Cars	Intercept	-4.87	-11.46, 1.73	59	-1.48	.15
	Age	0.78	-2.11, 3.67	59	0.54	.59

8. Individual level decoding analysis



Supplementary Figure 8. Illustration of the winner-takes-all (WTA) classifier. In each individual, the time series data is split into odd and even trials. We concatenate the time series data from 23 electrodes in the left occipitotemporal, occipital, and right occipitotemporal ROIs in to a pattern vector for each split half and each condition. The classifier is trained on one half of the data (i.e. odd or even trials) and tested on how well it could predict the rest half of the data (i.e. even or odd trials) for each individual. The bottom shows the representation similarity matrix (RSM) in an example infant. Each cell indicates the correlation between distributed responses to different images of the same (on diagonal) or different (off diagonal) categories. *F: Faces; L: Limbs; Corr: Corridors; Char: Characters; Car: Cars.*

Supplementary Movie 1. Movie showing gaussian lowpass filtered face stimuli shown in the experiment at 5 cycles per degree (cpd).

Supplementary Movie 2. Movie showing gaussian lowpass filtered limb stimuli shown in the experiment at 5 cycles per degree (cpd).

Supplementary Movie 3. Movie showing gaussian lowpass filtered corridor stimuli shown in the experiment at 5 cycles per degree (cpd).

Supplementary Movie 4. Movie showing gaussian lowpass filtered character stimuli shown in the experiment at 5 cycles per degree (cpd).

Supplementary Movie 5. Movie showing gaussian lowpass filtered car stimuli shown in the experiment at 5 cycles per degree (cpd).

SI References

1. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *Journal of Neuroscience* **35**, 12412–12424 (2015).
2. Willenbockel, V. *et al.* Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods* **42**, 671–684 (2010).
3. Xie, S. *et al.* Visual category representations in the infant brain. *Current Biology* **32**, 5422-5432.e6 (2022).
4. Li, G. *et al.* Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. *J Neurosci* **34**, 4228–4238 (2014).
5. Grieve, P. G., Emerson, R. G., Fifer, W. P., Isler, J. R. & Stark, R. I. Spatial correlation of the infant and adult electroencephalogram. *Clinical Neurophysiology* **114**, 1594–1608 (2003).
6. Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W. & Norcia, A. M. Cue-Invariant Networks for Figure and Background Processing in Human Visual Cortex. *J. Neurosci.* **26**, 11695–11708 (2006).
7. Blair, R. C. & Karniski, W. An alternative method for significance testing of waveform difference potentials. *Psychophysiology* **30**, 518–524 (1993).
8. Norcia, A. M. & Tyler, C. W. Spatial frequency sweep VEP: Visual acuity during the first year of life. *Vision Research* **25**, 1399–1408 (1985).
9. Norcia, A. M., Tyler, C. W. & Hamer, R. D. Development of contrast sensitivity in the human infant. *Vision Research* **30**, 1475–1486 (1990).
10. Norcia, A. M. Development of vision in infancy. in *Adler's Physiology of the Eye* 531–551 (Elsevier Health Sciences, 2011).