Rhythmic information sampling in the brain during visual recognition

Laurent Caplette[1,2]*, Karim Jerbi[1] & Frédéric Gosselin[1]

[1]Département de psychologie, Université de Montréal, Montréal, QC, Canada
[2]Department of psychology, Yale University, New Haven, CT, U.S.A.

*Corresponding author. Email: laurent.caplette@yale.edu

Abstract

Several studies observed behavioral oscillations in perceptual sensitivity across stimulus presentation time, and these fluctuations have been linked to oscillatory processing in the brain. However, whether specific brain areas show oscillations across stimulus time (i.e., oscillatory sampling) has not been investigated. Here, we randomly revealed features of face images across time and recorded participants' brain activity using magnetoencephalography (MEG) while they performed two recognition tasks. This allowed us to quantify how each snapshot of visual information coming from the stimulus is processed across time and across the brain. Oscillatory sampling was mostly visible in early visual areas, at theta and low beta frequencies. Furthermore, it explained a larger part of brain activity than oscillatory processing. Other parts of the brain showed distinct sampling patterns. These results advance our understanding of the oscillatory neural dynamics associated with visual processing and show the importance of considering the temporal dimension of stimuli when studying visual recognition.

Introduction

In addition to being processed through time in the brain, visual information is also received on the retina through time during the fixation of an object. This temporal aspect of perception is less discussed, but no less important (see VanRullen, 2011; King & Wyart, 2021). Depending on the moment at which it is received, information may be processed differently in a given brain area or even not processed at all (Caplette et al., 2020). These variations in processing across stimulus presentation time may be oscillatory, whereby information received at specific moments reoccurring periodically is processed more efficiently (i.e., a rhythmic sampling; Figure 1).

Previous studies have reported behavioral oscillations in perceptual and attentional sampling across stimulus presentation time (Blais et al., 2013; Dugué et al., 2016; Holcombe & Chen, 2013; Landau & Fries, 2012; Latour, 1967; Fiebelkorn et al., 2013; VanRullen et al., 2007). For instance, the visual threshold for detecting two successive flashes oscillates (around 33 Hz) as a function of the time interval between them (Latour, 1967), and the information used to recognize a face seems to be preferentially sampled at a frequency between 10 and 15 Hz (Blais, Arguin & Gosselin, 2013). In addition, when multiple stimuli are monitored simultaneously, attention seems to fluctuate between them at a 7-8 Hz frequency (Dugué, Roberts & Carrasco, 2016; Helfrich et al., 2018; Holcombe & Chen, 2013; Landau & Fries, 2012; Landau et al., 2015; Fiebelkorn, Saalman & Kastner, 2013; Fiebelkorn, Pinsk & Kastner, 2018; Re, Inbar, Richter & Landau, 2019; VanRullen, Carlson & Cavanagh, 2007); this rhythmic sampling also seems to occur when only one object is monitored (Holcombe & Chen, 2013; VanRullen et al., 2007).

These fluctuations have been linked to underlying brain oscillations. Among other findings, detection accuracy has been found to correlate with the phase or power of ongoing or prestimulus theta oscillations (Busch, Dubois & VanRullen, 2009; Busch & VanRullen, 2010; Hanslmayr, Volberg, Wimber, Dalal & Greenlee, 2013; Helfrich et al., 2018; Fiebelkorn et al., 2018) and neural correlates of perception have been found to depend on prestimulus alpha or theta phase (Gruber et al., 2014; Hanslmayr et al., 2013; Jansen & Brandt, 1991). However, sampling (whether oscillatory or not) in the brain has not been directly investigated. That is, whether specific brain areas show a rhythmic sampling of visual information across stimulus time remains unknown.

To investigate this phenomenon, we used a temporal reverse correlation paradigm in which random visual features were revealed at random moments across stimulus time, and we correlated

neural activity in each brain region and at each time point to the presentation of these features. This allowed us to render and quantify how each visual feature from each stimulus moment is processed across time and across the brain, and thus to characterize rhythmic sampling in the brain.

Results and Discussion

We recorded the brain activity of five neurotypical adults over five days each with MEG. On each trial, we showed face images to the participants, and they had to categorize either the gender (man vs woman) or the facial expression (happy vs neutral) of the face in alternating blocks of trials. Importantly, each face image was only partially visible: we revealed random parts of each face image at random moments across the 200 ms stimulus duration (Figure 1b-c; Videos S1-S2; Caplette et al., 2020; Vinette et al., 2004). Accurate categorizations mostly correlated with the presentation of the eyes in the gender discrimination task, while they mostly correlated with the presentation of the mouth in the expression discrimination task, and these correlations were largely maintained throughout stimulus presentation, with some fluctuations (Figure 1d-e). These results replicate previous findings (Dupuis-Roy et al., 2009, 2019; Faghel-Soubeyrand et al., 2019; Gosselin & Schyns, 2001; Schyns et al., 2002).

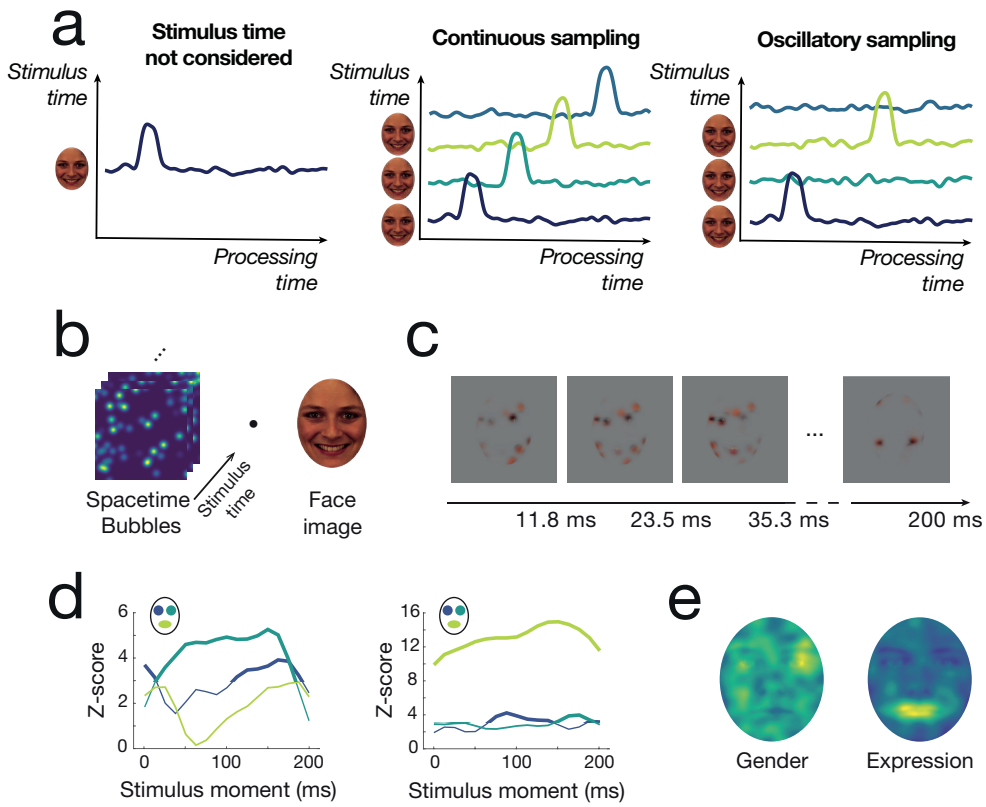

Figure 1. Framework, methods and behavioral results. **a)** (left) When stimulus duration is not considered, a stimulus is effectively conceptualized as an instantaneous event (only one point on the *y* axis). This visual information is then processed across time (the amplitude of the curve illustrating neural activity). (middle, right) In fact, stimuli last a certain amount of time (multiple points on the *y* axis). This entails that information received at each moment during stimulus fixation is processed through time. (middle) Information received at different moments is processed in the same way. (right) Information received at different moments is processed to a different degree, in a periodic fashion. **b)** Experimental stimuli were created by dot-multiplying a 3D tensor sparsely populated with 3D Gaussian apertures (Spacetime Bubbles) by a masked face image. **c)** Example stimulus. Random areas of a random face image were smoothly revealed in random video frames (i.e., *stimulus moments*) across 200 ms. See Videos S1-S2. **d)** Time-resolved behavioral group results. Each line depicts how much each face feature presented on each stimulus moment correlated with accurate responses. Bold segments indicate significant correlations ($Z > 3.12$; $p < .05$, one-tailed, FWER-corrected). (left) Gender discrimination task. (right) Expression discrimination task. **e)** Pattern of behavioral group results for the whole face averaged across time. (left) Gender discrimination task. (right) Expression discrimination task.

Characterizing sampling and processing across the brain

For each subject and task, ridge regressions were performed between the visibility of each face feature (eyes and mouth) at each stimulus moment and MEG activity at each time point and cortical source (Figure 2a). Stimulus time × MEG time maps were thus obtained for each subject, task, face feature and cortical source. These maps reveal an as-of-yet unseen dimension of visual

processing: they show how information received on the retina at different moments throughout stimulus presentation (i.e., *stimulus moments*) is simultaneously processed through time in a specific brain area. By visualizing both stimulus time and MEG time simultaneously, we can untangle the brain activity due to visual information received at different moments, and more fully characterize the dynamics of visual recognition in the brain (see examples of maps on Figure 2b).

In most of these maps, information seemed to be processed in a highly variable manner across stimulus moments. We quantified this variance in brain activity for all brain sources, looking at diagonal time windows where most activity occurred. Processing was highly variable across stimulus moments for multiple areas across the occipital, temporal and parietal lobes, and for both time windows investigated (90-150 ms and 150-210 ms). Significant variance was observed for both tasks and for all face features ($Z > 7.86$; $p < .05$, one-tailed, FWER-corrected; Fig. 2c). This finding confirms that the exact moment at which information is received on the retina matters for its processing in the brain. This variance could be caused by many different underlying processes, including oscillatory sampling, temporal integration, and top-down visual routines (Caplette et al., 2020).

We characterized the different kinds of activity patterns that were occurring across the brain. To do so, we first reduced the number of maps for each subject, task and face feature using nonnegative matrix factorization (NMF). We tested the significance of the activity in these NMF maps and confirmed that significant processing ($Z > 8.12$; $p < .05$, one-tailed, FWER-corrected) occurred in multiple brain areas, at different latencies and for various stimulus moments (see example maps on Figure 3). We then input all NMF maps to a density peak clustering algorithm (Rodriguez & Laio, 2014) to identify reliable types of maps, irrespectively of potential small differences in latencies or amplitudes. We identified three types of patterns. In maps of the first type, a clear diagonal trend was visible in the activations across stimulus moments, meaning that information received at all moments was more or less always processed with a constant delay after its reception. However, information from different stimulus moments was processed differently: information received at some moments elicited a greater activity than information received at other moments, in a somewhat rhythmic fashion (see also next section). This pattern was mostly visible in early visual areas (peaking in the occipital poles) and at early latencies (median of 112 ms)

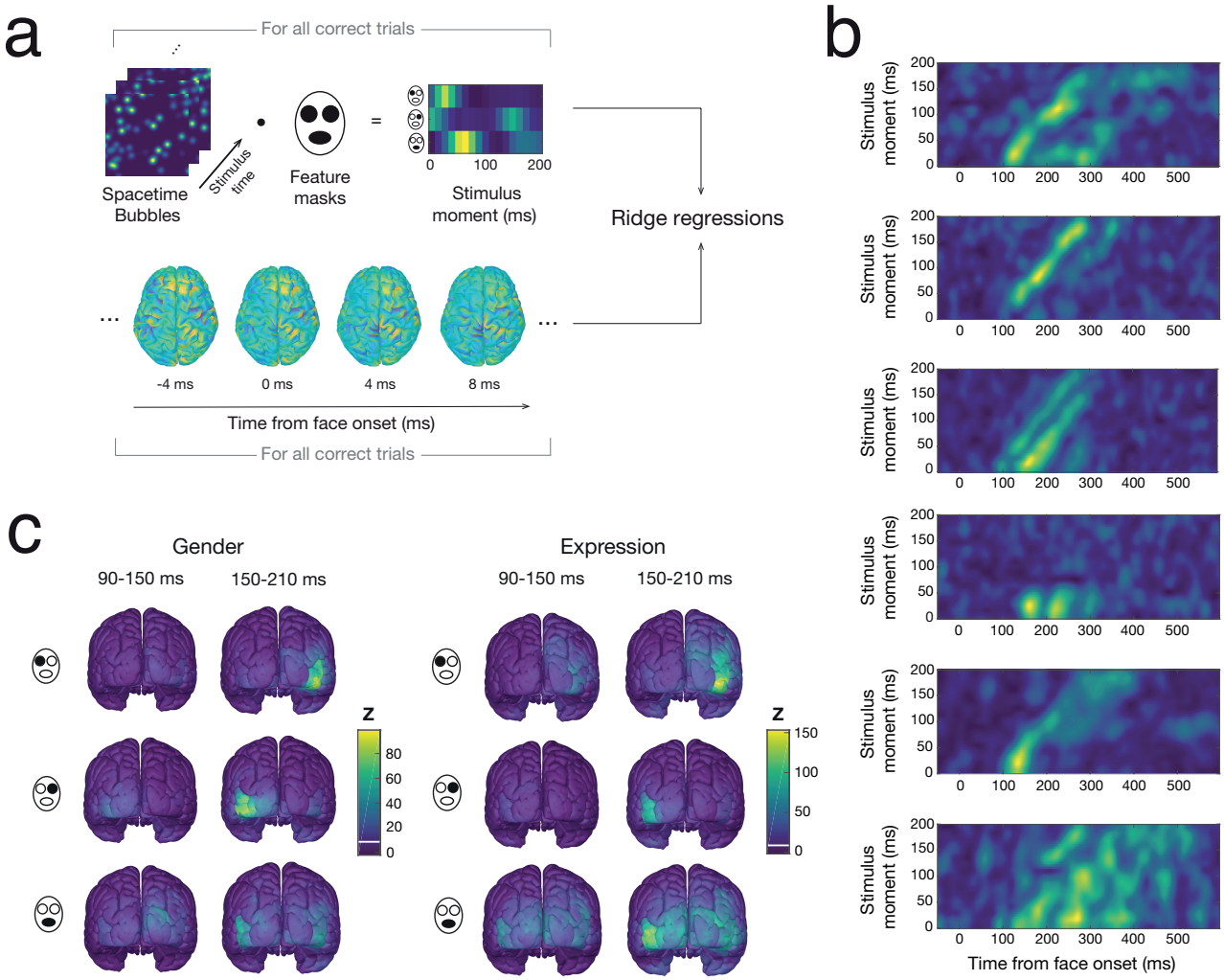

Figure 2. Characterizing sampling and processing across the brain. **a)** Regression analysis. Stimulus dimensionality is first reduced from the full image plane to the 3 main face features (the temporal dimension is unaltered). Then, ridge regressions are performed to uncover how each face feature on each stimulus moment (independent variables) explains brain activity at each cortical source and time point (dependent variables) on correct trials. **b)** Results from the regression can be rearranged in one stimulus time × MEG time map for each participant, cortical source, task, and face feature. Each map indicates how visual information received at each moment on the retina (*y* axis) is processed across time (*x* axis). Example maps are depicted (from top to bottom, these maps correspond to: (i) S1, gender task, right eye, left occipital pole; (ii) S3, gender task, right eye, left occipital pole; (iii) S3, gender task, mouth, left middle occipital gyrus; (iv) S4, gender task, left eye, right inferior temporal gyrus; (v) S4, expression task, mouth, right calcarine sulcus; (vi) S5, expression task, mouth, right middle occipital gyrus). Lighter colors indicate greater processing. **c)** Variance in the processing of each face feature in each task across stimulus moments, in two different time windows. Note that variance is computed in diagonal time windows on the time × time maps (i.e., take the larger delay from later stimulus moments into account; see also Figure 4b). Lighter colors indicate greater variance. White horizontal bars indicate the statistical significance threshold ($Z > 7.86$; $p < .05$, one-tailed, FWER-corrected).

(Figure 3, left). It is consistent with the idea that the earliest brain areas simply process all incoming sensory information in an automatic bottom-up fashion. Another type of pattern was visible in partly overlapping areas (occipital poles) but also extending to slightly more anterior loci (left inferior occipital gyrus and right middle occipital gyrus) and later latencies (median of 172 ms). In maps of this type, only information received in the first 50-75 ms of stimulation was processed, most often in one brief burst, and sometimes in two successive bursts (Figure 3, middle).

Preferential processing of these stimulus moments may occur because the information received early is generally prioritized, possibly because the total amount of useful information to come is uncertain (Nienborg & Cumming, 2009). The third and final cluster includes maps that show late (median of 448 ms) and relatively long-lasting processing of most stimulus moments. This pattern can be found in slightly more anterior brain areas on average, peaking in the left inferior occipital gyrus and extending up to the right parietal cortex (Figure 3, right). These maps likely represent a higher-level processing of sensory information that is occurring in later areas related to object recognition and evidence accumulation (Hanks et al., 2015; Woolnough et al., 2020).

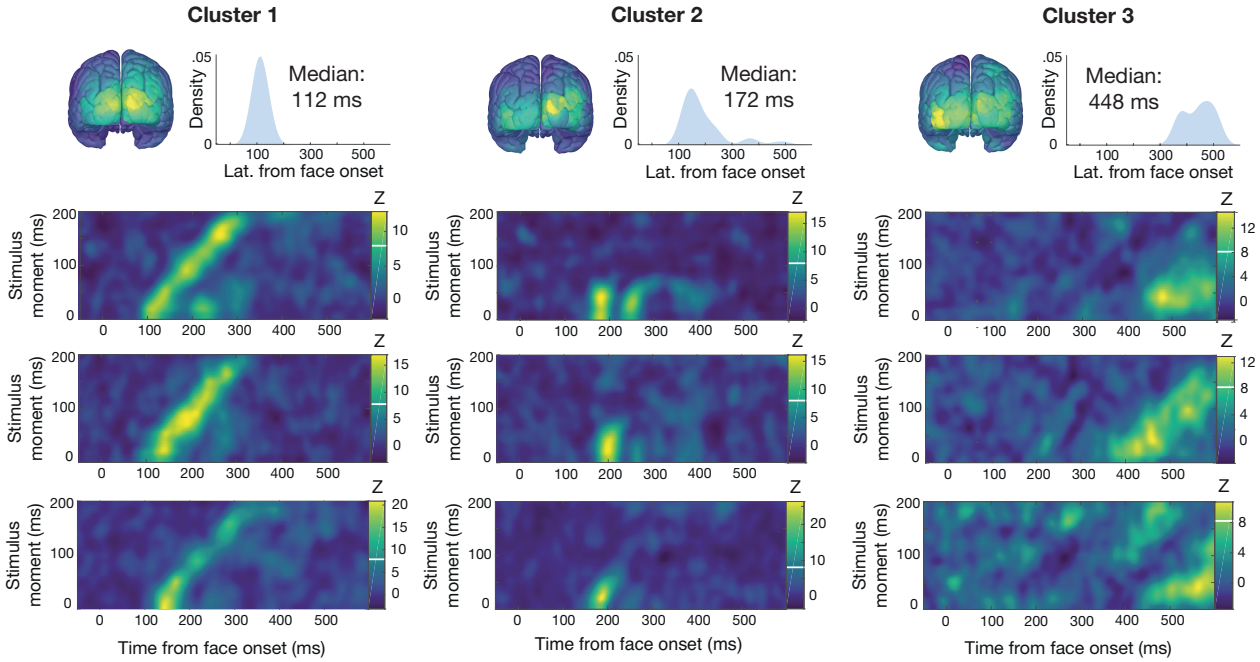

Figure 3. Types of sampling patterns, assessed as clusters in high-dimensional pattern space. For each cluster, the average brain location, the distribution of onset latencies (with median latency) and example NMF maps are depicted. White horizontal bars indicate the statistical significance threshold ($Z > 8.12$; $p < .05$, one-tailed, FWER-corrected). For cluster 1, the maps correspond, from top to bottom, to: S1, gender task, left eye; S3, gender task, mouth; S1, expression task, mouth. For cluster 2: S4, expression task, left eye; S1, expression task, mouth; S3, gender task, mouth. For cluster 3: S4, expression task, mouth; S5, expression task, mouth; S5, expression task, mouth.

Rhythmic sampling across the brain

Since rhythmic sampling is often observed behaviorally (Blais et al., 2013; Dugué et al., 2016; Holcombe & Chen, 2013; Landau & Fries, 2012; Latour, 1967; Fiebelkorn et al., 2013; VanRullen et al., 2007; see VanRullen, 2016, for a review), it should also exist in the brain – but where and when it does remains unclear, especially in the context of object recognition during a typical eye

fixation. Importantly, this oscillatory sampling must not be confused with oscillatory processing: while the latter refers to information that is processed non-continuously, in an oscillating manner, the former refers to information received at successive moments being processed and not processed, in a periodic, alternating way (Figure 4a).

To investigate oscillations in sampling, we performed a Fourier analysis of the activity depicted in the stimulus time × MEG time maps, across the stimulus time dimension (Figure 4b). We first focused on activity in a diagonal time window between 90 and 150 ms after feature onsets. We uncovered significant oscillations at most analyzed frequencies, with peaks at 6.8 Hz and 13.5-16.9 Hz. These oscillations were observed mainly in the occipital poles and contiguous areas but extended to the parietal and temporal cortex ($Z > 5.15$; $p < .05$, one-tailed, FWER-corrected; see example maps on Figure 4c and average magnitudes across the brain on Figure 4d). Significant oscillations were also found in a slightly later time window (150-210 ms), in more anterior areas, but these were predominantly occurring at the 6.8 Hz frequency, with monotonically decreasing power at other frequencies (Figure 4e; oscillations were also much less prevalent in later time windows). This is, to our knowledge, the first direct demonstration of oscillatory *sampling* in the brain. Similar oscillations in the theta and alpha range had previously been observed behaviorally during sustained attention tasks (e.g., Busch & VanRullen, 2010; Fiebelkorn et al., 2013; Landau & Fries, 2012). We also observed significant low beta oscillations between 13 and 17 Hz: these frequencies are similar to those revealed by Blais et al. (2013) who behaviorally assessed the frequencies at which information was sampled by human subjects in a face recognition task (see also Macdonald et al., 2014; VanRullen et al., 2005; Vinette, Gosselin & Schyns, 2004). In general, the presence of sampling oscillations between 3 and 20 Hz is consistent with findings from previous behavioral and electrophysiological studies about perceptual oscillations (VanRullen, 2016).

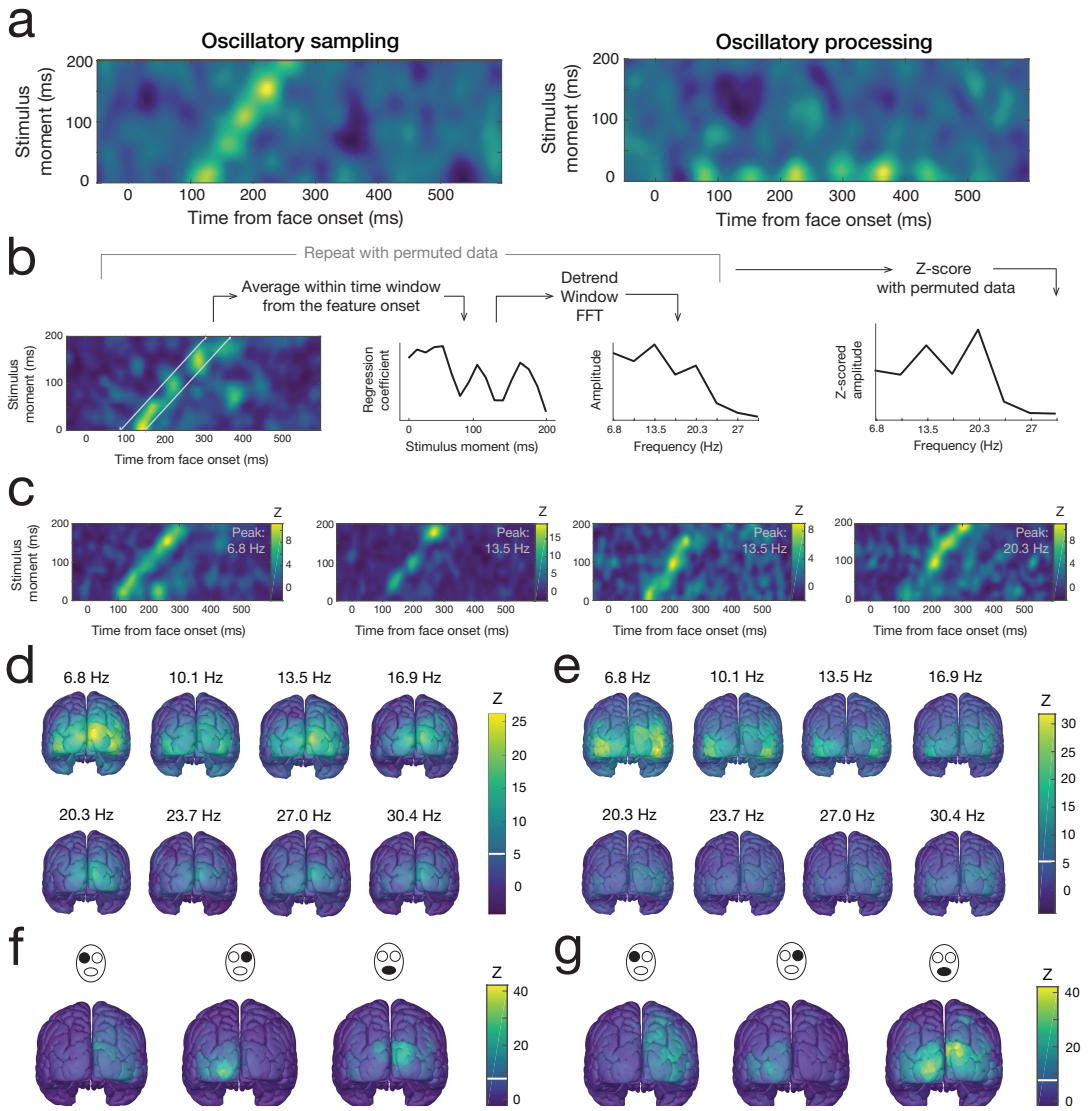

Figure 4. Rhythmic sampling across the brain. **a)** Simulated data illustrating the difference between sampling oscillations (left) and processing oscillations (right). Note that each map depicts only one possible manifestation of the oscillatory phenomenon illustrated. **b)** Sampling oscillations analysis. Activity is first averaged within a diagonal time window to obtain one coefficient per stimulus moment. This series of regression coefficients (including padding, not illustrated) is then detrended, windowed and Fourier-transformed. This analysis is repeated with the permutation distribution of maps and the observed Fourier coefficients are z-scored with the permutation Fourier coefficients. **c)** Example maps with significant sampling oscillations. The sampling frequency with the highest amplitude is displayed in the top right corner. **d-e)** Amplitude of sampling oscillations at each frequency, across the brain, averaged across tasks and face features, for the 90-150 ms time window (d) and the 150-210 ms time window (e). The white horizontal bar indicates the statistical significance threshold ($Z > 5.15$; $p < .05$, one-tailed, FWER-corrected). **f-g)** Sampling oscillations vs processing oscillations model fit, for each face feature, for the gender (f) and expression (g) tasks. Positive z-scores indicate a better fit of the sampling oscillations model. There are no significant negative z-scores. The white horizontal bar indicates the statistical significance threshold ($Z > 7.58$; $p < .05$, two-tailed, FWER-corrected).

Furthermore, these oscillations in sampling (across the vertical stimulus moment axis on the time × time maps) are much more prevalent than oscillations in processing (across the

horizontal MEG time axis on the time × time maps; Figure 4a). To quantify this, we fitted models that explained all the activity within each map with either processing oscillations or sampling oscillations (while accounting for the larger processing delays for later stimulus moments) and assessed which model explained the activity the best (amplitudes and phases were free parameters; see Methods). For all sources with a significant difference between the models, oscillations in sampling explained activity better ($Z > 7.58$; $p < .05$, two-tailed, FWER-corrected; Figure 4f-g). These results suggest that successive cycles of endogenous oscillations in early visual areas are allocated sequentially to information received at successive moments, instead of all coding information received at the same moment(s). This should be especially helpful to the organism if sensory stimuli were to change during the fixation, the visual cortex being continuously updated with the most recent sensory information. In sum, instead of being reflected as an oscillatory processing of some snapshot of the visual field, brain oscillations seem to lead to a rhythmic sampling of the continuous flow of visual information.

Considering stimulus time for a more complete understanding of visual recognition

We developed a novel method to isolate the processing of information received at specific moments across the brain. This allowed us to reveal a new dimension of visual processing and characterize more fully the dynamics of object recognition. We first showed that processing is highly dependent on when information is received on the retina. We further identified three types of patterns of sampling and processing that occur concurrently in the brain. In the first type, mostly visible in the earliest visual areas, information from all stimulus moments was processed around 100 ms after it was received, and processing oscillated as a function of stimulus moment. In the second type of pattern, predominant in similar but slightly more anterior occipital areas, only early stimulus moments were processed, at a latency around 170 ms. Finally, we identified a third kind of pattern: in these cases, information from most stimulus moments was processed around 440 ms and in more anterior regions such as the inferior occipital gyrus and the parietal cortex. Importantly, our method allowed us to reveal and quantify oscillations in sampling (i.e., different stimulus moments being processed to a different degree, in a periodic fashion). We showed that they occur mostly in early visual areas, in the theta, alpha and beta bands, and that they were more prevalent than processing oscillations (i.e., oscillations in the processing of information received at a given moment; e.g., Schyns, Thut & Gross, 2011). These findings further the understanding of the role

of brain oscillations in visual processing. More generally, our results also show that considering the temporal extent of stimuli and eye fixations, in addition to the temporal dimension of brain activity, is necessary to understand visual recognition more fully. Other phenomena could be investigated in future experiments using this paradigm, including temporal integration, attentional sampling and visual routines.

Methods

Participants

Five neurotypical adults (mean age = 27.60; SD = 3.21) were recruited on the campus of the University of Montreal; each came to the laboratory for a total of five experimental sessions (many trials per subject – 5,500 – were necessary to sample with sufficient signal-to-noise ratio the large search space – all image pixels and stimulus moments – explored in this study). Participants did not suffer from any psychiatric or psychological disorder and had no known history of head concussions. The experimental protocol was approved by the ethics board of the Faculty of Arts and Sciences of the University of Montreal and the study was carried in accordance with the approved guidelines. Written informed consent was obtained from all the participants after the procedure had been fully explained, and a monetary compensation was provided upon completion of each experimental session.

Materials

The experimental program ran on a Dell Precision computer with Windows XP in the Matlab environment, using custom scripts and functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner, Brainard & Pelli, 2007). Stimuli were projected on a screen using a Sanyo PLC-XP41L projector, calibrated to allow a linear manipulation of luminance, with a resolution of 1152 x 864 pixels and an 85 Hz refresh rate. A viewing distance of 55 cm was maintained throughout the experiment. MEG activity was recorded using a 275-sensor CTF scanner with a sampling rate of 1200 Hz. Fiducials at the nasion, and left and right temple were used to track head movements in the scanner. Vertical electro-oculogram was bipolarly registered above and below the eyes, and horizontal electro-oculogram was bipolarly registered at the outer canthi of both eyes, to detect blinks and eye movements. An electrocardiogram was used to detect heartbeats. Prior to each session, the surface of the scalp and the positions of the fiducials were digitized using a Polhemus device.

Stimuli

Two hundred and sixty-four color images of faces were selected from the image database *Karolinska Directed Emotional Faces* (Goeleven, De Raedt, Leyman & Verschuere, 2008); only

frontal-view faces were chosen. These were composed of 66 different identities (33 women and 33 men) each performing a happy and a neutral expression; two different pictures of each facial expression were used. Faces were aligned on twenty hand-annotated landmarks averaged to six mean coordinates for left and right eyes, left and right eyebrows, nose, and mouth, using a Procrustes transformation. Face images were then cropped and masked by a centered lightly smoothed elliptical mask to conceal the background, hair and shoulders; final masked faces had a horizontal diameter of 164 pixels, or 6 degrees of visual angle. The mean luminance and contrast of all masked faces were equalized, separately for each color channel, using the SHINE toolbox (Willenbockel et al., 2010).

On each Bubbles trial, random areas of a randomly selected exemplar face were gradually revealed at random moments across a total duration of 200 ms (Caplette et al., 2020; Vinette et al., 2004). In other words, different random parts of a face were revealed on different frames of the stimulus, and the revealed areas were gradually appearing and disappearing (Figure 1c; Videos S1-S2). A duration of 200 ms was chosen so that no saccade could occur during stimulus presentation on most trials. To create a Bubbles stimulus, we first generated a random 480 (pixels) × 480 (pixels) × 25 (frames) sparse tensor (including padding that was not used for the final stimulus) composed of zeros and a few ones. The probability of each element being one was adjusted on a trial-by-trial basis using a gradient descent algorithm to maintain accuracy around 75%. Three-dimensional Gaussian apertures (σ_{space} = 11.4 pixels; σ_{time} = 22.3 ms) were then centred on each one. On each trial, there was a minimum of one bubble in the (non-padding portion of the) stimulus (the average number of bubbles was 94.74). Superfluous padding was then removed so that the tensor was 320 (pixels) × 320 (pixels) × 17 (frames), and thresholding was applied so that no value exceeded one. Each frame of the tensor was finally multiplied elementwise with the same elliptical face mask described above. We called this 3D array *sampling tensor* and the value of each element determined the visibility of a given pixel in a given frame for this trial. Specifically, the face image (repeated across the temporal dimension) and the sampling tensor were multiplied elementwise together, the complement of the sampling tensor was multiplied elementwise with a mid-grey plane and both results were summed together, so that non-sampled parts of the face were replaced by a median grey. Note also that on each trial, the underlying face image had a 50% probability of being flipped vertically, to compensate for possible informational differences between the left and right sides of the face images used.

Experimental design

Each participant came to the MEG Laboratory of the Department of Psychology, University of Montreal, five times. They filled a personal information questionnaire (education, age, sex, hours of sleep, alertness, concussion history, mental illness history, etc.) on the first session. Participants completed between 1000 and 1300 trials on each session, in blocks of 250-275 trials separated by short breaks. On each block, participants performed either a gender discrimination task ("man or woman?") or an expression discrimination task ("happy or neutral?"). Tasks alternated on each successive block and the first task performed was counterbalanced across subjects. Overall, each participant completed 2350 trials of each task: 2250 Bubbles trials (see above) and 100 trials in which whole non-sampled faces were shown. These 100 whole-face trials per task were randomly intermixed with Bubbles trials within four blocks of each task (participants had to perform the same task and faces were also shown for 200 ms). Participants were instructed not to blink during the trials themselves. After every 5 trials, the screen automatically showed text indicating that the participants could take a few seconds to blink and rest their eyes before pressing a key to continue the experiment.

On each trial, a central black fixation cross was shown to the participants for a constant 1500 ms, after which the video stimulus appeared during 200 ms, superposed on the fixation cross, again followed by the fixation cross until the participant responded; a mid-grey background was always present. A fixed interval (1500 ms) between the response and the next stimulus was used so that participants could predict the onset of the trials. Participants had to respond as accurately and rapidly as possible with two keys on the keyboard (key-response associations were counterbalanced across participants).

Behavioral data analysis

Trials with a response time +/- 3 standard deviations away from the mean response time, below 100 ms or above 2000 ms were excluded from further analyses (1.37% of trials) (Caplette et al., 2020). Accuracies were transformed in z-scores across trials for each session. Sampling tensors were downsampled across the spatial dimensions to 64 × 64 pixels and transformed in z-scores across trials. Sampling tensors and accuracy vectors of all sessions were concatenated for each subject and task. An outer product was then performed between the vectorized sampling tensors

and the accuracies for all trials (given the random sampling, this is equivalent to a multiple linear regression). This results in a *classification image* indicating how each pixel correlates to accurate responses. The analysis was then repeated while randomly permuting the accuracies 1000 times to establish a null distribution of classification images, and results were averaged across subjects. To observe the pixel-resolution time-averaged pattern of correlations, all 17 frames were averaged together, and the observed results were z-scored using the distribution of null results. To observe the time-resolved patterns of correlations for the main face features, pixel ROIs were created using lightly smoothed circles (for the eyes) and ellipse (for the mouth). All three ROIs contained a similar number of pixels (within a 1% margin). A scalar product was then performed between the masks of the ROIs and the classification images. Finally, the observed and null results were transformed in z-scores using the distribution of null results, and the maximum statistic method was applied to establish a statistical threshold while correcting for multiple comparisons ($p < .05$, one-tailed, FWER-corrected; Holmes et al., 1996).

MEG preprocessing and source reconstruction

Participants had all been MRI scanned in various studies over the course of the past five years; these anatomical T1 MRIs were used for source reconstruction in the current study. All MRIs were obtained using a MP-RAGE sequence on a 3T Siemens (Trio or Prisma) scanner with a 1 mm × 1 mm × 1 mm spacing.

The cortical surface was extracted using Freesurfer (http://surfer.nmr.mgh.harvard.edu) and downsampled to 8000 vertices using Brainstorm. All further preprocessing was conducted using functions from the Brainstorm toolbox (Tadel, Baillet, Mosher, Pantazis & Leahy, 2011). Anatomical landmarks corresponding to the locations of the MEG fiducials were identified and the MRI was aligned with the MEG sensors. The digitized surface of the scalp was used to refine the coregistration.

When necessary, MEG runs were subdivided so that the head did not move more than 3.5 mm within each analyzed brain activity segment; short segments with large head movements were also excluded in the process. Data from all resulting segments were band-passed between 1 and 40 Hz (Coquelet et al., 2020; Zhan et al., 2019) using a Kaiser FIR Filter with 60 dB stopband attenuation and resampled to a 250 Hz sampling rate to reduce data dimensionality. Bad channels were visually identified and removed. Trials with noisy or anomalous segments, or with

blinks/saccades during or close to the stimulus, were removed following a visual inspection. Signal Space Projection (SSP; Tesche et al., 1995) was used to identify and remove remaining saccade, blink and heartbeat artifacts. When SSP failed to identify artifacts, an ICA and correlations of component time courses with ECG or EOG time courses were used. The data were then segmented into trials from –250 ms to 600 ms from the stimulus onset, and baseline corrected using the average activity between 250 ms and 0 ms before the stimulus onset.

For each run, a forward head model was constructed using overlapping spheres (Leahy, Mosher, Spencer, Huang & Lewine, 1998). A noise covariance matrix was then estimated from (filtered and resampled) empty-room noise recordings from the same day; it was regularized using optimal shrinkage during source estimation (Ledoit & Wolf, 2004). Estimation of the activity at 8,000 points of the cortical surface was performed for each trial using the minimum norm technique. One dipole, oriented normally to the cortical surface, was estimated at each vertex. Source maps were normalized using dynamical Statistical Parametric Mapping (dSPM; Dale et al., 2000).

Creation of stimulus time × MEG time maps

MEG activity was transformed in z-scores across non-excluded correct trials for each day, source and time point. Sampling tensors were then reduced to 3 (face features) × 17 (stimulus frames, or moments) matrices (see Behavioral data analysis). These samples were transformed in z-scores across all correct trials; they were then concatenated across all sessions, separately for each task. Source activity for all correct trials was also concatenated in the same way.

For each task, source, and time point, a ridge regression was performed between samples and source activity of all correct trials, using a regularization parameter of 20,000. Regression coefficients were then transformed into absolute values, and Gaussian-smoothed across the cortical surface (σ_{cortex} = 2.5 mm) and across time (σ_{time} = 12 ms). Analyses were repeated 250 times while randomly permuting sampling matrices across trials to establish an empirical null distribution, and the observed regression coefficients were z-scored using the distribution of null regression coefficients. To visualize the sampling and processing patterns, we rearranged these regression coefficients into one stimulus time × MEG time map for each subject, task, face feature and cortical source.

Analysis of activity variance across stimulus moments

To assess the variance across stimulus moments, we averaged regression coefficients between 90 and 150 ms from the feature onset, and between 150 and 210 ms from the feature onset, for each stimulus moment, and computed the variance across the 17 stimulus moments (diagonal time windows). We repeated these analyses with the permutation maps to obtain a null distribution of variance. We projected the results from individual subjects onto the MNI template brain (15,000 vertices) and averaged them. The observed and null results were transformed in z-scores using the distribution of null results and a statistical threshold ($p < .05$, one-tailed, FWER-corrected) was determined using the maximum statistic method (Holmes et al., 1996). Note that final results were further projected to a higher-resolution template brain (306,716 voxels) for illustration (this is also the case for all other analyses).

NMF and clustering analyses

We then investigated the different types of activity patterns visible across the brains of participants. We first reduced the large number of maps to a few representative exemplars for each subject, task and face feature using nonnegative matrix factorization (NMF). The NMF algorithm requires the *a priori* setting of a k parameter representing the number of components to fit. To choose the value of this parameter k, (i) we divided the trials in two random halves and repeated the regression analyses described above for each half to obtain two sets of time × time maps; (ii) for each possible parameter k between 1 and 12, we computed a R^2 value measuring how well the components extracted by the NMF algorithm on one half could reconstruct the original data on the other half; (iii) we repeated this analysis for both halves and averaged the results; (iv) we chose the k value that maximized the average R^2. We then applied the NMF on the complete dataset using the chosen k value for each subject, task, and feature. The weights thus obtained were finally reapplied to the observed maps and applied to the null maps, and both resulting observed and null NMF maps were transformed in z-scores using the null NMF maps. The resulting NMF maps are the maps depicted on Figure 3. A statistical threshold that corrects for multiple comparisons was established using the maximum statistic method ($p < .05$, one-tailed, FWER-corrected; Holmes et al., 1996).

To identify types of NMF maps, we used clustering by fast search and find of density peaks (Rodriguez & Laio, 2014). All maps resulting from the NMF analysis that had at least some significant activity were given to the clustering algorithm after reducing their dimensionality from

2788 (164 time points × 17 stimulus moments) to 24 (the number of principal components necessary to explain at least 90% of the variance) using PCA. The clustering was computed using the correlation distance between the maps to ensure that the global magnitude of each map was not considered. The points on the decision graph that had a *rho* value larger than 1 and a *delta* value two standard deviations above the mean were selected to form clusters. This resulted in three clusters. The latency distribution of each cluster was computed using the first latency above significance for each map and a Gaussian kernel with a fixed bandwidth of 30 ms.

Analysis of oscillatory sampling and model fit

For this analysis, regression coefficients were not smoothed across the cortical surface or across time. Coefficients within specific time windows from the feature onset (90–150 ms and 150–210 ms), where most brain activity occurs, were averaged together. For each source and face feature, resulting activity across all 25 (stimulus + padding) frames was linearly detrended, windowed and Fourier transformed. Given the short signals, we used a Tukey window with a taper value of 0.1. Fourier coefficients were then smoothed across the cortical surface (σ_{cortex} = 2.5 mm). Frequencies from 6.8 Hz to 30.4 Hz were analyzed; lower and higher frequencies could not be analyzed due to the stimulus length, number of frames and temporal smoothing. We projected the results from individual subjects onto the MNI template brain and averaged them. These analyses were repeated on the null distribution regression coefficients. Both observed and null distribution Fourier coefficients were z-scored with the null distribution Fourier coefficients, and a statistical threshold that corrects for multiple comparisons ($p < .05$, one-tailed, FWER-corrected) was determined using the maximum statistic method (Holmes et al., 1996). Note that our empirical null distribution has the same aperiodic temporal structure as our observed data (unlike if we were to use a time-shuffling method) and therefore our statistical threshold should not be liberal for any frequency (see Brookshire, in press).

To test whether sampling oscillations explained activity better than processing oscillations, we fitted two models to each time × time map: one representing oscillations across the horizontal processing time dimension (i.e., processing oscillations) and one representing oscillations across the vertical stimulus time dimension (i.e., sampling oscillations). To equalize the data on which both models were fit and to account for delays associated with later stimulus moments, the models were fit on maps in which activity for later stimulus moments was shifted left proportionally to

their expected processing delay (because of this necessary shift, only 400 ms of processing were included for each stimulus moment). Each model was created by summing seven sine waves of frequencies from 6.76 Hz to 27.04 Hz in one dimension and repeating the resulting wave across the other dimension. The magnitudes and phases of each sine wave were free parameters (14 free parameters for each model). Best-fit magnitudes were estimated using FFTs (using the same detrending and windowing procedure as above) across the relevant dimension. For the phases, initial values were first estimated using FFTs across the relevant dimension and the models were then fitted 100 times using random values pi radians around these initial values. For each model, the values that minimized mean squared error were chosen. To compute a measure of model fit, the final mean squared errors were smoothed across the cortical surface (σ_{cortex} = 2.5 mm) and the difference between the models was computed. This whole procedure was then repeated with the null time × time maps. For each subject, task, face feature, and cortical source, the observed and null final difference scores were z-scored using the null final difference scores. Results were finally averaged across subjects, and statistical thresholds that correct for multiple comparisons ($p < .05$, two-tailed, FWER-corrected) were established using the maximum statistic method (Holmes et al., 1996).

References

Blais, C., Arguin, M., & Gosselin, F. (2013). Human visual processing oscillates: evidence from a classification image technique. *Cognition, 128*(3), 353–362. doi:10.1016/j.cognition.2013.04.009

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Brookshire, G. (in press). Putative rhythms in attentional switching can be explained by aperiodic temporal structure. *Nature Human Behaviour.*

Bruers, S., & VanRullen, R. (2017). At What Latency Does the Phase of Brain Oscillations Influence Perception? *eNeuro, 4*(3). doi:10.1523/ENEURO.0078-17.2017

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16048–16053.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J Neurosci, 29*(24), 7869–7876. doi:10.1523/JNEUROSCI.0113-09.2009

Caplette, L., Ince, R. A. A., Jerbi, K., & Gosselin, F. (2020). Disentangling presentation and processing times in the brain. *NeuroImage*, 218:116994.

Coquelet, N., De Tiège, X., Destoky, F., Roshchupkina, L., Bourguignon, M., Goldman, S., Peigneux, P., & Wens, V. (2020). Comparing MEG and high-density EEG for intrinsic functional connectivity mapping. *NeuroImage*, *210*:116556.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, *26*(1), 55–67.

Dugue, L., Roberts, M., & Carrasco, M. (2016). Attention Reorients Periodically. *Curr Biol, 26*(12), 1595–1601. doi:10.1016/j.cub.2016.04.046

Dupuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *J Vis, 9*(2), 10 11–18. doi:10.1167/9.2.10

Dupuis-Roy, N., Faghel-Soubeyrand, S., & Gosselin, F. (2019). Time course of the use of chromatic and achromatic facial information for sex categorization. *Vision Research*, *157*, 36–43.

Faghel-Soubeyrand, S., Dupuis-Roy, N., & Gosselin, F. (2019). Inducing the use of right eye enhances face-sex categorization performance. *Journal of Experimental Psychology: General*, *148*(10), 1834–1841.

Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2018). A Dynamic Interplay within the Frontoparietal Network Underlies Rhythmic Spatial Attention. *Neuron*, *99*(4), 842–853.

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. *Curr Biol, 23*(24), 2553–2558. doi:10.1016/j.cub.2013.10.063

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*(6), 1094–1118.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature, 520*(7546), 220–223. doi:10.1038/nature14066

Hanslmayr, S., Volberg, G., Wimber, M., Dalal, S. S., & Greenlee, M. W. (2013). Prestimulus Oscillatory Phase at 7 Hz Gates Cortical Information Flow and Visual Perception. *Current Biology*, *23*(22), 2273–2278.

Helfrich, R. F., Fiebelkorn, I. C., Szczepanski, S. M., Lin, J. J., Parvizi, J., Knight, R. T., & Kastner, S. (2018). Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron, 99*(4), 854–865 e855. doi:10.1016/j.neuron.2018.07.032

Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *J Vis, 13*(1), 12. doi:10.1167/13.1.12

Holmes, A. P., Blair, R. C., Watson, J. D. G., & Ford, I. (1996). Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*, *16*(1), 7–22.

King, J.-R. & Wyart, V. (2021). The human brain encodes a chronicle of visual events at each instant of time through the multiplexing of traveling waves. *Journal of Neuroscience*, *41*(34):7224–7233.

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

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Curr Biol, 22*(11), 1000–1004. doi:10.1016/j.cub.2012.03.054

Landau, A. N., Schreyer, H. M., van Pelt, S., & Fries, P. (2015). Distributed Attention Is Implemented through Theta- Rhythmic Gamma Modulation. *Current Biology*, *25*(17), 2332–2337.

Latour, P. L. (1967). Evidence of internal clocks in the human operator. *Acta Psychologica, 27,* 341–348.

Leahy, R. M., Mosher, J. C., Spencer, M. E., Huang, M. X., & Lewine, J. D. (1998). A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Electroencephalography and Clinical Neurophysiology*, *107*(2):159–173.

Ledoit, O., & Wolf, M. B. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*, 365–411.

Macdonald, J. S., Cavanagh, P., & VanRullen, R. (2014). Attentional sampling of multiple wagon wheels. *Attention, Perception, & Psychophysics*, *76*(1), 64–72.

Nienborg, H., & Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, *459*(7243), 89–92

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Re, D., Inbar, M., Richter, C. G., & Landau, A. N. (2019). Feature-Based Attention Samples Stimuli Rhythmically. *Current Biology*, *29*(4), 693–699.

Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, *344*(6191), 1492–1496.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychol Sci, 13*(5), 402–409. doi:10.1111/1467-9280.00472

Schyns, P. G., Thut, G., & Gross, J. (2011). Cracking the code of oscillatory activity. *PLoS Biology*, *9*(5), e1001064.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*(3):879716.

Tesche, C. D., Uusitalo, M. A., Ilmoniemi, R. J., Huotilainen, M., Kajola, M., & Salonen, O. (1995). Signal-space projections of MEG data characterize both distributed and well-localized neuronal sources. *Electroencephalography and Clinical Neurophysiology*, *95*, 189–200.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*:365.

VanRullen, R. (2016). Perceptual Cycles. *Trends in Cognitive Sciences*, *20*(10), 723–735.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proc Natl Acad Sci U S A, 104*(49), 19204–19209. doi:10.1073/pnas.0707316104

VanRullen, R., Reddy, L., & Koch, C. (2005). Attention-driven discrete sampling of motic perception. *Proceedings of the National Academy of Sciences*, *102*(14), 5291–5296

Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. *Cognitive Science*, *28*(2), 289–301.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods, 42*(3), 671-684. doi:10.3758/BRM.42.3.671

Woolnough, O., Rollo, P. S., Forseth, K. J., Kadipasaoglu, C. M., Ekstrom, A. D., & Tandon, N. (2020). Category selectivity for face and scene recognition in human medial parietal cortex. *Current Biology*, *30*(14), 2707–2715.

Zhan, J., Ince, R. A. A., van Rijsbergen, N., & Schyns, P. G. (2019). Dynamic construction of reduced representations in the brain for perceptual decision behavior. *Current Biology*, *29*(2), 319–326.e4

Acknowledgements

We wish to thank Robin Ince and Nicolas Dupuis-Roy for help with the experimental design and analysis. We also thank the five participants without who this project would not have been possible. Finally, we thank our funding sources: NSERC (Discovery Grant and Alexander-Graham-Bell Doctoral Scholarship), FRQNT (Postdoctoral Fellowship) and University of Montreal (Doctoral Scholarship).

Author Contributions (CRediT)

L.C.: Conceptualization, methodology, software, project administration, investigation, data curation, formal analysis, visualization, writing – original draft, writing – review & editing.

K.J.: Methodology, supervision, validation, writing – review & editing.

F.G.: Conceptualization, methodology, supervision, validation, funding acquisition, writing – review & editing.