Sensory Target Detection at Local and Global Timescales Reveals a Hierarchy of Supramodal Dynamics in the Human Cortex

Maria Niedernhuber, Federico Raimondo, Jacobo D Sitt, Tristan A Bekinschtein

Behavioral/Cognitive

Sensory Target Detection at Local and Global Timescales Reveals a Hierarchy of Supramodal Dynamics in the Human Cortex

Maria Niedernhuber,[1,2*] Federico Raimondo,[3,4,5,6*] Jacobo D. Sitt,[7†] and Tristan A. Bekinschtein[1†]

[1]Cambridge Consciousness and Cognition Lab, Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, United Kingdom, [2]Body, Self, and Plasticity Lab, Department of Psychology, University of Zurich, Zurich, 8050, Switzerland, [3]Brain and Spine Institute, Pitiè Salpêtrière Hospital, Paris, 75013, France, [4]National Institute of Health and Medical Research, Paris, 75013, France, [5]Institute of Neuroscience and Medicine, Brain & Behaviour, Research Centre Jülich, Jülich, 52425, Germany, [6]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, 40225, Germany, and [7]Sorbonne Université, Institut du Cerveau-Paris Brain Institute-ICM, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, APHP, Hôpital de la Pitié Salpêtrière, Paris, 75013, France

To ensure survival in a dynamic environment, the human neocortex monitors input streams from different sensory organs for important sensory events. Which principles govern whether different senses share common or modality-specific brain networks for sensory target detection? We examined whether complex targets evoke sustained supramodal activity while simple targets rely on modality-specific networks with short-lived supramodal contributions. In a series of hierarchical multisensory target detection studies ($n = 77$, of either sex) using EEG, we applied a temporal cross-decoding approach to dissociate supramodal and modality-specific cortical dynamics elicited by rule-based global and feature-based local sensory deviations within and between the visual, somatosensory, and auditory modality. Our data show that each sense implements a cortical hierarchy orchestrating supramodal target detection responses, which operate at local and global timescales in successive processing stages. Across different sensory modalities, simple feature-based sensory deviations presented in temporal vicinity to a monotonous input stream triggered a mismatch negativity-like local signal which decayed quickly and early, whereas complex rule-based targets tracked across time evoked a P3b-like global neural response which generalized across a late time window. Converging results from temporal cross-modality decoding analyses across different datasets, we reveal that global neural responses are sustained in a supramodal higher-order network, whereas local neural responses canonically thought to rely on modality-specific regions evolve into short-lived supramodal activity. Together, our findings demonstrate that cortical organization largely follows a gradient in which short-lived modality-specific as well as supramodal processes dominate local responses, whereas higher-order processes encode temporally extended abstract supramodal information fed forward from modality-specific cortices.

Key words: **attention; electroencephalography; hierarchical predictive coding; multisensory processing; supramodal processing; target detection**

Significance Statement

Each sense supports a cortical hierarchy of processes tracking deviant sensory events at multiple timescales. Conflicting evidence produced a lively debate around which of these processes are supramodal. Here, we manipulated the temporal complexity of auditory, tactile, and visual targets to determine whether cortical local and global ERP responses to sensory targets share cortical dynamics between the senses. Using temporal cross-decoding, we found that temporally complex targets elicit a supramodal sustained response. Conversely, local responses to temporally confined targets typically considered modality-specific rely on early short-lived supramodal activation. Our finding provides evidence for a supramodal gradient supporting sensory target detection in the cortex, with implications for multiple fields in which these responses are studied (e.g., predictive coding, consciousness, and attention).

Received Apr. 4, 2022; revised June 24, 2022; accepted July 20, 2022.

Author contributions: M.N., F.R., J.D.S., and T.A.B. designed research; M.N. and F.R. performed research; M.N. and F.R. analyzed data; M.N. wrote the first draft of the paper; M.N., F.R., J.D.S., and T.A.B. edited the paper; M.N. wrote the paper.

This work was supported by BJA/RCoA WKR0-2018-0060 to T.A.B. J.D.S. was supported by Paris Brain Institute (France); Program "Investissements d'avenir" ANR-10- IAIHU-06; CARNOT Maturation Grant (Con&Heart) to J.D.S.; NeuroCatalyst Grant NeuroEnviroTech; and European Union PerBrain Study (ERA PerMed JTC2019 "PerBrain").

*M.N. and F.R. contributed equally to this work as co-first authors.

†T.A.B. and J.D.S. contributed equally to this work as co-last authors.

The authors declare no competing financial interests.

Correspondence should be addressed to Maria Niedernhuber at mn473@cam.ac.uk.

https://doi.org/10.1523/JNEUROSCI.0658-22.2022

Introduction

The ability to detect deviant sensory events in a stream of predictable stimuli is crucial for adaptive behavior. To enable this, each sense relies on a dedicated hierarchical system of cortices spanning from primary sensory to associative and frontal areas (Felleman and Van Essen, 1991; Rao and Ballard, 1999; Wacongne et al., 2011; Cornella et al., 2012; J. D. Murray et al., 2014; Chao et al., 2018; de Lange et al., 2018). A prominent proposal states each perceptual system encompasses a series of cortices arranged along a simple-to complex hierarchy. By this view, successive levels encode increasingly abstract sensory information along a temporal and topographical hierarchy (Hubel and Wiesel, 1962; Maunsell and Van Essen, 1983; Van Essen and Maunsell, 1983; Riesenhuber and Poggio, 1999; Hochstein and Ahissar, 2002; Kiebel et al., 2008; Nelken and Bar-Yosef, 2008; J. D. Murray et al., 2014; Taylor et al., 2015; Fardo et al., 2017; Ito et al., 2020; Raut et al., 2020; Wengler et al., 2020; Golesorkhi et al., 2021; Çatal et al., 2022).

Converging evidence from target detection studies in different sensory modalities supports the notion that target detection processes are organized along a unimodal-to-supramodal gradient from modality-specific to modality-independent cortices. A rich body of literature identified the prefrontal cortex as a key site for the detection of rule-based targets, such as the spatial conjunction of visual features (Donner et al., 2000; Kristjánsson et al., 2007; Miller, 2009). Conversely, early visual cortices were shown to support the detection of simple visual (Pazo-Alvarez et al., 2003; Ester et al., 2009; Maekawa et al., 2009). In the auditory system, studies mapping the neural basis of simple and complex auditory targets show that higher-order frontal cortices process complex sound patterns, whereas simple pitch deviations are processed in early auditory cortices (Bekinschtein et al., 2009; Wacongne et al., 2011; Cornella et al., 2012; Uhrig et al., 2014; Chennu et al., 2016). Together, earlier work suggests that the cortex processes complex, temporally extended targets while the detection of simple, temporally confined targets largely recruits modality-specific areas (Golesorkhi et al., 2021; Wolff et al., 2022).

Across sensory modalities, sensory targets are followed by two cortical responses, which can be located on successive levels of the cortical hierarchy based on their temporal and cognitive properties: the mismatch negativity (MMN) and the P300 (P3a/P3b) complex. The MMN is an early local negativity associated with temporally proximal sensory change detection (Mäntysalo and Näätänen, 1987; Shinozaki et al., 1998; Näätänen et al., 2001). Originally discovered in the auditory modality (Näätänen et al., 1982), the MMN can be detected in different sensory modalities (Czigler et al., 2006; Allen et al., 2016). Canonically considered to be pre-attentive (Tiitinen et al., 1994), the MMN is modulated by attention but resists distraction (Chennu et al., 2013; Auksztulewicz and Friston, 2015). The P3b is a late distributed temporally extended positivity indexing complex sensory targets in different sensory modalities (Yamaguchi and Knight, 1991a,b; Polich, 2007; Pegado et al., 2010). In contrast to the MMN, the P3b requires memory to track the sensory context in which irregularities are embedded. The P3b also disappears without attention to the sensory irregularity (Squires et al., 1977; Katayama and Polich, 1998; Polich, 2007). Based on these properties, the MMN and P3b can be placed in lower-order and higher-order hierarchical levels, respectively (Bekinschtein et al., 2009; Wacongne et al., 2011; Chao et al., 2018).

Considerable debate revolves around the extent to which local and global ERP responses to sensory targets rely on supramodal activation patterns. Various studies found the P3b to originate in supramodal, but also modality-specific, frontoparietal sources (Halgren et al., 1995, 1998; Katayama and Polich, 1998; Walz et al., 2013; Dreo et al., 2017). Sources in primary sensory and inferior frontal regions generate the MMN in different sensory modalities (R. Näätänen et al., 1982; Pazo-Alvarez et al., 2003; Akatsuka et al., 2007; Garrido et al., 2008, 2009; Ostwald et al., 2012; Shen et al., 2018). Evidence investigating supramodal contributions to the MMN is inconclusive (Chang et al., 2017; Mariola et al., 2019), leaving the question open whether local sensory mismatch responses share common neural signatures between the senses.

We hypothesized that global top-down-driven sensory novelty responses to complex targets in higher-order cortices might share neural signatures between the senses. In contrast, local bottom-up ERP responses to simple targets might be supported by early localized modality-specific activity with only few supramodal contributions (preregistered: https://osf.io/3mqvy/). Our approach exploits differences in the susceptibility of electrophysiological responses to bottom-up and top-down variables to dissociate their neural dynamics in different levels of the cortical hierarchy. Based on earlier work elucidating local and global cortical signaling in the auditory hierarchy (Bekinschtein et al., 2009; Wacongne et al., 2011; Sitt et al., 2014; Chennu et al., 2016; Phillips et al., 2016). we use multisensory versions of a hierarchical oddball paradigm ("local-global paradigm") in which local and global irregularities in the sensory environment elicit P3b-like global top-down responses as well as MMN-like local bottom-up responses (Shirazibeheshti et al., 2018). The local-global paradigm achieves this by manipulating the complexity of the sensory context in which a stimulus change appears, thereby mapping cortical and perceptual local-global hierarchies onto each other (Northoff et al., 2020).

Materials and Methods

Participants. We developed two multisensory versions of the local-global paradigm. In the bimodal version, separate blocks of somatosensory or auditory expectation violations were presented. In addition to purely visual, somatosensory, and auditory blocks, the trimodal version of the paradigm encompassed blocks in which trials combined inputs from two different sensory modalities, which could be visual, auditory, or somatosensory. Only individuals with no history of neurologic or psychiatric conditions and no tactile and auditory impairment were recruited into both studies. In addition, individuals with visual impairments were excluded from participation in the trimodal study. All participants gave written and informed consent. Data for the bimodal local-global paradigm were collected at the University of Cambridge and obtained ethical approval from the Department of Psychology (CPREC 2014.25). For the bimodal study, we invited individuals aged 18-35 to participate through the SONA participant database at the Department of Psychology. We recruited 30 individuals (15 of 15 female/male, mean ± SD age: 24.57 ± 4.52 years) who were paid £10 per hour for a duration of 3-3.5 h. The trimodal task was performed at the Center Nationale de la Recherche Scientifique in France. Participants for this study were invited through the CNRS RISC system. Only individuals aged 18-80 were asked to participate in the trimodal study. Of 54 participants (35 female, mean ± SD age: 25.20 ± 4.10 years) who participated in the trimodal study, 7 were excluded because of a recording error. Participants in the trimodal study were paid €40 for their effort.

Materials. In the trimodal study, auditory stimulation was applied using Etymotic noise-isolating insert earphones. Eccentric Rotating Mass motors controlled by two Texas Instruments DRV2605 haptic drivers were used to deliver vibrotactile stimulation to the wrist. Two independent 8×8 LED matrices implemented in virtual reality goggles

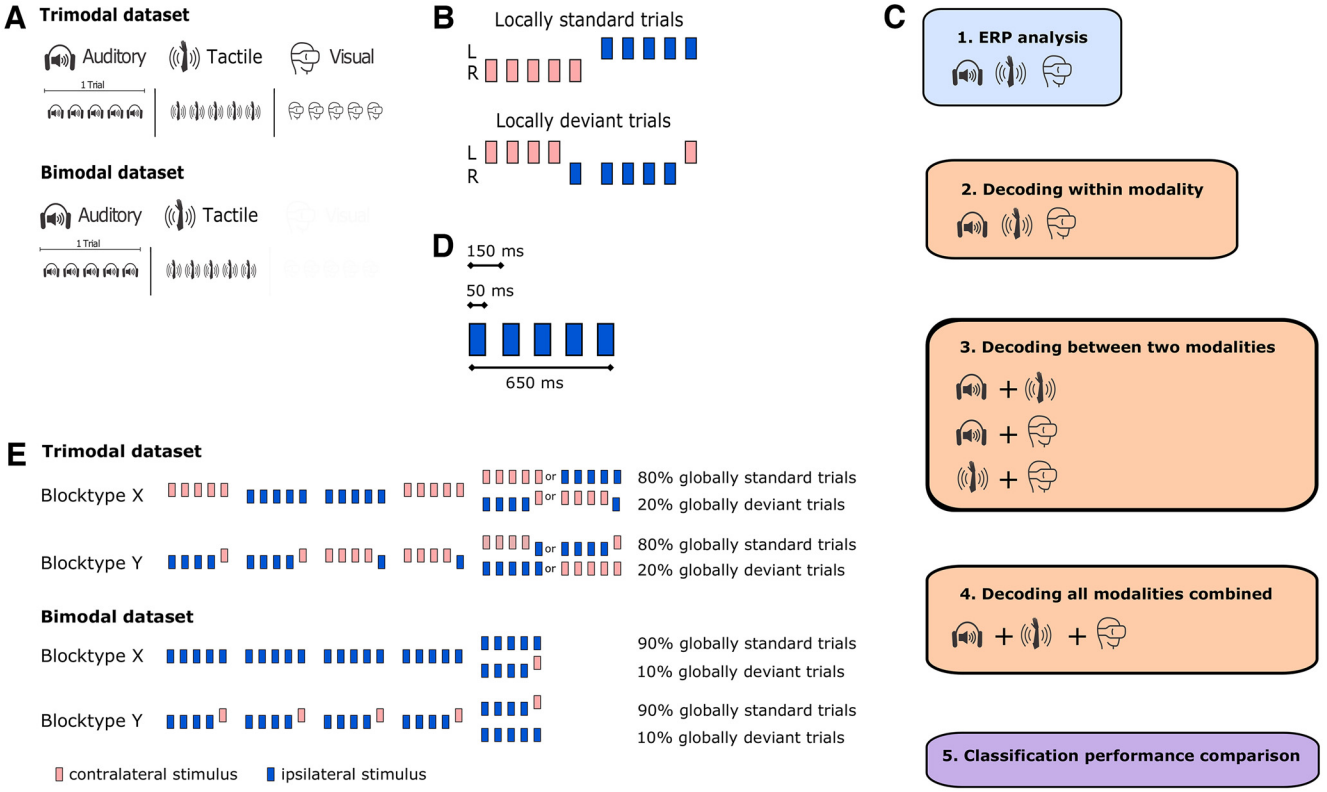

Figure 1. Each trial consists of five stimuli sampled from one sensory modality. Stimuli could be auditory, visual, or tactile in the trimodal study or auditory and tactile in the bimodal study. For each block, trials sampled from specific sensory modality are presented (***A***). Trial design setting up the local contrast between deviant and standard stimuli at trial end: Standard trials consisted of five identical stimuli applied to the same body hemisphere. Deviant trials were composed of four identical ipsilateral stimuli followed by a contralateral stimulus. In the trimodal study, sensory stimulation started on the left side in 50% of trials for each block. In the bimodal study, trials in 50% of blocks started on the left (***B***). We first analyzed ERP time courses to obtain local and global ERP responses for each contrast. Then, we performed a series of three main temporal decoding analyses for each dataset: First, we decoded the temporal evolution of each local and global ERP response within a sensory modality. We trained classifiers on one modality to test them on another for each modality pair. In a further step, we trained and tested classifiers on a combination of trial from all sensory modalities for local and global ERP responses separately. Finally, we performed comparisons of classification performance between local and global ERP responses (***C***). Each stimulus was presented for 50 ms with an interstimulus interval of 150 ms. Trials were presented with a jittered intertrial interval of 1450-1650 ms in the bimodal study and a fixed intertrial interval of 1400 ms for the trimodal study (***D***). Block design setting up the global contrast using nested stimulus groups: In block Type X, locally standard trials dominate the input stream and locally deviant trials occur only rarely at a global level; in block Type Y, this pattern is inverted (***E***).

were used to apply visual inputs isolating visual hemifields. The setup was controlled using an Arduino Zero board. The bimodal study used auditory inputs generated by mixing three sinusoidal signals of 500, 1000, and 2000 Hz (tone Type A) or 350, 700, and 1400 Hz (tone Type B) in MATLAB R2016 based on Chennu et al. (2013) and applied using EARTONE 3A insert earphones. Tactile stimulation was delivered using a custom-made device, which applies mechanical pins to the fingertip with a Saia-Burgess 12.7 mm stroke, 12 v, 4 W DC push-action solenoid with 0.3-0.6 N force, and no nominal delay after current onset) controlled by an Arduino Mega board.

Experimental design. We designed two multisensory variants of the local-global oddball paradigm depicted in Figure 1. In this paradigm, expectations about sensory inputs are violated either locally within trials or globally between trials. Here, trials were composed of five stimuli with a stimulus duration of 50 ms and a stimulus onset interval of 150 ms. Each trial consisted of four identical ipsilateral stimuli followed by a deviant contralateral stimulus (locally deviant trials) or another ipsilateral stimulus (locally standard trials). Contrasting locally deviant and standard trials reveals an MMN-like amplitude difference referred to as local ERP response in a time window between 50 and 250 ms after onset of the last stimulus in a trial. Global violations of sensory expectations are achieved when a frequently presented trial type is occasionally interspersed with a different trial type. A comparison of frequent (globally standard) trials and rare (globally deviant) trials uncovers a global ERP response, which manifests as a late distributed P3b-like positive wave (Bekinschtein et al., 2009; Wacongne et al., 2011; Chennu et al., 2013, 2016; Sitt et al., 2014; Phillips et al., 2016).

During each study, two block types were presented. The bimodal study consisted of 8 auditory and somatosensory blocks. Standard stimuli in 50% of blocks were presented on the left side and on the right side in the remaining blocks. Each block consisted of 78% globally standard trials and 22% globally deviant trials. In block Type X, locally standard trials were occasionally interrupted by 22% locally deviant trials which could equally likely be a locally deviant trial in which only the laterality or both laterality and stimulus type were varied. In block Type Y, a stream of locally deviant trials in which the last stimulus was applied to the contralateral body hemisphere was occasionally interrupted by locally standard trials or locally deviant trials in which the last stimulus differed in laterality and stimulus type. In the trimodal study, blocks were composed of 80% globally standard and 20% globally deviant trials. Trials randomly started on the left or right side within each block with 50% probability. In block Type X, locally standard trials consisting of five identical stimuli were interspersed with locally deviant trials in which the last stimulus was applied to the contralateral hemisphere. Block Type Y consisted of a sequence of locally deviant trials occasionally interrupted by locally standard trials. Blocks in both tasks started with a habituation phase in which the globally standard trial was repeated to establish an expectation of globally recurring stimulus patterns. We presented 24 repetitions of the globally standard in the trimodal study and 15 repetitions in the bimodal study.

In the bimodal study, somatosensory locally standard trials consisted of five touches ipsilaterally applied to the index finger. We introduced two types of locally deviant trials in which the last stimulus in a trial was

applied to the contralateral index finger or contralateral middle finger. In auditory blocks, locally standard trials presented as five identical sounds. Local deviations were introduced by varying either only the laterality of the ear which received the last sound in a trial or both the laterality and pitch of the last sound in a trial. The latter locally deviant trial type was globally deviant type in each block type and the former locally deviant trial type was globally standard in block Type Y and deviant in block Type X. The bimodal study consisted of 8 auditory and 8 somatosensory blocks. The block order was pseudo-randomized so that the experiment started with somatosensory block Type Y followed by somatosensory block Type X, no more than two consecutive blocks were presented in the same sensory modality, and each half of the experiment contained equal proportions of somatosensory and auditory blocks. Each block consisted of 158-160 trials and lasted ~4.5 min. Intertrial intervals were randomly sampled from a uniform distribution between 800 and 1000 ms in steps of 50 ms. In each block, 30-34 globally deviant trials were embedded in a sequence of 112 globally standard trials; and both deviant types occurred in equal proportions. Each globally deviant trial was preceded by 2, 3, 4, or 5 globally standard trials in equal proportions. Participants were exposed to white noise during somatosensory blocks to avoid auditory cues from the tactile stimulation device.

The trimodal study contained blocks with exclusively auditory, visual, or somatosensory stimulation. Participants in this study also underwent multisensory blocks in which expectations violations require to converge inputs from two of these sensory modalities and which were not further analyzed. Local deviations were introduced by varying the laterality of the last stimulus in a trial; 50% of trials applied standard stimuli to the left hemisphere (ear, visual hemifield, or wrist) and the last stimulus to the right hemisphere (and vice versa). Participants underwent three experimental sessions with a total duration of 20 min (4.5 min per session). The remaining two sessions applied cross-modal stimulation, and analyses were not included in this paper. Sixteen participants were presented with somatosensory blocks, 15 participants with auditory blocks, and 16 participants with visual blocks. Each block type was presented twice per session in a fixed order: X-Y-X-Y. 31 (~20%) trials included in each block were globally deviant. Each globally deviant trial was preceded by 3, 4, or 5 globally standard trials. To ensure that participants attend to the global regularity in sensory stimulation patterns, we instructed them to count the number of deviant stimulus groups occurring in a stimulus stream and report the number after each block. Blocks in which participants deviated from the true count by more than two were repeated.

Statistical analysis. EEG data acquisition was performed using a Net Amps 300 amplifier with an Electrical Geodesics 256-channel high density EEG net at the ICM in Paris for the trimodal study and an Electrical Geodesics 128-channel high density EEG net at the Department of Psychology, University of Cambridge for the bimodal study. We performed EEG data preprocessing using EEGLAB 2019 in MATLAB 2019b. In a first step, data were downsampled to 250 Hz, filtered between 0.5 and 30 Hz, and epoched with respect to the onset of the last stimulus in a trial. Habituation trials at the start of each block were removed. Having removed electrodes placed on the neck and cheek which record mostly muscle artifacts, we retained 91 electrodes in the bimodal dataset and 175 electrodes in the trimodal data set for further analysis. We performed baseline removal using a window of 100 ms before epoch onset. Noisy trials (with a variance of >350) and channels (with a variance of >500) were temporarily removed using a semiautomated procedure. We removed artifacts resulting from sweat, eye, and muscle movements using independent component analysis. Ultimately, we removed the remaining artifacts using trial-wise interpolation.

We performed cluster-based permutation analyses to test for differences in ERP amplitude time courses. We used the Common Average as a reference and performed a baseline correction in a time window of 100 ms before onset of the last stimulus in a trial. For each condition pair with unequal trial numbers, trials in the condition with a higher trial number were randomly deselected until the number of trials in both conditions was equal. Cluster-based permutation uses Monte Carlo partitioning to obtain a cluster-level t statistic. To perform Monte Carlo partitioning, data are pooled and randomly divided into two new datasets of equal size. We performed two-sided t tests on the subject averages time-channel pairs and retained only t values with $p < 0.05$. Spatiotemporally adjacent t values were summarized and the largest cluster-level summarized t value was identified. Having performed this procedure 3000 times, we determined the p value corresponding to the proportion of maximal cluster-level t values larger than the observed t value in the original comparison. Conditions were deemed to be different if $p < 0.05$ (Oostenveld et al., 2011).

We applied temporal decoding to examine whether two contrasts rely on similar cortical signatures. Temporal decoding is a machine learning procedure that assesses whether a classifier trained to discriminate two trial types at one time point will generalize to the remaining time points in a sample. We applied a bootstrapping procedure in which 5 trials were randomly sampled from each dataset until we reached 540 epoch averages for each trial type (deviant/standard), respectively. This procedure was repeated 50 times. Classification performance was assessed on each of the resulting datasets. To perform temporal decoding within a sensory modality, we used a fivefold stratified cross-validation procedure in which linear support vector machines were trained to optimally separate standard and deviant trials on 4/5 of the dataset and tested on the remaining data. For temporal decoding between sensory modalities, we fitted classifiers using training and testing datasets from two different sensory modalities. For each condition pair, we trained and tested classifiers on every time point in a time window of 600 ms after onset of the last stimulus in a trial. Classification performance was assessed using the area under the curve receiver operating characteristic (AUC-ROC), which is a nonparametric criterion-free measure of separability. This procedure results in a training time versus testing time temporal generalization matrix with AUC-ROC classification scores as cells. To identify adjacent AUC-ROC scores that differ from chance, we performed a Monte Carlo cluster-based permutation analysis with 1024 random partitions on classification score averages from each bootstrapped dataset and applied two-tailed paired t tests to identify clusters of AUC-ROC values ($p < 0.05$), which differ from chance (King and Dehaene, 2014; King et al., 2014).

Results

Hierarchically nested sensory deviations elicit local and global ERP responses in different sensory modalities

We investigated commonalities in cortical responses to rule-based global and feature-based local sensory deviations using multisensory versions of the local-global paradigm. In a first step, we demonstrated that hierarchical manipulations of sensory context elicit an MMN-like local ERP response and a P3b-like global ERP response in the auditory, somatosensory, and visual modality in both experiments. To that end, we performed cluster-based permutation of ERP amplitude time courses (Maris and Oostenveld, 2007) displayed in Figure 2. Replicating earlier findings (Bekinschtein et al., 2009; Chennu et al., 2013), the auditory local ERP response in the bimodal paradigm manifested as a frontotemporal bipolar two-peak difference wave (cluster $t = -12\,000$, $p < 0.001$) and a one-peak difference wave in the trimodal paradigm (cluster $t = -5396.6$, $p < 0.001$). A somatosensory local ERP response emerged as a central negativity between ~50 and 150 ms in the bimodal paradigm (cluster $t = -752.31$, $p < 0.001$) and as a two-peak negativity between ~100 and 350 ms the trimodal paradigm (cluster $t = -3298.4$, $p < 0.001$). We also identified a visual local ERP response as a negativity in a mid-range time window between ~100 and 350 ms (cluster $t = -6178.9$, $p < 0.001$) shown in Figure 2.

A comparison of globally deviant and standard trials revealed a positive difference wave in a time window from ~250 ms until the end of the trial regardless of sensory modality (with some shifts in onset of the effect). In the bimodal (cluster $t = 22{,}143$, $p < 0.001$) and trimodal study (cluster $t = 27{,}934$, $p < 0.001$), we

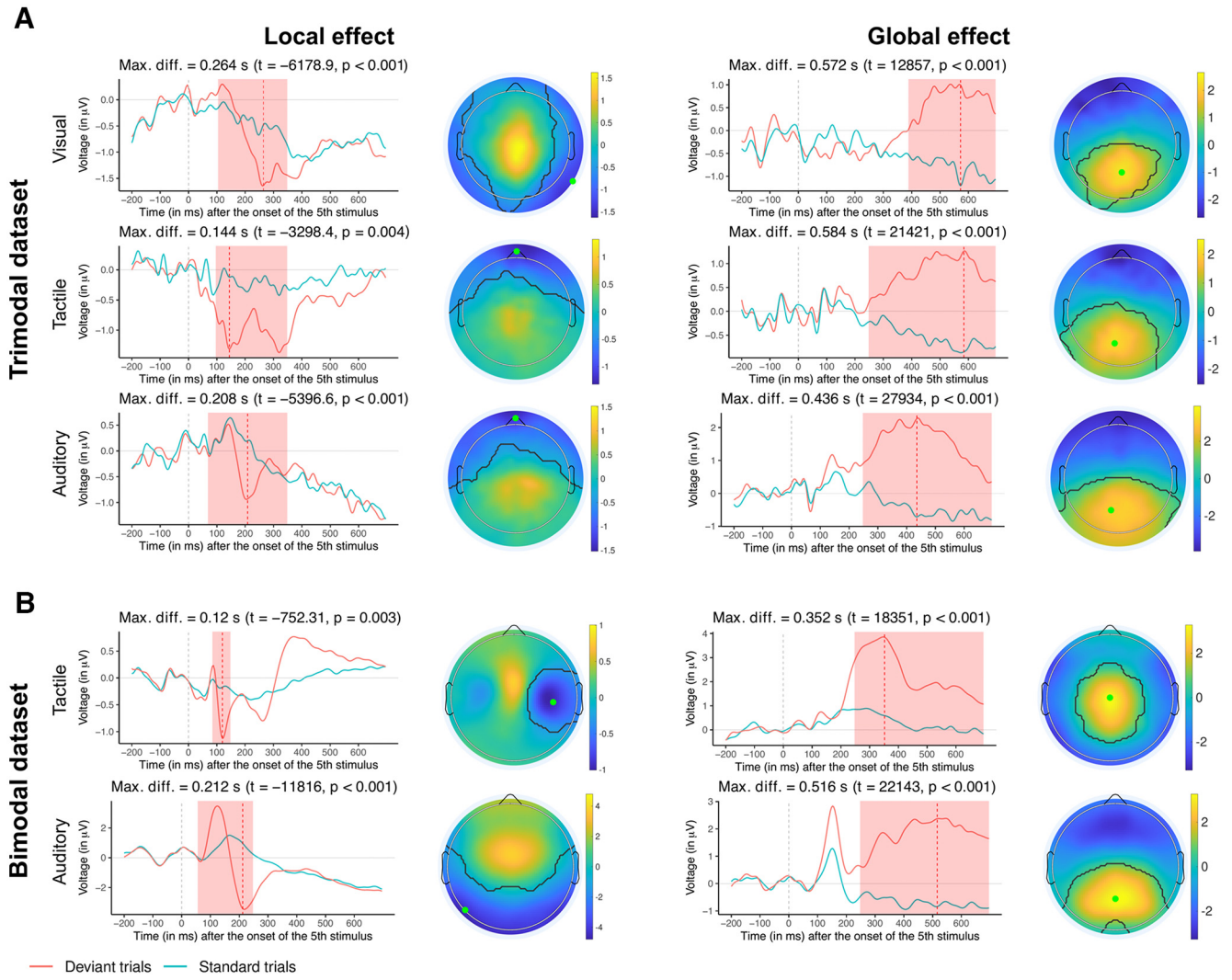

Figure 2. Cluster-based permutation test results for ERP amplitude differences showing that a local and global ERP response can be obtained in different sensory modalities for the trimodal (***A***) and the bimodal study (***B***). On the left, each panel represents ERP amplitude time courses of the corresponding deviant-standard condition pair. Next to the ERP voltage time course plot, the corresponding topographical map is shown. Time periods in which a cluster-based permutation test identified a difference between both conditions are highlighted in red in the ERP time course plot and delineated with a black line on the topographical map. Dotted red line indicates the time point at which the difference between both conditions is maximal in the ERP amplitude time course plot. Orange dot on the topographical map represents the electrode at which this difference was obtained.

revealed an auditory global ERP response as a positive deflection with a posterior distribution. A somatosensory global ERP response presented with a similar posterior distribution in both the bimodal (cluster $t = 18{,}351$, $p < 0.001$) and trimodal study (cluster $t = 21{,}421$, $p < 0.001$). Ultimately, a visual global ERP response emerged as a positive difference wave in a relatively late time window from ~400 ms until the end of the trial (cluster $t = 12{,}857$, $p < 0.001$) shown in Figure 2. These results show that complex targets that require the conscious tracking of sensory patterns across time elicit a global ERP response in different sensory modalities. Our finding that the global ERP response manifests as a large, late, and posterior positive deflection regardless of sensory modality complements previous studies which characterize the related P300 as a late positivity (Bledowski et al., 2004; Bekinschtein et al., 2009; Chennu et al., 2013; Walz et al., 2013). Together, our results show that the functional dissociation of local and global ERP responses is a supramodal property of target detection systems in different sensory domains.

Cortical responses to rule-based but not feature-based sensory targets are sustained across time in each sensory modality

Previous work has shown that functional differences between cortical responses to auditory rule-based and feature-based targets are reflected in the extent to which they are maintained in auditory networks. Cortical activation patterns in response to auditory rule-based targets are sustained in time, whereas feature-based auditory deviations decay quickly (King and Dehaene, 2014). Is the temporal evolution of cortical target detection responses a property common to different sensory domains? We examined whether global ERP responses are linked to a sustained cortical activation pattern, whereas local ERP responses are supported by short-lived activity regardless of sensory modality. We used temporal decoding to characterize neural activation patterns elicited by local and global ERP responses in different sensory modalities. In short, temporal decoding is a machine learning approach used to characterize the temporal evolution of cortical activation patterns linked to sensory events. A classifier

trained at a time point t is not only tested at t but at all other remaining time points. This leads to a temporal generalization matrix of classification performance scores. The shape of the temporal generalization matrix offers insights into the temporal dynamic of cognitive operations and their cortical generators (Dehaene and King, 2016).

We observed that local ERP responses could be decoded along the diagonal in a mid-range time window for each sensory modality (Fig. 3). Across different contrasts, local ERP responses were found to decay quickly. This finding is mostly consistent with a serial activation of different cortices dedicated to the sensory modality in which the local ERP response was applied (King et al., 2014). The visual local ERP response was maintained briefly between 200 and 400 ms (mean AUC = 0.51 ± 0.03, maximum AUC = 0.61 at 268 ms training time and 264 ms testing time, mean cluster $t = 3.8$, $p < 0.05$). In the trimodal study, the somatosensory local ERP response (mean AUC = 0.51 ± 0.02, maximum AUC = 0.58 at 212 ms training time and 216 ms testing time, mean cluster $t = 2.96$, $p < 0.05$) and the auditory local ERP response (mean AUC = 0.52 ± 0.03, maximum AUC = 0.67 at 220 ms training time and 212 ms testing time, mean cluster $t = 4.52$, $p < 0.05$) were maintained for ~100 ms from ~200 ms. In the bimodal study, the somatosensory effect was best decoded between 200 and 300 ms (mean AUC = 0.52 ± 0.03, maximum AUC = 0.63 at 208 ms training time and 208 ms testing time, mean cluster $t = 4.47 \pm 7.53$, $p < 0.05$). In this study, temporal decoding revealed a classification score matrix consistent with two distinct processes underpinning the auditory local ERP response. From ~100 ms, the auditory local ERP response can be decoded along the diagonal, which suggests a serial propagation of cortical activity along the auditory cortical hierarchy. However, cortical activity is sustained for ~150 ms from ~200 ms (mean AUC = 0.51 ± 0.04, maximum AUC = 0.71 at 188 ms training time and 188 ms testing time, mean cluster $t = 2.11 \pm 13.15$, $p < 0.05$). Although classification scores for the auditory and somatosensory local ERP response differed from chance across an extended time window, classifiers performed only slightly better than chance from ~350 ms for both effects.

We found that classifiers trained to distinguish globally deviant and standard trials from ~200 ms generalized across other time samples in the remaining trial window regardless of which sensory modality was tested (Fig. 3). This decoding procedure led to a rectangular classification score matrix for each comparison. In line with earlier ERP time course comparison results (Fig. 2), the visual global ERP response manifested relatively late from ~400 ms (mean AUC = 0.51 ± 0.06, maximum AUC = 0.66 at 516 ms training time and 520 ms testing time, mean cluster $t = 3.19$, $p < 0.05$). In the trimodal study, the somatosensory global ERP response appeared from ~350 ms (mean AUC = 0.52 ± 0.05, maximum AUC = 0.64 at 512 ms training time and 508 ms testing time, mean cluster $t = 5.56$, $p < 0.05$) and the auditory global ERP response from ~200 ms (mean AUC = 0.54 ± 0.07, maximum AUC = 0.69 at 532 ms training time and 524 ms testing time, mean cluster $t = 12.5$, $p < 0.05$). In the bimodal study, temporal generalization was found relatively early from ~150 ms for both the somatosensory (mean AUC = 0.56 ± 0.06, maximum AUC = 0.7 at 372 ms training time and 344 ms testing time, mean cluster $t = 19.32 \pm 19.07$, $p < 0.05$) and auditory global ERP response (mean AUC = 0.6 ± 0.05, maximum AUC = 0.7 at 432 ms training time and 428 ms testing time, mean cluster $t = 33.06 \pm 17.56$, $p < 0.05$). Cluster-based permutation tests comparing differences in classification scores between the local and global ERP response show that the global ERP response generalizes in a late time window, whereas the local ERP response does not (Fig. 3).

Collectively, our results indicate that cortical activation associated with the global ERP response starts no earlier than ~150 ms after onset of the last stimulus in a trial and is sustained over time until the trial ends. However, we observed some shifts in the onset latency between different global ERP responses with the visual global ERP response not appearing before 400 ms. Despite these shifts, this pattern suggests that a single cortical system is active in that time window (King and Dehaene, 2014). This finding leaves the question open whether this system is dedicated to a specific sensory modality or shared between different senses.

Supramodal activation is sustained for global but decays for local ERP responses

Here we provide evidence for the hypothesis that cortical hierarchies dedicated to each sense are organized along a gradient of supramodality. Building up on our finding that global ERP responses are associated with a sustained late cortical activation pattern, we further demonstrate that this sustained pattern is shared between different sensory modalities. We also show that local ERP responses in different sensory modalities rely on few, if any, common cortical signatures in comparison to global ERP responses.

Temporal decoding was used to examine whether cortical responses share neural dynamics between sensory modalities. To identify common activity patterns linked to evoked responses, temporal generalization analysis can be applied in two ways: Classifiers can be trained to separate a deviant-standard condition pair in a target sensory modality at a time point t and tested on deviant-standard condition pairs in a different sensory modality across all time points in a trial. Alternatively, a classifier can be trained to separate deviant-standard condition pairs when they are pooled across all sensory modalities separately for global and local ERP responses. We used a fivefold stratified cross-validation approach in which classifiers were trained on 4/5 of the data and tested on the remaining 1/5 for both analyses (for details, see Materials and Methods).

We initially trained a classifier to separate globally deviant from standard trials when trials for each condition pair are pooled across sensory modalities. This procedure revealed temporal generalization in a late time window across different global ERP responses in both datasets (Fig. 4). We provide evidence for shared activity supporting the auditory and visual global ERP response from ~400 ms regardless of whether classifiers were trained on the visual and tested on the auditory contrast (mean AUC = 0.52 ± 0.07, maximum AUC = 0.66 at 452 ms training time and 572 ms testing time, mean cluster $t = 4.62$, $p < 0.05$) or vice versa (mean AUC = 0.51 ± 0.06, maximum AUC = 0.65 at 676 ms training time and 496 ms testing time, mean cluster $t = 3.37$, $p < 0.05$). A similar pattern was found for somatosensory and visual effects when classifiers were trained on visual and tested on somatosensory deviant-standard pairs (mean AUC = 0.51 ± 0.05, maximum AUC = 0.64 at 540 ms training time and 464 ms testing time, mean cluster $t = 3.2$, $p < 0.05$) or vice versa (mean AUC = 0.51 ± 0.07, maximum AUC = 0.68 at 580 ms training time and 540 ms testing time, mean cluster $t = 2.94$, $p < 0.05$). Shared activity between the auditory and somatosensory global ERP response was found from ~350 ms when classifiers were trained on the auditory and tested on the somatosensory contrast (mean AUC = 0.53 ± 0.05, maximum AUC = 0.64 at 424 ms training time and 580 ms testing time, mean cluster $t = 10.21$, $p < 0.05$)

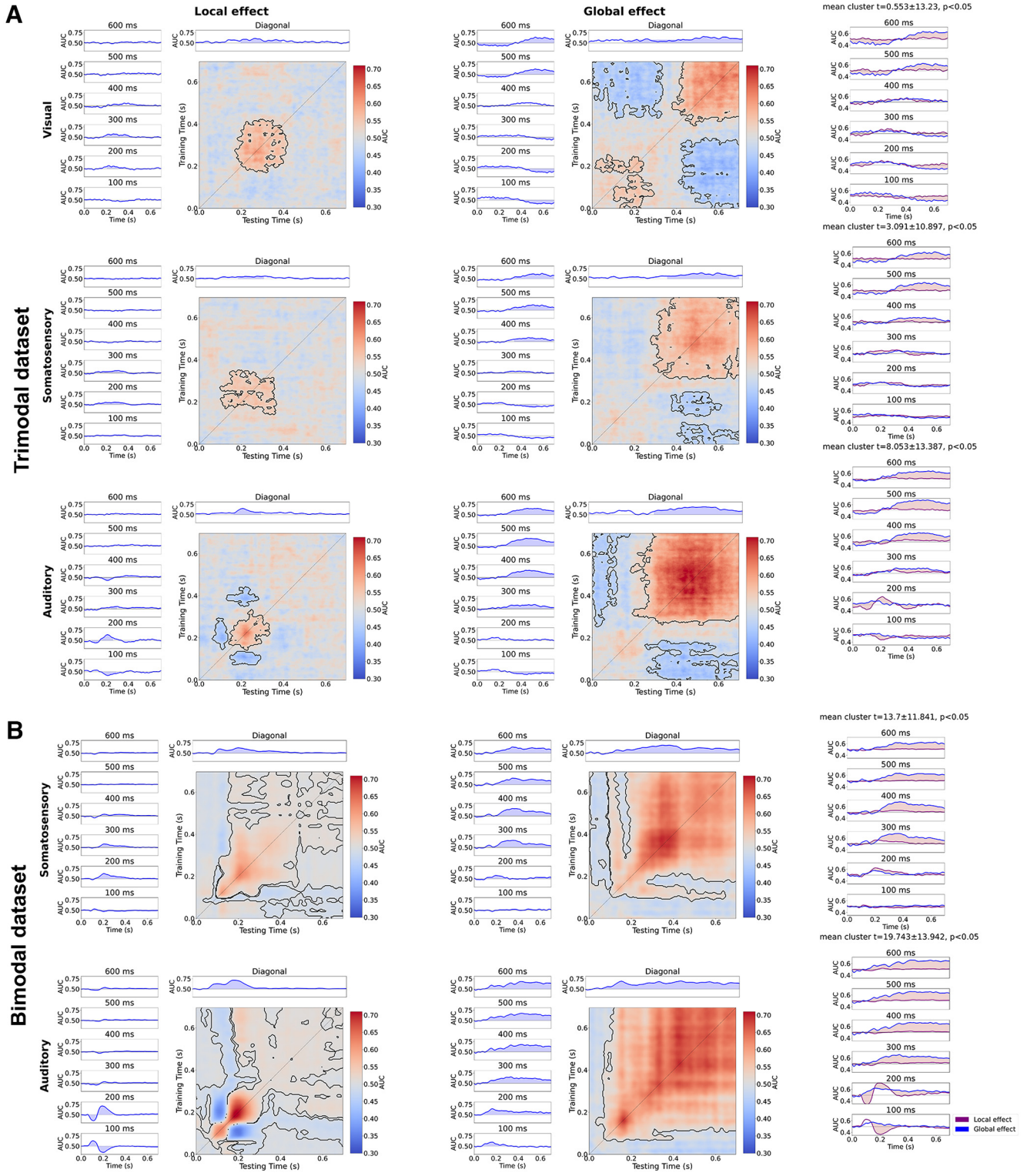

Figure 3. Panels represent temporal generalization results for the local (left) and global ERP response (right) in the trimodal (***A***) and bimodal dataset (***B***). Each panel represents results of a temporal generalization analysis in which a classifier is trained to distinguish deviant and standard trials at each time point and then tested on all remaining time points in a trial. Classification scores are displayed on a red-to-blue gradient. On top of each matrix, adjacent classification scores different from chance (0.5) in a cluster-based permutation test are highlighted with a black line. Next to each matrix, we show a series of time courses of classification scores produced by classifiers trained at the specified time point. On top of each matrix, we plotted the classification score time course for a decoding procedure in which a classifier is trained and tested on deviant-standard pairs at the same time point (corresponding to the matrix diagonal). Blue shade represents classification scores different from chance. On the right, results of a cluster-based permutation test contrasting classification score matrices of the corresponding local and global ERP response. Each subplot represents the classification score time courses of a classifier trained to separate global deviant-standard pairs at the specified time point and tested across the remaining time window, and its local counterpart. Red shade represents time periods in which local and global ERP response classification time courses differ.

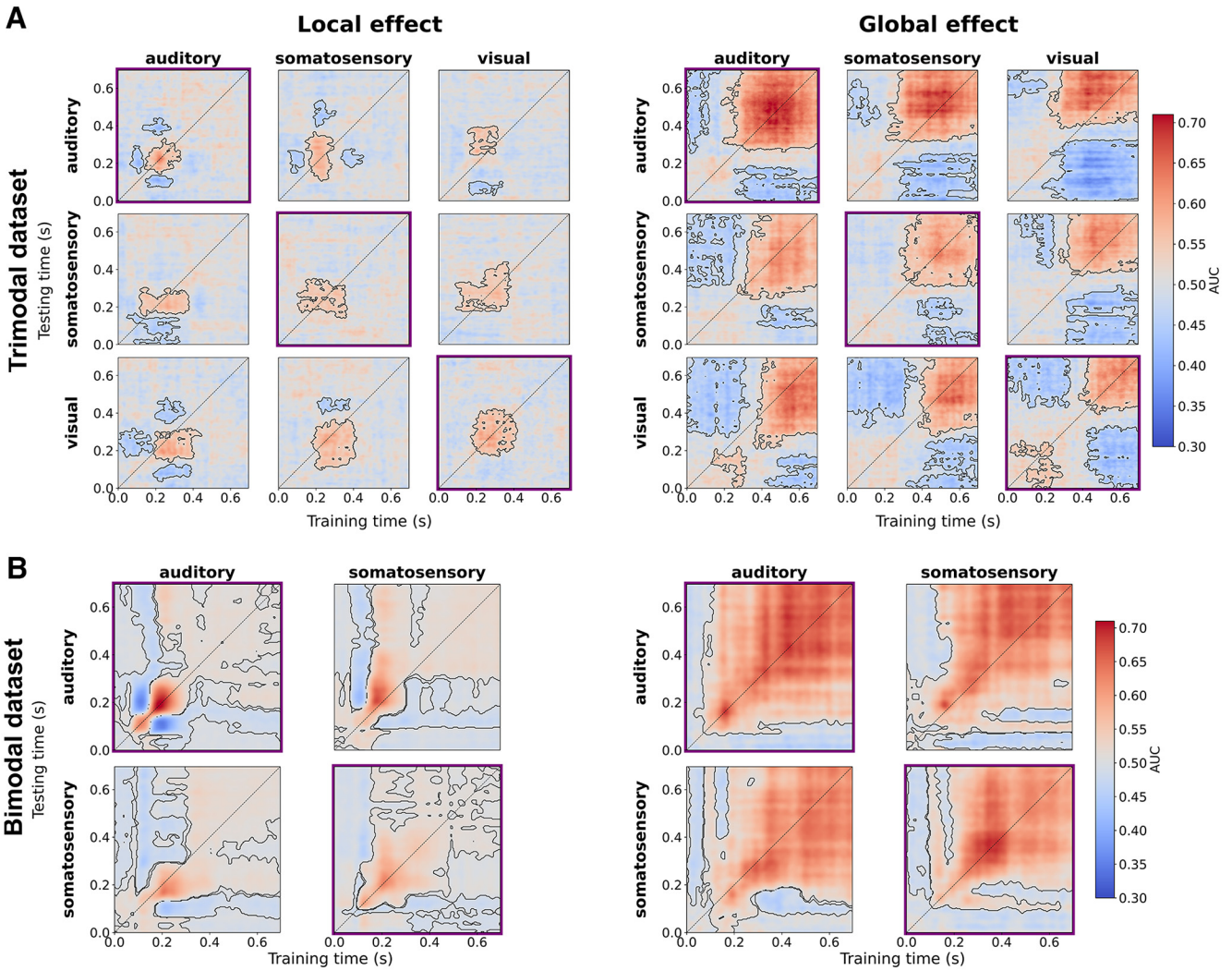

Figure 4. Each panel represents results from a series of temporal generalization analyses decoding the local ERP response (left) and the global ERP response (right) in the trimodal dataset (***A***) and bimodal dataset (***B***). Each matrix represents results of a temporal generalization analysis in which a classifier is trained to distinguish deviant and standard trials at each time point in the training modality and then tested on all remaining time points in a trial in the testing modality. For each experiment, this leads to an $n \times n$ matrix with n being the number of sensory modalities tested in a study (3 for the trimodal and 2 for the bimodal study). Temporal generalization matrices placed along the diagonal of a panel show results from a decoding analysis performed within a sensory modality and are highlighted with a purple frame. The remaining temporal generalization matrices represent results from a decoding procedure in which a classifier is trained to distinguish deviant and standard trials corresponding to the sensory modality indicated next to each row and tested on the sensory modality corresponding to the column label. AUC-ROC classification scores are shown on a red-to-blue gradient. Black line indicates classification score clusters that were found to be different from chance in a cluster-based permutation test; 0 ms indicates the onset of the last stimulus in a trial.

or vice versa (mean AUC = 0.53 ± 0.06, maximum AUC = 0.69 at 576 ms training time and 432 ms testing time, mean cluster $t = 10.26$, $p < 0.05$). This finding was replicated in the bimodal dataset in a more extensive time window from ~150 ms when classifiers were trained on the somatosensory and tested on the auditory effect (mean AUC = 0.57 ± 0.06, maximum AUC = 0.67 at 188 ms training time and 160 ms testing time, mean cluster $t = 22.26 \pm 19.85$, $p < 0.05$) and vice versa (mean AUC = 0.57 ± 0.05, maximum AUC = 0.66 at 292 ms training time and 348 ms testing time, mean cluster $t = 24.75 \pm 17.7$, $p < 0.05$). Together, these findings show that there is shared activation between global ERP responses regardless of which sensory modality is used for training and testing. The global ERP response consistently manifests in a rectangular classification score matrix, which suggests that common activation is maintained in cortical networks (with some temporal shifts in onset times).

Interestingly, temporal decoding of local ERP responses revealed short-lived temporal generalization at ~200 ms in both datasets. In the trimodal dataset, short-lived shared representations from ~200 to 350 ms were found to be associated with local ERP responses across all comparisons: auditory to somatosensory (mean AUC = 0.51 ± 0.03, maximum AUC = 0.58 at 204 ms training time and 252 ms testing time, mean cluster $t = 2.04$, $p < 0.05$), somatosensory to auditory (mean AUC = 0.5 ± 0.03, maximum AUC = 0.62 at 216 ms training time and 200 ms testing time, mean cluster $t = 1.01$, $p < 0.05$), auditory to visual (mean AUC = 0.5 ± 0.03, maximum AUC = 0.62 at 176 ms training time and 268 ms testing time, mean cluster $t = -0.57$, $p < 0.05$), visual to auditory (mean AUC = 0.52 ± 0.02, maximum AUC = 0.59 at 264 ms training time and 192 ms testing time, mean cluster $t = 4.64$, $p < 0.05$), somatosensory to visual (mean AUC = 0.5 ± 0.03, maximum AUC = 0.6 at 208 ms training time and 260 ms testing time, mean cluster $t = 1.34$, $p < 0.05$), and visual to somatosensory (mean AUC = 0.5 ± 0.02, maximum AUC = 0.55 at 260 ms training time and 216 ms testing time, mean cluster $t = -0.93$, $p < 0.05$). In the bimodal

dataset, we found evidence for shared representations from ~200 ms when classifiers were trained on somatosensory and tested on auditory deviant-standard pairs (mean AUC = 0.52 ± 0.03, maximum AUC = 0.68 at 208 ms training time and 176 ms testing time, mean cluster $t = 4.9 \pm 9.02$, $p < 0.05$) and vice versa (mean AUC = 0.51 ± 0.02, maximum AUC = 0.64 at 172 ms training time and 208 ms testing time, mean cluster $t = 2.99 \pm 8.74$, $p < 0.05$).

Our result demonstrates that local ERP responses are supported by supramodal transient activation patterns starting from ~200 ms. Although our finding is compatible with the idea that local ERP responses are processed in cortical hierarchies dedicated to the sensory modality in which the stimulation is applied, we provide evidence for some overlap in higher-order or associative regions between these hierarchies. In sum, these results support the hypothesis that global ERP responses share sustained neural activation patterns between the senses while local ERP responses share fewer, if any, activity (Fig. 4).

To corroborate these findings, we use temporal decoding to examine the temporal evolution of cortical activity when global deviant-standard pairs and local deviant-standard pairs are each pooled across sensory modalities. Our results show that the local ERP response is associated with some temporal generalization from ~180 to 250 ms in both the bimodal (mean AUC = 0.51 ± 0.02, maximum AUC = 0.65 at 204 ms training time and 204 ms testing time, mean cluster $t = 2.75 \pm 8.43$, $p < 0.05$) and the trimodal dataset (mean AUC = 0.51 ± 0.03, maximum AUC = 0.57 at 220 ms training time and 220 ms testing time, mean cluster $t = 3.87$, $p < 0.05$). For local ERP responses, our result is consistent with the involvement of a series of modality-specific neural generators in early levels of the cortical hierarchy and a contribution of supramodal regions in later stages. Again, temporal decoding revealed sustained shared activation starting from ~150 ms until trial end for the global ERP response in the bimodal dataset (mean AUC = 0.579 ± 0.057, maximum AUC = 0.667 at 340 ms training time and 336 ms testing time). In the trimodal dataset, shared activation between global ERP responses extended across the complete time window. We found evidence for sustained activity from ~350 until trial end. Our results also reveal a rectangular classification score cluster, which differs slightly but significantly from chance from the onset of the last stimulus in a trial to ~250 ms (mean AUC = 0.52 ± 0.06, maximum AUC = 0.66 at 520 ms training time and 544 ms testing time, mean cluster $t = 7.35$, $p < 0.05$). Based on work by King and Dehaene (2014), these temporal generalization results suggest that there is a common generator for global ERP responses in different sensory modalities. In contrast, multiple neural generators support the local ERP response. For the local ERP response, our evidence shows that these generators are modality-specific in early stages and supramodal in later stages of cortical processing.

Finally, we examined differences in decoding strength and cluster size between the local and global ERP response. For that, we compared maximum classification scores as well as the number of AUC scores with decoding performance above chance in clusters identified by the cluster-based permutation test of decoding performance scores drawn from all 14 temporal decoding analyses using Mann–Whitney U tests. Cluster size was enhanced for the global ERP response relative to the local ERP response ($U = 2$, $p < 0.001$), which suggests that activation in supramodal networks supporting the global ERP response is sustained in time, whereas supramodal signatures of the local ERP response decay quickly (Fig. 5). Peak decoding performance was also found to be larger for the global than the local ERP response ($U = 36$, $p = 0.001$), indicating that decoding results for supramodal activation linked to the global ERP response are relatively more informative.

Discussion

Hierarchically nested sensory targets elicit local and global ERP responses across sensory modalities

A long-standing debate in neuroscience revolves around which sensory and perceptual processes are supramodal or modality-specific (Driver and Noesselt, 2008; Walz et al., 2013; Faivre et al., 2018; Cao et al., 2019). Inspired by the notion that perceptual systems in the cortex are hierarchically organized on a simple-to-complex axis (Rao and Ballard, 1999; C. Murray, 2004; Kiebel et al., 2008; Dürschmid et al., 2016), we investigated whether cortical responses to sensory targets implemented in successive levels of the cortical hierarchy are ordered along a gradient of supramodality. Across a series of two local-global experiments combining evidence from the somatosensory, visual, and auditory modality, we first established that selective hierarchically nested divergences of sensory inputs can trigger an MMN-like local ERP response and a P3b-like global ERP response in different sensory modalities. Most research on the MMN and P3b concentrates on the auditory domain, and comparably less is known about the visual or somatosensory P3b or MMN and their temporal dynamics (Linden et al., 1999; Ostwald et al., 2012). As has been shown for the auditory modality (King et al., 2014), the global ERP response is maintained in higher-order cortical networks while the local ERP response is serially propagated along cortical areas, which locates both signals at successive stages in the cortical hierarchy. We show that cortical responses to sensory targets rely on activation patterns which are sustained in higher-order cortices across time only when sensory targets are complex and require the attentional tracking of the target for different sensory modalities. Conversely, the detection of targets which deviate from a short preceding stimulus stream and require only short-term memory produce a cortical signal that is propagated along cortical regions in a mid-latency time window regardless of sensory modality. Converging results from different temporal decoding analyses, we conclude that the prolonged maintenance of cortical activation elicited by the global ERP response and the serial propagation of the local ERP response are principles of cortical function found across sensory modalities. This demonstrates that cortical hierarchies implement target detection processes which track sensory irregularities in hierarchically nested different timescales at successive cortical processing stages for each sense.

Rule-based sensory targets elicit supramodal and sustained responses in the cortex

Our finding that the sustained common supramodal activation patterns support the P3-like global ERP response contributes evidence to a controversy around its putative supramodal underpinnings. Some studies investigating cortical activation linked to the auditory and visual P3 suggest a common network, including the insula and frontoparietal areas between these senses (Linden et al., 1999; Walz et al., 2013). However, other studies highlight a contribution of modality-specific higher-order regions to the P300 (Bledowski et al., 2004), leaving the question open whether and which sensory modalities share cortical networks to support P3-like global cortical signals. Our results support the notion that a supramodal (auditory, visual, somatosensory) network underpins the P3-like global ERP response while at the same time not ruling out contributions of modality-specific processes.

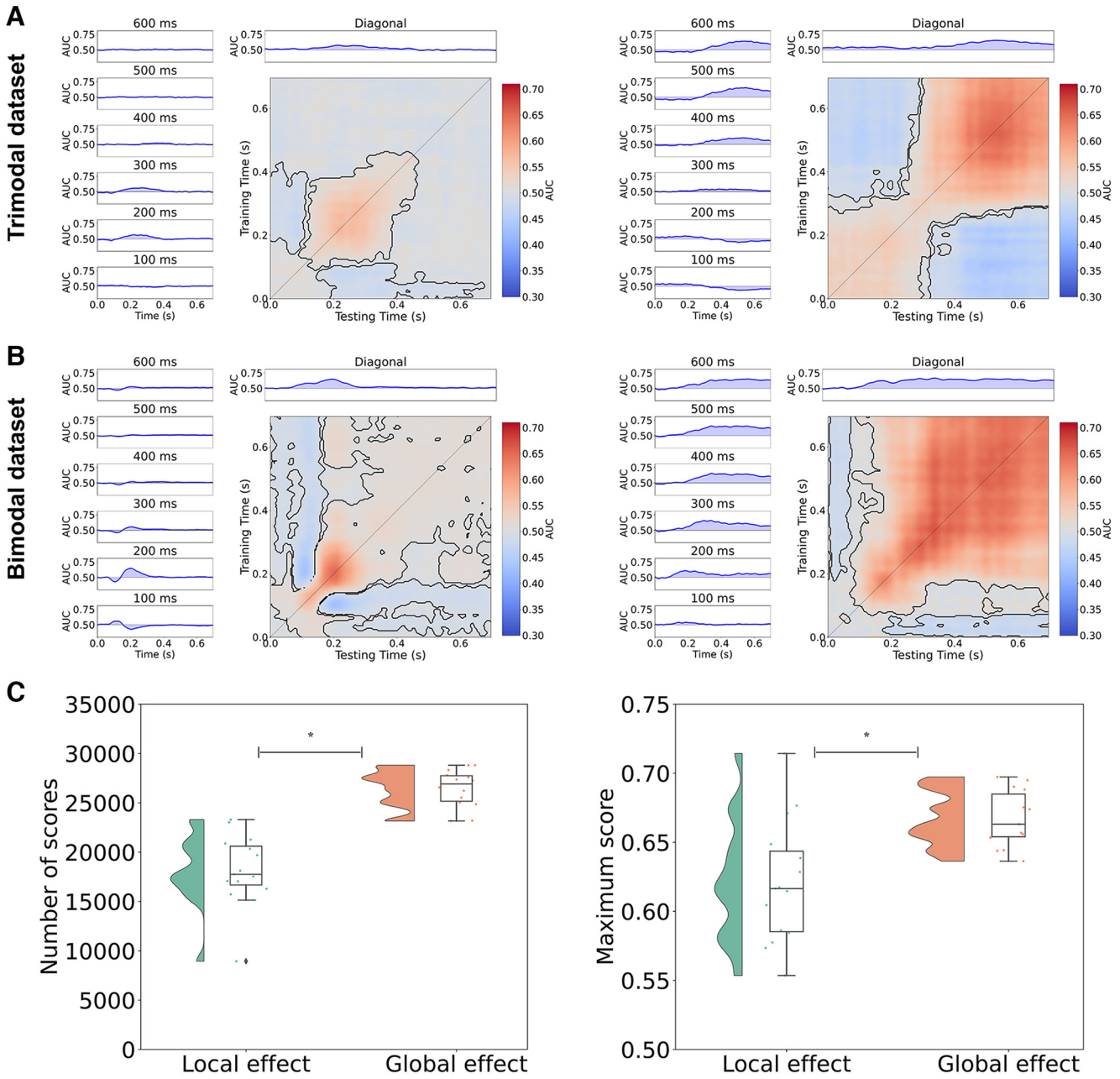

Figure 5. In a temporal decoding analysis, we pooled deviant trials and standard trials regardless of sensory modality and tested whether a classifier trained to discriminate deviant and standard trials at a specific time point might generalize to the remaining time points. Each panel represents the resulting matrix of AUC-ROC classification scores for the global (***A***) and local ERP response (***B***). In each panel: right, results for the global ERP response; left, results for the local ERP response. In conditions in which cluster-based permutation was performed, clusters that differ from chance are highlighted with a purple horizontal line for classification performance in intervals of 100 ms and green for diagonal classification performance; 0 ms indicates the onset of the last stimulus in a trial. Dotted line indicates clusters of AUC-ROC scores which differed from chance. Rain cloud plots supplemented with box-and-whisker plots represent the distribution of maximum classification scores (right) and number of classification scores above chance in each cluster (left) drawn from clusters across all 14 temporal decoding analyses for the local and global ERP response. Significant differences between the local and global ERP response were assessed with a Mann–Whitney *U* test and highlighted with an asterisk (***C***).

Complementing evidence that intrinsic neural timescales are linked to conscious information processing (Zilio et al., 2021), the global ERP response has also been proposed to be a cortical signal reflecting the conscious processing of incoming sensory stimulation (Bekinschtein et al., 2009). By this view, the global ERP response reflects recurrent information flow in a global neuronal workspace which maintains cortical signatures to become consciously accessible (Dehaene and Changeux, 2011). Sensory inputs from modality-specific cortices are fed forward to the global neuronal workspace, which broadcasts integrated multisensory information from the top down to the levels below (Mashour et al., 2020). Our observation that cortical signatures supporting the global ERP response are supramodal aligns with the theory that the global ERP response marks a supramodal top-down-driven process in which sensory information is amplified for conscious access via allocated attention (Chennu et al., 2013).

Local ERP responses to sensory targets are linked to short-lived modality-specific activation

A classic view of MMN-like local ERP responses states that they rely on modality-specific cortical networks involving primary

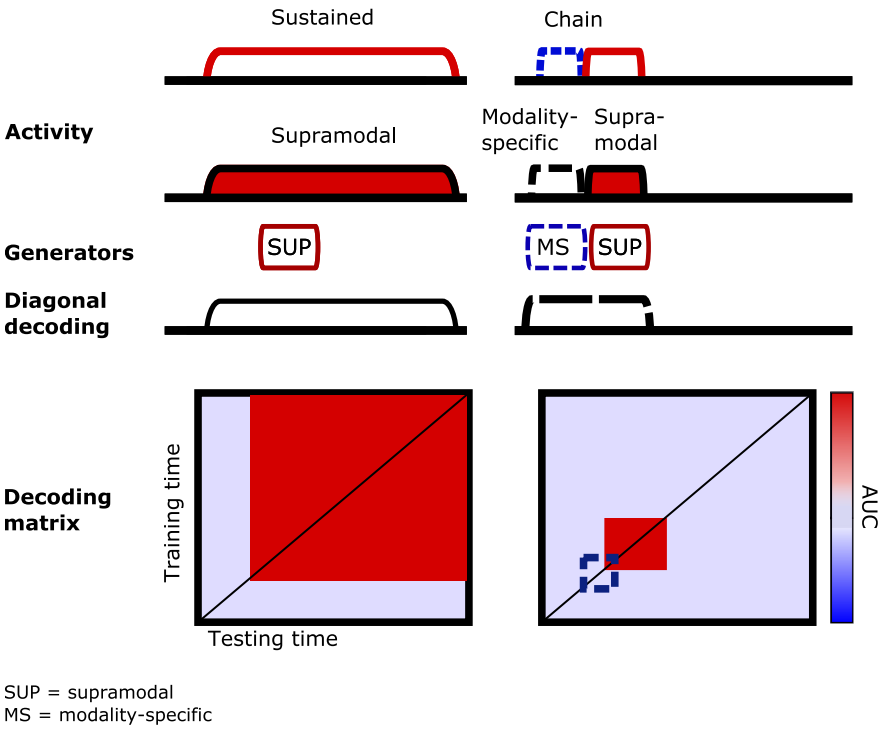

Figure 6. Supramodal and modality-specific aspects of the local and global ERP response. A temporal decoding analysis within sensory modality shows that the global ERP response is supported by one sustained process, whereas the local ERP response likely relies on a chain of processes. Follow-up temporal decoding analyses from one modality to the other, and combining all sensory modalities, revealed that the global ERP response activates a single supramodal network across time, whereas the local ERP response is propagated along the cortical hierarchy in a series of short-lived modality-specific and supramodal processes. We infer that a single supramodal generator contributes to the global ERP response, whereas the local effect is likely supported by a chain of modality-specific and supramodal generators. Cortical activity indexing the global ERP response leads to an extended rectangular classification score matrix and some classification score clusters ordered along the diagonal. Red represents supramodal processes found in temporal cross-decoding between sensory modalities. Blue represents modality-specific processes. We also show a list of supramodal and modality-specific properties of the local and global ERP response (King and Dehaene, 2014).

and secondary sensory regions (Nyman et al., 1990; Pazo-Alvarez et al., 2003). In our study, temporal cross-decoding analyses uncovered short-lived supramodal signatures for the local ERP response starting from ~200 ms after onset of the last stimulus in a trial. Interestingly, previous studies demonstrate that local ERP responses are supported by a network involving modality-specific and frontal regions in which neuronal messages are propagated forward to the inferior frontal gyrus after initial processing in primary sensory areas, raising the possibility that frontal contributions to the MMN might host supramodal signatures. Indeed, both the visual and auditory MMNs were found to consist of an earlier component in modality-specific early sensory cortices followed by an attention-modulated late frontal component from ~200 ms after oddball onset (Deouell, 2007; Hedge et al., 2015). Similarly, studies of effective connectivity underpinning the MMN show that the potential is likely supported by a network spanning primary and secondary sensory cortices as well as frontal regions in different sensory modalities (Garrido et al., 2009; Ostwald et al., 2012; Auksztulewicz and Friston, 2015; Chennu et al., 2016; Fardo et al., 2017). Finally, the temporal characteristics of supramodal signatures supporting the local ERP response are congruent with a contribution of frontal areas linked to attention and target detection (Garrido et al., 2009). Combined with earlier results, our results suggest that the MMN-like local ERP response might consist of short-lived modality-specific and supramodal components. In sum, this finding provides evidence for the idea that successive layers in the cortical hierarchy might support increasingly supramodal processes.

A gradient of supramodality as a principle of cortical organization

The canonical view of cortical function states that cortical hierarchies implement a strict unimodal-to-supramodal gradient. According to this view, supramodal processing is deferred to associative and frontal cortices (Felleman and Van Essen, 1991). Mounting evidence demonstrates that multisensory processes are ubiquitous in the cortical hierarchy and occur at all processing stages, which refutes the idea that early cortices are strictly unimodal (Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008). Integrating both views, our findings support the view that the cortex is hierarchically organized along a gradient of supramodality. Earlier studies used the local-global paradigm to demonstrate that the MMN-like local ERP response is generated in the primary auditory cortex whereas the global ERP response relies on activity in frontoparietal regions (Bekinschtein et al., 2009; Wacongne et al., 2011; Chennu et al., 2013; Uhrig et al., 2014; El Karoui et al., 2015; Chao et al., 2018). In our study, feature-based sensory irregularities triggering quickly decaying and early modality-specific processes were supplemented by a supramodal contribution (Fig. 6). Finally, a sustained late response to rule-based sensory irregularities shared between sensory modalities might reflect a recurrent supramodal process in higher-order cortical areas. However, our finding that local responses evolve into short-lived supramodal activation patterns provides evidence for the notion that early cortical function is not strictly specific to a sensory modality. Crucially, our finding that the P3b-like global ERP response relies on sustained supramodal cortical signatures while the local ERP response elicits early responses with short-lived commonalities between the senses supports the notion of a gradient of supramodality underpinning cortical hierarchies but also refutes the idea that early cortical target detection processes are strictly modality-specific.

Finally, our results can be interpreted as evidence for a predictive coding view of cortical function. Predictive coding states that cortical responses to irregular sensory information reflect a prediction error resulting from a reconciliation of actual sensory inputs and their predictions (Rao and Ballard, 1999; Friston, 2005; Hohwy, 2012; Clark, 2013). From this perspective, local and global ERP responses can be seen as manifestations of prediction errors located at temporally dissociable successive levels of a dedicated cortical hierarchy for each sense (Wacongne et al., 2011). A central idea in predictive coding is that higher-order levels of the cortical hierarchy converge information from different senses forwarded from the levels below to generate predictions about the sensory environment (Friston, 2005; Hohwy, 2012; Clark, 2013; de Lange et al., 2018). This aligns with our result that higher-order cortical responses share supramodal

signatures between the senses, while lower-order responses largely rely on modality-specific activation patterns. Extending these earlier findings, we deliver an integrative framework for cortical responses to sensory targets tracking different timescales at successive levels of the cortical hierarchy across different sensory modalities.

References

Akatsuka K, Wasaka T, Nakata H, Kida T, Kakigi R (2007) The effect of stimulus probability on the somatosensory mismatch field. Exp Brain Res 181:607–614.

Allen M, Fardo F, Dietz MJ, Hillebrandt H, Friston KJ, Rees G, Roepstorff A (2016) Anterior insula coordinates hierarchical processing of tactile mismatch responses. Neuroimage 127:34–43.

Auksztulewicz R, Friston K (2015) Attentional enhancement of auditory mismatch responses: a DCM/MEG Study. Cereb Cortex 25:4273–4283.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA 106:1672–1677.

Bledowski C, Prvulovic D, Hoechstetter K, Scherg M, Wibral M, Goebel R, Linden DE (2004) Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study. J Neurosci 24:9353–9360.

Cao Y, Summerfield C, Park H, Giordano BL, Kayser C (2019) Causal inference in the multisensory brain. Neuron 102:1076–1087.e8.

Çatal Y, Gomez-Pilar J, Northoff G (2022) Intrinsic dynamics and topography of sensory input systems. Cereb Cortex bhab504.

Chang A, Seth AK, Roseboom W (2017) Neurophysiological signatures of duration and rhythm prediction across sensory modalities. bioRxiv 183954.

Chao ZC, Takaura K, Wang L, Fujii N, Dehaene S (2018) Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. Neuron 100:1252–1266.e3.

Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. J Neurosci 33:11194–11205.

Chennu S, Noreika V, Gueorguiev D, Shtyrov Y, Bekinschtein TA, Henson R (2016) Silent expectations: dynamic causal modeling of cortical prediction and attention to sounds that weren't. J Neurosci 36:8305–8316.

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204.

Cornella M, Leung S, Grimm S, Escera C (2012) Detection of simple and pattern regularity violations occurs at different levels of the auditory hierarchy. PLoS One 7:e43604.

Czigler I, Winkler I, Pató L, Várnagy A, Weisz J, Balázs L (2006) Visual temporal window of integration as revealed by the visual mismatch negativity event-related potential to stimulus omissions. Brain Res 1104:129–140.

de Lange FP, Heilbron M, Kok P (2018) How do expectations shape perception? Trends Cogn Sci 22:764–779.

Dehaene S, Changeux JP (2011) Experimental and theoretical approaches to conscious processing. Neuron 70:200–227.

Dehaene S, King JR (2016) Decoding the dynamics of conscious perception: the temporal generalization method. In: Micro-, meso-and macro-dynamics of the brain, pp 85–97. New York: Springer.

Deouell LY (2007) The frontal generator of the mismatch negativity revisited. J Psychophysiol 21:188–203.

Donner T, Kettermann A, Diesch E, Ostendorf F, Villringer A, Brandt SA (2000) Involvement of the human frontal eye field and multiple parietal areas in covert visual selection during conjunction search. Eur J Neurosci 12:3407–3414.

Dreo J, Attia D, Pirtošek Z, Repovš G (2017) The P3 cognitive ERP has at least some sensory modality-specific generators: evidence from high-resolution EEG. Psychophysiology 54:416–428.

Driver J, Noesselt T (2008) Multisensory interplay reveals cross-modal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57:11–23.

Dürschmid S, Edwards E, Reichert C, Dewar C, Hinrichs H, Heinze HJ, Kirsch HE, Dalal SS, Deouell LY, Knight RT (2016) Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc Natl Acad Sci USA 113:6755–6760.

El Karoui I, King JR, Sitt J, Meyniel F, Van Gaal S, Hasboun D, Adam C, Navarro V, Baulac M, Dehaene S, Cohen L, Naccache L (2015) Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: an intracranial study in humans. Cereb Cortex 25:4203–4212.

Ester EF, Serences JT, Awh E (2009) Spatially global representations in human primary visual cortex during working memory maintenance. J Neurosci 29:15258–15265.

Faivre N, Filevich E, Solovey G, Kühn S, Blanke O (2018) Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. J Neurosci 38:263–277.

Fardo F, Auksztulewicz R, Allen M, Dietz MJ, Roepstorff A, Friston KJ (2017) Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. Neuroimage 153:109–121.

Felleman DJ, Van Essen D (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836.

Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM (2008) The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42:936–944.

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2009) Dynamic causal modeling of the response to frequency deviants. J Neurophysiol 101:2620–2631.

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol 120:453–463.

Ghazanfar A, Schroeder C (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278–285.

Golesorkhi M, Gomez-Pilar J, Tumati S, Fraser M, Northoff G (2021) Temporal hierarchy of intrinsic neural timescales converges with spatial core-periphery organization. Commun Biol 4:1–14.

Halgren E, Baudena P, Clarke JM, Heit G, Liégeois C, Chauvel P, Musolino A (1995) Intracerebral potentials to rare target and distractor auditory and visual stimuli: I. Superior temporal plane and parietal lobe. Electroencephalogr Clin Neurophysiol 94:191–220.

Halgren E, Marinkovic K, Chauvel P (1998) Generators of the late cognitive potentials in auditory and visual oddball tasks. Electroencephalogr Clin Neurophysiol 106:156–164.

Hedge C, Stothart G, Todd Jones J, Rojas Frías P, Magee KL, Brooks JC (2015) A frontal attention mechanism in the visual mismatch negativity. Behav Brain Res 293:173–181.

Hochstein S, Ahissar M (2002) View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36:791–804.

Hohwy J (2012) Attention and conscious perception in the hypothesis testing brain. Front Psychol 3:96.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154.

Ito T, Hearne LJ, Cole MW (2020) A cortical hierarchy of localized and distributed processes revealed via dissociation of task activations, connectivity changes, and intrinsic timescales. Neuroimage 221:117141.

Katayama J, Polich J (1998) Stimulus context determines P3a and P3b. Psychophysiology 35:23–33.

Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. PLoS Comput Biol 4:e1000209.

King JR, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 18:203–210.

King JR, Gramfort A, Schurger A, Naccache L, Dehaene S (2014) Two distinct dynamic modes subtend the detection of unexpected sounds. PLoS One 9:e85791.

Kristjánsson Á, Vuilleumier P, Schwartz S, Macaluso E, Driver J (2007) Neural basis for priming of pop-out during visual search revealed with fMRI. Cereb Cortex 17:1612–1624.

Linden DE, Prvulovic D, Formisano E, Völlinger M, Zanella FE, Goebel R, Dierks T (1999) The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. Cereb Cortex 9:815–823.

Maekawa T, Tobimatsu S, Ogata K, Onitsuka T, Kanba S (2009) Preattentive visual change detection as reflected by the mismatch negativity (MMN) for a memory-based process. Neurosci Res 65:107–112.

Mäntysalo S, Näätänen R (1987) The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. Biol Psychol 24:183–195.

Mariola A, Baykova R, Chang AY, Seth AK, Roseboom W (2019) Clear evidence for electrophysiological signatures of duration and rhythm prediction, but not across sensory modalities. bioRxiv 183954.
Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.
Mashour GA, Roelfsema P, Changeux JP, Dehaene S (2020) Conscious processing and the global neuronal workspace hypothesis. Neuron 105:776–798.
Maunsell JH, Van Essen D (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J Neurosci 3:2563–2586.
Miller EK (2009) Executive function and higher-order cognition: definition and neural substrates. Encyclopedia Neurosci 4:7.
Murray C (2004) An interpretative phenomenological analysis of the embodiment of artificial limbs. Disabil Rehabil 26:963–973.
Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Padoa-Schioppa C, Pasternak T, Seo H, Lee D, Wang XJ (2014) A hierarchy of intrinsic timescales across primate cortex. Nat Neurosci 17:1661–1663.
Näätänen R, Simpson M, Loveless NE (1982) Stimulus deviance and evoked potentials. Biol Psychol 14:53–98.
Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) Primitive intelligence in the auditory cortex. Trends Neurosci 24:283–288.
Nelken I, Bar-Yosef O (2008) Neurons and objects: the case of auditory cortex. Front Neurosci 2:107–114.
Northoff G, Wainio-Theberge S, Evers K (2020) Is temporo-spatial dynamics the 'common currency' of brain and mind? In Quest of 'Spatiotemporal Neuroscience.' Phys Life Rev 33:34–54.
Nyman G, Alho K, Laurinen P, Paavilainen P, Radil T, Reinikainen K, Sams M, Näätänen R (1990) Mismatch negativity (MMN) for sequences of auditory and visual stimuli: evidence for a mechanism specific to the auditory modality. Electroencephalogr Clin Neurophysiol 77:436–444.
Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open Source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:56869.
Ostwald D, Spitzer B, Guggenmos M, Schmidt TT, Kiebel SJ, Blankenburg F (2012) Evidence for neural encoding of Bayesian surprise in human somatosensation. Neuroimage 62:177–188.
Pazo-Alvarez P, Cadaveira F, Amenedo E (2003) MMN in the visual modality: a review. Biol Psychol 63:199–236.
Pegado F, Bekinschtein T, Chausson N, Dehaene S, Cohen L, Naccache L (2010) Probing the lifetimes of auditory novelty detection processes. Neuropsychologia 48:3145–3154.
Phillips HN, Blenkmann A, Hughes LE, Kochen S, Bekinschtein TA, Rowe JB (2016) Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. Cortex 82:192–205.
Polich J (2007) Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 118:2128–2148.
Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.
Raut RV, Snyder AZ, Raichle ME (2020) Hierarchical dynamics as a macroscopic organizing principle of the human brain. Proc Natl Acad Sci USA 117:20890–20897.
Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025.
Shen G, Smyk NJ, Meltzoff AN, Marshall PJ (2018) Using somatosensory mismatch responses as a window into somatotopic processing of tactile stimulation. Psychophysiology 55:e13030.
Shinozaki N, Yabe H, Sutoh T, Hiruma T, Kaneko S (1998) Somatosensory automatic responses to deviant stimuli. Brain Res Cogn Brain Res 7:165–171.
Shirazibeheshti A, Cooke J, Chennu S, Adapa R, Menon DK, Hojjatoleslami SA, Witon A, Li L, Bekinschtein T, Bowman H (2018) Placing meta-stable states of consciousness within the predictive coding hierarchy: the deceleration of the accelerated prediction error. Conscious Cogn 63:123–142.
Sitt JD, King JR, El Karoui I, Rohaut B, Faugeras F, Gramfort A, Cohen L, Sigman M, Dehaene S, Naccache L (2014) Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. Brain 137:2258–2270.
Squires K, Petuchowski S, Wickens C, Donchin E (1977) The effects of stimulus sequence on event related potentials: a comparison of visual and auditory sequences. Percept Psychophys 22:31–40.
Taylor P, Hobbs JN, Burroni J, Siegelmann HT (2015) The global landscape of cognition: hierarchical aggregation as an organizational principle of human cortical networks and functions. Sci Rep 5:18112.
Tiitinen H, May P, Reinikainen K, Näätänen R (1994) Attentive novelty detection in humans is governed by pre-attentive sensory memory. Nature 372:90–92.
Uhrig L, Dehaene S, Jarraya B (2014) A hierarchy of responses to auditory regularities in the macaque brain. J Neurosci 34:1127–1132.
Van Essen DC, Maunsell JH (1983) Hierarchical organization and functional streams in the visual cortex. Trends Neurosci 6:370–375.
Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci USA 108:20754–20759.
Walz JM, Goldman RI, Carapezza M, Muraskin J, Brown TR, Sajda P (2013) Simultaneous EEG-fMRI reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. J Neurosci 33:19212–19222.
Wengler K, Goldberg AT, Chahine G, Horga G (2020) Distinct hierarchical alterations of intrinsic neural timescales account for different manifestations of psychosis. eLife 9:e56151.
Wolff A, Berberian N, Golesorkhi M, Gomez-Pilar J, Zilio F, Northoff G (2022) Intrinsic neural timescales: temporal integration and segregation. Trends Cogn Sci 26:159–173.
Yamaguchi S, Knight RT (1991a) Anterior and posterior association cortex contributions to the somatosensory P300. J Neurosci 11:2039–2054.
Yamaguchi S, Knight RT (1991b) P300 generation by novel somatosensory stimuli. Electroencephalogr Clin Neurophysiol 78:50–55.
Zilio F, Gomez-Pilar J, Cao S, Zhang J, Zang D, Qi Z, Tan J, Hiromi T, Wu X, Fogel S, Huang Z, Hohmann MR, Fomina T, Synofzik M, Grosse-Wentrup M, Owen AM, Northoff G (2021) Are intrinsic neural timescales related to sensory processing? Evidence from abnormal behavioral states. Neuroimage 226:117579.