The infant brain rapidly entrains to visual statistical regularities during stimulus exposure

Chiara Capparini*[1,2], Lauréline Fourdin[1,2], Vincent Wens[1,3], Pauline Dontaine[4], Alec Aeby[2,4], & Julie Bertels[1,2]

[1] Laboratoire de Neuroimagerie et Neuroanatomie translationnelles (LN2T), ULB Neuroscience Institute (UNI), Université libre de Bruxelles (ULB), 1050 Brussels, Belgium

[2] ULBabyLab – Center for Research in Cognition and Neurosciences (CRCN), ULB Neuroscience Institute (UNI), Université libre de Bruxelles (ULB), 1050 Brussels, Belgium

[3] Service of Translational Neuroimaging, Hôpital Erasme, Hôpital Universitaire de Bruxelles (HUB), 1050 Brussels, Belgium

[4] Department of Pediatric Neurology, Hôpital Erasme and Hôpital Universitaire des Enfants Reine Fabiola (HUDERF), Hôpital Universitaire de Bruxelles (HUB), Université Libre de Bruxelles (ULB), 1050 Brussels, Belgium

*Corresponding author. Email address: chiara.capparini@ulb.be

Abstract

Statistical learning (SL) has been studied quite extensively in infancy. Still, most evidence relies on post-exposure behavioural tasks whose directionality (familiarity vs. novel effects) may not be straightforward to predict nor to interpret. In addition, these tasks do not tell anything about the online learning dynamics and may be influenced by memory effects. In this work, we investigated online SL mechanisms by tracking neural entrainment to visual regularities in a group of 4- to 6-month-old infants exposed to a stream of shapes presented at 6 Hz. Shapes were either organized in doublets or presented randomly. Results revealed neural entrainment at the base frequency of 6 Hz and harmonics did not differ across conditions, confirming that infants attended to the stimuli similarly across conditions. On the other hand, entrainment at the doublet frequency of 3 Hz and harmonics varied across conditions and trials. Infants showed greater occipital entrainment to the doublet frequency in the deterministic condition than in the random one, especially over the first trials of exposure. This suggests that the brain can detect visual regularities from early infancy. Further, this sensitivity emerged early over the exposure period and did not show a learning curve. Hence, considering its time course and the brain regions involved, neural entrainment at the doublet frequency seems to reflect a bottom-up detection mechanism rather than a learning process. These findings are crucial to better understand how infants extract regularities during stimulus exposure and what neural entrainment can reveal in a visual SL task.

Keywords: Statistical learning, Implicit learning, Infancy, EEG, Frequency tagging, Neural entrainment.

1. Introduction

Detecting and extracting regularities in our environment is of foremost importance from the earliest stages of development. The ability to identify structural and statistical relations in a continuous stream of sensory events, that is Statistical Learning (SL; Saffran et al., 1996), seems already present in newborns (Benjamin et al., 2022; Bulf et al., 2011; Flo et al., 2022; Gervain et al., 2008). It has a pivotal role for preverbal infants, as it allows them to learn regularities in a sensory array implicitly, without the need of any instruction. During development, this mechanism has clear implications for language acquisition and visual computations in the natural environment. For instance, it helps parsing a continuous speech that does not contain systematic boundaries into meaningful word units (Erickson & Thiessen, 2015; Pelucchi et al., 2009). Additionally, SL mechanisms play a role when tracking the visual regularities of human action sequences, allowing to organise continuous motion in a meaningful way (Stahl et al., 2014).

Over the past decades, SL has been mainly investigated in the auditory (e.g., Aslin et al., 1998; Saffran et al., 1996, 1999) and visual modalities (e.g., Fiser & Aslin, 2002; Kirkham et al., 2002; Slone & Johnson, 2015) via behavioural studies. Typically, the participant is exposed to a continuous stream of items that are regularly organised into units (often doublets or triplets), according to a higher transition probability (TP) within-unit and a lower TP between-unit. Different TPs should signal boundaries and units within an otherwise continuous and interrupted sequence of items, hence extraction of these statistics should facilitate the learning of a structure. After a familiarization of fixed duration (e.g., Tummeltshammer et al., 2017) or a habituation phase (e.g., Kirkham et al., 2002), the participant's SL abilities are tested comparing their attention towards novel vs. familiar units. A novelty preference has been evidenced in a number of seminal works on infant SL abilities (e.g., Bulf et al., 2011; Kirkham et al., 2002; Saffran et al., 1996). This preference is typically taken as a demonstration of discrimination and cognitive abilities in infancy (e.g., Fantz, 1964; Spelke, 1992), assuming that attention is deployed elsewhere if a familiar stimulus is already encoded and

represented in memory (Hunter & Ames, 1988). Nevertheless, this is not always the case and a familiarity preference may also arise at early stages of development (Wetherford & Cohen, 1973), with limited familiarization time (Hunter & Ames, 1988) or with increased stimulus complexity (Cohen, 2004). In SL studies, this reverse pattern of preference has been associated with the stimulus presentation modality (i.e., familiarity preference in a visual condition and novelty preference in an auditory condition; Emberson et al., 2019) and with the predictability of the stimulus and the environment (Bertels et al., 2021; Kidd et al., 2012). In this context, predicting whether to expect a novelty or familiarity effect is not straightforward and both directions of preference can be potentially taken as evidence of discrimination, raising interpretative challenges across several domains (Houston-Price & Nakai, 2004; Raz et al., 2023; Roder et al., 2000).

The field of infant neuroimaging may help to better characterize and support behavioural evidence, especially to disambiguate unclear findings. Neuroimaging data can provide invaluable insights into perceptual and learning mechanisms online as they unfold, as opposed to offline post-exposure behavioural measures such as novelty-familiarization scores, which can be confounded by memory effects (Batterink & Paller, 2017). Further, adult neural evidence of SL has been found to be more robust than behavioural familiarity measures (Turk-Browne et al., 2009). Thus far, event-related potentials (ERPs) have been adopted to show a N400 component associated with online segmentation of a visual sequence in adult participants (Abla & Okanoya, 2009). During development, atypical ERPs in response to visual regularities have been proposed as an early biomarker of Autism Spectrum Disorders (Jeste et al., 2015; Marin et al., 2020). In the auditory domain, online SL mechanisms have been investigated with ERPs even at earlier developmental stages, assessing infant-directed speech in sleeping newborns (Bosseler et al., 2016). An alternative and promising approach is to quantify the rhythmicity of the brain responses in the frequency domain via steady-state evoked potentials (SSEPs; also referred to as rhythmic stimulation, frequency tagging, fast periodic stimulation, or neural entrainment; see, for instance, Norcia et al., 2015; Tononi et al., 1998). This method implies presenting a periodic perceptual stimulation at a given frequency

and analysing the brain signal in the frequency domain, so that the response strength at the stimulated frequency and its higher harmonic frequencies can be tracked. This tool has proven to be particularly promising and powerful in developmental research (e.g., Kabdebon et al., 2022; Peykarjou, 2022). Compared with ERPs, the high rate of stimulus presentation of SSEPs provides responses particularly robust to artifacts, with high signal-to-noise ratio (SNR), and thus requires a short period of stimulation to differentiate them reliably and objectively from the noise level (Peykarjou, 2022). For its compliance to the limited attentional span in infancy, this approach has been recently adopted in developmental research to investigate several cognitive processes, such as attentional mechanisms (e.g., Christodoulou et al., 2018), face and object discrimination (e.g., de Heering & Rossion, 2015; Rekow et al., 2021) or categorisation (e.g., Bertels et al. 2020; Peykarjou et al., 2024). Interestingly, this approach can be applied to SL, whose paradigms are intrinsically characterised by a rhythmic and regular presentation of items. A recent study by Choi et al. (2020) paved this way in developmental research assessing SL in the auditory domain. The authors presented 6-month-old infants with a continuous speech and tracked their neural entrainment to embedded words. Participants not only showed neural entrainment at the syllable level, they also showed entrainment at the frequency of word presentation. The neural word learning index increased logarithmically over the time course of learning, presumably as the participants detected word units from isolated syllables (Choi et al., 2020).

Interestingly, while SSEPs evidence of SL processes has been increasing over the past years in the auditory modality for both infants (Buiatti et al., 2009; Choi et al., 2020; Cirelli et al., 2016; Edalati et al., 2023; Kabdebon et al., 2015) and adults (Batterink & Paller, 2017, 2019; Farthouat et al., 2016; Moser et al., 2021), the same approach has not been adapted to the visual domain to the same extent. To the best of our knowledge, only limited adult research applied frequency tagging during a visual SL task, either with intracranial recordings to investigate deep structures involved in rhythmic processing (Henin et al., 2021; Sherman et al., 2023) or in tasks assessing the formation of visual linguistic units to better understand reading processes (Saringer et al., 2024). At present, how the

developing brain responds during the acquisition time course of visual regularities remains largely unexplored. In this context, a frequency-tagging approach is promising also in the visual domain for several reasons. Overall, SSEPs were initially introduced to study low-level visual processing in non-verbal participants (e.g., Norcia & Tyler, 1985; Braddick et al., 1986) and, hence, they are well-established in the visual domain. Further, a frequency-tagging approach would allow to collect data over the initial exposure phase, namely to explore the dynamics of learning while it occurs. Lastly, SSEPs may also provide insights into the degree to which the learning mechanisms evidenced in the auditory modality by Choi and collaborators (2020) are domain-general and applicable across other sensory modalities. At present, post-exposure behavioural evidence of SL has been taken to demonstrate a general mechanism operating across modalities, even though modality effects have been observed (Emberson et al., 2019; Saffran, 2002) and there is an ongoing debate about the specificity of SL learning mechanisms (Frost et al., 2015).

In this work, we aimed to explore SL as it unfolds during stimulus exposure in a group of 4- to 6-month-old infants. To test this, visual stimuli were presented periodically as a continuous stream of rapidly presented shapes while recording the infant's brain activity with high-density electroencephalography (hdEEG). When participants looked at the stimulation, visual SSEPs were expected to appear at the frequency of stimulation (*F*, 6 Hz in this study) and higher harmonics (2*F*, 3*F*, etc. – 12 Hz, 18 Hz, etc.). On top of that, a doublet organization was present in two conditions, with either a within-doublet TP of 100% (doublet condition) or 33% (control condition). Hence, each doublet unit appeared at a frequency of *F*/2, 3 Hz. Despite no explicit segmentation cues indicating the doublet organization in the continuous visual stream, we hypothesized a neural peak at the doublet frequency and following harmonics (i.e., 9 Hz, 15 Hz, etc.), reflecting detection of these regularities already in infancy. Behavioural evidence of visual SL suggests that this ability should be in place in the age range investigated in the present work without signs of developmental changes (Marcovitch & Lewkowicz, 2009; Kirkham et al., 2002). We expected this behavioural evidence of SL to be reflected at the neural level, as already found with 6-month-old infants in the auditory domain

(Choi et al., 2020). Although the current frequency-tagging approach has not yet been adopted in visual SL tasks in infancy, 6-month-old infants seem to use neural learning strategies similar to adults to form expectations about the sensory inputs and track changes in the environment, sharing a fundamental architecture of top-down feedback neural strategies in visual areas (Emberson et al., 2015). Nevertheless, the developmental invariancy of SL during early infancy has been challenged (e.g., Slone & Johnson, 2015) and may differ across sensory domains (Emberson et al., 2019). For this reason, we targeted an age range that is well studied in the SL behavioural literature and, at the same time, we analysed exploratory the effect of age on our frequency-tagging measures of SL, applied here to the visual domain for the first time.

Overall, neural entrainment at the doublet frequency was predicted to increase over sequences of stimulus exposure, as participants progressively learn the item organization. Moreover, a response at the doublet frequency of stimulation was especially predicted in case of a deterministic transition (TP of 100%) among the elements forming the doublet, namely in the doublet condition. In the control condition, we explored whether infants tracked only the regular position of the first and second elements of the doublets, hence if tracking the position regularity is enough to elicit a significant activity at the doublet frequency. If present, this peak of neural activity in the control condition may potentially take more exposure to emerge and, in turn, may have a smaller amplitude than the one elicited in the doublet condition (which includes both a deterministic TP regularity and the position regularity). Further, if 4-6-month-olds could already detect non-exclusive pairings in the control condition, this may be evident at higher harmonics compared to a deterministic TP. In fact, more complex responses seem to require more harmonics (Retter et al., 2021). Lastly, the doublet and control conditions were compared with a random condition in which shapes followed a pseudo-randomly presentation without any regular organization. In this condition, we predicted comparable activity at the frequency of stimulation *F* as in the other two conditions, but no activity at the doublet frequency *F/2* as the TP across visual items did not change.

2. Methods

2.1 Participants

Thirty 4- to 6-month-old infants (M_{age} = 156.5 days, *SD* = 23.2 days; 11 females) from the surroundings of Brussels, Belgium, comprised the final sample of this study. An additional 2 infants were tested but excluded from analysis because of equipment error (*n* = 1) or behavioural issues that did not allow them to attend the task (*n* = 1). Participants were born full-term (> 37 weeks), had normal birthweight (2.5 – 4.5 kg) and no complications at birth (Apgar scores > 7). Infants were predominantly White (86.67%); the remaining participants were of Mixed racial background. At the time of testing, infants had no sensory impairment nor developmental concern. Before the beginning of the study, the caregiver provided informed written consent. The study was conducted according to the principles expressed in the Declaration of Helsinki. The protocol of the study was approved by the Ethics Committees of the Faculty of Psychological and Educational Sciences of the Université libre de Bruxelles (ULB) and of the Hôpital Universitaire de Bruxelles (HUB), ref. P2019/480, B406201941700.

2.2 Stimuli and apparatus

Stimuli consisted of eight colourful shapes (yellow circle, orange star, turquoise square, blue cross, green triangle, purple diamond, red hexagon, pink arrow) presented on a uniform grey background (Figure 1). Each shape was 800 x 800 pixels and subtended a visual angle of 13.5° at a distance of 85 cm. Stimuli were presented at the centre of the presentation screen with sinusoidal contrast modulation, one at a time in a continuous stream at a frequency of 6 Hz, in blocks of 20 s. Fading periods of 2s at the beginning and at the end of each block helped reducing saccades and blinks caused by abrupt stimulation onset and offset, leading to blocks of 24 s with a total of 144 shapes. An attention getting video with catching moving objects paired with sounds was presented in between blocks.

Stimuli were displayed on a 30-inch LCD monitor (HP z30i, 1920 x 1200 pixel resolution, 60 Hz refresh rate) located in the testing room. Stimulus presentation was run in MATLAB (version

R2017a, the MathWorks, Natick, MA, USA) with the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al. 2007) on a computer controlled by the experimenter from an adjacent room. Brain activity was recorded with an EGI dense-array EEG system (Electrical Geodesic Inc., Eugene, Oregon, USA) at a sampling rate of 1000 Hz. The analog EEG signal was referenced to the vertex (Cz), digitised by an EGI Net Amps 400 amplifier and online band-pass filtered in the range of 0.1-120 Hz. EEG data were acquired with Net Station software, version 5.4.2 (r29917) running on a MacBook Pro Retina laptop computer. A set of 128-channel HydroCel Geodesic Sensor Nets (covering head perimeters ranging from 40 to 44 cm) was adopted to test infant participants. The LCD monitor for stimulus presentation was equipped with an external video camera to monitor the participant's behaviour during the experimental procedure. The video feed was recorder synchronously with EEG data on the acquisition computer. The sound of the attention getting videos was played through two stereo external speakers located on either side of the LCD monitor in the testing room.

2.3 Procedure

After being capped with the EEG net, infant participants sat on their caregiver's lap at 85 cm distance from the LCD monitor. Caregivers were instructed to avoid speaking and interfering with the experimental procedure while maintaining their infant in a stable upright position in front of the monitor. Lights were switched off and the testing room was only lit by the computer monitor to limit potential distractions during the procedure. EEG sensor impedance was checked before the start of the experiment to ensure that it remained below 50 kΩ for all channels.

Participants were randomly assigned to one of three stimulus presentation conditions (doublet, control, or random; see Figure 2) in a between-subject design (10 per condition). The experimental procedure started with the presentation of the attention getting video. The experimenter controlled the presentation of the video and stopped it as soon as the participant was attentive, so that the stimulus presentation could start. Triggers were sent to the EEG system at each individual shape presentation, excluding the fading periods (i.e., for a total of 120 shapes per block). In the doublet condition, the continuous stream of eight shapes was organised in four doublets (for

instance, square followed by circle, diamond by cross, triangle by hexagon, and arrow by star). This yielded a TP within doublets of 1 and a TP between doublets of 0.33, with doublets pseudo-randomly ordered throughout the block of stimulation (i.e., no consecutive presentation of the same doublet was allowed throughout the visual presentation). In the control condition, four shapes could be presented in position *n* and the remaining four shapes could be presented in position *n+1*. This yielded a TP between items in positions *n* and *n*+1 of 0.25 and a TP between items in positions *n+1* and *n*+2 of 0.33. This condition resulted in a more challenging item pairing, in which a total of 16 doublets could be generated (hence, an item was not exclusively associated with another one, as in the doublet condition). Doublets were pseudo-randomly organized (no consecutive repetition of the same doublet) also in this condition. The control condition was meant to control whether infants monitored only the regular appearance of the first element composing the doublet (four shapes always appearing in the first position of the doublet, as in the doublet condition) instead of a deterministic doublet organization itself. If so, no difference was expected between doublet and control conditions. Whereas if infants were tracking the doublet organization itself, the neural entrainment in the doublet condition was expected to be higher than in the control condition. In both conditions, shape pairing was randomised by the computer for each participant at the beginning of the experiment, ensuring that any potential effect was not due to a specific doublet organization, but the same item pairing was maintained over blocks for the same participant. Lastly, in the random condition individual shapes appeared one following the other in pseudo-random order (no consecutive presentation of the same shape), without any doublet organisation. Hence, the TP between the eight shapes in the random condition was 0.14. For all conditions, blocks of visual stimulation were presented until the participant was no longer attentive. The experimental procedure could last 5 to 10 minutes, depending on the participant's looking behaviour.

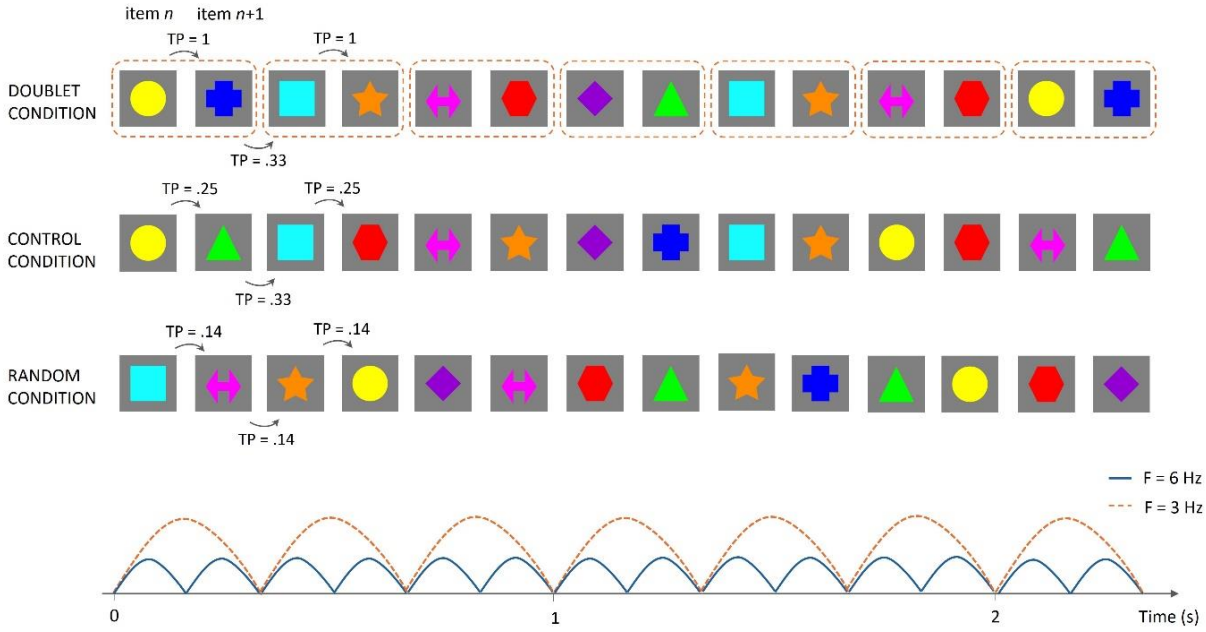


Figure 1. Schematic illustration of the experimental procedure across the three conditions (doublet, control and random). Stimuli were eight shapes presented by sinusoidal contrast modulation at a rate of six images per second (6 Hz; blue line in the bottom graph). Inter-shape transitional probability (TP) varied across conditions. In the doublet condition, shapes were organised in four doublets. In the control condition, four shapes could be presented in position *n* and the other four shapes could be presented in position *n*+1, yielding a more complex doublet organisation and a total of sixteen possible doublets. Doublets were presented pseudo-randomly (no consecutive repetition of the same doublet) in both doublet and control conditions. The doublet frequency was three doublets per second (3 Hz; orange dotted line in the bottom graph). In the random condition, all shapes were pseudo-randomly presented (no repetition of the same shape).

2.4 EEG data pre-processing and analysis

Data were pre-processed in MATLAB using the PREP pipeline (Bigdely-Shamlo et al., 2015). This pipeline allowed line noise removal, signal referencing relative to an estimate of the true average reference, and noisy channels detection and interpolation (Bigdely-Shamlo et al., 2015). The pipeline automatically reiterated until all noisy channels were interpolated, up to 3 times. On average, 18.67 (*SD* = 7.07) channels out of 128 were interpolated per infant participant over a mean of 1.93 iterations. Events containing stimulation triggers were extracted to obtain segmented blocks of 20 s, without including the fade in and fade out periods. A Fast Fourier Transformation (FFT) was applied to the 20 s blocks, leading to a frequency resolution of 0.05 Hz. The strength of the response can be quantified in terms of signal-to-noise ratio (SNR) between the power at the stimulated

frequency and the estimated background noise. Namely, the SNR was computed for each block as the ratio between the Fourier amplitude at the tagged frequency and the average amplitude of the 10 surrounding bins (5 bins on each side, excluding the two bins immediately adjacent to the tagged frequency, as recommended in the review by Peykarjou, 2022). At this stage, valid blocks were selected. As opposed to the auditory counterpart of this task (Choi et al., 2020), the visual domain holds extra challenges as the infant's visual attention naturally fluctuates during a task. Hence, a careful selection of the attended and valid sequences was required to ensure that brain responses were not undermined by inattention. Valid blocks were selected based on a SNR above 2 at the 6-Hz base frequency in at least one of ten pre-defined occipital channels covering the medial occipital cortex (i.e., electrodes 70, 83, 75 of the 128-channel nets – corresponding to O1, O2, and Oz in the 10-10 montage – as well as neighbouring electrodes 69, 73, 74, 81, 82, 88 and 89; see Supplementary Figure 1a). Further, the participant's behaviour was coded offline from video recordings to make sure that the infant looked at the screen for more than 7 s (i.e., at least one third of the block, hence about 42 shapes). The coder had no information on what the participant was looking at. This combination of electrophysiological and behavioural criterium ensured that the visual stimulation was sufficiently attended. Participants' data had to show at least two valid blocks in order to be considered for further analyses.

Tagged responses from valid sequences were analysed not only at the base frequency (6 Hz item presentation) and at the doublet frequency (3 Hz doublet presentation) but also at their harmonics (i.e., multiple integers of the frequency of interest). Only the harmonics not overlapping with base frequency harmonics were considered as purely doublet frequency harmonics (i.e., 9 Hz, 15 Hz, 21 Hz, etc.). Taking harmonic responses into account improves the quantification of periodic but non-sinusoidal responses and ensures an unbiased comparison of conditions (Peykarjou, 2022; Retter et al., 2021). To determine how many harmonics consider in further analysis, mean SNR data were visually inspected to detect a frequency range of interest over the SNR spectra. Within this range, harmonics selection was performed using a one-sample Wilcoxon test to compare the SNR

value at a given frequency to 1. To determine the significant harmonics of the base frequency (i.e., 6, 12, 18, 24 Hz, etc.), we considered averaged spectra from all three conditions since the frequency of stimulation did not vary across conditions and no differences were expected. Whereas to determine the significant harmonics at the doublet frequency (i.e., 3, 9, 15, 21 Hz, etc.), we combined the doublet and control conditions only, since in the random condition there was no doublet organization and no response at 3 Hz and harmonics was expected. The dominant frequency was defined for both the base and doublet harmonic frequencies as the frequency in which the highest amplitude response was elicited.

2.5 Statistics

At the individual level, maximum statistics were performed on whole-scalp SNR values to identify participants showing a significant neural response at the base or doublet frequency and the selected harmonics across conditions. A similar maximum statistics was performed on group-averaged whole-scalp SNR values, now to localise electrodes with significant SNR at the base and doublet frequencies and, in turn, were used as region of interest (ROI). For maximum statistics, a non-parametric permutation test was employed to generate surrogate data. Null distributions were based on 10000 permutations, testing the null hypothesis that the SNR was equal to 1. Surrogates were obtained by randomly replacing the whole-scalp SNR topography with a unit topography across different conditions (for individual analysis) or across subjects (for group analysis) using a random binary mask. Once ROIs for each base- and doublet-level frequency and selected harmonics were determined at the group level, SNR spectra were averaged across the determined electrodes of interest (See Results – Regions of interest, below). These SNR values were then averaged across selected harmonics for the following analyses, leading to one averaged value at the base frequency and another one at the doublet frequency. At this point, linear mixed-effects models (LMMs) were adopted to model the data, using the lme4 package by Bates and colleagues (2015) implemented in R (R Core Team, 2023). Notably, LMMs enabled us to incorporate an unbalanced dataset across subjects following valid trials selection and to model trial-level variability over time, which was

relevant to study any learning effect over trials. In this work, the SNR was our dependent variable. Separate models were run for the average SNR at the base frequency, for the average SNR at the doublet frequency and for a doublet learning index (the ratio between average SNR at the doublet frequency and average SNR at the base frequency per trial; a similar approach was used in Choi et al., 2020). This index was meant to measure the relative entrainment sensitivity to the doublet organization over trials of exposure and, in turn, should be indicative of statistical learning. A learning index increase over time would reflect higher entrainment to the doublet frequency relative to the base stimulation frequency, which is expected if the participant progressively becomes aware of the doublet organization. The average SNR at both the base and doublet frequency had a right-skewed distribution which was corrected with logarithmic transformation (log base 10) to assure model residuals normality for the LMMs analyses. We evaluated whether trial-level fixed effects (condition and trial order) and an individual-level covariate (age) were related to our variable of interest, i.e. the SNR. Trial order was computed following exclusion of invalid sequences. Participant ID was included in all the models as a random factor. To analyse the effects of condition, we used simple coding instead of the default dummy coding. This coding scheme allowed us to assess the deviation of each condition from the overall average effect. The intercept hence represents the grand mean of conditions rather than the mean of a reference condition. Confidence intervals (95% CI) were calculated using the Wald method. Models were compared and selected by likelihood ratio test and Akaike Information Criterion (AIC).

3. Results

3.1 Valid trials and harmonics selection

On average, each infant participant viewed 10.83 blocks (*SD* = 3.36, range = 4 - 18 blocks). Out of a total of 319 recorded blocks, 111 blocks belonged to the doublet condition, 106 blocks to the control condition, and 108 blocks to the random condition. Of these recorded blocks, 241 (75.55%) were considered valid and further analysed (see Methods section, EEG data pre-processing and analysis), leading to an average of 8.03 valid blocks of 120 images each per infant participant (*SD*

= 3.66, range = 2 – 17 blocks). In more detail, 85 blocks in the doublet condition, 83 blocks in the control condition, and 73 blocks in the random condition were valid and considered for the following analyses. Of note, average looking time during valid trials did not significantly differ across conditions (all p >.05).

Following visual inspection of the SNR spectra, data were extracted in the 2 – 36 Hz range and significant harmonics were evaluated with a one-sample Wilcoxon test. At the base frequency, the SNR at 6, 12, 18, and 24 Hz significantly differed from 1 (z > 18979, all p < .001). At the doublet frequency, the SNR at 3 Hz (z = 8722, p = 0.01) and at 9 Hz (z = 11843, p < .001) significantly differed from 1 (See Supplementary Figure 2 on harmonics selection). The following ROI analyses were limited only to these six selected frequencies of interest. In the LMMs analyses, when referring to the SNR at the base or doublet frequency we refer to the combined SNR obtained by averaging the above-described significant harmonics at the base (6, 12, 18, and 24 Hz) and doublet frequencies (3 and 9 Hz), respectively.

3.2 ROI analysis

At the individual level, clusters of significant electrodes emerged at the stimulation frequency (6 Hz) in the vast majority of participants (27 subjects out of 30, or 9 out of 10 per experimental condition), consistently localised in the occipital area. Most subjects (24 subjects out of 30) also showed significant occipital activity in at least one of the three base frequency harmonics of interest (12, 18, and 24 Hz). On the other hand, significant occipital activity at the dominant doublet frequency (9 Hz) was evidenced for 12 subjects out of 30 (7 subjects out of 10 in the doublet condition, 5 subjects out of 10 in the control condition, no one in the random condition). Less subjects (8 out of 30) showed significant occipital activity at 3 Hz (4 subjects out of 10 in the doublet condition, 3 subjects out of 10 in the control condition, and 1 subject out of 10 in the random condition).

At the group level, maximum statistics confirmed a medial occipital cluster of channels with significant SNR across both the base and doublet frequencies. At the base frequency and harmonics,

the dominant frequency was 6 Hz, with significant SNR values in a cluster of 11 medial occipital channels (channels 69, 70, 73, 74, 75, 76, 81, 82, 83, 88, and 89; SNR threshold $p < .05 = 3.295$). Of note, this area overlapped with the cluster of 10 occipital channels defined a-priori for valid trials selection (See EEG Data pre-processing and Analysis section). Base frequency harmonics showed significant activity in a smaller subcluster of medial occipital channels. In more detail, the same cluster of 5 channels (channels 70, 74, 75, 81, and 82) was significant at 12 Hz (SNR threshold $p < 0.05 = 2.353$) and at 18 Hz (SNR threshold $p < 0.05 = 1.693$). A smaller subcluster of 4 channels (74, 75, 82, and 88) was significant at 24 Hz (SNR threshold $p < 0.05 = 1.361$). On the other hand, the dominant frequency at the doublet frequency and harmonics was the first harmonic, hence 9 Hz. A cluster of 3 sensors in the medial occipital area reached significance at 9 Hz (channels 74, 75, and 82; SNR threshold $p < 0.05 = 2.095$). The doublet frequency of 3 Hz showed significant activity in a cluster of 2 channels (74 and 82; SNR threshold $p < 0.05 = 1.563$). Outside the medial occipital region, no other channel reached significance at the group level, both at the base and doublet frequencies (Figure 2). The following LMM analyses were run averaging SNR values obtained in a ROI of 3 occipital channels (namely 74, 75, and 82; See Supplementary Figure 1b) that popped out across the base and doublet frequencies of interest. The same analyses were conducted considering the bigger cluster of 10 occipital channels a-priori determined for valid trial selection (Supplementary Figure 1a) and results did not substantially change.

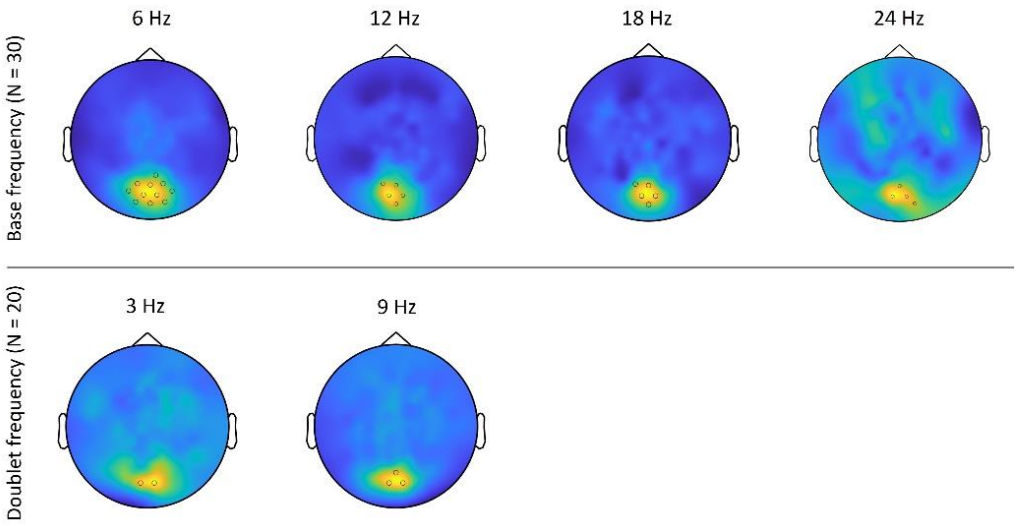


Figure 2. Topographical maps of brain activity at the base frequency and harmonics (top, all subjects) and at the doublet frequency and harmonics (bottom, subjects assigned to a condition with a doublet organization). For each frequency of interest, the electrodes whose SNR values overcame the significance threshold are highlighted in the figure. Topographical maps are shown at the maximal colour scale for each frequency of interest, with yellow representing the highest SNR values and dark blue representing the lowest SNR values.

3.3 Base stimulation responses (6 Hz and harmonics)

The random intercept model (AIC = 32.404, LogLik = -13.202) was significantly improved by the inclusion of trial order as fixed effect (AIC = 16.956, LogLik = -4.478, $X^2(1) = 17.448$, $p < .001$). The inclusion of condition and age as fixed effects did not improve the model significantly, nor did the inclusion of any interaction effect. Overall, the model with trial order as fixed effects was retained as it had the lowest AIC and provided the most parsimonious explanation of the data. The model outcome is reported in Supplementary Table 1. In this model, the interclass correlation coefficient (ICC), corresponding to the proportion of outcome variance due to individual differences, was 17.27%. The model revealed that the average SNR at the base frequency varied over valid trials, with a general SNR decrease as trials progressed in time (0.021 unit decrease of the estimate over time, 95% CI = -0.031, -0.012 , $t(239) = -4.384$, $p < .001$). Notably, the average SNR at the base frequency was not significantly affected by condition (Figure 3).

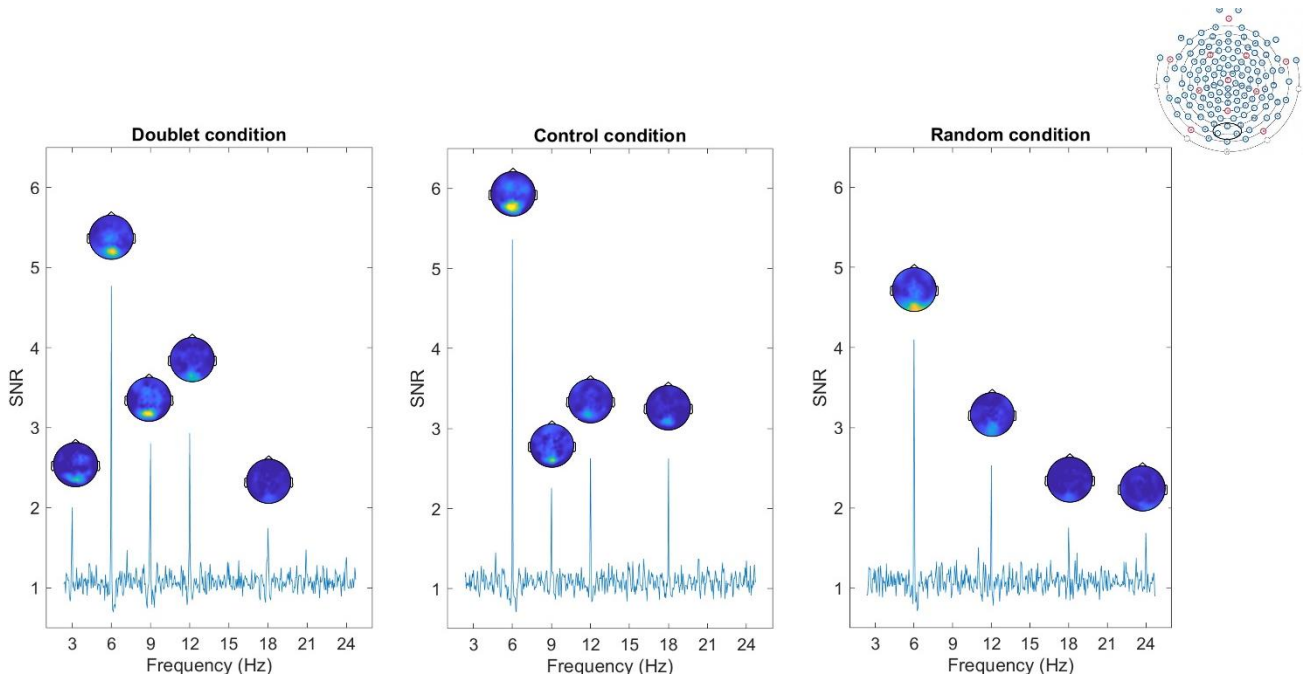


Figure 3. SNR spectra in a region of interest (ROI) of three medial occipital electrodes (electrodes 74, 75, 82, ROI marked in black in the top-right EEG net configuration) during the exposure to visual stimuli across the three different experimental conditions (doublet, control and random). At the base frequency and harmonics (6, 12, 18, and 24 Hz), average SNR values did not differ across conditions. On the other side, average SNR values at the doublet frequency and harmonics (3 and 9 Hz) differed across conditions. Topographical maps show the distribution of SNR values across the scalp and are displayed at those frequencies whose SNR significantly differed from 1. For the topographical maps, different scales are used for the base and doublet frequencies (base frequencies range: 1-5.5; doublet frequencies range: 1-3).

3.4 Doublet individuation responses (3 Hz and harmonics)

The random intercept model (AIC = 182.570, LogLik = -88.284) was significantly improved by the inclusion of condition as fixed effect (AIC = 171.970, LogLik = -80.986, $X^2(2) = 14.596$, $p < .001$). The model was further improved by the inclusion of both condition and trial order as fixed effects (AIC = 168.850, LogLik = -78.423, $X^2(1) = 5.124$, $p = .024$). On the other hand, adding age as fixed effect did not improve the model significantly. The model with condition and trial order as fixed effects was also significantly improved by an interaction term between these fixed effects (AIC = 161.090, LogLik = -72.547, $X^2(1) = 11.621$, $p = .003$). This last model was retained as it had the lowest

AIC, indicating a better fit. The model outcome is reported in Supplementary Table 2. In this model, the proportion of outcome variance due to individual differences, expressed by the ICC, was 8.86%. Results showed that, contrary to the base-level SNR, the doublet-level SNR was influenced by the experimental condition (Figure 3). The model revealed that the doublet-level SNR in the control condition was associated with lower SNR values than the grand mean (-0.315 log unit variation, $t(95) = -3.059$, 95% CI = -0.517, -0.113, $p = 0.003$). Similarly, the doublet-level SNR in the random condition was lower than the grand mean SNR across the three conditions (-0.558 log unit variation, $t(113) = -5.123$, 95% CI = -0.771, -0.344, $p < .001$). Further, condition interacted with trial order, with a difference in the progression of the doublet-level SNR over time (the slope of trial order) in the control condition (0.041 log unit variation of the estimate as trials progressed, 95% CI = 0.012, 0.069 , $t(230) = 2.775$, $p = .006$) and in the random condition (0.053 log unit variation of the estimate as trials progressed, 95% CI = 0.019, 0.087 , $t(230) = 3.079$, $p = .002$) when compared to the average slope (Figure 4a).

Post hoc pairwise comparisons adjusted with Tukey's method further explored the main effect of condition and the interaction between condition and trial order. Regarding the effect of condition, post hoc tests revealed that the doublet-level SNR in the doublet condition (M = 0.231, SE = 0.049) was significantly higher than the doublet-level SNR in the random condition (M = -0.043, SE = 0.052, 95% CI = 0.098, 0.451, $p = .002$). The difference between the control condition (M = 0.132, SE = 0.049) and the random condition approached significance (95% CI = -0.001, 0.352, $p = .051$). The doublet frequency SNR did not differ significantly between the doublet and control conditions ($p = .343$). To make the results more interpretable, the back-transformed mean SNR was 1.701 (95% CI = 1.349, 2.145) for the doublet condition, 1.356 (95% CI = 1.075, 1.711) for the control condition, and 0.905 (95% CI = 0.709, 1.154) for the random condition. The interaction between condition and trial order was explored with post hoc pairwise comparisons on the slopes of trial order across conditions. The slope for the doublet condition (estimate = - 0.043, SE = 0.011) was significantly more negative than for the control (estimate = - 0.002, SE = 0.010, $p = .018$) and for the random (estimate =

0.010, SE = 0.014, *p* = .007) conditions. The difference in slopes between control and random conditions was not significant (*p* = .744). As shown in Figure 4a, the doublet-level SNR in the doublet condition had higher values compared to the other conditions since the first valid trials, to then progressively reach similar SNR values across conditions over time. Accordingly, during the first valid trial of exposure the SNR in the doublet condition was already significantly higher (M = 0.417, SE = 0.067) compared to the control (M = 0.142, SE = 0.064, *p* = .011) and random (M = - 0.088, SE = 0.070, *p* < .001) conditions. Also the control condition marginally differed from the random condition (*p* = .045). No significant difference emerged from pairwise comparisons across condition at the last valid trial with data for all three conditions (trial 12; all *p* values > .05)

3.5 Doublet learning index

The random intercept model (AIC = 483.510, LogLik = -238.750) was significantly improved by the inclusion of condition as fixed effect (AIC = 479.090, LogLik = -234.540, $X^2(2) = 8.423$, *p* = .015). The inclusion of other fixed effects of interest (trial order and age) did not significantly improve the model with only condition as fixed effect, which was retained as it provided the most parsimonious explanation of the data. The model outcome is reported in Supplementary Table 3. In this model, the proportion of outcome variance due to individual differences, expressed by the ICC, was 11.98%. The model showed that the learning index in the random condition was expected to be lower than the gran mean of learning index (-0.413 log unit variation, *t*(26) = -2.840, 95% CI = -0.698, -0.128, *p* = 0.009). The learning index in the control condition also approached a significant difference from the mean learning index across conditions (-0.295 log unit variation, *t*(25) = -2.054, 95% CI = -0.576, -0.014, *p* = 0.051). Post hoc pairwise comparisons adjusted with Tukey's method confirmed that the learning index in the doublet condition (M = 1.034, SE = 0.102) was significantly higher than the learning index in the random condition (M = 0.621, SE = 0.105, 95% CI = 0.050, 0.776, *p* = .024). The learning index in the control condition (M = 0.739, SE = 0.102) lied in between, without any

significant differences from the doublet and random conditions (all $p > .05$). Interestingly, trial order had no significant effect on the learning index, which was stable over time (Figure 4b).

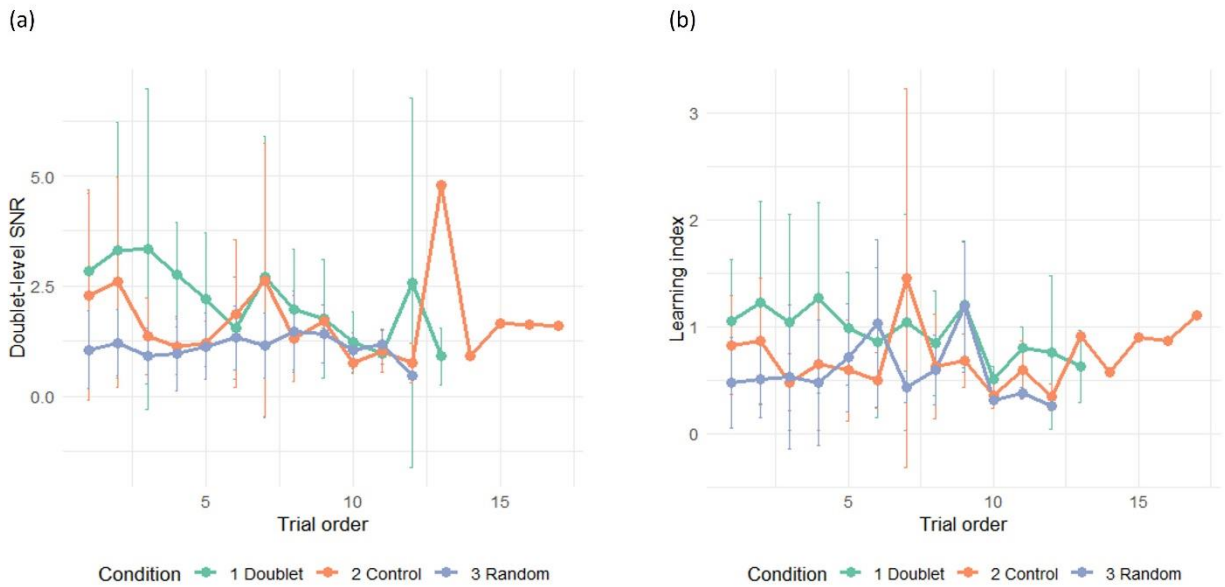


Figure 4. Mean doublet-level SNR (a) and doublet-learning index (b) as a function of trial order. The learning index is calculated as the ratio between the doublet-level SNR and the base-level SNR. Bars represent the standard deviation of each data point.

Since, contrary to our hypothesis, the doublet learning index did not seem to gradually increase over trials of exposure and the doublet-level SNR showed high values from the first valid trials, an additional analysis was conducted to examine whether the differences in learning index across conditions were already evident during the first valid trial. Of note, the first valid trial corresponded to the very first trial of exposure for the vast majority of participants (28 out of 30 infants). A linear model was fitted to this subset of data, with learning index as dependent variable and condition as predictor. Indeed, results revealed a main effect of condition (F (2, 27) = 3.498, p = .045). Specifically, at trial one the random condition had a lower learning index than the grand average index (estimate = -0.577, p = .014), while the control condition did not differ significantly (estimate = -0.228, p = .310). Post hoc pairwise comparisons using the Tukey method revealed that the index in the doublet condition (M = 1.052, SE = 0.155) was significantly higher than the random condition (M = 0.475, SE = 0.155, t(27) = 2.626, p = 0.036). Again, the learning index of the control condition (M = 0.824, SE = 0.155) did not

differ significantly from the doublet and random conditions (all *p* values >.05), lying somewhere in the middle.

4. Discussion

This study is the first to show that preverbal infants rapidly attune to visual statistical regularities using a frequency-tagging approach. While there has been growing interest in investigating auditory SL mechanisms during stimulus exposure with this approach (e.g., Buiatti et al., 2009; Choi et al., 2020; Kabdebon et al., 2015), this has not been transferred to the visual modality. Still, regularities in our environments are not limited to the auditory modality or language acquisition phenomena, and there is vast evidence of SL in the visual modality using post-exposure behavioural tasks (e.g., Fiser & Aslin, 2002; Kirkham et al., 2002; Slone & Johnson, 2015). To bridge this gap and investigate visual SL during stimulus exposure, we presented infants with a continuous stream of shapes at 6 Hz that either follow a deterministic doublet organization (doublet condition), a low-TP doublet organization (control condition), or had no clear organization (random condition) while recording their brain activity. Results showed similar neural entrainment at the base stimulation frequency (6 Hz) and its harmonics, which revealed comparable visual engagement across conditions. This is also in line with similar looking time values across conditions. Further, the base-level entrainment dropped over trials of exposure, possibly reflecting neural adaptation phenomena (Benda, 2021) or infants' limited attentional resources. This habituation effect of visual SSEPs in response to repeated information has been already found in infancy by Christodoulou and colleagues (2018) and can be reverted by presenting novel information, hence reflecting changes in overt visual attention. On the other side, entrainment at the doublet frequency (3 Hz) and its harmonics differed across experimental conditions and over trials of exposure. The SNR associated with a deterministic doublet organization of the visual stimuli was significantly higher than the SNR associated with a random stimulus organization, especially over the first valid trials of exposure. This confirmed that 4- to 6-month-old infants were able to efficiently detect a regular doublet organization in a stream of continuously presented visual stimuli. The peak of activity at the doublet frequency and harmonics

was not only evident at the group level, but was also a strong marker at the individual level, especially at the first harmonic (9 Hz). Interestingly, our data revealed that infants showed enhanced sensitivity to a deterministic doublet organization of visual items from the first valid trial of exposure. In turn, these detection and learning mechanisms seem to require limited exposure. This is in line with past evidence showing that visual statistical learning can operate very quickly and with little exposure in adult participants (Turk-Brown et al., 2009). The current data reveal that similar mechanisms may be available from early developmental stages.

While the average SNR in the doublet and random conditions differed significantly, the control condition lied in between. Interestingly, the doublet-level SNR in the control condition was predominantly driven by activity at 9 Hz, as it was the case for the doublet condition, but this time with little or no contribution at 3 Hz. Even though there is no consensus on what drives differences in terms of harmonics recruitment in different tasks, current frequency-tagging guidelines recommend considering the aggregated signal coming from the significant harmonics rather than focusing exclusively on the fundamental harmonic response (Peykarjou, 2022; Retter et al., 2021). In the current work, the dominant frequency at the doublet level was in fact the first harmonic (9 Hz), followed by the fundamental doublet frequency of 3 Hz. Accordingly, considering only the fundamental responses of interest of 3 Hz would have hidden predominant neural activity at further harmonics that was clearly associated with the task presentation. In fact, the peak of activity at 9 Hz in the control condition suggests that infants detected a regular organization in the visual stream, even though not as strongly as in the doublet condition. This peak of activity at the doublet frequency in the control condition showed that infants were extracting regularities even in a more challenging visual presentation, namely a non-deterministic doublet organization leading to 16 possible doublets (with inter- and intra-doublet TP of 33% and 25%, respectively). In terms of number of harmonics, neural responses recruited a limited number of harmonics in infancy, in line with past evidence comparing infant and adult participants (e.g., two harmonics in an infant sample

vs. seven harmonics in an adult sample undergoing the same visual discrimination paradigm in Baccolo et al., 2023).

The brain activity in response to the doublet frequency and harmonics was solely localised in medial occipital visual areas, as it was the case for the base frequency and harmonics. Accordingly, the mechanisms under investigation were largely supported by primary visual areas rather than more domain-general memory systems classically associated with implicit visual learning, such as the medial temporal lobe (Shapiro et al., 2012). This may be due to the fact that, contrary to a post-exposure task that largely involves a memory component, our task measured brain activity during the online exposure to the regularities itself, hence it revealed an initial perceptual component of statistical learning in which memory may play a more limited role. This is reflected by the conceptualization of two distinct components of SL by Batterink and Paller (2017), namely an initial sensitivity to a sensory regularity and a following commitment of this regularity to the memory system. Together with the occipital localization of the brain activity in response to the doublet frequency, the fact that a response at the doublet frequency emerged during the first trials is also coherent with a perceptual component explanation.

Contrary to our expectations, the current results did not reveal a progressive increase in the doublet-level entrainment over the base-level entrainment (the so-called learning index) as trials progressed in time. This index was meant to monitor the evolution of the doublet-level SNR in relation with the base-level SNR over time. An increase in the learning index would have suggested a progressive encoding of the regular item organization relative to the fundamental frequency of presentation. In the present work, the learning index was higher in the doublet condition than the random one but was stable over time, hence doublet-level and base-level entrainment changed proportionally over trials of exposure, not revealing a doublet learning curve. A number of past works have evidenced an increase in neural entrainment as the exposure to structured auditory input progressed (e.g., Batterink & Paller, 2017, 2019; Choi et al., 2020; Flo et al., 2022; Moreau et

al., 2022; Moser et al., 2021; Ordin et al., 2020). This evidence is currently based on auditory investigations and it has been framed in the context of continuous speech segmentation.

Given the fact that we did not find the same learning progression in the current study, an open question is whether this learning curve is a modality-dependent feature that specifically applies to structured auditory input. Accordingly, an increased neural entrainment over time may emerge only under some specific experimental conditions (e.g., exposure time, sensory modality, stimulus type). Potentially, limited attentional resources in infancy may play a role as trials progressed and may differentially affect auditory and visual inputs and, in turn, neural entrainment measures. Frequent attention fluctuations to the screen during infancy may limit the exposure to visual input to a higher extent than auditory input, which plays and can be encoded continuously. Overall, auditory exposure happens regardless of the participant's behavioral state, while visual exposure is not happening if the participant turns away from the screen, hence is less stable over time. Considering SL behavioural tasks, while early investigations went towards a domain-general SL account, recent evidence has highlighted modality (and sometimes even stimulus) specificity (Frost et al. 2015). Further, a direct behavioural comparison of auditory and visual SL in an infant sample has recently revealed relevant differences between modalities, concluding that there is weaker learning in the visual than in the auditory modality (Emberson et al., 2019). Going back to frequency-tagging tasks, different results in the present study and its auditory counterpart (Choi et al., 2020) may be partially explained by methodological differences between studies. Stimulus presentation and, in turn, measures of neural entrainment varied across modalities. Notably, the exposure to auditory stimulation played continuously (minimum 120 s of continuous exposure in Choi et al., 2020) with a dynamic visual attention getter on screen. Data could in turn be organized in continuous epochs and analysed with a sliding time window. Such continuous input exposure is not effective in the visual domain. The sequence duration in visual frequency-tagging paradigms with infants ranges from 15 to 30 s, in an effort to balance infants' limited attentional resources and a sufficient signal strength for SNR quantification (Peykarjou, 2022). Sequences are then interspersed with attention getting videos

and only sequences in which the participant was sufficiently attentive are considered, hence neural entrainment measures in visual paradigms are obtained from sequences that are more discontinuous in time than auditory paradigms. Potentially, a learning curve in the visual domain may be disguised by doublet entrainment levels that progressively increased within each trial and not across trials of exposure, again reflecting a detection mechanism rather than a progressive acquisition of knowledge. While visual frequency-tagging paradigms cannot successfully include several minutes of continuous exposure as auditory ones, it would be interesting to apply a more discontinuous stimulus exposure to auditory inputs and test whether a continuous input exposure is necessary to elicit a learning curve. Further investigations comparing sensory modalities and experimental designs are in turn needed to better understand to what extent neural entrainment can be compared across modalities.

Moreover, not having evidence in favour of a learning curve also raises the question of what neural entrainment is reflecting. The current data suggest that neural entrainment may reveal a bottom-up detection mechanism, reflecting a downstream consequence of the progressive exposure to regularities. In fact, we revealed that the conditions including a doublet organization showed higher neural entrainment rates already during the first valid trials of exposure and this entrainment difference across conditions faded over time, instead of showing a learning curve. Additionally, this detection account was also supported by a clear brain localization in medial occipital areas. On the other side, recent evidence suggests that neural entrainment in response to a speech stream of repeating words may functionally contribute to statistical learning rather than reflecting downstream perceptual consequences (Batterink et al., 2024). Further studies are needed to understand whether, under some experimental circumstances, neural entrainment plays a top-down functional role in supporting the progressive learning evolution also in the visual domain. On top of modality and stimulus differences which may play a role, it may be worth investigating how entrainment and its time course over trials of exposure evolves during developmental stages. SL has been shown to increase as a function of age during childhood, with different trajectories across visual and auditory

modalities (Ren & Wang, 2023). In infancy, evidence suggests that learning abstract spatial relations is initially made on a perceptual basis and become more abstract around 6 months of age (Casasola & Ahn, 2018; Casasola, 2018). Similarly, learning regularities may be a more bottom-up detection mechanism at the age investigated here (4 to 6 months of age), but it may be of interest understanding the developmental trajectory of visual SL and the associated tagged responses during the exposure phase at different stages of development.

In sum, the current data suggest that the infant brain becomes quickly sensitive to the TP in a stream of visual images, especially to a deterministic doublet organization. We believe that frequency-tagging is a valuable tool to investigate detection of regularities while it occurs and it provides further time-course perceptual insights that post-exposure behavioural tasks cannot measure.

References

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience research*, *64*(2), 185-190.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science*, *9*(4), 321-324.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31-45.

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, *115*, 56-71.

Batterink, L. J., Mulgrew, J., & Gibbings, A. (2024). Rhythmically modulating neural entrainment during exposure to regularities influences statistical learning. *Journal of Cognitive Neuroscience*, *36*(1), 107-127.

Benda, J. (2021). Neural adaptation. *Current Biology*, *31*(3), R110-R116.

Benjamin, L., Fló, A., Palu, M., Naik, S., Melloni, L., & Dehaene-Lambertz, G. (2023). Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Developmental science*, *26*(2), e13300.

Bertels, J., Bourguignon, M., De Heering, A., Chetail, F., De Tiège, X., Cleeremans, A., & Destrebecqz, A. (2020). Snakes elicit specific neural responses in the human infant brain. *Scientific reports*, *10*(1), 7443.

Bertels, J., San Anton, E., Boursain, E., Bulf, H., & Destrebecqz, A. (2022). Visual statistical learning in infancy: Discrimination of fine-grained regularities depends on early test trials. *Infancy*, *27*(3), 462-478.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in neuroinformatics*, *9*, 16.

Bosseler, A. N., Teinonen, T., Tervaniemi, M., & Huotilainen, M. (2016). Infant directed speech enhances statistical learning in newborn infants: An ERP study. *PloS one*, *11*(9), e0162177.

Braddick, O. J., Wattam-Bell, J., & Atkinson, J. (1986). Orientation-specific cortical responses develop in early infancy. *Nature*, *320*(6063), 617-619.

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spatial Vision, 10*, 433-436.

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*, *44*(2), 509-519.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Christodoulou, J., Leland, D. S., & Moore, D. S. (2018). Overt and covert attention in infants revealed using steady-state visually evoked potentials. *Developmental Psychology*, *54*(5), 803.

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, *31*(9), 1161-1173.

Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: effects of music background. *Frontiers in neuroscience*, *10*, 191178.

Cohen, L. B. (2004). Uses and misuses of habituation and related preference paradigms. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 349-352.

De Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *elife*, *4*, e06564.

Edalati, M., Wallois, F., Safaie, J., Ghostine, G., Kongolo, G., Trainor, L. J., & Moghimi, S. (2023). Rhythm in the premature neonate brain: Very early processing of auditory beat and meter. *Journal of Neuroscience*, *43*(15), 2794-2802.

Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences*, *112*(31), 9585-9590.

Emberson, L. L., Misyak, J. B., Schwade, J. A., Christiansen, M. H., & Goldstein, M. H. (2019). Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism (s). *Developmental Science*, *22*(6), e12847.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66-108.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention familiar patterns relative to novel ones. *Science*, *146*, 668-670.

Farthouat, J., Franco, A., Mary, A., Delpouve, J., Wens, V., Op de Beeck, M., ... & Peigneux, P. (2017). Auditory magnetoencephalographic frequency-tagged responses mirror the ongoing segmentation processes underlying statistical learning. *Brain topography*, *30*, 220-232.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in cognitive sciences*, *19*(3), 117-125.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222-14227.

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., ... & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science advances*, *7*(8), eabc4530.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 341-348.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*.

Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., ... & Johnson, S. P. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental science*, *18*(1), 90-105.

Kabdebon, C., Peña, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and language*, *148*, 25-36.

Kabdebon, C., Fló, A., De Heering, A., & Aslin, R. (2022). The power of rhythms: how steady-state evoked responses reveal early neurocognitive development. *NeuroImage*, *254*, 119150.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, *7*(5), e36399.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?. *Perception*, *36*, ECVP Abstract Supplement.

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science*, *12*(6), 1020-1025.

Marin, A., Hutman, T., Ponting, C., McDonald, N. M., Carver, L., Baker, E., ... & Jeste, S. S. (2020). Electrophysiological signatures of visual statistical learning in 3-month-old infants at familial and low risk for autism spectrum disorder. *Developmental Psychobiology*, *62*(6), 858-870.

Moser, J., Batterink, L., Hegner, Y. L., Schleger, F., Braun, C., Paller, K. A., & Preissl, H. (2021). Dynamics of nonlinguistic statistical learning: From neural entrainment to the emergence of explicit knowledge. *NeuroImage*, *240*, 118378.

Norcia, A. M., & Tyler, C. W. (1985). Spatial frequency sweep VEP: visual acuity during the first year of life. *Vision research*, *25*(10), 1399-1408.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of vision*, *15*(6), 4-4.

Peykarjou, S. (2022). Frequency tagging with infants: The visual oddball paradigm. *Frontiers in Psychology*, *13*, 1015611.

Peykarjou, S., Hoehl, S., & Pauen, S. (2024). The development of visual categorization based on high-level cues. *Child Development*, *95*(2), e122-e138.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child development*, *80*(3), 674-685.

Raz, G., Cao, A., Bui, M. K., Frank, M. C., & Saxe, R. (2023). No evidence for familiarity preferences after limited exposure to visual concepts in preschoolers and infants. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Ren, J., & Wang, M. (2023). Development of statistical learning ability across modalities, domains, and languages. *Journal of Experimental Child Psychology*, *226*, 105570.

Rekow, D., Baudouin, J. Y., Poncet, F., Damon, F., Durand, K., Schaal, B., ... & Leleu, A. (2021). Odor-driven face-like categorization in the human infant brain. *Proceedings of the National Academy of Sciences*, *118*(21), e2014979118.

Retter, T. L., Rossion, B., & Schiltz, C. (2021). Harmonic amplitude summation for frequency-tagging analysis. *Journal of Cognitive Neuroscience*, *33*(11), 2372-2393.

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4), 491-507.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172-196.

Sáringer, S., Kaposvári, P., & Benyhe, A. (2024). Visual linguistic statistical learning is traceable through neural entrainment. *Psychophysiology*, e14575.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of cognitive neuroscience*, *26*(8), 1736-1747.

Sherman, B. E., Aljishi, A., Graves, K. N., Quraishi, I. H., Sivaraju, A., Damisah, E. C., & Turk-Browne, N. B. (2023). Intracranial entrainment reveals statistical learning across levels of abstraction. *Journal of Cognitive Neuroscience*, *35*(8), 1312-1328.

Slone, L. K., & Johnson, S. P. (2015). Infants' statistical learning: 2-and 5-month-olds' segmentation of continuous visual sequences. *Journal of experimental child psychology*, *133*, 47-56.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological review*, *99*(4), 605.

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child development*, *85*(5), 1821-1826.

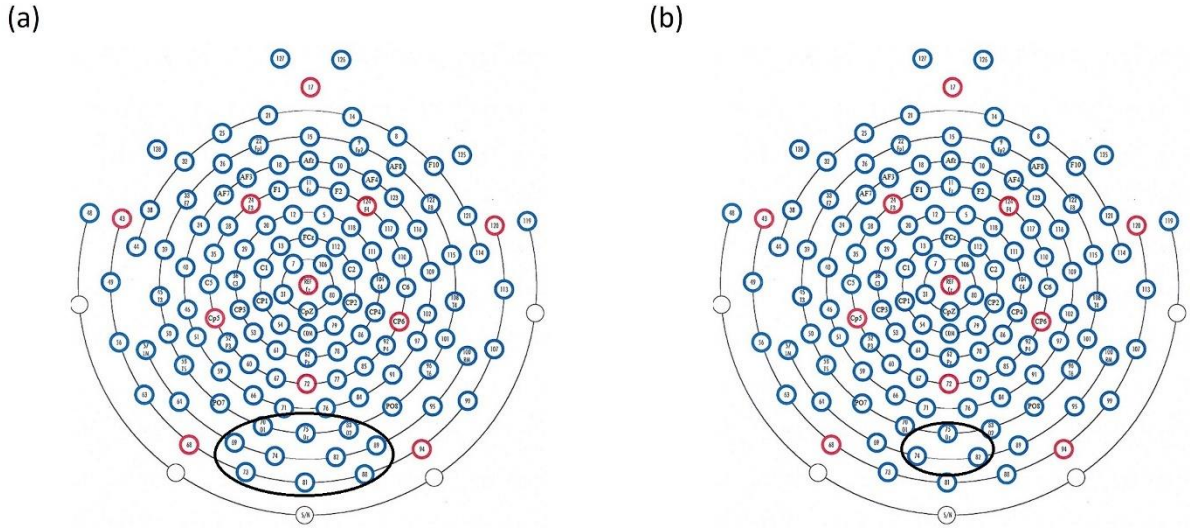
Tononi, G., Srinivasan, R., Russell, D. P., & Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proceedings of the National Academy of Sciences*, *95*(6), 3198-3203.

Tummeltshammer, K., Amso, D., French, R. M., & Kirkham, N. Z. (2017). Across space and time: Infants learn from backward and forward visual statistics. *Developmental Science*, *20*(5), e12474.

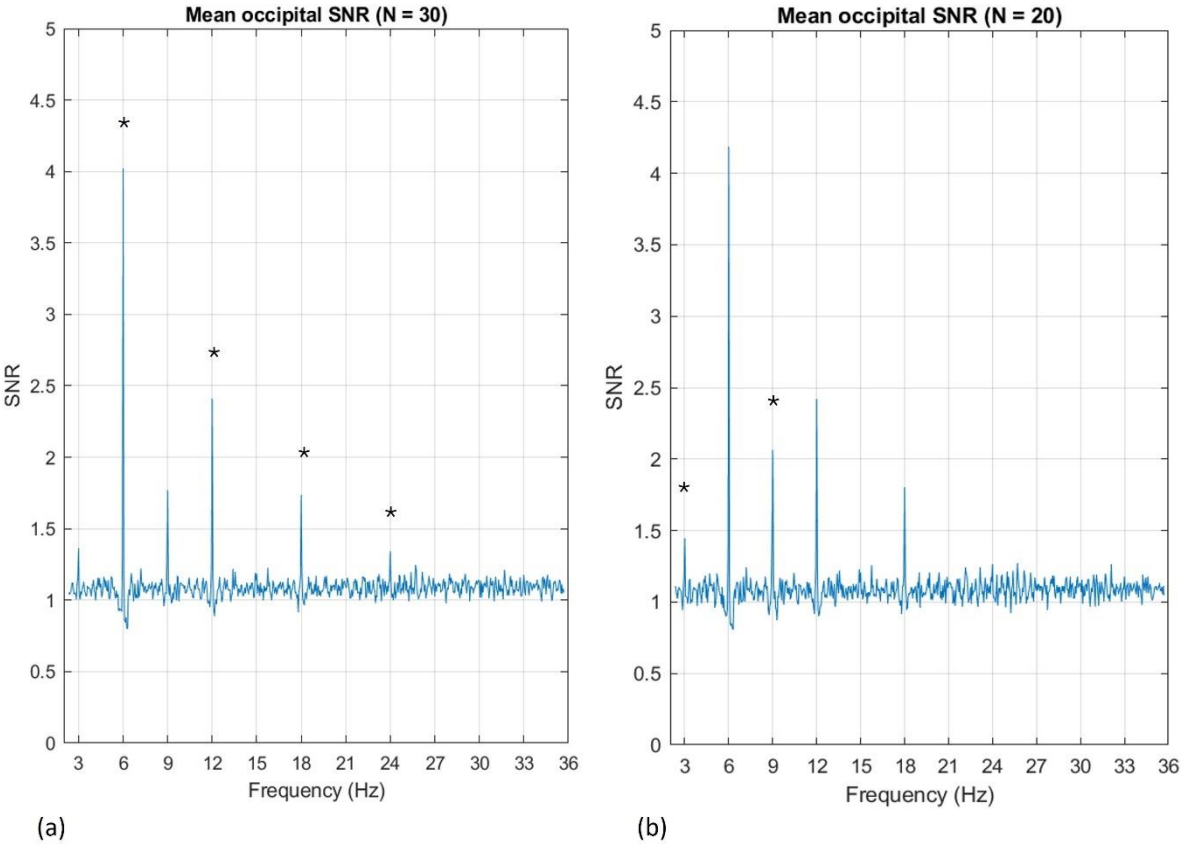
Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of cognitive neuroscience*, *21*(10), 1934-1945.

Wetherford, M. J., & Cohen, L. B. (1973). Developmental changes in infant visual preferences for novelty and familiarity. *Child Development*, 416-424.

Supplementary Materials



Supplementary Figure 1. Regions of interest (ROI) used for data processing and analyses. A 10-channel ROI was a-priori defined for valid trials selection (a). Statistical analyses were run in a smaller ROI of 3 medial occipital electrodes that appeared significant at the group level for both the base and doublet frequencies of interest (b).



Supplementary Figure 2. Mean SNR spectra in the 2 – 36 Hz frequency range. SNR values were averaged over an a-priori defined ROI of 10 occipital electrodes, either across all three experimental conditions (a) or in the doublet and control conditions (b). The base-level harmonics were selected across all conditions; the significant frequencies are marked with an asterisk in graph (a). The doublet-level harmonics were selected considering the doublet and control conditions; the significant frequencies are marked with an asterisk in graph (b).

			Fixed Effects		
	Estimate	**SE**	**95% CI**	**t value**	***p***
Intercept	0.457	0.034	0.389, 0.525	13.237	<.001 ***
Trial Order	-0.021	0.005	-0.031, 0.012	-4.384	<.001 ***
			Random Effects		
	Variance		**SD**		**95% CI**
Intercept	0.011		0.107		0.003, 0.025

Supplementary Table 1. The LMMs results of the base stimulation response (6 Hz and harmonics). Significance codes: "***" p-value [0, .001]. Confidence intervals calculated using the Wald method. Model equation: Base-level SNR ~ Trial Order + (1 | Participant).

Fixed Effects					
	Estimate	**SE**	**95% CI**	**t value**	***p***
Intercept	0.169	0.043	0.083, 0.254	3.884	<.001 ***
Condition (Control)	-0.315	0.103	-0.517, -0.113	-3.059	.003 **
Condition (Random)	-0.558	0.109	-0.771, -0.344	-5.123	<.001 ***
Trial Order	-0.012	0.007	-0.025, 0.001	-1.741	.083
Condition(Control): Trial Order	0.041	0.015	0.012, 0.069	2.775	.006 **
Condition(Random): Trial Order	0.053	0.017	0.019, 0.087	3.079	.002 **

Random Effects			
	Variance	**SD**	**95% CI**
Intercept	0.010	0.100	0.000, 0.023

Supplementary Table 2. The LMMs results of the doublet identification response (3 Hz and harmonics). Significance codes: "***" p-value [0, .001] and "**" p-value [.001, .010]. Confidence intervals calculated using the Wald method. Model equation: Doublet-level SNR ~ Condition + Trial Order + Condition : Trial Order + (1 | Participant).

Fixed Effects					
	Estimate	**SE**	**95% CI**	**t value**	***p***
Intercept	0.798	0.059	0.682, 0.914	13.504	<.001 ***
Condition (Control)	-0.295	0.143	-0.576, -0.014	-2.054	0.051
Condition (Random)	-0.413	0.145	-0.698, -0.128	-2840	.009 **
Random Effects					
	Variance		**SD**		**95% CI**
Intercept	0.052		0.228		0.007, 0.110

Supplementary Table 3. The LMMs results of the learning index, defined as the ratio between doublet-level SNR and base-level SNR. Significance codes: "***" p-value [0, .001] and "**" p-value [.001, .010]. Confidence intervals calculated using the Wald method. Model equation: Learning index ~ Condition + (1 | Participant).