MODELING OF VISUAL COGNITION, BODY SENSE, MOTOR CONTROL AND THEIR INTEGRATIONS

EDITED BY : Hong Qiao and Li Hu

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

MODELING OF VISUAL COGNITION, BODY SENSE, MOTOR CONTROL AND THEIR INTEGRATIONS

Topic Editors:
Hong Qiao, Institute of Automation, Chinese Academy of Sciences & Chinese Academy of Sciences Center for Excellence in Brain Science and Intelligence Technology & University of Chinese Academy of Sciences, China
Li Hu, Institute of Psychology, Chinese Academy of Sciences, China

The interdisciplinary studies between neuroscience and information science have greatly promoted the development of these two fields. The achievements of these studies can help humans understand the essence of biological systems, provide computational platforms for biological experiments, and improve the intelligence and performance of the algorithms in information science.

This research topic is focused on the computational modeling of visual cognition, body sense, motor control and their integrations. Firstly, the modeling and simulation of vision and body sense are achieved by 1) understanding neural mechanism underlying sensory perception and cognition, and 2) mimicking accordingly the structures and mechanisms of their signal propagation pathways. The achievement of this procedure could provide neural findings for better encoding and decoding visual and somatosensory perception of humans, and help robots or systems build humanoid robust vision, body sensing, and various emotions. Secondly, the modeling and simulation of the motor system of the primate are achieved by mimicking the coordination of bones, muscles and joints and the control mechanisms of the neural system in the brain and spinal cord. This procedure could help robots achieve fast, robust and accurate manipulations and be used for safe human-computer interaction. Finally, by integrating them, more complete and intelligent systems/robots could be built to accomplish various tasks self-adaptively and automatically.

Citation: Qiao, H., Hu, L., eds. (2017). Modeling of Visual Cognition, Body Sense, Motor Control and Their Integrations. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-109-8

Table of Contents

Section 5: Machine Learning Algorithms for Pain Prediction

Editorial: Modeling of Visual Cognition, Body Sense, Motor Control and Their Integrations

Hong Qiao[1,2,3]* and Li Hu[4]*

[1] State Key Lab of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [2] Chinese Academy of Sciences Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China, [3] University of Chinese Academy of Sciences, Beijing, China, [4] CAS Key Laboratory of Mental Health, Institute of Psychology (CAS), Beijing, China

Editorial on the Research Topic

Modeling of Visual Cognition, Body Sense, Motor Control and Their Integrations

The interdisciplinary studies between neuroscience and computer science have greatly promoted the development of these two fields. The achievements of these studies can help humans understand the essence of biological systems; provide computational platforms and intelligent methods for biological experiments; and improve the intelligence and performance of the algorithms in computer science.

We present 10 papers in this research topic, which are mainly focused on neural mechanisms underlying the perception of vision, motor and pain; computational modeling of visual processing; bio-inspired visual models; and novel machine learning algorithms to reliably predict pain.

1. NEURAL MECHANISMS RESEARCH OF VISION AND MOTOR

As our dominant sense, the ultimate purpose of visual processing is to support us in perception, cognition, learning, and activities.

One article (Perry et al.) gives a brief review of alterations in visual processing near hand, which supports the hypothesis that there exist parallel, and separate, effector-based attentional systems. Whereas the oculomotor system enhances visual responses through gain modulation, and near-hand attention system sharpens features (such as orientation) relevant to reaching and grasping. This article provides a potential structure for visual-motor interaction modeling in bio-inspired imitation learning.

2. COMPUTATIONAL MODELING OF VISUAL PROCESSING

The work of Galeazzi et al. is much related to the review paper (Perry et al.) on visual processing near hand. The authors analyzed the functions of neurons in VisNet model through a biologically plausible process of unsupervised competitive learning and self-organization both with realistic and natural images. The experiments showed that individual output cells of the network could develop single, localized, hand-centered receptive fields which are invariant to retinal location. Eguchi et al. modified VisNet to model the neural representation of object shape in the primate ventral visual system. By unsupervised visually-guided learning, the individual neurons show similar firing properties with V4 and TEO. The neurons in the higher layer of the network could learn to respond to localized boundary contour elements and show translation invariance across different retinal locations through the use of a trace learning rule.

Both of these two computational modeling methods simulate the principles and mechanisms of the visual pathway, and may inspire the future work in bio-inspired visual modeling for image processing applications.

3. BIO-INSPIRED VISUAL MODELS

Li et al. proposed an enhanced HMAX model for image categorization. By mimicking the attention modulation, memory processing and feature encoding mechanisms of visual cognition, a bottom-up saliency map, an unsupervised iterative clustering method and multi-feature fusion method are introduced to the HMAX model. The enhanced bio-inspired model with small memory size showed better accuracy than other unsupervised feature learning methods in Caltech101 dataset. Fu et al. proposed an CNN model for feature construction in text analysis. By modifying the CNN model to adapt to the text inputs, introducing similarity of asker-answer information as attention modulation, and bringing in reputation information to imitate memory, the improved CNN model showed better performance in answer recommendation task.

Different from the computational modeling of visual processing (Galeazzi et al.; Eguchi et al.), these two bio-inspired visual models had excellent performance in public datasets focusing on computer science application, which shows that biological research can promote the development of computer science.

4. NEURAL MECHANISMS UNDERLYING THE PERCEPTION OF PAIN

Pain is a subjective first-person experience, and self-report is the gold standard to determine pain in various clinical practice. Considering that self-report of pain is not available in some vulnerable populations, the development of an objective assessment of pain would be highly needed in clinical applications (Huang et al., 2013). To achieve this aim, we need to (1) identify neural activity that could serve as a cortical signature for pain perception in humans using non-invasive functional neuroimaging techniques, e.g., electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), and (2) develop novel algorithms that could reliably predict the perceived pain based on the identified pain-related neural responses (Hu and Iannetti, 2016).

Three articles provide recent advances to better understand the neural mechanism related to the central processing of pain perception. Guo et al. investigated the vigilance states of the brain when the subjects were suffering from acute pain or chronic pain, and demonstrated that the vigilance level to external sensory stimuli would be increased with acute pain, but decreased with chronic pain. These observations indicated that the study of pain-induced influences on cortical processing of non-nociceptive sensory information would be a doable way to differentiate acute pain and chronic pain, thus help monitor the progress of pain chronification in clinical practice (Guo et al.). In addition, Li et al. investigated the effects of placebo analgesia

on spontaneous brain oscillations during tonic muscle pain. They observed that placebo-induced decreases in the subjective pain perception significantly correlated with the increases of the amplitude of alpha oscillations, which suggested that alpha oscillations in frontal-central region could serve as the cortical indicator of placebo effect on tonic muscle pain (Li et al.). Finally, Peng and Tang provided a comprehensive summary of the functional properties of pain-induced modulations of ongoing cortical oscillations. In addition to the traditional methods, they proposed that novel approaches should be adopted to comprehensively explore the dynamics of oscillatory activities associated with pain perception and behavior. Based on these understandings, Peng and Tang pointed out the potential clinical applications of neurostimuation techniques (e.g., repeated transcranial magnetic stimulation (rTMS) and transcranial alternating current stimulation (tACS)) based on the modulation of pain-related cortical oscillations, which could help promote the establishment of rational therapeutic strategy in the framework of intelligent systems.

5. MACHINE LEARNING ALGORITHMS FOR PAIN PREDICTION

Two articles in this Research Topic developed novel techniques to improve the performance of pain prediction based on non-invasive functional neuroimaging signals. Bai et al. observed that pain-evoked EEG responses were significantly correlated with spontaneous EEG activities at interindividual level, and proposed a normalization approach to reduce the interindividual variability of pain-evoked EEG responses based on the spontaneous EEG activities for each subject. In addition, Bai et al. found that the relationship between pain-evoked EEG responses and pain perception was nonlinear, which inspired them to develop a novel two-stage pain prediction strategy, a binary classification of low-pain and high-pain trials followed by a continuous prediction of high-pain trials to significantly improve the prediction accuracy (Bai et al.). From a different aspect, Tu et al. provided evidences showing that the joint use of both pre-stimulus ongoing and post-stimulus evoked EEG/fMRI activities could significantly improve the performance of pain prediction compared to using just post-stimulus evoked brain responses. Both studies (Bai et al.; Tu et al.) shed new lights on the development of novel algorithms that could improve the prediction accuracy based on functional neuroimaging signals.

Taken together, this research topic provides a series of work in the interdisciplinary studies of vision, motor and pain. The biological findings and models of the topic could inspired the future studies both in biology and computer science.

AUTHOR CONTRIBUTIONS

HQ and LH are the organizers of the research topic "Editorial: Modeling of Visual Cognition, Body Sense, Motor Control and Their Integrations." For this editorial, we discussed and built the outline together. HQ was in charge of Section 1, 2 and 3 and

the figure. LH was in charge of Section 4, 5. Both of the authors commented on the manuscript.

REFERENCES

Hu, L., and Iannetti, G. D. (2016). Painful issues in pain prediction. *Trends Neurosci.* 39, 212–220. doi: 10.1016/j.tins.2016.01.004

Huang, G., Xiao, P., Hung, Y. S., Iannetti, G. D., Zhang, Z. G., and Hu, L. (2013). A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. *Neuroimage* 81, 283–293. doi: 10.1016/j.neuroimage.2013.05.017

An Eye in the Palm of Your Hand: Alterations in Visual Processing Near the Hand, a Mini-Review

Carolyn J. Perry [1,2,3]*, Prakash Amarasooriya [1,2] and Mazyar Fallah [1,2,3,4]

[1] Visual Perception and Attention Laboratory, York University, Toronto, ON, Canada, [2] Centre for Vision Research, York University, Toronto, ON, Canada, [3] School of Kinesiology and Health Science, York University, Toronto, ON, Canada, [4] Canadian Action and Perception Network, York University, Toronto, ON, Canada

Feedback within the oculomotor system improves visual processing at eye movement end points, also termed a visual grasp. We do not just view the world around us however, we also reach out and grab things with our hands. A growing body of literature suggests that visual processing in near-hand space is altered. The control systems for moving either the eyes or the hands rely on parallel networks of fronto-parietal regions, which have feedback connections to visual areas. Since the oculomotor system effects on visual processing occur through feedback, both through the motor plan and the motor efference copy, a parallel system where reaching and/or grasping motor-related activity also affects visual processing is likely. Areas in the posterior parietal cortex, for example, receive proprioceptive and visual information used to guide actions, as well as motor efference signals. This trio of information channels is all that would be necessary to produce spatial allocation of reach-related visual attention. We review evidence from behavioral and neurophysiological studies that support the hypothesis that feedback from the reaching and/or grasping motor control networks affects visual processing while noting ways in which it differs from that seen within the oculomotor system. We also suggest that object affordances may represent the neural mechanism through which certain object features are selected for preferential processing when stimuli are near the hand. Finally, we summarize the two effector-based feedback systems and discuss how having separate but parallel effector systems allows for efficient decoupling of eye and hand movements.

Keywords: attention, vision, sensorimotor integration, reaching and grasping, peripersonal space

INTRODUCTION

Accumulating behavioral evidence has shown that visual processing is altered near the hand. Speeded target detection and figure-ground assignment (Reed et al., 2006, 2010; Jackson et al., 2010), improvements in working memory (Tseng and Bridgeman, 2011), orientation processing (Craighero et al., 1999; Bekkering and Neggers, 2002; Hannus et al., 2005; Gutteling et al., 2011, 2013), target discrimination (Deubel et al., 1998), and in reaching and grasping precision (Brown et al., 2008), are just some of the effects seen when a reach places a hand near a visual stimulus. In addition, these alterations are seen whether the hand is nearby due to a sustained reach or if the hand is moved towards the visual stimulus during each trial in a more active manner. What remains a topic of debate is the mechanism by which these alterations in visual processing occur.

A number of studies suggest that visual processing near the hand is altered through spatial attention selection mechanisms (di Pellegrino and Frassinetti, 2000; Schendel and Robertson, 2004; Reed et al., 2006, 2010; Abrams et al., 2008). These studies have hypothesized that populations of fronto-parietal bimodal neurons underlie enhanced visual selection in near-hand space; however, these neurons are also thought to influence near-hand processing in the absence of spatial attention influences (Brown et al., 2008). More recently, enhanced magnocellular processing has been postulated as an alternative explanation for the near-hand effect (Gozli et al., 2012). For this review, we investigate the hypothesis that these effects are driven by a novel, effector specific, attentional selection mechanism that is different from either oculomotor-driven visual spatial or feature-based attention, and is mediated by feedback from fronto-parietal regions involved in reaching and grasping networks. We will first review the anatomical similarities between the oculomotor and the reaching/grasping networks, and provide evidence of feedback influences within the oculomotor system. We will then compare the neurophysiological alterations in visual processing near the hand to alterations in visual processing due to the oculomotor system and provide supporting evidence of feedback influences in the reaching and grasping system. We suggest that links between the visual system and the motor systems could drive enhanced processing of action-relevant object features, but that de-coupled eye and hand movements indicate the need for separate, effector-based selection mechanisms.

NEURAL CIRCUITRY

The reaching, grasping, and oculomotor systems all involve parallel networks of fronto-parietal areas (**Figure 1**). A dorsomedial stream, projecting from visual area V6 (Rizzolatti and Matelli, 2003; Passarelli et al., 2011), consisting of the medial intraparietal (MIP) area and area V6A in the superior parietal lobule (SPL), along with the dorsal premotor cortex (PMd) in the frontal lobe, which forms what is thought to be the neural network for reaching in the non-human primate (Caminiti et al., 1996; Culham et al., 2006; Filimon, 2010), with homologs in humans (Culham et al., 2006; Filimon, 2010). As with reaching, it has been suggested that there is a parallel dorsolateral circuit specialized for grasping (Fagg and Arbib, 1998; Luppino et al., 2001; Filimon, 2010) that projects from visual area MT/V5 (Rizzolatti and Matelli, 2003), and that this circuit is mainly dependent upon connections between the anterior intraparietal (AIP) region in the inferior parietal lobule (IPL) and the ventral premotor cortex (PMv), with homologous areas in humans (Fagg and Arbib, 1998; Culham et al., 2003, 2006; Frey et al., 2005). The reaching and grasping circuits however, appear to not be as completely functionally distinct as once thought as recent work has also found grasping related activity in the dorsomedial stream in non-human primate (Raos et al., 2003, 2004; Fattori et al., 2009, 2010, 2012) and human populations (Gallivan et al., 2011; Monaco et al., 2011). In fact, it has been suggested that the visual, somatosensory, and motor properties of V6A indicate a role for this area in the online error control
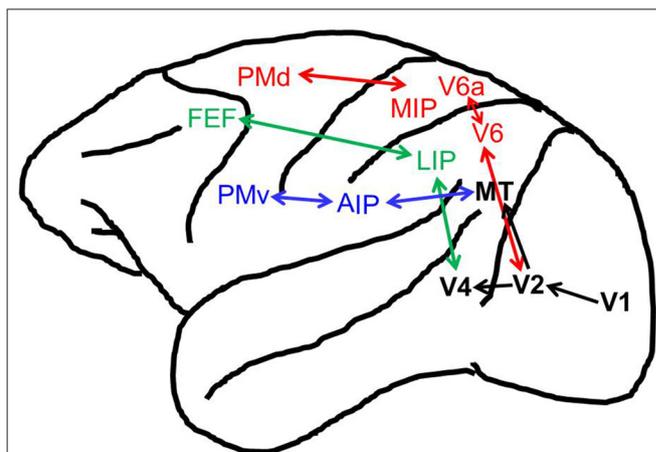
FIGURE 1 | Reach, grasp, and oculomotor control brain regions in the macaque. Shown are the cortical brain regions associated with the reach (in red), grasp (in blue), oculomotor (in green), and visual (in black) systems. Not pictured are anatomical cross-talk connections between the reaching and grasping networks (i.e., between V6A and anterior intraparietal (AIP)/Ventral premotor cortex (PMv), see Fattori et al., 2015).

for all of prehension, including reaching and grasping (Fattori et al., 2015). For movements of the eyes, the cortical oculomotor system in non-human primates and humans is comprised of the lateral intraparietal area (LIP)/parietal eye fields (PEF) and the frontal eye fields (FEF; Goldberg and Segraves, 1989; Bisley and Goldberg, 2003; Culham and Valyear, 2006; Culham et al., 2006). Due to the similarity between the anatomical components of these systems, we suggest that it is possible that oculomotor feedback mechanisms enhancing visual processing, could be replicated by the reaching and grasping networks to alter visual processing near the hand.

FEEDBACK IN THE OCULOMOTOR SYSTEM

The influence of feedback, from fronto-parietal motor related areas, on visual processing is already well-supported for the oculomotor system. Early psychophysical work established an indirect link between alterations in visual processing due to shifts in attention and saccade motor planning (Rizzolatti et al., 1987; Kowler et al., 1995; Sheliga et al., 1994; Deubel and Schneider, 1996; Kustov and Robinson, 1996; Nobre et al., 2000; Castet and Montagnini, 2006; van der Stigchel and Theeuwes, 2006; Baldauf and Deubel, 2008). In general, visual processing was improved when a visual target coincided with the endpoint of a planned saccade suggesting a close relationship between the oculomotor system and attention related changes in visual processing. These studies led to investigations that more causally associated activations of eye-movement related brain regions to shifts in spatial attention and consequently alterations in visual processing at the end points of planned saccades (Moore and Fallah, 2001, 2004; Moore and Armstrong, 2003; Müller et al., 2005; Neggers et al., 2007; Van Ettinger-Veenstra et al., 2009; Gutteling et al., 2010; Bosch et al., 2013). For example,

subthreshold microstimulation of the FEF resulted in increased visual sensitivity at the end-point of the unactivated motor plan behaviorally (Moore and Fallah, 2001, 2004) and within area V4 (Moore and Armstrong, 2003). This would suggest that recurrent connections between FEF and V4 allow for signals from FEF to feed back into the occipital lobe to influence subsequent visual processing (Armstrong et al., 2006; Armstrong and Moore, 2007; Ekstrom et al., 2008, 2009; Squire et al., 2012). Further evidence in primates comes from a study by Supèr et al. (2004) who found that in primary visual cortex neural activity corresponding to the location of the saccade target was enhanced approximately 100 ms before the onset of memory and visually-guided saccades. Studies in humans using transcranial magnetic stimulation (TMS) provide additional support for oculomotor feedback modulating visual processing. A single TMS pulse activates neurons in the targeted area. As such single pulse TMS over FEF enhances visual processing (Grosbras and Paus, 2003; Ruff et al., 2008; Van Ettinger-Veenstra et al., 2009) presumably by activating the feedback connections to visual processing areas. In contrast, a triple pulse disrupts the normal processing in an area. Triple pulse TMS used to disrupt the FEF results in impaired discrimination of a subsequently presented target (Neggers et al., 2007) suggesting that oculomotor feedback is necessary for spatial attention. Both the primate microstimulation studies and the human TMS studies support oculomotor feedback producing spatial attention effects behaviorally and within visual neurons. This would require attention signals to occur in the frontal lobe and propagate back to the occipital lobe. This is indeed what Van Ettinger-Veenstra et al. (2009) showed with EEG neuroimaging. They found that frontal activity associated with a saccade-go signal preceded activity in the occipital cortex associated with the appearance of a visual target. Thus, feedback projections from oculomotor-related frontal areas alter processing in posteriorly located visual areas.

VISUAL PROCESSING NEAR THE HAND

As mentioned previously, behavioral studies have provided indirect evidence suggesting that the space near the hand is prioritized. One prevailing theory suggests that alterations in visual processing occur as a result of attentional selection of near-hand space (di Pellegrino and Frassinetti, 2000; Schendel and Robertson, 2004; Reed et al., 2006, 2010; Abrams et al., 2008; Brown et al., 2008). Much like visual processing at the end point of a saccade is altered, the parallel within the reaching and grasping system would be a change in visual processing that occurs at the end point of a reach or grasp, i.e., in the workspace near the hand. One can imagine the benefit of this type of mechanism. This is especially true when reaching for an object while simultaneously viewing something in a different location that draws oculomotor driven spatial attention away from the object to be picked up. The underlying neural mechanisms that would drive altered visual processing near the hand have, as yet, not been well studied. A very recent neurophysiological study however, has shed light on the neural underpinnings of near-hand visual processing (Perry et al., 2015). Neuronal activity

was recorded from area V2 which is an area that is known to be selective for orientation (Motter, 1993), a feature important for reaching and grasping (Murata et al., 2000; Raos et al., 2004; Fattori et al., 2009), modulated by attention (Motter, 1993; Luck et al., 1997), and directly linked to fronto-parietal reaching and grasping areas (Gattass et al., 1997; Passarelli et al., 2011; Fattori et al., 2015). Instead of allocating classic visual spatial attention with a cue (Moran and Desimone, 1985; Motter, 1993; McAdams and Maunsell, 1999; Treue and Martinez-Trujillo, 1999), Perry et al. (2015) used the presence or absence of a nearby hand to determine the effects of near-hand attention on neuronal responses in area V2. Under these conditions, there was a significant increase in response at the preferred orientation when the hand was nearby. This is consistent with classic visual spatial studies which produce a "gain-modulation" of neuronal responses: responses are multiplied by the same factor regardless of selectivity (McAdams and Maunsell, 1999; Seidemann and Newsome, 1999; Treue and Martinez-Trujillo, 1999; McAdams and Reid, 2005). This results in a scaling of the tuning curve. However in contrast to gain modulation, there was no corresponding increase at the orthogonal orientation when the hand was near. Consequently, this produced a sharpening, instead of a scaling, of the orientation tuning curves when the hand was near, suggesting a different underlying mechanism than for oculomotor driven spatial attention. Sharpening of orientation tuning curves would result in greater orientation selectivity.

In addition to spatial attention, neuronal enhancement is also found with feature-based attention, where attending to a feature (such as a vertical bar) enhances processing of that specific feature (vertical), which aids greatly in visual search. Feature-based attention is described by the feature-similarity gain model of attention which predicts that enhancement of neuronal responses are strongest when the orientation of the grasp target (attended feature) and the orientation of the visual stimulus are matched, falling off as the difference in their orientations increased (Treue and Martinez-Trujillo, 1999). No such relationship was found. These results (Perry et al., 2015) suggest then that the attentional prioritization of near-hand space does not conform to known spatial or feature-based attentional mechanisms and that a novel, effector based, mechanism exists. This mechanism would preferentially process features (such as orientation) necessary for grasping, which would then improve the accuracy of an upcoming grasp.

EVIDENCE FOR FEEDBACK IN THE REACHING AND GRASPING SYSTEMS

While the effects of near-hand attention are seen in early visual areas, behaviorally these effects cannot be driven by the oculomotor system. The control system for near-hand attention, albeit separate from the oculomotor system, would likely be driven through the parallel feedback from fronto-parietal motor planning areas. It has been shown that neuronal response variability is reduced in premotor cortex during reaching (Churchland et al., 2010) and in the FEF during

oculomotor preparation (Purcell et al., 2012). Notably, neurons in V4 undergo a reduction in neuronal response variability prior to the onset of a saccade (Steinmetz and Moore, 2010). This suggests that reductions in oculomotor response variability propagate back to posteriorly located visual processing regions. If feedback from fronto-parietal reaching and grasping networks is the method through which neurons in V2 undergo alterations in their response properties (such as sharpened tuning—Perry et al., 2015), it would be expected that response variability would also be reduced. This is, in fact, what was found (Perry et al., 2015). Thus, both oculomotor and near-hand spatial attention rely on feedback projections which concomitantly reduce response variability.

In human populations, this premise of feedback connections mediating changes in visual response properties was tested by Gutteling et al. (2013). They investigated whether activation of the anterior portion of the intraparietal sulcus (aIPS) prior to a grasping or pointing movement improved orientation perception. aIPS has been shown to be part of a network of fronto-parietal areas that are involved in the control of grasping movements (Taira et al., 1990; Gallese et al., 1994; Sakata et al., 1995). Furthermore, aIPS has been shown to be selective for the orientation of the object to be grasped (Murata et al., 2000) and connected to occipital visual areas (Nakamura et al., 2001; Ruff et al., 2008; Blankenburg et al., 2010), including ventral stream regions (Borra et al., 2008) that would be sensitive to changes in orientation. Activation of aIPS during action preparation (Gutteling et al., 2013) improved orientation sensitivity, suggesting that aIPS is involved in modulating visual information during action planning. In addition, compared to pointing, grasping a 3-dimensional oriented bar, has been shown with electroencephalography to strengthen the N1 component and associated selection negativity in lateral occipital regions suggesting that the plan to grasp influences early ventral stream visual processing (orientation) of action-relevant features (Van Elk et al., 2010). Improved sensitivity and strengthened selection negativity is consistent with improved orientation tuning found in non-human primate V2 neurons when a hand is nearby (Perry et al., 2015).

Area V6A is another candidate area whose feedback could sharpen orientation tuning, as it has been found to be sensitive to the orientation of the wrist (Fattori et al., 2009), selective for grip type (Fattori et al., 2010), contains cells selective for orientation (Gamberini et al., 2011), and has direct connections to early visual processing areas (Passarelli et al., 2011). In addition, activity in V6A has been shown to be modulated by shifts in covert, oculomotor driven, spatial attention (Galletti et al., 2010), suggesting that it may play a similar role in hand driven attention.

Recurrent feedback loops between fronto-parietal and early visual processing areas (e.g., V2) would provide relevant corollary motor discharge information to enhance visual information relevant to reaching and grasping objects (i.e., sharpened orientation tuning) that would then update ongoing motor plans. As a movement progresses, sharpened orientation tuning information could be used to improve or correct hand shaping and wrist orientation resulting in improved reach and grasp accuracy. Given that V6A is thought to be involved in online error control of both reaching and grasping (Fattori et al., 2015), recurrent feedback loops between V2 and V6A are the likely candidate mechanism to underlie this process.

AFFORDANCES

Orientation is considered to be part of the processing that occurs in the ventral stream that results in object recognition. It is not thought to be necessary for processes in the dorsal stream that culminate in knowing where something is, for computations of complex motion of an object, or for execution of movement. Why then would orientation processing in V2 be improved simply because the hand is near? Close links between the visual and motor systems have been at the core of the *affordance* literature for years. Gibson (1979) suggested that one of the key functions of the visual system was to provide information to the motor system about the possible actions that could be implemented, or alternatively, the possible actions that the visual information *affords*. Since then, Tucker and Ellis (1998, 2001) and Ellis and Tucker (2001) have argued that the motor system itself could extract visually pertinent information that would produce affordances. In fact, they have used the term micro-affordances to refer to object properties that are action-relevant and could be used to inform subsequent movements to interact with the object of interest (Tucker and Ellis, 2001). Orientation is an object feature that informs the "graspability" of an object. For example, object orientation can either facilitate or impede response times depending on whether the object orientation produces a motor affordance (Tucker and Ellis, 1998). In other words, the orientation of an object informs the grasp that needs to be planned. Regions within the parietal lobe, integral to reaching and grasping movements, show selectivity for the size, shape and orientation of an object both during fixation and grasping movements (Taira et al., 1990; Gallese et al., 1994; Murata et al., 2000; Fattori et al., 2009, 2010, 2012; Breveglieri et al., 2015), suggesting these areas play a key role in the integration of visual and motor information and object affordances. Therefore, orientation is a feature necessary to grasp objects accurately and is processed within the fronto-parietal grasping network, especially within area AIP.

Even if there is not a representation of the object as a whole in the dorsal stream, the vision for action theory (Goodale and Milner, 1992; Goodale, 2008, 2013) would also suggest that there are features of an object that are action relevant and therefore worthy of preferential processing, or attentional selection, by the dorsal stream action system. Patients with visual agnosia, who can still scale and orient their hand to an object to be grasped in spite of being unable to recognize the object they are grasping, speak to this point (Goodale et al., 1991, 1994; Milner et al., 2012). Given that object features such as orientation have been shown to affect subsequent motor affordances, and that object properties are extracted to inform the scale and orientation of the hand in patients who cannot recognize objects, it logically follows that orientation be an object feature preferentially processed within

the dorsal stream in parallel to its processing within the ventral stream for object recognition.

ADVANTAGES OF SEPARATE EFFECTOR MECHANISMS

Being able to separate the deployment of attention between effectors allows for the decoupling of actions. Many examples exist of instances where we reach for one thing while looking elsewhere. In fact, optic ataxia, in which there is an inability to reach to peripheral targets, results from damage to the posterior parietal cortex (Milner and Goodale, 1995; Carey et al., 1997; Jackson et al., 2005). It has been shown that reaching to centrally located targets activates the MIP sulcus and PMd, while reaching to peripherally located targets additionally activates the parietal occipital junction and more rostral parts of PMd. These differentiated networks support dissociation between where gaze and grasp are deployed (Prado et al., 2005). Furthermore, recent work has shown that when a sequence of reaching movements are planned, visual discrimination is significantly enhanced not just at the first movement goal but also at the second (Baldauf et al., 2006; Baldauf and Deubel, 2008, 2009). So while an eye movement would be planned and then executed to the first target, the second is already enhanced suggesting that reach execution is separate from oculomotor planning and in turn, that movement planning and execution in the posterior parietal cortex already accommodates separate representations of gaze and reach targets (Jackson et al., 2009). These decoupled eye and hand movements are supported by the presence of neuronal populations in parietal areas that produce multiple types of reference frame transformations to encode targets in eye-centered or hand-/body-centered frames of reference (Lacquaniti et al., 1995; Batista et al., 1999, 2007; Buneo et al., 2002, 2008; Cohen and Andersen, 2002; Marzocchi et al., 2008; Chang et al., 2009; Chang and Snyder, 2010; McGuire and Sabes, 2011). As populations

encoding targets in either eye- or hand-centered reference frames support decoupled movements, it follows then that there should exist separate effector-based attentional mechanisms.

CONCLUSION

We have reviewed literature in support of the hypothesis that there exist parallel, but separate, effector-based attentional systems. Whereas the oculomotor system enhances visual responses through gain modulation, near-hand attention sharpens orientation tuning and, potentially, other features relevant to reaching and grasping. Thus, these effector-based systems may be specialized for the actions those effectors can perform. We suggest that improved orientation processing is a feature important for accurate reaching and grasping, and that separate effector-based attentional mechanisms allow for the decoupling of visual enhancements associated with eye and hand movements. Future investigations are needed to further support this hypothesis for example, by systematically testing grasp-relevant and irrelevant features. In addition, testing whether both the reaching and grasping or grasping alone is involved in near-hand attention which will provide details regarding which fronto-parietal networks may be involved and what other object features may be preferentially processed.

AUTHOR CONTRIBUTIONS

CJP, PA, and MF all contributed to the writing and revision of this article.

ACKNOWLEDGMENTS

REFERENCES

Abrams, R. A., Davoli, C. C., Du, F., Knapp, W. H., and Paull, D. (2008). Altered vision near the hands. *Cognition* 107, 1035–1047. doi: 10.1016/j.cognition.2007.09.006

Armstrong, K. M., Fitzgerald, J. K., and Moore, T. (2006). Changes in the visual receptive fields with microstimulation of the frontal cortex. *Neuron* 50, 791–798. doi: 10.1016/j.neuron.2006.05.010

Armstrong, K. M., and Moore, T. (2007). Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field. *Proc. Natl. Acad. Sci. USA* 104, 9499–9504. doi: 10.1073/pnas.0701104104

Baldauf, D., and Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Exp. Brain Res.* 184, 411–425. doi: 10.1007/s00221-007-1114-x

Baldauf, D., and Deubel, H. (2009). Attentional selection of multiple goal positions before rapid hand movement sequences: An ERP study. *J. Cogn. Neurosci.* 21, 18–29. doi: 10.1162/jocn.2008.21021

Baldauf, D., Wolf, M., and Deubel, H. (2006). Deployment of visual attention before sequences of goal-directed hand movements. *Vision Res.* 46, 4355–4374. doi: 10.1016/j.visres.2006.08.021

Batista, A. P., Buneo, C. A., Snyder, L. H., and Andersen, R. A. (1999). Reach plans in eye-centered coordinates. *Science* 285, 257–260. doi: 10.1126/science.285.5425.257

Batista, A. P., Santhanam, G., Yu, B. M., Ryu, S. I., Afshar, A., and Shenoy, K. V. (2007). Reference frames for reach planning in macaque dorsal premotor cortex. *J. Neurophysiol.* 98, 966–983. doi: 10.1152/jn.00421.2006

Bekkering, H., and Neggers, F. S. W. (2002). Visual search is modulated by action intentions. *Psychol. Sci.* 13, 370–374. doi: 10.1111/j.0956-7976.2002.00466.x

Bisley, J. W., and Goldberg, M. E. (2003). The role of the parietal cortex in the neural processing of saccadic eye movements. *Adv. Neurol.* 93, 141–157.

Blankenburg, F., Ruff, C. C., Bestmann, S., Bjoertomt, O., Josephs, O., Deichmann, R., et al. (2010). Studying the role of human parietal cortex in visuospatial attention with concurrent TMS-fMRI. *Cereb. Cortex* 20, 2702–2711. doi: 10.1093/cercor/bhq015

Borra, E., Belmalih, A., Calzavara, R., Gerbella, M., Murata, A., Rozzi, S., et al. (2008). Cortical connections of the macaque anterior intraparietal (AIP) area. *Cereb. Cortex* 18, 1094–1111. doi: 10.1093/cercor/bhm146

Bosch, S. E., Neggers, S. F. W., and van der Stigchel, S. (2013). The role of the frontal eye fields in oculomotor competition: image-guided TMS enhances contralateral target selection. *Cereb. Cortex* 23, 824–832. doi: 10.1093/cercor/bhs075

Breveglieri, R., Galletti, C., Bosco, A., Gamberini, M., and Fattori, P. (2015). Object affordance modulates visual responses in the macaque medial posterior parietal cortex. *J. Cogn. Neurosci.* 27, 1447–1455. doi: 10.1162/jocn_a_00793

Brown, L. E., Kroliczak, G., Demonet, J. F., and Goodale, M. A. (2008). A hand in blindsight: hand placement near target improves size perception in the blind

visual field. *Neuropsychologia* 46, 786–802. doi: 10.1016/j.neuropsychologia.2007.10.006

Buneo, C. A., Batista, A. P., Jarvis, M. R., and Andersen, R. A. (2008). Time-invariant reference frames for parietal reach activity. *Exp. Brain Res.* 188, 77–89. doi: 10.1007/s00221-008-1340-x

Buneo, C. A., Jarvis, M. R., Batista, A. P., and Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature* 416, 632–636. doi: 10.1038/416632a

Caminiti, R., Ferraina, S., and Johnson, P. B. (1996). The sources of visual information to the primate frontal lobe: a novel role for the superior parietal lobule. *Cereb. Cortex* 6, 319–328. doi: 10.1093/cercor/6.3.319

Carey, D. P., Coleman, R. J., and Della Sala, S. (1997). Magnetic misreaching. *Cortex* 33, 639–652. doi: 10.1016/s0010-9452(08)70722-6

Castet, J. S., and Montagnini, A. (2006). Dynamics of attentional deployment during saccadic programming. *J. Vis.* 6, 196–212. doi: 10.1167/6.3.2

Chang, S. W., Papadimitriou, C., and Snyder, L. H. (2009). Using a compound field to compute a reach plan. *Neuron* 64, 744–755. doi: 10.1016/j.neuron.2009.11.005

Chang, S. W., and Snyder, L. H. (2010). Idiosyncratic and systematic aspects of spatial representations in the macaque parietal cortex. *Proc. Natl. Acad. Sci. U S A* 107, 7951–7956. doi: 10.1073/pnas.0913209107

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378. doi: 10.1038/nn.2501

Cohen, Y. E., and Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nat. Rev. Neurosci.* 3, 553–562. doi: 10.1038/nrn873

Craighero, L., Fadiga, L., Rizzolatti, G., and Umiltà, C. (1999). Action for perception: a motor-visual attentional effect. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1673–1692. doi: 10.1037/0096-1523.25.6.1673

Culham, J. C., Cavina-Pratesi, C., and Singhal, A. (2006). The role of parietal cortex in visuomotor control: what have we learned from neuroimaging? *Neuropsychologia* 44, 2668–2684. doi: 10.1016/j.neuropsychologia.2005.11.003

Culham, J. C., Danckert, S. L., DeSouza, J. F., Gati, J. S., Menon, R. S., and Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp. Brain Res.* 153, 180–189. doi: 10.1007/s00221-003-1591-5

Culham, J. C., and Valyear, K. F. (2006). Human parietal cortex in action. *Curr. Opin. Neurobiol.* 16, 205–212. doi: 10.1016/j.conb.2006.03.005

Deubel, H., and Schneider, W. X. (1996). Saccade target selection and object recognition-evidence for a common attentional mechanism. *Vision Res.* 36, 1827–1837. doi: 10.1016/0042-6989(95)00294-4

Deubel, H., Schneider, W. X., and Paprotta, I. (1998). Selective dorsal and ventral processing: Evidence for a common attentional mechanism in reaching and perception. *Vis. Cogn.* 5, 81–107. doi: 10.1080/713756776

di Pellegrino, G., and Frassinetti, F. (2000). Direct evidence from parietal extinction of enhancement of visual attention near a visible hand. *Curr. Biol.* 10, 1475–1477. doi: 10.1016/s0960-9822(00)00809-5

Ekstrom, L. B., Roelfsema, P. R., Arsenault, J. T., Bonmassar, G., and Vanduffel, W. (2008). Bottom-up dependent gating of frontal signals in early visual cortex. *Science* 321, 414–417. doi: 10.1126/science.1153276

Ekstrom, L. B., Roelfsema, P. R., Arsenault, J. T., Kolster, H., and Vanduffel, W. (2009). Modulation of the contrast response function by electrical microstimulation of the macaque frontal eye field. *J. Neurosci.* 29, 10683–10694. doi: 10.1523/jneurosci.0673-09.2009

Ellis, R., and Tucker, M. (2001). Micro-affordance: The potentiation of components of action by seen objects. *Br. J. Psychol.* 91, 451–471. doi: 10.1348/000712600161934

Fagg, A. H., and Arbib, M. A. (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Netw.* 11, 1277–1303. doi: 10.1016/s0893-6080(98)00047-1

Fattori, P., Breveglieri, R., Bosco, A., Gamberini, M., and Galletti, C. (2015). Vision for prehension in the medial parietal cortex. *Cereb. Cortex* [Epub ahead of print] doi: 10.1093/cercor/bhv302

Fattori, P., Breveglieri, R., Marzocchi, N., Filippini, D., Bosco, A., and Galletti, C. (2009). Hand orientation during reach-to-grasp movements modulates

neuronal activity in the medial posterior parietal area V6A. *J. Neurosci.* 29, 1928–1936. doi: 10.1523/jneurosci.4998-08.2009

Fattori, P., Breveglieri, R., Raos, V., Bosco, A., and Galletti, C. (2012). Vision for action in the macaque medial posterior parietal cortex. *J. Neurosci.* 32, 3221–3234. doi: 10.1523/jneurosci.5358-11.2012

Fattori, P., Raos, V., Breveglieri, R., Bosco, A., Marzocchi, N., and Galletti, C. (2010). The dorsomedial pathway is not just for reaching: Grasping neurons in the medial parieto-occipital cortex of the macaque monkey. *J. Neurosci.* 30, 342–349. doi: 10.1523/jneurosci.3800-09.2010

Filimon, F. (2010). Human cortical control of hand movements: parieto-frontal networks for reaching, grasping and pointing. *Neuroscientist* 16, 388–407. doi: 10.1177/1073858410375468

Frey, S. H., Vinton, D., Norlund, R., and Grafton, S. T. (2005). Cortical topography of human anterior intraparietal cortex active during visually guided grasping. *Brain Res. Cogn. Brain Res.* 23, 397–405. doi: 10.1016/j.cogbrainres.2004.11.010

Gallese, V., Murata, A., Kaseda, M., Niki, N., and Sakata, H. (1994). Deficit of hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport* 5, 1525–1529. doi: 10.1097/00001756-199407000-00029

Galletti, C., Breveglieri, R., Lappe, M., Bosco, A., Ciavarro, M., and Fattori, P. (2010). Covert shift of attention modulates the ongoing neural activity in a reaching area of the macaque dorsomedial visual stream. *PLoS One* 5:e15078. doi: 10.1371/journal.pone.0015078

Gallivan, J. P., McLean, D. A., Valyear, K. F., Pettypiece, C. E., and Culham, J. C. (2011). Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *J. Neurosci.* 31, 9599–9610. doi: 10.1523/jneurosci.0080-11.2011

Gamberini, M., Galletti, C., Bosco, A., Breveglieri, R., and Fattori, P. (2011). Is the medial posterior parietal area V6A a single functional area? *J. Neurosci.* 31, 5145–5157. doi: 10.3410/f.11420956.12442054

Gattass, R., Sousa, A. P., Mishkin, M., and Ungerleider, L. G. (1997). Cortical projections of area V2 in the macaque. *Cereb. Cortex* 7, 110–129. doi: 10.1093/cercor/7.2.110

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Goldberg, M. E., and Segraves, M. A. (1989). "The visual and frontal cortices.", in *The Neurobiology of Saccadic Eye Movements, Reviews of Oculomotor Research Series Vol. 3, (Amsterdam: Elsevier)* eds Goldberg M. E. and Wurtz R. H. 283–313.

Goodale, M. A. (2008). Action without perception in human vision. *Cogn. Neuropsychol.* 25, 891–919. doi: 10.1080/02643290801961984

Goodale, M. A. (2013). Separate visual systems for perception and action: a framework for understanding cortical visual impairment. *Dev. Med. Child Neurol.* 55, 9–12. doi: 10.1111/dmcn.12299

Goodale, M. A., Meenan, J. P., Bulthoff, H. H., Nicolle, D. A., Murphy, K. J., and Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Curr. Biol.* 4, 604–610. doi: 10.1016/s0960-9822(00)00132-9

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.

Goodale, M. A., Milner, A. D., Jakobson, L. S., and Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature* 349, 154–156.

Gozli, D. G., West, G. L., and Pratt, J. (2012). Hand position alters vision by biasing processing through different visual pathways. *Cognition* 124, 244–250. doi: 10.1016/j.cognition.2012.04.008

Grosbras, M. H., and Paus, T. (2003). Transcranial magnetic stimulation of the human frontal eye field facilitates visual awareness. *Eur. J. Neurosci.* 18, 3121–3126. doi: 10.1111/j.1460-9568.2003.03055.x

Gutteling, T. P., Kenemans, J. L., and Neggers, S. F. (2011). Grasping preparation enhances orientation change detection. *PLoS One* 6:e17675. doi: 10.1371/journal.pone.0017675

Gutteling, T. P., Park, S. Y., Kenemans, J. L., and Neggers, S. F. W. (2013). TMS of the anterior intraparietal area selectively modulates orientation change detection during action preparation. *J. Neurophysiol.* 110, 33–41. doi: 10.1152/jn.00622.2012

Gutteling, T. P., van Ettinger-Veenstra, H. M., Kenemans, J. L., and Neggers, S. F. W. (2010). Lateralized frontal eye field activity precedes occipital activity shortly before saccades: evidence for cortico-cortical feedback as a

mechanism underlying covert attention shifts. *J. Cogn. Neurosci.* 22, 1931–1943. doi: 10.1162/jocn.2009.21342

Hannus, A., Cornelissen, F. W., Lindemann, O., and Bekkering, H. (2005). Selection-for-action in visual search. *Acta. Psychol.* 118, 171–191. doi: 10.1016/j.actpsy.2004.10.010

Jackson, C. P., Miall, R. C., and Balslev, D. (2010). Spatially valid proprioceptive cues improve the detection of visual stimulus. *Exp. Brain Res.* 205, 31–40. doi: 10.1007/s00221-010-2330-3

Jackson, S. R., Newport, R., Husain, M., Fowlie, J. E., O'Donoghue, M., and Bajaj, N. (2009). There may be more to reaching than meets the eye: re-thinking optic ataxia. *Neuropsychologia* 47, 1397–1408. doi: 10.1016/j.neuropsychologia.2009.01.035

Jackson, S. R., Newport, R., Mort, D., and Husain, M. (2005). Where the eye looks, the hand follows; limb-dependent magnetic misreaching in optic ataxia. *Curr. Biol.* 15, 42–46. doi: 10.1016/j.cub.2004.12.063

Kowler, E., Anderson, E., Dosher, B., and Blaser, E. (1995). The role of attention in the programming of saccades. *Vis. Res.* 35, 1897–1916. doi: 10.1016/0042-6989(94)00279-u

Kustov, A. A., and Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature* 384, 74–77. doi: 10.1038/384074a0

Lacquaniti, F., Guigon, E., Bianchi, L., Ferraina, S., and Caminiti, R. (1995). Representing spatial information for limb movement: role of area 5 in the monkey. *Cereb. Cortex* 5, 391–409. doi: 10.1093/cercor/5.5.391

Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2 and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42.

Luppino, G., Calzavara, R., Rozzi, S., and Matelli, M. (2001). Projections from the superior agranular frontal cortex in the temporal sulcus to the macaque. *Eur. J. Neurosci.* 14, 1035–1040. doi: 10.1046/j.0953-816x.2001.01734.x

Marzocchi, N., Breveglieri, R., Galletti, C., and Fattori, P. (2008). Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movement? *Eur. J. Neurosci.* 27, 775–789. doi: 10.1111/j.1460-9568.2008.06021.x

McAdams, C. J., and Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19, 431–441.

McAdams, C. J., and Reid, R. C. (2005). Attention modulates the responses of simple cells in monkey primary visual cortex. *J. Neurosci.* 25, 11023–11033. doi: 10.1523/jneurosci.2904-05.2005

McGuire, L. M., and Sabes, P. M. (2011). Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *J. Neurosci.* 31, 6661–6673. doi: 10.1523/jneurosci.2921-10.2011

Milner, A. D., Ganel, T., and Goodale, M. A. (2012). Does grasping in patient D.F. depend on vision? *Trends Cogn. Sci.* 16, 256–257. doi: 10.1016/j.tics.2012.03.004

Milner, A. D., and Goodale, M. A. (1995). *The Visual Brain in Action.* New York: Oxford.

Monaco, S., Cavina-Pratesi, C., Sedda, A., Fattori, P., Galletti, C., and Culham, J. C. (2011). Functional magnetic resonance adaptation reveals the involvement of the dorsomedial stream in hand orientation for grasping. *J. Neurophysiol.* 106, 2248–2263. doi: 10.1152/jn.01069.2010

Moore, T., and Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421, 370–373. doi: 10.1038/nature01341

Moore, T., and Fallah, M. (2001). Control of eye movements and spatial attention. *Proc. Natl. Acad. Sci. USA* 98, 1273–1276. doi: 10.1073/pnas.021549498

Moore, T., and Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *J. Neurophysiol.* 91, 152–162. doi: 10.1152/jn.00741.2002

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science* 229, 782–784. doi: 10.1126/science.4023713

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2 and V4 in the presence of competing stimuli. *J. Neurophysiol.* 70, 909–919.

Müller, J. R., Philiastides, M. G., and Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proc. Natl. Acad. Sci. USA* 102, 524–529. doi: 10.1073/pnas.0408311101

Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. (2000). Selectivity for the shape, size and orientation of objects for grasping in neurons of monkey parietal area AIP. *J. Neurophysiol.* 83, 2580–2601.

Nakamura, H., Kuroda, T., Wakita, M., Kusunoki, M., Kato, A., Mikami, A., et al. (2001). From three-dimensional space vision to prehensile hand movements: the lateral intraparietal area links the area V3A and the anterior intraparietal area in macaques. *J. Neurosci.* 21, 8174–8187.

Neggers, S. F., Huijbers, W., Vrijlandt, C. M., Vlaskamp, B. N., Schutter, D. J., and Kenemans, J. L. (2007). TMS pulses on the frontal eye fields break coupling between visuospatial attention and eye movements. *J. Neurophysiol.* 98, 2765–2778. doi: 10.1152/jn.00357.2007

Nobre, A. C., Sebestyen, G. N., and Miniussi, M. (2000). The dynamics of shifting visuospatial attention revealed by event-related potentials. *Neuropsychologia* 38, 964–974. doi: 10.1016/s0028-3932(00)00015-4

Passarelli, L., Rosa, M. G., Gamberini, M., Bakola, S., Burman, K. J., Fattori, P., et al. (2011). Cortical connections of area V6Av in the macaque: a visual-input node to the eye/hand coordination system. *J. Neurosci.* 31, 1790–1801. doi: 10.1523/jneurosci.4784-10.2011

Perry, C. J., Sergio, L. E., Crawford, J. D., and Fallah, M. (2015). Hand placement near the visual stimulus improves orientation selectivity in V2 neurons. *J. Neurophysiol.* 113, 2859–2870. doi: 10.1152/jn.00919.2013

Prado, J., Clavagnier, S., Otzenberger, H., Scheiber, C., Kennedy, H., and Perenin, M. T. (2005). Two cortical systems for reaching in central and peripheral vision. *Neuron* 48, 849–858. doi: 10.1016/j.neuron.2005.10.010

Purcell, B. A., Heitz, R. P., Cohen, J. Y., and Schall, J. D. (2012). Response variability of frontal eye field neurons modulates with sensory input and saccade preparation but not visual search salience. *J. Neurophysiol.* 108, 2737–2750. doi: 10.1152/jn.00613.2012

Raos, V., Franchi, G., Gallese, V., and Fogassi, L. (2003). Somatotopic organization of the lateral part of area F2 (dorsal premotor cortex) of the macaque monkey. *J. Neurophysiol.* 89, 1503–1518. doi: 10.1152/jn.00661.2002

Raos, V., Umiltá, M. A., Gallese, V., and Fogassi, L. (2004). Functional properties of grasping- related neurons in the dorsal premotor area F2 of the macaque monkey. *J. Neurophysiol.* 92, 1990–2002. doi: 10.1152/jn.00154.2004

Reed, C. L., Betz, R., Garza, J. P., and Roberts, R. J. (2010). Grab it! Biased attention in functional hand tool space. *Atten. Percept. Psychophys.* 72, 236–245. doi: 10.3758/app.72.1.236

Reed, C. L., Grubb, J. D., and Steele, C. (2006). Hands up: attentional prioritization of space near the hand. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 166–177. doi: 10.1037/0096-1523.32.1.166

Rizzolatti, G., and Matelli, M. (2003). Two different streams from the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157. doi: 10.1007/s00221-003-1588-0

Rizzolatti, G., Riggio, L., Dascola, I., and Umilta, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia* 25, 31–40. doi: 10.1016/0028-3932(87)90041-8

Ruff, C. C., Bestmann, S., Blankenburg, F., Bjoertomt, O., Josephs, O., Weiskopf, N., et al. (2008). Distinct causal influences of parietal versus frontal areas on human visual cortex: evidence from concurrent TMS-fMRI. *Cereb. Cortex* 18, 817–827. doi: 10.1093/cercor/bhm128

Sakata, H., Taira, M., Murata, A., and Mine, S. (1995). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cereb. Cortex* 5, 429–438. doi: 10.1093/cercor/5.5.429

Schendel, K., and Robertson, L. C. (2004). Reaching out to see: arm position can attenuate human visual loss. *J. Cogn. Neurosci.* 16, 935–943. doi: 10.1162/0898929041502698

Seidemann, E., and Newsome, W. T. (1999). Effect of spatial attention on the responses of area MT neurons. *J. Neurophysiol.* 81, 1783–1794.

Sheliga, B. M., Riggio, L., and Rizzolatti, G. (1994). Orienting of attention and eye movements. *Exp. Brain Res.* 98, 507–522.

Squire, R. F., Steinmetz, N. A., and Moore, T. (2012). Frontal eye field. *Scholarpedia* 7:5341. doi: 10.4249/scholarpedia.5341

Steinmetz, N. A., and Moore, T. (2010). Changes in the response rate and response variability of area V4 neurons during the preparation of saccadic eye movements. *J. Neurophysiol.* 103, 1171–1178. doi: 10.1152/jn.00689.2009

Supèr, H., van der Togt, C., Spekreijse, H., and Lamme, V. A. (2004). Correspondence of presaccadic activity in the monkey primary visual cortex

with saccadic eye movements. *Proc. Natl. Acad. Sci. USA* 101, 3230–3235. doi: 10.1073/pnas.0400433101

Taira, M., Mine, S., Georgopoulos, A. P., Murata, A., and Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Exp. Brain Res.* 83, 29–36.

Treue, S., and Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 575–579.

Tseng, P., and Bridgeman, B. (2011). Improved change detection with nearby hands. *Exp. Brain Res.* 209, 257–269. doi: 10.1007/s00221-011-2544-z

Tucker, M., and Ellis, R. (1998). On the relations between seen objects and components of potential actions. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 830–846. doi: 10.1037/0096-1523.24.3.830

Tucker, M., and Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Vis. Cogn.* 8, 769–800. doi: 10.1080/13506280042000144

van der Stigchel, S., and Theeuwes, J. (2006). Our eyes deviate away from a location where a distractor is expected to appear. *Exp. Brain Res.* 169, 338–349. doi: 10.1007/s00221-005-0147-2

Van Elk, M., Van Schie, H. T., Neggers, S. F., and Bekkering, H. (2010). Neural and temporal dynamics underlying visual selection for action. *J. Neurophysiol.* 104, 972–983. doi: 10.1152/jn.01079.2009

Van Ettinger-Veenstra, H. M., Huijbers, W., Gutteling, T. P., Vink, M., Kenemans, J. L., and Neggers, S. F. (2009). fMRI-guided TMS on cortical eye fields: the frontal but not intraparietal eye fields regulate the coupling between visuospatial attention and eye movements. *J. Neurophysiol.* 102, 3469–3480. doi: 10.1152/jn.00350.2009

CrossMark

The Development of Hand-Centered Visual Representations in the Primate Brain: A Computer Modeling Study Using Natural Visual Scenes

Juan M. Galeazzi*, Loredana Minini and Simon M. Stringer

Department of Experimental Psychology, Oxford Centre for Theoretical Neuroscience and Artificial Intelligence, University of Oxford, Oxford, UK

Neurons that respond to visual targets in a hand-centered frame of reference have been found within various areas of the primate brain. We investigate how hand-centered visual representations may develop in a neural network model of the primate visual system called VisNet, when the model is trained on images of the hand seen against natural visual scenes. The simulations show how such neurons may develop through a biologically plausible process of unsupervised competitive learning and self-organization. In an advance on our previous work, the visual scenes consisted of multiple targets presented simultaneously with respect to the hand. Three experiments are presented. First, VisNet was trained with computerized images consisting of a realistic image of a hand and a variety of natural objects, presented in different textured backgrounds during training. The network was then tested with just one textured object near the hand in order to verify if the output cells were capable of building hand-centered representations with a single localized receptive field. We explain the underlying principles of the *statistical decoupling* that allows the output cells of the network to develop single localized receptive fields even when the network is trained with multiple objects. In a second simulation we examined how some of the cells with hand-centered receptive fields decreased their shape selectivity and started responding to a localized region of hand-centered space as the number of objects presented in overlapping locations during training increases. Lastly, we explored the same learning principles training the network with natural visual scenes collected by volunteers. These results provide an important step in showing how single, localized, hand-centered receptive fields could emerge under more ecologically realistic visual training conditions.

Keywords: hand-centered, neural networks, self-organization, reference frames, posterior parietal cortex, area 5d, premotor

1. INTRODUCTION

The brain seems to represent the location of objects in space using a variety of coordinate systems. Consistent with this, several neurophysiological recordings have reported neurons encoding the location of visual targets in different frames of reference. Visual targets are represented initially in a retinocentric or eye-centered frame of reference and in later stages of processing this information is recoded into more abstract, non-retinal coordinate maps that are more suitable to

guide our behavior. For example, head-centered, body-centered, hand-centered as well as mixed representations have been reported in different parts of the posterior parietal cortex and adjacent areas (Andersen et al., 1985; Brotchie et al., 1995; Buneo et al., 2002; Pesaran et al., 2006; Bremner and Andersen, 2012).

Similarly, a number of electrophysiological recordings in macaques have also reported neurons with localized and selective responses to stimuli shown in localized regions near the body or parts of the body (i.e., peri-personal space and peri-hand space; Hyvärinen and Poranen, 1974; Rizzolatti et al., 1981, 1988; Graziano and Gross, 1993; Graziano et al., 1994, 1997; Fogassi et al., 1996, 1999; Graziano and Gross, 1998; Graziano, 1999). The visual responding regions of these cells seem to extend from the skin and could be found anchored to different parts of the body (e.g., around the hand, mouth and face). Their response properties do not seem to change with eye movements and the target does not have to necessarily touch the skin to elicit a response.

Cells representing the location of visual targets in hand-centered coordinates have been reported in multiple areas, mostly in the parietal cortex and premotor areas. For planning reach vectors, hand-centered coordinates seem to be the dominant representation in area 5d (Buneo and Andersen, 2006; Bremner and Andersen, 2012). Other hand-centered receptive fields have been found also in ventral premotor areas (Graziano et al., 1997; Graziano, 1999). These cells fire maximally to the location of the target relative to the hand, irrespective of where on the retina this fixed spatial configuration appears. A number of neurophysiological and behavioral studies with human subjects have similarly shown evidence of hand-centered encoding of the location of visual objects near the hands (peri-hand space) in parietal and premotor areas (Makin et al., 2009, 2007; Brozzoli et al., 2011, 2012; Gentile et al., 2011).

Different theoretical approaches have been proposed to reflect the different stages of coordinate transformations and explain some of the response properties found in some neurons of the PPC and premotor areas. A variety of neural network models have been suggested to account for the development of these supra-retinal representations (e.g., head-centered, hand-centered; Zipser and Andersen, 1988; Pouget and Sejnowski, 1997; Blohm et al., 2009; Chang et al., 2009). Some of these models have focused on the development of head-centered responses and despite the computational advantages of these different theoretical efforts, most of this work has been based on supervised learning algorithms, which cannot provide a biologically plausible account of how these properties develop in the cortex. Other computational approaches have suggested a different way of implementing these transformations using neurons behaving like basis function units that could provide an immediate read-out of multiple frames of reference (Pouget and Sejnowski, 1997).

A self-organizing hypothesis to account for how hand-centered representations could occur has been recently proposed (Galeazzi et al., 2013). Here, it was suggested that while the eyes are exploring a visual scene involving a target object in a fixed position with respect to the hand, a form of trace learning would allow the network to associate different views of the same

hand-object spatial configuration. This hypothesis was tested using a biologically plausible neural network model, VisNet, of the primate visual system. The architecture of VisNet consisted of a hierarchy of competitive neural layers, with unsupervised learning taking place in the feedforward connections between the layers. These simulation results showed how output cells could learn to respond selectively to the location of targets with respect to the hand, irrespective of where on the retina this spatial configuration was shown.

The simulations presented previously by our laboratory (Galeazzi et al., 2013) involved showing only a hand and single circular object at any one time during training. However, in the real world we rarely encounter one object at the time. In fact, our visual system is mostly confronted with a complex environment consisting of multiple objects. Moreover, in real-world visual scenes the various objects that we encounter throughout our sensory-motor experiences have different shapes and sizes. Nevertheless, cells in the dorsal visual system seem to be able to generalize and form delineated hand-centered visual receptive fields. In this paper we explore whether our model would still be able to develop output cells with single, localized, hand-centered receptive fields when the network is exposed to more realistic images. In the initial simulations presented in Experiments 1 and 2, the training images were comprised of a variety of everyday objects presented simultaneously around a realistic hand. In Experiment 3, we increased the realism further by presenting the hand against a range of completely natural backgrounds during training.

Early research with VisNet (Stringer and Rolls, 2000) has revealed the difficulty for the network to build transform (e.g., position) invariant representations of individual objects when it is trained on cluttered backgrounds. How could the network develop neurons that respond selectively to a single object when it is trained with cluttered images always containing more that one object at a time? Later work has shown that VisNet can in fact form representations of individual objects even when they are never seen in isolation during training (Stringer et al., 2007; Stringer and Rolls, 2008). The statistical decoupling between the different objects works when there is a sufficiently large number of objects and the network is presented with many different combinations of these objects during training. Any particular combination of objects will be seen together only rarely which prevents individual neurons in the output layer from learning to respond to the particular combinations of objects seen during training. Instead, the neurons are forced to learn to respond to the individual objects themselves. The fundamental principle is that competitive learning binds together the features that are seen more often than other less frequent combinations of features in the environment. Thus, the network does not need any prior knowledge of which features belong to a particular object; it self-organizes by learning to respond to the combinations of features that co-occur the most.

We hypothesized that a similar mechanism of statistical decoupling may produce visual neurons that have learned to respond to single object locations in a hand-centered frame of reference. Let us assume that during training the network model is exposed to many images containing the hand with

multiple other objects, but where the objects occur in different combinations of hand-centered locations in the different images. Because the objects are always seen with the hand, this forces each of the output neurons to learn to respond to some combination of the hand and hand-centered object locations. However, over many different images there will be a relatively weak statistical link between any two particular hand-centered object locations. These statistics will drive the development of output neurons that have learned to respond to particular spatial configurations of the hand and a single object. That is, these neurons will respond to the presence of an object in only one localized hand-centered receptive field.

To test this learning hypothesis and increase the ecological plausibility of our simulations, three experiments are presented. We first show how the model can develop hand-centered representations using more realistic training images composed of the hand with pairs of objects presented in different hand-centered locations. Many images with different combinations of hand-centered locations were used to ensure adequate statistical decoupling between the different object locations. In a second experiment, we explored whether the output cells of our model developed hand-centered receptive fields that were also somewhat selective to the shape of the object, as well as evaluating how this shape selectivity is affected as the network is trained with more objects. Lastly, in the third experiment we explore whether the network could still develop localized hand-centered receptive fields when the hand is shown against a large collection of different natural background scenes during training. In this case, the background scenes used were entirely natural with no careful control of what objects were present and where they were located.

2. MATERIALS AND METHODS

2.1. VisNet Model

The experiments presented in this paper were conducted using the VisNet model of the primate visual system (**Figure 1**). VisNet is composed of four feedforward layers of competitive neural networks. Each neuronal layer incorporates lateral competition between neurons which is implemented by local graded inhibition. The synaptic connections between the successive layers of neurons are updated using associative learning. Although VisNet has been often used to model invariance in the ventral visual stream, it has been subsequently applied to simulate visual processes occurring in the dorsal stream (Rolls and Stringer, 2007; Galeazzi et al., 2013; Rolls and Webb, 2014). Both ventral and dorsal streams share architectural similarities, each consisting of a hierarchical series of neuronal layers with competition mediated by inhibitory interneurons within each layer (Rolls and Webb, 2014). The VisNet model is described in the Appendix, more detailed descriptions can be found in Rolls (2008).

In this study the model implements *trace learning*, in which a temporal trace of the previous activity of the neuron is incorporated in the learning rule. This learning mechanism encourages individual neurons to respond to subsets of input

FIGURE 1 | Stylized image of the VisNet four-layered network. The architecture of the network shows a hierarchical organization which can be found in the dorsal visual system. Convergence through the network is designed to provide fourth-layer neurons with information from across the entire input retina.
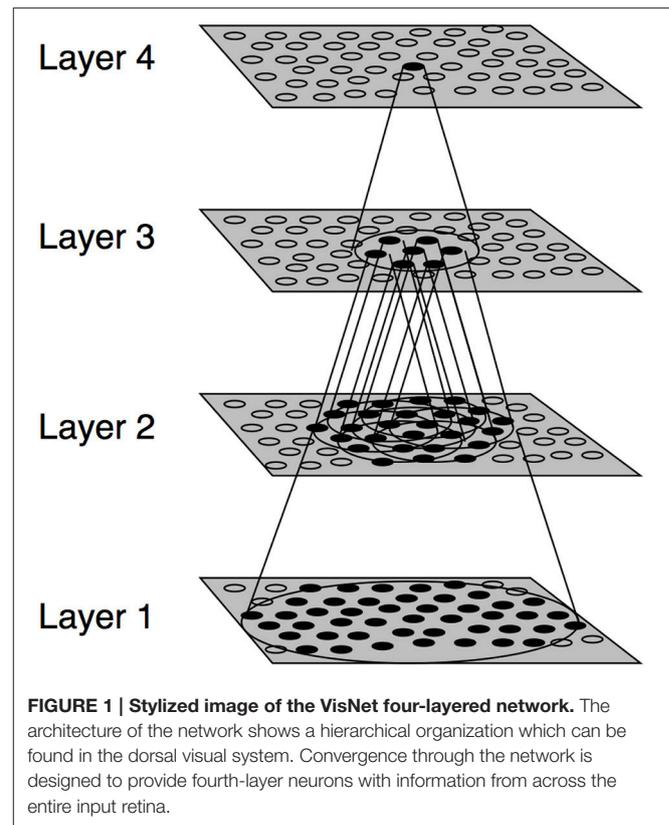
images that occur close together in time. We have previously shown how trace learning may allow neurons to develop responses that are selective for the location of visual targets with respect to the hand but invariant to the position of the hand-object configuration on the retina. In particular, we suggested that while the eyes are exploring a visual scene containing a target object in a fixed position with respect to the hand, trace learning would associate together different views (retinal locations) of the same hand-object configuration onto the same subset of output neurons. In this way, different output cells would learn to respond selectively to different positions of the visual objects with respect to the hand, where the neuronal responses were invariant across different retinal locations (Galeazzi et al., 2013).

2.2. Information Measures

In addition to the response profile of individual neurons, we assessed the network performance using single and multiple cell information theoretic measures. These measures have been used extensively to analyse the performance of the VisNet model in previous work (See Appendix). In this particular case, these measures are used to evaluate whether individual cells in the output layer are able to respond to a specific target location in a hand-centered frame of reference over a number of different retinal locations.

The single cell information metric computes the amount of information conveyed by an individual output layer cell

about which of the stimuli has been shown during testing. In this study, a stimulus is defined as one of the different hand-object configurations presented to the network during testing. For example, if an output neuron developed a localized hand-centered receptive field, then it would respond maximally and selectively to the location of an object in a particular position with respect to the hand across all tested retinal locations in which this configuration appears.

On the other hand, the maximal cell information computes the amount of information conveyed by the output population about all of the possible hand-object configurations. This measure verifies whether there is information about all of the testing stimuli across the output layer. For example, if the maximal multiple cell information is reached, this would mean that all the tested hand-object configurations are represented independently by separate output neurons. In other words, the network would develop a variety of hand-centered output cells, each of them with their own localized hand-centered receptive field. These cells would then respond selectively to the location of an object in a particular position with respect to the hand, and all of the tested locations would be represented in the output layer. More details about how these metrics are applied and calculated for this study are provided in Appendix.

2.3. Model Parameters

For these simulations we used an up-scaled version of the model "retina" (i.e., 256 × 256). Increasing the size of the retina, significantly improves the resolution and therefore the

performance of the model. The rest of the parameters are described in Appendix.

3. TRAINING AND TESTING PROCEDURES

3.1. Experiment 1: Presentation of the Hand with Pairings of Natural Objects

In the first experiment, VisNet was trained on images portraying various spatial configurations of the hand with pairs of natural objects, which were presented against different textured backgrounds. Each of these training images was shifted across different retinal locations during training. We investigated whether these training images could produce output layer neurons with single, localized, hand-centered receptive fields, and which responded invariantly as the neuron's preferred hand-object configuration was shifted across different retinal locations.

3.1.1. Stimuli

The training images for the first experiment consisted of a hand and two natural objects in different spatial configurations surrounding the hand, all of which were presented against different textured backgrounds. The images of the hand, objects and backgrounds were selected from open source pictures on the internet. The templates were designed, scaled and arranged using Adobe Photoshop software. The images were generated in RGB color and subsequently converted to monochrome using the MATLAB function `rgb2gray`. **Figure 2** shows a sample of some of the training images of hand-object configurations that were generated for this study.
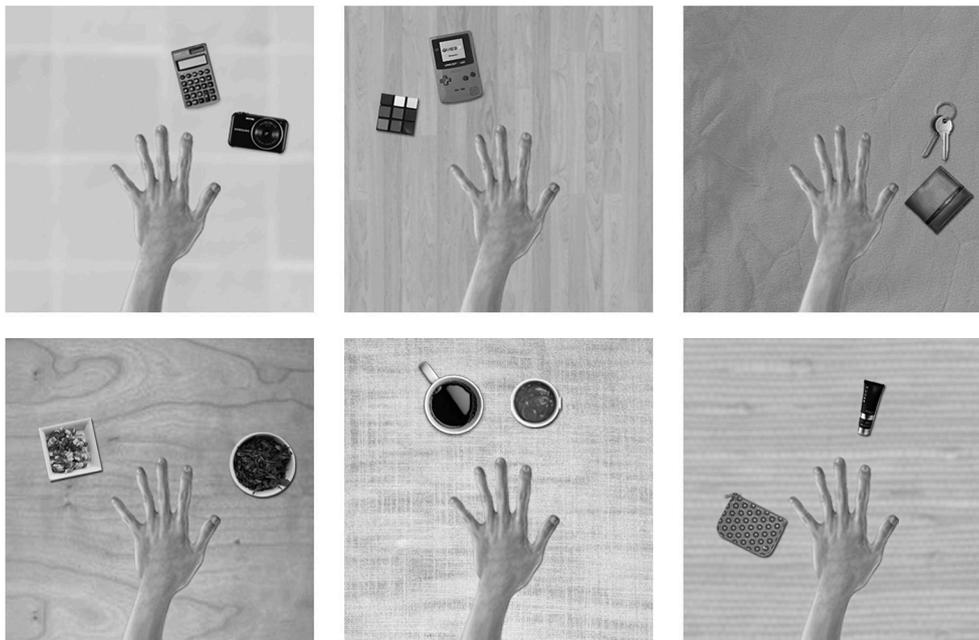
FIGURE 2 | These are examples of some of the training images used in the first experiment. A pair of objects would appear in different hand-centered locations simultaneously. The eyes would move exploring the visual scene producing different views of the same configuration across different retinal locations. The six figures represent a sample of the possible images of object pairings generated from the pool of natural objects and textured backgrounds. The relative positions of the hand and the pair of objects are unchanged during the eye movements.

The backgrounds of the images were extended to 512×512 pixels for the preprocessing stage. The filtered outputs were then cropped back to the original size 256×256. This step is important to avoid possible artifacts or edge effects from the filters in the initial layer of the network.

There was a pool of 42 natural objects to be presented with the hand during training. The centers of all the objects were distributed along a semicircle in six different possible locations around the hand. The images showed all possible pairings of the six hand-centered object locations. The number of possible pairs of object locations may be calculated by

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} \tag{1}$$

where $n = 6$ and $r = 2$, which gives a total of 15 pairings of object locations. For each such pair of object locations we randomly selected two objects from the pool to be presented in that pair of locations. However, each such pair of objects was presented in both possible arrangements: i.e., object 1 in location 1 and object 2 in location 2, and then object 1 in location 2 and object 2 in location 1. This led to a total of 30 hand-object configurations. Then each of the blocks of these 30 hand-object configurations was presented against one of the 21 different textured backgrounds. This generated a total of 630 images. In order to present the hand-object configurations in different retinal positions, each of these configurations was translated by six pixels at a time across VisNet's retina. During training, the hand was always shown surrounded by a pair of objects and never with a single object in isolation. After training was completed, the images used during testing consisted of the hand and a novel object in a specific position relative to the hand. In the test images, the novel objects were shown in one of the same six hand-centered object locations that were used during training.

3.1.2. Training
The training procedure for this experiment consisted of presenting VisNet with pairs of non-overlapping natural objects displayed around the hand on a textured background. During training, the objects were presented in pairs and never in isolation, (see **Figure 2**). In order to develop invariant responses across different retinal views, each of the images representing a particular configuration of a hand and objects was trained across five different retinal locations. The image sequences were meant to arise from a series of eye movements and the resulting shifts in the position of the hand and visual objects on the 256×256 "retina." During each of the image sequences, the fixed spatial configuration of the hand and pair of objects was translated six pixels at a time. During the visual exploration of a particular spatial configuration the natural background was always the same. A new background was only used when a new configuration of a hand and objects was presented.

During training, each image was presented to the network in turn. The image was first convolved with the input Gabor filters and the outputs of the Gabor filters are then passed to the first layer of neurons. Next, the firing rates of neurons in the first layer were calculated with soft competition as described in

Appendix. Next, the weights of the afferent synaptic connections were updated according to the trace rule given by Equation (A11). This process was then repeated for each subsequent layer of the network. The network was thus trained one layer at the time, starting with layer 1 and finishing in layer 4.

One training epoch consisted of the presentation of all 30 object pairings shown against one of the 21 different textured backgrounds, with each of these images presented across five different retinal locations. **Figure 2** shows examples of six training images, each composed of the hand with two natural objects. In these simulations the network was trained for fifty epochs per layer. The learning rates used were 0.1, 0.1, 0.1, and 0.1 in each layer. The number of epochs and learning rates used are the same in all the experiments. For more details on the VisNet parameters, see Appendix.

3.1.3. Testing
Throughout the testing phase, the synaptic weights were not changed. **Figure 3** shows the six images presented to the network during testing. In order to test whether VisNet has developed translation invariant neurons with a single, localized, hand-centered receptive field, the network was tested with images of the hand and a single circular object presented in only one of the six hand-centered locations at a time. Furthermore, because the goal is to test whether neurons respond to a specific hand-centered location irrespective of the object form, the test images used a simple textured object as shown in **Figure 3**. During testing, the responses of the output layer neurons were recorded for each of the hand-object configurations shown in **Figure 3** presented in each of the five retinal locations.

Lastly, a recent addition to the inspection tools of VisNet enables the user to select an output cell after training and then trace back the connections through layers that have been strengthen by learning. This process can be repeated up until the point that we reach the bank of Gabor filters in the input layer. This permits us to identify which visual features of the input images the selected output cell is responding to most strongly.

3.2. Experiment 2: Decay of Object-Selectivity with Increased Visual Training
In the second experiment, we investigated how the shape-selectivity of hand-centered output layer neurons depended on the amount of visual training that the network had received. Specifically, we explored the hypothesis that neurons would become less shape selective as they were trained on larger numbers of objects at their preferred hand-centered location.

3.2.1. Stimuli
The training images for this experiment consisted of the hand presented with a single natural object at a time. The natural object was always presented at the same location with respect to the hand. The objects were drawn randomly from the same pool of 42 natural objects used in the first experiment. The images were generated in RGB color and subsequently converted to monochrome using the MATLAB function `rgb2gray`. Different simulations were run with increasing numbers (1–8)
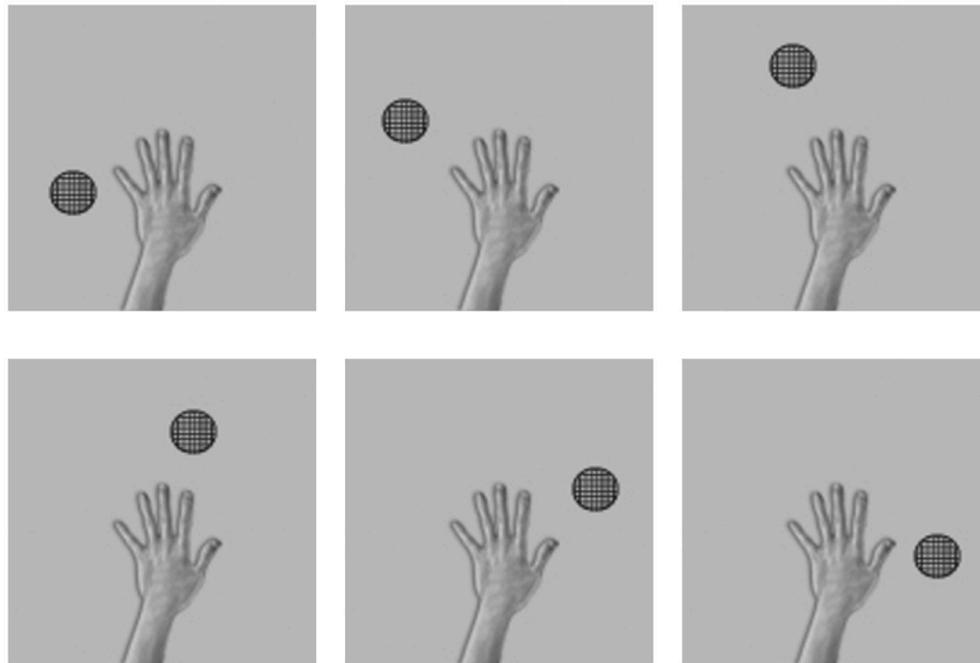
FIGURE 3 | These are examples of six testing images used in the first experiment to determine the hand-centered receptive fields of output neurons after training. Unlike the training condition where two objects were presented simultaneously, in this case a single textured stimulus was presented in six different hand-centered locations. The hand-centered object locations were the same as those used during training in the first experiment. Each hand-object configuration would be tested in five different retinal locations.

of natural objects used during training. For each simulation, the network was tested with images of the hand and each of the 100 different novel objects presented in the same hand-centered location on which the network was trained. The objects used during training and testing were not the same. **Figure 4** shows examples of the pool of objects used for training and testing. At testing, we recorded the percentage of the 100 test objects that the output neurons responded to. This allowed us to assess the shape selectivity of these neurons.

3.2.2. Training and Testing
For this experiment we were interested in exploring whether the output cells that developed visual hand-centered receptive fields could also show shape selectivity, and how this shape selectivity depended on the amount of visual training with different natural objects. We started by training the network with an image of the hand with a single natural object in a particular position with respect to the hand. We then tested the network with a pool of 100 novel objects presented in the same hand-centered location as used during training. Then across further simulations we systematically increased the number of objects that appeared in the same hand-centered location during training. One training epoch consisted of presenting images of the hand with each of the training objects that were used for that particular simulation. After training was completed, the network was tested with the same set of 100 images showing the hand with one of the novel objects. The aim was to investigate how the shape selectivity of neurons that learned to respond to that hand-centered location

was affected by the number (1–8) of natural objects seen there during training.

This experiment was not focused on the development of invariant neuronal responses across different retinal locations, and so we trained each image of the hand and object in only a single retinal location. Consequently, we updated the synaptic weights between layers according to the simpler Hebb rule (See Equation A10 in Appendix).

3.3. Experiment 3: Presentation of the Hand Against Natural Backgrounds
In the third experiment, VisNet was trained on images with the hand presented against completely natural backgrounds, which were also shifted across different retinal locations. We investigated whether output layer neurons learned to respond to objects presented in single hand-centered locations, and whether these responses were invariant as the neuron's preferred hand-object configuration was shifted across the retina.

3.3.1. Stimuli
In order to generate our pool of natural visual scenes, we asked four volunteers to provide 10–12 photographs of natural visual scenes from their everyday life in which they would normally use their hands to manipulate objects. All of the volunteers were naive and unaware of the purpose of the study. We provided several examples (e.g., using cutlery in a meal, grasping a cup, etc.) and provided three sample photos in order to give them a general idea of the nature of the scenes we
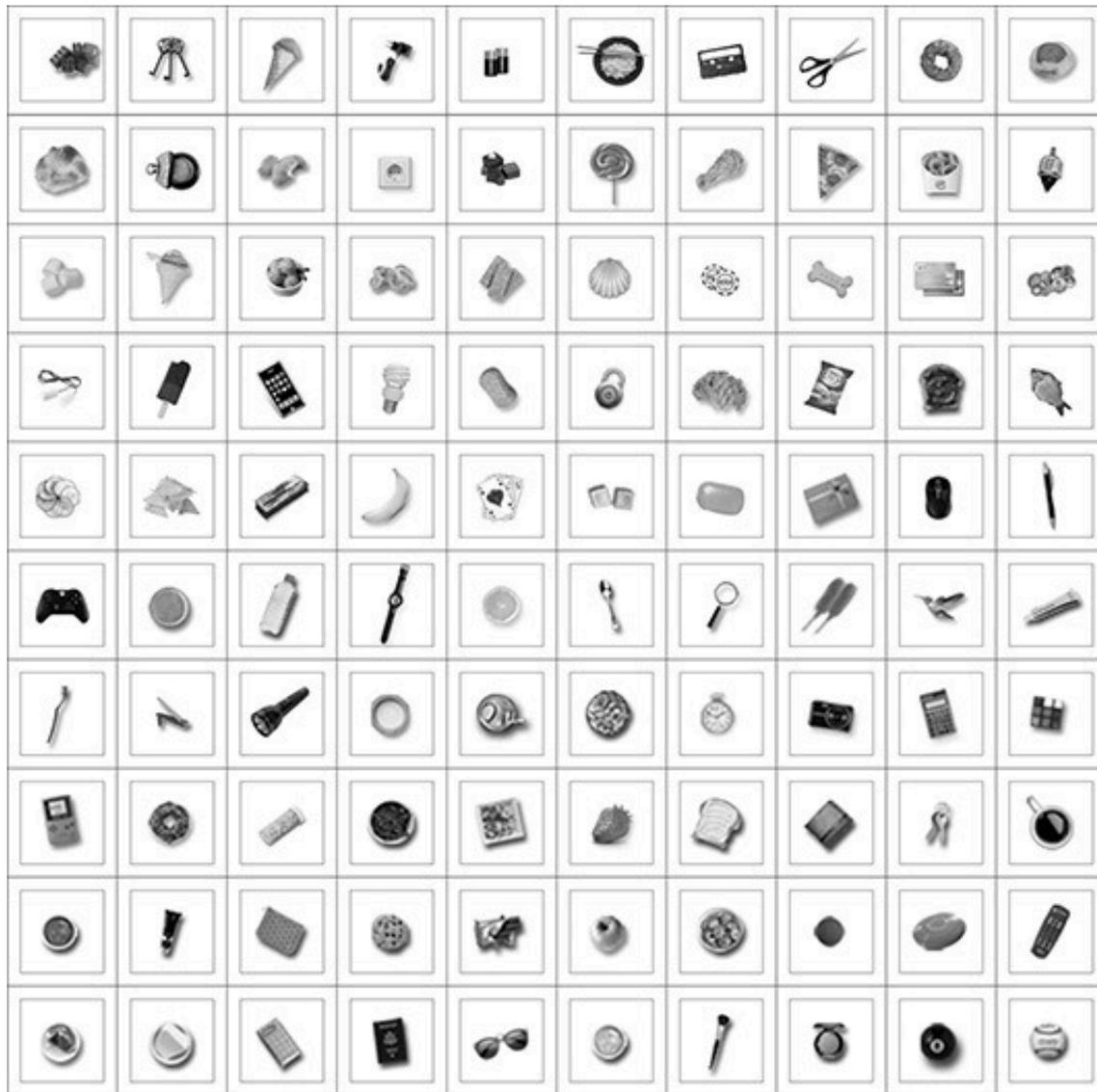
FIGURE 4 | The figure shows various examples of natural objects that were presented in the same location with respect to the hand during training and testing for the second experiment. Different objects were used during training and testing. The objects were always presented in the same hand-centered location in order to explore how the shape selectivity of neurons representing that location was affected by the number of objects presented there during training.

were interested in collecting. We provided further instructions regarding the angle and distance at which the photos should have been taken. The pictures were meant to be taken from a first person point of view and the distance between the objects and the camera had to be at arm's length. Additionally we asked them not to include the image of their own hand in the picture.

The training stimuli for this experiment consisted of images showing a picture of a real hand that was superimposed in all of the natural visual scenes collected by our participants. The templates were scaled and arranged using Adobe Photoshop software. The images were generated in RGB color and subsequently converted to monochrome using the MATLAB

function `rgb2gray` and then resized to a 256 × 256 matrix. **Figure 5** shows a sample of some of the training images that were generated. A total of 48 natural images were collected and used for the experiment. In order to present the configurations of the hand and objects in different retinal positions, each of the fixed spatial configurations was translated by five pixels at a time across VisNet's retina within a 3 by 2 grid. That is, for this experiment the sequences included horizontal as well as vertical shifts on the network's retina.

After training was completed, the stimuli used during testing consisted of images showing the hand and a novel textured object in five different positions relative to the hand as shown in **Figure 6**.
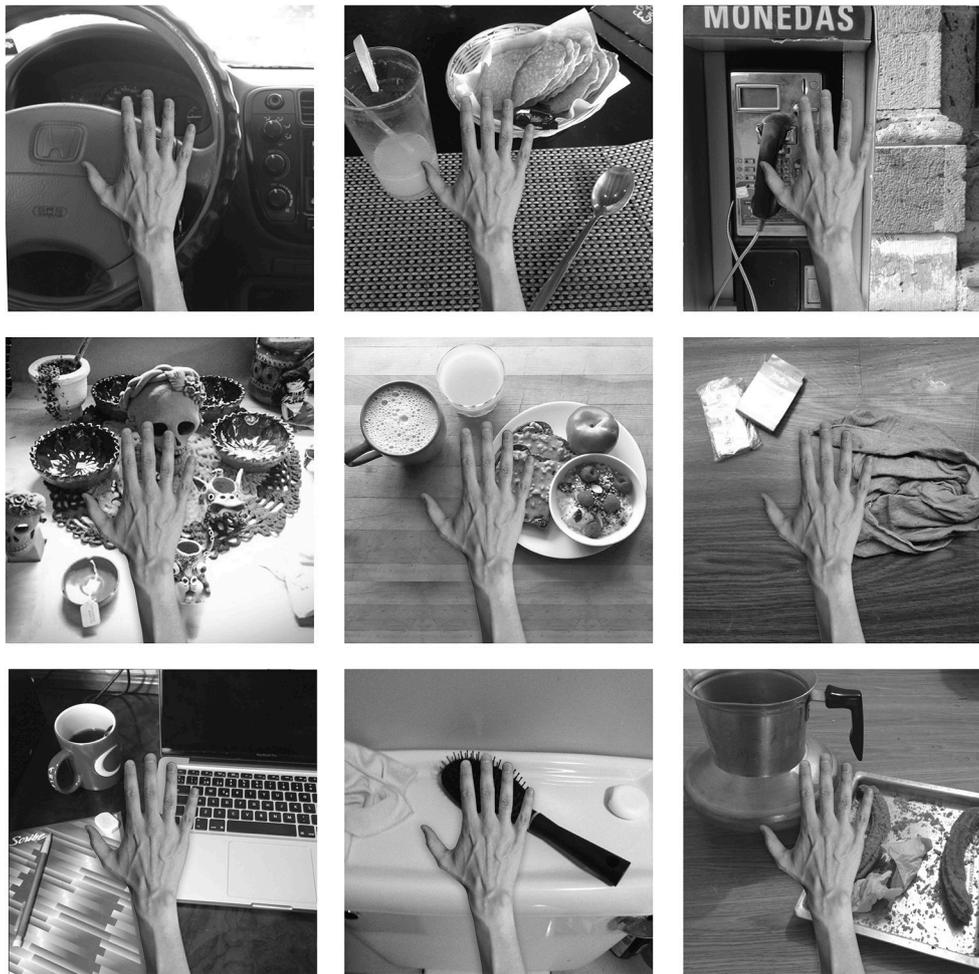
FIGURE 5 | The figure shows various examples of the hand presented against different natural backgrounds during training in the third experiment.
The position of the hand within each of the backgrounds is unchanged during the eye movements.

3.3.2. Training and Testing

The training procedure for this experiment consisted of presenting VisNet with images of the hand embedded within 48 different natural scenes containing a variety of objects as shown in **Figure 5**. As in previous simulations, image sequences were meant to arise from a series of eye movements and the resulting shifts in the position of the hand and visual objects on the 256 × 256 "retina." During each of the image sequences, the fixed spatial configuration of the visual scene is translated both horizontally and vertically by five pixels at a time across a 3 by 2 grid of retinal locations. In the first experiment we shifted the images only horizontally. However, in order to increase the ecological validity of this third experiment, we included a vertical shift of five pixels as well. In this experiment, the synaptic weights were updated according to the trace rule given by Equation (A11). One training epoch consisted of presenting all 48 images in all 6 retinal locations.

Figure 6 shows the images used to test the network after training. In order to test whether VisNet has developed

translation invariant neurons with a single, localized, hand-centered receptive field, the network was tested with images consisting of the hand with only a single textured object presented in one of five different hand-centered locations. The responses of the output neurons are recorded with each of these hand-object configurations presented in all six of the retinal locations used during training.

4. RESULTS

4.1. Experiment 1: Presentation of the Hand with Pairings of Natural Objects

We studied the responses of the output (fourth) layer cells in VisNet before and after the network was trained on the images of hand-object configurations shown in **Figure 2**. After the network was trained, the network was tested on the images shown in **Figure 3** to determine whether cells in the output layer had developed single, localized hand-centered receptive fields and responded invariantly across the different retinal locations.
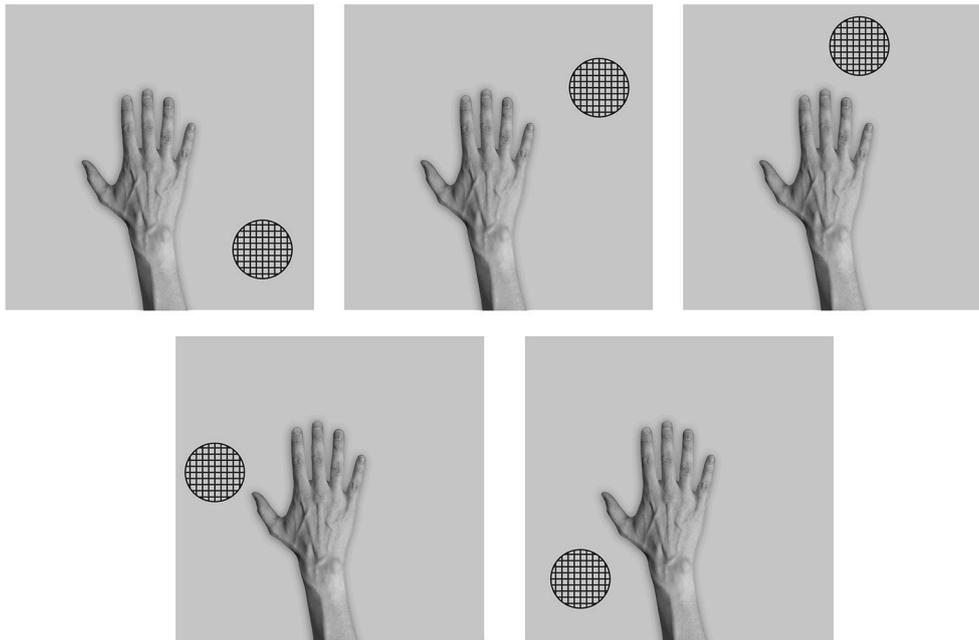
FIGURE 6 | These are the five test images used after the network had been trained with the hand presented against natural background scenes in the third experiment. That is, the network was tested with a single textured stimulus presented in five different hand-centered locations. Each hand-object configuration was tested in the six different retinal locations that the hand was originally trained in.

Information analysis was then conducted on the responses of the cells to all of the test images.

In previous simulations in which VisNet was trained on all possible pairings of objects, it was reported that as the number of objects increased, the statistical decoupling between the objects started to force the network to learn to represent the objects individually (Stringer et al., 2007). However, in the new simulations carried out here the image of the hand was always present with the objects. In this case, the most correlated features would correspond to a combination of features of the hand and features of the trained objects presented in a particular location with respect to the hand. Therefore, individual cells should learn to respond to a particular spatial configuration of the hand and a single hand-centered object location.

Figure 7 shows the response profiles of six neurons in the output layer of VisNet before training. Following the same conventions of Galeazzi et al. (2013), each of the six columns of plots contains the firing responses of a particular output cell, which are labeled at the top of the column. Whereas the six rows of plots show the responses of the cells to each of the six hand-object configurations presented during testing.

Each plot shows the responses of the given cell to the particular hand-object configuration over the five retinal locations. The x axis in each plot represents the five retinal locations of the hand-object configuration on which the neuron was tested, while the y axis represents the corresponding firing rate of the output neuron. The top row shows the cell responses when a single textured object is presented in the first of the testing locations

with respect to the hand. This corresponds to the upper left image in **Figure 3**. The following rows show the cell responses when the visual object is presented in successive test locations with respect to the hand. The last row corresponds to the configuration displayed in the bottom right image of **Figure 3**.

In **Figure 7** we can see that before training, all of the six cells responded rarely and randomly to the different hand-object configurations. The responses do not have a particular ordered structure. In **Figure 8** we can see the response profiles of the same six neurons in the output layer of VisNet after training. In this case it can be seen that, after training, each of the six cells has learned to respond to just one of the hand-object configurations, and responds to that configuration over all five tested retinal locations. Furthermore, we can see here already that each of the six hand-object configurations was represented by one of the cells.

In order to have an overview of how these configurations are represented across the output cell population, we present the information analysis measures. **Figure 9** shows the single and multiple information measures for the output (fourth) layer neurons before and after training with all of the hand-object configurations. The single cell information analysis (**Figure 9** top) shows that, after training, 115 neurons conveyed the maximal single cell information of 2.58 bits. These output cells responded to only a single position of the test object with respect to the hand, and responded irrespective of retinal location. The multiple cell information analysis (**Figure 9** bottom) shows that, before training, the multiple cell information does not reach the maximal value of 2.58 bits. However, after training we can see

FIGURE 7 | Firing rate responses from the first experiment before training. Each of the six columns shows the firing responses of a particular cell. Each row shows the responses of the six cells to one of the six hand-object configurations (shown on the left) over all five different retinal locations shown along the abscissae. It can be seen that each of the six cells initially responds randomly to each of the hand-object configurations over the different retinal locations.

that multiple cell information asymptotes to the maximal value, which means that all six of the hand-object configurations are represented by separate cells in the output layer. **Figure 8** shows examples of neurons representing each of the six hand-object configurations.

We traced the strengthened connections from each one of the output cells through successive layers to the input Gabor filters driving that cell. **Figure 10** shows the Gabor input filters with strengthened connections to a trained output neuron that had learned to respond to one of the hand-centered locations. On the left side of **Figure 10** we can see the Gabor filters that are most strongly driving the responses of the particular output cell. In this example, we show a cell that is representing a subset of Gabor filtered inputs corresponding to the hand, as well as a subset of inputs representing a visual location near the hand. Tracing back the synaptic connectivity in this way enables us to inspect the nature and extension of the hand centered visual receptive field developed by the output cell after training. We can thus determine not only the ability of the cell to represent an individual region with respect to the hand, but also the input features that were extracted from the set of objects shown.

Altogether, the individual cell firing rate responses, the information analysis and the inspection of connectivity in this experiment demonstrate that VisNet is able to develop neurons with a single, localized, hand-centered visual receptive fields even when trained on more realistic images with multiple natural objects shown with the hand against various textured backgrounds. In particular, the principles of statistical decoupling continue to operate successfully under these more ecological training conditions. That is, after extensive training, the output cells learn to respond to the features that are seen more frequently together throughout training. This is a basic property of competitive learning. Since the network is trained on multiple natural objects with the hand against various textured backgrounds, the features that appear more frequently together are the hand (which is always present) and a subset of features that are associated with a particular object location. Consequently, individual output neurons learn to represent a particular configuration of the hand and one object location with separate neurons responding to different hand-centered object locations. However, the statistical decoupling between any two object locations is too weak to allow individual output cells to learn to respond to more than one hand-centered location.

FIGURE 8 | Firing rate responses from the first experiment after training. Response profiles of the same six neurons from **Figure 7** after training on the images shown in **Figure 3**. Conventions as in **Figure 7**. It can be seen that each of the six cells responds selectively to just one of the hand-object configurations, and responds to that configuration over all five retinal positions shown along the abscissae. Moreover, each of the six hand-object configurations is represented by one of the cells.

Additionally, the trace learning mechanism enables the network to encode these representations across different retinal locations. Thus, these cells will respond to the same hand-object configuration irrespective of the position of the hand with respect to the body and regardless of the gaze direction. These hand-centered cells will fire maximally as long as the spatial configuration of the hand and an object is the same.

4.2. Experiment 2: Decay of Object-Selectivity with Increased Visual Training

In Experiment 1, we were not interested in developing hand-centered cells that were selective to specific objects. On the contrary, we were primarily interested in the development of hand-centered receptive fields where the neuron would respond to the presence of almost any object as long as it was presented within the receptive field. These cells are thought to mostly provide information about the location of an object with respect to the hand, rather than representing the detailed features of the object. However, our simulations do not preclude

the possibility that some shape selectivity could arise after training.

In the second experiment, we investigated whether the hand-centered output neurons showed selectivity to the shapes of objects presented with the hand, and how this shape selectivity depended on the amount of training that the network had received with other objects. By testing the network on images with a variety of novel objects in the same hand-centered location used during training, it was possible to assess whether the cells that had learned to respond to that hand-centered location would fire selectively to objects of a particular shape. A number of experiments were performed with sampling different objects during training. The results presented here are taken from one of these experiments and are typical of the effects we observed.

In Experiment 2, eight separate simulations were conducted. Successive simulations used increasing numbers of training objects from 1 to 8, which were always presented at the same location with respect to the hand during training. For each simulation, after training we identified the subpopulation of output neurons that had learned to respond to that hand-centered location. The criterion for classifying a cell as responsive
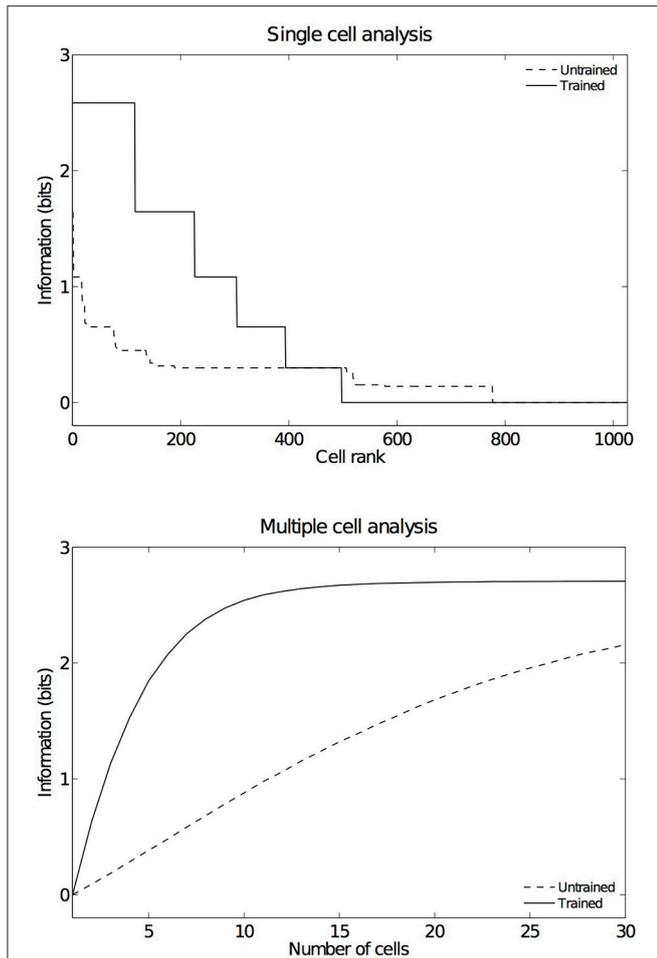
FIGURE 9 | Information results from the first experiment before and after training. The upper plot shows the amount of single cell information carried by individual output cells in rank order. After training, it was found that 115 cells reached the maximum amount of single cell information of 2.58 bits. These cells responded perfectly to just one of the six tested hand-object configurations, and responded to that configuration across all five different retinal locations. In the untrained condition no cells reached maximal information. The lower plot shows the multiple cell information measures calculated across 30 cells with maximal single cell information. It can be seen that, after training, the multiple cell information asymptotes to the maximal value of 2.58 bits. This confirms that all six tested hand-object configurations are represented by the output cells.

FIGURE 10 | Tracing back the synaptic connections from a trained output cell to the input Gabor filters in the first experiment. The left side shows the input Gabor filters that an output cell has learned to respond after training. This is an example of a neuron that represents a hand-object configuration with the object above the hand. In this image the Gabor filters with the strongest connectivity through the layer to the output cell are plotted, where each Gabor filter is weighted by the strengths of the feed-forward connections from that filter through the successive layers to the output neuron. It can be seen that this neuron receives the strongest inputs from a subset of Gabor filters that represent the location of the target on top of the hand. The right side shows the image of the hand and the overlapped images of all the training objects that appeared during training in this hand-centered location.
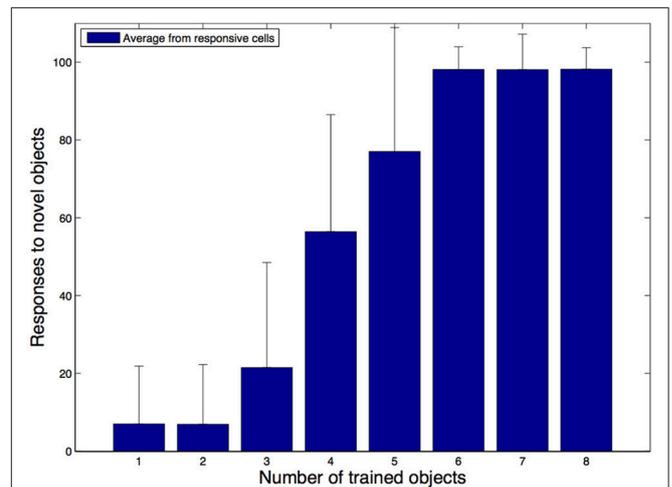
FIGURE 11 | Simulation results for the second experiment. In these simulations we explored how the shape selectivity of a subpopulation of hand-centered output neurons is affected as the network is trained with an increasing number of natural objects at their preferred hand-centered location. The plot shows the average number of novel test objects that the subpopulation of output cells respond to as the network is trained with an increasing number of the training objects. It is evident that as the network is exposed to more objects during training, most cells start to lose their shape selectivity and respond to a larger percentage of the novel objects.

was that its firing rate should reach a threshold of 0.5. Then we tested the network on 100 images of the hand with different novel objects at the same hand-centered location. Each time we recorded whether each of the neurons responded to the new object at that hand-centered location. This procedure was used to reveal how the shape selectivity of the output neurons changed as the network was trained with increasing numbers of objects at their preferred hand-centered location.

Figure 11 shows the average number of novel objects that the hand-centered cells in the network responded to after training as a function of the number of objects that the network has

seen at that hand-centered location during training. The ordinate corresponds to the percentage of novel objects that the cells respond to while the abscissa corresponds to the number of objects seen during training. We can see from these simulations that the cells with hand-centered receptive fields started to lose their shape selectivity as they got trained with more and more objects in the same hand-centered location. Even when we still found a few shape selective cells, the proportion of highly selective cells was substantially reduced as the training is increased. This means that most of the cells would respond to the presence of an object in a region of space near the hand regardless of the form of the object.

What learning mechanism leads to a reduction in the shape selectivity of neurons as the network is trained on increasing numbers of objects at the same hand-centered location? When the first object is presented with the hand during training, a small subset of output neurons will win the competition and respond. Then Hebbian associative learning in the feedforward connections within the network will increase the tuning of these cells to respond to that particular object in that hand-centered location. However, when another object is presented in the same hand-centered location, the two objects may share some features in common. The activation of these common features may then cause the same subset of output neurons to respond again because the relevant feedforward connections were strengthened during training with the first object. The effect of this will be to associate the features of the new object with the same output neurons. This process may be repeated with a number of successive different objects presented with the hand. All of the features of these objects will become associated with the same output neurons. Thus, the output neurons gradually lose their selectivity to the form of the objects, and merely respond to any object presented in that hand-centered location. This would produce receptive fields that represent the locations in which the objects appear with respect to the hand, without being particularly selective about the differences between the features of these objects. Thus, as the results show, as the network is trained with more and more objects, the localized hand-centered receptive fields start to lose their shape selectivity and respond to a variety of novel objects as long as they appear within the hand-centered receptive field. This learning process is somewhat similar to continuous transformation (CT) learning (Stringer et al., 2006), which drives the development of invariant neuronal responses by exploiting the similarities between visual stimuli.

Consistent with our results, when we make a comparison at a single-cell neuron level between high-level ventral regions that are shape selective, such as the anterior inferotemporal cortex (AIT) and high level dorsal regions that have been also reported as shape selective (e.g., LIP), it has been found that AIT neurons on average had higher shape selectivity than those of LIP (Lehky and Sereno, 2007). AIT neurons also had significantly more units that were highly selective to shape, whereas LIP had very few neurons that were highly selective to shape.

4.3. Experiment 3: Presentation of the Hand Against Natural Backgrounds

In the third experiment we investigated whether output neurons developed localized hand-centered receptive fields when the network was trained on images containing a hand presented against a natural background scene as shown in **Figure 5** and then tested on the images shown in **Figure 6**.

Figures 12, **13** show the response profiles of five neurons in the output layer of VisNet before training and after training, respectively. Following the same conventions of the response profiles in Experiment 1, each of the five columns of plots contains the firing responses of a particular output cell, which is labeled at the top of the column. The five rows show the responses of the cells to each of the five hand-object configurations presented during testing. Each plot shows the responses of the given cell to the particular hand-object configuration over six different retinal locations. Before training (**Figure 12**) none of the cells responded exclusively to any of the hand-object configurations; in fact they responded rarely. However, after training, in **Figure 13** we can see that each of the five cells learned to respond exclusively to one specific hand-object configuration, and that these responses were invariant to different retinal locations.

As in the other two experiments presented here, an information analysis was carried out to investigate how these hand-object configurations are represented across the whole population of output cells. **Figure 14** shows the single and multiple cell information measures for the output (fourth) layer neurons before and after training the network on images of the hand presented against natural backgrounds. The information analysis was performed by testing the network on the five hand-object configurations shown in **Figure 6**, where each such configuration was presented in six retinal locations.

Figure 14 (top) shows the single cell information measures for the output layer of neurons. We can see here that, before training none of the cells reached the maximum information. However, after training 49 neurons reached the maximal single cell information of 2.32 bits. This means that these 49 output cells responded selectively to a single localized position of the test object with respect to the hand, and that this response was invariant to retinal location. In **Figure 14** (bottom) it is evident that before training the multiple cell information did not reach the maximal value of 2.32 bits. However, after training we can see that the multiple cell information asymptotes to the maximal value, which means that all of the possible hand-object configurations are successfully represented by separate cells in the output layer. In fact, the five cell response profiles after training shown in **Figure 13** already confirmed that the network was able to represent each of the five hand-object configurations. The multiple cell analysis simply reaffirms that all five hand-object configurations are represented invariantly across all retinal locations by separate output neurons.

For this simulation we again traced the strengthened connections from each one of the output cells through successive layers to the input Gabor filters driving that cell. In **Figure 15** we can see the Gabor input filters with strengthened connections to a trained output neuron that had learned to respond to one of the hand-centered locations. On the left side of **Figure 15** we can see the Gabor filters that are most strongly driving the responses of the particular output cell. This cell is representing a subset of Gabor filtered inputs corresponding to the hand, as well as a subset of inputs representing a localized region near the hand.

FIGURE 12 | Firing rate responses from the third experiment before training. Each of the five columns shows the firing responses of a particular cell. Each row shows the responses of the five cells to one of the five hand-object configurations (shown on the left) over all six different retinal locations shown along the abscissae. It can be seen that each of the five cells initially responds randomly to each of the hand-object configurations over the different retinal locations.

The right side of **Figure 15** shows the image of the hand with the hand-centered receptive field of the neuron shown in blue.

5. DISCUSSION

In the simulations presented in this paper we have investigated whether VisNet could still self-organize and develop neurons with single, localized hand-centered receptive fields, as the network is trained under more realistic visual training conditions. In these experiments, we have systematically improved the realism of the visual training stimuli in order to test the robustness of the proposed learning mechanism that relies on a combination of statistical decoupling between hand-centered object locations and trace learning in order to drive the development of hand-centered visual representations.

We have shown how some neurons learn to respond to particular spatial configurations of the hand and an object location. Such neurons represent the location of a visual object in the reference frame of the hand. This learning process exploits the statistical decoupling that will exist between different hand-centered object locations across many different images. Furthermore, these neuronal responses can become invariant across different retinal locations by trace learning. This learning rule binds together input patterns which tend to occur close

together in time. If the eyes typically saccade around a visual scene faster than the hand moves, then trace learning will bind together the same hand-object configuration across different retinal locations.

In Section 4.1 we began to address how the network might develop neurons with single, localized, hand-centered receptive fields if it is trained on more realistic images containing multiple objects presented simultaneously with the hand. Specifically, we showed that presenting the objects in many different pairs of hand-centered locations during training facilitated the statistical decoupling between different object locations, which in turn forced output neurons to develop localized hand-centered receptive fields. This allowed us to train the network with more than one object presented at a time with the hand.

In Section 4.2 we investigated how the shape selectivity of neurons was affected by the number of objects that the network was trained on at a particular hand-centered location. We proposed that whenever a new object is shown at a particular hand-centered location, then there will likely be some overlap of features with previous objects presented at that location. In such a case, it is likely that some of the same output cells will fire again to the presence of the new object. These cells would get their synaptic weights from the features of the new

FIGURE 13 | Firing rate responses from the third experiment after training. Response profiles of the same five neurons from **Figure 12** after training on the images shown in **Figure 6**. Conventions as in **Figure 12**. It can be seen that each of the five cells has learned to respond selectively to just one of the hand-object configurations, and responds to that configuration over all six retinal positions shown along the abscissae. Moreover, each of the five hand-object configurations is represented by one of the cells.

object strengthened. As the network is trained on more and more objects at the same hand-centered location, this subset of cells gradually learn to respond to most object features at that location and hence lose their shape selectivity. Our simulations suggest the possibility that hand-centered neurons in area 5d and other parts of the posterior parietal cortex may in fact display a range of different degrees of object shape selectivity. The responses of some neurons may be still somewhat selective to shape, while other neurons respond to almost all objects placed within their hand-centered receptive field. Such a heterogeneous population of neurons was in fact observed in our simulations.

Lastly, in Section 4.3 we further increased the realism of the simulations by training VisNet on images of the hand presented against natural visual scenes. Unlike the previous simulations where the hand-centered object locations were carefully controlled, this time the objects could appear in any location around the hand. Furthermore, there was also more variability in the relative size of the objects and their distance to the hand. Given the richness of the visual training scenes in Experiment 3, the output cells showed more spatial heterogeneity in their receptive fields. For example, as shown in **Figure 15**, one of the particularly interesting

differences in this simulation result was that the localized receptive fields near the hand had irregular and idiosyncratic shapes, some of them covering larger areas surrounding the hand.

Altogether, the results from the experiments presented here showed how individual output cells could develop single, localized, hand-centered visual receptive fields which are invariant to retinal location. This occurred even when the network was trained on more realistic visual scenes with multiple objects presented simultaneously with the hand, or even with the hand presented against complex natural backgrounds. This is an important step to show how these hand-centered representations could emerge from the natural statistics of our visual experiences and under more realistic training conditions. More importantly, we showed that this can be achieved using an unsupervised learning mechanism where the synaptic weights are updated in a biologically plausible manner using locally available information such as the pre- and post-synaptic neuronal activities.

5.1. Future Directions

In the simulations described in this paper, the hand was always presented to the network in the same pose. In future work,

FIGURE 14 | Information results from the third experiment before and after training. The information analysis was carried out by testing the network on the five hand-object configurations shown in **Figure 6**, where each such configuration was presented in six retinal locations. The upper plot shows the amount of single cell information carried by individual output cells in rank order. After training, it was found that 49 cells reached the maximum amount of single cell information of 2.32 bits. These cells responded perfectly to just one of the five tested hand-object configurations, and responded to that configuration across all six different retinal locations. In the untrained condition no cells reached maximal information. The lower plot shows the multiple cell information measures calculated across 25 cells with maximal single cell information. It can be seen that, after training, the multiple cell information asymptotes to the maximal value of 2.32 bits. This confirms that all five tested hand-object configurations are represented by the output cells.

FIGURE 15 | Tracing back the synaptic connections from a trained output cell to the input Gabor filters in the third experiment after the network was trained on images of the hand presented against natural scenes. The left side shows the input Gabor filters that an output cell has learned to respond to after training. In this image the Gabor filters with the strongest connectivity through the layers to the output cell are plotted, where each Gabor filter is weighted by the strengths of the feed-forward connections from that filter through the successive layers to the output neuron. The right side shows the image of the hand with the hand-centered receptive field of the neuron shown in blue.

we plan to run simulations in which the hand is seen in different postures. For example, the network might be trained on sequences of images as the hand rotates to pick up a series of objects. In this case, we hypothesize that neurons may develop a diverse range of response properties. Some neurons may become selectively tuned to the presence of a visual target with respect to just one pose of the hand, while other neurons could develop pose invariant responses through an invariance learning mechanism such as trace learning (Földiák, 1991; Rolls, 1992) or continuous transformation learning (Stringer et al., 2006).
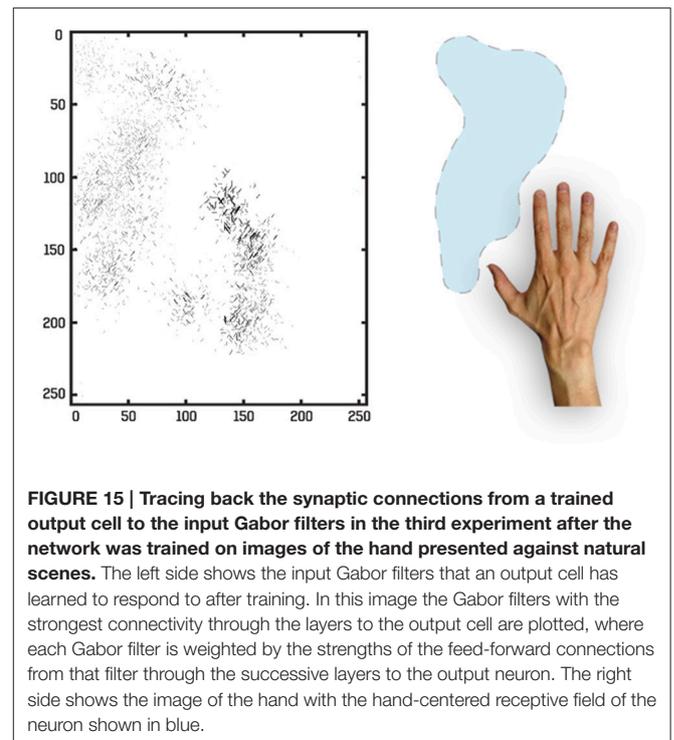
In this paper we were primarily interested in the visual development of such hand-centered representations using a self-organizing approach. Therefore, the input provided to the network about the location of the hand and target was presented visually. However, in the brain the positional information of the location of the hand is integrated using inputs from different modalities, including tactile and proprioceptive signals. In this study we did not explore the role of these different incoming signals. Nevertheless, we hypothesize that they could in some cases facilitate the statistical decoupling and formation of localized hand-centered receptive fields. For example, tactile feedback from the touch of an object will be generally congruent with visual signals representing the hand-centered location of the visual object. In future work, we plan to integrate signals from other modalities such as tactile and proprioceptive information to explore their role in the development of hand-centered representations.

As we mentioned in the Introduction, a variety of regions have been reported as encoding target positions in a hand-centered frame of reference. However, there might be functional differences between these different hand-centered representations (De Vignemont and Iannetti, 2015). It is, for example, unclear how the hand-centered encoding of reach vectors reported in area 5d by Bremner and Andersen (2012) may relate or differ from other hand-centered and peri-hand representations reported in different regions (Graziano et al., 1994, 1997; Graziano and Gross, 1998; Graziano, 1999). The intention to reach to a desired location might be crucial for the hand-centered cells in area 5d, while the mere presence of an object near the hand could be sufficient to elicit a response

from a hand-centered cell in PMv even if there is no intention to interact with it. Some of the behavioral tasks and data analysis from these different studies are not immediately comparable and involve a limited set of experimental conditions. This makes it difficult to disentangle not only the frame of reference in which a particular cell encodes the location of a target, but also how visual, proprioceptive, tactile and motor signals are weighted and integrated during the task. Furthermore, many of these cells may very well have interesting dynamical properties in which the frame of reference could be varying during different moments of the task (Bremner and Andersen, 2014).

REFERENCES

Andersen, R. A., Essick, G. K., and Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science* 230, 456–458. doi: 10.1126/science.4048942

Blohm, G., Keith, G., and Crawford, J. (2009). Decoding the cortical transformations for visually guided reaching in 3d space. *Cereb. Cortex* 19, 1372–1393. doi: 10.1093/cercor/bhn177

Bremner, L. R., and Andersen, R. A. (2012). Coding of the reach vector in parietal area 5d. *Neuron* 75, 342–351. doi: 10.1016/j.neuron.2012.03.041

Bremner, L. R., and Andersen, R. A. (2014). Temporal analysis of reference frames in parietal cortex area 5d during reach planning. *J. Neurosci.* 34, 5273–5284. doi: 10.1523/JNEUROSCI.2068-13.2014

Brotchie, P., Andersen, R. A., Snyder, L. H., and Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* 375, 232–235. doi: 10.1038/375232a0

Brozzoli, C., Gentile, G., and Ehrsson, H. H. (2012). That's near my hand! parietal and premotor coding of hand-centered space contributes to localization and self-attribution of the hand. *J. Neurosci.* 32, 14573–14582. doi: 10.1523/JNEUROSCI.2660-12.2012

Brozzoli, C., Gentile, G., Petkova, V. I., and Ehrsson, H. H. (2011). Fmri adaptation reveals a cortical mechanism for the coding of space near the hand. *J. Neurosci.* 31, 9023–9031. doi: 10.1523/JNEUROSCI.1172-11.2011

Buneo, C., and Andersen, R. (2006). The posterior parietal cortex: sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia* 44, 2594–2606. doi: 10.1016/j.neuropsychologia.2005.10.011

Buneo, C., Jarvis, M., Batista, A., and Andersen, R. (2002). Direct visuomotor transformations for reaching. *Nature* 416, 632–636. doi: 10.1038/416632a

Chang, S., Papadimitriou, C., and Snyder, L. (2009). Using a compound gain field to compute a reach plan. *Neuron* 64, 744–755. doi: 10.1016/j.neuron.2009.11.005

De Vignemont, F., and Iannetti, G. (2015). How many peripersonal spaces? *Neuropsychologia* 70, 327–334. doi: 10.1016/j.neuropsychologia.2014.11.018

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., and Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area f4). *J. Neurophysiol.* 76, 141–157.

Fogassi, L., Raos, V., Franchi, G., Gallese, V., Luppino, G., and Matelli, M. (1999). Visual responses in the dorsal premotor area f2 of the macaque monkey. *Exp. Brain Res.* 128, 194–199. doi: 10.1007/s002210050835

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Comput.* 3, 194–200.

Galeazzi, J. M., Mender, B. M., Paredes, M., Tromans, J. M., Evans, B. D., Minini, L., et al. (2013). A self-organizing model of the visual development of hand-centred representations. *PLoS ONE* 8:e66272. doi: 10.1371/journal.pone.0066272

Gentile, G., Petkova, V. I., and Ehrsson, H. H. (2011). Integration of visual and tactile signals from the hand in the human brain: an fmri study. *J. Neurophysiol.* 105, 910–922. doi: 10.1152/jn.00840.2010

Graziano, M. S. (1999). Where is my arm? the relative role of vision and proprioception in the neuronal representation of limb position. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10418–10421. doi: 10.1073/pnas.96.18.10418

Graziano, M. S., and Gross, C. G. (1993). A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Exp. Brain Res.* 97, 96–109. doi: 10.1007/BF00228820

Graziano, M. S., and Gross, C. G. (1998). Spatial maps for the control of movement. *Curr. Opin. Neurobiol.* 8, 195–201. doi: 10.1016/S0959-4388(98)80140-2

Graziano, M. S., Hu, X., and Gross, C. (1997). Visuospatial properties of ventral premotor cortex. *J. Neurophysiol.* 77, 2268–2292.

Graziano, M. S., Yap, G. S., and Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science* 266, 1054–1057. doi: 10.1126/science.7973661

Hertz, J., Krogh, A., and Palmer, R. (1991). *Introduction to the Theory of Neural Computation*, Vol. 1. Redwood City, CA: Westview Press; Addison-Wesley.

Hyvärinen, J., and Poranen, A. (1974). Function of the parietal associative area 7 as revealed from cellular discharges in alert monkeys. *Brain* 97, 673–692. doi: 10.1093/brain/97.1.673

Lehky, S. R., and Sereno, A. B. (2007). Comparison of shape encoding in primate dorsal and ventral visual pathways. *J. Neurophysiol.* 97, 307–319. doi: 10.1152/jn.00168.2006

Makin, T. R., Holmes, N. P., Brozzoli, C., Rossetti, Y., and Farne, A. (2009). Coding of visual space during motor preparation: approaching objects rapidly modulate corticospinal excitability in hand-centered coordinates. *J. Neurosci.* 29, 11841–11851. doi: 10.1523/JNEUROSCI.2955-09.2009

Makin, T. R., Holmes, N. P., and Zohary, E. (2007). Is that near my hand? multisensory representation of peripersonal space in human intraparietal sulcus. *J. Neurosci.* 27, 731–740. doi: 10.1523/JNEUROSCI.3653-06.2007

Pesaran, B., Nelson, M., and Andersen, R. (2006). Dorsal premotor neurons encode the relative position of the hand, eye, and goal during reach planning. *Neuron* 51:125. doi: 10.1016/j.neuron.2006.05.025

Pouget, A., and Sejnowski, T. (1997). Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.* 9, 222–237. doi: 10.1162/jocn.1997.9.2.222

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., and Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Exp. Brain Res.* 71, 491–507. doi: 10.1007/BF00248742

Rizzolatti, G., Scandolara, C., Matelli, M., and Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. ii. visual responses. *Behav. Brain Res.* 2, 147–163. doi: 10.1016/0166-4328(81)90053-X

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical areas. *Philos. Trans. R. Soc. Lond. B* 335, 11–21. doi: 10.1098/rstb.1992.0002

Rolls, E. T. (2008). *Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach, 1st Edn.* Oxford: Oxford University Press.

Rolls, E. T., and Milward, T. (2000). A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition and information-based performance measures. *Neural Comput.* 12, 2547–2572. doi: 10.1162/089976600300014845

Rolls, E. T., and Stringer, S. M. (2001). Invariant object recognition in the visual system with error correction and temporal difference learning. *Network* 12, 111–129. doi: 10.1080/net.12.2.111.129

Rolls, E. T., and Stringer, S. M. (2007). Invariant global motion recognition in the dorsal visual system: a unifying theory. *Neural Comput.* 19, 139–169. doi: 10.1162/neco.2007.19.1.139

Rolls, E. T., and Treves, A. (1990). The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* 1, 407–421. doi: 10.1088/0954-898X/1/4/002

Rolls, E. T., and Treves, A. (1998). *Neural Networks and Brain Function*. Oxford: Oxford University Press.

Rolls, E. T., Treves, A., and Tovee, M. J. (1997a). The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. *Exp. Brain Res.* 114, 177–185. doi: 10.1007/PL00005615

Rolls, E. T., Treves, A., Tovee, M. J., and Panzeri, S. (1997b). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J. Comput. Neurosci.* 4, 309–333.

Rolls, E. T., and Webb, T. J. (2014). Finding and recognizing objects in natural scenes: complementary computations in the dorsal and ventral visual systems. *Front. Comput. Neurosci.* 8:85. doi: 10.3389/fncom.2014.00085

Royer, S., and Paré, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* 422, 518–522. doi: 10.1038/nature01530

Stringer, S., Perry, G., Rolls, E., and Proske, J. (2006). Learning invariant object recognition in the visual system with continuous transformations. *Biol. Cybern.* 94, 128–142. doi: 10.1007/s00422-005-0030-z

Stringer, S., and Rolls, E. (2000). Position invariant recognition in the visual system with cluttered environments. *Neural Netw.* 13, 305–315. doi: 10.1016/S0893-6080(00)00017-4

Stringer, S., and Rolls, E. (2008). Learning transform invariant object recognition in the visual system with multiple stimuli present during training. *Neural Netw.* 21, 888–903. doi: 10.1016/j.neunet.2007.11.004

Stringer, S. M., Rolls, E. T., and Tromans, J. M. (2007). Invariant object recognition with trace learning and multiple stimuli present during training. *Network* 18:161. doi: 10.1080/09548980701556055

Wallis, G. M., and Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Prog. Neurobiol.* 51, 167–194. doi: 10.1016/S0301-0082(96)00054-8

Zipser, D., and Andersen, R. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* 331, 679–684. doi: 10.1038/331679a0

APPENDIX

A. VisNet Architecture and Parameters

A.1. VisNet

The VisNet model consists of a hierarchical series of four feedforward layers of competitive networks. Within each neuronal layer there is lateral competition between neurons implemented by local graded inhibition. During training, there is associative learning at the synaptic connections between the successive layers of neurons (See **Figure 1**). In VisNet, natural visual images are first passed through an array of filters mimicking the response properties of V1 simple cells, and subsequently these images are fed to the first layer of the network architecture. The forward connections to individual cells are derived from a topologically corresponding region of the preceding layer, using a Gaussian distribution of connection probabilities. These distributions are defined by a radius which will contain approximately 67% of the connections from the preceding layer. This leads to an increase in the receptive field size of neurons through successive layers of the network hierarchy. The network dimensions used for this study are shown in **Table A1**. The architecture captures the hierarchical organization of competitive neuronal layers that is common in both the dorsal and ventral visual systems.

The simulations were conducted utilizing an updated version of the VisNet model (Rolls and Milward, 2000; Rolls, 2008). Before presenting the stimuli to VisNet's input layer, they are pre-processed by an initial layer representing V1 with a dimension of 256×256 where each x, y-location contains a bank of Gabor filter outputs g corresponding to a hypercolumn generated by

$$g\left(x, y; \lambda, \theta, \psi, \sigma, \gamma\right) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \tag{A1}$$

$$x' = x \cos\theta + y \sin\theta \tag{A2}$$

$$y' = -x \sin\theta + y \cos\theta \tag{A3}$$

for all combinations of $\lambda = 2, \gamma = 0.5, \sigma = 0.56\lambda, \theta \in \{0, \pi/4, \pi/2, 3\pi/4\}$ and $\psi \in \{0, \pi, -\pi/2, \pi/2\}$.

The activation h_i of each neuron i in the network is set equal to a linear sum of the inputs y_j from afferent neurons j weighted by the synaptic weights w_{ij}. That is,

$$h_i = \sum_j w_{ij} y_j \tag{A4}$$

where y_j is the firing rate of the presynaptic neuron j in the preceding layer, and w_{ij} is the strength of the synapse from neuron j to neuron i.

Within each layer competition is graded rather than winner-take-all, and is implemented in two stages. First, to implement lateral inhibition the activation of neurons within a layer are convolved with a spatial filter, I, where δ controls the contrast and σ controls the width, and a and b index the distance away from the center of the filter

$$I_{a,b} = \begin{cases} -\delta e^{-\frac{a^2 + b^2}{\sigma^2}} & \text{if } a \neq 0 \text{ or } b \neq 0, \\ 1 - \sum_{\substack{a \neq 0 \\ b \neq 0}} I_{a,b} & \text{if } a = 0 \text{ and } b = 0. \end{cases} \tag{A5}$$

Typical lateral inhibition parameters are given in **Table A2**.

Next, contrast enhancement is applied by means of a sigmoid activation function

$$y = f^{sigmoid}(r) = \frac{1}{1 + e^{-2\beta(r - \alpha)}} \tag{A6}$$

where r is the activation (or firing rate) after lateral inhibition, y is the firing rate after contrast enhancement, and α and β are the sigmoid threshold and slope respectively. The parameters α and β are constant within each layer, although α is adjusted to control the sparseness of the firing rates. The sparseness a of the firing within a layer can be defined, by extending the binary notion of the proportion of neurons that are firing, as

$$a = \frac{(\sum\limits_{i=1}^{N} y_i/N)^2}{\sum\limits_{i=1}^{N} y_i^2/N} \tag{A7}$$

where y_i is the firing rate of the ith neuron in the set of N neurons (Rolls and Treves, 1990, 1998; Rolls, 2008). For the simplified case of neurons with binarised firing rates $= 0/1$, the sparseness is the proportion $\in [0, 1]$ of neurons that are active. For example, to set the sparseness to, say, 5%, the threshold is set to the value of the 95th percentile point of the activations within the layer. Typical parameters for the sigmoid activation function are shown in **Table A3**.

For these simulations we used a trace learning rule (Földiák, 1991; Rolls, 1992) to adjust the strengths of the feed-forward

TABLE A1 | Network dimensions showing the number of connections per neuron and the radius in the preceding layer from which 67% are received.

	Dimensions	Number of connections	Radius
Layer 4	32×32	100	12
Layer 3	32×32	100	9
Layer 2	32×32	100	12
Layer 1	64×64	100	12
Retina	$256 \times 256 \times 16$	–	–

TABLE A2 | Lateral inhibition parameters.

Layer	1	2	3	4
Radius, σ	1.38	2.7	4.0	6.0
Contrast, δ	1.5	1.5	1.6	1.4

TABLE A3 | The sigmoid parameters used to control the global inhibition within each layer of the model.

Layer	1	2	3	4
Percentile	99	98	88	95
Slope β	190	40	75	26

synaptic connections between the layers during training. The trace rule incorporates a trace $\bar{y}^{\tau}$ of recent neuronal activity into the postsynaptic term. The trace term reflects the recent activity of the postsynaptic cell. The effect of this is to encourage the postsynaptic cell to learn to respond to input patterns that tend to occur close together in time.

The equation of the original trace learning rule as used by Wallis and Rolls (1997) is the following

$$\Delta w_j = \alpha \bar{y}^{\tau} x_j^{\tau} \qquad (A8)$$

where the trace $\bar{y}^{\tau}$ is updated according to

$$\bar{y}^{\tau} = (1 - \eta) y^{\tau} + \eta \bar{y}^{\tau - 1} \qquad (A9)$$

and we have the following definitions

x_j:	j^{th} input to the neuron.	y:	Output from the neuron.
$\bar{y}^{\tau}$:	Trace value of the output of the neuron at time step τ.	α:	Learning rate. Annealed between unity and zero.
w_j:	Synaptic weight between j^{th} input and the neuron.	η:	Trace value. The optimal value varies with presentation sequence length.

The parameter η may be set in the interval $[0, 1]$. For our simulations the trace learning η is set to 0.8. If $\eta = 0$ then the Equation (A8) becomes the standard Hebb rule

$$\Delta w_j = \alpha y^{\tau} x_j^{\tau}. \qquad (A10)$$

However, the version of the trace rule used in this paper only includes the trace of activity from the immediately preceding timestep, as used in other studies (Rolls and Milward, 2000; Rolls and Stringer, 2001) for improving the performance of the standard trace rule and enhancing the effect of the invariance representation. Thus, the rule takes now the following form

$$\Delta w_j = \alpha \bar{y}^{\tau - 1} x_j^{\tau}. \qquad (A11)$$

Neuronal mechanisms that might support trace learning in the brain have been previously discussed (Rolls, 1992; Wallis and Rolls, 1997). To restrict and limit the growth of each neuron's synaptic weight vector, $\mathbf{w}_i$ for the ith neuron, its length is normalized at the end of each timestep during training as is usual in competitive learning (Hertz et al., 1991). Normalization is required to ensure that the same set of neurons do not always win the competition. Neurophysiological evidence for synaptic weight normalization has been presented (Royer and Paré, 2003).

A.2. Information Theory Measures

Single and multiple cell information theoretic measures are used to assess the network's performance. Both measures help to determine whether individual cells in the output layer are able to respond to a specific target location in a hand-centered frame of reference over a number of different retinal locations. In previous VisNet studies, the single cell information measure has been applied to individual cells in the last layer of the network and measures how much information is available from the response of a single cell about which stimulus was shown. In this current study, a stimulus is defined as one of the different hand-object configurations. If an output neuron responds to just one of the spatial configurations, and the cell responds to this configuration across all tested retinal locations, then the cell will convey maximal single cell information. The amount of information carried by a single cell about a stimulus is computed using the following formula

$$I(s, R) = \sum_{r \in R} P(r|s) \log_2 \frac{P(r|s)}{P(r)} \qquad (A12)$$

where the stimulus-specific information $I(s, R)$ is the amount of information the set of responses R of a single cell has about a specific stimulus (i.e., target location with respect to the hand) s, while the set of responses R corresponds to the firing rate y of a cell to each of the stimuli (i.e., hand-object configurations) presented in all tested retinal locations. Further details of how the single cell information is calculated are provided in the literature (Rolls et al., 1997a; Rolls and Milward, 2000; Rolls, 2008).

The maximum single cell information measure is

$$\text{Max. single cell info.} = \log_2(\text{Number of stimuli}). \qquad (A13)$$

For example, when we present 5 stimuli during testing, (i.e., spatial configurations of the hand and the test object), the maximum single cell information measure is 2.32 bits. When we present 6 target stimuli, the maximum single cell information measure is 2.58 bits. The cell reaches the maximal information when it responds selectively to just one of the hand-object spatial configurations, and responds to that spatial configuration across all the tested retinal positions.

On the other hand, the multiple-cell information computes the average amount of information about which stimulus was presented obtained from the responses of all the output cells. This procedure is used to verify whether, across the population of cells, there is information about all of testing stimuli (i.e., hand-object configurations) shown. Procedures for calculating the multiple cell information measure have been described in detail by Rolls et al. (1997b), Rolls and Milward (2000). In brief, from a single presentation of a stimulus, we calculate the average amount of information obtained from the responses of all the cells regarding which stimulus is shown. This is achieved through a decoding procedure that estimates which stimulus s' gives rise to the particular firing rate response vector on each trial. A probability table of the real stimuli s and the decoded stimuli s' is then constructed. From this probability table, the mutual

information is calculated as

$$I(S, S') = \sum_{s,s'} P(s, s') \log_2 \frac{P(s, s')}{P(s)P(s')}. \qquad \text{(A14)}$$

Multiple cell information values are calculated for the subset of cells which, according to the single cell analysis, have the most information about which stimulus (i.e., hand-object configuration) is shown. In particular, the multiple cell information is calculated from five cells for each stimulus that had the most single cell information about that stimulus. For example, in simulations with six target locations this results in a population of 30 cells. Previous research (Stringer and Rolls, 2000) found this to be a sufficiently large subset to demonstrate that shift invariant representations of each stimulus presented during testing were formed, and that each stimulus could be uniquely identified.

A.3. Data Sharing
The VisNet simulator can be downloaded from https://github.com/bedeho/VisBack.

Computational modeling of the neural representation of object shape in the primate ventral visual system

Akihiro Eguchi[1]*, Bedeho M. W. Mender[1], Benjamin D. Evans[1], Glyn W. Humphreys[2] and Simon M. Stringer[1]

[1] Department of Experimental Psychology, Oxford Centre for Theoretical Neuroscience and Artificial Intelligence, Oxford University, Oxford, UK, [2] Department of Experimental Psychology, Oxford Cognitive Neuropsychology Centre, Oxford University, Oxford, UK

Neurons in successive stages of the primate ventral visual pathway encode the spatial structure of visual objects. In this paper, we investigate through computer simulation how these cell firing properties may develop through unsupervised visually-guided learning. Individual neurons in the model are shown to exploit statistical regularity and temporal continuity of the visual inputs during training to learn firing properties that are similar to neurons in V4 and TEO. Neurons in V4 encode the conformation of boundary contour elements at a particular position within an object regardless of the location of the object on the retina, while neurons in TEO integrate information from multiple boundary contour elements. This representation goes beyond mere object recognition, in which neurons simply respond to the presence of a whole object, but provides an essential foundation from which the brain is subsequently able to recognize the whole object.

1. Introduction

1.1. Hierarchical Representations in the Primate Ventral Visual Pathway

Over successive stages of processing, the primate ventral visual pathway develops neurons that respond selectively to objects of increasingly complex visual form (Kobatake and Tanaka, 1994), going from simple orientated line segments in area V1 (Hubel and Wiesel, 1962) to whole objects or faces in the anterior inferotemporal cortex (TE) (Perrett et al., 1982; Tsunoda et al., 2001; Tsao et al., 2003). In addition, in higher layers of the ventral pathway, the responses of neurons to objects and faces show invariance to retinal location, size, and orientation (Tanaka et al., 1991; Rolls et al., 1992; Perrett and Oram, 1993; Rolls, 2000; Rolls and Deco, 2002). These later stages of processing carry out object recognition by integrating information from more elementary visual features represented in earlier layers (Brincat and Connor, 2004). Thus, in order to understand visual object recognition in the primate brain, we need also to understand the encoding of more elementary features in the early and middle stages of the ventral visual pathway. In particular, many theories suppose that object recognition operates through the computation of intermediate representations which reflect the spatial relations between the parts of objects (Giersch, 2001; Pasupathy and Connor, 2001; Brincat and Connor, 2004).

Experimental studies have shown that neurons in successive stages of the primate ventral visual pathway encode the spatial structure of visual objects and their parts. For example, single

unit recording studies carried out by Pasupathy and Connor (2001) have shown that, within an intermediate stage of the ventral visual pathway, area V4, there are neurons that respond selectively to the shape of a local boundary element at a particular position in the frame of reference of the object. Some of these V4 neurons also maintain their response properties as an object shifts across different locations on the retina (Pasupathy and Connor, 2002). Further experimental studies have shown that neurons in the later stages of the ventral visual pathway, TEO and posterior TE, integrate information from multiple boundary contour elements (Brincat and Connor, 2004). This representation of the detailed spatial form of the separate parts of each object may provide a necessary foundation for the subsequent recognition of whole objects. That is, object selective cells at the end of the ventral visual pathway may learn to respond to unique distributed representations of object shape in earlier areas (Booth and Rolls, 1998).

1.2. Computer Modeling Study

A number of modeling studies have tried to reproduce the observed shape selective and translation invariant firing properties of neurons in area V4 (Cadieu et al., 2007; Rodríguez-Sánchez and Tsotsos, 2012). However, these past models have not utilized biologically plausible local learning rules, which use pre- and post-synaptic cell quantities to drive modification of the synaptic connections during visually-guided learning. Therefore, it still remains a challenge to understand exactly how V4 neurons develop their shape selective response properties through learning. The purpose of this paper is to provide a biologically plausible theory of this learning process. More generally, we investigate through computer simulation how the cell firing properties reported in visual areas V4, TEO, and posterior TE may develop through visually-guided learning, and thus how the primate ventral visual pathway learns to represent the spatial structure of objects.

The simulation studies presented below are conducted with an established neural network model of the primate ventral visual

pathway, VisNet (Wallis and Rolls, 1997), shown in **Figure 1**. The standard network architecture consists of a hierarchy of four competitive neural layers (Rumelhart and Zipser, 1985) corresponding to successive stages of the ventral visual pathway. The VisNet architecture is feed-forward with lateral interactions within layers. Many engineering approaches to efficiently solve similar problems extensively rely their architectures on top-down information flows, mainly for their supervised learning. However, our aim is to pin down the simplest form of core-mechanisms in intermediate vision, that is sufficient to explain a specific brain function. In fact, in other feature hierarchical neural network modeling studies, such top-down information transfer is often excluded (Olshausen et al., 1993; Riesenhuber and Poggio, 1999; Serre et al., 2005, 2007; Wallis, 2013).

The researchers involved in these last publications acknowledge the extensive presence of such back projections in the visual cortex; however, they also think the exact roles of these projections still remain a matter of debate. For example, it has been proposed that the role of these feedback pathways is to relay the interpretations of higher cortical areas to lower cortical areas in order to verify the high-level interpretation of a scene (Mumford, 1992) or to refine the tuning characteristics of lower-level cortical cells based upon the interpretations made in higher cortical areas (Tsotsos, 1993). On the other hand, numerous physiological studies have also reported that only short time spans are required for various selective responses to appear in monkey IT cells, which imply that feedback processes may not be critical for coarse, rapid recognition (Perrett et al., 1992; Hung et al., 2005; vanRullen, 2008).

We also stand on the similar point of view, and learning mechanisms implemented in the current model are a direct extension of previous papers in the field (Rumelhart and Zipser, 1985). In our paper, we have applied these established learning mechanisms to the important new problem of how the primate ventral visual system learns to represent the shapes of objects.

FIGURE 1 | Left: Stylized image of the four layer VisNet architecture. **Right:** Convergence in the visual system. The diagonal lines show the convergence of feed-forward connections through successive layers of the ventral visual system leading to an increase in receptive field size from 1.3° in V1 to 50° in TE (Figures excerpted from Wallis and Rolls, 1997).
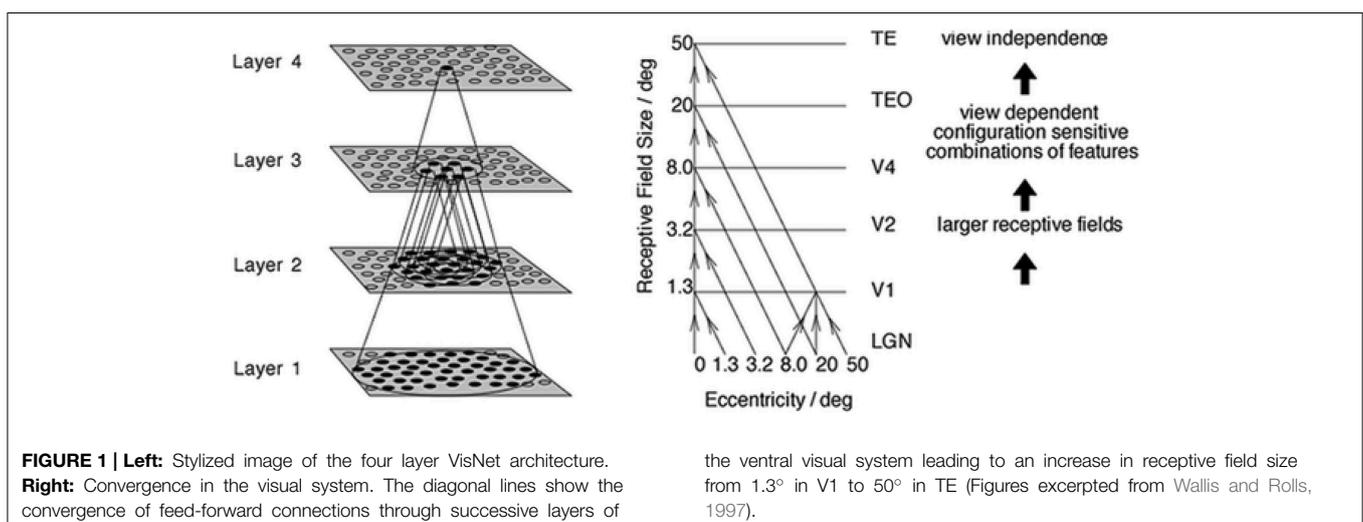
1.3. Hypothesis

In this paper, we consider how biologically plausible neuronal and synaptic learning mechanisms may be applied to the challenge of explaining (i) how neurons in V4 learn to respond selectively to the shape of localized boundary contour elements in the frame of reference of the object, (ii) how neurons in areas TEO and posterior TE learn to respond to localized combinations of boundary contour elements, and (iii) how these neurons learn to respond with translation invariance as the object is shifted through different retinal locations. In particular, we hypothesize that a biologically plausible solution may be provided by combining the statistical decoupling (Stringer et al., 2007; Stringer and Rolls, 2008) that will occur between different forms of boundary contour element over a large population of different object shapes, with the use of a temporal trace learning rule to modify synaptic weights as objects shift across different retinal locations (Wallis and Rolls, 1997; Rolls, 2000).

1.3.1. Neurons Learn to Respond to Individual Boundary Contour Elements by Exploiting Statistical Decoupling

In previous work, we have investigated how VisNet may learn transform invariant representations of individual objects if the network is always presented with multiple objects simultaneously during training (Stringer et al., 2007; Stringer and Rolls, 2008). We have found that if VisNet is trained on different combinations of objects on different occasions and as long as there are enough objects in the total pool of objects, this will result in statistical decoupling between any two objects. This statistical decoupling forces neurons in the higher competitive layers of VisNet to learn to respond to the individual objects, rather than the combinations of objects on which the network is actually trained.

This is because a competitive neural network has a capacity limit in terms of the number of object categories that can be represented in a non-overlapping manner in the output layer. **Figure 2A** provides some insight into the learning mechanisms driving the formation of neurons encoding individual object.

Consider the highly simplified situation where, a winner-take-all competitive network with $64 \times 64 = 4096$ output neurons is presented with n different objects, which are presented in pairs to VisNet during training. With winner-take-all competition, the network is able to develop 4096 non-overlapping output representations. **Figure 2A** shows how the number of individual objects, $y_1 = n$, and the number of possible objects comprised of pairs of objects $y_2 =_n C_2 = n(n-1)/2$, rise quadratically with increasing n. Because of this, y_2 reaches the capacity limit of the network much more quickly than y_1. Therefore, for $n > 91$, individual output neurons are forced to switch from representing the objects to representing the individual objects. Although, of course, the output layer as a whole will still provide unique representations of the pairs of objects, themselves, but in a distributed, overlapping manner.

We now propose that a similar learning mechanism may operate to enable the network to learn to represent the individual boundary contour elements within objects. For example, consider the simplified case shown in **Figure 3A**. This figure shows a set of four sided shapes, where each side has one of three possible conformations: concave, straight, or convex. Therefore, there are 4 sides $\times$ 3 side types $= 12$ different boundary contour elements (each defined by a unique combination of position and shape), which may be used to construct a total of $3^4 = 81$ different whole objects. We demonstrate that, when VisNet is trained on such a population of different object shapes constructed from different combinations of boundary contour elements, there is statistical decoupling between any two boundary contour elements.

Figure 2B provides an illustration of how the capacity limit forces output neurons to learn to represent individual boundary elements. **Figure 2B** (left) shows two different object shapes that share a boundary element at the bottom, which are presented to the network during training. Each of the two objects stimulates a subset of output neurons, and those neurons learn to represent each shape through associative learning in the feed-forward synaptic connections. The situation in the figure supposes that

FIGURE 2 | (A) The capacity limit of a competitive neural network forces individual output neurons to switch from representing object shapes to representing the boundary elements as the number of object shapes on

which the network is trained increases. **(B)** Illustration of how the network model develops neurons that have learned to respond to individual boundary elements of 2D object shapes.

FIGURE 3 | (A) Form of visual objects used to train and test VisNet for Study 1. Each object has a fixed number of sides (*n*), each of which has a fixed number of possible boundary conformations (*p*). **(B)** Form of visual objects used to train and test VisNet in Study 2.

These visual stimuli are similar to those used in the original neurophysiological experiments of Pasupathy and Connor (2001). **(C)** Examples of some of the realistic visual objects used for training VisNet in Study 3.

the subset of the neurons activated partly overlap. In this situation, the boundary element at the bottom becomes especially strongly associated with the subset of output neurons at the intersection of the two object shape representations. **Figure 2B** (right) shows that during testing this intersecting subset of output neurons will respond whenever the network is presented with an object shape containing the given boundary element. In this manner, without any top-down information transfer, the network should be able to develop representations of localized boundary elements. This kind of the distributed coding of 2D object shape, utilizing an alphabet of localized boundary elements, may be used to represent the shape of any object.

1.3.2. Neurons Develop Translation Invariant Responses Through Trace Learning (Temporal Association)

Another key property of the neurons reported by Pasupathy and Connor (2001) in area V4, and neurons reported by Brincat and Connor (2004) in areas TEO and posterior TE, is that they respond with translation invariance as an object shifts across different locations over the receptive field. The question is how these neurons might learn to respond in such a translation invariant manner?

One possible explanation is that the brain uses temporal associative learning to develop such transformation invariant representations. The theory assumes that, every now and then, a primate will make a series of fixations at different points on the same visual object before moving onto

another object; much experimental work has studied the statistics of saccades and fixations across natural visual scenes (Findlay and Gilchrist, 2003). Of particular relevance is how the eyes saccade around natural visual scenes containing multiple objects. Seminal psychophysical studies of how human subjects move their gaze around pictures of natural scenes were carried out by Yarbus (1967). It was indeed evident from this work that there was a tendency for observers to shift their fixation to a number of different points on a salient object, such as a person, before moving onto the next object.

Therefore, we assume that eye movements would be sufficiently small so that the same object is always projected within the simulated receptive field when learning it. We believe this constraint is reasonable to simulate recent physiological findings. For example, Li and DiCarlo (2008) conducted a study where monkeys are trained to track an object on a screen where a object with identity A is originally placed on the one of two possible retinal positions ($+3°$ or $-3°$) and later shifted to the center ($0°$). In the experimental condition, the identity of the object is swapped from A to B when it is shifted to the center, and the eyes saccade to it. As a results, individual neurons in primate IT that are originally translation-invariantly selective to identity A start to respond also to object with identity B only at the specific retinal location. This finding does not exclude the possibility of the temporal association learning which may occur at larger eye movement; however, it provided a reasonable evidence for the translation invariance learning mechanism within IT (Isik et al., 2012).

Accordingly, our proposed solution is *temporal trace learning* (Foldiak, 1991; Wallis and Rolls, 1997; Rolls and Milward, 2000). An example of such a learning rule is given in Section 2. If the eyes shift about a visual scene more rapidly than the objects change within the scene, then the images of an object in different locations on the retina will tend to be clustered together in time. In this case, a trace learning rule will encourage neurons in higher layers to learn to respond with translation invariance to specific objects or features across different retinal locations.

This rule is biologically plausible in terms of the way it utilizes only locally available biological quantities, that is, the present and recent activities of the pre- and post-synaptic neurons, respectively. Also, it has been shown that this type of temporal associative learning arises naturally within biophysically realistic spiking neural networks when longer time constants for synaptic conductance are introduced (Evans and Stringer, 2012).

Our past research has shown that this trace learning rule may be combined with the mechanism of statistical decoupling described above to produce translation invariant representations of statistically independent visual objects (Stringer et al., 2007; Stringer and Rolls, 2008). We hypothesize that the same trace learning rule could encourage neurons representing boundary contour elements to respond with translation invariance across different retinal locations.

1.4. Overview of Simulation Studies Carried Out in this Paper

Study 1 provides a proof-of-principle analysis. VisNet was trained on artificial visual objects similar to those shown in **Figure 3A**. These carefully constructed objects allowed us to explore how the statistical decoupling between different boundary contour elements influences the neuronal firing properties that develop during learning. We also showed how the capacity of the network to represent many different boundary contour conformations can be increased by introducing a Self-Organizing Map (SOM) architecture within each layer. Finally, we used the same artificial visual stimuli to confirm that trace learning can produce neurons that respond to individual boundary contour elements with translation invariance across different retinal locations.

In Study 2, the sets of visual stimuli presented to VisNet during training and testing were similar to those used in the original physiological experiments of Pasupathy and Connor (2001). Examples are shown in **Figure 3B**. This allowed for a direct comparison between the performance of the VisNet model and real neurons recorded in area V4 of the primate ventral visual pathway.

In Study 3, we trained VisNet on a large number of realistic visual objects with different boundary shapes. A sample of these objects is shown in **Figure 3C**. This generated a more realistic and demanding test of the underlying theory.

2. Materials and Methods

2.1. Hierarchical Neural Network Architecture of the Model

VisNet is a hierarchical neural network model of the primate ventral visual pathway, which was originally developed by Wallis and Rolls (1997). The standard network architecture is shown in **Figure 1**. It is based on the following: (i) a series of hierarchical competitive layers with local graded lateral inhibition. (ii) Convergent connections to each neuron from a topologically corresponding region of the preceding layer. (iii) Synaptic plasticity based on a biologically-plausible local learning rule such as the Hebb rule or trace rule, which are explained in Section 2.4.

In past work, the hierarchical series of four neuronal layers of VisNet have been related to the following successive stages of processing in the ventral visual pathway: V2, V4, the posterior inferior temporal cortex, and the anterior inferior temporal cortex. In this paper, we model for the first time neuronal response properties observed within a series of intermediate layers. Due to the relatively course-grained four-layer architecture of VisNet, we do not wish to emphasize a specific correspondence between the layers of VisNet and particular stages of the ventral pathway. However, as our main focus was on the neuronal properties reported in V4 and TEO, we mostly focused on the first three layers of VisNet.

In VisNet, the forward connections to individual cells are derived from a topologically corresponding region of the preceding layer, using a Gaussian distribution of connection probabilities. These distributions are defined by a radius which contained approximately 67% of the connections from the preceding layer. The values employed in the current studies are given in **Table 1**, which have been proposed to be realistic in Wallis and Rolls (1997). However, to deal with more complex images, the size of the layer was extended to 128 × 128 neurons from 32 × 32 neurons. The gradual increase in the receptive field of cells in successive layers reflects the known physiology of the primate ventral visual pathway (Pettet and Gilbert, 1992; Pasupathy, 2006; Freeman and Simoncelli, 2011).

2.2. Pre-processing of the Visual Input by Gabor Filters

Before the visual images are presented to VisNet's input layer 1, they are pre-processed by a set of input filters that accord with the general tuning profiles of simple cells in V1. The filters provide a unique pattern of filter outputs for each transform of each visual object, which is passed through to the first layer of VisNet. In this paper, the input filters were matching the firing properties of V1 simple cells, which respond to local oriented bars and edges within the visual field (Jones and Palmer, 1987; Cumming

TABLE 1 | VisNet parameters.

Layer	Dimensions	Number of connections	Radius
Layer 4	128 × 128	100	12
Layer 3	128 × 128	100	9
Layer 2	128 × 128	100	6
Layer 1	128 × 128	201	6
Retina	256 × 256 × 16		

and Parker, 1999). The input filters used are computed by the following equations (Daugman, 1985):

$$g(x, y, \lambda, \sigma, \theta, \psi, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \tag{1}$$

with the following definitions:

$$\begin{aligned} x' &= x \cos\theta + y \sin\theta \\ y' &= -x \sin\theta + y \cos\theta \end{aligned} \tag{2}$$

where x and y specify the position of a light impulse in the visual field (Petkov and Kruizinga, 1997). The parameter λ is the wavelength, σ is the standard deviation which is a function of λ and spatial bandwidth b, θ defines the orientation of the feature, ψ defines the phase offset, and γ sets the aspect ratio. In each experiment, an array of Gabor filters is generated at each of 256×256 retinal locations with the parameters given in **Table 2**.

The outputs of the Gabor filters are passed to the neurons in layer 1 of VisNet according to the synaptic connectivity given in **Table 1**. Each layer 1 neuron received connections from 201 randomly chosen Gabor filters localized within a topologically corresponding region of the retina. In the original VisNet model (Wallis and Rolls, 1997), the input filters were tuned to the four different spatial wavelengths 2, 4, 8, and 16 pixels. The shortest wavelength filters provided the highest resolution information about the image. The neurons in the first layer of VisNet were thus assigned most of their afferent inputs from the shortest wavelength filters. In the current simulations reported here, the model used inputs from only the shortest wavelength filters, which was found to be sufficient to represent the simple visual objects. For consistency with past VisNet simulations, each neuron in the first layer of VisNet received afferent connections from 201 of the short wavelength filters.

2.3. Calculation of Cell Activations within the Network
Within each of the neural layers 1–4 of the network, the activation h_i of each neuron i was set equal to a linear sum of the inputs y_j from afferent neurons j in the preceding layer weighted by the synaptic weights w_{ij}. That is,

$$h_i = \sum_j w_{ij} y_j \tag{3}$$

where y_j is the firing rate of neuron j, and w_{ij} is the strength of the synapse from neuron j to neuron i.

2.4. Lateral Interaction between Neurons Within each Layer
In the simulations reported below, the lateral interaction between the neurons within each neuronal layer was implemented in one of two different ways. The simplest approach was to implement a competitive network architecture (Rolls and Treves, 1998), in which neurons inhibited all of their neighbors. However, in some simulations we also implemented a more complex SOM architecture (Kohonen, 2000), which included both short range excitation and longer range inhibition between neurons (i.e., a "Mexican hat" connectivity). A SOM architecture leads to a map-like arrangement of neuronal response characteristics across a layer after training, with nearby cells responding to similar inputs. In particular, we investigated the hypothesis that the SOM architecture could increase the capacity of the network by enabling neurons in the higher layers to discriminate between more boundary contour shapes. Parameters shown in **Tables 3**, **4** were selected based on those that previously optimized performance (Rolls and Milward, 2000; Tromans et al., 2011).

2.4.1. Competitive Network Architecture
The original VisNet model implemented a competitive network within each layer. Within each layer, competition was graded rather than winner-take-all. To implement lateral competition, the activations h_i of neurons within a layer were convolved with a spatial filter, I_{ab}, where δ controlled the contrast and σ controlled the width, and a and b indexed the distance away from the center of the filter:

$$I_{a,b} = \begin{cases} -\delta \exp\left(-\frac{a^2 + b^2}{\sigma^2}\right) & a \neq 0 \text{ or } b \neq 0 \\ 1 - \displaystyle\sum_{a \neq 0, b \neq 0} I_{a,b} & a = 0 \text{ and } b = 0 \end{cases} \tag{4}$$

The lateral inhibition parameters for the competitive network architecture are given in **Table 3**.

TABLE 3 | Lateral inhibition parameters for the competitive network architecture.

Layer	1	2	3	4
Radius (σ)	1.38	2.7	4.0	6.0
Contrast (δ)	1.5	1.5	1.6	1.4

TABLE 4 | SOM parameters.

Layer	1	2	3	4
Excitatory radius (σ_E)	1.4	1.1	0.8	1.2
Excitatory contrast (δ_E)	5.35	33.15	117.57	120.12
Inhibitory radius (σ_I)	2.76	5.4	8.0	12.0
Inhibitory contrast (δ_I)	1.5	1.5	1.6	1.4

TABLE 2 | Parameters for Gabor input filters.

Parameter (Symbol)	Values
Wavelength(λ)	2
Spatial bandwidth (b)	1.5 octaves
Orientation(θ)	$0, \pi/4, \pi/2, 3\pi/4$
Phase shift (ψ)	0: white on black bar
	π: black on white bar
Aspect ratio (γ)	0.5

2.4.2. Self-organizing Map
In this paper, we have also run simulations with a SOM (von der Malsburg, 1973; Kohonen, 1982) implemented within each layer. In the case of the SOM architecture, short-range excitation and long-range inhibition are combined to form a Mexican-hat spatial profile and is constructed as a difference of two Gaussians as follows:

$$I_{a,b} = -\delta_I \exp\left(-\frac{a^2 + b^2}{\sigma_I^2}\right) + \delta_E \exp\left(-\frac{a^2 + b^2}{\sigma_E^2}\right) \quad (5)$$

To implement the SOM, the activations h_i of neurons within a layer were convolved with a spatial filter, I_{ab}, where δ_I controlled the inhibitory contrast and δ_E controlled the excitatory contrast. The width of the inhibitory radius was controlled by σ_I and the width of the excitatory radius by σ_E. The parameters a and b indexed the distance away from the center of the filter. The lateral inhibition and excitation parameters used in the SOM architecture are given in **Table 4**.

2.5. Contrast Enhancement of Neuronal Firing Rates within Each Layer
Next, the contrast between the activities of neurons with each layer was enhanced by passing the activations of the neurons through a sigmoid transfer function (Rolls and Treves, 1998) as follows:

$$y = f^{sigmoid}(r) = \frac{1}{1 + \exp\left(-2\beta(r - \alpha)\right)} \quad (6)$$

where r is the activation after applying the lateral competition or SOM filter, y is the firing rate after contrast enhancement, and α and β are the sigmoid threshold and slope, respectively. The parameters α and β are constant within each layer, although α is adjusted within each layer of neurons to control the sparseness of the firing rates. For example, to set the sparseness to 4%, the threshold is set to the value of the 96th percentile point of the activations within the layer. The parameters for the sigmoid activation function are shown in **Table 5**. These are the standard parameter values that have been used in past VisNet studies (Stringer et al., 2006, 2007; Stringer and Rolls, 2008).

2.6. Training the Network: Visually-guided Learning of Synaptic Weights
The outputs of the Gabor filters were passed to layer 1 of VisNet. Activity was then propagated sequentially through layers 2 to 4 using the same mechanisms at each layer. During training with visual objects, the strengths of the feed-forward synaptic connections between successive neuronal layers are modified by local learning rules, where the change

in the strength of a synapse depends on the current or recent activities of the pre- and post-synaptic neurons. Two such learning rules were implemented with different learning properties.

2.6.1. The Hebb Learning Rule
One simple well-known learning rule is the Hebb rule:

$$\delta w_{ij} = k r_i^\tau r_j^\tau \quad (7)$$

where δw_{ij} is the change of synaptic weight w_{ij} from pre-synaptic neuron j to post-synaptic neuron i, r_i^τ is the firing rate of post-synaptic neuron i at timestep τ, r_j^τ is the firing rate of pre-synaptic neuron j at timestep τ, and k is the learning rate constant.

2.6.2. The Trace Learning Rule
An alternative learning rule that, in addition to producing neurons that respond to individual contour elements, can also drive the development of translation invariant neuronal responses is the trace learning rule (Foldiak, 1991; Wallis and Rolls, 1997), which incorporates a memory trace of recent neuronal activity:

$$\delta w_{ij} = k \bar{r}_i^{\tau-1} r_j^\tau \quad (8)$$

where $\bar{r}_i^\tau$ is the trace value of the firing rate of post-synaptic neuron i at timestep τ. The trace term is updated at each timestep according to

$$\bar{r}_i^\tau = (1 - \eta) r_i^\tau + \eta \bar{r}_i^{\tau-1} \quad (9)$$

where η may be set anywhere in the interval $[0, 1]$, and for the simulations described below, η was set to 0.8. The effect of this learning rule is to encourage neurons to learn to respond to visual input patterns that tend to occur close together in time. If the eyes shift about a visual scene containing a static object, then the trace learning rule will tend to bind together successive images corresponding to that object in different retinal locations.

In our simulations, natural eye movements are simulated implicitly during training by shifting each visual object in turn across a number of retinal locations. That is, to simulate natural rapid eye movements during visual inspection of each object, the visual object itself is shifted across the retina. After an object shifted through all of the retinal locations, the next object was presented across the same locations.

To prevent the same few neurons always winning the competition, the synaptic weight vector $\mathbf{w}_i$ of each neuron i is renormalized to unit length after each learning update for each training pattern by setting

$$\mathbf{w}_i = \frac{\mathbf{w}_i}{||\mathbf{w}_i||} \quad (10)$$

where $||\mathbf{w}_i||$ is the length of the vector $\mathbf{w}_i$ given by

$$||\mathbf{w}_i|| = \sqrt{\sum_j w_{ij}^2} \quad (11)$$

TABLE 5 | Parameters for Sigmoid activation function.

Layer	1	2	3	4
Percentile	99.2	98	88	91
Slope (β)	190	40	75	26

2.7. Testing the Network

After the synaptic weights were established by training the network on a set of visual objects, the learned response properties of neurons through successive layers were tested. This was done by presenting visual objects constructed from a pool of different boundary contour elements, with the objects being similar or different to those used during training. A number of tests are applied to the recorded neuronal responses, including information theory, which are described below. We also analyzed the learned response properties of an output cell by plotting the subset of input Gabor filters with the strongest feed-forward connections to that output cell after training.

2.8. Information Analysis

To quantify the performance in transformation invariance learning with VisNet, the techniques of Shannon's information theory have previously been used (Rolls and Treves, 1998). In particular, a single cell information measure was applied to analyse the responses of individual cells. In order to keep the notation consistent with past publications (Rolls et al., 1997; Rolls and Milward, 2000), we have here denoted the neuronal firing rates by r.

To be informative in the context of this study, the responses of a given neuron (r) should be specific to a particular contour that appears at a particular side (s), and independent of the remaining global form of the object or retinal location. The amount of stimulus-specific information that a certain cell transmits is calculated from the following formula with details given by Rolls and Milward (2000).

$$I(s, \vec{R}) = \sum_{r \in \vec{R}} P(r|s) \log_2 \frac{P(r|s)}{P(r)} \tag{12}$$

Here s is a particular stimulus (i.e., a specific contour, at a specific side) and $\vec{R}$ is the set of responses of the cell to the set of objects that contain the contour at that particular side.

In past research with VisNet, this single-cell information analysis was used when only one object was presented to the network at a time. Therefore, the maximum information that an ideally developed cell could carry was log_2(number of stimuli). However, in this study, the complete object shape (composed of n contours) is presented. Therefore, this is conceptually equivalent to always presenting n stimuli simultaneously, thus altering the maximum attainable value of the single-cell information to $log_2(p)$ bits of information.

3. Results

3.1. Study 1: VisNet Simulations with Artificial Visual Objects Constructed from Multiple Boundary Elements

In Study 1, VisNet was trained on artificial visual objects similar to those shown in **Figure 3A**. For each simulation, these visual objects had a fixed number of sides (n), and the curvature of each side was selected from a fixed number of different boundary conformations or elements (p) and were projected on 256 × 256

pixels of simulated retina. Therefore, for each simulation there were p^n complete objects constructed from all combinations of the $n \times p$ contour elements. These artificially constructed objects allowed us to investigate how the learned neuronal response properties are affected by n and p. We then investigated the development of translation invariance as objects are shifted by 10 pixels at a time over a grid of four different locations on the retina by utilizing the trace learning mechanism discussed above.

3.1.1. Development of Neurons that Respond to Localized Boundary Conformation

We began by demonstrating how neurons in the output layer learn to respond to individual boundary contour elements when VisNet, implemented with competitive network, is trained on whole objects comprised of a number of such boundary elements. During training, the feed-forward synaptic connections were modified using the Hebb learning rule.

VisNet was first trained on a set of stimuli with $n = 3$ sides: top, left, and right. Each side has two possible boundary conformations: concave and convex. This gave a total of $2^3 = 8$ objects. As conceptually the third layer of VisNet may represent TEO, the VisNet architecture we used consisted of three competitive network layers in this simulation.

Figure 4A shows the learned responses y, given by Equation (6), of a typical output cell in layer 3 of VisNet, which developed selectivity to a concave contour situated at the top of each object after training; the criteria of the selectivity is whether the cell responds with a firing rate, r, approximately equal to 1 ($1.00000 \geq y \geq 0.99995$) across a set of whole objects containing a concave contour on the top while the cell responds with a firing rate approximately equal to 0 ($0.00005 > y \geq 0.00000$) across a set of whole objects not containing a concave contour on the top.

Figure 4A (top) shows a histogram of the average firing rate responses of the neuron to six (overlapping) subsets of objects, where each subset contains all those objects that incorporate a particular one of the six contour elements. **Figure 4A** (bottom) shows the actual subsets of objects that correspond to the six data points shown in the histogram. The results confirm that the neuron responds selectively.

Figure 4B shows the input Gabor filters that the same output cell in layer 3 has learned to respond to after training. In this case, the neuron receives the strongest inputs from a subset of Gabor filters that represent a concave contour on the top of each object. Such neuronal representations about each contour shape were found across the layer in the trained network. The distribution was quantified later in Sections 3.1.3 and 3.1.4.

3.1.2. How the Responses of Neurons to their Preferred Boundary Elements Depend on the Position of the Boundary Element in the Frame of Reference of the Object

Additional simulations investigated how the responses of neurons to their preferred boundary element depended on the position of the boundary element with respect to the object. In these simulations, VisNet, implemented with competitive networks, was trained on objects constructed with $n = 4$ sides: top, bottom, left, and right. Each side had $p = 3$ possible
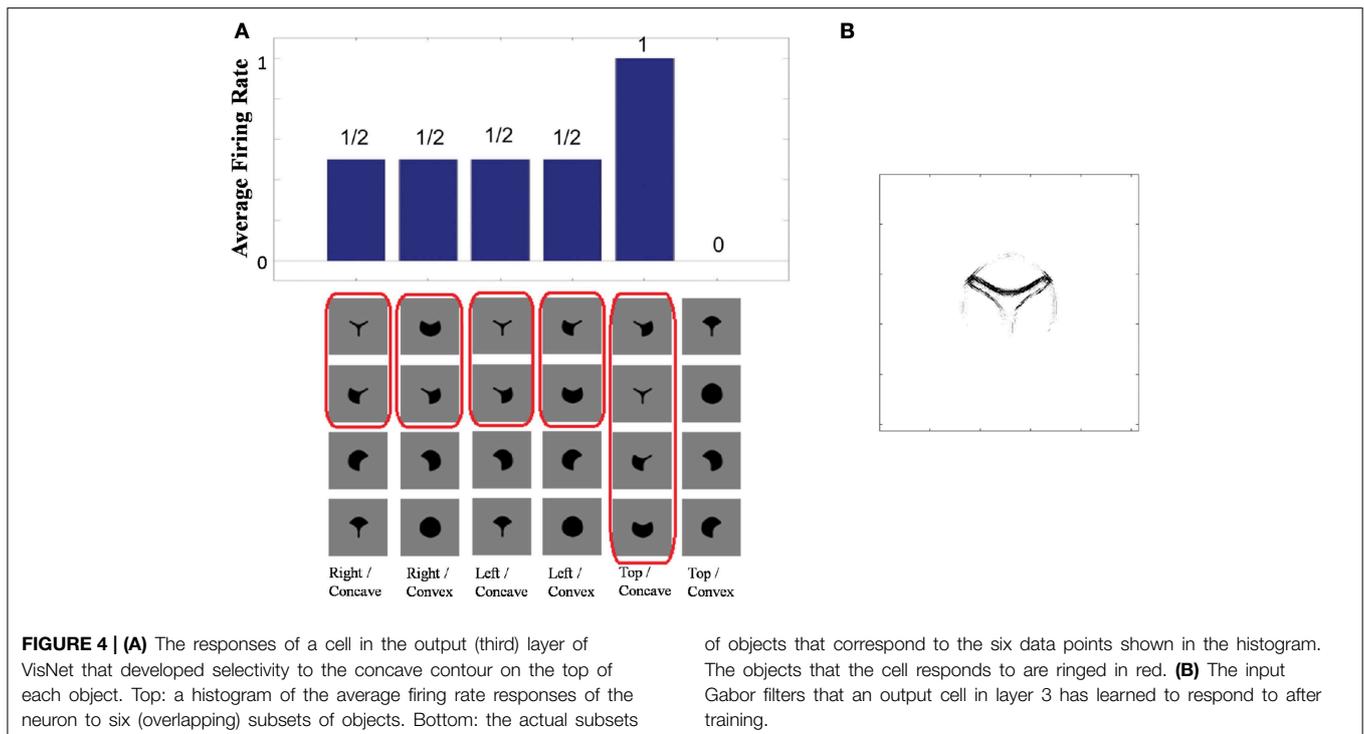
FIGURE 4 | (A) The responses of a cell in the output (third) layer of VisNet that developed selectivity to the concave contour on the top of each object. Top: a histogram of the average firing rate responses of the neuron to six (overlapping) subsets of objects. Bottom: the actual subsets of objects that correspond to the six data points shown in the histogram. The objects that the cell responds to are ringed in red. **(B)** The input Gabor filters that an output cell in layer 3 has learned to respond to after training.

boundary conformations: concave, straight and convex. During training, the feed-forward synaptic connections were modified using the Hebb learning rule.

VisNet was tested with two sets of objects. The first set contained those four-sided objects from the original training set that had at least one straight contour element, either on the right, bottom, left, or top. The second set contained mirror images of the first set of objects. The mirror images were constructed by reflecting the original trained objects around the retinal location of the vertical straight contour on the right of the training objects so that the vertical straight contours on the right and left of the two objects are aligned on the retina as shown in **Figure 5A**. If the neuron has learned about the local image context represented by nearby input filers, the neuron should respond only to the original images with a vertical straight contour on the right.

This effect is confirmed in **Figures 5B,C**. **Figure 5B** shows a histogram of the average firing rate response of the neuron to the four subsets of trained objects that contain a straight contour at one of the sides: right, bottom, left, and top (conventions as in **Figure 4A**). The histogram confirms that the neuron has learned to respond to a vertical straight contour on the right of each of the trained objects. **Figure 5C** shows similar results for the mirror image objects. Here it can be seen that the neuron fails to respond to any of the mirror image objects, including those mirror image objects with a vertical straight contour on the left.

Figure 5D shows the input Gabor filters that had strong connectivity through the layers to such a neuron. The plot is dominated by a strong vertical straight bar on the right hand side. This shows that the neuron has learned to respond to a straight contour on the right of each object. However, the activity of the neuron will also be influenced by other less strong filters shown in

the plot. These additional filters extend furthest to the left of the dominating vertical straight bar. In particular, the strong input filters to the left of the vertical straight bar represent boundary contour features that could co-occur within an object with the vertical straight contour on the right. The same is not true for the curve on the right of the vertical straight bar, which joins the same two vertices linked by the vertical straight bar and so would have to be an alternative contour element to the vertical straight bar. The effect of this pattern of additional input filters is that the neuron may require the presence of additional object contours to the left of the vertical straight contour in order for the neuron to respond. That is, the neuron will only respond to a vertical straight contour when that particular contour shape is on the right hand side of an object rather than the left of the object.

3.1.3. How the Number of Object Sides (n) and the Number of Possible Boundary Elements at Each Side (p) Affect the Learned Neuronal Response Properties

We investigated how the neuronal firing properties that develop in the network depend on the number of object sides (n) and the number of possible boundary contour elements (p) at each side. Each simulation was run with a fixed value of n and p. Across simulations, the number of sides, n, was varied from 3 to 8, while the number of possible boundary elements, p, was varied from 2 to 4. For each simulation, the network was trained on the full set of objects that could be constructed given the fixed values of n and p for that simulation; however, simulations with $p^n > 1000$ were omitted for practical reasons. During training, the feed-forward synaptic connections were modified using the Hebb learning rule within VisNet implemented with competitive networks.

FIGURE 5 | Demonstration of neuronal response tuning in an object centered frame of reference. Example stimuli used for testing are shown in **(A)**. **(B)** Shows a histogram of the average firing rate response of the neuron to the four subsets of trained objects that contain a straight contour at one of the sides: right, bottom, left, and top. **(C)** Shows similar results for the mirror image objects. **(D)** The input Gabor filters that had strong connectivity through the layers to a neuron that had learned to respond to a straight contour on the right of each object.

For each combination of n and p, **Figure 6A** (top) gives the number of neurons that learned to respond selectively to all objects that contained one particular type of boundary contour element, but not to objects that did not contain that boundary element.

It was found that the last layer of the untrained network already contained a small number of cells that were selective for objects that contained one type of boundary element. This was because this simulation task was relatively easy in that it did not require the output neurons to respond invariantly as objects were translated across different retinal locations. In simulations reported later in Section 3.1.6, the output neurons were tested with the objects presented in different retinal locations. In these simulations, training was indeed required to produce any neurons that responded selectively to objects containing one kind of boundary element.

In the trained network, it can be seen that all simulations produced large numbers of neurons that were selective for objects that contained one particular type of boundary element. Secondly, the number of object sides, n, did not have a significant systematic effect on the performance of the network. In contrast,

as the number of possible boundary elements at each side, p, increased, the number of neurons that learned to respond selectively to objects containing one type of boundary element declined.

We hypothesize that this is due to the effective increase in the density of the boundary contour elements at each side, which increases the difficulty of neurons in the higher layers developing separate representations of these more similar boundary conformations. In particular, an invariance learning mechanism known as Continuous Transformation (CT) learning (Stringer et al., 2006) may cause neurons in higher layers to learn to respond to a number of similar boundary conformations at each side; CT learning is able to bind smoothly varying input patterns, such as a continuum of different possible boundary conformations at one of the object sides, onto the same postsynaptic neuron. In this way, CT learning may dramatically reduce the selectivity of neurons for particular boundary conformations.

Typical network behavior for a relatively large value of p is shown in **Figure 7**. In this example, the network was trained on objects with $n = 3$ sides, each of which had $p = 4$

FIGURE 6 | Results of simulations in which VisNet is trained and tested on objects constructed with a fixed number of sides (*n*) and number of possible boundary elements at each side, *p*. (A) For each simulation, the table records the number of neurons (in the third layer) that learned to respond selectively to all objects that contained one particular type of boundary contour element, but not to objects that did not contain that

boundary element. Results are given before training (left) and after training (right). **(B)** Single cell information analysis results. For each simulation, the single cell information measures for all output (third) layer neurons are plotted in rank order according to how much information they carry. For each simulation, results are presented before training (dotted line) and after training with competitive network (broken dashed line) and with SOM (solid line).

possible boundary elements. The figure shows results for a typical output cell that failed to learn to respond selectively to objects containing one particular type of boundary contour. **Figure 7** (left) shows the input Gabor filters that had strong connectivity through the layers to the neuron. The neuron has strong connections from three similar boundary elements on the lower right. **Figure 7** (right) shows the average firing rate response of the neuron to the 12 subsets of objects that contain one of the different boundary elements. The neuron responds

maximally to the first three subsets of objects, which contain the three boundary elements that are strongly represented in the left plot. Thus, the neuron has learned to respond equally strongly to all of these three boundary elements and is unable to distinguish between them. This observed behavior is typical when the number (density) of boundary contour elements at each side is increased. Investigation into the responses of neurons across the output layer after the training the network on objects where each side had a relatively high number of possible boundary

FIGURE 7 | Simulation showing the failure of an output neuron to discriminate between a relatively large number of boundary contours at one object side. Left: The input Gabor filters that had strong connectivity through the layers to the output neuron. **Right:** histogram showing average firing rate response of the neuron to the 12 subsets of objects that contain one of the different boundary elements. That is, each of the data points (1–12) represents the average firing rate of the neuron across the 16 objects containing the following boundary elements: (1) right/sharp-convex, (2) right/convex, (3) right/concave, (4) right/sharp-concave, (5) left/sharp-convex, (6) left/convex, (7) left/concave, (8) left/sharp-concave, (9) top/sharp-convex, (10) top/convex, (11) top/concave, (12) top/sharp-concave.

element contours, p, showed that many cells were unable to distinguish between differently shaped contours on the same sides.

The simulations at this juncture show that a biologically plausible neural network can learn to code relative position information for visual elements, but has limited capacity. In the next section, we show how introducing a SOM architecture within each layer of VisNet can enhance the selectivity of neurons for individual boundary elements when the number of boundary elements at each side, p, is large, overcoming the capacity limitation.

3.1.4. The Effect of a Self-Organizing Map (SOM) Architecture on Learned Neural Selectivity for Boundary Contour Elements

We compared the performance of the standard competitive network architecture in each layer with performance when a SOM was introduced. We hypothesized that the SOM architecture could increase the capacity of the network to represent and distinguish between a larger number of finer variations in boundary contour curvature.

As discussed in the previous section, a competitive network may have difficulty in forming separate output representations of similar input patterns. In particular, CT learning (Stringer et al., 2006) may encourage the same output neurons to learn to respond to similar input patterns representing boundary contour elements of slightly different shape, or even bind together a continuum of input patterns covering the space of all possible boundary shapes at a particular object-centered boundary location.

The SOM architecture is specifically designed to encourage the output neurons to develop a fine-scaled representation of a continuum of smoothly varying input patterns (Kohonen, 2000). A SOM has additional short range lateral excitatory connections

between neurons within each layer. These connections encourage nearby output neurons to learn to respond to similar input patterns, which in turn leads to a map-like arrangement of neuronal response characteristics across the layer after training. In particular, slightly different input patterns will be distributed across different output neurons. Thus, the effect of these additional short range excitatory connections is to influence learning in the network to spread the representations of a continuum of overlapping input patterns over a map of output neurons. This should allow the network to develop a more fine-grained representation of the space of possible boundary contour shapes.

We therefore hypothesized that the introduction of a SOM architecture within each layer of VisNet would spread out the representations of many different boundary contour curvatures (p) at a particular side of the object over a map of output neurons. This would help to produce distinct neural representations of a large number of different boundary contour elements in the output layer, and effectively increase the capacity of the network to represent finer variations in boundary contour curvature.

During training, the feed-forward synaptic connections were again modified using the Hebb learning rule, and the simulation results with the SOM architecture implemented within each layer are presented in **Figure 6A** (bottom). The network was tested on objects constructed with a fixed number of sides, n, and different numbers of possible boundary elements at each side, p. For each simulation, the heatmap shows the number of neurons that learned to respond selectively to all objects that contained one particular type of boundary contour element, but not to other objects. These results should be compared with **Figure 6A** (top), which gives the corresponding results with a competitive network architecture implemented within each layer. As hypothesized, the introduction of SOM architecture within each layer led to many more neurons learning to respond selectively to objects containing a particular boundary contour element. This effect is particularly pronounced for larger numbers of n and p.

These effects can also be seen by examining the amount of information carried by neurons about the presence of particular types of boundary elements within the objects presented to VisNet. We have previously used information theoretic measures to assess the amount of information carried by neurons about the presence of whole object stimuli within a scene, where the objects may be presented under different transforms such as changes in retinal position or orientation (Wallis and Rolls, 1997; Rolls and Milward, 2000; Stringer et al., 2007; Stringer and Rolls, 2008). A neuron that responds selectively to one particular stimulus across a large number of transforms will carry a high level of information about the presence of that object within a scene. In this current paper, we were instead interested in the amount of information carried by neurons about the presence of particular boundary elements within an object.

Figure 6B present the single cell information analysis results for simulations in which VisNet was tested on objects with different numbers of sides, n, and numbers of possible boundary elements at each side, p. The results are presented before training (dotted line), after training with the competitive network architecture (broken dashed line) and with the SOM architecture

(solid line). The single cell information measures for all output layer neurons are plotted in rank order according to how much information they carry. In all simulations, training the network on the set of p^n whole objects led to many top layer neurons attaining the maximal level of single cell information of $log_2(p)$ bits. These results imply that training the network on the whole objects led to many output neurons learning to respond selectively to all of the objects that contained a particular one of the boundary contour elements, but not to objects that do not contain that boundary element. That is, these neurons had learned to respond to the presence of that particular boundary contour element within any object. In all simulations, many top layer neurons attained the maximal level of single cell information of $log_2(p)$ bits. However, consistent with our hypothesis, the incorporation of a SOM architecture typically led to a significant increase in the number of neurons that attained the maximal level of single cell information.

Furthermore, different sub-populations of cells that carry maximum single-cell information about each contour element were mapped onto the corresponding locations within the layer. This extended analysis has revealed that using a SOM led to a feature map as shown in **Figure 8**. This result was consistent with various physiological findings that indicate the topographic organization within ventral visual pathway (Larsson and Heeger, 2006; Hansen et al., 2007; Silver and Kastner, 2009).

3.1.5. Response properties of Neurons through Successive Layers of VisNet

We subsequently investigated how the response properties of neurons vary through successive layers of VisNet, which is implemented with SOM, before and after training. For all of the simulations performed, the feed-forward synaptic connections were modified using the Hebb learning rule.

Table 6 presents simulation results showing the responses of neurons through layers 1 to 3. The results are presented for a simulation with $n = 4$ sides and $p = 2$ contour elements per side and compared before and after training. Each sub-table

gives the number of neurons that responded selectively to either objects containing a single boundary element, objects containing a combination of two boundary elements, or a single whole object. It can be seen that, in all three layers, training the network led to a substantial increase in the number of neurons that responded to objects containing a single boundary element. The numbers of neurons that learned to respond to individual boundary elements increased through successive layers of VisNet.

For the simulation reported in **Table 6**, training did not lead to a similarly large increase in the numbers of neurons that responded to either a combination of two boundary elements, or a single whole object. This contrasts with experimental studies showing that neurons in the later stages of the ventral visual pathway, TEO and posterior TE, integrate information from multiple boundary contour elements (Brincat and Connor, 2004). We, therefore, investigated how neurons might learn to respond to localized clusters of boundary contour elements and also to whole objects. In fact, by examining the input Gabor filters that had a strong connectivity to these types of neuron, we were able to show that some neurons in VisNet were indeed learning to respond to either a combination of two boundary elements, or a whole object. These results are shown in **Figure 9A**.

Figure 9A compares the response properties of trained and untrained neurons in simulations with the SOM architecture. The network is presented with objects containing $n = 4$ sides, where each side has $p = 2$ possible boundary elements. Results are shown for four neurons. For each neuron, we show the input Gabor filters that had strong connectivity through the layers to the neuron (left), and a histogram showing average firing rate response of the neuron to the objects that contain one of the 8 boundary elements (right). The four neurons shown in the **Figure 9A** had the following characteristics. (top-left) A trained neuron that has learned to respond to a combination of two adjacent boundary contour elements: top convex and right convex. The Gabor filter plot shows that the feed-forward synaptic weights have been strengthened selectively from the two boundary elements only. (top-right) A trained neuron that has learned to respond to a whole object. The preferred object is comprised of two concave on top and right and two convex on bottom and left. The Gabor filter plot shows that the neuron has learned to respond to the complete set of boundary elements

FIGURE 8 | Simulation results demonstrating that the SOM architecture leads to a feature map in the output layer. (A) Left: contour plots showing the amount of single cell information carried by the 128×128 layer of output neurons for six boundary elements after training on objects with $n = 3$ and $p = 2$. The different colored contour plots correspond to the following boundary elements: top/convex (pink), top/concave (light green), right/convex (blue), right/concave (yellow), left/convex (light blue), and left/concave (red). **(B)** Right: similar results for the case $n = 3$ and $p = 3$.

TABLE 6 | Simulation results showing the responses of neurons through layers 1 to 3 with the SOM architecture.

Layer	1 Contour	2 Contours	Object
N4P2 experiment (SOM)			
UNTRAINED NETWORK			
3	856	270	538
2	418	104	82
1	293	0	0
TRAINED NETWORK			
3	4216	92	89
2	2440	10	10
1	540	8	0
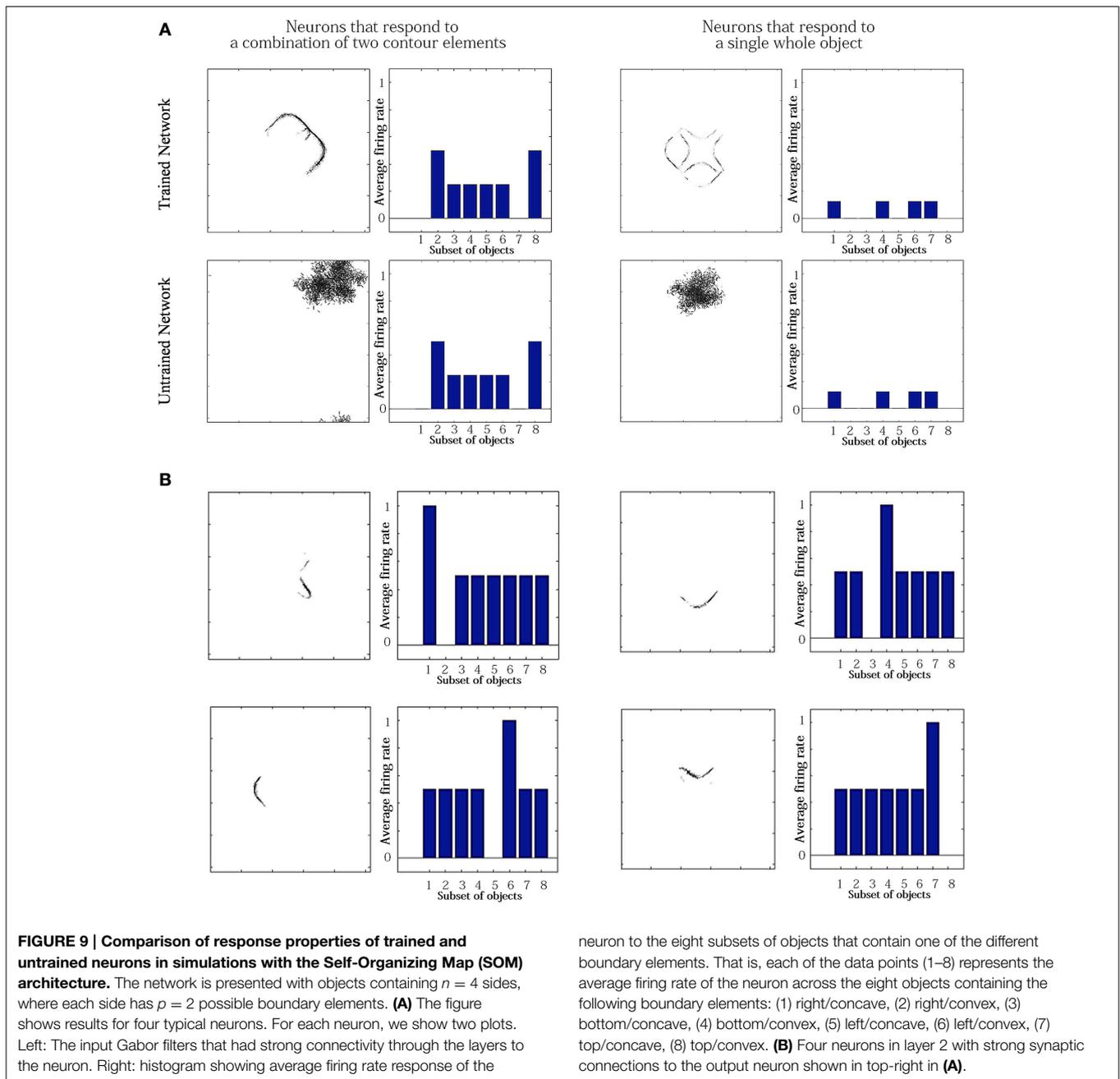
FIGURE 9 | Comparison of response properties of trained and untrained neurons in simulations with the Self-Organizing Map (SOM) architecture. The network is presented with objects containing $n = 4$ sides, where each side has $p = 2$ possible boundary elements. **(A)** The figure shows results for four typical neurons. For each neuron, we show two plots. Left: The input Gabor filters that had strong connectivity through the layers to the neuron. Right: histogram showing average firing rate response of the neuron to the eight subsets of objects that contain one of the different boundary elements. That is, each of the data points (1–8) represents the average firing rate of the neuron across the eight objects containing the following boundary elements: (1) right/concave, (2) right/convex, (3) bottom/concave, (4) bottom/convex, (5) left/concave, (6) left/convex, (7) top/concave, (8) top/convex. **(B)** Four neurons in layer 2 with strong synaptic connections to the output neuron shown in top-right in **(A)**.

comprising the preferred object. (bottom-left) An untrained neuron that happens to respond selectively during testing to two adjacent boundary elements. However, the Gabor filter plot shows that a random collection of Gabor filters have strong feed-forward connections to the neuron. This means that across a richer diversity of test images, this neuron would not maintain such a strict selectivity, and would in fact be most effectively stimulated by the random constellation of Gabor filters shown. (bottom-right) An untrained neuron that responds selectively to a whole object. The Gabor filter plot shows that the neuron receives strong connections from a random collection of Gabor

filters. This neuron would not maintain a strict selectivity to the object when tested on a greater diversity of images.

The conclusion of the results shown in **Figure 9A** is that although **Table 6** appeared not to show an increase during training in the numbers of neurons that responded to combinations of two boundary elements or a whole object, in fact training did lead to an increase in the numbers of neurons that had specifically learned to respond to whole stimuli. However, in **Table 6**, this effect had been masked by the existence of many untrained cells that already responded by chance to combinations of two boundary elements or a whole object, but which in fact

had random inputs from a large randomized collection of Gabor filters. Such untrained neurons are unlikely to be selective for combinations of two boundary elements or a particular object if the network were tested on a richer diversity of images. In particular, these untrained neurons would respond more selectively for images corresponding to the random constellations of Gabor filters shown in the bottom of **Figure 9A**. In contrast, the trained neurons on the top have strengthened connections specifically from combinations of two boundary elements or a whole object, and would therefore maintain their selectivity more robustly across a greater variety of test images.

We also found that output neurons in layer 3 learned to respond to whole objects by combining inputs from neurons in the preceding layer that responded to the individual boundary elements. This can be seen by examining the strengths of the synaptic connections from neurons in layer 2 to output neurons in layer 3 after training. Output neurons that had learned to respond to a particular object received the strongest synaptic connections from neurons in layer 2 that represented the constituent boundary elements of that object. **Figure 9B** shows four neurons in layer 2 with strong synaptic connections to a whole shape selective neuron reported in the top-left of **Figure 9A**. the output neuron shown in **Figure 9A**. Each of the four neurons in layer 2 had learned to respond to a different one of the boundary elements which were contained in the object that the output neuron had learned to respond to. This example shows that neurons in the later stages of the model are able to integrate information from multiple boundary contour elements, as consistent with neurophysiological results for areas TEO and posterior TE of the primate ventral visual pathway (Brincat and Connor, 2004).

3.1.6. Translation Invariance of Neuronal Responses as Objects are Shifted Across Different Locations on the Retina

The neurons reported by Pasupathy and Connor (2001) in area V4, and neurons reported by Brincat and Connor (2004) in areas TEO and posterior TE, respond with translation invariance as an object is shifted across different retinal locations. In this section we show how these translation invariant neuronal responses may be set up by training the network with the trace learning rule. The trace learning rule encourages individual postsynaptic neurons to learn to respond to subsets of input patterns that tend to occur close together in time. Therefore, in the simulation described below, during training we selected each object in turn and presented that object in a number of different retinal locations before moving on to the next object.

For this simulation, VisNet had four layers with a SOM architecture implemented within each layer. The visual objects had $n = 4$ sides, where each side has $p = 3$ possible boundary elements. Each of the visual objects was presented in a 2×2 grid of four different retinal locations, which were separated by horizontal and vertical shifts of 10 pixels.

Figure 10A shows the results after training for a typical output neuron in layer 4. **Figure 10B** shows the input Gabor filters that had strong connectivity through the layers to the output neuron. It can be seen that the neuron has strong connections from a
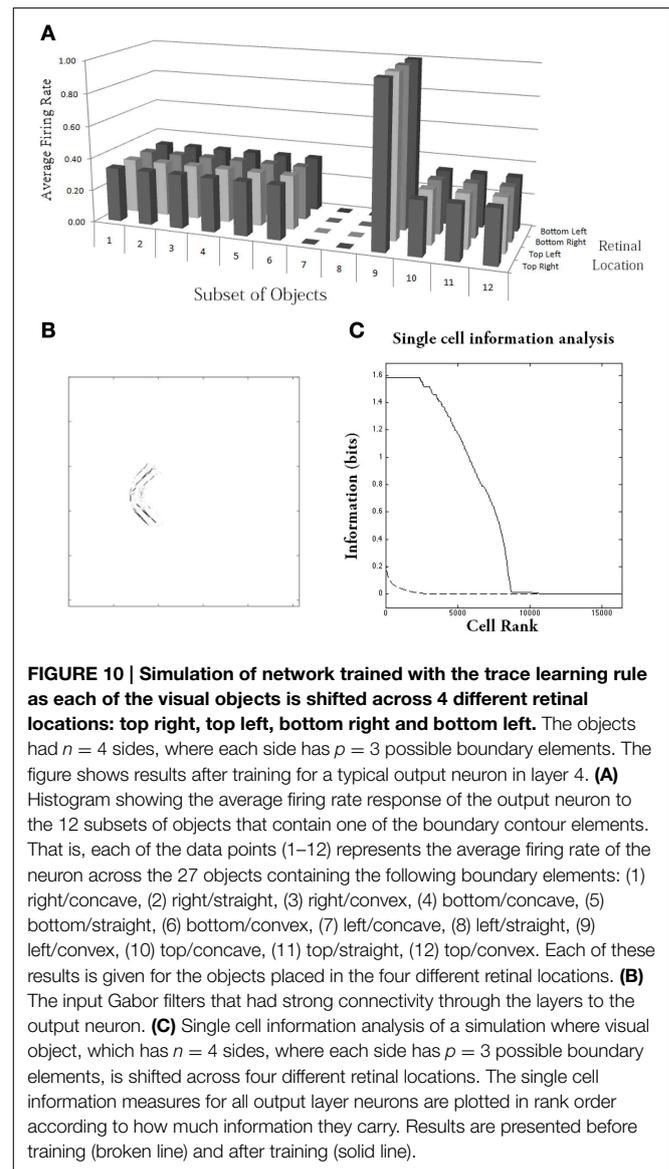
FIGURE 10 | Simulation of network trained with the trace learning rule as each of the visual objects is shifted across 4 different retinal locations: top right, top left, bottom right and bottom left. The objects had $n = 4$ sides, where each side has $p = 3$ possible boundary elements. The figure shows results after training for a typical output neuron in layer 4. **(A)** Histogram showing the average firing rate response of the output neuron to the 12 subsets of objects that contain one of the boundary contour elements. That is, each of the data points (1–12) represents the average firing rate of the neuron across the 27 objects containing the following boundary elements: (1) right/concave, (2) right/straight, (3) right/convex, (4) bottom/concave, (5) bottom/straight, (6) bottom/convex, (7) left/concave, (8) left/straight, (9) left/convex, (10) top/concave, (11) top/straight, (12) top/convex. Each of these results is given for the objects placed in the four different retinal locations. **(B)** The input Gabor filters that had strong connectivity through the layers to the output neuron. **(C)** Single cell information analysis of a simulation where visual object, which has $n = 4$ sides, where each side has $p = 3$ possible boundary elements, is shifted across four different retinal locations. The single cell information measures for all output layer neurons are plotted in rank order according to how much information they carry. Results are presented before training (broken line) and after training (solid line).

convex boundary element on the left of an object. The separate contours that can be seen in the plot correspond to the different retinal locations in which the objects are trained. **Figure 10A** shows a histogram presenting the average firing rate response of the output neuron to the 12 subsets of objects that contain one of the boundary contour elements. The neuron responds maximally to the subset of objects containing a convex boundary element on the left. Notably, the neuron responds maximally to this subset of objects over all four retinal locations. Thus, the neuron has learned to respond to objects containing the convex boundary element on the left regardless of where the object is presented on the retina. These translation invariant neuronal responses are a result of training the network with the trace learning rule.

Figure 10C shows findings from the single cell information analysis. The results are presented before training (broken line) and after training (solid line). Training the network on the set of p^n whole objects over the four retinal locations led to

many top layer neurons attaining the maximal level of single cell information of $log_2(p)$ bits. Neurons carrying maximal single cell information responded selectively to a subset of objects containing one particular type of boundary element, and with translation invariance as the objects also were shifted over all four retinal locations. In these simulations with translation invariance, the information is dramatically increased after training. This is because it is very unlikely for untrained neurons to both respond selectively to a single boundary contour element across all objects, and be able to respond with translation invariance as these objects are shifted across the retina. Therefore, training will lead to a much more significant difference between the performances of the untrained and trained networks.

3.2. Study 2: VisNet Simulations with Visual Stimuli of Pasupathy and Connor

In Study 2, the visual stimuli presented to VisNet were similar to the artificial stimuli used in the neurophysiological experiments of Pasupathy and Connor (2001) shown in **Figure 3B**. This allowed direct comparison between the learned response characteristics of the neurons in the VisNet model and the experimentally observed cell responses encoding local boundary information reported.

The stimuli were constructed by systematically combining sharp convex, medium convex, broad convex, medium concave and broad concave boundary elements to form closed shapes. We also vary the angular separations of the vertices used to construct the stimuli on 256×256 pixels of the simulated retina as shown in **Figure 3B**. Furthermore, we also rotated the visual stimuli through $360°$ in a single central location on the retina in steps of $10°$ during training to provide more natural visual training. This meant that there was not such a clean statistical decoupling between the boundary elements as for Study 1. Nevertheless, we expected that with the new objects used in Study 2 there would still be sufficient statistical decoupling between the boundary elements to ensure that the network developed neurons during visually guided learning that responded to a localized region of boundary curvature.

For all simulations in Study 2, the VisNet architecture consisted of three layers of SOM, where each layer is composed of 64×64 neurons. During training, the feed-forward synaptic weights are modified using the trace learing rule, which is needed to develop translation invariant neuronal responses.

3.2.1. Development of Neurons Encoding Local Boundary Conformation in an Object-centered Frame of Reference

Figure 11 shows a comparison between the responses of a neuron recorded in area V4 of the primate ventral visual pathway by Pasupathy and Connor (2001) and a neuron recorded from our simulation, which exhibits a similar degree of selectivity. The neuron recorded by Pasupathy and Connor (2001) responds selectively to object shapes with an acute convex curvature at the top right of the object. Many other neurons in the output layer of VisNet learned to respond selectively to particular combinations of local boundary curvature and position with respect to the center of mass of the object. The network accomplished this even

though the statistical independence of the boundary contour elements was not perfect.

To analyse the detailed firing properties of each output neuron and quantified the distributions, we recorded its response to all objects as they were rotated through $360°$. Next we segmented the boundary contour of each object into multiple elements based on the positions where the rate of change of the curvature exceeded a fixed threshold. This then enabled us to calculate the average response of the neuron to each particular combination of local boundary curvature and angular position where that boundary curvature appears, where the average is computed over all orientations of all objects. **Figure 12A** shows a heatmap of the average responses of the output neuron shown on the right of **Figure 11** to different combinations of boundary conformation and angular position. The result indicates that this neuron responds maximally to object shapes with an acute convex curvature at the top right. The correlation coefficient between the result and a predicted result of a modeled V4 neuron based on Gaussian distribution, which is tuned to acute contours at $70°$ is strong (0.798) and confirms the selectivity. **Figures 12B–D** show examples of different trained cells.

For each neuron, we then analyzed the number of local peaks in the heatmap of average firing rate against curvature and angular position, as shown in **Figure 12**. Specifically, for each neuron we counted the number of local peaks that were greater than 60% of the average firing rate across the heatmap. Before training, 176 cells had one peak, 98 cells had two peaks, 63 cells had three peaks, and 44 cells had four peaks. After training, the distributions were 319 cells, 460 cells, 414 cells, and 374 cells. (These distributions were significantly different, $\chi^2 = 17.58$, $df = 3$, $P \ll 0.01$.) Thus, training led to a large increase in the number of neurons that were selectively tuned to either one or just a few boundary contour elements. The simulation results also predict the existence of individual neurons that are tuned to boundary elements in multiple locations. Consistent with this, Brincat and Connor (2004) have reported that some neurons in TEO and posterior TE do indeed respond to the co-occurrence of multiple adjacent contour elements.

3.2.2. Development of Translation Invariant Neuronal Responses

Pasupathy and Connor (2001) and Brincat and Connor (2004) reported that neurons encoding the boundary conformation of objects also respond with translation invariance as an object is shifted across different retinal locations. In this section we confirm that neurons in VisNet also develop translation invariant responses when the network is trained on the stimuli shown in **Figure 3B**. To cope with the larger computational resource requirements, only the stimuli with an angular separation between vertices of $135°/135°/90°$ were used, and the size of the image was reduced to 128×128 pixels. During training, the trace learning rule was used to modify the synaptic weights.

In this simulation, during training each object was shifted across a 3×3 grid of nine different retinal locations, which are separated by horizontal and vertical intervals of 10 pixels. At each pixel location, the objects are presented in all orientations

FIGURE 11 | Comparison between the single neuron recording data of Pasupathy and Connor (2001) and corresponding results from VisNet simulations. On the left of the figure are shown the responses of a neuron recorded in area V4 of the primate ventral visual pathway and shown in Figure 5A of Pasupathy and Connor (2001). Each object shape shown to the monkey is represented by a white icon, and the firing rate response of the neuron is represented by the surrounding shading with high firing denoted by black. Each row shows a different object shape, with each column corresponding to a different orientation of the object. It can be seen that the neuron responds selectively to object shapes with an acute convex curvature at the top right of the object. On the right of the figure are shown corresponding results for an output cell in layer 3 of VisNet, which has learned to respond with similar selectivity.

through 0°–360° in 10° steps. This means that during training the objects underwent two different kinds of transformation, both translation and rotation. We assume that typically the eyes shift about a visual scene more rapidly than the objects rotate on the retina. To simulate this effect, VisNet was trained as follows. During training, the orientation of each object was kept fixed at some initial angle while the object was shifted across all of the different retinal locations. Then the orientation of the object was adjusted by, for example, 10° and the object was again shifted across all of the retinal locations. This procedure was repeated for all object orientations from 0° to 360° in steps of 10°. This training procedure ensured that images of each object in the same orientation but different retinal locations were closely clustered together in time.

Figure 13 shows results for a typical output neuron after training. Each subplot shows the average responses of the neuron to different combinations of local boundary curvature and angular position. The top subplot shows the average neuronal responses over all nine retinal locations, while the remaining subplots show the average neuronal responses to each of the nine separate retinal locations.

In order to quantify the distribution of such cells, the number of peaks of responses for each cell were calculated. Before training, 91 cells had one peak, 61 cells had two peaks, 34 cells had three peaks, and 24 cells had four peaks. After training, the distributions were 288 cells, 253 cells, 119 cells, and 158 cells. (These distributions were significantly different, $\chi^2 = 1.99e+03$, $df = 3$, $P \ll 0.01$.)

FIGURE 12 | Heatmap showing the average responses of four output neurons to different combinations of local boundary curvature and angular position where the boundary curvature appears. The average is computed over all orientations (0°–360°) of all objects. **(A)** The neuron responds maximally to object shapes with an acute convex curvature at the top-right. This is the same neuron that was shown on the right of **Figure 11**. **(B–D)** Three other cells that show different firing patterns are also plotted to show the variability in the network.

It is evident that the neuron displays a pattern of selectivity for boundary curvature and angular position that is similar across the nine retinal locations. Thus, the responses of the neuron exhibit translational invariance, similar to the neurons reported in the neurophysiology experiments of Pasupathy and Connor (2001) and Brincat and Connor (2004).

3.3. Study 3: VisNet Simulations with Images of Natural Objects

In Study 3, VisNet was trained with images of natural objects in order to demonstrate that the learning mechanisms elucidated in this paper and tested with artificially constructed visual stimuli in sections of Study 1 and 2 will indeed work effectively on real world visual objects. We hypothesize that across many images of natural objects with different boundary shapes, there will be an effective statistical decoupling between localized boundary elements, which are defined by local curvature and angular position with respect to the center of mass of the object. This should force the neurons in higher layers of the network to learn to respond to the individual boundary elements rather than the whole objects.

Some examples of the natural objects used in these simulations are shown in **Figure 3C**. The set of stimuli used in the simulations is composed of 177 realistic three dimensional objects. Various kinds of three dimensional objects are downloaded from Google 3D Warehouse, converted into gray-scaled images, and rescaled to fit on the center of 256×256 retina. In order to enhance the realism of the visual images used to train VisNet, during training each of the natural objects is rotated in plane through 360° in steps of 10°. After training, the neuronal responses in

the network were examined with the test stimuli used for Study 2 (**Figure 3B**).

3.3.1. Development of Neurons Encoding Local Boundary Conformation in an Object-centered Frame of Reference

Figure 14 shows the responses of a typical output neuron after training. This neuron learned to respond to an acute convex curvature at the bottom left of an object. Moreover, although not shown, many other neurons in the output layer of VisNet learned to respond selectively to particular combinations of local boundary curvature and angular position of the boundary element.

In order to quantify the distribution of such cells, the number of peaks of responses for each cell were calculated. Before training, 176 cells had one peak, 98 cells had two peaks, 63 cells had three peaks, and 44 cells had four peaks. After training, the distributions were 232 cells, 141 cells, 125 cells, and 103 cells. (These distributions were significantly different, $\chi^2 = 176.82$, $df = 3$, $P \ll 0.01$.)

This result showed that VisNet was able to develop these neuronal responses even though the network had been trained on many natural visual objects without artificially constructing the boundary shapes from artificially predefined elements.

3.3.2. Development of Translation Invariant Neuronal Responses

We then tested whether neurons in VisNet can also develop translation invariant responses when the network was trained on the natural objects shown in **Figure 3C**. Each of the natural objects was shifted across a 3×3 grid of nine different retinal locations, which were separated by horizontal and vertical intervals of 10 pixels. At each pixel location, the objects were presented in different orientations through 0°–360° in 10° steps. The temporal sequencing of these two kinds of transforms was the same as described in Section 3.2.2. During training, the trace learning rule was used to modify the synaptic weights.

Figure 15 shows results for a typical output neuron after training. Each subplot shows the average responses of the neuron to different combinations of local boundary curvature and angular position. The top subplot shows the average neuronal responses over all nine retinal locations, while the remaining subplots show the average neuronal responses to each of the nine separate retinal locations. It can be seen that the neuron responds selectively to objects with a high convex curvature at the top-left. Moreover, the responses of the neuron are similar across all nine retinal locations.

In order to quantify the distribution of such cells, the number of peaks of responses for each cell were calculated. The distributions were that before training, 97 cells had one peak, 38 cells had two peaks, 25 cells had three peaks, and 31 cells had four peaks, whereas after training, the distributions were 349 cells, 148 cells, 90 cells, and 109 cells. (These distributions were significantly different, $\chi^2 = 1.34\text{e} + 03$, $df = 3$, $P \ll 0.01$). Thus, the responses of the neuron are reasonably translation invariant, similar to the neurons reported in the neurophysiology
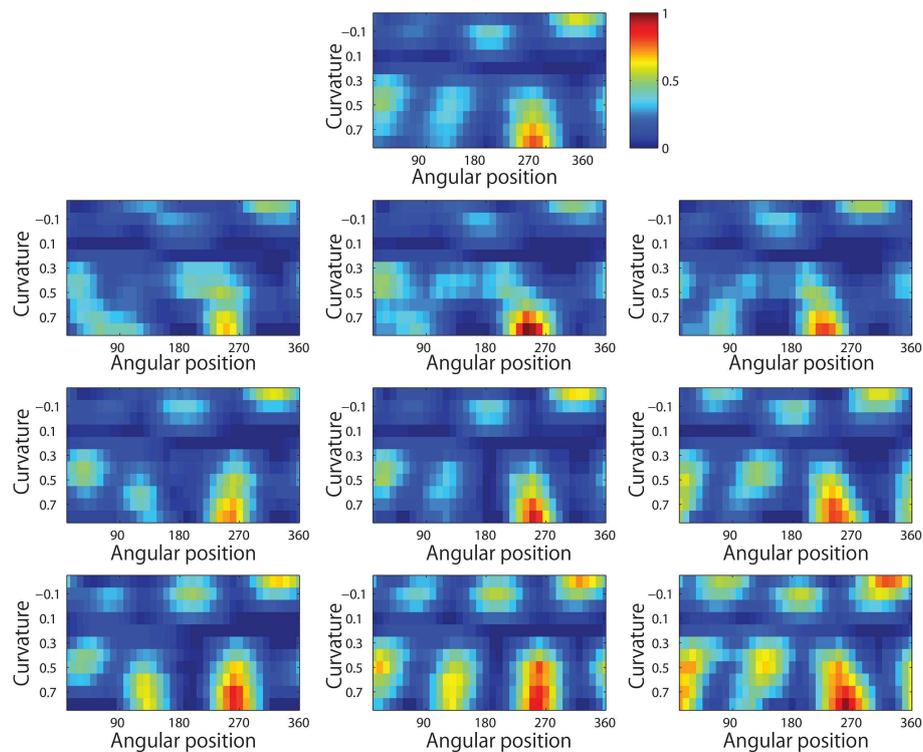
FIGURE 13 | Demonstration of translation invariance after training the network on the visual stimuli shown in Figure 3B. The figure shows the average responses of a typical output neuron over the nine retinal locations after training with the trace learning rule. Each of the ten subplots shows the average responses of the neuron to different combinations of local boundary curvature and angular position where the boundary curvature appears. The top subplot shows the average neuronal responses computed over all nine retinal locations. While the bottom subplots show the average neuronal responses computed separately for each of the nine retinal locations.

experiments of Pasupathy and Connor (2001) and Brincat and Connor (2004).

In conclusion, the above results thus demonstrate that even when VisNet is trained on realistic natural visual objects, where the boundary shapes have not been carefully constructed from a pool of artificial elements, the network still develops neurons that respond selectively to the curvature and location of localized boundary contour elements in the frame of reference of the object. Moreover, with the help of the trace learning rule, these neuronal responses are also translation invariant as an object shifts across different retinal locations.

4. Discussion

In this paper, we have demonstrated that when a neural network model, VisNet, of the primate ventral visual pathway is trained on many objects with different boundary shapes, the neurons in the higher layers of the network learn to respond to localized boundary contour elements, which are defined by the curvature and location of the boundary element in the frame of reference of the object. Interestingly, neurons learn to respond to these boundary elements rather than learning to respond to the whole objects that were actually presented during training. Moreover, the neurons were able to learn to respond with translation

invariance as visual objects are shifted across different retinal locations. This was shown to be successful when VisNet was trained with either the artificially constructed visual stimuli used in Studies 1 and 2, or with images of natural visual objects in Study 3.

The primary contribution of this paper is to elucidate and test two key biologically plausible learning mechanisms that can combine to promote the development of these neuronal response characteristics. First, similar to the results shown in the previous study with multiple-objects (Stringer et al., 2007; Stringer and Rolls, 2008), if the network is trained on many objects with different boundary shapes, where each boundary is comprised of a different constellation of contour elements, then this leads to a statistical decoupling between the boundary elements. This is sufficient to allow the competitive layers of VisNet to develop neurons that respond to individual boundary elements defined by curvature and position within the object, which are similar to the neurons reported in the physiological experiments conducted by Pasupathy and Connor (2001). Secondly, consistent with previous simulation studies (Wallis and Rolls, 1997; Rolls and Milward, 2000), neurons learned to respond with translation invariance across different retinal locations through the use of a trace learning rule. This kind of learning places constraints on the statistics of how the eyes move and visual objects change or
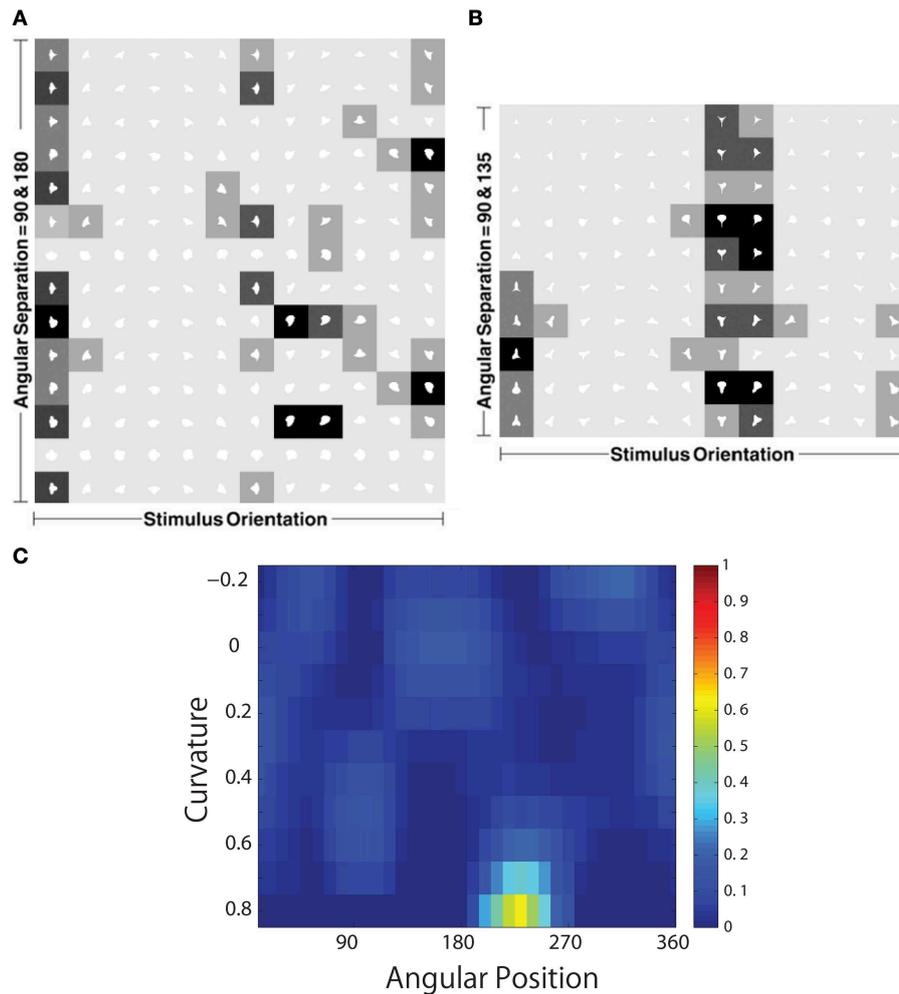
FIGURE 14 | The development of neuronal responses in VisNet that encode combinations of local boundary curvature and rotational position after the network has been trained with images of natural objects as shown in Figure 3C. The figure shows the responses of an output neuron which has learned to respond to an acute convex curvature at the bottom left (225°) of an object. **(A,B)** Shows the responses of the neuron to objects with an angular separation between the vertices of 135°/135°/90° and 180°/90°/90°, respectively. Each

object shape is represented by a white icon, and the firing rate response of the neuron is represented by the surrounding shading with high firing denoted by black. Each row shows a different object shape, with each column corresponding to a different orientation of the object. **(C)** Shows a heatmap of the average responses of the neuron to different combinations of local boundary curvature and angular position where the boundary curvature appears. The average is computed over all orientations (0°–360°) of all objects.

transform on the retina. These two mechanisms together provide a biologically plausible account of how neurons in the primate ventral visual pathway may learn to represent localized boundary contour elements of objects as revealed by Pasupathy and Connor (2001).

Furthermore, neurophysiological experiments carried out by Brincat and Connor (2004) have shown that neurons in the later stages of the ventral visual pathway, TEO and posterior TE, integrate information from multiple boundary contour elements. In our simulations, the number of cells that were tuned to combinations of multiple contours increased in the higher layers. Tracing back the feed-forward synaptic connectivity to these output neurons confirmed that their selectivities were

built by combining inputs from neurons representing each local boundary contour in the preceding layer.

The simulations reported in this present work are the first to show how neuronal responses encoding the local boundary conformation of objects may develop through a biologically plausible process of visually-guided learning. Both the Hebb learning rule and trace learning rule used above are biologically plausible in that they are "local" learning rules, which only use locally available biological quantities, such as the activity of the pre- and post-synaptic neurons, to modify the synaptic weights. This is in sharp contrast to other modeling studies that manually set up the synaptic weights in a non-local manner. In particular, the trace learning rule drives the development of translation
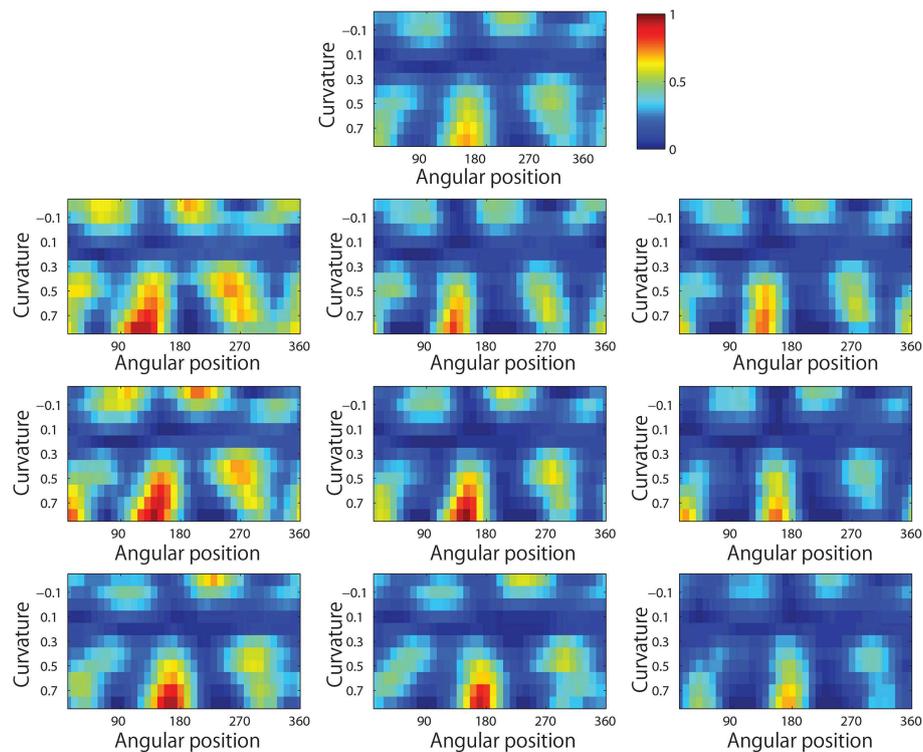
FIGURE 15 | The development of translation invariant neuronal responses in the output layer of VisNet after the network has been trained with images of natural objects as shown in Figure 3C. The network was trained with the trace learning rule in order to promote invariance learning across nine different retinal locations. The figure shows the average responses of a typical output neuron over the nine retinal locations after training. Each of the ten subplots shows the average responses of the neuron to different combinations of local boundary curvature and angular position where the boundary curvature appears. The top subplot shows the average neuronal responses computed over all nine retinal locations. While the bottom subplots show the average neuronal responses computed separately for each of the nine retinal locations.

invariant neuronal responses. Convincing experimental evidence for the presence of trace learning in the primate visual system has been provided by Cox et al. (2005), and a plausible account of the synaptic basis of trace learning has been provided by simulations of biologically detailed integrate and fire neural networks carried out by Evans and Stringer (2012). Furthermore, the trace learning rule can be implemented in the afferent synaptic connections to all neuronal layers in the network, which avoids the biologically implausible need for separate layers for template learning and invariance learning as has been implemented in previous models. Another important factor that underpins the biological plausibility of the simulations carried out in this paper is that the network model was always trained on whole objects rather than carefully pre-segmented and isolated parts of objects corresponding to local boundary elements. Indeed, in Study 3, VisNet was trained on a random assortment of whole natural visual objects. Nevertheless, the network was still able to develop neurons that were specifically tuned to localized boundary segments of objects. We also found the performance of the model to be extremely robust, which gives additional credence to the learning mechanisms explored in this paper.

4.1. Future Work

The version of the VisNet architecture used in this paper incorporated associative learning only in the bottom-up (feed-forward) connections between successive layers of the network. Furthermore, no top-down connections were included in the model even though these are known to exist in the primate ventral visual pathway. The rationale for using this simplified architecture in the current study was that it is sufficient to replicate how neurons in V4, TEO, and posterior TE are able to learn to encode the conformation of boundary contour elements at a particular position within an object. However, Zhou et al. (2000) have shown that the responses of neurons in earlier stages of visual processing such as V1 and V2, which have preferred responses to oriented edges, are also modulated by which side of a figure the edge occurs on. This is the case even when the figure/background cues lie well-outside the classical receptive field of the neuron. This suggests that global image context specifying border ownership modulates the activity these neurons. This contextual information must be conveyed to these early stage visual neurons by some combination of top-down connections between layers and recurrent connections within layers.

Another question is whether the approach proposed here can be extended to 3D shape. Yamane et al. (2008) have demonstrated the existence of neurons that encode the 3D configuration of localized surface fragments defined by their conformation, orientation and position with respect to the center of mass of the object. A population of such neurons provides a distributed representation of an object's 3D shape. The response characteristics of these neurons are also invariant as the object is shifted through different locations on the retina. It will be important to evaluate if a model such as VisNet, trained using stereoscopic input, can begin to capture the partonomic structure of 3D objects. Furthermore, it will be critical to assess whether learning rules, such as trace learning, can still be used to generate translationally invariant recognition processes.

However, theorists have long posited that the visual system in fact represents complex three-dimensional shapes, such as a table or a chair, by decomposing it into volumetric parts with axial symmetry (Biederman, 1987). A recent fMRI study in humans has provided evidence for this at the level of the neuronal population, where it was found that the visual system explicitly represents the relationships between the medial axes of linked object parts (Lescroart and Biederman, 2013). Consequently, more recently, Hung et al. (2012) have investigated medial axis shape coding in the inferotemporal cortex. This work extended their studies of parts-based spatial representations to "skeletal" representations involving a configuration of volumetric parts, where each part has an axis of radial symmetry or medial axis. The three-dimensional structure of an object may then be represented by a combination of the relationships between the medial axes of the object parts as well as the conformations of the surfaces of the object parts. Hung et al. (2012) confirmed that individual neurons in IT do in fact encode a configuration of both medial axis and surface fragments. In future work, we shall investigate whether the computational learning mechanisms demonstrated in this paper may also give rise to these kinds of skeletal representations.

References

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147. doi: 10.1037/0033-295X.94.2.115

Booth, M. C., and Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. cortex* 8, 510–523. doi: 10.1093/cercor/8.6.510

Brincat, S. L., and Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7, 880–886. doi: 10.1038/nn1278

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., and Poggio, T. (2007). A model of v4 shape selectivity and invariance. *J. Neurophysiol.* 98, 1733–1750. doi: 10.1152/jn.01265.2006

Cox, D. D., Meier, P., Oertelt, N., and DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nat. Neurosci.* 8, 1145–1147. doi: 10.1038/nn1519

Cumming, B. G., and Parker, A. J. (1999). Binocular neurons in v1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.* 19, 5602–5618.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* 2, 1160–1169. doi: 10.1364/JOSAA.2.001160

Evans, B. D., and Stringer, S. M. (2012). Transformation-invariant visual representations in self-organizing spiking neural networks. *Front. Comput. Neurosci.* 6:46. doi: 10.3389/fncom.2012.00046

Findlay, J. M., and Gilchrist, I. D. (2003). "Natural scenes and activities," in *Active Vision: The Psychology of Looking and Seeing,* Vol. 26 (Oxford : Oxford University Press), 129–150. doi: 10.1093/acprof:oso/9780198524793.001.0001

Foldiak, P. (1991). Learning invariance from transformation sequences. *Neural Comput.* 3, 194–200. doi: 10.1162/neco.1991.3.2.194

Freeman, J., and Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nat. Neurosci.* 14, 1195–1201. doi: 10.1038/nn.2889

Giersch, A. (2001). The effects of lorazepam on visual integration processes: how useful for neuroscientists? *Vis. Cogn.* 8, 549–563. doi: 10.1080/13506280143000115

Hansen, K. A., Kay, K. N., and Gallant, J. L. (2007). Topographic organization in and near human visual area v4. *J. Neurosci.* 27, 11896–11911. doi: 10.1523/JNEUROSCI.2991-07.2007

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154.2. doi: 10.1113/jphysiol.1962.sp006837

Hung, C. C., Carlson, E. T., and Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron* 74, 1099–1113. doi: 10.1016/j.neuron.2012.04.029

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866. doi: 10.1126/science.1117593

Isik, L., Leibo, J. Z., and Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Front. Comput. Neurosci.* 6:37. doi: 10.3389/fncom.2012.00037

Jones, J. P., and Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58, 1187–1211.

Kobatake, E., and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* 71, 856–867.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69. doi: 10.1007/BF00337288

Kohonen, T. (2000). *Self-Organizing Maps, 3rd Edn.* New York, NY: Springer.

Larsson, J., and Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *J. Neurosci.* 26, 13128–13142. doi: 10.1523/JNEUROSCI.1657-06.2006

Lescroart, M. D., and Biederman, I. (2013). Cortical representation of medial axis structure. *Cereb. Cortex* 23, 629–637. doi: 10.1093/cercor/bhs046

Li, N., and DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* 321, 1502–1507. doi: 10.1126/science.1160028

Mumford, D. (1992). On the computational architecture of the neocortex. *Biol. Cybern.* 66, 241–251. doi: 10.1007/BF00198477

Olshausen, B. A., Anderson, C. H., and Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* 13, 4700–4719.

Pasupathy, A. (2006). Neural basis of shape representation in the primate brain. *Progr. Brain Res.* 154, 293–313. doi: 10.1016/S0079-6123(06)54016-6

Pasupathy, A., and Connor, C. E. (2001). Shape representation in area v4: position-specific tuning for boundary conformation. *J. Neurophysiol.* 86, 2505–2519. Available online at: http://jn.physiology.org/content/86/5/2505.article-info

Pasupathy, A., and Connor, C. E. (2002). Population coding of shape in area v4. *Nat. Neurosci.* 5, 1332–1338. doi: 10.1038/972

Perrett, D. I., Hietanen, J. K., Oram, M. W., and Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 23–30. doi: 10.1098/rstb.1992.0003

Perrett, D. I., and Oram, M. W. (1993). Neurophysiology of shape processing. *Image Vis. Comput.* 11, 317–333. doi: 10.1016/0262-8856(93)90011-5

Perrett, D. I., Rolls, E. T., and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* 47, 329–342. doi: 10.1007/BF00239352

Petkov, N., and Kruizinga, P. (1997). Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. *Biol. Cybern.* 76, 83–96. doi: 10.1007/s004220050323

Pettet, M. W., and Gilbert, C. D. (1992). Dynamic changes in receptive-field size in cat primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 89, 8366–8370. doi: 10.1073/pnas.89.17.8366

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Rodríguez-Sánchez, A. J., and Tsotsos, J. K. (2012). The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. *PLoS ONE* 7:e42058. doi: 10.1371/journal.pone.0042058

Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27, 205–218. doi: 10.1016/S0896-6273(00)00030-1

Rolls, E. T., Cowey, A., and Bruce, V. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas [and discussion]. *Philos. Trans. Biol. Sci.* 335, 11–21. doi: 10.1098/rstb.1992.0002

Rolls, E. T., and Deco, G. (2002). *Computational Neuroscience of Vision, 1st Edn.* Oxford: Oxford University Press.

Rolls, E. T., and Milward, T. (2000). A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Comput.* 12, 2547–2572. doi: 10.1162/089976600300014845

Rolls, E. T., and Treves, A. (1998). *Neural Networks and Brain Function, , 1st Edn.* Oxford: Oxford University Press.

Rolls, E. T., Treves, A., Tovee, M. J., and Panzeri, S. (1997). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J. Comput. Neurosci.* 4, 309–333. doi: 10.1023/A:1008899916425

Rumelhart, D. E., and Zipser, D. (1985). Feature discovery by competitive learning*. *Cogn. Sci.* 9, 75–112. doi: 10.1207/s15516709cog0901.5

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Poggio, T. (2005). *A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex.* Cambridge, MA: MIT CSAIL.

Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429. doi: 10.1073/pnas.0700622104

Silver, M. A., and Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cogn. Sci.* 13, 488–495. doi: 10.1016/j.tics.2009.08.005

Stringer, S. M., Perry, G., Rolls, E. T., and Proske, J. H. (2006). Learning invariant object recognition in the visual system with continuous transformations. *Biol. Cybern.* 94, 128–142. doi: 10.1007/s00422-005-0030-z

Stringer, S. M., and Rolls, E. T. (2008). Learning transform invariant object recognition in the visual system with multiple stimuli present during training. *Neural Netw.* 21, 888–903. doi: 10.1016/j.neunet.2007.11.004

Stringer, S. M., Rolls, E. T., and Tromans, J. M. (2007). Invariant object recognition with trace learning and multiple stimuli present during training. *Network* 18, 161–187. doi: 10.1080/09548980701556055

Tanaka, K., Saito, H., Fukada, Y., and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* 66, 170–189.

Tromans, J. M., Harris, M., and Stringer, S. M. (2011). A computational model of the development of separate representations of facial identity and expression in the primate visual system. *PLoS ONE* 6:e25616. doi: 10.1371/journal.pone.0025616

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* 6, 989–995. doi: 10.1038/nn1111

Tsotsos, J. K. (1993). "An inhibitory beam for attentional selection," in *Proceedings of the 1991 York Conference on Spacial Vision in Humans and Robots*, (New York, NY: Cambridge University Press), 313–331.

Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.* 4, 832–838. doi: 10.1038/90547

vanRullen, R. (2008). The power of the feed-forward sweep. *Adv. Cogn. Psychol.* 3, 167–176. doi: 10.2478/v10053-008-0022-3

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 14, 85–100.

Wallis, G. (2013). Toward a unified model of face and object recognition in the human visual system. *Front. Psychol.* 4:497. doi: 10.3389/fpsyg.2013.00497

Wallis, G., and Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progr. Neurobiol.* 51, 167–194. doi: 10.1016/S0301-0082(96)00054-8

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., and Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360. doi: 10.1038/nn.2202

Yarbus, A. L. (1967). *Eye Movements During Perception of Complex Objects.* New York, NY: Springer US.

Zhou, H., Friedman, H. S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.* 20, 6594–6611.

Enhanced HMAX model with feedforward feature learning for multiclass categorization

Yinlin Li[1], Wei Wu[1], Bo Zhang[2]* and Fengfu Li[2]

[1] State Key Lab of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [2] Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

In recent years, the interdisciplinary research between neuroscience and computer vision has promoted the development in both fields. Many biologically inspired visual models are proposed, and among them, the Hierarchical Max-pooling model (HMAX) is a feedforward model mimicking the structures and functions of V1 to posterior inferotemporal (PIT) layer of the primate visual cortex, which could generate a series of position- and scale- invariant features. However, it could be improved with attention modulation and memory processing, which are two important properties of the primate visual cortex. Thus, in this paper, based on recent biological research on the primate visual cortex, we still mimic the first 100–150 ms of visual cognition to enhance the HMAX model, which mainly focuses on the unsupervised feedforward feature learning process. The main modifications are as follows: (1) To mimic the attention modulation mechanism of V1 layer, a bottom-up saliency map is computed in the S1 layer of the HMAX model, which can support the initial feature extraction for memory processing; (2) To mimic the learning, clustering and short-term memory to long-term memory conversion abilities of V2 and IT, an unsupervised iterative clustering method is used to learn clusters with multiscale middle level patches, which are taken as long-term memory; (3) Inspired by the multiple feature encoding mode of the primate visual cortex, information including color, orientation, and spatial position are encoded in different layers of the HMAX model progressively. By adding a softmax layer at the top of the model, multiclass categorization experiments can be conducted, and the results on Caltech101 show that the enhanced model with a smaller memory size exhibits higher accuracy than the original HMAX model, and could also achieve better accuracy than other unsupervised feature learning methods in multiclass categorization task.

Keywords: HMAX, biologically inspired, feedforward, saliency map, middle level patch learning, feature encoding, multiclass categorization

1. Introduction

Image categorization is a critical issue in computer vision and neuroscience research. As the natural images have a lot of variations in lighting, scale, shape, position and occlusion, extracting intrinsic features, which are not only invariant within same class but also discriminative between different classes, is the principle of the algorithms for image categorization. And the mechanisms and

structures of the visual cortex, which support the robust recognition, are also the key points of neuroscience for visual cognition research. Traditional computer vision algorithms are far from perfect due to the aforementioned variations, while the visual system of the primates shows good performance in daily life. Thus, mimicking the structures, mechanisms and functions of the primate visual cortex to design visual algorithms will highlight computer vision researches, help to get an insight of the visual cortex and further promote the interdisciplinary study of computer vision and neuroscience.

In the last decades, many kinds of features have been proposed to represent the natural images in the field of computer vision. On the one hand, many global image representation methods are proposed, such as the subspace analysis methods including Principal Components Analysis (PCA) (Turk and Pentland, 1991) and Fishers Linear Discriminant Analysis (LDA) (Belhumeur et al., 1997), which can achieve compact holistic encoding but cannot deal well with partial occlusion or strong view changes; On the other hand, many elaborated local feature representation methods are designed, such as SIFT (Lowe, 2004) and SUFT (Bay et al., 2008), which are scale-invariant and robust to moderate viewpoint variations.

Moreover, a middle level representation method—Bag of Words (BoW) (Sivic and Zisserman, 2003), has achieved good performance for image-level classification. It extracts a collection of unordered local patches of a test image, and maps them to discrete visual words learned by k-means vector quantization (VQ), and then obtains a histogram feature vector for classification. As the BoW model does not encode spatial information, it can be invariant to position and pose, but lose discrimination in some conditions. In Lazebnik et al. (2006), Spatial Pyramid Matching (SPM) kernel is introduced to BoW, in which spatial information are encoded in different scales and better performance is obtained in scene classification task.

When compared with primate visual cortex, a majority of the traditional methods could be called as *flat processing methods*, in which features are designed and processed by task-dependent learning algorithms (Krüger et al., 2013), but the primate visual cortex is organized in a hierarchical structure, and has good generality and robustness in a various of visual tasks.

Thus, it could be meaningful to mimic primate visual cortex to design hierarchical computer vision algorithms. In this interdisciplinary research field, the groundbreaking work is the Nobel Prize work of Hubel and Wiesel (1959, 1962). Based on biological experiments on cats striate cortex (V1), they described a circuit model with simple cells and complex cells, in which the complex cell has a similar response characteristic as the simple cell, but has a larger receptive field and a higher level tolerance to variations. After that, many biologically inspired computational models for visual cognition are proposed, including the Neocognitron (Fukushima, 1988), the saliency-based visual attention model (Itti et al., 1998) and the HMAX model (Riesenhuber and Poggio, 1999; Serre et al., 2007), etc. Among them, the HMAX model is a feedforwad hierarchical feature learning model for classification task. It tries to mimic

the structures and functions of the ventral stream of the primate visual cortex in the first 100–150 ms of visual cognition, and includes four layers (S1, C1, S2, C2) corresponding to the V1 to PIT layers of the primate visual cortex. By alternating between convolution operation in S layers and max-pooling operation in C layers, the model finally generates a set of position- and scale-invariant features.

However, the HMAX model has its shortages. Firstly, a random patch/prototype sampling method in C1 layer is used. The representation and discrimination ability of these patches are not guaranteed, and it doesn't mimic the higher level learning ability of the visual cortex (Gross, 2008; López-Aranda et al., 2009). Secondly, the model is only designed for binary classification task. A high feature dimension will be generated for its application in multiclass categorization task, as patches need to be sampled in each object class respectively, which decreases its generalization ability and is different from the memory process of the visual cortex (Gross, 2008; Tyler et al., 2013).

In recent years, many researchers tried to modify the HMAX model to improve its performance or introduce more biological mechanisms into it. Mutch and Lowe (2006), Huang et al. (2011b) refined the model with sparsification, lateral inhibition and feedback based feature selection for image classification. While Mutch and Lowe (2006) achieved patch selection based on the weights of SVM classifier, and Huang et al. (2011b) used a boosting method to learn discriminative patch. Both of them didn't consider the possibility of learning patch in an unsupervised manner. Walther et al. (2002) merged the saliency-based attention model (Itti et al., 1998) with the HMAX model to modify the response characteristics of the S2 layer, while we will try to introduce attention modulation in an early layer S1 to support the patch learning in the next layer (C1). Thériault et al. (2013) extended the coding and pooling mechanisms of the HMAX model with more scale and spatial information for robust image classification, but it didn't achieve patch learning as the original HMAX model. In addition, other modifications of the HMAX model demonstrated good performance in face recognition (Liao et al., 2013; Qiao et al., 2014a,b), scene classification (Huang et al., 2011a), and handwritten digit recognition (Hamidi and Borji, 2010). The corresponding properties of the HMAX and the BoW model to the human visual cortex were also investigated by Ramakrishnan et al. (2015).

Meanwhile, Deep Neural Networks (DNN), such as the Convolutional Deep Belief Network (CDBN) (Lee et al., 2009) and the Convolutional Neural Networks (CNN) (Krizhevsky et al., 2012), are also organized in a hierarchical mode. Although their correspondences to the structures and mechanisms of the visual cortex are not quite clear, they have shown good performance in image categorization task. However, these models are difficult to train because very large training sets are required to avoid overfitting, and most of the CNN models with the best performance (Girshick et al., 2014; Schroff et al., 2015) are supervised models.

Thus, in this paper, based on related biological researches (see more details in Section 2), we mainly focus on the first

100–150 ms feedforward feature learning process of the primate visual cortex (Lamme and Roelfsema, 2000; Pascual-Leone and Walsh, 2001) to extend the original HMAX model in the following aspects:

(1) **Attention Modulation:** To mimic the bottom-up attention modulation (Theeuwes, 2010; Zhang et al., 2012) and the response characteristics of neurons in V1 layer (Chatterjee and Callaway, 2003; Donk and van Zoest, 2008), a saliency map is computed by combing the orientation and Lab color-space information together in the S1 layer of the HMAX model based on the concept of local feature contrast. The generated salient regions mainly correspond to salient objects, and the boundary and resolution of objects are well kept. The proposed salient regions are taken as the initial candidate regions for feature extraction.

(2) **Memory Processing:** To mimic the conversion of short-term memory to long-term memory of V2 (López-Aranda et al., 2009) and the learning, selectivity and clustering ability in distributed regions of inferotemporal cortex (IT) (Gross, 2008), a memory processing method with two steps is proposed to replace the random prototype sampling method in the HMAX model. Firstly, multiscale middle level patches are densely extracted in the salient regions. Secondly, the patches are selected with an unsupervised iterative clustering method. During which, rare and meaningless patches are deleted, and similar patches are grouped in the same cluster, and a classifier for each cluster is also learned. Thus, each cluster can be taken as a distributed region of IT layer, which contains neurons with similar selectivity of memory. Furthermore, the patches in each cluster mainly correspond to critical parts of objects, which are discriminative and representative. Due to the unsupervised learning mode, similar patches from different objects are shared, which would support the multiclass categorization task with less memory.

(3) **Feature Encoding and Multiclass Categorization:** Corresponding to the distributed memory regions with similar discrimination ability (Gross, 2008), the Gaussian-like operation in S2 layer of the HMAX model is replaced by classification operation of each cluster. To mimic the feature encoding in Milner and Goodale (2008), the maximal activation of each cluster in the C2 layer of the HMAX model and its relative spatial position are cascaded as the final feature vector. Softmax is taken as the decision layer for multiclass categorization, and each output corresponds to the distributed associated regions of different objects for visual cognition in the cortex (Tyler et al., 2013).

The remaining parts of this paper are organized as follows. In Section 2, the related biological researches supporting the work of this paper are discussed. In Section 3, a brief introduction of the HMAX model is given, and the detailed improvements and methods of our work are proposed. In Section 4, multiclass categorization results on Caltech101 are given, and comparison experiments with other models are also discussed. Finally, in Section 5, we conclude this paper and discuss the results and our future work.

2. Related Biological Researches

As the HMAX model and its modifications in this paper try to mimic the structures and mechanisms of the ventral stream of primate visual cortex, the review of related biological researches in anatomy, neurobiology and cognitive science that support the whole HMAX framework and the modifications are discussed respectively as below.

2.1. Biological Researches of the HMAX Framework

The ventral stream of primate visual cortex is associated with complex shape discrimination, object recognition, attention and long-term memory (Merigan, 1996; De Weerd et al., 1999; Nassi and Callaway, 2009). It is organized in a hierarchical way, after getting its inputs from the lateral geniculate nucleus (LGN), the visual information goes through V1, V2, V4 to areas of IT: PIT, Central inferotemporal(CIT), and anterior inferotemporal (AIT) successively.

In the ventral stream, as receptive fields of neurons in one visual layer together represent the entire visual field, each layer contains a full representation of the visual space. During the processing, visual information is propagated from a local region to its succeeding hierarchical region, in which the receptive field size of a neuron is approximately 2.5 times larger than the input layer. Such convergent connectivity overlaps continuously with each other and ensures the invariant representation of visual stimuli. Please refer to Serre et al. (2007) for more detailed biological evidence of the HMAX model.

2.2. Biological Researches of the Modifications
2.2.1. Neuronal Response Characteristic and Feature Encoding Mode
The orientation, position and color information are critical for feature encoding in visual cognition.

2.2.1.1. Orientation and Location
The neuronal responses of V1 can discriminate small changes in visual orientations and spatial frequencies, and the spatial location of visual information is well retained. V2 and V4 are similar with V1, but have more tuning properties. The responses of V2 neurons could also be modulated by the orientation of illusory contours, and discriminate whether the stimulus is part of the foreground or the background (Qiu and von der Heydt, 2005). V4 is tuned for object features of intermediate complexity, like simple geometric shapes. IT layer is associated with the representation of complex object features.

2.2.1.2. Color
The processing of color information begins in the retina with three types of cones cells-L, M, S, which have different responses to different wavelength lights (Hunt, 2005). Then the signals are transmitted through LGN to V1. The color cells in LGN and V1 are only sensitive along two axes, roughly red-cyan and

blue-yellow (Wiesel and Hubel, 1966; Chatterjee and Callaway, 2003; Field et al., 2007). In V1, there are double-opponent neurons which compute local color contrast and color constancy (Danilova and Mollon, 2006; Kentridge et al., 2007). V1 color cells are clustered within cytochrome-oxidase blobs, and then project to the cytochrome-oxidase thin stripes of V2, which in turn project to globs in PIT. Glob cells achieve the perception of hue including red, green, blue, and to some extent yellow (Conklin, 1973). The final processing of color signals takes place in IT, which may help with shape decision making (Matsumora et al., 2008; Conway, 2009).

Finally, the visual inputs are transformed into representations that embody the enduring characteristics of objects and their spatial relationship (Milner and Goodale, 2008).

2.2.2. Attention Modulation
Attention modulation includes two modes: bottom-up and top-down. Visual selection is completely stimulus-driven in the first 150 ms, and the salience of objects can be modulated by bottom-up priming in a passive automatic way. In the later time (N150 ms), through massive recurrent feedback processing, active volitional control based on expectancy and task will bias visual selection in a top-down manner (Theeuwes, 2010).

In this paper, we focus on the bottom-up attention modulation, which is associated with salience. It is computed on the basis of the detection of locations which have significant local feature contrast, along some dimension or combination of dimensions (Itti and Koch, 2001; Donk and van Zoest, 2008). Firstly, a bottom-up saliency map can be created in V1 (Theeuwes, 2010; Zhang et al., 2012), and lateral connections (Gilbert and Wiesel, 1983; Rockland and Lund, 1983) between V1 neurons help mutual suppression between neurons tuned to similar input features. In addition, V2 is mainly responsive to top-down modulations (Beck and Kastner, 2005). In V4, bottom-up saliency and top-down control converge, and finally generate an overall saliency map (Töellner et al., 2011a,b).

2.2.3. Distributed Memory and Association Structure
The regions in the ventral stream have distributed memory and association structures.

Layer 6 of V2 are found to be important in the storage of object recognition memory and the conversion of short-term object memories into long-term object memories (López-Aranda et al., 2009). IT is connected with other memory associated areas, namely the hippocampus, the amygdala and the prefrontal cortex. Gross (2008) revealed that neurons in IT with similar selectivity of memory are clustered together and they also display learning ability over time. For example, different neural populations appear to be selectively tuned to particular components (e.g., face, eyes, hands, legs) of the same biological object.

Moreover, discrete object categories are even associated with different regions: objects with many shared features (typical of living things) are associated with activities in the lateral fusiform gyri, whereas objects with fewer shared features (typical of nonliving things) are associated with activities in the medial fusiform gyri. While Perirhinal cortex (PRC) in the anteromedial temporal lobe (aMTL) is associated with discrimination between highly similar objects (Tyler et al., 2013). In addition, the Parahippocampal Place Area (PPA) could differentiate between scenes and objects, and the Fusiform Face Area (FFA) is more sensitive to facial and body recognition rather than to objects (Spiridon et al., 2006).

3. Methods and Detailed Implementation

In this part, the HMAX model is firstly reviewed. Secondly, based on the biological researches stated above, our enhanced model, focusing on the first 100–150 ms unsupervised feedforward cognitive process of the primate visual cortex, is proposed. And the modifications and methods are discussed in details.

3.1. The HMAX Model
During the hierarchical processing, the HMAX model progressively increases its selectivity and invariance for recognition. The function of each layer in the HMAX model is discussed briefly in the following.

3.1.1. S1 Layer
This layer mimics the simple cells in V1, which have a Gabor-like response characteristic. The grayscale input image is processed by a convolution operation with multidimensional array of S1 cells, and the S1 cells act with Gabor function as follows

$$G(x, y) = \exp\left(-\frac{x_0^2 + \gamma^2 y_0^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} x_0\right) \qquad (1)$$

where $x_0 = x\cos\theta + y\sin\theta$ and $y_0 = -x\sin\theta + y\cos\theta$. 4 orientations θ (0°, 45°, 90°, and 135°) and 16 scales s are selected, and other parameters are also tuned to generate 64 (= 4 × 16) S1 layer feature maps $\mathbf{FM_{S1}}$, see Table I in Serre et al. (2007), for more details.

3.1.2. C1 Layer
This layer mimics the complex cells in V1, which have larger receptive fields than simple cells in V1 (S1 layer) and show some degree of tolerance to shift and scale. Each C1 layer feature map is generated by max-pooling local neighborhoods ($L_S \times L_S$) in the same scale band with a step overlap, as Equation (2). Here, one scale band is formed by two feature maps with adjacent scales in S1 layer. Thus, some degree of shift and scale invariance is achieved in C1 layer, and 32 (= 4 × 8) C1 layer feature maps $\mathbf{FM_{C1}}$ are obtained.

$$\mathbf{FM_{C1}}(x, y)^{s,\theta} = \max_{\mathbf{u}_{x,y} \in \mathbf{B\_FM_{S1}}^{s,\theta}} \mathbf{u}_{x,y} \qquad (2)$$

where $\mathbf{u}_{x,y}$ are the local neighborhoods centered at point (x, y) in one of the orientation map of one scale band of S1 layer— $\mathbf{B\_FM_{S1}}^{s,\theta}$.

3.1.3. Prototype Sampling
In this stage, M prototypes $\{P\}$ are extracted from the C1 layer across all four orientations ($n \times n \times 4$), and $n = (4, 8, 12, 16)$

is the prototype size. Only a random sampling method is used for prototype extracting. For binary classification task, the prototypes are only sampled from the positive training set.

3.1.4. S2 Layer
This layer corresponds to the cells in V4 and IT layer. For all positions and orientations of each scale band, the difference of the one feature map patch $X^s \in \mathbf{FM_{C1}}^s$ centered at (x, y) and each prototype $P^m \in \{P\}$ is computed in a Gaussian-like way as Equation (3).

$$\mathbf{FM_{S2}}(x, y)^s_m = \exp(-\beta \|X^s - P^m\|) \qquad (3)$$

Where β defines the sharpness of the tuning. Here, as all the four orientations are computed together, $8 \times M$ S2 layer feature maps $\mathbf{FM_{S2}}$ are computed.

3.1.5. C2 Layer
In this layer, for the $\mathbf{FM_{S2}}$ corresponding to one prototype P^m, its C2 layer response is computed by taking a global maximum over all scales and positions. Thus, the final feature vector consists of M C2 values, which is a position- and scale- invariant representation of an image.

3.2. The Enhanced HMAX Model
Given a set of training images $\mathcal{D}$ and $\mathcal{N}$, where $\mathcal{D}$ is a "discovery dataset" comprising a variety of object classes, and $\mathcal{N}$ is the "natural world dataset" including many other common objects and scenes. The goal of the enhanced HMAX model is to mimic the first 100–150 ms feedforward visual cognition procedure with the images in $\mathcal{D}$ and $\mathcal{N}$ by introducing attention modulation, memory processing and position encoding into the original HMAX model, and finally achieve multiclass categorization. The whole framework of this paper is given in **Figure 1**. All the modifications of the original HMAX model are discussed in the following, which correspond to related biological researches that stated in Section 2.

3.2.1. Attention Modulation—Saliency Map Generation
In this step, the original HMAX model is extended with attention modulation in S1 level, in which a bottom-up saliency map is generated based on color and orientation contrast, which corresponds to the biological evidence of attention modulation in V1 layer (Gilbert and Wiesel, 1983; Donk and van Zoest, 2008; Theeuwes, 2010). Only the dataset $\mathcal{D}$ is processed in this step, as it contains the object class to be learned. The generated saliency map will support the prototype learning in next stage.

Different from the gray input images in the original HMAX model, we use color input images and convert them to Lab images, as this color space is mostly consistent with the characters of LGN and V1 cells, which are sensitive along two axes, roughly red-cyan and blue-yellow (Danilova and Mollon, 2006; Kentridge et al., 2007).

For a color image, based on the work of Itti et al. (1998) and Achanta et al. (2009), firstly, the S1 layer orientation feature map with 12 orientations θ and 16 Gabor scales s are computed

based on the L channel in Lab color space. Since all the feature maps have the same size of the original image, we can directly compute the orientation saliency map by difference operation as Equation (4) rather than the downsampling and interpolation operation in Itti et al. (1998). Here, the first 8 scales are selected to compute the orientation saliency map. The scale interval Δs for the difference operation is 4, and the difference of all the scales and orientations are added together to get $\mathbf{SFM_O}$. Then, by computing the mean value $avg()$ and the standard deviation $std()$ of $\mathbf{SFM_O}$, the normalized orientation saliency map $\overline{\mathbf{SFM_O}}$ is obtained.

$$\mathbf{SFM_O} = \sum_{s=1}^{4} \sum_{\theta=1}^{12} (\mathbf{FM_O}^{s,\theta} - \mathbf{FM_O}^{s+\Delta s,\theta}) \qquad (4)$$

$$\overline{\mathbf{SFM_O}} = (\mathbf{SFM_O} - avg(\mathbf{SFM_O}))/std(\mathbf{SFM_O})$$

Secondly, the Lab color feature map $\mathbf{FM_C}$ is obtained by gaussian filtering of the original Lab image, and the color saliency map $\mathbf{SFM_C}$ is computed as Equation (5). $avg(\mathbf{FM_C^i})$ computes the mean value of the ith channel of $\mathbf{FM_C}$, and the normalized color saliency map $\overline{\mathbf{SFM_C}}$ is computed in the same way as $\overline{\mathbf{SFM_O}}$.

$$\mathbf{SFM_C} = \sum_{i=l,a,b} (\mathbf{FM_C}^i - avg(\mathbf{FM_C}^i))^T (\mathbf{FM_C}^i - avg(\mathbf{FM_C}^i)) \ (5)$$

Where l, a, b corresponds to the three channels of Lab color space, respectively.

Finally, the normalized saliency feature maps of color and orientation are combined together as $\overline{\mathbf{SFM}} = \lambda_1 \cdot \overline{\mathbf{SFM_O}} + \lambda_2 \cdot \overline{\mathbf{SFM_C}}$ to get the final saliency map ($\lambda_1 = 0.4, \lambda_2 = 0.6$). The procedure of saliency map generation is illustrated in the S1 layer of **Figure 1**. Furthermore, the salient points are also sorted according to their values in $\overline{\mathbf{SFM}}$.

3.2.2. Memory Processing—Prototype Learning
The prototype selection of the original HMAX model (Serre et al., 2007) is based on random sampling. The representation and discrimination ability of these prototypes are not guaranteed. While in other modified HMAX models (Mutch and Lowe, 2006; Huang et al., 2011b), prototypes are selected or learned in each object class, respectively in a one vs. all manner, which is a supervised procedure.

However, we try to mimic the first 100–150 ms in visual cognition, which is an unsupervised feedforward procedure. Thus, we modify the unsupervised middle level patch (prototype) discovery method in Singh et al. (2012) to adapt to the HMAX framework. In the new model, patches belonging to multiclass can be learned without image label in an iterative way. During this procedure, similar patches are clustered together and one classifier is learned for each cluster for discrimination. This procedure corresponds to the memory processing function of V2 and IT, as the layer 6 of V2 are found important for the conversion of short-term memories to long-term memories (López-Aranda et al., 2009), and neurons in IT with similar selectivity of memory are clustered together and they also display learning ability over time (Gross, 2008).

In the new model, the datasets $\mathcal{D}$ and $\mathcal{N}$ are divided into two equal, non-overlapping subsets (D_1, N_1 and D_2, N_2) for cross-validation. The unsupervised prototype learning can be achieved

FIGURE 1 | The whole framework of the enhanced HMAX model. Each layer corresponds to the region in visual cortex. Besides the functions and structures of the original HMAX model, the modifications are marked with dotted green bounding box. The image with red bounding box on the bottom is the current processing image. LGN is sensitive to Lab color space; S1 (simple cells in V1 layer): The bottom-up saliency map based on orientation and color contrast is computed; C1 (complex cells in V1 layer): The initial patches are sampled by centering on the selected salient points; Prototype learning (V2 and IT): An iterative clustering method is used to learn discriminative patch clusters and their classifiers, which corresponds to memory processing; S2 (V4 and IT): Each cluster classifier is used as a detector to generate S2 layer; C2: Final features are integrated with orientation, position (and color) information; Visual task (IT): Multiclass categorization with softmax are achieved.

in two phases: *initial sampling* and *iterative learning*. The iterative learning is alternately processed between two steps: clustering and training classifiers on the two subsets. In addition, multi-scale patches are extracted, and the patches with different size $n (=16, 28)$ are processed independently in the prototype learning procedure, and finally integrated together in the C2 layer.

In the initial sampling phase, the patches from N_1 are taken as negative samples and selected in a random sampling manner with an overlap constraint, which filtrates the randomly sampled centers by making the distance between the any two centers no smaller than $\frac{1}{4}$ of the patch size n. The patches from D_1 are sampled in the salient regions. We discuss the initial sampling method in D_1 in the following.

Firstly, 8 C1 layer feature maps $\mathbf{FM3_{C1}}$ are computed with Equation (2). As the patches are sampled in the first scale band of C1 layer $\mathbf{FM_{C1}}^1$, the corresponding positions of the sorted salient points in C1 layer are computed. The final salient points are selected sequentially with an overlap constraint, which is the same as the constraint of the random sampling method on N_1. Then, S middle level patches $\{P^D\}$ in $\mathbf{FM_{C1}}^1$ are extracted by taking the final selected salient points as centers, which could guarantee a good cover of the whole salient region as well as avoid big overlap between patches.

Furthermore, due to the bigger size of middle level patches and more orientations computed than those of the original HMAX model, the feature dimension of a patch is high, which could be difficult for the SVM training of each cluster in the iterative

learning step, as there are very little positive training data. Thus, a dimension reducing method is proposed, which is similar to the design of HoG features (Dalal and Triggs, 2005) (illustrated in **Figure 2**). One patch is divided into 3×3 blocks with an overlap, and the orientation histogram of each block is computed, normalized with L2 norm, and cascaded to form the final feature vector of a patch, which is an effective and concise representation of a patch. In some cases, since the IT layer is sensitive to the RGB color space (Conklin, 1973), the RGB color histogram can also be computed in the same way of orientation histogram (dividing into 2×2 blocks), and added to the final feature vector.

In the iterative learning phase, the initial sampled patches are further learned and clustered.

Since the traditional k-means clustering method is not fruitful for the middle level patches due to its low level distance metric, in order to learn discriminative patches and avoid overfitting, an iterative learning method is used.

Secondly, by taking the patches of a cluster as positive features and all randomly sampled patches $\{P^N\}$ in N_1 as negative features, a weighted linear SVM classifier is learned for each cluster. And the SVM classifier is used as a detector in the first C1 scale band of N_1 to find hard negative patches, which are then used to retrain the SVM classifier of each cluster. Then, the learned SVM classifier of each cluster is used as detector in D_2, and only the top q (=5) ranked patches are taken to update the corresponding cluster to keep the purity. If the top ranked patches are less than 3, the cluster is deleted. Then, the subsets D_1, N_1 and D_2, N_2 are switched and a new iteration with SVM training and cluster updating are processed. In experiments, the algorithm converges in 4–5 iterations.

Moreover, the purity and discriminativeness of each learned cluster K_i is computed as Equation (6).

$$purity(K_i) = \frac{1}{r}\sum_{j=1}^{r} Score_{SVM}(P_j), \ \ P_j \in K_i$$

$$discri(K_i) = FireNum_D/(FireNum_D + FireNum_N) \quad (6)$$

Where $Score_{SVM}(P_j)$ is the score of the *jth* patches in the *ith* cluster K_i computed with the corresponding SVM classifier, and r is set to 10 ($r > q$) to evaluate the generalization of the cluster. $FireNum_D$ and $FireNum_N$ are the firing rates of the SVM classifier of cluster K_i in the datasets $\mathcal{D}$ and $\mathcal{N}$, respectively.

The purity and discriminativeness are normalized in the same way as Equation (4), and the general score of each cluster is computed with the normalized purity and discriminativeness, defined as $score(K_i) = \overline{purity(K_i)} + \lambda_3 \cdot \overline{discri(K_i)}$. Finally, the top ranked clusters and their corresponding classifiers are represented as $\Omega^n = \{K_i, C_i\}_{i=1}^{\Gamma_n}$ (Γ_n is the number of patches with size $n = 16, 28$), and all the clusters with different size n are stored together as $\Omega = \{K_i, C_i\}_{i=1}^{\Gamma}$, $\Gamma = \Gamma_{16} + \Gamma_{28}$.

The whole prototype learning algorithm is given in **Algorithm 1**.

3.2.3. Feature Integration with Position Encoding

In this part, the final feature vector in C2 layer with orientation and spatial position is computed.

Firstly, for each cluster $\{K_i, C_i\}$ in Ω, its corresponding S2 layer feature maps are generated by using C_i as detector in all the scale bands of the C1 layer. Each unit in the S2 layer is a SVM score, which could intuitively represent the discrimination ability of the *ith* cluster that corresponds to a distributed memory region of object component in IT (Gross, 2008). Finally, $8 \times \Gamma$ S2 layer feature maps are obtained.

Then, the C2 layer features are computed in the same way of the original HMAX. But the relative position coordinate $(x_{max}/W, y_{max}/L)$ of the maximum score of each cluster classifier is also added to the final feature vector, and W, L are the width and length of the S2 layer feature map with the maximum score in it. Thus, the length of the C2 layer feature vector of an image is $3 \times \Gamma$. Here, by integrating appearance features and loose spatial constraint together, more representative and
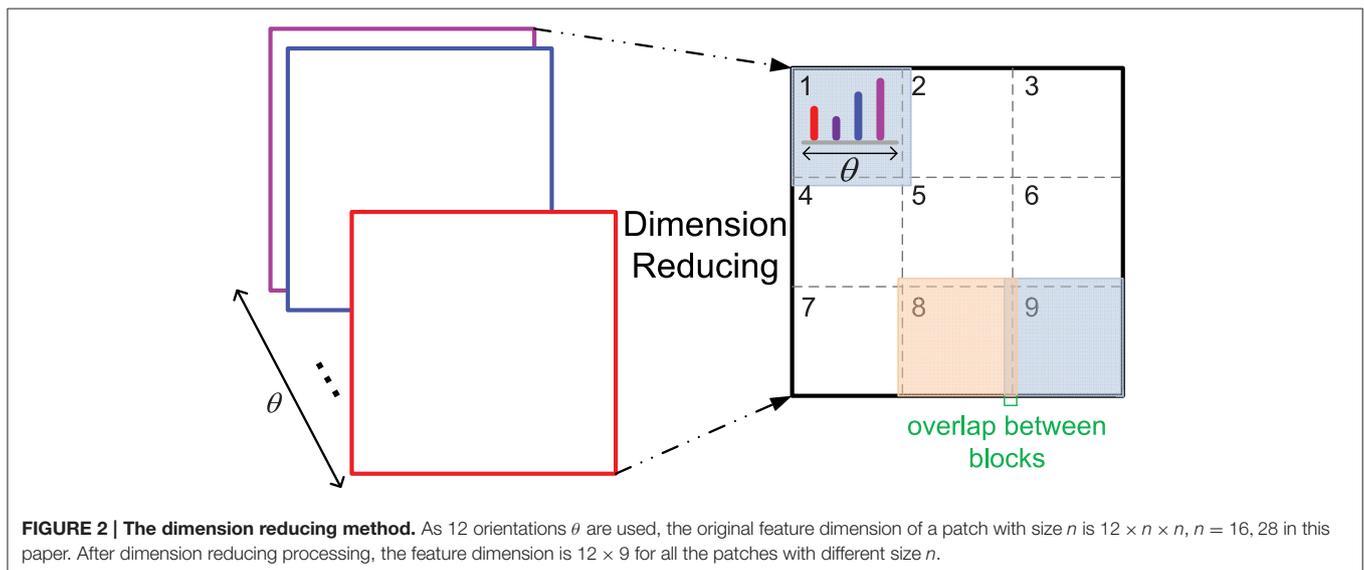
FIGURE 2 | The dimension reducing method. As 12 orientations θ are used, the original feature dimension of a patch with size n is $12 \times n \times n$, $n = 16, 28$ in this paper. After dimension reducing processing, the feature dimension is 12×9 for all the patches with different size n.

Algorithm 1 Unsupervised Prototype Learning Algorithm

Input: Training set $\mathcal{T}$ including $\mathcal{D}$ and $\mathcal{N}$
Output: The top ranked clusters and their corresponding clusters $\Omega = \{K_i, C_i\}_{i=1}^{\Gamma}$

  1: $\mathcal{D} \Rightarrow \{D_1, D_2\}$; $\mathcal{N} \Rightarrow \{N_1, N_2\}$ ▷ Split $\mathcal{D}$ and $\mathcal{N}$ into equal sized disjoint subsets
  2: Compute $\mathbf{FM_{C1}}$ with Equation (2)   ▷ Compute C1 layer feature maps
  3: **for** one patch size n in $\{16, 28\}$ **do**
  4:      Select S points from the sorted salient points   ▷ Operate in the first scale band of $\mathbf{FM_{C1}}$ of D_1
  5:      Extract S patches $\{P^D\}$ with dimension reduction
  6:      $\{K_i\}_{i=1}^{S/5} \Leftarrow Kmeans(\{P^D\})$ ▷ Use Kmeans to divide patches to $S/5$ clusters
  7:      **while** not converged **do**
  8:          **for** all i that $size(K_i) \geq 3$ **do**   ▷ Maintain clusters with enough patches
  9:             $C_i \Leftarrow SVM\_train(K_i, N_1)$ ▷ Use weighted SVM to train classifier for each cluster
10:             $Hard\_N_1 \Leftarrow hard\_mine(C_i, N_1)$ ▷ Find the hard negative patches in N_1
11:             $C_i^{new} \Leftarrow SVM\_retrain(K_i, Hard\_N_1)$ ▷ Retrain the classifier with $Hard\_N_1$
12:             $K_i^{new} \Leftarrow detect\_top(C_i^{new}, D_2, q)$ ▷ Find top $q = 5$ patches in D_2
13:          **end for**
14:          $K \Leftarrow K^{new}$; $C \Leftarrow C^{new}$
15:          $swap(D_1, D_2)$; $swap(N_1, N_2)$
16:      **end while**
17:      compute $score(K_i) = \overline{purity(K_i)} + \lambda_3 \cdot \overline{discri(K_i)}$ based on Equation (6)
18:      $\Omega^n = \{K_i, C_i\}_{i=1}^{\Gamma_n} \Leftarrow select\_top(C, score, \Gamma_n)$ ▷ Select the top Γ_n clusters of each patch size
19: **end for**
20: Unite all the top ranked cluster Ω^n with different patch size n to $\Omega = \{K_i, C_i\}_{i=1}^{\Gamma}$, $\Gamma = \Gamma_{16} + \Gamma_{28}$

---

discriminative features are learned, which is consistent with the function of the ventral visual stream (Milner and Goodale, 2008).

3.2.4. MultiClass Categorization

Based on the unsupervisedly learned features in C2 layer together with the image labels, a softmax layer is added on the top of the C2 layer to achieve the multiclass categorization task. Each output of the softmax layer corresponds to a distributed association region of an object class (Tyler et al., 2013). In addition, due to the unsupervised iterative learning manner of $\Omega = \{K_i, C_i\}_{i=1}^{\Gamma}$, similar patches from same object class are gathered together, and in some conditions, similar patches from different object class are also clustered together. The features from multiclass are shared, and the memory storage could be small. Meanwhile, the discriminativeness and purity are also guaranteed. Thus, the final feature vector is compact and suitable for multiclass categorization task.

4. Results

Multiclass categorization experiments on Caltech101 are carried out. The implementation of each modification and the final categorization result of the proposed model are evaluated and discussed. Furthermore, the comparison experiments with the original HMAX model and other unsupervised feature learning methods on multiclass categorization are also conducted and analyzed.

4.1. Dataset

Caltech101 (Fei-Fei et al., 2007) is a dataset with 102 classes (101 object class and 1 background). Here, 10 object classes are

selected, and 30 color images are randomly sampled in each class to form the "discovery dataset" $\mathcal{D}$ (positive training set). The 437 color images in the background class are taken as the "natural world dataset" $\mathcal{N}$ (negative training set). During the testing process, another 20 color images in each of the 10 object classes are selected to form the testing set.

4.2. Saliency Map Generation and Salient Point Selection

In this part, we discuss the role of saliency map in S1 layer (corresponding to V1 layer). Firstly, the V1 layer does have the ability of bottom-up saliency map generation based on local contrast. Secondly, the saliency map in S1 layer could provide a good initial region for patch selection. In **Figure 3**, some images, their corresponding saliency maps, and initially selected patches with different methods are given. We can see that the generated salient regions of our saliency map computation method (column 2) correspond to object regions in images, and the boundary and content are well kept. The proposed initial patch sampling method based on salient points (column 3) has a dense cover of the whole object region as well as avoid big overlap between patches, while the random sampling method with only overlap constraint (column 4) has a wider cover of the whole image, which extracts some meaningless patches in the background. Moreover, the purely random sampling method (column 5) has extracted some highly overlap patches, which is redundant, and can not guarantee a good cover of the whole object region.

For images with more complicated backgrounds, some saliency maps generated by the proposed method are also given in **Figure 4**. Although some points in the backgrounds are also activated, the object regions still have more salient and continuous activations.

FIGURE 3 | Some image examples, their saliency maps, and the initially sampled patches (red bounding boxes) with different methods. The 1st column includes original images, the 2nd column includes saliency maps computed based on Equations (4) and (5). The 3rd column includes initially sampled patches extracted by taking the final selected salient points as centers, which is used in this paper. The 4th column includes randomly sampled patches but with the overlap constraint (same with the constraint of 3rd column). The 5th column includes purely random sampled patches.

FIGURE 4 | Images with complicated backgrounds (left) and their saliency maps (right). Although some points in the complicated backgrounds are activated, the dominant object regions still have more salient and continuous activations.

4.3. Memory Processing—Prototype Learning

By processing the initially sampled patches with the unsupervised iterative patch clustering method in **Algorithm** 1, similar middle level patches are clustered together, and their corresponding SVM classifiers are also obtained. The convergence procedure of two clusters is given in **Figure 5**. Before the first iteration, the cluster is generated by k-means clustering, and there are some noises because of the low level distance metric. After 4 iterations, the middle level patches that clustered together become more similar.

FIGURE 5 | The iterative learning procedure of two clusters. The initial clustering with k-means is not quite meaningful because the patches in one cluster don't belong to same or similar part of objects. With the unsupervised iterative prototype learning method, the patches in one cluster become more and more similar. After 4th iterations, the patches in one cluster correspond to same critical part of objects.

Some examples of the final learned clusters are given in **Figure 6**. For each cluster in **Figure 6A**, the middle level patches correspond to a kind of key parts of an object class, which are representative and discriminative. While in **Figure 6B**, although the patches in same cluster are from different object classes, their appearances in orientation feature space are similar, which indicates that the similar middle level patches from different object class could be shared. Finally, by combining the middle level patches and the corresponding SVM classifier together, each cluster could be taken as a distributed region selective to one kind of object parts in the IT layer of the visual cortex.

4.4. Categorization Results and Comparisons

In this section, the multiclass categorization results of the enhanced HMAX model (eHMAX) are discussed in a various of conditions and compared with the original HMAX model (oHMAX). In addition, because the features of the eHMAX are learned in an unsupervised way, and each learned cluster could be considered as a true visual word (see **Figure 6**), and in the C2 layer the relative position coordinate of each cluster is also encoded into the final features, we could see that the framework of the eHMAX is similar with the BOW and SPM framework. Thus, the comparison experiments of the eHMAX and the representative models with BOW and SPM framework are also conducted, which includes KSPM (Lazebnik et al., 2006), ScSPM (Yang et al., 2009), and LLC (Wang et al., 2010).

Firstly, the categorization results of the eHMAX and the oHMAX with different sizes and different numbers of patches are given in **Figure 7**. Here, the number of patches in the eHMAX

corresponds to the number of clusters, as each cluster generates one feature map in the S2 layer, which is same with function of one patch (prototype) in the oHMAX.

As shown in **Figure 7**, with same number of patches, the patches with bigger size have shown higher accuracy in both models. It is because that the patch size 28 is much closer to the middle level patches, which always correspond to critical parts of object. While the patch size $n = 4, 8$ is too small to contain enough discriminative information. Moreover, the eHMAX model has shown better accuracy than the oHMAX model almost in all the conditions. For example, when the number of patches is 100, the accuracy of the eHMAX with patch size 16 and 28 is 83 and 88%, respectively, which is 9.5 and 13% higher than the oHMAX with 100 patches sized at 16 and 28. This indicates that the learned clusters in the eHMAX are more discriminative and representative. In order to achieve higher accuracy, more number of patches is needed for the oHMAX. And in some conditions, the increase of number of patches can not improve the accuracy a lot because of the low discrimination ability of randomly sampled patches. For example, the accuracy of the oHMAX model with 1000 patches sized at 16 and 28 is 80.5 and 81.5%, respectively. The improvements are not that dramatic comparing with the performance with the configuration of 100 patches. In a word, the memory storage and feature representation of the eHMAX model is more compact and effective.

In addition, We find that without encoding the relative spatial position information, the accuracy of the eHMAX model with patch size $n = 16, 28$ (100 clusters) drops to 79 and 83.5%,

FIGURE 6 | Some examples of the final learned clusters. In **(A)** (6 clusters), for each cluster, its patches correspond to same critical part of objects, which indicates the prototype learning method has learned representative features; In **(B)** (4 clusters), similar patches from different objects are clustered together, which shares the memories of different object class and helps to save memory size.

respectively. It is obvious that besides the learned discriminative and representative clusters, the good performance of the eHMAX model is also partly dependent on position encoding.

Secondly, according to the numbers of selected top clusters in different patch size, the final results of the eHMAX by combing mutiscale clusters are given in **Table 1**, and the results of other models are also listed. In the eHMAX Model, by combining 100 clusters sized at 28 and 500 clusters sized at 16, the best performance is obtained as 92.5%, while the oHMAX model with same number and scale of patches has an accuracy of 83%. For the oHMAX in Serre et al. (2007) with 4 patch sizes [4,8,12,16] and 800 patches of each size, the accuracy is only 78.5%. In addition, by setting the dictionary size of KSPM, LLC and ScSPM model to 600, which equals to the number of clusters in the eHMAX model, the ScSPM model achieves the best performance as 91%, but the accuracies of these three models are still lower than the eHMAX.

5. Discussion

Different from the original HMAX model with a random patch/prototype sampling method, and other modified HMAX models with selection of patches in a supervised manner, we focus on the first 100-150 ms feedforward/unsupervised cognitive processing to enhance the HMAX model, its success mainly depends on attention modulation, memory processing and feature encoding abilities, which are designed based on the related biological researches.

In the experiments, it is clear that the attention modulation could generate saliency maps with high quality, and provide good candidate salient regions/points for patch learning. The memory processing procedure could learn discriminative and representative middle level patches in an unsupervised iterative manner. Meaningless patches are deleted and similar patches from same/different object classes can be gathered in a same cluster during the procedure, which indicates the memory selectivity, sharing and clustering ability of the enhanced HMAX model.

As for the multiclass categorization experiments on Caltech101, the performance of the enhanced HMAX model and the original HMAX model with different size and number of patches is evaluated. Both of the models could achieve higher categorization accuracies with bigger size of patches, which indicates the middle level patches ($n = 28$) contain more discriminative information. The categorization accuracies of the two models have no significant improvement when the number of the patches is bigger than 100. For the enhanced HMAX model, the reason may be that the purity and discrimination of the new clusters are lower than that of the first 100 clusters. For the original HMAX model, the reason may be the new randomly sampled patches are meaningless or redundant. Furthermore, the enhanced HMAX always has a better performance than the original HMAX model with the same size and number of patches, with the reason that the enhanced HMAX model learns more discriminative middle level patches and also encodes relative position information into features. All in all, the enhanced

FIGURE 7 | Categorization accuracy of 10 classes in Caltech101 with different methods. The size number of each line corresponds to the used patch size of each model. The bigger the patch size, the higher accuracy can be achieved for all the models. The eHMAX with patch size 28 has the highest accuracy in all the conditions, which indicates that the memory storage and feature representation of the eHMAX model is more compact and effective.

TABLE 1 | Categorization accuracy of 10 classes in Caltech101 with different models.

Model	Parameters	Accuracy (%)
eHMAX	Patch size: [16,28], Number of clusters: [500,100]	**92.5**
oHMAX	Patch size: [16,28], Number of patches: [500,100]	83
oHMAX	Patch size: [4,8,12,16], Number of patches: [800,800,800,800]	78.5
KSPM	Dictionary size: 600	85
ScSPM	Dictionary size: 600	91
LLC	Dictionary size: 600	89.5

The best accuracy is achieved by the eHMAX model as 92.5%, and it is bold to be more striking.

HMAX model can achieved higher performance with smaller memory storage.

In addition, the comparison experiments of the HMAX model and three representative BOW and SPM models are conducted, which include KSPM, ScSPM, and LLC model. These three models also learn features in an unsupervised way, and their dictionary/codebook is similar to the patch cluster in the enhanced HMAX model. But the visual words in the KSPM and the ScSPM models are SIFT descriptors with patch size $n = 16$, and the visual words in LLC model are HOG descriptors with three sizes, $n = 16, 25, 31$, respectively. They are all extracted from the original image level, and these three models are flat processing method.

The experiment results indicate that the enhanced HMAX model has a higher accuracy than the above three models, which may owe to its hierarchical modeling and the discriminative middle level patches. Firstly, the hierarchical modeling helps to achieve some kind of invariance. Secondly, the size of the middle level patches is $n = 16, 28$ in the C1 layer (C1 layer is five times smaller than the original image), and the middle level patches mainly correspond to critical parts of objects, which are much bigger than the SIFT and HOG descriptors.

6. Conclusion

In this paper, based on recent biological research findings, we modified the original HMAX model by mimicking the first 100–150 ms unsupervised feedforward visual cognition process. The main contributions include:

(1) A bottom-up saliency map is generated based on local orientation and color contrast in S1 layer, which mimics the attention modulation ability of V1 layer of the visual cortex. The boundary and content of salient object are well kept, and the points in the salient regions are selected to support the initial sampling of patches.

(2) An unsupervised iterative clustering method is used to learn more representative and discriminative middle level patches, which mimics the learning, clustering and short-term memory to long-term memory conversion abilities of V2 and IT layer. After a few iterations, the patches in each cluster almost correspond to the same or similar key parts of object class, and one classifier of each cluster is also learned to distinguish it from others.

(3) The feature vector is computed in C2 layer, which is the cascade of the maximum activation value of each cluster and their corresponding relative spatial position. Finally, a softmax decision layer is used to achieve the multiclass categorization. This process mimics the feature encoding mode and distributed associated regions of different objects in the visual cortex.

Experiments on multiclass categorization task have demonstrated the effectiveness of the enhanced HMAX model.

In the future, on the one hand, we will investigate the reinforcement learning ability and the recurrent feedback processing of the visual cortex, and mimic the related structures and mechanisms to build new biologically inspired visual models. With the labels of images, the saliency map generation and memory learning can be further reinforced in a supervised manner, and a higher accuracy and robustness could be expected. With the ground-truth bounding box of objects, the relative position of each patch to the center of each object could be encoded to support categorization as well as detection task. On the other hand, it will also be meaningful to find a way to achieve multiple visual tasks, such as classification, detection and segmentation, in an unsupervised or weakly supervised way, since this way requires less human labor and the primate visual cortex does have such ability.

Author Contributions

YL prepared the methods of attention modulation, memory processing and position encoding. WW provided the related biological researches, which inspired the design of the whole framework. YL and FL conducted the experiments. YL, WW, and BZ prepared the manuscript. BZ initiated this study and supervised all aspects of the work. All authors discussed the results and commented on the manuscript.

Acknowledgment

References

Achanta, R., Hemami, S., Estrada, F., and Susstrunk, S. (2009). "Frequency-tuned salient region detection," in *Proc. CVPR* (Miami, FL: IEEE), 1597–1604.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Comp. Vis. Image Underst.* 110, 346–359. doi: 10.1016/j.cviu.2007.09.014

Beck, D. M., and Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nat. Neurosci.* 8, 1110–1116. doi: 10.1038/nn1501

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 711–720. doi: 10.1109/34.598228

Chatterjee, S., and Callaway, E. M. (2003). Parallel colour-opponent pathways to primary visual cortex. *Nature* 426, 668–671. doi: 10.1038/nature02167

Conklin, H. C. (1973). Basic color terms: their universality and evolution - berlin,b and kay,p. *Am. Anthropol.* 75, 931–942. doi: 10.1525/aa.1973.75.4.02a00010

Conway, B. R. (2009). Color vision, cones, and color-coding in the cortex. *Neuroscientist* 15, 274–290. doi: 10.1177/1073858408331369

Dalal, N., and Triggs, B. (2005). "Histograms of oriented gradients for human detection," in *Proc. CVPR*, Vol. 1 (San Diego, CA: IEEE), 886–893.

Danilova, M. V., and Mollon, J. D. (2006). The comparison of spatially separated colours. *Vis. Res.* 46, 823–836. doi: 10.1016/j.visres.2005.09.026

De Weerd, P., Peralta, M. R. III, Desimone, R., and Ungerleider, L. G. (1999). Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. *Nat. Neurosci.* 2, 753–758. doi: 10.1038/11234

Donk, M., and van Zoest, W. (2008). Effects of salience are short-lived. *Psychol. Sci.* 19, 733–739. doi: 10.1111/j.1467-9280.2008.02149.x

Fei-Fei, L., Fergus, R., and Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Comput. Vis. Image Underst.* 106, 59–70. doi: 10.1016/j.cviu.2005.09.012

Field, G. D., Sher, A., Gauthier, J. L., Greschner, M., Shlens, J., Litke, A. M., et al. (2007). Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *J. Neurosci.* 27, 13261–13272. doi: 10.1523/JNEUROSCI.3437-07.2007

Fukushima, K. (1988). Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Netw.* 1, 119–130. doi: 10.1016/0893-6080(88)90014-7

Gilbert, C. D., and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *J. Neurosci.* 3, 1116–1133.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR* (Columbus, OH: IEEE), 580–587.

Gross, C. G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia* 46, 841–852. doi: 10.1016/j.neuropsychologia.2007.11.009

Hamidi, M., and Borji, A. (2010). Invariance analysis of modified c2 features: case study-handwritten digit recognition. *Mach. Vis. Appl.* 21, 969–979. doi: 10.1007/s00138-009-0216-9

Huang, K., Tao, D., Yuan, Y., Li, X., and Tan, T. (2011a). Biologically inspired features for scene classification in video surveillance. *IEEE Trans. Syst. Man Cybern. B* 41, 307–313. doi: 10.1109/TSMCB.2009.2037923

Huang, Y., Huang, K., Tao, D., Tan, T., and Li, X. (2011b). Enhanced biologically inspired model for object recognition. *IEEE Trans. Syst. Man Cybern. B* 41, 1668–1680. doi: 10.1109/TSMCB.2011.2158418

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cats striate cortex. *J. Physiol. Lond.* 148, 574–591. doi: 10.1113/jphysiol.1959.sp006308

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. Lond.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Hunt, R. W. G. (2005). *The Reproduction of Colour*. New York, NY: John Wiley & Sons.

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259. doi: 10.1109/34.730558

Kentridge, R. W., Heywood, C. A., and Weiskrantz, L. (2007). Color contrast processing in human striate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15129–15131. doi: 10.1073/pnas.0706603104

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS* (Lake Tahoe, CA), 1097–1105.

Krüger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piáter, J., et al. (2013). Deep hierarchies in the primate visual cortex: what can we learn for computer vision? *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1847–1871. doi: 10.1109/TPAMI.2012.272

Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X

Lazebnik, S., Schmid, C., and Ponce, J. (2006). "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proc. CVPR*, Vol. 2 (New York, NY: IEEE), 2169–2178.

Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. (2009). "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. ICML* (Brisbane: ACM), 609–616.

Liao, Q., Leibo, J. Z., and Poggio, T. (2013). "Learning invariant representations and applications to face verification," in *Proc. NIPS* (Lake Tahoe, CA), 3057–3065.

López-Aranda, M. F., López-Tellez, J. F., Navarro-Lobato, I., Masmudi-Martín, M., Gutiérrez, A., and Khan, Z. U. (2009). Role of layer 6 of v2 visual cortex in object-recognition memory. *Science* 325, 87–89. doi: 10.1126/science.1170869

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110. doi: 10.1023/B:VISI.0000029664.99615.94

Matsumora, T., Koida, K., and Komatsu, H. (2008). Relationship between color discrimination and neural responses in the inferior temporal cortex of the monkey. *J. Neurophysiol.* 100, 3361–3374. doi: 10.1152/jn.90551.2008

Merigan, W. H. (1996). Basic visual capacities and shape discrimination after lesions of extrastriate area v4 in macaques. *Vis. Neurosci.* 13, 51–60. doi: 10.1017/S0952523800007124

Milner, A. D., and Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia* 46, 774–785. doi: 10.1016/j.neuropsychologia.2007.10.005

Mutch, J., and Lowe, D. G. (2006). "Multiclass object recognition with sparse, localized features," in *Proc. CVPR*, Vol. 1 (New York, NY: IEEE), 11–18.

Nassi, J. J., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10, 360–372. doi: 10.1038/nrn2619

Pascual-Leone, A., and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512. doi: 10.1126/science.1057099

Qiao, H., Li, Y., Tang, T., and Wang, P. (2014a). Introducing memory and association mechanism into a biologically inspired visual model. *IEEE Trans. Cybern.* 44, 1485–1496. doi: 10.1109/TCYB.2013.2287014

Qiao, H., Xi, X., Li, Y., Wu, W., and Li, F. (2014b). Biologically inspired visual model with preliminary cognition and active attention adjustment. *IEEE*

Trans. Syst. Man Cybern. B. doi: 10.1109/TCYB.2014.2377196. [Epub ahead of print].

Qiu, F. T. T., and von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron* 47, 155–166. doi: 10.1016/j.neuron.2005.05.028

Ramakrishnan, K., Scholte, H. S., Groen, I. I., Smeulders, A. W., and Ghebreab, S. (2015). Visual dictionaries as intermediate features in the human brain. *Front. Comput. Neurosci.* 8:168. doi: 10.3389/fncom.2014.00168

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Rockland, K. S., and Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *J. Comp. Neurol.* 216, 303–318. doi: 10.1002/cne.902160307

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). "Facenet: a unified embedding for face recognition and clustering," in *Proc. CVPR* (Boston, MA: IEEE).

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426. doi: 10.1109/TPAMI.2007.56

Singh, S., Gupta, A., and Efros, A. A. (2012). "Unsupervised discovery of mid-level discriminative patches," in *Proc. ECCV* (Firenze: Springer), 73–86.

Sivic, J., and Zisserman, A. (2003). "Video google: a text retrieval approach to object matching in videos," in *Proc. ICCV* (Nice: IEEE), 1470–1477.

Spiridon, M., Fischl, B., and Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum. Brain Mapp.* 27, 77–89. doi: 10.1002/hbm.20169

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychol.* 135, 77–99. doi: 10.1016/j.actpsy.2010.02.006

Thériault, C., Thome, N., and Cord, M. (2013). Extended coding and pooling in the hmax model. *IEEE Trans. Image Process.* 22, 764–777. doi: 10.1109/TIP.2012.2222900

Töellner, T., Zehetleitner, M., Gramann, K., and Mueller, H. J. (2011a). Stimulus saliency modulates pre-attentive processing speed in human visual cortex. *PLoS ONE* 6:e16276. doi: 10.1371/journal.pone.0016276

Töellner, T., Zehetleitner, M., Krummenacher, J., and Mueller, H. J. (2011b). Perceptual basis of redundancy gains in visual pop-out search. *J. Cogn. Neurosci.* 23, 137–150. doi: 10.1162/jocn.2010.21422

Turk, M. A., and Pentland, A. P. (1991). "Face recognition using eigenfaces," in *Proc. CVPR* (Maui: IEEE), 586–591

Tyler, L. K., Chiu, S., Zhuang, J., Randall, B., Devereux, B. J., Wright, P., et al. (2013). Objects and categories: feature statistics and object processing in the ventral stream. *J. Cogn. Neurosci.* 25, 1723–1735. doi: 10.1162/jocn/a/00419

Walther, D., Itti, L., Riesenhuber, M., Poggio, T., and Koch, C. (2002). "Attentional selection for object recognitiona gentle way," in *Proc. BMCV Workshop* (Tübingen: Springer), 472–479.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., and Gong, Y. (2010). "Locality-constrained linear coding for image classification," in *Proc. CVPR* (San Francisco, CA: IEEE), 3360–3367.

Wiesel, T. N., and Hubel, D. H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J. Neurophysiol.* 29, 1115–1156.

Yang, J., Yu, K., Gong, Y., and Huang, T. (2009). "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. CVPR* (Miami, FL: IEEE), 1794–1801.

Zhang, X., Zhaoping, L., Zhou, T., and Fang, F. (2012). Neural activities in v1 create a bottom-up saliency map. *Neuron* 73, 183–192. doi: 10.1162/jocn/a/00419

Visual Cortex Inspired CNN Model for Feature Construction in Text Analysis

Hongping Fu[1], Zhendong Niu[1]*, Chunxia Zhang[2], Jing Ma[1] and Jie Chen[1]

[1] School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China, [2] School of Software, Beijing Institute of Technology, Beijing, China

Recently, biologically inspired models are gradually proposed to solve the problem in text analysis. Convolutional neural networks (CNN) are hierarchical artificial neural networks, which include a various of multilayer perceptrons. According to biological research, CNN can be improved by bringing in the attention modulation and memory processing of primate visual cortex. In this paper, we employ the above properties of primate visual cortex to improve CNN and propose a biological-mechanism-driven-feature-construction based answer recommendation method (BMFC-ARM), which is used to recommend the best answer for the corresponding given questions in community question answering. BMFC-ARM is an improved CNN with four channels respectively representing questions, answers, asker information and answerer information, and mainly contains two stages: biological mechanism driven feature construction (BMFC) and answer ranking. BMFC imitates the attention modulation property by introducing the asker information and answerer information of given questions and the similarity between them, and imitates the memory processing property through bringing in the user reputation information for answerers. Then the feature vector for answer ranking is constructed by fusing the asker-answerer similarities, answerer's reputation and the corresponding vectors of question, answer, asker, and answerer. Finally, the Softmax is used at the stage of answer ranking to get best answers by the feature vector. The experimental results of answer recommendation on the Stackexchange dataset show that BMFC-ARM exhibits better performance.

Keywords: convolutional neural networks, biologically inspired feature construction, feature encoding, answer recommendation, community question answering, text analysis

1. INTRODUCTION

Community Question Answering (CQA) has attracted a lot of attentions from both research and industry communities in recent years. A fundamental problem in CQA is answer recommendation, which recommends the best answer of a question to the asker who posts the question.

Most previous research takes this problem as a ranking task and employs learning-to-rank algorithms to rank answers. Then the answer in the top of the answer list is recommended to users. To achieve this, most researchers focus on constructing complex and novel features (e.g., lexical features, syntactic features, and semantic features) to improve the recommendation performance. For example, Surdeanu et al. (2011) use linguistically motivated features to rank answers to non-factoid questions. They exploit natural language processing such as named-entity identification, syntactic parsing, and semantic role labeling to construct similarity features, translation features, density features, and frequency features.

However, feature construction is a time-consuming and labor-consuming problem which needs huge priori knowledge and experience, especially with the increasingly huge amount of questions and corresponding answers in CQA. Nowadays, many researchers focus on constructing features automatically using neural network. They only focus on the information of questions and answers, which is not suitable for CQA containing huge social information. It is worth to leverage the social information presented in CQA that users tend to focus or vote answers according to the relation with others.

Since, biologically inspired models are gradually proposed to solve the problem in text analysis recently and the biological research of primate visual cortex, traditional CNN can be improved by introducing attention modulation and memory processing of primate visual cortex. In this paper, we employ the attention modulation and memory processing of primate visual cortex to enhance the CNN model, and propose a biological-mechanism-driven-feature-construction based answer recommendation method (BMFC-ARM) to recommend the best answer for given questions in community question answering. In order to support feature construction, BMFC-ARM imitates the attention modulation property by introducing the asker-answerer information of given questions and computing the similarity between them, and then brings in the user reputation information of users who have answered the questions, which imitates the memory processing property. After feature construction, the Softmax is used at the stage of answer ranking to get the best answer. The experimental results of answer recommendation on the Stackexchange dataset show that the BMFC-ARM exhibits better performance.

The rest of this paper is organized as follows. Section 2 describes the related work. The proposed BMFC-ARM is introduced in Section 3, which contains biological mechanism driven feature construction and answer ranking. Section 4 gives details of experiments and corresponding results. Finally, Section 5 summarizes conclusion and future work.

2. RELATED WORK

2.1. Answer Recommendation

Answer recommendation is the basis research in CQA, which is designed to recommend the best answer to users. Wang et al. (2009) and Tu et al. (2009) proposed an analogical reasoning-based method to model question-answer relations to rank answers. Hieber and Riezler (2011) focused on the challenge of identifying high quality content caused by the inherent noisiness of user generated data. They proposed a series of features to model answer quality and expended the query, then used perceptron and Ranking SVM to rank answers. To recommend a reasonable answer to users, Liu et al. (2014) recognized questions in microblog and used collaborative filtering methods with integrated standard features and contextual features which are extracted from auxiliary resources. Beyond textural features used in previous works, user information is also investigated in answer ranking. Zhou et al. (2012) took advantage of three kinds of user ptofile information: level-related, engagement-related, and authority-related, and employed SVMRank and

ListNet for ranking answers. To avoid the manual quality control mechanisms, Dalip et al. (2013) proposed a learning to rank approach, and used textual and non-textual features which can represent the quality of query and answer pairs to rank answers. Specifically, the non-textual features contain user, review, and user-graph features.

2.2. Deep Learning for Text Analysis

Kalchbrenner et al. (2014) proposed a dynamic convolutional neural network with a dynamic k-Max pooling to model sentences. Hu et al. (2015) adapted the convolutional strategy in vision and speech, and then proposed convolutional neural network models for matching two sentences.

In classification tasks, Wang et al. (2016) proposed a framework to expand short texts based on word embedding clustering and convolutional neural network to overcome the worse classification performance caused by data sparsity and semantic sensitivity. Lai et al. (2015) introduce a recurrent neural network for text classification. Zeng et al. (2014) exploit convolutional neural network to extract word level features and sentence level features. Santos et al. (2015) proposed a a new pairwise ranking loss function and used convolutional neural network for relation classification.

Deep learning has been proven to be effective for many text analysis tasks. Recently some researchers brought deep learning into question answering. Bordes et al. (2014a,b) used an embedding model to project question-answer pairs into a joint space. tau Yih et al. (2014) used convolutional neural network to measure the similarity of entity and relation of a question with those in knowledge base for single-relation question answering. Iyyer et al. (2014) introduced a recursive neural network to model textual composition for factoid question answering. Zhou et al. (2016) aimed to find answers of previous queries to new queries, and used neural network architecture to learn the semantic representation of queries and answers in community question answering retrieval.

Different from previous deep learning methods only focusing on the semantics of question-answer pairs to rank answers, we also take user information into account in this paper, which is an very important aspect in community question answering.

3. METHODOLOGY

In this section, we present the proposed approach BMFC-ARM, which contains biological mechanism driven feature construction (BMFC) and answer ranking. First, an overview of the framework of BMFC-ARM is given. Then we describe the BMFC method and answer ranking in detail.

3.1. Overview of BMFC-ARM

Answer recommendation can be viewed as a ranking problem. Given a set of questions Q in a community question answering (CQA) system, each question $q_i \in Q$ contains a list of answers $A_i = \{a_{i1}, a_{i2}, \ldots, a_{ib}, \ldots, a_{in}\}$, where a_{ib} is the best answer selected by asker or CQA systems, our goal is to learn a ranker

according to these question-answer pairs, then recommend the best answer to any additional questions.

The proposed BMFC-ARM consists of two stages: BMFC and answer ranking which shown in **Figure 1.** BMFC method is to automatically construct features by introducing the attention modulation and memory processing, which contains three parts: text model, user model, and feature fusion. First, questions and their corresponding answers are passed through text model to get their feature vectors which contain semantic information. At the same time, the corresponding asker information and answerer information are passed through user model to get their feature vectors. In order to introduce the attention modulation and memory processing properties, BMFC imitates the attention modulation property by introducing the asker information and answerer information of given questions through user model

and computing the similarity between them, and then brings in the user reputation information of user who answered the questions, which imitates the memory processing property. After getting the feature representation of questions, answers, askers and answerers, feature fusion is used to combine those features into a single vector. After feature construction, answer ranking employs Softmax to recommend the best answer.

3.2. Biological Mechanism Driven Feature Construction (BMFC)

For the openness of CQA, all users can answer questions, which results in the unstable quality of answers. For the sociality of CQA, users get more interaction with each other when they
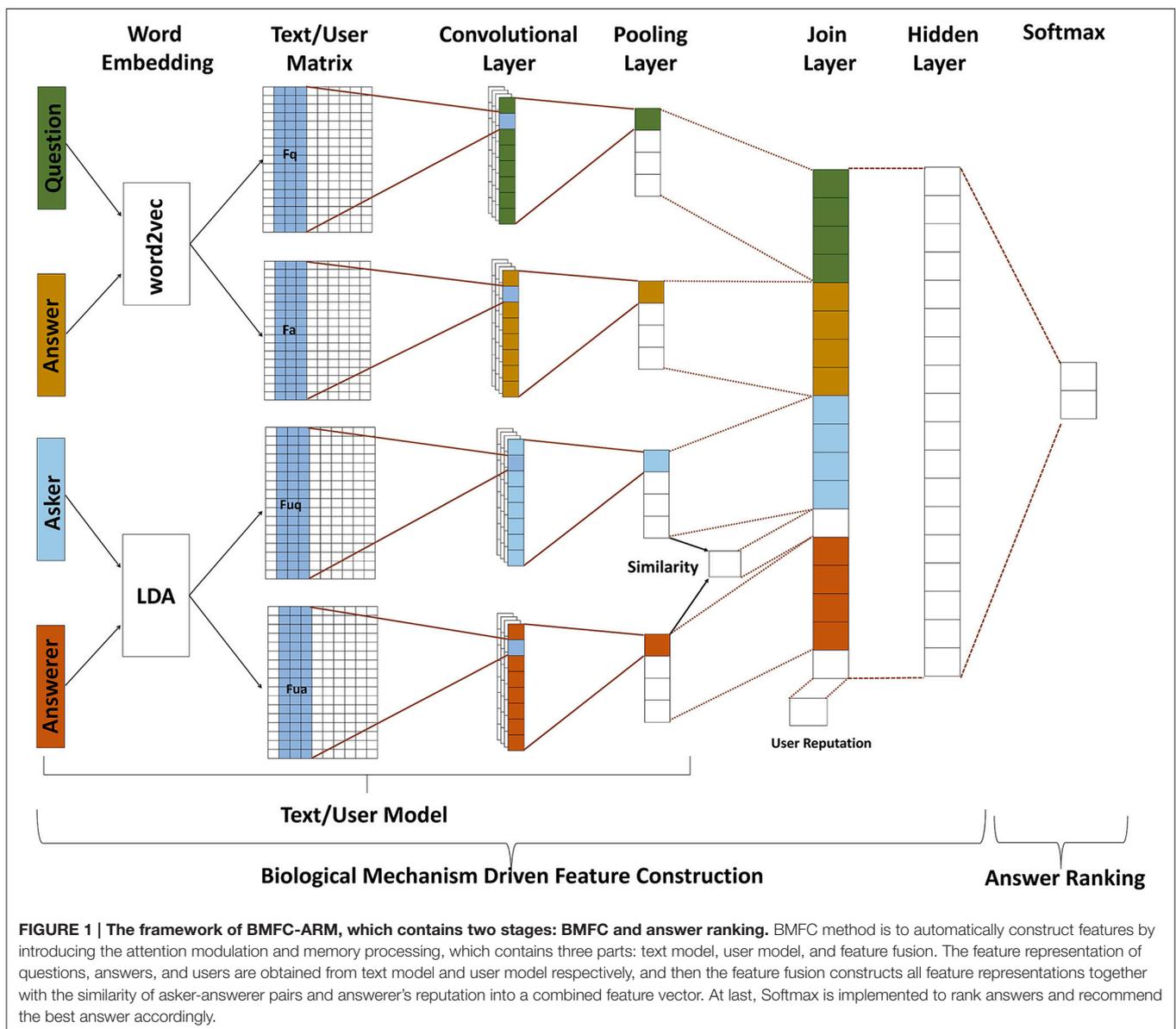
FIGURE 1 | The framework of BMFC-ARM, which contains two stages: BMFC and answer ranking. BMFC method is to automatically construct features by introducing the attention modulation and memory processing, which contains three parts: text model, user model, and feature fusion. The feature representation of questions, answers, and users are obtained from text model and user model respectively, and then the feature fusion constructs all feature representations together with the similarity of asker-answerer pairs and answerer's reputation into a combined feature vector. At last, Softmax is implemented to rank answers and recommend the best answer accordingly.

are similar, and may select the answer that provided by the answerer who is similar with them as the best answer. Therefore, in this paper, we assume that when users choose an answer as the best answer in CQA, their thinking process have two properties: (1) whether the answer is related to the question; (2) whether the answerer is the person they care about or familiar with.

According to the assumption, we introduce attention modulation and memory processing of primate visual cortex, and propose a biological mechanism driven feature construction (BMFC) method. As users may choose an answer which answered by the person similar to them as the best answer, BMFC imitate the attention modulation property by computing the similarity between askers and answerers of given questions based on user model to reflect the relation between askers and answerers. The reputation information represents the quality of answers user answered. In order to reflect the relevance of answers and questions, BMFC method

introduces user reputation to imitate the the memory processing property. BMFC method contains text model, user model and feature fusion. The flow of BMFC method is shown in **Figure 2**.

3.2.1. Text Model
The text model in BMFC is based on convolutional neural network which is shown in **Figure 3**. It contains two channels to model question and answer respectively, and each channel contains a convolution layer followed by a simple pooling layer.

3.2.1.1. Text Matrix
Our text model first transforms the original text into vectors. Inspired by Kalchbrenner et al. (2014), we use word2vec that takes advantage of the context of the word which contains more semantic information to do the word embedding for each word in a text, and then construct the text

FIGURE 2 | The BMFC method, which contains three parts: text model, user model, and feature fusion. First, questions and their corresponding answers are passed through text model to get their feature vectors which contain semantic information. At the same time, the corresponding asker and answerer are passed through user model to get user feature vector. In order to introduce the attention modulation and memory processing properties, BMFC imitate the attention modulation property by introducing the asker-answerer information of given questions through user model and computing the similarity between them, and then bring in the user reputation information of user who answered the questions, which imitates the memory processing property. After getting the feature representation of questions, answers and users, feature fusion is used to combine those features into a single vector.

FIGURE 3 | The text model is used to map text into its corresponding feature representations. We use word2vec to tranform texts into vectors, and then use two channel convolutional neural network to model questions and answers. All texts pass through a convolutional layer followed by a pooling layer. After text model, texts are presented by their corresponding features.

matrix $T \in R^{d \times |t|}$ shown bellow:

$$T = \begin{bmatrix} | & | & | & | & | \\ w_1 & \cdots & w_i & \cdots & w_{|t|} \\ | & | & | & | & | \end{bmatrix}$$

where $w_i \in R^d$ is the word embedding of a word in the text which contains $|t|$ words and i is the position of the word in the text.

Then we will give a description of convolutional layer and pooling layer used in each channel in next sections.

3.2.1.2. Convolutional layer

Convolutional layer is to convolve a matrix of weights with the matrix of activations at the layer below, which has two kinds of convolution: narrow convolution and wide convolution Kalchbrenner et al. (2014). In our framework, we use wide convolution which can deal with words at boundaries, and give equal attention to words in different positions. And we use ReLU as the activation function $f(\cdot)$. Given the text matrices $T \in R^{d \times |t|}$ and a convolution filter $k \in R^m$, the convolution operation between them results in a vector $c \in R^{|t| + m - 1}$. Each element of c is computed as follows:

$$c_i = f(T_{i-m+1:i}^T \cdot \mathbf{k} + b) \tag{1}$$

where $|t|$ is the number of text word, b is the bias, m is the width of convolutional filter.

3.2.1.3. Pooling layer

After convolutional layer, the input texts are represented by the extracted features, and then passed through the pooling layer. Pooling layer is used to reduce the dimension of features obtained through the convolutional layer and aggregate feature information from different parts. There are three commonly used

pooling methods: average-pooling, max-pooling, and stochastic-pooling. Boureau et al. (2010) compared average-pooling and max-pooling in detail. In this paper, we use max pooling which is the most widely used pooling methods. It chooses the feature with the maximum value in an area as shown in Equation (2).

$$c_p = max\{c\} \tag{2}$$

Then, the text matrix, convolutional layer and pooling layer form our text model which builds rich feature representations of the input question and answer.

Unlike previous works which just map the question and answer into a vector space, BMFC takes user information into account modeling asker and answerer into the same vector space, and evaluates the relatedness of asker-answerer pairs based on user model.

3.2.2. User Model

To introduce the attention modulation and memory processing property into BMFC, we propose user model which represents user information. In this paper, we use users' self descriptions as user information.

Same as text model, user model is based on CNN which contains two channels to represent askers and answerers, respectively. And each channel has a convolutional layer and a pooling layer shown in **Figure 4**. Since users' self descriptions are very short, e.g., some just contain keywords, we use Latent Dirichlet Allocation (LDA) Blei et al. (2003) to generate user matrix $U \in R^{d \times |u|}$.

$$U = \begin{bmatrix} | & | & | & | & | \\ w_{u1} & \cdots & w_{ui} & \cdots & w_{u|u|} \\ | & | & | & | & | \end{bmatrix}$$

FIGURE 4 | The user model for mapping user information into its corresponding feature representation. We use LDA to generate user matrix, and then use two channel convolutional neural network to model askers and answerers. All user matrixes pass through a convolutional layer followed by a pooling layer. After user model, askers, and answerers are presented by their corresponding features.

where d is the dimension of word vector, $\boldsymbol{w}_{ui} \in \boldsymbol{R}^d$ is the word representation of a word in user self description, $|u|$ is the number of words and i is the position of $\boldsymbol{w}_{ui}$.

The convolutional layer and pooling layer in user model are similar with those in text model.

3.2.3. Feature Fusion
After text model and user model, the information of questions, answers, and corresponding askers and answerers is represented by numeric vectors $\boldsymbol{v}_q$, $\boldsymbol{v}_a$, $\boldsymbol{v}_{uq}$, and $\boldsymbol{v}_{ua}$, respectively. Then, we compute the similarity between asker and answerer to represent their relations. Here, we use cosine similarity shown as follows:

$$s_{uqua} = \frac{\boldsymbol{v}_{uq} \cdot \boldsymbol{v}_{ua}}{\|\boldsymbol{v}_{uq}\| \times \|\boldsymbol{v}_{ua}\|} \qquad (3)$$

where s_{uqua} is the similarity between asker and answerer, $\|\boldsymbol{v}_{uq}\|$ is the Euclidean norm of $\boldsymbol{v}_{uq} = v_{uq1}, v_{uq2}, \cdots, v_{uqn}$ defined as $\sqrt{v_{uq1}^2 + v_{uq2}^2 + \cdots + v_{uqn}^2}$. Similarly, $\|\boldsymbol{v}_{ua}\|$ is the Euclidean norm of $\boldsymbol{v}_{ua}$.

Then, BMFC method concatenates the asker-answerer similarities s_{uqua}, answerer's reputation v_r, and corresponding vectors of question, answer, asker, and answerer into a single vector which can be represented as $\boldsymbol{v} = [\boldsymbol{v}_q{}^T; \boldsymbol{v}_a{}^T; \boldsymbol{v}_{uq}{}^T; s_{uqua}; \boldsymbol{v}_{ua}{}^T; v_r]$. Then, BMFC uses a hidden layer to interact the different parts of $\boldsymbol{v}$ to construct the final feature to represent samples:

$$\phi(\mathbf{w} \cdot \boldsymbol{v} + b)$$

where $\boldsymbol{w}$ is the weight vector of the hidden layer, b is the bias, and $\phi(\cdot)$ is the *tanh* function.

3.3. Answer Ranking
After feature construction using BMFC method, question-answer pairs, and their corresponding users' information are represented through a vector $\mathbf{V}$. In our method, we use a simple pointwise ranking method to rank answers. Softmax is often used in classification problem, which gives a probability of the sample belongs to each class. Given the sample vector $\mathbf{V}$, the probability that it belongs to class j ($j = 1, \ldots, K$) is computed by Equation (4). Then, answers are ranked according to this probability.

$$P(y = j|\mathbf{V}) = \frac{e^{\mathbf{V}^T \mathbf{W}_j}}{\sum_{k=1}^{K} e^{\mathbf{V}^T \mathbf{W}_k}} \qquad (4)$$

where $\mathbf{W}_k$ is the weight vector of the kth class.

4. EXPERIMENT

4.1. Experiment Setting
4.1.1. Dataset
In our experiments, the raw data we use are from Stack Exchange Data Dump[1], which is an anonymized dump of all user-contributed content on the Stack Exchange network[2]. The dataset contains 238 sites and each site consists of questions and corresponding answers of each question. We select around 840 resolved questions in four sites: movies, sports, travel, and music.

We split the dataset of 2385 question-answer pairs into a training set (train, 80%), a development set (dev, 10%), and a testing set (text, 10%) by randomly selecting 669 questions for training set, 87 questions for development set, and 84 questions for testing set, which are shown in **Table 1**. Here, each pair of

[1]https://archive.org/details/stackexchange
[2]http://stackexchange.com

TABLE 1 | Summary of the answer recommendation dataset.

Data	# Questions	# QA Pairs	# Askers	# Answerers	# Users
Train	669	1908	363	597	787
Dev	87	239	71	160	209
Test	84	238	76	173	229
Total	840	2385	428	689	912

#Question, #QA pairs, #Askers, and #Answerers are the number of questions, question-answer pairs, askers, and answerers respectively. #Users is the total number of askers and answerers except the overlap of them.

question and its answer together with the corresponding asker and answerer constitutes an example. The example with best answer is considered as the most relevant example among all examples with other answers of the same question. This setup is used in training set, development set, and testing set.

4.1.2. Word Embeddings

In our experiments, we use word2vec[3] to get word embeddings for questions and answers in text model, while using LDA to generate user representation for askers and answerers in user model.

For text model which represents question and answer information, we use word2vec to get word embeddings, which contains more semantic information by making use of the context of words. Similar with Kim (2014), Yu et al. (2014), we use the fixed word embeddings trained on all sites of Stack Exchange Data Dump. And we use the skipgram model with window size 5 to train word embeddings. Then words are represented by 50-dimensional vectors.

Due to the brief self description of users, we use JGibbLDA[4] trained by Gibbs sampling to generate word embeddings for the user model. The parameter α is set as 0.5, β is set as 0.1, topic number is set as 100, and each topic contains 50 words.

4.1.3. Parameters

The width m_t of convolutional filter of the text model is set to 5, and the width m_u of the user model is set to 2 according to experimental results. The convolutional maps of both models are 100, and the depth of the convolutional filter is set to 50 which is equal to the dimension of word vectors. We use ReLU as the activation function and max pooling method.

Similarly with Kim (2014), we use stochastic gradient descent over mini-batches to train the BMFC-ARM where batch size is set to 50.

4.1.4. Evaluation

For the task of answer recommendation, top answers in ranking list determine users' satisfaction. Therefore, we use Precision@N and Mean Reciprocal Rank(MRR) as metrics to evaluate our proposed method, which consider the position factor. Both of them are commonly used in information retrieval and question answering. Since we want to recommend the best answer to users, we use Precision@1($N = 1$) in this paper, which means that we

[3]https://code.google.com/archive/p/word2vec/
[4]http://jgibblda.sourceforge.net/

only focus on the precision of the first answer. Then Precision@1 is set to 1 if the best answer is ranked as first, 0 otherwise.

MRR takes the position of relevant answers into consideration. Where Reciprocal Rank is the multiplicative inverse of the rank of the first correct answer, and Mean Reciprocal Rank is the average of Reciprocal Rank that taken over all questions. MRR is computed as

$$MRR = \frac{1}{|Q|} \sum_{q=1}^{|Q|} \frac{1}{rank(q)}$$

where $|Q|$ is the number of questions in test dataset, $rank(q)$ is the position of the best answer in the resulting answer list.

4.2. Results

In this section we report the results of answer recommendation obtained by BMFC-ARM, and give a comparison among different methods (CNN-1, CNN-2, CNN-4, CNN-4M, CNN-4A, and BMFC-ARM). CNN-1 method just considers the information of questions and answers, which is passed through a single CNN network to obtain the corresponding features. CNN-2 method is a CNN network with two channels, which means that question information and answer information are passed through one channel respectively, and then obtains their corresponding features. CNN-2 just considers the information of questions and answers, which is similar with CNN. CNN-4 method considers both question-answer information and user information, which means that the information of question, answer, asker and answerer is passed through four channels of CNN network respectively, and then obtains their corresponding features. Based on CNN-4, CNN-4M brings in the answerer's reputation to imitate memory processing property, and CNN-4A introduces the similarity between askers and answerers. The proposed BMFC-ARM imitates the attention modulation property by introducing the asker-answerer information of given questions and computing the similarity between them, and brings in the user reputation information for users who answered the questions to imitate the memory processing property. The details of data used in this experiment are shown in **Table 1**. The evaluation results measured by MRR and P@1 are reported over this random split.

When users' information is added, we compare the effects of different widths m_u of convolution filter due to the brief self description of users. Unlike setting 5 as the width of convolution filter in the text model of representing questions and answers, our experiments compare the user model setting 2, 3, 4, and 5 as the width of convolution filter, respectively. **Table 2** gives the answer recommending results using different convolution filter widths in the user model. As seen from **Table 2**, when $m_u = 2$, the value of MRR and P@1 of all methods are higher than the cases of $m_u = 3$, 4, and 5. The reason behind this result may be due to the brief user information. Therefore, in the subsequent experiments of this paper, the convolution filter width of the user model is set to 2.

Figures 5, **6** show the recommendation results with different methods with memory processing property and without memory

TABLE 2 | Results with different widths of convolutional filter in user model.

Model	$m_u = 2$		$m_u = 3$		$m_u = 4$		$m_u = 5$	
	MRR	P@1	MRR	P@1	MRR	P@1	MRR	P@1
CNN-4	0.7498	0.5595	0.7403	0.5357	0.7409	0.5357	0.7240	0.5119

The CNN-4 method considers both question-answer information and user information, which means that the information of question, answer, asker, and answerer is passed through four channels of CNN network respectively, and then obtains their corresponding features. m_u means the convolution filter width in user model.
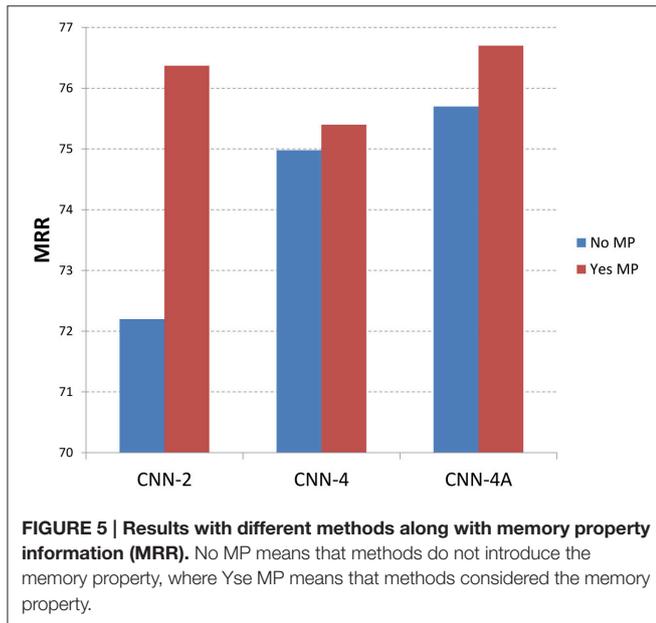
FIGURE 5 | Results with different methods along with memory property information (MRR). No MP means that methods do not introduce the memory property, where Yse MP means that methods considered the memory property.

FIGURE 6 | Results with different methods along with memory property information (P@1). No MP means that methods do not introduce the memory property, where Yse MP means that methods considered the memory property.

TABLE 3 | Results with different methods (CNN-1, CNN-2, CNN-4, CNN-4M, CNN-4A, and BMFC-ARM).

Model	MRR	P@1
CNN-1	0.6933	0.4524
CNN-2	0.7220	0.5000
CNN-4	0.7498	0.5595
CNN-4M	0.7540	0.5712
CNN-4A	0.7567	0.5714
BMFC-ARM (Our model)	0.7673	0.5952

CNN-1 method just considers the information of questions and answers, which is passed through a single CNN network to obtain the features. CNN-2 method is a CNN network with two channels, which means that question information and answer information are passed through one channel respectively. CNN-4 method considers both question-answer information and user information, which means that the information of question, answer, asker, and answerer is passed through four channels of CNN network respectively. Based on CNN-4, CNN-4M brings in the answerer's reputation to imitate memory processing property, and CNN-4A introduces the similarity between askers and answerers. BMFC-ARM imitates the attention modulation property by introducing the asker-answerer information of given questions and computing the similarity between them, and brings in the user reputation information which imitates the memory processing property.

processing property. **Figure 5** shows the results with MRR measure and **Figure 6** gives the P@1 measure. In these two figures, the blue histogram represents methods which do not consider the memory process mechanism, while the red histogram represents methods that considered the memory process mechanism. From **Figure 5** we can see that CNN-2, CNN-4, and CNN-4A obtain better performance by introducing the memory process mechanism, which shows that the memory processing mechanism through user reputation is useful to recommend best answers. For CNN-2 which just considers question information and answer information through two channels of CNN, the recommendation result performs better through adding user reputation, which shows that memory processing mechanism plays an important role in answer recommendation. From the recommendation results shown in **Figure 6** we can find that P@1 measure has the same tendency with MRR measure, which also shows that methods with memory processing mechanism get better performance than those without memory processing mechanism.

The recommendation results of different methods (BMFC-ARM, CNN-1, CNN-2, CNN-4, CNN-4M, and CNN-4A) with evaluation metrics of MRR and P@1 are shown in **Table 3**.

From **Table 3**, it is promising to observe that the proposed BMFC-ARM outperforms those CNN-1, CNN-2, CNN-4, CNN-4M, and CNN-4A with MRR and P@1 measure. It is probably because that BMFC-ARM takes user information into account introducing the attention modulation property and memory processing property. From methods CNN-2 and CNN-4, we can find that CNN-4 with user information performs better than CNN-2 which just uses question and answer information. This shows the importance of user information for answer recommendation. Therefore, when recommending the best answer to users in CQA, we need to take the relation information between askers and answerers into account rather than just considering question and answer information. The phenomenon

that CNN-4M performs better than CNN-4 may be caused by the introduced memory processing property. This indicates that our memory processing property is useful by introducing user reputation information. And the phenomenon that CNN-4A performs better than CNN-4 shows that recommendation results can be improved by considering the similarity between askers and answerers which brings in users' relation. The method CNN-4A outperforms CNN-2 shows that through introducing the attention modulation property, represented by user information and the similarity between askers and answers, can improve the recommendation results.

5. CONCLUSION

Convolutional neural networks (CNN) are hierarchical artificial neutral networks, which are popularly used in natural language processing. In this paper, we propose the BMFC-ARM to recommend best answers for given questions in community question answering, which is an improved CNN by introducing attention modulation and memory processing of primate visual cortex. In order to support the feature construction, we imitate the attention modulation property by computing the similarity of asker-answerer information of given questions, and bring in the user reputation information for users who answered

the questions, which imitates the memory processing property. Softmax is used at the stage of answer ranking to get the best answer. The answer recommendation experimental results on the Stackexchange dataset show that BMFC-ARM exhibits better performance.

In the future, we will investigate how to bring the users' sentiment information of questions into our framework and find a novel way to represent the text.

AUTHOR CONTRIBUTIONS

HF prepared the methods of feature construction and answer ranking. JC provided the related researches. HF and JM conducted the experiments. HF prepared the manuscript. ZN and CZ initiated this study and supervised all aspects of the work. All authors discussed the results and commented on the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022. doi: 10.1162/jmlr.2003.3.4-5.993

Bordes, A., Chopra, S., and Weston, J. (2014a). "Question answering with subgraph embeddings," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP* (Doha), 615–620.

Bordes, A., Weston, J., and Usunier, N. (2014b). "Open question answering with weakly supervised embedding models," in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases - Vol. 8724*, ECML PKDD 2014, (New York, NY: Springer-Verlag New York, Inc), 165–180.

Boureau, Y. L., Bach, F., LeCun, Y., and Ponce, J. (2010). "Learning mid-level features for recognition," in *2010 IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (San Francisco, CA), 2559–2566.

Dalip, D. H., Gonçalves, M. A., Cristo, M., and Calado, P. (2013). "Exploiting user feedback to learn to rank answers in q&a forums: a case study with stack overflow," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, (New York, NY: ACM), 543–552.

Hieber, F., and Riezler, S. (2011). "Improved answer ranking in social question-answering portals," in *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents*, SMUC '11, (New York, NY: ACM), 19–26.

Hu, B., Lu, Z., Hang, L., and Chen, Q. (2015). "Convolutional neural network architectures for matching natural language sentences" in *Advances in Neural Information Processing Systems 27 – 28th Annual Conference on Neural Information Processing Systems 2014, NIPS 2014*, (Montreal, QC).

Iyyer, M., Boyd-Graber, J., Claudino, L., Socher, R., and Iii, H. D. (2014). "A neural network for factoid question answering over paragraphs," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP* (Doha), 633–644.

Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014). "A convolutional neural network for modelling sentences," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Baltimore, MD), 655–665.

Kim, Y. (2014). "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP (Doha), 1746–1751.

Lai, S., Xu, L., Liu, K., and Zhao, J. (2015). "Recurrent convolutional neural networks for text classification," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI, 15 (Austin, TX), 2267–2273.

Liu, X., Xie, R., Lin, C., and Cao, L. (2014). Question microblog identification and answer recommendation. *Multimedia Syst.* 22, 487–496. doi: 10.1007/s00530-014-0411-z

Santos, C. N. D., Xiang, B., and Zhou, B. (2015). "Classifying relations by ranking with convolutional neural networks," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Beijing), 626–634.

Surdeanu, M., Ciaramita, M., and Zaragoza, H. (2011). Learning to rank answers to non-factoid questions from web collections. *Comput. Linguist.* 37, 351–383. doi: 10.1162/COLI_a_00051

tau Yih, W., He, X., and Meek, C. (2014). "Semantic parsing for single-relation question answering," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Baltimore, MD), 643–648.

Tu, X., Wang, X.-J., Feng, D., and Zhang, L. (2009). "Ranking community answers via analogical reasoning," in *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, (New York, NY: ACM), 1227–1228.

Wang, P., Xu, B., Xu, J., Tian, G., Liu, C.-L., and Hao, H. (2016). Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification. *Neurocomputing* 174, 806–814. doi: 10.1016/j.neucom.2015.09.096

Wang, X.-J., Tu, X., Feng, D., and Zhang, L. (2009). "Ranking community answers by modeling question-answer relationships via analogical reasoning," in *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, (New York, NY: ACM), 179–186.

Yu, L., Hermann, K. M., Blunsom, P., and Pulman, S. (2014). "Deep learning for answer sentence selection" in *Proceedings of the Deep Learning and Representation Learning Workshop: NIPS-2014* (Montreal, QC).

Zeng, D., Liu, K., Lai, S., Zhou, G., and Zhao, J. (2014). "Relation classification via convolutional deep neural network," in *the 25th International Conference on Computational Linguistics: Technical Papers*, COLING 2014 (Dublin), 2335–2344.

Zhou, G., Zhou, Y., He, T., and Wu, W. (2016). Learning semantic representation with neural networks for community question answering retrieval. *Knowl. Based Syst.* 93, 75–83. doi: 10.1016/j.knosys.2015.11.002

Zhou, Z.-M., Lan, M., Niu, Z.-Y., and Lu, Y. (2012). "Exploiting user profile information for answer ranking in cqa," in *Proceedings of the 21st International Conference Companion on World Wide Web*, WWW '12 Companion (New York, NY: ACM), 767–774.

A Brain Signature to Differentiate Acute and Chronic Pain in Rats

Yifei Guo [1,2], Yuzheng Wang [1,2], Yabin Sun [1,2] and Jin-Yan Wang [1]*

[1] Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [2] School of Humanities, University of Chinese Academy of Sciences, Beijing, China

The transition from acute pain to chronic pain entails considerable changes of patients at multiple levels of the nervous system and in psychological states. An accurate differentiation between acute and chronic pain is essential in pain management as it may help optimize analgesic treatments according to the pain state of patients. Given that acute and chronic pain could modulate brain states in different ways and that brain states could greatly shape the neural processing of external inputs, we hypothesized that acute and chronic pain would show differential effects on cortical responses to non-nociceptive sensory information. Here by analyzing auditory-evoked potentials (AEPs) to pure tones in rats with acute or chronic pain, we found opposite influences of acute and chronic pain on cortical responses to auditory inputs. In particular, compared to no-pain controls, the N100 wave of rat AEPs was significantly enhanced in rats with acute pain but significantly reduced in rats with chronic pain, indicating that acute pain facilitated cortical processing of auditory information while chronic pain exerted an inhibitory effect. These findings could be justified by the fact that individuals suffering from acute or chronic pain would have different vigilance states, i.e., the vigilance level to external sensory stimuli would be increased with acute pain, but decreased with chronic pain. Therefore, this auditory response holds promise of being a brain signature to differentiate acute and chronic pain. Instead of investigating the pain system *per se*, the study of pain-induced influences on cortical processing of non-nocicpetive sensory information might represent a potential strategy to monitor the progress of pain chronification in clinical applications.

Keywords: acute pain, chronic pain, auditory-evoked potentials (AEPs), sensory processing, animal models

INTRODUCTION

Acute pain, which serves as a warning signal of injury or illness, normally comes on quickly and lasts for a short time (Carr and Goudas, 1999; Apkarian et al., 2009). If not treated properly, acute pain can develop into chronic pain in which the pain persists even after the initial injury or illness is healed (Merskey and Bogduk, 1994). When this happens, considerable changes occur in both the peripheral and central nervous systems (CNS) as well as in the psychological profiles of individuals (May, 2008). An accurate differentiation between acute and chronic pain is essential in pain management as it may help optimize analgesic treatments according to the pain state of patients (Loeser and Melzack, 1999; Chou and Huffman, 2007a,b). It is, however, very difficult to make such a differentiation during pain chronification, and a commonly-used operational approach for this purpose is purely based on the duration of pain (e.g., pain that lasts for more than 3 or 6 months is defined as chronic pain; Merskey and Bogduk, 1994). This approach can be highly unreliable

because it ignores the substantial individual differences in the process of pain chronification (Lavand'homme, 2011). Nor can questionnaires be relied on to distinguish acute pain from chronic pain, since patients may sometimes describe the two pain states with equivalent characteristics (Hashmi et al., 2013).

Some recent studies have found that information about the transition from acute pain to chronic pain could be documented by changes in brain structure and function (May, 2008; Apkarian et al., 2009, 2011), for example, a large-scale reorganization of brain activities towards emotional circuits could occur during the chronification of back pain (Hashmi et al., 2013) and brain structural and functional connectivity may be able to predict that process (Baliki et al., 2012; Mansour et al., 2013). Importantly, the modulated brain structure and function could influence cortical processing of various sensory information—not only nociceptive information (Apkarian et al., 2005; Wiech et al., 2008) but also non-nociceptive information (Gilbert and Sigman, 2007; Fontanini and Katz, 2008). Consistent with this, besides the large number of studies that focused on the functionality of the nociceptive system *per se* in pain states, there are reports of distorted cortical processing of non-nociceptive sensory inputs (e.g., auditory and visual stimuli) in individuals with acute pain (Johnson and Adler, 1993; Lorenz and Bromm, 1997; Bingel et al., 2007) or chronic pain (Lorenz et al., 1997; Wang et al., 1999; Blomhoff et al., 2000; Ambrosini et al., 2003; Carrillo-de-la-Peña et al., 2006; Casale et al., 2008) in experimental settings. Given that acute pain and chronic pain may modulate brain activities in different ways (Apkarian et al., 2009, 2011), we hypothesized that the pain-related distortions of non-nociceptive sensory processing could be differently represented when pain shifts from acute to chronic states. If this hypothesis is valid, it would suggest that examining the pain-related distortions of non-nociceptive sensory processing might be a viable strategy for monitoring pain chronification and thus could be potentially applied in clinical practice.

Here we tested this hypothesis by investigating the different influences of acute pain and chronic pain on auditory-evoked potentials (AEPs) using rat models. An acute inflammatory pain model was produced by intraplantar injection of formalin, and a chronic inflammatory pain model was produced by intraplantar injection of complete Freund's adjuvant (CFA). In both pain models, multi-channel AEPs elicited by pure tones in freely-moving rats were recorded and compared.

MATERIALS AND METHODS

Animals

Sixty-four male Sprague-Dawley rats (weight at arrival: 180–200 g; Laboratory Animal Center, Academy of Military Medical Sciences, Beijing, China) were used in the experiments. Animals were housed individually under controlled temperature (22 ± 2°C) and humidity (50 ± 10%) conditions with a reversed 12 h light/dark cycle (light on at 7:00 PM). They were handled daily for a week before electrode implantation

surgery. All experimental procedures were in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the ethics committee of the Institute of Psychology, Chinese Academy of Sciences.

Electrode Implantation

Animals were anesthetized with sodium pentobarbital (50 mg/kg, i.p.) and then secured on a stereotaxic apparatus (Stoelting, Wood Dale, IL, USA). Twelve recording electrodes (stainless steel screws, 1.0 mm in diameter) were implanted symmetrically on the rat skull over both hemispheres according to the following coordinates: (1) electrodes L1 and R1, 5.0 mm anterior to bregma (5.0 A), ± 1.5 mm lateral to midline (± 1.5 L); (2) electrodes L2 and R2, 3.0 A, ± 1.0 L; (3) electrodes L3 and R3, −1.5 A, ± 2.5 L; (4) electrodes L4 and R4, −4.5 A, ± 1.0 L; (5) electrodes L5 and R5, 0.0 A, ± 4.5 L; and (6) electrodes L6 and R6, −4.5 A, ± 5.0 L. A reference and a ground electrode were placed at the midline, 2.0 and 4.0 mm posterior to lambda, respectively. Insulated wires connected the electrodes to a miniature connector, and the whole assembly was firmly attached to the skull with dental cement. After receiving penicillin (160,000 U, i.p.), animals were allowed at least 1 week to recover from the surgery.

Auditory Stimuli

Auditory stimuli were generated digitally using custom MATLAB (Mathworks, Natick, MA, USA) scripts, amplified by a power amplifier (A-S300, YAMAHA, Hamamatsu, Japan), and delivered through a loudspeaker (H1189–27TDFC, SEAS, Oslo, Norway) mounted in the ceiling of an anechoic sound-attenuated chamber. Recordings were carried out in a Plexiglas cage (L: 23 cm, W: 22 cm, H: 36 cm) situated in the sound-attenuated chamber. The loudspeaker was approximately 1 m from the middle of the test cage. The acoustic system was calibrated with a condenser microphone (C01U, Samson, Hauppauge, NY, USA) and a sound level meter (1350A, TES, Taipei, Taiwan) before the experiments.

AEPs were elicited by pure tones of either 8000 or 8800 Hz presented at 75 dB SPL with 100 ms duration and 500 ms inter-stimulus interval (auditory oddball paradigm). In accordance with previous studies recording auditory responses of the rat brain (Shinba, 1997; Lazar and Metherate, 2003; Jung et al., 2013; Witten et al., 2014), auditory stimuli with higher frequency than those commonly used in human AEP studies (Sambeth et al., 2003) were used in the present study, since rats exhibit more robust electrophysiological responses to higher pitch tones (with a maximum between 8 and 20 KHz) than to lower pitch tones (Knight et al., 1985). Each recording session contained eight stimulation blocks presented in random order with an approximately 1 min break between successive blocks. In four of these blocks, the lower frequency tone served as the standard (i.e., frequent stimuli, 85%) and the higher frequency tone as the deviant (i.e., rare stimuli, 15%). In the other four blocks, the roles of the lower and higher frequency tones were switched. In each block, 260 tones were presented in a pseudorandom order with the constraint that at least two standards were delivered

before each deviant. The first 10 stimuli in each block were excluded from off-line analysis in order to minimize the potential influence of switching between different types of blocks on the measured auditory responses (Nakamura et al., 2011). Each block lasted about 2.5 min and an entire session took less than 30 min.

EEG Recording

Rats were individually placed in the test cage 15–20 min before EEG data collection to familiarize them with the test environment. For EEG recording, a headstage was attached to the connector mounted on the rat's head and connected to an EEG amplifier (UEA-16BZ, SYMTOP, Beijing, China) via a flexible multi-strand cable. EEG signals were recorded continuously from the 12 recording electrodes, sampled at 1000 Hz, and low-pass filtered at 120 Hz. Rats were allowed to move freely in the test cage throughout the recording session.

Experimental Procedures

Acute Pain Model

Forty rats were randomly divided into four groups: 1% formalin ($n = 11$), 5% formalin ($n = 10$), normal saline (NS) control ($n = 10$), and no-treatment (NT) control ($n = 9$) group. After the rats were placed in the test cage for approximately 20 min, they were injected with 1% formalin, 5% formalin, or NS (50 μL each) subcutaneously into the plantar surface of their left hindpaw according to the group they belonged to. Immediately after injection, the rats were returned to the test cage. Nociceptive behaviors were video-recorded over the following 60 min and quantified by measuring the time spent licking the injected paw within each 5 min period. Rats in the NT group were treated by the same operations but without any injection. AEPs were repeatedly recorded 24 h before (baseline), 20–50 min, 90–120 min, and 24 h after injection. One rat in the 5% formalin group did not show any nociceptive behavior following injection and thus was excluded from further analyses.

Chronic Pain Model

Twenty-four rats were randomly divided into two groups: CFA group ($n = 12$) and NS group ($n = 12$). The rats were subcutaneously injected with either 100 μL CFA (Sigma-Aldrich, St. Louis, MO, USA) or NS into the plantar surface of their right hindpaw. AEPs were repeatedly recorded 1 day before (baseline), 1, 3, 7, 14, and 28 days after injection. Thermal nociceptive thresholds, quantified using the paw withdrawal latencies (PWLs) to radiant heat, of the injected and non-injected hindpaws were measured on each test day (PWL test started at least 2 h after the end of EEG recording).

The thermal nociceptive threshold test was adapted from Hargreaves et al. (1988). Rats were placed individually in Plexiglas chambers on an elevated glass floor and habituated to the test apparatus for at least 20 min. Focused radiant heat generated by a 100 W projector lamp was applied through the glass floor to the plantar surface of the stimulated hindpaw. PWL was defined as the time from the onset of heat stimulation to the withdrawal of the hindpaw. A cut-off time of 22 s was employed to avoid tissue damage. Five trials separated by at least 5 min were conducted on each hindpaw. To ensure that the rats were familiarized to the stimulation procedure and to increase the reliability of the measurement, latency of the first trial was discarded, and latencies of the following four trials were averaged to give a mean PWL.

Three rats (one in the CFA group, two in the NS group) did not show any movement of the stimulated hindpaws during the 22 s test period and thus were excluded from further analyses.

EEG Data Analysis

EEG data were preprocessed using EEGLAB (Delorme and Makeig, 2004), an open source toolbox running in the MATLAB environment, and custom MATLAB scripts. Continuous EEG signals were band-pass filtered between 1 and 30 Hz and segmented into epochs extending from −50 ms to +350 ms relative to the stimulus onset. EEG segments were baseline-corrected using the pre-stimulus interval, and trials contaminated by gross artifacts were manually rejected by visual inspection. Since we aimed to assess the influence of pain states (acute or chronic pain) on AEPs, our analysis was focused on standard-related cortical response due to its higher signal-to-noise ratio than deviant-related cortical response (the number of trials of standard was much larger than that of deviant). For each group, single-trial responses to standard stimuli were averaged for each rat and session. Single-rat average waveforms were subsequently averaged to obtain group-level waveforms for each session. Three distinct components in AEPs were identified, which consisted of an initial negative deflection peaking at ∼40 ms after stimulus onset (N40), followed by a positive deflection peaking at ∼60 ms (P60) and another negative deflection peaking at ∼100 ms (N100). For each group, peak latency and baseline-to-peak amplitude of each component were measured for each rat and session from the electrodes where the deflection reached its maximum. Grand-average scalp topographies at their peak latencies were plotted using a rat head model according to the rat brain atlas of Paxinos and Watson (2007).

Statistical Analysis

Data are expressed as mean ± standard error (SE). Statistical analyses were performed with STATISTICA 10 (StatSoft, Tulsa, OK, USA) and GraphPad Prism 5.0 (GraphPad Software, La Jolla, CA, USA). Statistical significance was set as $p < 0.05$.

For acute pain model, the licking time was compared using a two-way analysis of variance (ANOVA) with group (three levels: NS, 1% and 5% formalin) as a between-subject factor and time (12 levels: every 5 min during the first hour following injection) as a within-subject factor. The cumulative licking time within the 20–50 min interval after injection was compared among the three injected groups using a one-way ANOVA. For each AEP component, peak latencies were compared using a two-way ANOVA with group (four levels: NT, NS, 1% and 5% formalin) as a between-subject factor and time (four levels: baseline, 20–50 min, 90–120 min, and 24 h after

injection) as a within-subject factor. Baseline-to-peak amplitudes of each AEP component were normalized for each subject by dividing the value in each session by the value in the baseline session, and the normalized amplitudes of each AEP component were compared using a two-way ANOVA with group (four levels: NT, NS, 1% and 5% formalin) as a between-subject factor and time (three levels: 20–50 min, 90–120 min, and 24 h after injection) as a within-subject factor. Note that the baseline data were not included in this analysis since in the baseline session the normalized amplitudes, all of which were 1, had no variance. Fisher's protected least significant difference test was used for *post hoc* comparisons.

For chronic pain model, the paw withdrawal latencies to radiant heat were compared using a three-way ANOVA with group (two levels: NS and CFA) as a between-subject factor, and time (six levels: baseline, 1, 3, 7, 14, and 28 days after injection) and stimulation site (two levels: left and right hindpaws) as within-subject factors. Fisher's protected least significant difference test was used for *post hoc* comparisons. For each AEP component, peak latencies were compared using a two-way ANOVA with group (two levels: NS and CFA) as a between-subject factor and time (six levels: baseline, 1, 3, 7, 14, and 28 days after injection) as a within-subject factor. Consistent with the analysis for the acute pain model, baseline-to-peak amplitudes of each AEP component were normalized for each subject, and the normalized amplitudes of each AEP component were compared using a two-way ANOVA with group (two levels: NS and CFA) as a between-subject factor and time (five levels: 1, 3, 7, 14, and 28 days after injection) as a within-subject factor. Fisher's protected least significant difference test was used for *post hoc* comparisons.

RESULTS

The Influence of Acute Pain on AEPs

Nociceptive behaviors, quantified by measuring the time spent licking the injected paw within each 5 min period, are summarized in **Figure 1A** (left). Rats injected with 1% or 5% formalin, but not those injected with NS, exhibited a typical biphasic pattern of licking behavior (phase I: 0–5 min; phase II: 15–60 min). This observation is consistent with that of previous reports on the temporal profile of formalin-induced acute pain (licking behaviors usually subside within 1 h, while some other spontaneous nociceptive behaviors may last up to approximately 2 h; Dubuisson and Dennis, 1977; Porro and Cavazzuti, 1993), which justifies the validity of the acute pain model. Two-way ANOVA revealed that the licking time was significantly modulated by "group" ($F_{(2,27)}$ = 38.4, p < 0.0001), "time" ($F_{(11,297)}$ = 16.7, p < 0.0001), and their interaction ($F_{(22,297)}$ = 4.8, p < 0.0001). The cumulative licking time within the 20–50 min interval after injection was significantly different among the three injected groups ($F_{(2,27)}$ = 45.3, p < 0.0001, one-way ANOVA; **Figure 1A**, right). *Post hoc* comparisons revealed that the cumulative licking time was significantly different between each pair of the injected groups (NS vs. 1% formalin: p < 0.001;

NS vs. 5% formalin: p < 0.001; 1% formalin vs. 5% formalin: p < 0.01).

The group-level average AEP waveforms were characterized by three distinct components: N40, P60, and N100. Whereas the N40 and P60 waves were maximal over the frontal and bilateral temporal regions respectively, the N100 wave displayed a negative maximum over the fronto-central area (**Figure 2A**). Therefore, in the subsequent analyses, peak latencies and amplitudes of these waves were measured from the waveforms averaged across the following electrodes: L1 and R1 for N40; L6 and R6 for P60; L1, R1, L2, and R2 for N100 (**Figure 2B**).

Peak latencies and amplitudes of N40, P60, and N100 for different groups and sessions are summarized in **Table 1**. Two-way ANOVA revealed that peak latencies of N40 were not significantly modulated by "group", "time", or their interaction (detailed statistics are summarized in **Table 2**). Peak latencies of P60 and N100 were only significantly modulated by "time" (P60: $F_{(3,105)}$ = 8.1, p < 0.0001; N100: $F_{(3,105)}$ = 25.1, p < 0.0001). Normalized N40 amplitudes were not significantly modulated by "group", "time", or their interaction. Normalized P60 amplitudes were only significantly modulated by "time" ($F_{(2,70)}$ = 4.4, p = 0.016). In contrast, normalized N100 amplitudes were significantly modulated by "group" ($F_{(3,35)}$ = 6.2, p = 0.002) and "time" ($F_{(2,70)}$ = 8.1; p = 0.0007), but not by their interaction ($F_{(6,70)}$ = 0.5; p = 0.81; **Figure 2C**). *Post hoc* comparisons revealed that normalized N100 amplitudes in the 1% and 5% formalin groups were significantly larger than those in the NS and NT groups for any post-injection session (p < 0.05 for all comparisons except for the marginal significance (p = 0.06) between 1% formalin and NS groups during the 20–50 min interval after injection). Normalized N100 amplitudes were not significantly different between NS and NT groups, as well as between 1% and 5% formalin groups for any post-injection session (p > 0.05 for all comparisons).

These results demonstrated that N100 amplitude of AEPs was significantly enhanced in rats with acute pain (1% and 5% formalin groups) compared to control rats (NT and NS groups), which indicated that acute pain would facilitate the cortical processing of auditory information in rats. Such facilitation effect existed not only when formalin-injected rats exhibited robust nociceptive behaviors but also when the apparent nociceptive behaviors had subsided, e.g., 24 h after formalin injection.

The Influence of Chronic Pain on AEPs

Nociceptive thresholds, quantified by measuring PWLs to radiant heat of the injected and non-injected hindpaws, are summarized in **Figure 1B**. Rats injected with CFA exhibited pronounced thermal hyperalgesia that developed within 1 day and persisted through 14 days following injection. This observation is similar to that of previous reports on the temporal profile of thermal hyperalgesia in CFA-induced chronic pain model (Wang et al., 2009; Li et al., 2012). Three-way ANOVA revealed that PWLs were significantly modulated by "group" ($F_{(1,19)}$ = 5.8, p = 0.03), "time" ($F_{(5,95)}$ = 7.5, p < 0.0001), "stimulation site" ($F_{(1,19)}$ = 43.8, p < 0.0001), interactions between two factors ("group" × "stimulation site": $F_{(1,19)}$ = 41.0,
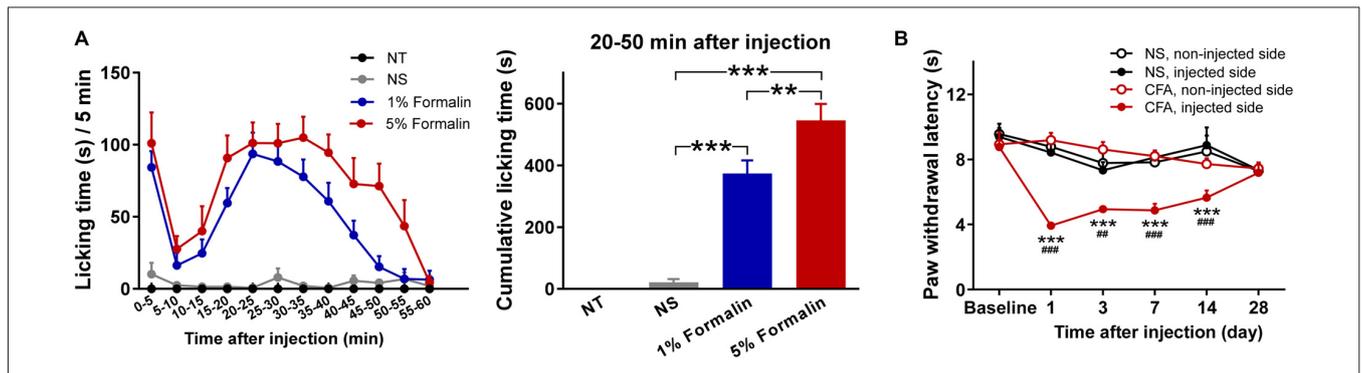
FIGURE 1 | Nociceptive behaviors of rats in acute and chronic pain models. (A) Formalin-induced acute pain behaviors. Left: Time spent licking the injected hindpaws within each 5 min period (from 0 to 60 min following the injection). Rats injected with 1% or 5% formalin showed a typical biphasic pattern of licking behavior (phase I: 0 to 5 min; phase II: 15 to 60 min), which was not observed in the NS and NT groups. Right: The cumulative licking time within the 20-50 min interval after injection. During this time interval, rats in the 1% and 5% formalin groups spent significantly longer time to lick their injected paws than rats in the NS group. Rats in the 5% formalin group also showed significantly longer licking time than rats in the 1% formalin group. NT: no treatment; NS, normal saline. **$p < 0.01$; ***$p < 0.001$. NT: $n = 9$; NS: $n = 10$; 1% formalin: $n = 11$; 5% formalin: $n = 9$. **(B)** Complete Freund's adjuvant (CFA)-induced chronic thermal hyperalgesia. Before injection (Baseline), paw withdrawal latency (PWL) to radiant heat stimuli was not significantly different between the NS and CFA groups, nor between the left and right hindpaws. From day 1 to day 14 after injection, PWLs of the injected hindpaw were significantly decreased in the CFA group compared to the NS group. Moreover, in the CFA group, PWLs of the injected hindpaw were significantly decreased compared to those of the non-injected hindpaw. NS, normal saline; CFA, complete Freund's adjuvant. For the comparison between CFA and NS groups of the injected hindpaw, ##$p < 0.01$, ###$p < 0.001$. For the comparison between injected and non-injected hindpaws in the CFA group, ***$p < 0.001$. NS: $n = 10$; CFA: $n = 11$. Data are expressed as mean ± standard error (SE).
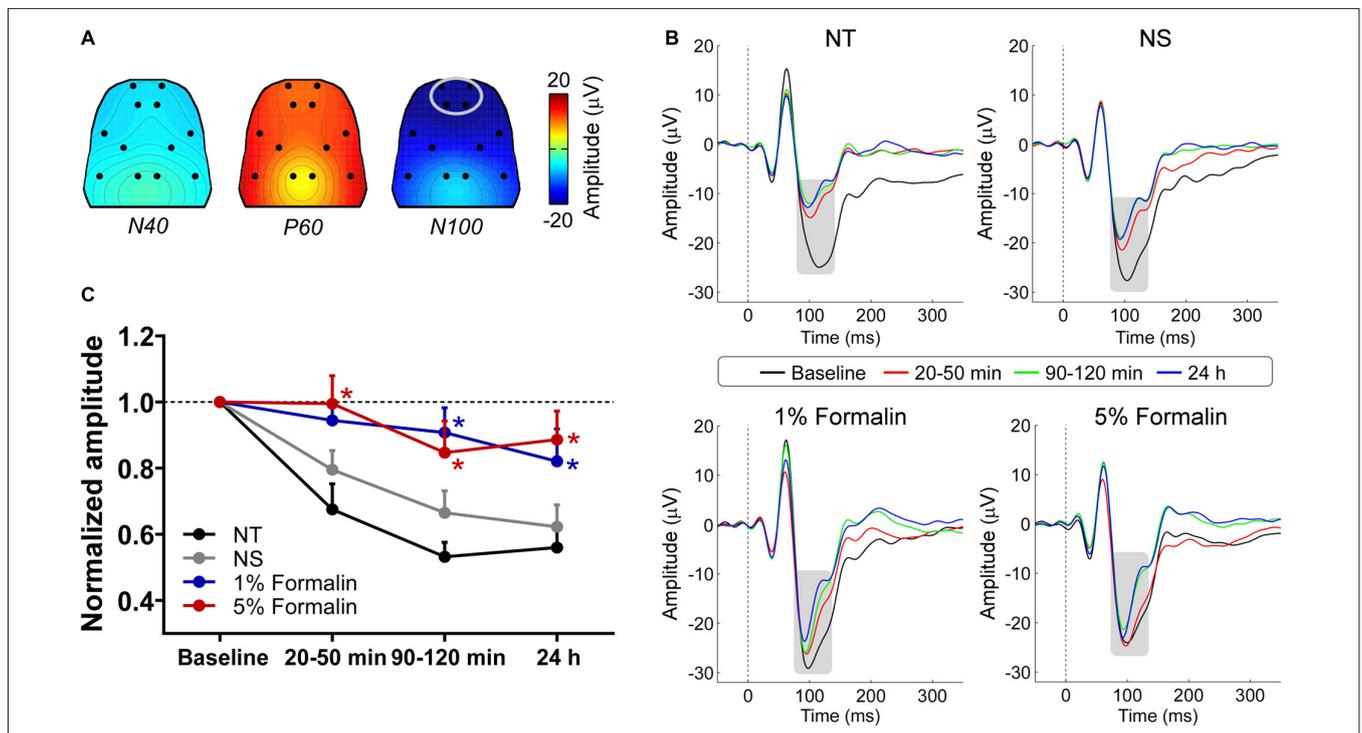
FIGURE 2 | The influence of acute pain on AEPs. (A) Grand-average scalp topographies of N40, P60, and N100 waves. **(B)** For each group, AEP waveforms from different sessions are plotted in different colors and superimposed. Displayed waveforms were measured from fronto-central electrodes (L1, R1, L2, and R2; enclosed by the light gray ellipse in **(A)**, where the N100 wave (marked using gray rectangles) displayed a negative maximum. **(C)** After injection, the normalized N100 amplitudes in the 1% and 5% formalin groups were significantly larger than those in the NS and NT groups for any post-injection session. NT: no treatment; NS, normal saline. *$p < 0.05$, compared to the NS group. NT: $n = 9$; NS: $n = 10$; 1% formalin: $n = 11$; 5% formalin: $n = 9$. Data are expressed as mean ± SE.

$p < 0.0001$; "time" × "stimulation site": $F_{(5,95)} = 9.5$, $p < 0.0001$; "group" × "time": marginal significance, $F_{(5,95)} = 2.2$, $p = 0.06$), and the interaction between three factors ($F_{(5,95)} = 7.6$,

$p < 0.0001$). *Post hoc* comparisons revealed that PWLs of the injected hindpaw were significantly shorter in the CFA group than in the NS group ($p < 0.01$ for all comparisons 1, 3, 7, and

TABLE 1 | Latency and amplitude of AEP components for different groups and sessions (acute pain model).

	Latency (ms)			Amplitude (μV)		
	N40	**P60**	**N100**	**N40**	**P60**	**N100**
NT group (n = 9)						
Baseline	38.3 ± 0.9	61.9 ± 0.6	112.0 ± 3.5	−8.1 ± 1.4	17.7 ± 2.9	−26.3 ± 4.8
20–50 min	38.2 ± 0.7	61.2 ± 0.6	101.9 ± 2.2	−6.3 ± 0.9	12.5 ± 1.6	−16.7 ± 3.8
90–120 min	38.8 ± 0.6	61.6 ± 0.5	102.3 ± 2.9	−6.5 ± 0.9	13.5 ± 1.2	−13.3 ± 2.7
24 h	38.6 ± 1.0	61.0 ± 0.6	99.3 ± 2.5	−7.0 ± 0.8	12.7 ± 1.9	−14.3 ± 3.3
NS group (n = 10)						
Baseline	39.4 ± 0.4	60.6 ± 0.5	105.1 ± 2.6	−7.5 ± 0.6	12.4 ± 1.1	−29.1 ± 5.3
20–50 min	39.6 ± 0.4	60.5 ± 0.6	98.5 ± 1.8	−7.1 ± 0.9	12.3 ± 1.5	−22.9 ± 4.3
90–120 min	39.5 ± 0.3	60.4 ± 0.7	97.4 ± 2.1	−7.8 ± 0.7	12.1 ± 1.5	−20.3 ± 5.4
24 h	39.2 ± 0.3	60.6 ± 0.7	96.1 ± 1.8	−7.1 ± 0.7	11.5 ± 1.1	−20.4 ± 5.6
1% Formalin group (n = 11)						
Baseline	38.7 ± 0.4	60.9 ± 0.7	102.2 ± 3.1	−6.9 ± 0.7	21.1 ± 2.9	−29.8 ± 5.4
20–50 min	38.5 ± 0.7	59.4 ± 0.7	97.7 ± 1.9	−6.3 ± 0.8	14.9 ± 2.0	−27.2 ± 4.0
90–120 min	38.4 ± 0.4	60.4 ± 0.5	95.1 ± 1.3	−7.1 ± 0.8	20.0 ± 2.8	−26.5 ± 4.9
24 h	38.6 ± 0.5	60.5 ± 0.7	91.7 ± 1.0	−6.8 ± 0.9	17.0 ± 2.4	−24.1 ± 4.7
5% Formalin group (n = 9)						
Baseline	38.3 ± 0.6	60.6 ± 0.5	103.2 ± 4.3	−7.6 ± 1.0	15.0 ± 1.8	−25.8 ± 5.1
20–50 min	37.8 ± 0.7	58.9 ± 0.5	98.0 ± 1.9	−5.8 ± 0.7	11.9 ± 1.1	−25.4 ± 3.7
90–120 min	38.2 ± 0.7	60.1 ± 0.3	95.6 ± 1.7	−5.2 ± 1.3	15.1 ± 2.2	−22.6 ± 4.9
24 h	38.3 ± 0.7	60.2 ± 0.6	97.3 ± 3.9	−6.9 ± 1.0	15.0 ± 1.9	−24.1 ± 6.0

Data are expressed as mean ± SE. NT, no treatment; NS, normal saline.

14 days after injection). In the CFA group, PWLs of the injected hindpaw were significantly shorter than those of the non-injected hindpaw ($p < 0.001$ for all comparisons 1, 3, 7, and 14 days after injection).

The group-level average AEPs of the chronic pain rats consisted of three distinct components (N40, P60, and N100), whose polarity and order were markedly similar to AEPs of the acute pain rats (**Figures 3A,B**). A comparison between **Figures 2A**, **3A** revealed high consistency in scalp topographies of the AEPs between the acute and chronic pain conditions, indicating that changes in pain state may not alter the spatial features of the auditory evoked cortical responses.

Peak latencies and amplitudes of N40, P60, and N100 for different groups and sessions are summarized in **Table 3**. Two-way ANOVA revealed that peak latencies of N40 and N100 were only significantly modulated by "time" (N40: $F_{(5,95)} = 3.1$, $p = 0.01$; N100: $F_{(5,95)} = 26.2$, $p < 0.0001$; **Table 4**). Peak latencies of P60 were not significantly modulated by "group", "time", or their interaction (**Table 4**). Normalized N40 and P60 amplitudes were not significantly modulated by "group", "time", or their

interaction. In contrast, normalized N100 amplitudes were significantly modulated by "group" ($F_{(1,19)} = 5.0$, $p = 0.038$) and "time" ($F_{(4,76)} = 15.4$; $p < 0.0001$), but not by their interaction ($F_{(4,76)} = 0.2$; $p = 0.95$; **Figure 3C**). *Post hoc* comparisons revealed that normalized N100 amplitudes in the CFA group were significantly reduced compared to those in the NS group for any post-injection session ($p < 0.05$ for all comparisons).

These results showed that N100 amplitude of AEPs was significantly reduced in rats with chronic pain (CFA group) compared to control rats (NS group), which indicated that chronic pain would inhibit the cortical processing of auditory information in rats. This inhibitory effect persisted throughout the observation period of 28 days.

DISCUSSION

We observed opposite influences of acute and chronic pain on cortical responses to auditory inputs using rat models. On one hand, N100 wave of rat AEPs was significantly enhanced in rats with acute pain compared to no-pain controls, suggesting that acute pain facilitated cortical processing of

TABLE 2 | Two-way ANOVA exploring the effect of "group" and "time" on latency and amplitude of AEP components (acute pain model).

	Latency						Normalized amplitude					
	N40		**P60**		**N100**		**N40**		**P60**		**N100**	
	F	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**
Group	1.0	0.43	1.3	0.30	2.2	0.10	0.8	0.49	2.1	0.12	6.2	**0.002**
Time	0.5	0.70	8.1	**<0.0001**	25.1	**<0.0001**	0.5	0.61	4.4	**0.02**	8.1	**0.0007**
Interaction	0.7	0.71	1.8	0.08	0.8	0.60	1.0	0.41	2.0	0.08	0.5	0.81

p values in boldface indicate statistically significant results.

FIGURE 3 | The influence of chronic pain on AEPs. (A) Grand-average scalp topographies of N40, P60, and N100 waves. **(B)** For each group, AEP waveforms from different sessions are plotted in different colors and superimposed. Displayed waveforms were measured from fronto-central electrodes (L1, R1, L2, and R2; enclosed by the light gray ellipse in **(A)**, where the N100 wave (marked using gray rectangles) displayed a negative maximum. **(C)** After injection, the normalized N100 amplitudes in the CFA group were significantly smaller than those in the NS group for any post-injection session. NS, normal saline; CFA, complete Freund's adjuvant. *$p < 0.05$, compared to the NS group. NS: $n = 10$; CFA: $n = 11$. Data are expressed as mean $\pm$ SE.

auditory information. On the other hand, N100 wave of rat AEPs was significantly reduced in rats with chronic pain compared to no-pain controls, suggesting that chronic pain inhibited cortical processing of auditory information. Our observations could not be explained by the direct interaction between nociceptive and non-nociceptive sensory inputs, since such interaction could not yield the opposite effects of acute and chronic pain. Instead, our observations could be justified by the fact that individuals who are suffering from acute or chronic pain would have different vigilance states, i.e., the level of vigilance to external sensory stimuli would be increased with acute pain, but decreased with chronic pain. Since the neural processing of auditory information was biased by acute and chronic pain in opposite directions, AEPs might

be used as a representative brain response to distinguish acute pain from chronic pain and to monitor the progress of pain chronification.

Acute Pain Facilitates Cortical Processing of Auditory Information

Pain, in its acute state, serves as a warning signal of tissue damage and induces protective responses that facilitate recuperation (Woolf, 1995; Millan, 1999; Milligan and Watkins, 2009). The presence of acute pain can result in a remarkably heightened level of general arousal and vigilance of the suffered individual (Millan, 1999; Price, 2000), which could be reflected by the increased attention to potential threats or dangers in the environment (Oken et al., 2006). Note that the increased

TABLE 3 | Latency and amplitude of AEP components for different groups and sessions (chronic pain model).

	Latency (ms)			Amplitude (μV)		
	N40	**P60**	**N100**	**N40**	**P60**	**N100**
NS group ($n = 10$)						
Baseline	39.0 ± 0.3	59.8 ± 0.5	105.9 ± 3.0	−7.9 ± 0.6	13.6 ± 1.5	−26.7 ± 4.2
Day 1	39.2 ± 0.3	59.9 ± 0.5	100.6 ± 2.8	−7.8 ± 0.5	12.7 ± 1.0	−25.8 ± 3.7
Day 3	39.5 ± 0.5	60.4 ± 0.5	96.7 ± 1.5	−7.3 ± 0.7	14.7 ± 1.2	−24.3 ± 3.6
Day 7	39.0 ± 0.3	59.3 ± 0.4	95.1 ± 1.7	−7.6 ± 0.6	14.2 ± 1.5	−24.2 ± 3.0
Day 14	38.2 ± 0.5	59.5 ± 0.4	92.2 ± 1.1	−7.5 ± 0.6	14.3 ± 1.8	−27.5 ± 2.9
Day 28	38.8 ± 0.4	60.0 ± 0.3	93.9 ± 1.4	−7.8 ± 0.6	16.7 ± 1.9	−34.6 ± 4.6
CFA group ($n = 11$)						
Baseline	39.0 ± 0.5	60.0 ± 0.6	109.7 ± 4.3	−7.8 ± 0.7	9.7 ± 1.6	−22.0 ± 2.5
Day 1	38.9 ± 0.4	59.7 ± 0.6	99.3 ± 2.4	−6.1 ± 0.9	11.1 ± 1.4	−16.0 ± 2.6
Day 3	39.1 ± 0.5	60.5 ± 0.6	94.5 ± 1.9	−6.0 ± 0.8	11.2 ± 1.4	−16.3 ± 2.7
Day 7	38.6 ± 0.4	60.7 ± 0.4	95.5 ± 1.8	−6.3 ± 0.9	12.7 ± 2.2	−16.6 ± 2.3
Day 14	38.4 ± 0.5	60.7 ± 0.6	94.4 ± 1.4	−6.1 ± 0.4	12.3 ± 1.6	−20.5 ± 2.7
Day 28	38.9 ± 0.5	60.4 ± 0.6	93.0 ± 1.6	−6.2 ± 0.7	12.2 ± 1.8	−25.4 ± 4.3

Data are expressed as mean ± SE. NS, normal saline; CFA, complete Freund's adjuvant.

TABLE 4 | Two-way ANOVA exploring the effect of "group" and "time" on latency and amplitude of AEP components (chronic pain model).

	Latency						Normalized amplitude					
	N40		P60		N100		N40		P60		N100	
	F	p	F	p	F	p	F	p	F	p	F	p
Group	0.07	0.80	0.9	0.38	0.02	0.90	3.5	0.08	1.4	0.26	5.0	**0.038**
Time	3.1	**0.01**	1.1	0.39	26.20	**<0.0001**	0.5	0.75	1.3	0.29	15.40	**<0.0001**
Interaction	0.4	0.83	2.0	0.09	1.1	0.36	0.2	0.96	0.6	0.65	0.2	0.95

p values in boldface indicate statistically significant results.

attention to external changes would be important as it allows the suffered individual to respond properly in life-threatening situations.

Here, we observed a significant enhancement of cortical response to auditory stimuli in rats experiencing acute pain compared to no-pain controls (**Figure 2C**), which indicated that acute pain could facilitate brain responses to external sensory inputs likely through triggering a surge in vigilance. Consistently, as demonstrated in some human brain imaging studies (Peyron et al., 1999, 2000), activations of bilateral thalamus and upper brainstem in response to acute pain were assumed to partly reflect a generalized arousal enhancement. In addition, neural processing of sensory inputs is highly susceptible to fluctuations in vigilance/arousal (Mackworth, 1968; Davis and Whalen, 2001; Oken et al., 2006), demonstrating an enhanced processing at an increased vigilance level (van Marle et al., 2009; Shackman et al., 2011). All these lines of evidence justify the significant influence of acute pain on brain state (i.e., increased vigilance level/attending to external changes), which would subsequently enhance the cortical processing of non-nociceptive sensory information.

Even though we have provided evidence showing that acute pain could influence the brain state (i.e., the vigilance level) significantly, we believe that their relationship is not straightforward. First, we showed that the facilitatory effect of acute pain could be sustained even when the prominent nociceptive behaviors had subsided. This observation would indicate the dissociation between acute pain and brain state (represented by the facilitatory effect) in the perspective of *duration*. Second, although the 5% formalin group showed clearly more intense nociceptive behaviors (**Figure 1A**), the normalized N100 amplitudes of AEPs were not significantly different between rats injected with 1% formalin and those injected with 5% formalin (**Figure 2C**). This observation would demonstrate the dissociation between acute pain and brain state in the perspective of *intensity*. Indeed, the detailed relationship between acute pain and brain state (or the facilitatory effect) should be investigated in the future.

Chronic Pain Inhibits Cortical Processing of Auditory Information

It is well documented that sleep disturbance and fatigue, consequent to the suffering of chronic pain, are of the most common complaints among chronic pain patients (Ashburn and Staats, 1999; Hart et al., 2000; Smith and Haythornthwaite, 2004). As demonstrated by electrophysiological activities and/or

vigilance-related cognitive performance (Belyavin and Wright, 1987; Cajochen et al., 1995, 1999; Cote et al., 2003; Ziino and Ponsford, 2006; Lim and Dinges, 2008), both factors would considerably reduce one's level of vigilance. Following, the attenuated level of vigilance (modulated brain state) could affect the brain responses to sensory inputs (Fruhstor and Bergström, 1969; Corsi-Cabrera et al., 1999; Cote et al., 2003). Moreover, in contrast to acute pain that would increase the individuals' attention to potential threats or dangers in the environment, chronic pain would lead to excessive attention to the internal changes (e.g., hypervigilance to pain and other somatic signals (Eccleston and Crombez, 1999, 2007; Crombez et al., 2005)) in patients. The focus on internal changes in chronic pain state would also result in a decreased level of vigilance to the external environment, and thus lead to a detrimental effect on the processing of pain-irrelevant, external signals.

Here, we observed significant attenuation of cortical response to auditory stimuli in rats with chronic pain compared to no-pain controls (**Figure 3C**). Note that our observation is consistent with some previous reports of sensory impairments in chronic pain patients (Evers et al., 1997; Lorenz et al., 1997; Buodo et al., 2004; Firat et al., 2006; Veldhuijzen et al., 2006; Casale et al., 2008; Korostenskaja et al., 2011). In addition, relevant phenomena have been found in rat models of chronic pain with either inflammatory (Millecamps et al., 2004) or neuropathic (Low et al., 2012) origin, which showed that rats in chronic pain state exhibited decreased ability to perceive small changes in the environment. All these evidences demonstrated that chronic pain could greatly influence the brain state (i.e., decreased vigilance level to external changes), which subsequently attenuated the cortical processing of non-nociceptive sensory information.

Similar to the relationship between acute pain and brain state, the relationship between chronic pain and brain state is also not straightforward. Although we did not assess the influence of chronic pain on brain state in the perspective of *intensity* in the present study, we showed that the inhibitory effect of chronic pain on auditory processing still existed on day 28 (**Figure 3C**) when the thermal hyperalgesia was abolished (**Figure 1B**). This observation would suggest the dissociation between chronic pain and brain state (represented by the inhibitory effect) in the perspective of *duration*. Note that this dissociation would be crucial as it implied that the treatment of chronic pain should not only aim to relieve patients from pain, but also be designed to eliminate possible co-morbidities of chronic pain (e.g., alterations in brain state), especially considering that some

of the co-morbidities could persist even when the chronic pain has been released (Chapman and Dunbar, 1998).

Although we found that the inhibitory effect of chronic pain could be reliably observed throughout our observation period of 28 days, we have also noticed an increase in N100 amplitude on day 14 and day 28 compared to those in the previous sessions in both the pain and no-pain groups (**Figure 3C**). We conjecture that the pronounced restoration or enhancement of N100 amplitude in the last two sessions was due most likely to the prolonged inter-session intervals (1 or 2 weeks) from day 7 to day 28 in contrast to the 2- or 4-day intervals for the previous sessions, which is consistent with the dishabituation effect on event-related potentials after longer intervals during repeated tests as reported previously (Kinoshita et al., 1996).

Transition of Brain States During Pain Chronification

The transition from acute pain to chronic pain has been proven to involve large-scale reorganizations of brain functions (Baliki et al., 2006; Geha et al., 2007; Malinen et al., 2010; Farmer et al., 2011; Parks et al., 2011; Weissman-Fogel et al., 2011). In general, whereas acute pain largely activates brain regions involved in nociceptive information processing (Apkarian et al., 2005), chronic pain is consistently and substantially encoded by brain regions related to emotional and motivational states of patients (Apkarian et al., 2011). A recent longitudinal study illuminated how such a change in brain activation pattern emerged during pain chronification in a group of patients with subacute back pain (Hashmi et al., 2013). It showed that the brain representation for the perception of back pain underwent large-scale reorganization from nociceptive processing regions (including insula, thalamus and anterior cingulate cortex) to emotional relevant circuits (including medial prefrontal cortex and amygdala) as the pain transitioned from subacute state into persistence over a 1-year period (Hashmi et al., 2013). This finding was confirmed by the results of an inter-subject comparison between acute/subacute and chronic back pain patients (Hashmi et al., 2013), as well as the results obtained from other cross-sectional analyses (Apkarian et al., 2005; Baliki et al., 2006, 2010).

Apkarian et al. (2005) pointed out that the increased engagement of cognitive/emotional circuits in chronic pain conditions indicated that chronic pain is different from acute pain in terms of the cognitive, emotional, and introspective components of pain. They expounded this notion in later works, suggesting that a transition in the salience of pain—from viewing a pain perception as an index of external threat to a representation of an internal disease state—is involved in the transition from acute to chronic pain (Apkarian et al., 2009), and may be sufficient to drive the shift in brain representations

of pain perception from acute to chronic conditions (Hashmi et al., 2013). Therefore, acute and chronic pain should not be simply described by different duration of pain, but actually represent two distinct states of the system. Our observation that acute pain enhanced the neural processing of auditory information while chronic pain suppressed it would represent such transition of brain states. For this reason, the AEPs, as a representative brain response to monitor the efficiency of the system to process external sensory inputs, may be potentially used to differentiate the brain states related to acute and chronic pain.

Limitations and Future Directions

We investigated the influences of acute and chronic pain on neural responses to auditory inputs using rat models. Indeed, these influences were observed at limited time points, which hampered us to continually monitor the progress of pain chronification. A longitudinal study that encompasses acute and chronic pain stages, as well as the critical period within which the acute-chronic transition occurs, would be necessary in the future to provide a fine-grained temporal profile of how the brain response changes during pain chronification. In addition, the sensitivity and specificity of using non-nociceptive brain response, e.g., AEPs, to discriminate between acute and chronic pain states should be characterized before using this response to monitor the progress of pain chronification. Importantly, even though animal models have been used to improve our understanding of pain mechanisms, we are aware that the information obtained from animal models cannot be directly applied to humans, and our findings should be replicated in human pain conditions for potential use in the clinic.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Ambrosini, A., Rossi, P., De Pasqua, V., Pierelli, F., and Schoenen, J. (2003). Lack of habituation causes high intensity dependence of auditory evoked cortical potentials in migraine. *Brain* 126, 2009–2015. doi: 10.1093/brain/awg206

Apkarian, A. V., Baliki, M. N., and Geha, P. Y. (2009). Towards a theory of chronic pain. *Prog. Neurobiol.* 87, 81–97. doi: 10.1016/j.pneurobio.2008.09.018

Apkarian, A. V., Bushnell, M. C., Treede, R. D., and Zubieta, J. K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *Eur. J. Pain* 9, 463–484. doi: 10.1016/j.ejpain.2004.11.001

Apkarian, A. V., Hashmi, J. A., and Baliki, M. N. (2011). Pain and the brain: specificity and plasticity of the brain in clinical chronic pain. *Pain* 152, S49–S64. doi: 10.1016/j.pain.2010.11.010

Ashburn, M. A., and Staats, P. S. (1999). Management of chronic pain. *Lancet* 353, 1865–1869. doi: 10.1016/S0140-6736(99)04088-X

Baliki, M. N., Chialvo, D. R., Geha, P. Y., Levy, R. M., Harden, R. N., Parrish, T. B., et al. (2006). Chronic pain and the emotional brain: specific brain activity associated with spontaneous fluctuations of intensity of chronic back pain. *J. Neurosci.* 26, 12165–12173. doi: 10.1523/jneurosci.3576-06.2006

Baliki, M. N., Geha, P. Y., Fields, H. L., and Apkarian, A. V. (2010). Predicting value of pain and analgesia: nucleus accumbens response to noxious stimuli changes in the presence of chronic pain. *Neuron* 66, 149–160. doi: 10.1016/j.neuron.2010.03.002

Baliki, M. N., Petre, B., Torbey, S., Herrmann, K. M., Huang, L. J., Schnitzer, T. J., et al. (2012). Corticostriatal functional connectivity predicts transition to chronic back pain. *Nat. Neurosci.* 15, 1117–1119. doi: 10.1038/nn.3153

Belyavin, A., and Wright, N. A. (1987). Changes in electrical-activity of the brain with vigilance. *Electroencephalogr. Clin. Neurophysiol.* 66, 137–144. doi: 10.1016/0013-4694(87)90183-0

Bingel, U., Rose, M., Gläscher, J., and Büchel, C. (2007). fMRI reveals how pain modulates visual object processing in the ventral visual stream. *Neuron* 55, 157–167. doi: 10.1016/j.neuron.2007.05.032

Blomhoff, S., Jacobsen, M. B., Spetalen, S., Dahm, A., and Malt, U. F. (2000). Perceptual hyperreactivity to auditory stimuli in patients with irritable bowel syndrome. *Scand. J. Gastroenterol.* 35, 583–589. doi: 10.1080/003655200750023534

Buodo, G., Palomba, D., Sarlo, M., Naccarella, C., and Battistella, P. A. (2004). Auditory event-related potentials and reaction times in migraine children. *Cephalalgia* 24, 554–563. doi: 10.1111/j.1468-2982.2003.00716.x

Cajochen, C., Brunner, D. P., Krauchi, K., Graw, P., and Wirz-Justice, A. (1995). Power density in theta/alpha frequencies of the waking EEG progressively increases during sustained wakefulness. *Sleep* 18, 890–894.

Cajochen, C., Khalsa, S. B., Wyatt, J. K., Czeisler, C. A., and Dijk, D. J. (1999). EEG and ocular correlates of circadian melatonin phase and human performance decrements during sleep loss. *Am. J. Physiol.* 277, R640–R649.

Carr, D. B., and Goudas, L. C. (1999). Acute pain. *Lancet* 353, 2051–2058. doi: 10.1016/S0140-6736(99)03313-9

Carrillo-de-la-Peña, M. T., Vallet, M., Pérez, M. I., and Gómez-Perretta, C. (2006). Intensity dependence of auditory-evoked cortical potentials in fibromyalgia patients: a test of the generalized hypervigilance hypothesis. *J. Pain* 7, 480–487. doi: 10.1016/j.jpain.2006.01.452

Casale, M. S., Baratto, M., Cervera, C., Gallamini, M., Lynch, G., Gjini, K., et al. (2008). Auditory evoked potential abnormalities in cluster headache. *Neuroreport* 19, 1633–1636. doi: 10.1097/WNR.0b013e328314e0dd

Chapman, C. R., and Dunbar, P. J. (1998). Measurement in pain therapy: is pain relief really the endpoint? *Curr. Opin. Anaesthesiol.* 11, 533–537. doi: 10.1097/00001503-199810000-00015

Chou, R., Huffman, L. H., American Pain Society, and American College of Physicians. (2007a). Medications for acute and chronic low back pain: a review of the evidence for an american pain society/american college of physicians clinical practice guideline. *Ann. Intern. Med.* 147, 505–514. doi: 10.7326/0003-4819-147-7-200710020-00008

Chou, R., Huffman, L. H., American Pain Society, and American College of Physicians. (2007b). Nonpharmacologic therapies for acute and chronic low back pain: A review of the evidence for an American pain Society/American college of physicians clinical practice guideline. *Ann. Intern. Med.* 147, 492–504. doi: 10.7326/0003-4819-147-7-200710020-00007

Corsi-Cabrera, M., Arce, C., Del Río-Portilla, I. Y., Pérez-Garci, E., and Guevara, M. A. (1999). Amplitude reduction in visual event-related potentials as a function of sleep deprivation. *Sleep* 22, 181–189.

Cote, K. A., Milner, C. E., Osip, S. L., Ray, L. B., and Baxter, K. D. (2003). Waking quantitative electroencephalogram and auditory event-related potentials following experimentally induced sleep fragmentation. *Sleep* 26, 687–694.

Crombez, G., Van Damme, S., and Eccleston, C. (2005). Hypervigilance to pain: an experimental and clinical analysis. *Pain* 116, 4–7. doi: 10.1016/j.pain.2005.03.035

Davis, M., and Whalen, P. J. (2001). The amygdala: vigilance and emotion. *Mol. Psychiatry* 6, 13–34. doi: 10.1038/sj.mp.4000812

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Dubuisson, D., and Dennis, S. G. (1977). The formalin test: a quantitative study of the analgesic effects of morphine, meperidine and brain stem stimulation in rats and cats. *Pain* 4, 161–174. doi: 10.1016/0304-3959(77)90130-0

Eccleston, C., and Crombez, G. (1999). Pain demands attention: a cognitive-affective model of the interruptive function of pain. *Psychol. Bull.* 125, 356–366. doi: 10.1037/0033-2909.125.3.356

Eccleston, C., and Crombez, G. (2007). Worry and chronic pain: a misdirected problem solving model. *Pain* 132, 233–236. doi: 10.1016/j.pain.2007.09.014

Evers, S., Bauer, B., Suhr, B., Husstedt, I. W., and Grotemeyer, K. H. (1997). Cognitive processing in primary headache: a study on event-related potentials. *Neurology* 48, 108–113. doi: 10.1212/wnl.48.1.108

Farmer, M. A., Chanda, M. L., Parks, E. L., Baliki, M. N., Apkarian, A. V., and Schaeffer, A. J. (2011). Brain functional and anatomical changes in chronic prostatitis/chronic pelvic pain syndrome. *J. Urol.* 186, 117–124. doi: 10.1016/j.juro.2011.03.027

Firat, Y., Ozturan, O., Bicak, U., Yakinci, C., and Akarcay, M. (2006). Auditory brainstem response in pediatric migraine: during the attack and asymptomatic period. *Int. J. Pediatr. Otorhinolaryngol.* 70, 1431–1438. doi: 10.1016/j.ijporl.2006.02.013

Fontanini, A., and Katz, D. B. (2008). Behavioral states, network states and sensory response variability. *J. Neurophysiol.* 100, 1160–1168. doi: 10.1152/jn.90592.2008

Fruhstor, H., and Bergström, R. M. (1969). Human vigilance and auditory evoked responses. *Electroencephalogr. Clin. Neurophysiol.* 27, 346–355. doi: 10.1016/0013-4694(69)91443-6

Geha, P. Y., Baliki, M. N., Chiaivo, D. R., Harden, R. N., Paice, J. A., and Apkarian, A. V. (2007). Brain activity for spontaneous pain of postherpetic neuralgia and its modulation by lidocaine patch therapy. *Pain* 128, 88–100. doi: 10.1016/j.pain.2006.09.014

Gilbert, C. D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696. doi: 10.1016/j.neuron.2007.05.019

Hargreaves, K., Dubner, R., Brown, F., Flores, C., and Joris, J. (1988). A new and sensitive method for measuring thermal nociception in cutaneous hyperalgesia. *Pain* 32, 77–88. doi: 10.1016/0304-3959(88)90026-7

Hart, R. P., Martelli, M. F., and Zasler, N. D. (2000). Chronic pain and neuropsychological functioning. *Neuropsychol. Rev.* 10, 131–149. doi: 10.1023/A:1009020914358

Hashmi, J. A., Baliki, M. N., Huang, L. J., Baria, A. T., Torbey, S., Hermann, K. M., et al. (2013). Shape shifting pain: chronification of back pain shifts brain representation from nociceptive to emotional circuits. *Brain* 136, 2751–2768. doi: 10.1093/brain/awt211

Johnson, M. R., and Adler, L. E. (1993). Transient impairment in P50 auditory sensory gating induced by a cold-pressor test. *Biol. Psychiatry* 33, 380–387. doi: 10.1016/0006-3223(93)90328-b

Jung, F., Stephan, K. E., Backes, H., Moran, R., Gramer, M., Kumagai, T., et al. (2013). Mismatch responses in the awake rat: evidence from epidural recordings of auditory cortical fields. *PLoS One* 8:e63203. doi: 10.1371/journal.pone.0063203

Kinoshita, S., Inoue, M., Maeda, H., Nakamura, J., and Morita, K. (1996). Long-term patterns of change in ERPs across repeated measurements. *Physiol. Behav.* 60, 1087–1092. doi: 10.1016/0031-9384(96)00130-8

Knight, R. T., Brailowsky, S., Scabini, D., and Simpson, G. V. (1985). Surface auditory evoked potentials in the unrestrained rat: component definition. *Electroencephalogr. Clin. Neurophysiol.* 61, 430–439. doi: 10.1016/0013-4694(85)91035-1

Korostenskaja, M., Pardos, M., Kujala, T., Rose, D. F., Brown, D., Horn, P., et al. (2011). Impaired auditory information processing during acute migraine: a magnetoencephalography study. *Int. J. Neurosci.* 121, 355–365. doi: 10.3109/00207454.2011.560312

Lavand'homme, P. (2011). The progression from acute to chronic pain. *Curr. Opin. Anaesthesiol.* 24, 545–550. doi: 10.1097/ACO.0b013e32834a4f74

Lazar, R., and Metherate, R. (2003). Spectral interactions, but no mismatch negativity, in auditory cortex of anesthetized rat. *Hear. Res.* 181, 51–56. doi: 10.1016/s0378-5955(03)00166-7

Li, S. G., Wang, J. Y., and Luo, F. (2012). Adult-age inflammatory pain experience enhances long-term pain vigilance in rats. *Plos One* 7:e36767. doi: 10.1371/journal.pone.0036767

Lim, J., and Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Ann. N Y Acad. Sci.* 1129, 305–322. doi: 10.1196/annals.1417.002

Loeser, J. D., and Melzack, R. (1999). Pain: an overview. *Lancet* 353, 1607–1609. doi: 10.1016/S0140-6736(99)01311-2

Lorenz, J., and Bromm, B. (1997). Event-related potential correlates of interference between cognitive performance and tonic experimental pain. *Psychophysiology* 34, 436–445. doi: 10.1111/j.1469-8986.1997.tb02387.x

Lorenz, J., Beck, H., and Bromm, B. (1997). Cognitive performance, mood and experimental pain before and during morphine-induced analgesia in patients with chronic non-malignant pain. *Pain* 73, 369–375. doi: 10.1016/s0304-3959(97)00123-1

Low, L. A., Millecamps, M., Seminowicz, D. A., Naso, L., Thompson, S. J., Stone, L. S., et al. (2012). Nerve injury causes long-term attentional deficits in rats. *Neurosci. Lett.* 529, 103–107. doi: 10.1016/j.neulet.2012.09.027

Mackworth, J. F. (1968). Vigilance, arousal and habituation. *Psychol. Rev.* 75, 308–322. doi: 10.1037/h0025896

Malinen, S., Vartiainena, N., Hlushchuk, Y., Koskinen, M., Ramkumar, P., Forss, N., et al. (2010). Aberrant temporal and spatial brain activity during rest in patients with chronic pain. *Proc. Natl. Acad. Sci. U S A* 107, 6493–6497. doi: 10.1073/pnas.1001504107

Mansour, A. R., Baliki, M. N., Huang, L. J., Torbey, S., Herrmann, K. M., Schnitzer, T. J., et al. (2013). Brain white matter structural properties predict transition to chronic pain. *Pain* 154, 2160–2168. doi: 10.1016/j.pain.2013.06.044

May, A. (2008). Chronic pain may change the structure of the brain. *Pain* 137, 7–15. doi: 10.1016/j.pain.2008.02.034

Merskey, H., and Bogduk, N. (Ed.) (1994). *Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms.* 2nd Edn. Seattle: IASP Press.

Millan, M. J. (1999). The induction of pain: an integrative review. *Prog. Neurobiol.* 57, 1–164. doi: 10.1016/s0301-0082(98)00048-3

Millecamps, M., Etienne, M., Jourdan, D., Eschalier, A., and Ardid, D. (2004). Decrease in non-selective, non-sustained attention induced by a chronic visceral inflammatory state as a new pain evaluation in rats. *Pain* 109, 214–224. doi: 10.1016/j.pain.2003.12.028

Milligan, E. D., and Watkins, L. R. (2009). Pathological and protective roles of glia in chronic pain. *Nat. Rev. Neurosci.* 10, 23–36. doi: 10.1038/nrn2533

Nakamura, T., Michie, P. T., Fulham, W. R., Todd, J., Budd, T. W., Schell, U., et al. (2011). Epidural auditory event-related potentials in the rat to frequency and duration deviants: evidence of mismatch negativity? *Front. Psychol.* 2:367. doi: 10.3389/fpsyg.2011.00367

Oken, B. S., Salinsky, M. C., and Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin. Neurophysiol.* 117, 1885–1901. doi: 10.1016/j.clinph.2006.01.017

Parks, E. L., Geha, P. Y., Baliki, M. N., Katz, J., Schnitzer, T. J., and Apkarian, A. V. (2011). Brain activity for chronic knee osteoarthritis: dissociating evoked pain from spontaneous pain. *Eur. J. Pain* 15, 843.e1–843.e14. doi: 10.1016/j.ejpain.2010.12.007

Paxinos, G., and Watson, C. (2007). *The Rat Brain in Stereotaxic Coordinates.* 6th Edn. London, Amsterdam, Burlington: Academic Press.

Peyron, R., García-Larrea, L., Grégoire, M. C., Costes, N., Convers, P., Lavenne, F., et al. (1999). Haemodynamic brain responses to acute pain in humans–Sensory and attentional networks. *Brain* 122, 1765–1780. doi: 10.1093/brain/122.9.1765

Peyron, R., Laurent, B., and García-Larrea, L. (2000). Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiol. Clin.* 30, 263–288. doi: 10.1016/s0987-7053(00)00227-6

Porro, C. A., and Cavazzuti, M. (1993). Spatial and temporal aspects of spinal cord and brainstem activation in the formalin pain model. *Prog. Neurobiol.* 41, 565–607. doi: 10.1016/0301-0082(93)90044-s

Price, D. D. (2000). Neuroscience–Psychological and neural mechanisms of the affective dimension of pain. *Science* 288, 1769–1772. doi: 10.1126/science.288.5472.1769

Sambeth, A., Maes, J. H. R., van Luijtelaar, G., Molenkamp, I. B. S., Jongsma, M. L. A., and Van Rijn, C. M. (2003). Auditory event-related potentials in humans and rats: effects of task manipulation. *Psychophysiology* 40, 60–68. doi: 10.1111/1469-8986.00007

Shackman, A. J., Maxwell, J. S., McMenamin, B. W., Greischar, L. L., and Davidson, R. J. (2011). Stress potentiates early and attenuates late stages of visual processing. *J. Neurosci.* 31, 1156–1161. doi: 10.1523/JNEUROSCI.3384-10.2011

Shinba, T. (1997). Event-related potentials of the rat during active and passive auditory oddball paradigms. *Electroencephalogr. Clin. Neurophysiol.* 104, 447–452. doi: 10.1016/s0168-5597(97)00047-6

Smith, M. T., and Haythornthwaite, J. A. (2004). How do sleep disturbance and chronic pain inter-relate? Insights from the longitudinal and cognitive-behavioral clinical trials literature. *Sleep Med. Rev.* 8, 119–132. doi: 10.1016/s1087-0792(03)00044-3

van Marle, H. J., Hermans, E. J., Qin, S., and Fernandez, G. (2009). From specificity to sensitivity: how acute stress affects amygdala processing of biologically salient stimuli. *Biol. Psychiatry* 66, 649–655. doi: 10.1016/j.biopsych.2009.05.014

Veldhuijzen, D. S., Kenemans, J. L., van Wijck, A. J. M., Olivier, B., Kalkman, C., and Volkerts, E. R. (2006). Processing capacity in chronic pain patients: a visual event-related potentials study. *Pain* 121, 60–68. doi: 10.1016/j.pain.2005.12.004

Wang, W., Wang, Y. H., Fu, X. M., Sun, Z. M., and Schoenen, J. (1999). Auditory evoked potentials and multiple personality measures in migraine and post-traumatic headaches. *Pain* 79, 235–242. doi: 10.1016/s0304-3959(98)00168-7

Wang, N., Wang, J. Y., and Luo, F. (2009). Corticofugal outputs facilitate acute, but inhibit chronic pain in rats. *Pain* 142, 108–115. doi: 10.1016/j.pain.2008.12.016

Weissman-Fogel, I., Moayedi, M., Tenenbaum, H. C., Goldberg, M. B., Freeman, B. V., and Davis, K. D. (2011). Abnormal cortical activity in patients with temporomandibular disorder evoked by cognitive and emotional tasks. *Pain* 152, 384–396. doi: 10.1016/j.pain.2010.10.046

Wiech, K., Ploner, M., and Tracey, I. (2008). Neurocognitive aspects of pain perception. *Trends Cogn. Sci.* 12, 306–313. doi: 10.1016/j.tics.2008.05.005

Witten, L., Oranje, B., Mørk, A., Steiniger-Brach, B., Glenthøj, B. Y., and Bastlund, J. F. (2014). Auditory sensory processing deficits in sensory gating and mismatch negativity-like responses in the social isolation rat model of schizophrenia. *Behav. Brain Res.* 266, 85–93. doi: 10.1016/j.bbr.2014.02.048

Woolf, C. J. (1995). Somatic pain–pathogenesis and prevention. *Br. J. Anaesth.* 75, 169–176. doi: 10.1093/bja/75.2.169

Ziino, C., and Ponsford, J. (2006). Vigilance and fatigue following traumatic brain injury. *J. Int. Neuropsychol. Soc.* 12, 100–110. doi: 10.1017/s1355617706060139

Placebo Analgesia Changes Alpha Oscillations Induced by Tonic Muscle Pain: EEG Frequency Analysis Including Data during Pain Evaluation

Linling Li [1,2†], Hui Wang [1,2†], Xijie Ke [3†], Xiaowu Liu [1], Yuan Yuan [1], Deren Zhang [3], Donglin Xiong [3] and Yunhai Qiu [1]*

[1] Research Center for Neural Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [2] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China, [3] Department of Pain, Shenzhen Sixth People's Hospital (Nanshan Hospital), Guangdong Medical College, Shenzhen, China

Placebo exhibits beneficial effects on pain perception in human experimental studies. Most of these studies demonstrate that placebo significantly decreased neural activities in pain modulatory brain regions and pain-evoked potentials. This study examined placebo analgesia-related effects on spontaneous brain oscillations. We examined placebo effects on four order-fixed 20-min conditions in two sessions: isotonic saline-induced control conditions (with/without placebo) followed by hypertonic saline-induced tonic muscle pain conditions (with/without placebo) in 19 subjects using continuous electroencephalography (EEG) recording. Placebo treatment exerted significant analgesic effects in 14 placebo responders, as subjective intensity of pain perception decreased. Frequency analyses were performed on whole continuous EEG data, data during pain perception rating and data after rating. The results in the first two cases revealed that placebo induced significant increases and a trend toward significant increases in the amplitude of alpha oscillation during tonic muscle pain compared to control conditions in frontal-central regions of the brain, respectively. Placebo-induced decreases in the subjective intensity of pain perception significantly and positively correlated with the increases in the amplitude of alpha oscillations during pain conditions. In conclusion, the modulation effect of placebo treatment was captured when the pain perception evaluating period was included. The strong correlation between the placebo effect on reported pain perception and alpha amplitude suggest that alpha oscillations in frontal-central regions serve as a cortical oscillatory basis of the placebo effect on tonic muscle pain. These results provide important evidence for the investigation of objective indicators of the placebo effect.

Keywords: placebo, EEG, tonic muscle pain, pain perception, alpha oscillation

INTRODUCTION

Placebo effects on pain perception were characterized using numerous hemodynamic (e.g., functional magnetic resonance imaging (fMRI) and positron emission tomography (PET)) and electrophysiological (e.g., electroencephalography (EEG) and magnetoencephalography (MEG)) in previous studies (Wager et al., 2004, 2006; Lorenz et al., 2005; Zubieta et al., 2005;

Scott et al., 2008; Tracey, 2010). Most of these studies demonstrated that placebo analgesia significantly decreased neural activities in pain modulatory brain regions, including thalamus, insula, and anterior cingulate cortex (ACC; Wager and Fields, 2011). Laser-evoked potentials (LEPs) are one of the best tools to assess the function of nociceptive pathways in physiological and clinical settings (Bromm and Treede, 1991; Iannetti et al., 2001), and LEPs were used in previous studies to investigate placebo analgesia (Wager et al., 2006; Watson et al., 2007). These studies demonstrated a clear decrease in P2 amplitude using LEPs (Wager et al., 2006), which suggests that the placebo treatment affected early nociceptive processing (e.g., attention and affect). One recent study reported that placebo analgesia during phasic pain was associated with changes in pain-evoked potentials but not oscillatory activities (Tiemann et al., 2015).

Reports of placebo effects in healthy subjects were primarily based on duration limited phasic pain (Atlas et al., 2009; Benedetti, 2009). Phasic pain provides some important methodological benefits (e.g., safe and easy to apply repeatedly), but it is too short to faithfully simulate clinical pain, which is rarely brief and exhibits an explicit onset of pain perception. Therefore, several studies proposed tonic pain models, which are crucial to model the pain experience in clinical settings (Le Pera et al., 2000; Chang et al., 2001, 2002, 2003, 2004; Huber et al., 2006; Dowman et al., 2008; Nir et al., 2010). One tonic pain model uses pain originating from deep tissue, such as intramuscular infusions of capsaicin or hypertonic saline, which is most frequently encountered in clinical practice pain (Apkarian et al., 2005). The present study used a prolonged muscle infusion of hypertonic saline to generate tonic muscle pain (Stohler and Kowalski, 1999). Hypertonic saline was continuously infused to maintain a relatively stable pain sensation based on real-time feedback of subjective pain intensity (Stohler, 1992).

FIGURE 1 | Experimental design and subjective pain intensity perception. (A) The experiment consisted of four order-fixed 20-min conditions in two sessions. Session 1: (I) control, (II) pain; Session 2: (III) control with placebo, (IV) pain with placebo. Innocuous and noxious stimulations were respectively applied in control conditions (I and III) and pain conditions (II and IV). **(B)** The subjective pain intensity (mean ± SEM) was collected every 15 s for each condition from all placebo responders ($N = 14$).

We collected continuous EEG data during tonic muscle pain to assess the effect of placebo treatment on: (1) the subjective perception of tonic pain; (2) the electrophysiological oscillatory activities; and (3) the correlations between changes in pain perception and oscillatory activities.

MATERIALS AND METHODS

Subjects
The study included 19 subjects (3 females and 16 males, mean age: 23 ± 2 years). All subjects were nonsmokers with no personal history of any neurological or psychiatric disease. None of the subjects had any history of chronic or acute pain up to 4 weeks before and during the study period, and none of the subjects was on any medication. All subjects provided informed consent, and the Human Research Ethics Committee of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences approved the experimental procedures.

Experimental Design
The experiment consisted of four order-fixed 20-min conditions in two sessions (**Figure 1A**): session 1: (I) control, (II) pain; and session 2: (III) control with placebo, (IV) pain with placebo. Subjects were informed that the impending sequential intramuscular injections were possibly painful or non-painful before each session. Experiments were conducted in a silent and separate room, and subjects were comfortably seated in a chair. Subjects were required to rate the intensity of pain perception every 15 s on a computer-controlled visual analog scale (VAS) ranging from 0 to 10 (0: no pain; 10: the most pain intensity imaginable) during all conditions. A moving bar was used to indicate VAS ratings, which were displayed on a monitor in front of the subjects. Subjects indicated the intensity of pain perception by pressing a keyboard key to stop the moving bar with their left hand (the moving bar ascended one score per second). Subjects were asked to arbitrarily choose given scores on the VAS every 15 s until their response was sufficiently accurate to familiarize subjects with the rating paradigm.

We used an automated stimulus delivery system in this study. We used two 24-gauge needles, and each needle was attached to a syringe through a disposable tube. The outline of the masseter muscle was established during clenching. The needles were inserted in bilateral masseter muscles to a depth of approximately 1 cm. Prolonged innocuous stimulation was introduced during control conditions (I and III) via infusions of medication-grade isotonic saline (0.9% NaCl) in the right masseter muscle. During pain conditions (II and IV), prolonged noxious stimulation was introduced by infusing hypertonic saline (5% NaCl) in the left masseter muscle. Automated syringe infusion pumps controlled the infusions. Isotonic saline was infused at a constant speed of 75 μl/min during innocuous stimulation (1500 μl in total). Noxious stimulation included a 0.2-ml bolus infusion over 15 s at the beginning and subsequent continuous infusions at variable speeds (2134 ± 930 μl in total). The speed of infusion was adjusted using a computer-

controlled closed-loop system based on the real-time feedback of pain perception to ensure perceived pain intensity maintained at an approximate VAS level of 5 (Zhang et al., 1993; Stohler and Kowalski, 1999). The adaptive controller identified the system dynamic response and proportional-integral-derivative (PID) controller parameters from the subjects' initial response to the bolus infusion (Zhang et al., 1993). The intramuscular infusion of hypertonic saline produced a deep aching sensation that was similar to chronic muscle pain, and the generated pain sensation disappeared 5–10 min after cessation of the hypertonic saline infusion (Stohler and Kowalski, 1999; Zubieta et al., 2005). Consecutive sessions were separated by at least 10 min.
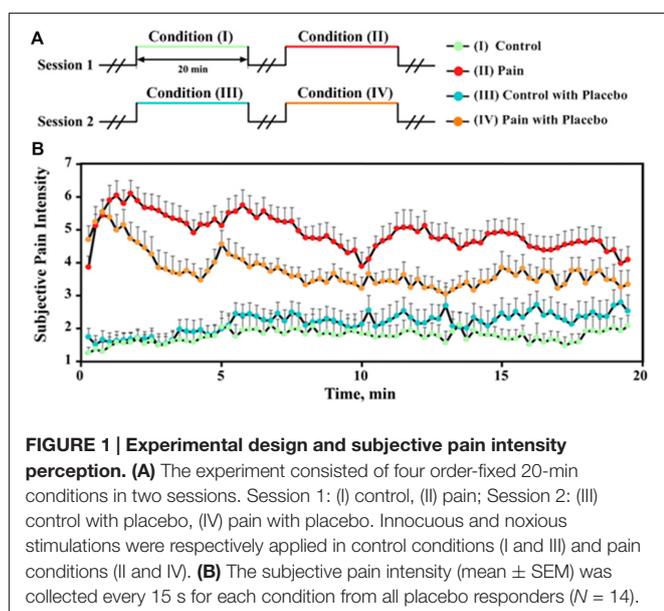
Subjects were infused with isotonic saline (0.9% NaCl) via an antecubital intravenous port in their right upper limb during all four conditions. However, the subjects were told that the isotonic saline was replaced by a novel medication named "Entacapone" prior to conditions with placebo (III and IV), and they were further given the following clinical trial-type instruction: "we are studying the analgesic effect of a novel medication named "Entacapone", and it may or may not ease your pain" (Zubieta et al., 2005; Scott et al., 2007). The same infusion profile of noxious stimulation was applied for pain (II) and pain with placebo (IV) for each subject (Scott et al., 2008).

Subjects were instructed to fill out the Chinese version of the Positive and Negative Affective Scale (PANAS; Watson et al., 1988) and McGill Pain Questionnaire (SF-MPQ; Melzack, 1987) after each condition (I–IV) to provide details of their subjective perceptions of pain. The Chinese version of these questionnaires exhibits acceptable reliability and validity (Huang et al., 2003; Li et al., 2013).

Behavioral Data Analysis

The average rating of pain intensity across all rating points (once every 15 s) was calculated for each subject during each condition. Subjects who reported an increase in the average rating of the intensity of pain perception to noxious stimulation after the placebo treatment (II vs. IV) were classified as nocebo responders, and the other subjects were classified as placebo responders (Scott et al., 2008). Previous studies reported that placebo and nocebo effects were associated with opposite responses of dopamine and endogenous opioid neurotransmission in a distributed network of cortical and subcortical regions (Scott et al., 2008). Therefore, psychophysical and electrophysiological data from nocebo responders were excluded from subsequent analyses.

Psychophysical data analyses were performed as follows. The ratings of pain perception, positive affect ratings (PANAS-P) and negative affect ratings (PANAS-N) were compared across all four conditions using a two-way repeated-measures analysis of variance (RM ANOVA), with "pain" (two levels: control vs. pain) and "placebo" (two levels: without vs. with placebo) as factors. *Post hoc* tests were performed when the interaction effect was significant. Not all subjects finished the SF-MPQ questionnaire after control conditions (I and III), so the total MPQ sensory (MPQ-S) and affective (MPQ-A) scores of only pain conditions

(II and IV) were calculated for each subject. The scores were compared between two pain conditions using a two-tailed paired sample *t*-test.

EEG Recording and Data Analysis

Continuous EEG data were recorded using a Neuroscan® Scan 4.2 (Neuroscan, Charlotte, NC, USA) amplifier and 128 Ag/AgCl electrodes mounted on an elastic cap (Quickap®, Neuromedical supplies, Charlotte, NC, USA) according to the extended international 10–20 system (Aslaksen et al., 2007). The reference channel was located at the vertex, and all channel impedances were kept lower than 10 kΩ. Extracranial activity was continuously recorded with a 0.05 Hz and 100 Hz band-pass filter and was digitized at a sampling rate of 1000 Hz. A notch filter was set to 50 Hz to reduce electrical interference. Electro-oculographic (EOG) signals were simultaneously recorded from four surface electrodes (one pair over the upper and lower eyelids; the other pair placed 1 cm lateral to the outer corner of the left and right orbits) to monitor ocular movements and eye blinks. Subjects were instructed to relax and keep their eyes open during each condition.

Preprocessing

EEG data were analyzed using Matlab (The Mathwork, Natick, MA, USA) and EEGLAB[1], which is an open source toolbox running under the Matlab environment. Continuous EEG data for each condition were down-sampled to 500 Hz and band-pass filtered between 1 and 100 Hz. Continuous EEG data contaminated by eye-blinks and movements were corrected using an independent component analysis (ICA) algorithm (Makeig et al., 1997; Jung et al., 2001; Delorme and Makeig, 2004). The de-noised EEG data were re-referenced to a common average reference. EEG data collected during a short period of 30 s at the beginning and end of each condition were discarded to exclude possible brain responses related to the sudden change in stimulation.

EEG Spectral Analysis

Nineteen minutes of continuous EEG data from each subject and condition were transformed to the frequency domain using a discrete Fourier transform to yield amplitude spectra (in μV) ranging from 1 to 100 Hz. The amplitudes of EEG oscillations in the delta (0−4 Hz), theta (4−8 Hz), alpha (8−12 Hz), beta (12−30 Hz), and gamma (30−100 Hz) bands were calculated for each condition and electrode, and the first group of amplitude spectra was obtained.

Previous studies generally used verbal pain perception ratings. EEG data during ratings were excluded because of the possible confounding factor of speaking. Subjects in this experiment indicated the intensity of pain perception by pressing a keyboard key to stop a moving bar with their left hand. This pain rating procedure required a longer time than verbal pain rating because the moving bar ascended one score per second. We investigated whether the inclusion of the EEG

---

[1]http://sccn.ucsd.edu/eeglab

data during pain perception was important for the extraction of placebo-related modulation effects. Therefore, additional separate analyses were performed with the EEG data during pain perception rating and EEG data after rating. Pain perception ratings were repeated once in every 15 s. Subjects pressed a button when the moving VAS bar indicated their pain intensity. Therefore, we partitioned the EEG data based on the time point when the pain intensity rating was completed. The original preprocessed continuous EEG data were segmented into EEG epochs of 1 s, and the segmented EEG epochs were transformed to the frequency domain for each subject and each condition to facilitate the partition. The obtained single-epoch amplitude spectra according to time period during rating were averaged for each electrode and condition to provide another group of amplitude spectra. The numbers of segments during ratings were 2.60 ± 1.17, 5.39 ± 1.19, 2.25 ± 1.15 and 4.19 ± 0.70 in conditions I, II, III, and IV, respectively. The obtained single-epoch amplitude spectra according to time periods after VAS rating were also extracted to provide a third group of amplitude spectra. The numbers of segments after ratings were 12.40 ± 1.17, 9.61 ± 1.19, 12.75 ± 1.15 and 10.81 ± 0.70 in conditions I, II, III, and IV, respectively.

All three groups of amplitude spectra were compared across all four conditions using point-by-point two-way RM ANOVA with "pain" (two levels: control vs. pain) and "placebo" (two levels: without vs. with placebo) as factors. Considering the two-by-two experimental design, significant interaction effect indicated the placebo effect. A permutation test with 5000 iterations was used to construct the null distribution of the max F-statistic across electrodes to control for multiple comparisons. We identified the F-statistic that corresponded to the 5% most extreme parts of the maximal F distribution. We thresholded our original statistical maps at that 5% level from the maximal F distribution (Maris and Oostenveld, 2007). Compared with this F-statistic, higher F value represented significant result after correction. Besides, we calculated the corrected P value of our observed F value by counting the proportion of the permutation distribution as or more extreme than F. Results of main effects and *post hoc* tests were presented when the interaction effect was significant.

Correlation Analysis

The correlation coefficients and significance of placebo responders were calculated between changes in the amplitude of alpha oscillation measured at frontal-central electrode FCz after placebo treatment (IV–II) and changes in: (1) subjective intensity of pain perception to noxious stimulation; and (2) psychophysical scores (i.e., PANAS and MPQ scores; II–IV). Besides, in order to keep consistent with the two-by-two experimental design, correlation analysis was also performed with changes which were calculated according to the interaction effect ((IV–II)–(III–I)) for alpha amplitude; ((II–IV)–(I–III)) for subjective intensity).

RESULTS

Psychophysical Results

The subjective intensity of pain perception to noxious stimulation increased after the placebo treatment (IV vs. II) in five subjects (nocebo responders) and decreased in the remaining 14 subjects (placebo responders). The intensity of pain perception to noxious stimulation for placebo responders revealed an overall declining tendency with increased stimulus duration (II and IV; **Figure 1B**), which may be due to the limitation of the maximum speed of hypertonic saline infusion. In contrast, the intensity of pain perception to innocuous stimulation was approximately a VAS level of 2 and increased slightly with increased stimulus duration (I and III; **Figure 1B**), which may be caused by the needle effect (Veerasarn and Stohler, 1992).

Table 1 summarizes the average ratings of subjective pain intensity, PANAS scores, and MPQ scores for placebo responders. The intensity of pain perception was significantly modulated by the factors "placebo" ($F_{(1,13)} = 25.889$, $P = 0.000$) and "pain" ($F_{(1,13)} = 105.663$, $P = 0.0000$) and the interaction between two factors ($F_{(1,13)} = 5.748$, $P = 0.032$; **Figure 2C**). The decrease in pain intensity to noxious stimulation was significant after placebo treatment (II vs. IV; $P = 0.000$), but only marginally significant to innocuous stimulation (I vs. III; $P = 0.058$). The PANAS-P scores were not significantly modulated by the factor "pain" ($F_{(1,13)} = 1.050$, $P = 0.324$) or "placebo" ($F_{(1,13)} = 2.444$, $P = 0.142$), or the interaction between the two factors ($F_{(1,13)} = 0.918$, $P = 0.356$). The PANAS-N scores were significantly modulated by the factor "placebo" ($F_{(1,13)} = 8.050$, $P = 0.014$) but not the factor "pain" ($F_{(1,13)} = 1.518$, $P = 0.240$) or the interaction between the two factors ($F_{(1,13)} = 1.194$, $P = 0.294$). MPQ-S scores decreased significantly in condition IV compared to condition II ($t_{(13)} = 2.230$, $P = 0.044$). In contrast, the MPQ-A scores were not

TABLE 1 | Psychophysical responses of placebo responders.

	Pain Intensity	PANAS-P	PANAS-N	MPQ-S	MPQ-A
Condition I	2.16 ± 1.17	23.14 ± 6.13	14.64 ± 3.88	–	–
Condition II	4.93 ± 1.21	21.36 ± 6.49	16.50 ± 5.49	9.93 ± 6.06	6.21 ± 3.95
Condition III	1.79 ± 1.11	20.64 ± 6.38	13.43 ± 3.16	–	–
Condition IV	3.75 ± 0.71	20.50 ± 6.47	13.36 ± 2.95	3.36 ± 2.27	2.07 ± 1.64

Mean ± 1 SD of psychophysical measures during conditions in the absence and presence of placebo in 14 placebo responders. Pain intensity refers to the average ratings of momentary pain acquired every 15 s.

significantly different between conditions II and IV ($t_{(13)}$ = 1.906, P = 0.079).

Electrophysiological Results

Frequency analyses of the 19-min continuous EEG data revealed that the group level scalp topographies of alpha oscillations were maximal at bilateral posterior parietal and occipital regions in all four conditions (**Figure 2A**). Point-by-point two-way RM ANOVA revealed that electrode FCz exhibited a significant interaction effect on the amplitudes of alpha oscillations after correction for multiple comparisons (**Figure 2B**). The amplitudes of alpha oscillations at FCz were 6.56 ± 2.19 µV, 5.70 ± 1.80 µV, 6.52 ± 2.23 µV,

and 6.18 ± 2.04 µV in conditions I, II, III, and IV respectively. The amplitudes of alpha oscillations at FCz were significantly modulated by the factor "pain" ($F_{(1,13)}$ = 13.886, P = 0.040, corr.) and the interaction between the two factors ($F_{(1,13)}$ = 13.003, P = 0.046, corr.; **Figure 2D**), but not by the factor "placebo" ($F_{(1,13)}$ = 1.483, P = 0.864, corr). *Post hoc* tests revealed that the amplitudes of alpha oscillations were significantly larger in condition IV than condition II (P = 0.005), but no significant difference was observed between the amplitudes of alpha oscillations in conditions I and III (P = 0.846).

Frequency analyses results of EEG data during pain perception ratings revealed that electrode FCz exhibited a

FIGURE 2 | Evidence showing the effect of placebo treatment from behavioral and EEG data. (A) Group level scalp topographies of alpha oscillations (8–12 Hz) of different experimental conditions. **(B)** Group level spectra (measured at FCz) of different experimental conditions. Scalp topography showing the significant interaction between the factors "pain" and "placebo" on the amplitudes of alpha oscillations at FCz is displayed in the insert. **(C)** Significant interaction effect between the factors "pain" and "placebo" was observed on the average ratings of pain intensity across all rating points (once every 15 s; left). **(D)** The amplitudes of alpha oscillation (measured at FCz). Each dot represents the mean value from one condition, and error bars represent, for each condition, ± SEM across subjects (F: F value of the interaction effect between the factors "pain" and "placebo"; corr.: corrected for multiple comparisons). **(E)** Significant correlation was observed between decrease in pain intensity during noxious stimulation after placebo treatment (II–IV) and the increase in the amplitude of alpha oscillation measured at FCz (IV–II). Each dot represents a value from each subject, and black line represents the best linear fit.

trend toward significant interaction effect between the factors "pain" and "placebo" on the amplitudes of alpha oscillation ($F_{(1,13)} = 8.065$, $P = 0.014$, uncorr., $P = 0.138$, corr.; left and middle panels of **Figure 3A**). *Post hoc* tests revealed that the amplitudes of alpha oscillations were significantly larger in condition IV than condition II ($P = 0.004$), but no significant difference was observed between conditions I and III ($P = 0.638$). Analysis results of EEG data after VAS ratings revealed that the interaction effect was not significant ($F_{(1,13)} = 4.564$, $P = 0.222$, corr.; left and middle panels of **Figure 3B**).

Correlation between Psychophysical and Electrophysiological Data

First, the correlation analysis was performed with changes between two pain conditions (II vs. IV). Significant positive correlation was observed between increases in the amplitudes of alpha oscillations measured at FCz after placebo treatment and decreases in: (1) subjective intensity of pain perception ($R = 0.611$, $P = 0.020$; **Figure 2E**); (2) MPQ-S scores ($R = 0.641$, $P = 0.014$); and (3) MPQ-A scores ($R = 0.594$, $P = 0.025$) when EEG data of the entire 19-min continuous EEG data were included. The correlation between alpha oscillation increases and pain perception decreases was also significant when only EEG data during pain perception rating were included ($R = 0.584$,

$P = 0.028$). The correlation was marginally significant when only EEG data after pain perception rating were included ($R = 0.524$, $P = 0.055$). Secondly, no significant correlation was observed when the correlation analysis was performed with the interaction terms ($P > 0.05$).

DISCUSSION

The present study described an active placebo effect on electrophysiological alpha oscillations during 20 min of tonic muscle pain. We observed placebo effects on the subjective intensity of pain perception to noxious stimulation. Placebo induced significant increases or a trend toward significant increases in the amplitude of alpha oscillation during tonic muscle pain in frontal-central regions when EEG data during pain perception ratings were not excluded. The decreases in the subjective intensity of pain perception to noxious stimulation after placebo treatment and the increases in the amplitude of alpha oscillation were significantly correlated. These findings suggest that placebo modulation in cognitive appraisal/experience of tonic muscle pain were effectively indexed by electrophysiological alpha oscillations, which served as additional evidence for the expectancy-based placebo mechanism (Wager et al., 2004; Zubieta et al., 2005; Scott et al., 2007; Atlas and Wager, 2012).
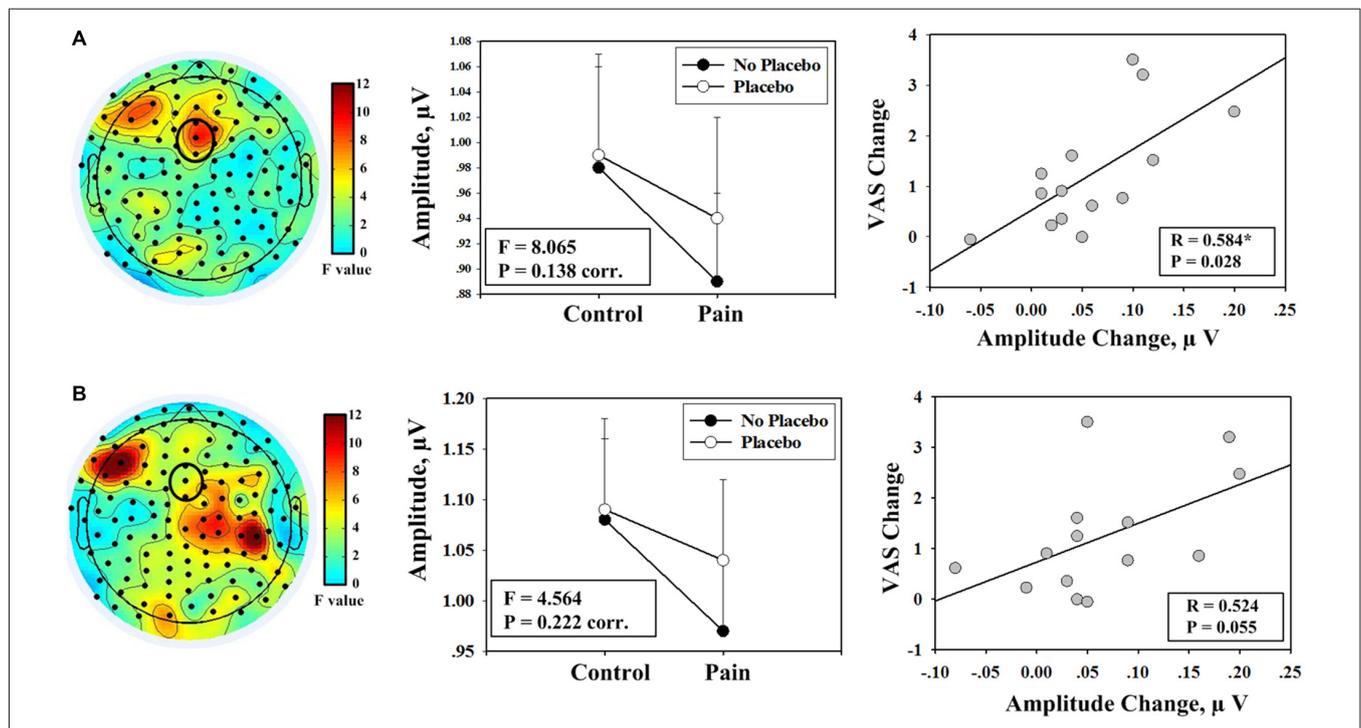
FIGURE 3 | Evidence showing the effect of placebo treatment from the partitions of EEG data during pain perception rating and after rating.
(A) A trend toward significant interaction effect was identified from frequency analyses including EEG data during the rating period (left panel); the mean values of the amplitudes of alpha oscillation (measured at FCz) from each condition are shown (F: *F* value of the interaction effect between the factors "pain" and "placebo"; corr.: corrected for multiple comparisons; middle panel); a significant correlation was observed between decrease in pain intensity during noxious stimulation after placebo treatment (II–IV) and the increase in the amplitude of alpha oscillation measured at FCz (IV–II; right panel). **(B)** No significant interaction effect was identified from frequency analyses that included EEG data of the time periods after rating (left panel); the corresponding results of mean amplitude values are shown (middle panel); marginally significant correlation was observed (right panel).

Numerous neuroimaging studies, including fMRI and PET studies of healthy subjects and clinical patients, revealed several cortical and subcortical regions that were mediated by placebo treatment (Meissner et al., 2011). The placebo analgesia also suppressed pain-induced responses in thalamus, insula, and ACC (Wager et al., 2004; Bingel et al., 2006; Kong et al., 2006; Price et al., 2007; Eippert et al., 2009). Assessments of the placebo effect to LEPs revealed a significant decrease in P2 amplitude, which was partially explained by the reduction in reported pain perception (Wager et al., 2006). The P2 in LEPs is highly likely generated from the ACC (Garcia-Larrea et al., 2003), and the decrease in P2 amplitude is consistent with the suppression of pain-induced responses in the ACC, which provides solid evidence that placebo analgesia is likely achieved via modulation of the emotional and cognitive components of pain (primarily coded by the ACC; Wiech et al., 2008; Tracey, 2010).

The placebo modulation effect that we observed supports the existence of a placebo effect on brain oscillation. The placebo treatment-induced changes in alpha oscillatory activities were maximal at frontal-central electrodes, which suggests the contribution of ACC to the generation of placebo-induced changes in alpha oscillations and confirms the modulation of placebo on the affective and cognitive components of pain that were observed to previous fMRI and PET studies (Wiech et al., 2008; Zubieta and Stohler, 2009; Tracey, 2010). Notably, the suppression of alpha amplitudes may reflect cortical activation or disinhibition of the corresponding neural networks (Pfurtscheller et al., 1996; Pfurtscheller and Lopes da Silva, 1999; Hu et al., 2013). For example, increased cellular excitability in thalamo-cortical systems was reflected by a decrease in alpha amplitude in EEG (Steriade and Llinás, 1988). Thus, the significant increase of alpha amplitude at frontal-central regions after placebo treatment may indicate an inhibition of cortical areas (including ACC) that are involved in pain processing (e.g., cognitive appraisal of tonic pain). However, we cannot make any firm conclusions about the contribution of ACC to the generation of placebo-induced changes in alpha oscillations without source analyses. We also cannot exclude the possible contribution of other neural sources (e.g., operculo-insular cortex) despite the performance of an EEG source analysis because of the limited spatial resolution of the EEG technique and the inverse problem in EEG source analysis (Michel et al., 2004). Hopefully, these issues may be effectively solved using the simultaneous EEG-fMRI technique, which was effectively used to extract fMRI activations that were significantly modulated by the alpha amplitude in EEG (Feige et al., 2005).

Only two published studies reported placebo treatment effects on brain alpha oscillatory activity. One study related alpha activity to placebo analgesia and reported a placebo-associated increase in alpha oscillations (Huneke et al., 2013). However, this study recorded alpha activity during resting states after placebo induction (Huneke et al., 2013). Another study reported that phasic pain-induced alpha responses were not sensitive to placebo manipulation using changes in stimulus intensity (Tiemann et al., 2015). This study did not include EEG data during pain perception (Tiemann et al., 2015). The placebo effect was derived from the cognitive and affective processing of pain perception, which may be more promising during the rating period. The alpha suppression in response to tonic pain primarily reflects high-level cognitive processing, and attention modulation may significantly affect it (Peng et al., 2014). Placebo treatment-related modulation effects of alpha oscillations may be better captured when subjects are asked to focus on their pain perception and report their pain intensity. Consequently, we observed significant modulation effects of placebo treatment and a positive correlation between placebo-induced pain decrease and increase in alpha amplitude when the pain perception evaluating period was included.

This study generated tonic muscle pain via an intramuscular infusion of hypertonic saline to produce a deep aching that was similar to the muscle pain experienced in clinical situations (Stohler and Kowalski, 1999). Our understanding of the neural mechanisms of pain were primarily based on the brain activation of phasic cutaneous pain, which involves fewer methodological challenges (e.g., easier to present several times to achieve a high signal-to-noise ratio of the brain responses) compared to tonic pain (Apkarian et al., 2011). However, chronic pain is normally prolonged and originates from deep tissue (e.g., muscle and viscera) in clinical practice (Apkarian et al., 2005; Schreckenberger et al., 2005). Therefore, the tonic muscle pain achieved by intramuscular infusion of hypertonic saline was used in the present study. The automated stimulus delivery system produced a prolonged, relatively stable muscle pain and achieved a better simulation of the pain experience in clinical settings, which may be important to establish the connection between placebo analgesic studies conducted in experimental settings (healthy subjects) and clinical practice (chronic pain patients).

There are several limitations to this study. First, this study consisted of fixed-order sessions (session 1: conditions I and II, session 2: conditions III and IV). Session 1 was always performed before session 2 because the individual infusion profiles used in condition IV should be identical to condition II. We cannot exclude the confounding factors of mental fatigue-induced alpha oscillation changes in this fixed-order and longer-lasting experiment. Experiments with prolonged stimulation are difficult to control as well as experiments using phasic stimulation. Mental fatigue and its influence on the measures of brain oscillation should be carefully considered. Spectral measures of brain oscillations were investigated to reflect changes in mental state in longer-lasting experiments. Several EEG measures were proposed to be valid and reliable indicators of mental fatigue, including a characterized shift of EEG power towards lower-frequency bands (delta, theta and alpha) and decrease in higher-frequency bands (Lal and Craig, 2002; Wascher et al., 2014). The amount of alpha suppression declined with time on task (Wascher et al., 2014). An increase in alpha power may reflect the increased effort and the difficulty of the subjects to maintain a state of alert wakefulness (Wascher et al., 2014). The significant correlation between the differences in pain perception and alpha amplitude was observed when the differences were calculated between two pain conditions II and IV. But no significant correlation could be observed when the differences were calculated according

to the interaction effect. Small sample size and fixed-order design might be some of those factors that contributed to this problem. Therefore, the correlation between the effect size of placebo analgesia in pain perception and alpha amplitude require further investigation using a randomized design and within-subject correlation analysis may offer more solid evidence. Second, the saline infusions in the control conditions and the pain conditions occurred on different sides. Therefore, we could only focus on the results of central electrodes in this study. The acquisitions of EEG data involve up to a few hundred electrodes positioned on the scalp, which together with volume conduction through the head results in a poor spatial resolution (Michel et al., 2004). The spreading effect from the lateral electrodes should be taken into consideration when interpreting the observed effects at the central electrodes. Third, the number of segments was different for different conditions when performing additional analyses with EEG data during and after ratings. This difference may be a confounding factor for comparisons of the amplitude spectrum among four conditions. Fourth, we only performed multiple comparisons correction for the number of electrodes (Schulz et al., 2015), but the correction for point-by-point analysis should account for the number of electrodes and the number of frequency bands

(Peng et al., 2014). Previous studies reported an association of placebo and nocebo effects with opposite responses of dopamine and endogenous opioid neurotransmission in a distributed network of cortical and subcortical regions (Scott et al., 2008), and possible electrophysiological responses that are oppositely involved in placebo and nocebo effects should be assessed in the future.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Apkarian, A. V., Bushnell, M. C., Treede, R. D., and Zubieta, J. K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *Eur. J. Pain* 9, 463–484. doi: 10.1016/j.ejpain.2004.11.001

Apkarian, A. V., Hashmi, J. A., and Baliki, M. N. (2011). Pain and the brain: specificity and plasticity of the brain in clinical chronic pain. *Pain* 152, S49–S64. doi: 10.1016/j.pain.2010.11.010

Aslaksen, P. M., Myrbakk, I. N., Høifødt, R. S., and Flaten, M. A. (2007). The effect of experimenter gender on autonomic and subjective responses to pain stimuli. *Pain* 129, 260–268. doi: 10.1016/j.pain.2006.10.011

Atlas, L. Y., and Wager, T. D. (2012). How expectations shape pain. *Neurosci. Lett.* 520, 140–148. doi: 10.1016/j.neulet.2012.03.039

Atlas, L. Y., Wager, T. D., Dahl, K. P., and Smith, E. E. (2009). "Placebo effects," in *Handbook of Neuroscience for the Behavioral Sciences*, eds G. Gary and J. T. C. Berntson (Hoboken, NJ: John Wiley and Sons, Inc.), 63–86.

Benedetti, F. (2009). *Placebo Effects: Understanding the Mechanisms in Health and Disease.* New York, NY: Oxford University Press.

Bingel, U., Lorenz, J., Schoell, E., Weiller, C., and Büchel, C. (2006). Mechanisms of placebo analgesia: rACC recruitment of a subcortical antinociceptive network. *Pain* 120, 8–15. doi: 10.1016/j.pain.2005.08.027

Bromm, B., and Treede, R. (1991). Laser-evoked cerebral potentials in the assessment of cutaneous pain sensitivity in normal subjects and patients. *Rev. Neurol. (Paris)* 147, 625–643.

Chang, P. F., Arendt-Nielsen, L., and Chen, A. C. (2002). Dynamic changes and spatial correlation of EEG activities during cold pressor test in man. *Brain Res. Bull.* 57, 667–675. doi: 10.1016/s0361-9230(01)00763-8

Chang, P. F., Arendt-Nielsen, L., Graven-Nielsen, T., and Chen, A. C. (2003). Psychophysical and EEG responses to repeated experimental muscle pain in humans: pain intensity encodes EEG activity. *Brain Res. Bull.* 59, 533–543. doi: 10.1016/s0361-9230(02)00950-4

Chang, P. F., Arendt-Nielsen, L., Graven-Nielsen, T., Svensson, P., and Chen, A. C. (2001). Different EEG topographic effects of painful and non-painful intramuscular stimulation in man. *Exp. Brain Res.* 141, 195–203. doi: 10.1007/s002210100864

Chang, P. F., Arendt-Nielsen, L., Graven-Nielsen, T., Svensson, P., and Chen, A. C. (2004). Comparative EEG activation to skin pain and muscle pain induced by

capsaicin injection. *Int. J. Psychophysiol.* 51, 117–126. doi: 10.1016/j.ijpsycho.2003.01.001

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Dowman, R., Rissacher, D., and Schuckers, S. (2008). EEG indices of tonic pain-related activity in the somatosensory cortices. *Clin. Neurophysiol.* 119, 1201–1212. doi: 10.1016/j.clinph.2008.01.019

Eippert, F., Bingel, U., Schoell, E. D., Yacubian, J., Klinger, R., Lorenz, J., et al. (2009). Activation of the opioidergic descending pain control system underlies placebo analgesia. *Neuron* 63, 533–543. doi: 10.1016/j.neuron.2009.07.014

Feige, B., Scheffler, K., Esposito, F., Di Salle, F., Hennig, J., and Seifritz, E. (2005). Cortical and subcortical correlates of electroencephalographic alpha rhythm modulation. *J. Neurophysiol.* 93, 2864–2872. doi: 10.1152/jn.00721.2004

Garcia-Larrea, L., Frot, M., and Valeriani, M. (2003). Brain generators of laser-evoked potentials: from dipoles to functional significance. *Neurophysiol. Clin.* 33, 279–292. doi: 10.1016/j.neucli.2003.10.008

Hu, L., Peng, W., Valentini, E., Zhang, Z., and Hu, Y. (2013). Functional features of nociceptive-induced suppression of alpha band electroencephalographic oscillations. *J. Pain* 14, 89–99. doi: 10.1016/j.jpain.2012.10.008

Huang, L., Yang, T., and Ji, Z. (2003). Applicability of the positive and negative affect scale in Chinese. *Chin. J. Ment. Health* 17, 54–56.

Huber, M. T., Bartling, J., Pachur, D., Woikowsky-Biedau, S., and Lautenbacher, S. (2006). EEG responses to tonic heat pain. *Exp. Brain Res.* 173, 14–24. doi: 10.1007/s00221-006-0366-1

Huneke, N. T., Brown, C. A., Burford, E., Watson, A., Trujillo-Barreto, N. J., El-Deredy, W., et al. (2013). Experimental placebo analgesia changes resting-state alpha oscillations. *PLoS One* 8:e78278. doi: 10.1371/journal.pone.0078278

Iannetti, G., Truini, A., Galeotti, F., Romaniello, A., Manfredi, M., and Cruccu, G. (2001). Usefulness of dorsal laser evoked potentials in patients with spinal cord damage: report of two cases. *J. Neurol. Neurosurg. Psychiatry* 71, 792–794. doi: 10.1136/jnnp.71.6.792

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* 14, 166–185. doi: 10.1002/hbm.1050

Kong, J., Gollub, R. L., Rosman, I. S., Webb, J. M., Vangel, M. G., Kirsch, I., et al. (2006). Brain activity associated with expectancy-enhanced placebo analgesia as

measured by functional magnetic resonance imaging. *J. Neurosci.* 26, 381–388. doi: 10.1523/jneurosci.3556-05.2006

Lal, S. K., and Craig, A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology* 39, 313–321. doi: 10.1017/s0048577201393095

Le Pera, D., Svensson, P., Valeriani, M., Watanabe, I., Arendt-Nielsen, L., and Chen, A. C. (2000). Long-lasting effect evoked by tonic muscle pain on parietal EEG activity in humans. *Clin. Neurophysiol.* 111, 2130–2137. doi: 10.1016/s1388-2457(00)00474-0

Li, J., Feng, Y., Han, J., Fan, B., Wang, J., Wei, X., et al. (2013). Multi-centered linguistic adaptation and validation of short-form McGill pain questionnaire-2. *Chin. J. Pain Med.* 19, 42–46.

Lorenz, J., Hauck, M., Paur, R. C., Nakamura, Y., Zimmermann, R., Bromm, B., et al. (2005). Cortical correlates of false expectations during pain intensity judgments—a possible manifestation of placebo/nocebo cognitions. *Brain Behav. Immun.* 19, 283–295. doi: 10.1016/j.bbi.2005.03.010

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. U S A* 94, 10979–10984. doi: 10.1073/pnas.94.20.10979

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Meissner, K., Bingel, U., Colloca, L., Wager, T. D., Watson, A., and Flaten, M. A. (2011). The placebo effect: advances from different methodological approaches. *J. Neurosci.* 31, 16117–16124. doi: 10.1523/JNEUROSCI.4099-11.2011

Melzack, R. (1987). The short-form McGill pain questionnaire. *Pain* 30, 191–197. doi: 10.1016/0304-3959(87)91074-8

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., and Grave de Peralta, R. (2004). EEG source imaging. *Clin. Neurophysiol.* 115, 2195–2222. doi: 10.1016/j.clinph.2004.06.001

Nir, R.-R., Sinai, A., Raz, E., Sprecher, E., and Yarnitsky, D. (2010). Pain assessment by continuous EEG: association between subjective perception of tonic pain and peak frequency of alpha oscillations during stimulation and at rest. *Brain Res.* 1344, 77–86. doi: 10.1016/j.brainres.2010.05.004

Peng, W., Hu, L., Zhang, Z., and Hu, Y. (2014). Changes of spontaneous oscillatory activity to tonic heat pain. *PLoS One* 9:e91052. doi: 10.1371/journal.pone.0091052

Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Pfurtscheller, G., Stancak, A., Jr., and Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: a review. *Int. J. Psychophysiol.* 24, 39–46. doi: 10.1016/s0167-8760(96)00066-9

Price, D. D., Craggs, J., Verne, G. N., Perlstein, W. M., and Robinson, M. E. (2007). Placebo analgesia is accompanied by large reductions in pain-related brain activity in irritable bowel syndrome patients. *Pain* 127, 63–72. doi: 10.1016/j.pain.2006.08.001

Schreckenberger, M., Siessmeier, T., Viertmann, A., Landvogt, C., Buchholz, H. G., Rolke, R., et al. (2005). The unpleasantness of tonic pain is encoded by the insular cortex. *Neurology* 64, 1175–1183. doi: 10.1212/01.wnl.0000156353.17305.52

Schulz, E., May, E. S., Postorino, M., Tiemann, L., Nickel, M. M., Witkovsky, V., et al. (2015). Prefrontal gamma oscillations encode tonic pain in humans. *Cereb. Cortex* 25:bhv043. doi: 10.1093/cercor/bhv043

Scott, D. J., Stohler, C. S., Egnatuk, C. M., Wang, H., Koeppe, R. A., and Zubieta, J. K. (2007). Individual differences in reward responding explain placebo-induced expectations and effects. *Neuron* 55, 325–336. doi: 10.1016/j.neuron.2007.06.028

Scott, D. J., Stohler, C. S., Egnatuk, C. M., Wang, H., Koeppe, R. A., and Zubieta, J. K. (2008). Placebo and nocebo effects are defined by opposite

opioid and dopaminergic responses. *Arch. Gen. Psychiatry* 65, 220–231. doi: 10.1001/archgenpsychiatry.2007.34

Steriade, M., and Llinás, R. R. (1988). The functional states of the thalamus and the associated neuronal interplay. *Physiol. Rev.* 68, 649–742.

Stohler, C. S. (1992). An experimental model of jaw muscle pain in man. *Biol. Mech. Tooth Mov. Craniofac. Adapt.* 503–511.

Stohler, C. S., and Kowalski, C. J. (1999). Spatial and temporal summation of sensory and affective dimensions of deep somatic pain. *Pain* 79, 165–173. doi: 10.1016/s0304-3959(98)00171-7

Tiemann, L., May, E. S., Postorino, M., Schulz, E., Nickel, M. M., Bingel, U., et al. (2015). Differential neurophysiological correlates of bottom-up and top-down modulations of pain. *Pain* 156, 289–296. doi: 10.1097/01.j.pain.0000460309.94442.44

Tracey, I. (2010). Getting the pain you expect: mechanisms of placebo, nocebo and reappraisal effects in humans. *Nat. Med.* 16, 1277–1283. doi: 10.1038/nm.2229

Veerasarn, P., and Stohler, C. S. (1992). The effect of experimental muscle pain on the background electrical brain activity. *Pain* 49, 349–360. doi: 10.1016/0304-3959(92)90242-4

Wager, T. D., and Fields, H. (2011). "Placebo analgesia," in *Wall and Melzack's Textbook of Pain*, eds S. B. McMahon and M. Koltzenburg (Oxford, England: Churchill Livingstone), 362–374.

Wager, T. D., Matre, D., and Casey, K. L. (2006). Placebo effects in laser-evoked pain potentials. *Brain Behav. Immun.* 20, 219–230. doi: 10.1016/j.bbi.2006.01.007

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., et al. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science* 303, 1162–1167. doi: 10.1126/science.1093065

Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol.* 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Watson, A., El-Deredy, W., Vogt, B. A., and Jones, A. K. P. (2007). Placebo analgesia is not due to compliance or habituation: EEG and behavioural evidence. *Neuroreport* 18, 771–775. doi: 10.1097/wnr.0b013e3280c1e2a8

Wiech, K., Ploner, M., and Tracey, I. (2008). Neurocognitive aspects of pain perception. *Trends Cogn. Sci.* 12, 306–313. doi: 10.1016/j.tics.2008.05.005

Zhang, X., Ashton-Miller, J. A., and Stohler, C. S. (1993). A closed-loop system for maintaining constant experimental muscle pain in man. *IEEE Trans. Biomed. Eng.* 40, 344–352. doi: 10.1109/10.222327

Zubieta, J. K., Bueller, J. A., Jackson, L. R., Scott, D. J., Xu, Y., Koeppe, R. A., et al. (2005). Placebo effects mediated by endogenous opioid activity on μ-opioid receptors. *J. Neurosci.* 25, 7754–7762. doi: 10.1523/jneurosci.0439-05.2005

Zubieta, J. K., and Stohler, C. S. (2009). Neurobiological mechanisms of placebo responses. *Ann. N Y Acad. Sci.* 1156, 198–210. doi: 10.1111/j.1749-6632.2009.04424.x

Pain Related Cortical Oscillations: Methodological Advances and Potential Applications

Weiwei Peng[1]* and Dandan Tang[2]

[1] Key Laboratory of Cognition and Personality (Ministry of Education), Faculty of Psychology, Southwest University, Chongqing, China, [2] School of Education Science, Zunyi Normal College, Guizhou, China

Alongside the time-locked event-related potentials (ERPs), nociceptive somatosensory inputs can induce modulations of ongoing oscillations, appeared as event-related synchronization or desynchronization (ERS/ERD) in different frequency bands. These ERD/ERS activities are suggested to reflect various aspects of pain perception, including the representation, encoding, assessment, and integration of the nociceptive sensory inputs, as well as behavioral responses to pain, even the precise details of their roles remain unclear. Previous studies investigating the functional relevance of ERD/ERS activities in pain perception were normally done by assessing their latencies, frequencies, magnitudes, and scalp distributions, which would be then correlated with subjective pain perception or stimulus intensity. Nevertheless, these temporal, spectral, and spatial profiles of stimulus induced ERD/ERS could only partly reveal the dynamics of brain oscillatory activities. Indeed, additional parameters, including but not limited to, phase, neural generator, and cross frequency couplings, should be paid attention to comprehensively and systemically evaluate the dynamics of oscillatory activities associated with pain perception and behavior. This would be crucial in exploring the psychophysiological mechanisms of neural oscillation, and in understanding the neural functions of cortical oscillations involved in pain perception and behavior. Notably, some chronic pain (e.g., neurogenic pain and complex regional pain syndrome) patients are often associated with the occurrence of abnormal synchronized oscillatory brain activities, and selectively modulating cortical oscillatory activities has been showed to be a potential therapy strategy to relieve pain with the application of neurostimulation techniques, e.g., repeated transcranial magnetic stimulation (rTMS) and transcranial alternating current stimulation (tACS). Thus, the investigation of the oscillatory activities proceeding from phenomenology to function, opens new perspectives to address questions in human pain psychophysiology and pathophysiology, thereby promoting the establishment of rational therapeutic strategy.

Keywords: pain, cortical oscillations, event-related desynchronization (ERD), event-related synchronization (ERS), electroencephalography (EEG)

INTRODUCTION

Pain, affecting the well beings of millions of individuals and imposing a severe financial burden upon our societies, is a major public healthcare problem. Pain relief, especially for the patients with pathological chronic pain, still remains a very problematic challenge to the physicians. The progress in understanding of the neural representation of pain in humans is not only important for basic neuroscience research, but also critical to develop effective strategies for the diagnosis and management of the pathological pain conditions. Specifically, this constitutes the understandings of: (1) the physiological mechanisms of the nociceptive system in healthy populations, particularly the cortical processes underlying the perception of pain and (2) the pathophysiological mechanisms of the nociceptive system in chronic pain patients, particularly the peripheral and central mechanisms leading to chronic pain. Thus, for a better understanding of the physiology and pathophysiology of pain in humans, novel approaches should be developed to identify the neural activities related to the processing of noxious inputs in humans, as well as characterize their functional roles in subjective pain perception.

In both physiological (Iannetti et al., 2003) and pathophysiological (Treede, 2003; Treede et al., 2003) studies, laser-evoked potentials (LEPs) have been extensively used to investigate the peripheral and central processing of nociceptive somatosensory inputs, and are currently considered as the best available diagnostic tool to assess the function of nociceptive pathways in patients (Cruccu et al., 2010). The radiant heat pulses that selectively excite nociceptive nerve endings in the epidermis (Bromm et al., 1984), can elicit a number of electrical brain responses, some of which can be detected with the electroencephalography (EEG) recording techniques (Carmon et al., 1976; Mouraux et al., 2003). Note that the EEG response is time-locked if it manifests the same pattern at roughly the same time on each trial after the stimulus onset, and the EEG response is phase-locked if it takes the same phase angle on each trial after the stimulus onset (Mouraux and Iannetti, 2008). The time-locked and phase-locked LEPs could be commonly obtained by an across-trial averaging procedure. Several deflections have been identified in LEPs (**Figure 1**), including: (1) an early component of a small negative deflection (N1, peaking at approximately 160 ms when stimulating the hand dorsum), with maximal distribution over the central temporal region contralateral to the stimulated side (Valentini et al., 2012); (2) the largest deflection of a negative-positive vertex potential (N2-P2 complex, peaking at approximately 160 and 390 ms when stimulating the hand dorsum), with maximal scalp distribution over the central region (Iannetti et al., 2008); and (3) a late component of a positive deflection (P4, approximately 390 ms when stimulating the hand dorsum), with maximal scalp distribution over the central–parietal region contralateral to the stimulated side (Hu et al., 2014a). As revealed by dipole modelings of scalp, subdural recordings, and direct intracranial recordings (Tarkka and Treede, 1993; Bromm and Chen, 1995; Lenz et al., 2000; Garcia-Larrea et al., 2003; Valentini et al., 2012), LEPs were

showed to be generated from a combination of cortical and subcortical structures, including the primary and secondary somatosensory cortex (S1 and S2), insula, and anterior/mid-cingulate cortex (ACC/MCC), as well as parietal operculum. Functionally, recent evidences (Iannetti et al., 2008; Mouraux and Iannetti, 2009) showed that these laser-evoked EEG responses represent an indirect readout of the function of nociceptive system, mainly determined by the saliency of the eliciting nociceptive stimulus, i.e., the ability to capture attention, instead of the specific neural processes underlying pain perception.

Alongside the ERPs, sensory stimuli could also induce transient modulations of the ongoing oscillatory activities in different frequency bands (Pfurtscheller and Lopes da Silva, 1999). Since these oscillatory activities are normally time-locked but not phase-locked to the onset of the stimulus, they would be eliminated by the classical across-trial averaging procedures that are typically used to reveal ERPs (Mouraux and Iannetti, 2008). Alternative signal processing techniques, based on the joint time-frequency decompositions of signals, are often adopted to explore the neurophysiological mechanisms of brain oscillations. These modulations are characterized by either transient enhancement (event-related synchronization, ERS) or transient suppression (event-related desynchronization, ERD) of the oscillation power, usually confined to a specific frequency band (Pfurtscheller and Lopes da Silva, 1999). The functional significance of ERS and ERD differs according to the frequency band within which they occur. For example, ERD in the alpha band (frequencies ranging from 8–13 Hz) has been hypothesized to reflect cortical activation or disinhibition (Pfurtscheller and Lopes da Silva, 1999; Schnitzler et al., 2000; Hu et al., 2013), while ERS in the gamma band (frequencies ranging from 30–100 Hz) has been hypothesized to play a crucial role in cortical integration and perception (Tallon-Baudry and Bertrand, 1999; Gross et al., 2007; Fries, 2009; Hipp et al., 2011).

By performing time–frequency analysis on the EEG signals elicited by nociceptive somatosensory stimuli, several electrophysiological responses (ERPs) related to the activation of nociceptive fibers have been disclosed (**Figure 1**), including: (1) a suppression of the alpha oscillations, i.e., α-ERD, globally across somatosensory, motor, and visual areas, reflecting a widespread change of cortical function and excitability, and relating to the special alerting function of pain (Mouraux et al., 2003; Ploner et al., 2006b; Hu et al., 2013); (2) a suppression of beta oscillations (∼20 Hz in frequency), i.e., β-ERD, predominantly over the contralateral primary motor cortex without an obvious beta oscillation rebound followed (Raij et al., 2004), indicating the prolonged excitations of neurons within motor cortex, which may be associated with the facilitation of the voluntary movements to prevent tissue damage in pain processing; and (3) enhancement of gamma oscillations, i.e., γ-ERS, over contralateral somatosensory cortex, particularly relating to subjective pain intensity (Gross et al., 2007; Zhang et al., 2012; Hu et al., 2014b), and reflecting the internal representations of behaviorally relevant stimuli that should receive enhanced/preferred processing.
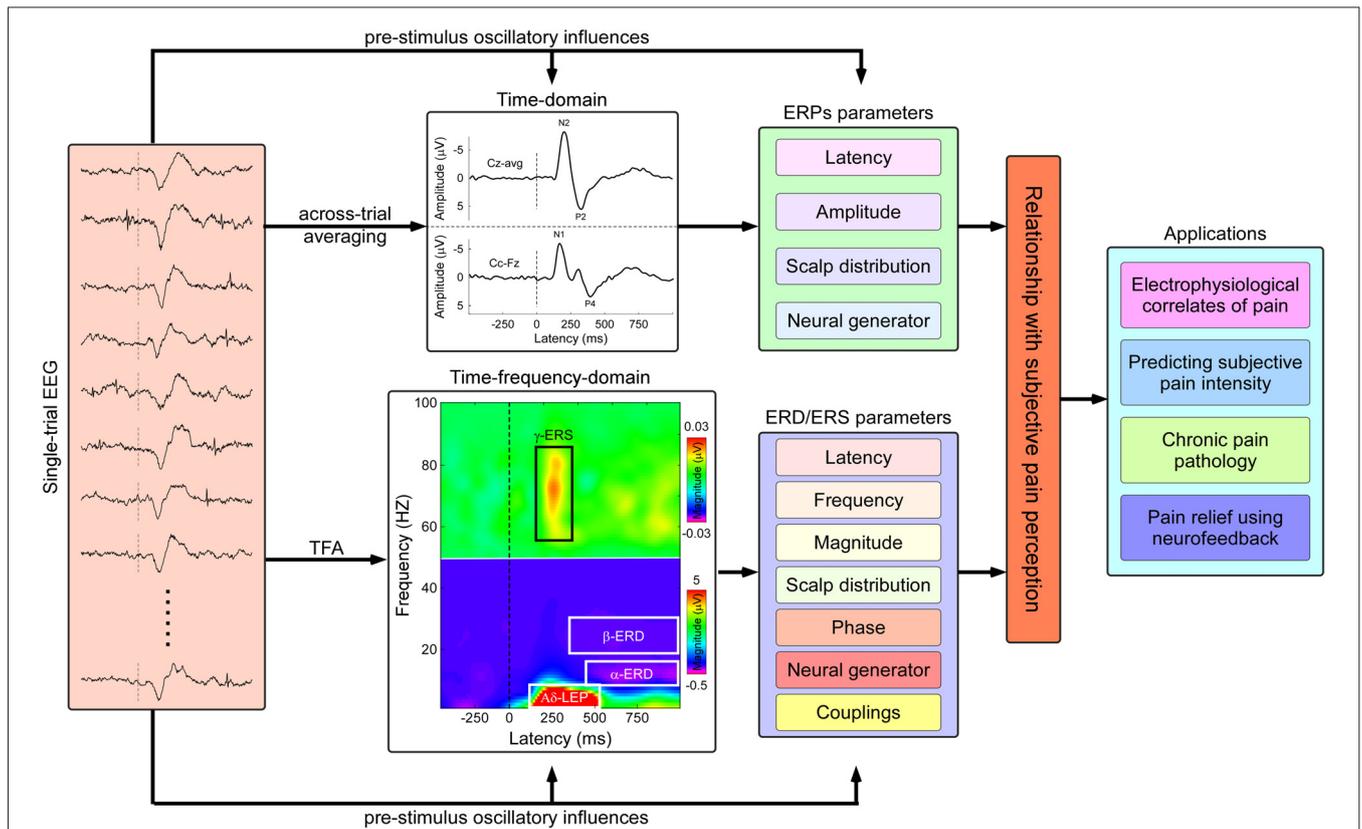
FIGURE 1 | Flowchart describing the identification of pain related electrophysiological features and applications in both basic and clinical pain study.
Nociceptive somatosensory inputs can elicit transient changes in the ongoing electroencephalography (EEG) activities, including phase-locked event-related potentials (ERPs) and non-phase-locked modulations of ongoing oscillatory activities in multiple frequency bands (appeared as event-related desynchronization or synchronization, ERD or ERS). The phase-locked ERPs activities could be obtained by the classical across-trial averaging process, characterized by their peak latency, amplitude, scalp topography, and neural generators, while the non-phase-locked ERD/ERS activities could be identified using time-frequency analysis (TFA), characterized by several parameters including latency, frequency, magnitude, scalp topography, phase, neural generator, casual information flows, and cross-frequency couplings (CFCs). The assessment of the relationship between human pain perception and electrophysiological responses has the potential applications in both basic and clinical pain study, including: (1) exploring electrophysiological signatures coding subjective pain perception; (2) predicting subjective pain intensity; (3) exploring pathological mechanisms of chronic pain; and (4) relieving pain modulating the cortical oscillatory activities using neurofeedback techniques.

These painful stimulus induced ERD/ERS responses, occurring in painful information processing, have been suggested to be associated with the perception of pain (Babiloni et al., 2006; Gross et al., 2007; Zhang et al., 2012) and with endogenous or exogenous attention to the painful stimuli (Mouraux et al., 2003; Hauck et al., 2007; Hu et al., 2013). However, it is still not clear whether these somatic sensory pain-related oscillatory activities are pain-specific opposed to non-painful somatosensory stimuli, or the salience of the stimuli presentation (Iannetti et al., 2008; Mouraux and Iannetti, 2009). Even though, these stimulus induced ERD/ERS activities could indeed provide plentiful information related to brain processing, which is different from those cortical activities reflected by stimulus-evoked ERPs (Mouraux and Iannetti, 2008). Previous studies have indicated that nociceptive somatosensory stimuli induced ERD/ERS activities in multiple frequency bands could reflect various aspects of pain perception (e.g., representation, encoding, assessment, and integration of the nociceptive sensory stimuli, as

well as the behavioral responses to pain), even the precise details of their roles remain unclear. Indeed, investigating the cortical oscillatory activities involved in human pain perception and establishing the oscillatory basis of pain opened a new window to study the cortical process underlying pain perception. Thus, in this article, we will: (1) highlight several methodological recommendations on investigating brain oscillations related to pain and (2) summarize the potential applications in both basic and clinical pain study.

METHODOLOGICAL RECOMMENDATIONS TO EXTRACT PAIN RELATED BRAIN OSCILLATORY ACTIVITIES

The transient modulations of cortical oscillatory activities induced by the nociceptive somatosensory stimuli are normally characterized by their peak frequency, latency, magnitude,

and topography distribution, relative to the baseline period (using subtraction or percentage approach). Nevertheless, the traditionally temporal, spectral, and spatial profiles can only partly reveal the dynamics of brain oscillatory activities. Investigating novel parameters comprehensively characterizing brain oscillations could help explore the psychophysiological mechanisms of neural oscillations, as well as the neural functions of cortical oscillations involved in sensory perception and behavior. In addition, pre-stimulus ongoing EEG oscillation could influence both post-stimulus electrophysiological activities and sensory perception (Thut et al., 2006; Romei et al., 2008; Fellinger et al., 2011; Lange et al., 2012; Tu et al., 2016), suggesting the importance of dissecting the contributions of pre- and post-stimulus oscillation to the variabilities of painful stimulus induced ERD/ERS activities. Based on these understandings, from the methodological aspect, we encourage the researchers of pain field to: (1) utilize novel parameters to comprehensively characterize pain related oscillations and (2) dissect the contributions of pre- and post-stimulus oscillations, when they are investigating the dynamics of brain oscillatory activities associated with pain perception and behavior.

Utilization of Novel Parameters to Comprehensively Characterize Pain Related Oscillations

Apart from the appearing frequency, latency, magnitude, and scalp topography, several other parameters, including but not limited to, the phase, neural generator, and cross-frequency coupling (CFC) of pain related oscillations, should also be further investigated, for a comprehensive and systemic understanding of the brain oscillations associated with pain.

Phase

Much of the research on oscillations in human EEG has focused on the dynamics of oscillations magnitudes. Nevertheless, the phase of the oscillatory activities at a given frequency band reflects cyclic fluctuations of a network's excitability and varies on a much faster timescale than the sluggish amplitude fluctuations at the same frequency band (Buzsáki and Draguhn, 2004; Lakatos et al., 2005; Rajkai et al., 2008), the phase of the oscillations may provide deep insights into the fine-grained neural mechanisms underlying sensory perception (Buzsáki and Draguhn, 2004; Busch et al., 2009). Indeed, it is suggested that phase synchronization between alpha oscillations in different brain areas allows for an effective network communication and information transmission regulation (von Stein and Sarnthein, 2000; Palva and Palva, 2011; Saalmann et al., 2012).

A growing body of studies on EEG oscillations have shown that the phase of ongoing theta and alpha frequency oscillations prior to the onset of stimuli could influence both the subsequent ERPs (e.g., Haig and Gordon, 1998; Kruglikov and Schiff, 2003; Gruber et al., 2005; Fellinger et al., 2011) and sensory stimulus perception (Busch et al., 2009; Mathewson et al., 2009). As shown in the target auditory oddball data, the amplitude of ERPs (e.g., N100 amplitude) as well as the reaction times (RTs) were both significantly modulated by the phase synchronization of

the alpha oscillation that was evaluated by the angular variance of the oscillation (Haig and Gordon, 1998). Using identical visual stimuli at the individual detection threshold (Busch et al., 2009), the phase of ongoing oscillation (in theta and alpha frequency bands) accounted for about 16% of variabilities of visual detection performance (hits or misses) and allowed the prediction of sensory performance on the single-trial level. In other words, the phase of ongoing oscillations reflects the cortical processing of threshold visual stimuli, thus providing a direct link between phase of oscillations and sensory perception and behavior.

These evidences of a relationship between spontaneous oscillation phase and the amplitude of subsequent ERPs, manual responses, and sensory perception, are in line with the cellular level concept that the neuronal oscillations reflect the cyclic variations of neuronal excitability (Buzsáki and Draguhn, 2004; Rajkai et al., 2008). Even the dynamics of phase information in cortical oscillatory activities have been shown to be functionally relevant in stimulus processing and perception of auditory, visual, and even somatosensory modalities, the modulations of pain elicited ERPs as well as pain perception and behavior by the phase of the oscillatory activities, still remain unclear. It therefore needs further investigation, which could broaden the understanding regarding how the ongoing oscillations shape our sensory painful perception.

Neural Generators

The spatial characteristics of stimulus induced ERD/ERS activities could be based on their scalp topographies, but the effects of active references in EEG recordings could not be denied. Whether the reference problems in assessing ERD/ERS oscillatory activities could be reduced by approximately standardizing the reference of scalp EEG recordings to a point at infinity, which was ever proposed in assessing evoked potentials by Yao (2001), should be further investigated. Nevertheless, the fact that the equivalent sources of evoked potentials and oscillatory activities are actually independent from the choice of a particular reference, suggests the importance of identifying neural generators of stimulus-induced ERD/ERS activities. With accumulating evidence showing the functions of the oscillatory brain activities in various aspects of pain perception (Mouraux et al., 2003; Ploner et al., 2006a,b; Gross et al., 2007; Zhang et al., 2012; Hu et al., 2013), identifying sources of oscillatory activities is an essential step to directly determine the relation of EEG oscillations to brain function and sensory process, thus revealing how the different cortical areas function as a network involved in human pain perception. For example, alpha oscillations close to the occipito-parietal midline is closely linked to coherent objects (Vanni et al., 1997), suggesting that the function of oscillatory activity in occipitoparietal visual areas in modulating visual shape processing. However, until now, identifying the sources of oscillations in human brain is still a challenging problem due to the low spatial resolution of EEG/MEG recording techniques.

Source localization techniques have been proposed to identify the responsible neural generators (Pascual-Marqui et al., 1994; Cheyne et al., 2003; Hoechstetter et al., 2004; Jurkiewicz et al.,

2006; Doesburg et al., 2009), e.g., dipole and distributed source modelings, as well as beamformer technique, and have been adopted in localizing the neural generators of pain related oscillations (Raij et al., 2004; Ploner et al., 2006a,b; Gross et al., 2007; Peng et al., 2012). Gross et al. (2007) computed the painful stimuli induced high-frequency oscillations in the electrical activity of the human S1 using a linearly constrained minimum variance spatial filtering approach, and then the relationships between stimulus induced gamma ERS and objective stimulus intensity as well as subjective pain intensity were established on the source level, making it possible to evaluate the functional relevance of gamma oscillations in pain perception more directly. However, these source localization models are typically ill-posed inverse problems since infinite number of sources could explain a given scalp topography and additional information as constraints is needed to obtain a unique solution. For example, the beamformer source localization technique, which uses an adaptive spatial filter to estimate the activity everywhere in the brain (Gaetz and Cheyne, 2003; Cheyne et al., 2003), is based on minimizing the source power (or variance) at a given location, and assumes that sources in different parts of the brain are not temporally correlated, which does not make sense physiologically sometimes.

Alternative approaches based on the simultaneous recordings of functional magnetic resonance imaging (fMRI) and EEG (Laufs et al., 2003a; Lei et al., 2011; Dong et al., 2014) have also been proposed to explore the neural sources of EEG oscillations by identifying fMRI blood oxygenation level-dependent (BOLD) signal changes related to spontaneous EEG power fluctuations. Even it combines the high spatial resolution in fMRI and high temporal resolution in EEG, such a method of correlating continuously band-specific EEG power with fMRI-BOLD signal changes, is actually an indirect way to identify source of oscillations. Indeed, monitoring the large-scale neuronal firing patterns and the generated local field potentials (LFPs) in animal models (e.g., behaving rodents) serves a direct and effective way to investigate the generators of these various oscillations as well as their spatial and temporal relationships.

CFC

As a statistical relationship between oscillatory activities in two different frequency bands, CFCs (may be appeared as phase-to-phase, phase-to-power, or power-to-power couplings) have been proposed to reflect the coordination of neural dynamics across temporal and spatial scales (Canolty and Knight, 2010; Canolty et al., 2006), and have been observed in many species and brain regions. As revealed by the LFPs on monkeys, the phase of low-frequency oscillations was shown to modulate the amplitude of gamma oscillations (Wang et al., 2012), and such CFC was suggest to integrate long-range neural interactions mediated by low-frequency rhythms (e.g., theta/alpha) with local computations mediated by high frequencies (i.e., gamma). Importantly, the abnormal CFC is linked to several cognitive processes and disease states (Schlee et al., 2009; López-Azcárate et al., 2010; Miskovic et al., 2011; de Hemptinne et al., 2013). Couplings between β-phase (13–30 Hz) and γ-amplitude (50–200 Hz) in primary motor cortex showed to be exaggerated

for Parkinson patients compared with healthy subjects without motor disorders, and such excessive coupling could be reduced by therapeutic subthalamic nucleus stimulation (de Hemptinne et al., 2013), suggesting the dysfunction of CFC in disease states.

With the evidences showing: (1) the potential relevance of CFC for understanding psychophysiological and pathological brain functions (Canolty et al., 2006; Schlee et al., 2009; Canolty and Knight, 2010; López-Azcárate et al., 2010; Miskovic et al., 2011; de Hemptinne et al., 2013) and (2) nociceptive somatosensory stimuli induced modulations of oscillations in multiple frequency bands (Schulz et al., 2011; Zhang et al., 2012; Hu et al., 2015), we believe that the oscillatory activities in different frequency bands are functioning interactively within the cortical network, and CFCs involved in pain could provide complemented information for the establishment of the cortical oscillatory bases of pain perception. However, it should be noted that the couplings measured anywhere in the brain can be potentially explained by the influence of external sensory inputs or internal cognitive events, on the phase and amplitude of the oscillations, rather than reflecting the actual modulations in different frequency bands. For example, the coupling of theta phase and gamma power observed in rodents (Wang et al., 2011), which was interpreted as a reflection of the storage and processing of nociceptive information, actually can be explained by the common effects of the nociceptive sensory inputs on both theta phase and gamma power, instead of the actual CFC. Therefore, whether the observed correlation between two bands (e.g., phase-amplitude) is due to the common drive, e.g., generated by external or internal input, or whether the correlation is due to a causal interaction between rhythms should be distinguished in the future study.

Dissection of Pre- and Post-Stimulus Oscillations

The traditional approach to estimate ERD/ERS activities relies on time-frequency decomposition methods to transform the single-trial electrocortical signals into time-frequency distributions (TFDs), and then the resulting TFDs are typically expressed as a percentage change relative to pre-stimulus EEG power to highlight the stimulus-induced changes in power within specific frequency bands (Ploner et al., 2006b; Iannetti et al., 2008; Hu et al., 2013). However, a recent study (Hu et al., 2014b) demonstrated that such baseline percentage approach would introduce a significant bias in estimating ERD/ERS magnitudes, i.e., resulting in an overestimation of ERS and underestimation of ERD, and pointed out that such bias could be avoided using a single-trial baseline subtraction approach.

Importantly, the pre-stimulus oscillatory activities in different frequency bands, reflecting the dynamics of brain states, can influence both the post-stimulus ERPs and sensory perception. For example, the pre-stimulus α-power could significantly modulate the nociceptive-induced α-ERD magnitude (Hu et al., 2013), by showing the nociceptive-induced α-ERD magnitude was significantly more dependent on the

pre-stimulus than on the post-stimulus α-power. A more recent study (Tu et al., 2016) showed that the pre-stimulus EEG oscillations in both alpha and gamma frequency bands could significantly modulate the subjective perception of painful stimuli, and importantly, the pre-stimulus alpha and gamma oscillatory activities could provide distinctive information in predicting subjective pain perception. Nevertheless, the single-trial baseline correction approaches (both percentage and subtraction methods) would confuse the contribution of pre- and post-stimulus EEG power, since the baseline corrected ERD/ERS activities reflect the mixed variabilities of changes in the state of the system (reflected as the pre-stimulus oscillations in different frequency bands; Laufs et al., 2003b; Del Percio et al., 2006; Hu et al., 2013) and changes induced by the stimulus and task (reflected as the post-stimulus oscillations).

Thus, it is crucial to dissect the contributions of pre- and post-stimulus power to the variability of ERD/ERS, which reflect different psychophysiological mechanisms. It is proposed to dissect and quantify the relationship between behavioral variables (e.g., RTs and subjective pain intensity) and pre- and post-stimulus EEG activities, e.g., based on a multivariate linear regression model with the combination of partial least square (PLS) regression (Hu et al., 2014b), thus allowing for a full exploration of electrocortical oscillations involved in pain perception.

POTENTIAL APPLICATIONS IN BASIC AND CLINICAL PAIN STUDIES

By comprehensively investigating neural oscillatory activities relating to the nociceptive sensory inputs (both transient and tonic stimuli) on healthy subjects, it is likely to establish an oscillatory basis of human pain perception and identify how a network of cortical areas involves in human pain experience. The identification of electrophysiological parameters or signatures encoding how the cortex processes the nociceptive inputs and how the experience of pain may emerge from this complex processing, could indeed open a window to study the cortical process underlying pain function as well as the physiology mechanism of nociceptive systems in humans. In clinical practice, this understanding also would make it possible to predict/measure subjective pain intensity objectively, and definitely help (1) explore the pathological mechanisms of chronic pain and (2) achieve pain relief by modulation the oscillatory activities using neurofeedback techniques, with the investigation of cortical oscillatory activities on chronic pain patients.

Identifying the Electrophysiological Signatures of Pain Perception

In the last decades, a large number of EEG/MEG studies (Gross et al., 2007; Iannetti et al., 2008; Schulz et al., 2011; Zhang et al., 2012; Hu et al., 2013, 2014a) have extensively investigated the neural activities in response to the various kinds of nociceptive stimuli, with focusing specifically on temporal aspects of nociceptive processing. LEPs have been used extensively in the past decades for a progress in the understanding of the cortical processes underlying pain perception, with the assumption that they reflect, at least partly, neural activities specifically involved in processing nociceptive somatosensory inputs. However, Mouraux and Iannetti (2009) demonstrated that nociceptive laser-evoked brain potentials do not reflect nociceptive-specific neural activity by showing: (1) LEPs could be entirely explained by a combination of multimodal neural activities and somatosensory-specific neural activities and (2) the magnitudes of the multimodal activities were significantly correlated with subjective ratings of saliency regardless the sensory modalities.

Nevertheless, with recent evidence showed that: (1) pain induced gamma oscillations over S1 covaried with objective stimulus intensity as well as subjective pain intensity (Gross et al., 2007); (2) the magnitudes of laser induced gamma band oscillations could always predict the subjective pain intensity regardless of the stimulus repetition when applying trains of three laser stimuli with constant 1 s interval (Zhang et al., 2012); and (3) tonic heat pain induced gamma oscillations could significantly predict subjective pain intensity (Peng et al., 2014; Schulz et al., 2015), we speculate that the gamma oscillation may be a candidate of the electrophysiological signatures reflecting nociceptive specific neural activities, even further investigation should be done.

Predicting Subjective Pain Intensity

Even pain is a subjective first-person experience, and self-report is considered as the golden standard for the evaluation of pain intensity in clinical situations (Cruccu et al., 2010), self-reports of pain intensity are not available in some vulnerable populations which may lead to inadequate or suboptimal treatment of pain. An objective measurement of pain intensity that can complement self-reports, e.g., to monitor the effect of analgesic drug or the recovery of nociceptive system for non-communicative patients, is in demanding in clinical practice. Even it would be optimal to use pain-specific electrophysiological signatures in predicting subjective pain intensity, using the electrophysiological features that are pain-related but not directly specific to pain processing, could also achieve a relatively high accuracy. For example, for an objective evaluation of pain intensity, Huang et al. (2013) used the evoked potentials information (N2 and P2 latencies and amplitudes) of single-trial LEPs, which are considered to mainly reflect attention capture and arousal to the painful stimuli (Iannetti et al., 2008; Mouraux and Iannetti, 2009), with prediction accuracy of ∼86.3% and ∼80.3% at within-individual and cross-individual level respectively.

Considering (1) the close association between time-frequency oscillatory features (e.g., gamma ERS) with subjective pain intensity (Gross et al., 2007; Zhang et al., 2012); (2) the oscillatory features could provide complementary information of cortical processing that is different from those reflected by evoked potentials (Mouraux and Iannetti, 2008); and (3) the fluctuations of pre-stimulus oscillations could influence and modulate the subsequent sensory perception

(Mathewson et al., 2009; Tu et al., 2016), we propose that the prediction of subjective pain intensity is promising to obtain a better performance with the combination information of stimulus-evoked ERPs, stimulus-induced ERD/ERS, and pre-stimulus oscillation in different frequency bands.

Investigating the Pathological Mechanisms of Chronic Pain: Abnormal Oscillatory Activities in Chronic Pain Patients

Clinical studies have revealed that some chronic pain patients are associated with the occurrence of abnormal cortical oscillatory activities (Sarnthein et al., 2006; Drewes et al., 2008; Sarnthein and Jeanmonod, 2008; Schlee et al., 2009; Walton et al., 2010). By comparing power spectra of the resting EEG of neurogenic pain patients and healthy controls, the patient group exhibited higher resting-EEG power over the frequency range of 2–25 Hz, and the maximal difference appeared in theta frequency band in all electrodes (Sarnthein et al., 2006). Importantly, the excessive theta power gradually decreased and approached normal values after thalamic surgery, suggesting that both EEG and neurogenic pain may be determined by tightly coupled thalamocortical loops (Sarnthein et al., 2006). In addition, the patients with visceral (Drewes et al., 2008) and somatic pain syndromes such as complex regional pain syndrome and neurogenic pain (Sarnthein and Jeanmonod, 2008; Walton et al., 2010) also showed higher baseline levels of delta and/or theta EEG oscillations compared with the healthy controls, localized to the somatosensory cortex corresponding to the pain localization, and to orbitofrontal-temporal cortices related to the affective pain perception. Hepatic encephalopathy patients showed a decreased peak frequency of alpha activity and a delayed alpha rebound in painful stimulus processing over the somatosensory cortex, compared with healthy controls (May et al., 2014). The alternations of the oscillatory activities in chronic pain patients may reflect a dysfunctioned local communication or long-range communication between the functionally specialized assemblies formed by a huge number of neurons in the human brain (Schnitzler and Gross, 2005). Studying the abnormal oscillations in chronic pain patients, could provide insights about the pathological mechanisms underlying chronic pain situations, thus at last leading to a rational basis for the management of pain.

Relieving Pain by Modulating Cortical Oscillatory Activities using Neurofeedback Techniques

With the evidences showing the association between the ongoing oscillatory activities and subsequent sensory perception and behaviors (Rahn and Başar, 1993a,b; Babiloni et al., 2008; Romei et al., 2008; Lange et al., 2012; Tu et al., 2016), the application of neurostimulation techniques outside the skull, such as repetitive transcranial magnetic stimulation (rTMS) and transcranial alternating current stimulation (tACS) that could selectively modulate the oscillatory activities at specific brain areas (e.g., sensorimotor cortex), is promising to relieve pain (Klein et al., 2015). Using these online stimulation techniques could not only reveal the causal roles of the oscillatory brain activities and subjective pain perception, but also may be considered as effective strategies for clinical pain relief.

Indeed, with the delivery of 20 Hz rTMS over S2, patients with chronic visceral pain exhibited significant analgesic effects (Fregni et al., 2005). In addition, subthreshold motor cortex rTMS at 10 Hz to the chronic neuropathic pain patients, could significantly reduce pain intensity and thermal sensory thresholds in the painful zone, and the pain relief showed to be correlated with the improvement of warmth sensory thresholds (Lefaucheur et al., 2008). They interpret the action of rTMS to patients with chronic pain could induce changes of cortical excitability, thus for a restoration of defective intracortical GABAergic inhibitory processes and the normalization of neuronal activity in thermal sensory relays, since chronic neuropathic pain was associated with the motor cortex disinhibition, which may be related to the impairment of GABAergic neurotransmission responsive to some aspects of pain symptom or to the underlying sensory or motor disturbance. In addition, by testing the effectiveness of tACS over S1 at a wide frequency band (ranging from 2–70 Hz), the tACS over S1 could elicit tactile sensation in a frequency-dependent manner (Feurra et al., 2011), with obvious effects at stimulus frequency within both alpha (10–14 Hz) and high gamma (52–70 Hz) ranges, indicating that online stimulation techniques could be used to reveal the causal roles of the brain oscillations.

SUMMARY

Besides ERPs, the nociceptive somatosensory inputs could also induce modulations of cortical oscillations, appeared as ERD or ERS in different frequency bands. These ERD/ERS activities are suggested to be involved in different aspects of pain perception (e.g., sensory perception and behavior), even though the details of their functional roles remain unclear. From a methodological perspective, apart from the temporal, spectral, and spatial profiles of the oscillatory activities, it is instructive to adopt novel parameters (e.g., phase, neural generator, and CFC) to comprehensively evaluate the dynamics of cortical oscillations, thus allowing a full exploration of the neuronal oscillations involved in pain perception. Identifying pain related oscillatory activities and establishing an oscillatory basis of pain perception, could lead new insights into the physiological mechanisms of nociceptive systems in humans. In clinical practice, this also offers exciting prospects for the investigation of pathological mechanisms of chronic pain, thus promoting the development of rational therapeutic strategy.

AUTHOR CONTRIBUTIONS

WWP wrote and revised the manuscript. DDT revised the manuscript.

FUNDING

REFERENCES

Babiloni, C., Brancucci, A., Del Percio, C., Capotosto, P., Arendt-Nielsen, L., Chen, A. C., et al. (2006). Anticipatory electroencephalography alpha rhythm predicts subjective perception of pain intensity. *J. Pain* 7, 709–717. doi: 10.1016/j.jpain. 2006.03.005

Babiloni, C., Del Percio, C., Brancucci, A., Capotosto, P., Le Pera, D., Marzano, N., et al. (2008). Pre-stimulus alpha power affects vertex N2–P2 potentials evoked by noxious stimuli. *Brain Res. Bull.* 75, 581–590. doi: 10.1016/j.brainresbull. 2007.09.009

Bromm, B., and Chen, A. C. (1995). Brain electrical source analysis of laser evoked potentials in response to painful trigeminal nerve stimulation. *Electroencephalogr. Clin. Neurophysiol.* 95, 14–26. doi: 10.1016/0013-4694(95)00032-t

Bromm, B., Jahnke, M. T., and Treede, R. D. (1984). Responses of human cutaneous afferents to CO_2 laser stimuli causing pain. *Exp. Brain Res.* 55, 158–166. doi: 10.1007/bf00240510

Busch, N. A., Dubois, J., and VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876. doi: 10. 1523/JNEUROSCI.0113-09.2009

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., et al. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313, 1626–1628. doi: 10.1126/science.1128115

Canolty, R. T., and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515. doi: 10.1016/j.tics.2010.09.001

Carmon, A., Mor, J., and Goldberg, J. (1976). Evoked cerebral responses to noxious thermal stimuli in humans. *Exp. Brain Res.* 25, 103–107. doi: 10. 1007/bf00237330

Cheyne, D., Gaetz, W., Garnero, L., Lachaux, J. P., Ducorps, A., Schwartz, D., et al. (2003). Neuromagnetic imaging of cortical oscillations accompanying tactile stimulation. *Brain Res. Cogn. Brain Res.* 17, 599–611. doi: 10.1016/s0926-6410(03)00173-3

Cruccu, G., Sommer, C., Anand, P., Attal, N., Baron, R., Garcia-Larrea, L., et al. (2010). EFNS guidelines on neuropathic pain assessment: revised 2009. *Eur. J. Neurol.* 17, 1010–1018. doi: 10.1111/j.1468-1331.2010. 02969.x

de Hemptinne, C., Ryapolova-Webb, E. S., Air, E. L., Garcia, P. A., Miller, K. J., Ojemann, J. G., et al. (2013). Exaggerated phase-amplitude coupling in the primary motor cortex in Parkinson disease. *Proc. Natl. Acad. Sci. U S A* 110, 4780–4785. doi: 10.1073/pnas.1214546110

Del Percio, C., Le Pera, D., Arendt-Nielsen, L., Babiloni, C., Brancucci, A., Chen, A. C., et al. (2006). Distraction affects frontal alpha rhythms related to expectancy of pain: an EEG study. *Neuroimage* 31, 1268–1277. doi: 10.1016/j. neuroimage.2006.01.013

Doesburg, S. M., Green, J. J., Mcdonald, J. J., and Ward, L. M. (2009). Rhythms of consciousness: binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. *PLoS One* 4:e6142. doi: 10.1371/journal. pone.0006142

Dong, L., Gong, D. K., Valdes-Sosa, P. A., Xia, Y., Luo, C., Xu, P., et al. (2014). Simultaneous EEG-fMRI: trial level spatio-temporal fusion for hierarchically reliable information discovery. *Neuroimage* 99, 28–41. doi: 10. 1016/j.neuroimage.2014.05.029

Drewes, A. M., Gratkowski, M., Sami, S. A., Dimcevski, G., Funch-Jensen, P., and Arendt-Nielsen, L. (2008). Is the pain in chronic pancreatitis of neuropathic origin? Support from EEG studies during experimental pain. *World J. Gastroenterol.* 14, 4020–4027. doi: 10.3748/wjg. 14.4020

Fellinger, R., Klimesch, W., Gruber, W., Freunberger, R., and Doppelmayr, M. (2011). Pre-stimulus alpha phase-alignment predicts P1-amplitude. *Brain Res. Bull.* 85, 417–423. doi: 10.1016/j.brainresbull.2011.03.025

Feurra, M., Paulus, W., Walsh, V., and Kanai, R. (2011). Frequency specific modulation of human somatosensory cortex. *Front. Psychol.* 2:13. doi: 10. 3389/fpsyg.2011.00013

Fregni, F., Dasilva, D., Potvin, K., Ramos-Estebanez, C., Cohen, D., Pascual-Leone, A., et al. (2005). Treatment of chronic visceral pain with brain stimulation. *Ann. Neurol.* 58, 971–972. doi: 10.1002/ana.20651

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* 32, 209–224. doi: 10. 1146/annurev.neuro.051508.135603

Gaetz, W. C., and Cheyne, D. O. (2003). Localization of human somatosensory cortex using spatially filtered magnetoencephalography. *Neurosci. Lett.* 340, 161–164. doi: 10.1016/s0304-3940(03)00108-3

Garcia-Larrea, L., Frot, M., and Valeriani, M. (2003). Brain generators of laser-evoked potentials: from dipoles to functional significance. *Neurophysiol. Clin.* 33, 279–292. doi: 10.1016/j.neucli.2003.10.008

Gross, J., Schnitzler, A., Timmermann, L., and Ploner, M. (2007). Gamma oscillations in human primary somatosensory cortex reflect pain perception. *PLoS Biol.* 5:e133. doi: 10.1371/journal.pbio.0050133

Gruber, W. R., Klimesch, W., Sauseng, P., and Doppelmayr, M. (2005). Alpha phase synchronization predicts P1 and N1 latency and amplitude size. *Cereb. Cortex* 15, 371–377. doi: 10.1093/cercor/bhh139

Haig, A. R., and Gordon, E. (1998). Prestimulus EEG alpha phase synchronicity influences N100 amplitude and reaction time. *Psychophysiology* 35, 591–595. doi: 10.1017/s0048577298970512

Hauck, M., Lorenz, J., and Engel, A. K. (2007). Attention to painful stimulation enhances gamma-band activity and synchronization in human sensorimotor cortex. *J. Neurosci.* 27, 9270–9277. doi: 10.1523/jneurosci.2283-07.2007

Hipp, J. F., Engel, A. K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* 69, 387–396. doi: 10.1016/j. neuron.2010.12.027

Hoechstetter, K., Bornfleth, H., Weckesser, D., Ille, N., Berg, P., and Scherg, M. (2004). BESA source coherence: a new method to study cortical oscillatory coupling. *Brain Topogr.* 16, 233–238. doi: 10.1023/b:brat.0000032857. 55223.5d

Hu, L., Peng, W., Valentini, E., Zhang, Z., and Hu, Y. (2013). Functional features of nociceptive-induced suppression of alpha band electroencephalographic oscillations. *J. Pain* 14, 89–99. doi: 10.1016/j.jpain.2012.10.008

Hu, L., Valentini, E., Zhang, Z. G., Liang, M., and Iannetti, G. D. (2014a). The primary somatosensory cortex contributes to the latest part of the cortical response elicited by nociceptive somatosensory stimuli in humans. *Neuroimage* 84, 383–393. doi: 10.1016/j.neuroimage.2013.08.057

Hu, L., Xiao, P., Zhang, Z. G., Mouraux, A., and Iannetti, G. D. (2014b). Single-trial time-frequency analysis of electrocortical signals: baseline correction and beyond. *Neuroimage* 84, 876–887. doi: 10.1016/j.neuroimage.2013.09.055

Hu, L., Zhang, Z. G., Mouraux, A., and Iannetti, G. D. (2015). Multiple linear regression to estimate time-frequency electrophysiological responses in single trials. *Neuroimage* 111, 442–453. doi: 10.1016/j.neuroimage.2015.01.062

Huang, G., Xiao, P., Hung, Y. S., Iannetti, G. D., Zhang, Z. G., and Hu, L. (2013). A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. *Neuroimage* 81, 283–293. doi: 10.1016/j.neuroimage.2013. 05.017

Iannetti, G. D., Hughes, N. P., Lee, M. C., and Mouraux, A. (2008). Determinants of laser-evoked EEG responses: pain perception or stimulus saliency? *Neuroimage* 100, 815–828. doi: 10.1152/jn.00097.2008

Iannetti, G. D., Truini, A., Romaniello, A., Galeotti, F., Rizzo, C., Manfredi, M., et al. (2003). Evidence of a specific spinal pathway for the sense of warmth in humans. *J. Neurophysiol.* 89, 562–570. doi: 10.1152/jn.00393.2002

Jurkiewicz, M. T., Gaetz, W. C., Bostan, A. C., and Cheyne, D. (2006). Post-movement beta rebound is generated in motor cortex: evidence from neuromagnetic recordings. *Neuroimage* 32, 1281–1289. doi: 10.1016/j. neuroimage.2006.06.005

Klein, M. M., Treister, R., Raij, T., Pascual-Leone, A., Park, L., Nurmikko, T., et al. (2015). Transcranial magnetic stimulation of the brain: guidelines for pain treatment research. *Pain* 156, 1601–1614. doi: 10.1097/j.pain. 0000000000000210

Kruglikov, S. Y., and Schiff, S. J. (2003). Interplay of electroencephalogram phase and auditory-evoked neural activity. *J. Neurosci.* 23, 10122–10127.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10. 1152/jn.00263.2005

Lange, J., Halacz, J., van Dijk, H., Kahlbrock, N., and Schnitzler, A. (2012). Fluctuations of prestimulus oscillatory power predict subjective perception of tactile simultaneity. *Cereb. Cortex* 22, 2564–2574. doi: 10.1093/cercor/bhr329

Laufs, H., Kleinschmidt, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., et al. (2003a). EEG-correlated fMRI of human alpha activity. *Neuroimage* 19, 1463–1476. doi: 10.1016/s1053-8119(03)00286-6

Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., et al. (2003b). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc. Natl. Acad. Sci. U S A* 100, 11053–11058. doi: 10.1073/pnas.1831638100

Lefaucheur, J. P., Drouot, X., Menard-Lefaucheur, I., Keravel, Y., and Nguyen, J. P. (2008). Motor cortex rTMS in chronic neuropathic pain: pain relief is associated with thermal sensory perception improvement. *J. Neurol. Neurosurg. Psychiatry* 79, 1044–1049. doi: 10.1136/jnnp.2007.135327

Lei, X., Xu, P., Luo, C., Zhao, J. P., Zhou, D., and Yao, D. Z. (2011). fMRI functional networks for EEG source imaging. *Hum. Brain Mapp.* 32, 1141–1160. doi: 10.1002/hbm.21098

Lenz, F. A., Krauss, G., Treede, R. D., Lee, J. L., Boatman, D., Crone, N., et al. (2000). Different generators in human temporal-parasylvian cortex account for subdural laser-evoked potentials, auditory-evoked potentials and event-related potentials. *Neurosci. Lett.* 279, 153–156. doi: 10.1016/s0304-3940(99)00986-6

López-Azcárate, J., Tainta, M., Rodríguez-Oroz, M. C., Valencia, M., González, R., Guridi, J., et al. (2010). Coupling between beta and high-frequency activity in the human subthalamic nucleus may be a pathophysiological mechanism in Parkinson's disease. *J. Neurosci.* 30, 6667–6677. doi: 10.1523/JNEUROSCI.5459-09.2010

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732. doi: 10.1523/JNEUROSCI.3963-08.2009

May, E. S., Butz, M., Kahlbrock, N., Brenner, M., Hoogenboom, N., Kircheis, G., et al. (2014). Hepatic encephalopathy is associated with slowed and delayed stimulus-associated somatosensory alpha activity. *Clin. Neurophysiol.* 125, 2427–2435. doi: 10.1016/j.clinph.2014.03.018

Miskovic, V., Moscovitch, D. A., Santesso, D. L., McCabe, R. E., Antony, M. M., and Schmidt, L. A. (2011). Changes in EEG cross-frequency coupling during cognitive behavioral therapy for social anxiety disorder. *Psychol. Sci.* 22, 507–516. doi: 10.1177/0956797611400914

Mouraux, A., Guérit, J. M., and Plaghki, L. (2003). Non-phase locked electroencephalogram (EEG) responses to CO_2 laser skin stimulations may reflect central interactions between A partial partial differential- and C-fibre afferent volleys. *Clin. Neurophysiol.* 114, 710–722. doi: 10.1016/s1388-2457(03)00027-0

Mouraux, A., and Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magn. Reson. Imaging* 26, 1041–1054. doi: 10.1016/j.mri.2008.01.011

Mouraux, A., and Iannetti, G. D. (2009). Nociceptive laser-evoked brain potentials do not reflect nociceptive-specific neural activity. *J. Neurophysiol.* 101, 3258–3269. doi: 10.1152/jn.91181.2008

Palva, S., and Palva, J. M. (2011). Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Front. Psychol.* 2:204. doi: 10.3389/fpsyg.2011.00204

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65. doi: 10.1016/0167-8760(84)90014-x

Peng, W. W., Hu, L., Zhang, Z. G., and Hu, Y. (2012). Causality in the association between P300 and Alpha event-related desynchronization. *PLoS One* 7:e34163. doi: 10.1371/journal.pone.0034163

Peng, W., Hu, L., Zhang, Z., and Hu, Y. (2014). Changes of spontaneous oscillatory activity to tonic heat pain. *PLoS One* 9:e91052. doi: 10.1371/journal.pone.0091052

Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Ploner, M., Gross, J., Timmermann, L., Pollok, B., and Schnitzler, A. (2006a). Oscillatory activity reflects the excitability of the human somatosensory system. *Neuroimage* 32, 1231–1236. doi: 10.1016/j.neuroimage.2006.06.004

Ploner, M., Gross, J., Timmermann, L., Pollok, B., and Schnitzler, A. (2006b). Pain suppresses spontaneous brain rhythms. *Cereb. Cortex* 16, 537–540. doi: 10.1093/cercor/bhj001

Rahn, E., and Başar, E. (1993a). Enhancement of visual-evoked potentials by stimulation during low prestimulus EEG stages. *Int. J. Neurosci.* 72, 123–136. doi: 10.3109/00207459308991620

Rahn, E., and Başar, E. (1993b). Prestimulus EEG-activity strongly influences the auditory evoked vertex response: a new method for selective averaging. *Int. J. Neurosci.* 69, 207–220. doi: 10.3109/00207459309003331

Raij, T. T., Forss, N., Stancák, A., and Hari, R. (2004). Modulation of motor-cortex oscillatory activity by painful Aδ- and C-fiber stimuli. *Neuroimage* 23, 569–573. doi: 10.1016/j.neuroimage.2004.06.036

Rajkai, C., Lakatos, P., Chen, C. M., Pincze, Z., Karmos, G., and Schroeder, C. E. (2008). Transient cortical excitation at the onset of visual fixation. *Cereb. Cortex* 18, 200–209. doi: 10.1093/cercor/bhm046

Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., and Thut, G. (2008). Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. *Cereb. Cortex* 18, 2010–2018. doi: 10.1093/cercor/bhm229

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., and Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756. doi: 10.1126/science.1223082

Sarnthein, J., and Jeanmonod, D. (2008). High thalamocortical theta coherence in patients with neurogenic pain. *Neuroimage* 39, 1910–1917. doi: 10.1016/j.neuroimage.2007.10.019

Sarnthein, J., Stern, J., Aufenberg, C., Rousson, V., and Jeanmonod, D. (2006). Increased EEG power and slowed dominant frequency in patients with neurogenic pain. *Brain* 129, 55–64. doi: 10.1093/brain/awh631

Schlee, W., Hartmann, T., Langguth, B., and Weisz, N. (2009). Abnormal resting-state cortical coupling in chronic tinnitus. *BMC Neurosci.* 10:11. doi: 10.1186/1471-2202-10-11

Schnitzler, A., and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nat. Rev. Neurosci.* 6, 285–296. doi: 10.1038/nrn1650

Schnitzler, A., Gross, J., and Timmermann, L. (2000). Synchronised oscillations of the human sensorimotor cortex. *Acta Neurobiol. Exp. (Wars)* 60, 271–287.

Schulz, E., May, E. S., Postorino, M., Tiemann, L., Nickel, M. M., Witkovsky, V., et al. (2015). Prefrontal Gamma oscillations encode tonic pain in humans. *Cereb. Cortex* 25, 4407–4414. doi: 10.1093/cercor/bhv043

Schulz, E., Tiemann, L., Schuster, T., Gross, J., and Ploner, M. (2011). Neurophysiological coding of traits and states in the perception of pain. *Cereb. Cortex* 21, 2408–2414. doi: 10.1093/cercor/bhr027

Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci.* 3, 151–162. doi: 10.1016/s1364-6613(99)01299-1

Tarkka, I. M., and Treede, R. D. (1993). Equivalent electrical source analysis of pain-related somatosensory evoked potentials elicited by a CO_2 laser. *J. Clin. Neurophysiol.* 10, 513–519. doi: 10.1097/00004691-199310000-00009

Thut, G., Nietzel, A., Brandt, S. A., and Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* 26, 9494–9502. doi: 10.1523/jneurosci.0875-06.2006

Treede, R. D. (2003). Neurophysiological studies of pain pathways in peripheral and central nervous system disorders. *J. Neurol.* 250, 1152–1161. doi: 10.1007/s00415-003-0237-7

Treede, R. D., Lorenz, J., and Baumgärtner, U. (2003). Clinical usefulness of laser-evoked potentials. *Neurophysiol. Clin.* 33, 303–314. doi: 10.1016/j.neucli.2003.10.009

Tu, Y. H., Zhang, Z. G., Tan, A., Peng, W. W., Hung, Y. S., Moayedi, M., et al. (2016). Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming stimuli. *Hum. Brain Mapp.* 37, 501–514. doi: 10.1002/hbm.23048

Valentini, E., Hu, L., Chakrabarti, B., Hu, Y., Aglioti, S. M., and Iannetti, G. D. (2012). The primary somatosensory cortex largely contributes to the early part of the cortical response elicited by nociceptive stimuli. *Neuroimage* 59, 1571–1581. doi: 10.1016/j.neuroimage.2011.08.069

Vanni, S., Revonsuo, A., and Hari, R. (1997). Modulation of the parieto-occipital alpha rhythm during object detection. *J. Neurosci.* 17, 7141–7147.

von Stein, A., and Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta

synchronization. *Int. J. Psychophysiol.* 38, 301–313. doi: 10.1016/s0167-8760(00)00172-0

Walton, K. D., Dubois, M., and Llinás, R. R. (2010). Abnormal thalamocortical activity in patients with Complex Regional Pain Syndrome (CRPS) type I. *Pain* 150, 41–51. doi: 10.1016/j.pain.2010.02.023

Wang, J., Li, D., Li, X. L., Liu, F. Y., Xing, G. G., Cai, J., et al. (2011). Phase-amplitude coupling between theta and gamma oscillations during nociception in rat electroencephalography. *Neurosci. Lett.* 499, 84–87. doi: 10.1016/j.neulet.2011.05.037

Wang, L., Saalmann, Y. B., Pinsk, M. A., Arcaro, M. J., and Kastner, S. (2012). Electrophysiological low-frequency coherence and cross-frequency coupling contribute to BOLD connectivity. *Neuron* 76, 1010–1020. doi: 10.1016/j.neuron.2012.09.033

Yao, D. Z. (2001). A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiol. Meas.* 22, 693–711. doi: 10.1088/0967-3334/22/4/305

Zhang, Z. G., Hu, L., Hung, Y. S., Mouraux, A., and Iannetti, G. D. (2012). Gamma-band oscillations in the primary somatosensory cortex-a direct and obligatory correlate of subjective pain intensity. *J. Neurosci.* 32, 7429–7438. doi: 10.1523/JNEUROSCI.5877-11.2012

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Normalization of Pain-Evoked Neural Responses Using Spontaneous EEG Improves the Performance of EEG-Based Cross-Individual Pain Prediction

Yanru Bai [1,2], Gan Huang [2], Yiheng Tu [3], Ao Tan [3], Yeung Sam Hung [3] and Zhiguo Zhang [1,2]*

[1] School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore, Singapore, [2] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China, [3] Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, Hong Kong

An effective physiological pain assessment method that complements the gold standard of self-report is highly desired in pain clinical research and practice. Recent studies have shown that pain-evoked electroencephalography (EEG) responses could be used as a readout of perceived pain intensity. Existing EEG-based pain assessment is normally achieved by cross-individual prediction (i.e., to train a prediction model from a group of individuals and to apply the model on a new individual), so its performance is seriously hampered by the substantial inter-individual variability in pain-evoked EEG responses. In this study, to reduce the inter-individual variability in pain-evoked EEG and to improve the accuracy of cross-individual pain prediction, we examined the relationship between pain-evoked EEG, spontaneous EEG, and pain perception on a pain EEG dataset, where a large number of laser pulses (>100) with a wide energy range were delivered. Motivated by our finding that an individual's pain-evoked EEG responses is significantly correlated with his/her spontaneous EEG in terms of magnitude, we proposed a normalization method for pain-evoked EEG responses using one's spontaneous EEG to reduce the inter-individual variability. In addition, a nonlinear relationship between the level of pain perception and pain-evoked EEG responses was obtained, which inspired us to further develop a new two-stage pain prediction strategy, a binary classification of low-pain and high-pain trials followed by a continuous prediction for high-pain trials only, both of which used spontaneous-EEG-normalized magnitudes of evoked EEG responses as features. Results show that the proposed normalization strategy can effectively reduce the inter-individual variability in pain-evoked responses, and the two-stage pain prediction method can lead to a higher prediction accuracy.

Keywords: pain-evoked EEG, spontaneous EEG, normalization, cross-individual prediction, pain prediction

INTRODUCTION

Pain is an unpleasant experience related to substantive or potential tissue damage (Loeser and Treede, 2008; Brown et al., 2011). Self-report is the gold standard to determine the presence, absence, and the degree of pain perception in clinic practice (Cruccu et al., 2010; Haanpää et al., 2011), but it may fail in certain patient populations, e.g., patients who suffer from consciousness disorders or are in coma (Schnakers and Zasler, 2007). Lack of accurate pain assessment in these populations can lead to inadequate or suboptimal treatment of pain. Therefore, it is of high importance to develop a physiology-based pain assessment method that is independent of participants' subjective rating (Brown et al., 2011; Terhaar et al., 2011; Huang et al., 2013b).

As a sensory perception that involves a complex set of brain activities, pain has been under intensive investigations using brain imaging techniques, such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). A variety of neural signatures of pain have been identified from brain imaging data, and the possibility to utilize pain-related neural signatures for pain assessment has been explored in many studies (Bromm and Treede, 1983; Iannetti et al., 2003; Marquand et al., 2010; Brown et al., 2011; Brodersen et al., 2012; Schulz et al., 2012; Zhang et al., 2012). For example, a support vector machine (SVM) trained on fMRI data was verified to be possible for pain assessment (Brown et al., 2011). Schulz et al. also applied a multivariate pattern analysis to predicted individual's pain sensitivity using single-trial pain-related EEG (Schulz et al., 2012).

Particularly, EEG-based pain assessment has attracted a growing interest in recent years, not only because the EEG technique is cheap, easy-to-use, and non-invasive, but also because the relationship between EEG responses and pain perception has been relatively well recognized. In basic research of pain, EEG activities elicited by nociceptive laser heat pulses are widely used to assess neural processing of nociceptive pain (Bromm and Treede, 1983; Iannetti et al., 2003; Treede et al., 2003). A positive relationship between the intensity of perceived pain and a variety of components (such as N2 and P2) in laser-evoked EEG responses has been well documented (Kakigi et al., 1989; Bromm and Treede, 1990; Beydoun et al., 1993; Arendt-Nielsen, 1994; Garcí-Larrea et al., 1997; Iannetti et al., 2005; Huang et al., 2013a). Based on the existing knowledge on relationship between laser-evoked EEG and pain, we have developed a method to predict the level of subjective pain perception using single-trial laser-evoked EEG potentials (Huang et al., 2013a), and achieved good predictive accuracy.

Pain prediction using evoked EEG can be realized at two levels: within-individual pain prediction (the classifier and the prediction model were trained on and applied to the same group of individuals) and cross-individual prediction (the classifier and the prediction model were trained on a group of individuals but applied to different individuals). Since within-individual pain prediction requires real pain ratings for new individuals and is not applicable to people who are unable to reliably express their pain perception (Brodersen et al., 2012; Schulz et al.,

2012), cross-individual pain prediction is more desired in clinical uses. However, our previous work (Huang et al., 2013b) showed that cross-individual pain prediction has a significantly lower performance than within-individual prediction, mainly because of the inherent inter-individual variability in pain perception and neural responses. Therefore, incorporating individual factors that are particularly related to inter-individual variability of pain perception or pain-evoked neural activities into the cross-individual pain prediction model is crucial in pain assessment and pain therapy (Davis, 2011), and it is also the objective of the present study.

In an EEG-based pain prediction model that links EEG signals and subjective pain ratings, substantial inter-individual variability is involved in both EEG and ratings. However, it is difficult to improve the performance of cross-individual pain prediction by means of reducing the inter-individual variability of subjective pain ratings, because pain ratings for a new individual, as the unknown variables to be predicted, are not accessible. Therefore, this study is focused exclusively on decreasing the inter-individual variability of pain-related EEG responses: we aim to explore how the pain-related EEG responses vary between individuals at different levels of pain and how to normalize pain-related EEG responses across individuals for a more accurate EEG-based cross-individual pain prediction.

In the present study, we hypothesize that an individual's spontaneous EEG activity can be used to normalize his/her pain-evoked EEG responses so as to improve the accuracy of cross-individual pain prediction. This hypothesis is induced by strong and consistent evidence showing that the magnitudes of a variety of pain-evoked EEG responses are highly correlated with that of spontaneous EEG of the same individual. Actually, the magnitudes of both spontaneous and pain-evoked EEG activities are altered by the difference in individual-specific factors, such as cortical anatomy (e.g., the thickness of the skin and skull) and experimental conditions (e.g., electrode position and scale-electrode impedances; Klistorner and Graham, 2001; You et al., 2012). Therefore, the magnitude of spontaneous EEG has the potential to serve as an individual scale to normalize the magnitude of pain-evoked EEG responses for a reduced inter-individual variability. The normalized magnitudes of pain-evoked EEG are used as features in the subsequent pain prediction. Next, a two-stage cross-individual pain prediction method is developed: a binary classifier to discriminate low-pain (NRS ≤ 4) and high-pain (NRS > 4) followed by a linear prediction model to predict the pain ratings (4–10) for high-pain trials only. The results showed that the proposed spontaneous EEG based normalization can effectively decrease the inter-individual variability in the classifiers and prediction models, and consequently, can increase the accuracy of pain prediction, as compared with the prediction based on raw pain-evoked EEG responses.

MATERIALS AND METHODS

Participants

Thirty-four healthy volunteers (17 females and 17 males), aged 18–25 years (Mean ± SD: 21.6 ± 1.7), without a history of

chronic pain, participated in the study. All volunteers gave their written informed consent and were paid for their participation. The experiment procedures were approved by the local ethics committee. Before the experiment, they were familiarized with the experimental setup and task.

Experimental Design

Nociceptive-specific radiant-heat stimuli were generated by an infrared neodymium yttrium aluminiumperovskite (Nd:YAP) laser with a wavelength of 1.34 μm (Electronical Engineering, Italy). At this wavelength, laser pulses activate directly nociceptive terminals in the most superficial skin layers. The laser beam was transmitted via an optic fiber and its diameter was set at ∼ 7 mm (≈38 mm^2) by focusing lenses. Laser pulses were directed at the medial side of the dorsum of left hand, between the first and third metacarpus. A He-Ne laser pointed to the area to be stimulated. The duration of the laser pulse was fixed at 4 ms. After each stimulus, the target of the laser beam was shifted by more than 1 cm in a random direction, to avoid nociceptor fatigue or sensitization.

Participants were asked to report the intensity of perceived pain elicited by the laser stimulus, using a numerical rating scale (NRS) ranging from 0 (no pain) to 10 (pain as bad as it could be). Prior to EEG data collection, the highest energy of the laser stimulation, used in the following experiment, was individually determined using the method of limits (from 1 J in step of 0.25 J) until a rating of 8 was reached. No withdrawal reflexes and motor contractions were observed until the stimulation intensity increased to 3.75–4.5 J. During the EEG data collection, 12–15 different levels of laser stimulation energies (from 1 to 3.75–4.5 J, in step of 0.25 J) were adopted, and 10 laser pulses at each energy level, for a total of 120–150 pulses, were delivered in two blocks. Before each block, the surface temperatures of hand dorsum for each participant were measured using an infrared thermometer. The order of stimulus energies was pseudo-randomized. The inter-stimulus interval (ISI) varied randomly between 10 and 15 s (uniformly distributed). An auditory tone delivered between 3 and 6 s after the laser pulse (uniformly distributed) prompted the participants to rate the intensity of pain. The dataset with 12–15 levels of stimulation energy enables a more comprehensive and detailed investigation of the relationship between pain-evoked EEG responses and spontaneous EEG activities and the relationship between pain-evoked EEG responses and subjective pain ratings.

EEG Recording

Participants were seated in a comfortable chair in a silent, temperature-controlled room. They wore protective goggles and were asked to focus their attention on the stimuli and relax their muscles. The EEG data were recorded using a 64-channel EEG cap with Ag-AgCl scalp electrodes placed according to the international 10–20 system (Brain Products GmbH, Munich, Germany; pass band: 0.01–100 Hz; sampling rate: 1000 Hz). The nose was used as the reference electrode, and the impedances of all electrodes were kept lower than 10 kΩ. Electrooculographic (EOG) signals were simultaneously recorded using surface electrodes to monitor ocular movements and eye blinks.

EEG Data Analysis
Preprocessing

Continuous EEG data from Cz channel were band-pass filtered between 1 and 30 Hz. EEG epochs were extracted using a window analysis time of 1 s (from 0.5 s pre-stimulus to 0.5 s post-stimulus), and baseline corrected using the pre-stimulus interval (−0.5 to 0 s). Artifacts due to eye blinks or eye movements were subtracted using independent component analysis. In all datasets, the independent components which had a large EOG channel contribution and a frontal scalp distribution were removed. The above EEG data preprocessing were realized using EEGLAB (Delorme and Makeig, 2004), an open source toolbox running in MATLAB environment.

Feature Extraction

EEG trials recorded at Cz (nose referenced) were used for prediction of pain perception. Each EEG trial consists of two segments: the pre-stimulus trial (−0.5 to 0 s) is spontaneous EEG (sEEG) and the post-stimulus trial (0–0.5 s) is dominated by pain-evoked EEG (pEEG) or, more precisely, Aδ-fiber pain-evoked EEG responses. The magnitude of sEEG or pEEG trial is quantified by root mean square (RMS)

$$RMS = \sqrt{\frac{1}{K} \sum_{k=1}^{K} x_k^2}, \tag{1}$$

where x_k is the k-th sample of the trial, and K is the number of data samples. The RMS of sEEG or pEEG are denoted as RMS_S or RMS_P and will be used as features in subsequent investigation of the relationship between pain and EEG and in pain prediction.

Relationship between sEEG and pEEG

To test whether sEEG can serve as a baseline to normalize pEEG for a smaller inter-individual variability, we examined the relationship between RMS_S and RMS_P. We assume that an individual's RMS_S and RMS_P (both of which are averaged across all trials at each pain intensity level) are normally distributed, and calculate the mean and standard deviation (SD) of RMS_S and RMS_P across all trials at each pain intensity level. Then, cross-individual correlation between these mean and SD values were estimated.

Relationship between Pain and pEEG

In our experiments, participants were asked to report the level of pain perception with four as the pinprick pain threshold (i.e., NRS > 4 refers to feeling of pinprick pain). Thus, NRS = 4 serves as a threshold to differentiate low-pain and high-pain (i.e., low-pain: NRS≤ 4, high-pain: NRS > 4). To investigate the relationship between the rating of perceived pain and evoked EEG responses, RMS_P, which was averaged across trials with each identical pain level for each individual, was fitted using two models: a global linear model and a two-piecewise linear model (two segments with NRS = 4 as the break point). The global linear model is based on the assumption that magnitude of pEEG is linearly increased with the pain rating, while the piecewise linear model assumes that the relationship between pain rating and pEEG is different for low-pain (NRS ≤ 4) and high-pain (NRS

> 4). The performance of the two models are quantified with the mean square error (MSE) and compared across individuals using a paired sample *t*-test.

Feature Normalization Based on sEEG

There are different normalization methods available, and here we normalized the magnitude of a pEEG trial as the *z*-score of the population defined by sEEG trials. For each individual, the magnitude of the *i*-th pEEG trial, $RMS_P(i)$, was normalized by RMS_S of all sEEG trials as

$$nRMS_P(i) = \frac{RMS_P(i) - \mu(RMS_S)}{\sigma(RMS_S)}, \qquad (2)$$

where $nRMS_P(i)$ is the normalized magnitude of the *i*-th pEEG trial, μ and σ are respectively the mean and the SD of RMS_S of all trials of this individual.

To examine whether the inter-individual variability of pEEG magnitudes was decreased by the sEEG-based normalization, we performed an ANOVA *F*-test on RMS_P and $nRMS_P$. More precisely, at each pain level, an ANOVA *F*-test was performed on RMS_P or $nRMS_P$ of all trials with this pain level across all individuals to check whether the means of RMS_P or $nRMS_P$ of this group of individuals are the same, and the resultant *F*-statistics denote the inter-individual variability relative to the within-individual variability of the variable under test. Then, we compared the *F*-statistics between RMS_P and $nRMS_P$ at each pain level, and it is expected that $nRMS_P$ has a smaller *F*-statistic than RMS_P.

We next investigate whether the proposed sEEG-based normalization method can reduce the cross-individual variability in the relationship between intensity of pain perception and the magnitude of pEEG. Firstly, for each individual, we calculated an optimal threshold of RMS_P or $nRMS_P$ that can best classify (with the highest accuracy) the individual' trials into low-pain (NRS $\leq$ 4) and high-pain (NRS $>$ 4). The inter-individual variability of the binary classification thresholds of RMS_P and $nRMS_P$ were measured by variance. If the variance of the threshold obtained from $nRMS_P$ is smaller than that from RMS_P, the effectiveness of sEEG-based normalization in reducing the individual difference in the binary classification can be validated. A two-sample *F*-test was also conducted to check whether the thresholds of RMS_P and $nRMS_P$ have the same variance. Secondly, we consider the relationship between pain ratings and RMS_P (or $nRMS_P$) of high-pain (NRS $>$ 4) trials to be a linear model specific to each individual, then the inter-individual variability in this relationship is indicated by the cross-individual variance of slope and intercept of the linear model. So, slopes and intercepts of all individuals were calculated using two sets of features (RMS_P and $nRMS_P$), and their inter-individual variability were compared. If the cross-individual variance of slopes or intercepts obtained using $nRMS_P$ is smaller than that obtained using RMS_P, the effectiveness of sEEG-based normalization in reducing the individual variability in the continuous prediction model can be validated. Similarly, a two-sample *F*-test was also conducted to check whether the slopes or intercepts of RMS_P and $nRMS_P$ have the same variance.

Binary Classification (Low-Pain vs. High-Pain)

A linear discriminant analysis (LDA) classifier was adopted to classify low-pain and high-pain trials using leave-one-individual-out cross validation. The classifier was first trained with RMS_P or $nRMS_P$ of training trials, which divided into two categories (low-pain: NRS $\leq$ 4, and high-pain: NRS $>$ 4), and then applied to the test trials to predict labels (low-pain vs. high-pain) from the corresponding RMS_P or $nRMS_P$. The classification performance was evaluated by accuracy, and the accuracies obtained from classification using RMS_P and $nRMS_P$ were compared using paired sample *t*-test.

Continuous Prediction of Pain Levels for High-Pain Trials

After binary classification, only high-pain trials are involved in continuous pain prediction, because there is no significant correlation between pain ratings and pEEG of low-pain trials. To prove that the sEEG-based normalization is effective for continuous pain prediction regardless of the results of the preceding binary classification, we performed continuous pain prediction for trials predicted as high-pain from binary classification as well as for real high-pain trials (NRS $>$ 4).

Relationship between single-trial RMS_P (or $nRMS_P$) and the corresponding intensity of pain perception was modeled by linear regression. For the *i*-th pEEG trial, the pain rating R_i can be estimated as

$$R_i = \alpha \cdot RMS_P(i) + c, \qquad (3)$$
$$R_i = \alpha \cdot nRMS_P(i) + c, \qquad (4)$$

where α and c are slope and intercept of the linear regression model. The model of (Equations 3 and 4) was trained and tested using leave-one-individual-out cross validation.

The prediction performance of the linear regression model was evaluated by Mean Absolute Error (MAE),

$$MAE = \frac{1}{N} \sum_{n=1}^{N} \left| R_i - \hat{R}_i \right|, \qquad (5)$$

where N is the total number of test trials, R_i and $\hat{R}_i$ are respectively the real and predicted rating value for the *i*-th trial. MAE provides a straightforward measure on how precisely the generated linear regression model can represent the relationship between pain ratings and pEEG magnitudes. The MAE values obtained from prediction using RMS_P and $nRMS_P$ were compared using the paired sample *t*-test.

RESULTS

Relationship between sEEG and pEEG

For each participant, the mean and SD of RMS_S and RMS_P at each pain intensity level were calculated. Since NRS $>$ 8 was not available for some participants, we use a combined level of "NRS $>$ 8" to denote all trials with an NRS $>$ 8. It can be clearly seen from **Figure 1** and **Table 1** that, a significant correlation ($p \leq 0.007$) between the mean values of RMS_S and RMS_P was
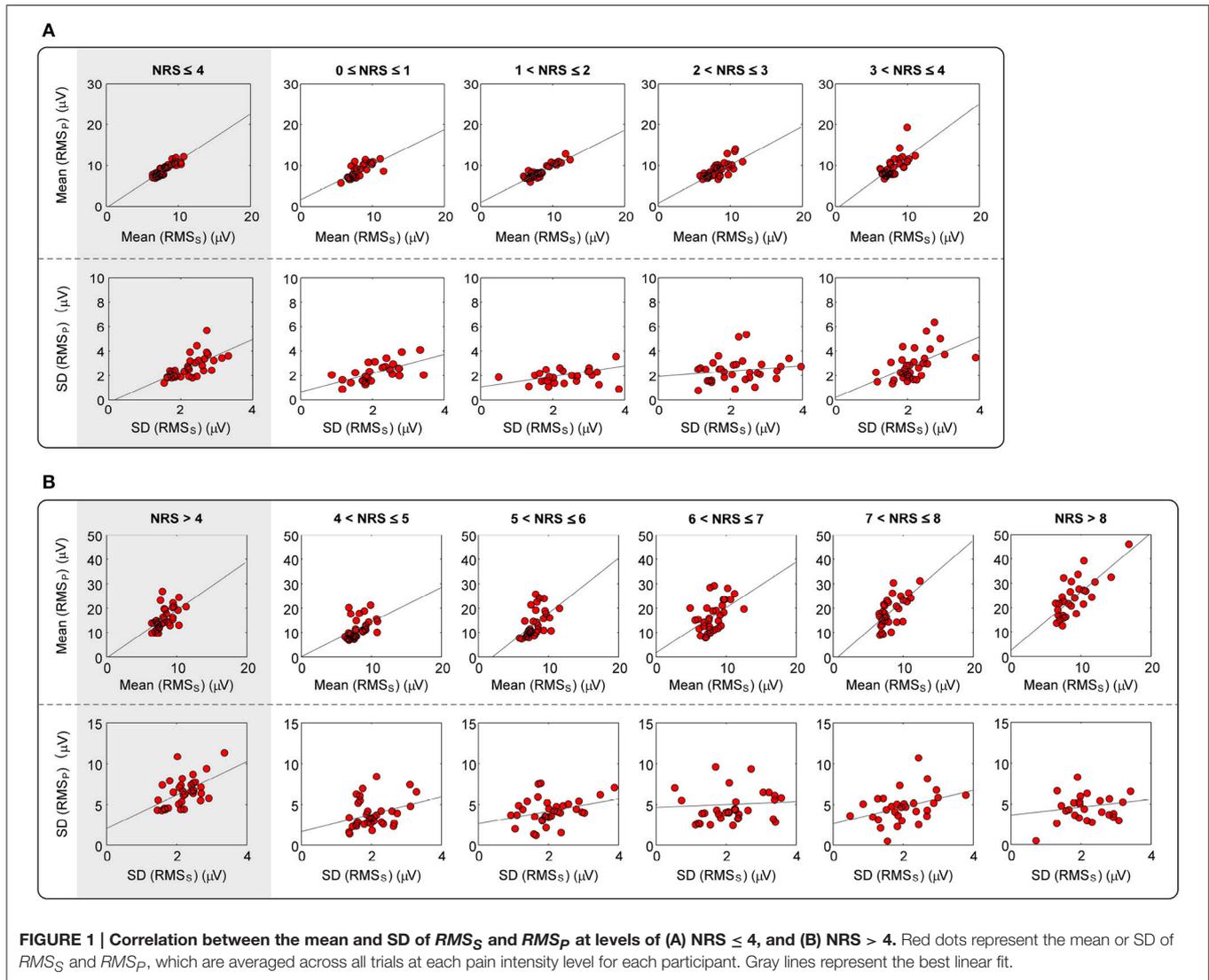
FIGURE 1 | Correlation between the mean and SD of RMS_S and RMS_P at levels of (A) NRS ≤ 4, and (B) NRS > 4. Red dots represent the mean or SD of RMS_S and RMS_P, which are averaged across all trials at each pain intensity level for each participant. Gray lines represent the best linear fit.

TABLE 1 | Correlation between the mean and SD of RMS_S and RMS_P at each pain level.

NRS		(0, 4]	(0, 1]	(1, 2]	(2, 3]	(3, 4]	
Mean	R	0.891	0.721	0.930	0.725	0.666	
	P-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	
SD	R	0.659	0.614	0.466	0.171	0.537	
	P-value	< 0.001	< 0.001	0.013	0.335	0.001	
NRS		**(4, 10]**	**(4, 5]**	**(5, 6]**	**(6, 7]**	**(7, 8]**	**(8, 10]**
Mean	R	0.550	0.454	0.525	0.570	0.633	0.555
	P-value	< 0.001	0.007	0.001	< 0.001	< 0.001	0.003
SD	R	0.519	0.303	0.340	0.065	0.402	0.193
	P-value	0.002	0.082	0.049	0.718	0.025	0.346

obtained at each intensity level of pain perception. In addition, a significant correlation between the SD values of RMS_S and RMS_P was also obtained ($p \leq 0.02$) for overall intensity level of low-pain (NRS ≤ 4) and high-pain (NRS > 4), though some individual intensity level is not significant (such as intensity level at 2–3, 4–5, 6–7, and 8–10). To conclude, the distributions of RMS_S and

FIGURE 2 | (A) Relationship between pain ratings and RMS_P (from one participant). Colored dots represent mean ± SD of RMS_P averaged across trials at different level of pain perception. The red line represents the fitted global linear model, while the blue lines represent the fitted two-piecewise linear model. **(B)** Comparison of MSE (mean ± SD) of all participants between two fitting models.

TABLE 2 | Comparison of *F*-statistics obtained from the ANOVA *F*-test on RMS_P and $nRMS_P$.

NRS	(0, 1]	(1, 2]	(2, 3]	(3, 4]	(4, 5]	(5, 6]	(6, 7]	(7, 8]	(8, 10]
RMS_P	5.279	6.476	5.802	10.553	17.022	13.501	17.278	10.420	18.718
$nRMS_P$	1.182	1.475	1.582	4.795	11.624	11.332	13.792	8.172	9.310

RMS_P are highly correlated, which verifies our hypothesis that magnitudes of pEEG are highly correlated with the magnitude of sEEG. This observation supports the idea that RMS_S could serve as an individual scale to normalize his/her RMS_P to reduce inter-individual variability of pain-related features in pain classification and prediction models.

Relationship between Pain and pEEG

Figure 2A shows the relationship between pain rating and the magnitude of pEEG (mean ± SD) of one participant. Overall, pain rating and RMS_P are positively related, but RMS_P does not increase significantly when the subjective pain ratings is ≤ 4 (referred to as "low-pain"); when the subjective pain rating is >4 (referred to as "high-pain"), RMS_P is linearly increased with pain ratings. MSE of the global linear model (red line) or the two-piecewise linear model (blue line) was adopted to measure the accuracy of fitting, as shown in **Figure 2B**. It can be seen from the group-level results in **Figure 2B** that, the fitting error of the piecewise linear model is significantly smaller than that of the global linear model. Therefore, the piecewise linear model can better describe the relationship between pain perception and RMS_P. The nonlinear relationship motivates us to develop the two-stage pain prediction (i.e., to classify low- and high-pain first, then to predict the pain rating for high-pain trials only).

Feature Normalization Based on sEEG

We first confirmed that the magnitudes of sEEG trials and pEEG trials of each individual approximately follow a normal

distribution ($p < 0.0001$ for all individuals, one-sample Kolmogorov–Smirnov test). Therefore, an individual's sEEG trials could form a distribution for normalizing pEEG trials into *z*-scores. **Table 2** shows that, at each pain level, the *F*-statistic obtained from the ANOVA *F*-test on $nRMS_P$ is lower than that obtained from RMS_P, which proves that the sEEG-based normalization method can effectively reduce the inter-individual variability of pEEG trials.

Figure 3 shows the optimal binary classification thresholds and slopes/intercepts of linear regression models for high-pain trials, which were obtained from available trials and ratings of all individuals. We can clearly see that, after sEEG-based normalization, the variance of all above three parameters were remarkably decreased, which illustrates that the proposed sEEG-based normalization method can effectively reduce the inter-individual variability in classification and prediction models.

Table 3 further shows that, cross-individual variances of all three classifier/model parameters were decreased after sEEG-based normalization. The two-sample *F*-test also demonstrates that the variances of all three classifier/model parameters are significantly different between using RMS_P and using $nRMS_P$ as features ($p < 0.0001$ for all).

Pain Prediction

The mean and SD of accuracy for binary pain classification (low-pain vs. high-pain) using RMS_P (i.e., pEEG features) and using $nRMS_P$ (i.e., sEEG-normalized pEEG features)

FIGURE 3 | Effects of sEEG normalization on inter-individual variability of (A) binary classification thresholds, (B) slopes of linear regression models for high-pain trials, (C) intercepts of linear regression models for high-pain trials. Noted that mean values were removed from these parameters for illustration. The box plots show the minimu, lower quartile, median, upper quartile, and maximum values of one group of variables.

TABLE 3 | Comparison of binary classifier thresholds and model parameters between using RMS_P and using $nRMS_P$.

		RMS_P	$nRMS_P$
Binary classfication thresholds	Cross-individual variance	4.284	1.797
	P-value (F-test for equal variance)	<0.001	
Slopes of linear models	Cross-individual variance	1.340	0.522
	P-value (F-test for equal variance)	<0.001	
Intercepts of linear models	Cross-individual variance	7.525	3.254
	P-value (F-test for equal variance)	<0.001	

TABLE 4 | Accuracy of binary classification and prediction error (MAE) of continuous prediction.

	RMS_P	$nRMS_P$	P-value (paired t-test)
Accuracy of binary classification (%)	68.95 ± 12.91	70.36 ± 14.18	0.092
MAE of continuous prediction (on predicted high-pain trials)	1.838 ± 0.602	1.625 ± 0.446	0.002
MAE of continuous prediction (on real high-pain trials)	1.235 ± 0.278	1.173 ± 0.278	0.003

are summarized in **Table 4**. Results show that $nRMS_P$ can improve the classification accuracy, though the performance improvement is only close to significant ($p = 0.092$).

The mean and SD of MAE for continuous pain prediction using RMS_P (i.e., pEEG responses) and using $nRMS_P$ (i.e., sEEG-normalized pEEG responses) for trials predicted as high-pain from binary classification and for real high-pain trials (NRS > 4) are summarized in **Table 4**. Results indicate that the proposed sEEG-based normalization method can significantly improve the prediction accuracy for both predicted high-pain trials

($p = 0.002$) and real high-pain trials ($p = 0.003$) in continuous pain prediction.

DISCUSSION

In this study, we proposed to normalize pain-evoked EEG responses using spontaneous EEG to improve the performance of EEG-based pain prediction. Pain-related EEG responses have been used to predict the level of subjective pain, but the large inter-individual variability seriously degrades the performance of cross-individual pain prediction. In this work, we began by performing a comprehensive and detailed investigation of the relationship between pEEG responses and sEEG activities as well as the relationship between subjective pain ratings and pEEG responses. Our results revealed a strong inter-individual correlation between the magnitude of pEEG and sEEG. Besides, our results also confirmed a nonlinear relationship between pEEG and subjective pain ratings. Based on above observations, we proposed a new two-stage approach for pain prediction: (1) a binary classification to differentiate low-pain and high-pain trials; (2) a continuous regression to predict pain ratings of high-pain trials. In both steps, the normalization strategy based on sEEG was used to reduce the inter-individual variability in the magnitude of pEEG, so that a higher classification accuracy and a lower prediction error were achieved. The new sEEG-based normalization strategy has the potential to contribute to an applicable and reliable tool for pain assessment.

Relationship between sEEG and pEEG

An individual's spontaneous EEG has been shown to be related to his/her genetic code, implying its uniqueness (Tran et al., 2001; Doležal et al., 2005; Anokhin et al., 2006; Marcel and Del Millan, 2007; Näpflin et al., 2007; Zietsch et al., 2007). A strong inter-individual correlation between the magnitude of sEEG and pEEG was also obtained from our database. A potential interpretation for this phenomenon may be due to the skull thickness, the orientation of the gray matter and so forth. These anatomical factors are specific to each person and will remain relatively stable

for adults. Experimental conditions, such as electrode position and scale-electrode impedances, could also contribute to the phenomenon, because they may influence the magnitudes of both pEEG and sEEG.

Relationship between Pain Rating and Pain-Evoked EEG Responses

Numerous previous studies have shown that the perceived pain intensity is strongly correlated with the amplitude of a number of evoked EEG responses (Iannetti et al., 2005; Huang et al., 2013b). In most of these works, the level of pain perception was assumed or found to be linearly correlated with the evoked EEG responses, but such a linear relationship has been challenged by growing evidence showing the nonlinearity between pain level and the neural responses. The assumption and observation of the linear relationship may due to the limited range of painful stimulus intensities used in most of pain experiments, which further limited the range of perceived pain intensity. To solve this problem, our experiment was designed to deliver a large number of laser pulses (>100) with a wide energy range (from 1 to 3.75–4.5 J; 12−15 levels) to each participant. The result that the fitting error of a two-piecewise linear model (with a break point of NRS = 4) was significantly smaller than that of a global linear model indicated a nonlinear relationship between pain level and the evoked EEG responses.

Feature Normalization Based on sEEG

To normalize the magnitudes of pEEG trials of one individual, we proposed to estimate their z-scores in the population defined by sEEG trials of this individual. Although other normalization methods exist, the proposed z-scores can achieve better prediction results than other normalization methods, such as dividing RMS_P with the mean of sEEG or subtracting the mean of sEEG from RMS_P (results are not shown here). Besides its good performance, the proposed z-score normalization also reflects certain physiological meanings. It has been revealed that the variability of spontaneous neural activity can reflect the "dynamic range" of possible neural responses to incoming stimuli and can provide a powerful and accessible measure for understanding various individual difference variables (Barlow, 1960; Rodin et al., 1965; Rogers, 1980; Polich, 1997; Ramos-Loyo et al., 2004; Lee et al., 2011; Nash et al., 2012; Garrett et al., 2013; Schiller et al., 2014). Therefore, the distribution defined by the magnitudes of sEEG is indicative of the possible range of magnitudes of evoked pEEG and it can be used as a baseline distribution to normalize pEEG magnitudes to z-scores.

Applicability of the sEEG-Normalization Based Pain Prediction

Predictive power of pEEG responses for decoding the intensity of subjective pain perception has been well documented in previous studies (Kakigi et al., 1989; Bromm and Treede, 1990; Garcí-Larrea et al., 1997; Iannetti et al., 2005; Huang et al., 2013b), which further led to several cross-individual pain prediction methods, which do not need any subjective pain rating for new individuals and thus more promising for clinical uses.

However, the accuracy of cross-individual pain prediction is still not satisfactory because of the inherent inter-individual variability in either pain evoked responses or pain ratings. A practical solution to this problem is to incorporate individual traits that are related to inter-individual variability into the pain prediction model (Davis, 2011). In our previous study (Huang et al., 2013b), single-trial evoked EEG features were normalized by subtracting the mean and dividing by the SD of the individual's evoked EEG features, and single-trial ratings of pain perception were rescaled within the range from 0 to 10 (defining 0 as the lowest pain rating and 10 as the highest pain rating for each participant). Although above normalization on both evoked EEG features and pain ratings can significantly increase the prediction accuracy, its drawback was obvious. First, the normalization was based on the distribution of evoked EEG features, which can only be obtained from a large number of painful stimuli and may not be accepted by participants. Second, it still needs subjective pain rating from a new individual, which is not suitable for participants with communication impairments. As compared with the conventional normalization strategy (Huang et al., 2013b), the proposed method has two main advantages: first, it will not introduce any pain experience to a new participant because the normalization is based on spontaneous EEG; second, it can well deal with the difficult situation that no reliable pain rating is available because no subjective rating is needed. Therefore, the proposed sEEG-based normalization method is more practical and feasible for clinical research and applications.

Limitation and Future Work

The proposed normalization strategy focused solely on features of pain-evoked EEG responses, simply because real values of pain perception are generally considered to be unknown in clinical scenarios. However, not only EEG responses but also the pain ratings are characterized by tremendous inter-individual variability. Different individuals perceive different pain perception in response to the same painful stimulus. For example, we have found pronounced sex-dependent difference in pain perception as well as in pain-evoked EEG responses (see Supplementary Materials). Taking into account gender difference (such as using sex as a predictor) may lead to a more accurate pain prediction. Mechanisms contributing to inter-individual differences in pain sensitivity include genetic, environmental, psychological, and cognitive factors (Nielsen et al., 2009; Coghill, 2010; Schulz et al., 2012), and it may be caused at any stage in pain processing from the skin to the brain. Highly sensitive individuals may activate stronger neural responses and/or pain experience than insensitive individuals (Coghill et al., 2003; Coghill, 2010). Variations in pain sensitivity is an important issue worthy of further investigation, because understanding the contributing factors of pain sensitivity will help greatly in developing a more accurate and practical method for diagnosis of pain (Edwards, 2005; Nielsen et al., 2009). Our future study is aimed to address above difficult problems, such as how to normalize pain ratings and pain sensitivity and how to incorporate personal traits and environmental factors in the prediction model, to develop a more accurate and practical EEG-based prediction assessment method.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

REFERENCES

Anokhin, A. P., Müller, V., Lindenberger, U., Heath, A. C., and Myers, E. (2006). Genetic influences on dynamic complexity of brain oscillations. *Neurosci. Lett.* 397, 93–98. doi: 10.1016/j.neulet.2005.12.025

Arendt-Nielsen, L. (1994). Characteristics, detection, and modulation of laser-evoked vertex potentials. *Acta Anaesthesiol. Scand. Suppl.* 101, 7–44. doi: 10.1111/j.1399-6576.1994.tb04027.x

Barlow, J. S. (1960). Rhythmic activity induced by photic stimulation in relation to intrinsic alpha activity of the brain in man. *Electroencephalogr. Clin. Neurophysiol.* 12, 317–326. doi: 10.1016/0013-4694(60)90005-5

Beydoun, A., Morrow, T. J., Shen, J. F., and Casey, K. L. (1993). Variability of laser-evoked potentials: attention, arousal and lateralized differences. *Electroencephalogr. Clin. Neurophysiol.* 88, 173–181. doi: 10.1016/0168-5597(93)90002-7

Brodersen, K. H., Wiech, K., Lomakina, E. I., Lin, C.-S., Buhmann, J. M., Bingel, U., et al. (2012). Decoding the perception of pain from fMRI using multivariate pattern analysis. *Neuroimage* 63, 1162–1170. doi: 10.1016/j.neuroimage.2012.08.035

Bromm, B., and Treede, R. (1983). Nerve fibre discharges, cerebral potentials and sensations induced by CO2 laser stimulation. *Hum. Neurobiol.* 3, 33–40.

Bromm, B., and Treede, R. (1990). Laser-evoked cerebral potentials in the assessment of cutaneous pain sensitivity in normal subjects and patients. *Rev. Neurol. (Paris).* 147, 625–643.

Brown, J. E., Chatterjee, N., Younger, J., and Mackey, S. (2011). Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. *PLoS ONE* 6:e24124. doi: 10.1371/journal.pone.0024124

Coghill, R. C. (2010). Individual differences in the subjective experience of pain: new insights into mechanisms and models. *Headache* 50, 1531–1535. doi: 10.1111/j.1526-4610.2010.01763.x

Coghill, R. C., Mchaffie, J. G., and Yen, Y.-F. (2003). Neural correlates of interindividual differences in the subjective experience of pain. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8538–8542. doi: 10.1073/pnas.1430684100

Cruccu, G., Sommer, C., Anand, P., Attal, N., Baron, R., Garcia-Larrea, L., et al. (2010). EFNS guidelines on neuropathic pain assessment: revised 2009. *Eur. J. Neurol.* 17, 1010–1018. doi: 10.1111/j.1468-1331.2010.02969.x

Davis, K. D. (2011). Neuroimaging of pain: what does it tell us? *Curr. Opin. Support. Palliat. Care* 5, 116–121. doi: 10.1097/SPC.0b013e3283458f96

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Doležal, J., Šťastný, J., and Sovka, P. (2005). "Recognition of direction of finger movement from EEG signal using markov models," in *IFMBE Proceedings* (Prague), 1727–1983.

Edwards, R. R. (2005). Individual differences in endogenous pain modulation as a risk factor for chronic pain. *Neurology* 65, 437–443. doi: 10.1212/01.wnl.0000171862.17301.84

Garcí-Larrea, L., Peyron, R., Laurent, B., and Mauguière, F. (1997). Association and dissociation between laser-evoked potentials and pain perception. *Neuroreport* 8, 3785–3789. doi: 10.1097/00001756-199712010-00026

Garrett, D. D., Samanez-Larkin, G. R., MacDonald, S. W. S., Lindenberger, U., Mcintosh, A. R., and Grady, C. L. (2013). Moment-to-moment brain signal variability: a next frontier in human brain mapping? *Neurosci. Biobehav. Rev.* 37, 610–624. doi: 10.1016/j.neubiorev.2013.02.015

Haanpää, M., Attal, N., Backonja, M., Baron, R., Bennett, M., Bouhassira, D., et al. (2011). NeuPSIG guidelines on neuropathic pain assessment. *PAIN®* 152, 14–27. doi: 10.1016/j.pain.2010.07.031

Huang, G., Xiao, P., Hu, L., Hung, Y. S., and Zhang, Z. (2013a). "Single-trial laser-evoked potentials feature extraction for prediction of pain perception," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE* (Osaka: IEEE), 4207–4210. doi: 10.1109/embc.2013.6610473

Huang, G., Xiao, P., Hung, Y. S., Iannetti, G. D., Zhang, Z. G., and Hu, L. (2013b). A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. *Neuroimage* 81, 283–293. doi: 10.1016/j.neuroimage.2013.05.017

Iannetti, G., Truini, A., Romaniello, A., Galeotti, F., Rizzo, C., Manfredi, M., et al. (2003). Evidence of a specific spinal pathway for the sense of warmth in humans. *J. Neurophysiol.* 89, 562–570. doi: 10.1152/jn.00393.2002

Iannetti, G., Zambreanu, L., Cruccu, G., and Tracey, I. (2005). Operculoinsular cortex encodes pain intensity at the earliest stages of cortical processing as indicated by amplitude of laser-evoked potentials in humans. *Neuroscience* 131, 199–208. doi: 10.1016/j.neuroscience.2004.10.035

Kakigi, R., Shibasaki, H., and Ikeda, A. (1989). Pain-related somatosensory evoked potentials following CO 2 laser stimulation in man. *Electroencephalogr. Clin. Neurophysiol.* 74, 139–146. doi: 10.1016/0168-5597(89)90019-1

Klistorner, A. I., and Graham, S. L. (2001). Electroencephalogram-based scaling of multifocal visual evoked potentials: effect on intersubject amplitude variability. *Invest. Ophthalmol. Vis. Sci.* 42, 2145–2152.

Lee, T.-W., Younger, W., Wu, H.-C., and Chen, T.-J. (2011). Do resting brain dynamics predict oddball evoked-potential? *BMC Neurosci.* 12:121. doi: 10.1186/1471-2202-12-121

Loeser, J. D., and Treede, R.-D. (2008). The Kyoto protocol of IASP basic pain terminology. *Pain* 137, 473–477. doi: 10.1016/j.pain.2008.04.025

Marcel, S., and Del Millan, J. R. (2007). Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *Pattern Anal. Mach. Intell. IEEE Trans.* 29, 743–752. doi: 10.1109/TPAMI.2007.1012

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., and Mourão-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *Neuroimage* 49, 2178–2189. doi: 10.1016/j.neuroimage.2009.10.072

Näpflin, M., Wildi, M., and Sarnthein, J. (2007). Test–retest reliability of resting EEG spectra validates a statistical signature of persons. *Clin. Neurophysiol.* 118, 2519–2524. doi: 10.1016/j.clinph.2007.07.022

Nash, K., Inzlicht, M., and Mcgregor, I. (2012). Approach-related left prefrontal EEG asymmetry predicts muted error-related negativity. *Biol. Psychol.* 91, 96–102. doi: 10.1016/j.biopsycho.2012.05.005

Nielsen, C. S., Staud, R., and Price, D. D. (2009). Individual differences in pain sensitivity: measurement, causation, and consequences. *J. Pain* 10, 231–237. doi: 10.1016/j.jpain.2008.09.010

Polich, J. (1997). On the relationship between EEG and P300: individual differences, aging, and ultradian rhythms. *Int. J. Psychophysiol.* 26, 299–317. doi: 10.1016/S0167-8760(97)00772-1

Ramos-Loyo, J., Gonzalez-Garrido, A. A., Amezcua, C., and Guevara, M. A. (2004). Relationship between resting alpha activity and the ERPs obtained during a highly demanding selective attention task. *Int. J. Psychophysiol.* 54, 251–262. doi: 10.1016/j.ijpsycho.2004.05.008

Rodin, E. A., Grisell, J. L., Gudobba, R., and Zachary, G. (1965). Relationship of EEG background rhythms to photic evoked responses. *Electroencephalogr. Clin. Neurophysiol.* 19, 301–304. doi: 10.1016/0013-4694(65)90211-7

Rogers, L. (1980). Ipsilateral and contralateral correlations between EEG and EP principal components. *Electroencephalogr. Clin. Neurophysiol.* 50, 441–448. doi: 10.1016/0013-4694(80)90010-3

Schiller, B., Gianotti, L. R., Nash, K., and Knoch, D. (2014). Individual differences in inhibitory control—Relationship between baseline activation in lateral PFC and an electrophysiological index of response inhibition. *Cereb. Cortex* 24, 2430–2435. doi: 10.1093/cercor/bht095

Schnakers, C., and Zasler, N. D. (2007). Pain assessment and management in disorders of consciousness. *Curr. Opin. Neurol.* 20, 620–626. doi: 10.1097/WCO.0b013e3282f169d9

Schulz, E., Zherdin, A., Tiemann, L., Plant, C., and Ploner, M. (2012). Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. *Cereb. Cortex* 22, 1118–1123. doi: 10.1093/cercor/bhr186

Terhaar, J., Viola, F. C., Franz, M., Berger, S., Bär, K.-J., and Weiss, T. (2011). Differential processing of laser stimuli by Aδ and C fibres in major depression. *PAIN®* 152, 1796–1802. doi: 10.1016/j.pain.2011.03.027

Tran, Y., Craig, A., and Mcisaac, P. (2001). Extraversion–introversion and 8–13 Hz waves in frontal cortical regions. *Pers. Individ. Dif.* 30, 205–215. doi: 10.1016/S0191-8869(00)00027-1

Treede, R.-D., Lorenz, J., and Baumgärtner, U. (2003). Clinical usefulness of laser-evoked potentials. *Neurophysiol. Clin. Neurophysiol.* 33, 303–314. doi: 10.1016/j.neucli.2003.10.009

You, Y., Thie, J., Klistorner, A., Gupta, V. K., and Graham, S. L. (2012). Normalization of visual evoked potentials using underlying electroencephalogram levels improves amplitude reproducibility in rats. *Invest. Ophthalmol. Vis. Sci.* 53, 1473–1478. doi: 10.1167/iovs.11-8797

Zhang, Z. G., Hu, L., Hung, Y. S., Mouraux, A., and Iannetti, G. (2012). Gamma-band oscillations in the primary somatosensory cortex—a direct and obligatory correlate of subjective pain intensity. *J. Neurosci* 32, 7429–7438. doi: 10.1523/JNEUROSCI.5877-11.2012

Zietsch, B. P., Hansen, J. L., Hansell, N. K., Geffen, G. M., Martin, N. G., and Wright, M. J. (2007). Common and specific genetic influences on EEG power bands delta, theta, alpha, and beta. *Biol. Psychol.* 75, 154–164. doi: 10.1016/j.biopsycho.2007.01.004

Decoding Subjective Intensity of Nociceptive Pain from Pre-stimulus and Post-stimulus Brain Activities

Yiheng Tu[1,2†], Ao Tan[1,2†], Yanru Bai[1,3], Yeung Sam Hung[2] and Zhiguo Zhang[1]*

[1] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China, [2] Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China, [3] School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore, Singapore

Pain is a highly subjective experience. Self-report is the gold standard for pain assessment in clinical practice, but it may not be available or reliable in some populations. Neuroimaging data, such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), have the potential to be used to provide physiology-based and quantitative nociceptive pain assessment tools that complements self-report. However, existing neuroimaging-based nociceptive pain assessments only rely on the information in pain-evoked brain activities, but neglect the fact that the perceived intensity of pain is also encoded by ongoing brain activities prior to painful stimulation. Here, we proposed to use machine learning algorithms to decode pain intensity from both pre-stimulus ongoing and post-stimulus evoked brain activities. Neural features that were correlated with intensity of laser-evoked nociceptive pain were extracted from high-dimensional pre- and post-stimulus EEG and fMRI activities using partial least-squares regression (PLSR). Further, we used support vector machine (SVM) to predict the intensity of pain from pain-related time-frequency EEG patterns and BOLD-fMRI patterns. Results showed that combining predictive information in pre- and post-stimulus brain activities can achieve significantly better performance in classifying high-pain and low-pain and in predicting the rating of perceived pain than only using post-stimulus brain activities. Therefore, the proposed pain prediction method holds great potential in basic research and clinical applications.

Keywords: pre-stimulus brain activity, EEG, fMRI, pain perception, machine learning, feature selection

INTRODUCTION

Pain assessment is a crucial clinical practice. Inaccurate pain assessment can lead to inadequate pain management, and even misleads diagnosis and treatment (Brown et al., 2011). As a multidimensional and highly subjective experience, pain perception is primarily measured by means of self-report [e.g., Visual Analog Scales (VAS) and Numeric Rating Scales (NRS)] in clinical applications (Cruccu et al., 2008; Haanpää et al., 2011). However, the subjectivity of self-report limits its application to people with impaired consciousness (e.g., patients in a coma, vegetative state or minimal conscious state; Schnakers and Zasler, 2007) or limited cognitive capacity (e.g., young children, the elderly, patients with cognitive impairment; Wong and Baker, 1988; Herr et al., 2004; Buffum et al., 2007), and people who are

unwilling to reliably communicate the feeling of pain. In addition, self-report provides limited understandings of the underlying neurophysiological processes of pain perception, which is important for the development of targeting treatments (Wager et al., 2013).Therefore, developing a neurophysiology-based pain assessment tool is highly necessary in basic pain research and clinical applications.

Non-invasive neuroimaging techniques, such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), enable us to readily obtain brain responses to nociceptive inputs. A variety of neural correlates of nociceptive pain [e.g., laser-evoked EEG potentials (LEPs), fMRI responses within the "pain matrix"] have been identified, allowing for neurophysiology-based pain assessments (Huang et al., 2013; Wager et al., 2013). Assessing subjective intensity of nociceptive pain perception with non-invasive neuroimaging data has gained emerging interest in recent years. For example, based on the strong correlation between the amplitudes/latencies of LEP and subjective pain perception, single-trial LEP features were used to predict pain intensity with high accuracy (Huang et al., 2013). Rapid developments of neuroimaging data analytics also lead to novel and effective algorithms for pain prediction. Machine learning, which can identify brain activation patterns corresponding to pain perception, has also been used in pain prediction due to its high sensitivity to multidimensional neuroimaging patterns (Schulz et al., 2011; Brodersen et al., 2012; Wager et al., 2013).

However, conventional neuroimaging-based pain prediction methods only make use of pain-related information encoded in brain activities evoked by nociceptive stimulation, while completely overlook ongoing brain activities prior to nociceptive stimulation. Actually, pre-stimulus ongoing brain activities contain important information that is predictive of forthcoming perception of pain, because previous studies have convincingly shown that pain perception is largely modulated by ongoing cognitive states (e.g., expectation, attention, reappraisal; Terkelsen et al., 2004; Quevedo and Coghill, 2007; Wiech et al., 2008; Tu et al., 2016). Babiloni et al. (2006) firstly revealed a strong negative correlation between pre-stimulus alpha-band EEG power and subjective pain rating. The relationship between pre-stimulus brain responses and pain perception was also observed in fMRI studies. It was reported that the variability in pain perception under identical stimuli was positively correlated with the fluctuation of baseline blood oxygenation level dependent (BOLD) signal in medial thalamus, lateral fronto-parietal network, and negatively correlated with BOLD in posterior cingulate and temporo-parietal cortices (Boly et al., 2007). The pre-stimulus functional connectivity between anterior insula cortex and brainstem was also found to negatively modulate pain perception (Ploner et al., 2010). Our recent work has introduced how the ongoing fluctuations of intrinsic cortical networks (as reflected by EEG spectrogram and BOLD-fMRI responses) determine the dynamic state of the brain and influence the pain perception (Tu et al., 2016). These findings suggested that pre- and post-stimulus brain activities provide complementary information for pain encoding: post-stimulus brain activities reflect the nociceptive information while

pre-stimulus brain activities are responsible for trial-to-trial variability in baseline cognitive and emotional states.

In the present work, we hypothesize that combining the information embedded in pre- and post-stimulus brain activities can lead to a more accurate prediction of nociceptive pain perception. To validate this hypothesis, we collected EEG and fMRI data in laser-evoked pain experiments and used machine learning algorithms to decode perceived pain intensity from both pre-stimulus ongoing and post-stimulus evoked brain activities. Temporal-spectral EEG spectrogram and BOLD-fMRI magnitudes (in both pre- and post-stimulus periods) comprise two high-dimensional feature sets used for pain decoding. A popular supervised machine learning method, partial least square regression (PLSR; Hu et al., 2014), was used to reduce the dimensionality of EEG or fMRI feature sets by detecting a subset of features that are closely correlated with pain. These features form a number of pre- and post-stimulus pain-related brain patterns (temporal-spectral patterns for EEG and spatial patterns for fMRI). Support vector machine (SVM; Cortes and Vapnik, 1995) was then used to decode the intensity of pain perception from the pre- and post-stimulus EEG or fMRI brain patterns. In both EEG and fMRI datasets, the proposed pain decoding method using both pre- and post-stimulus activities achieved higher prediction performance than conventional pain decoding methods using post-stimulus information only. In addition, the predictive power of pre- and post-stimulus brain patterns for pain decoding was individually assessed and ranked, helped to build a more concise prediction model and provided an understanding of to what extent the extracted pain-related patterns contribute to pain perception.

MATERIALS AND METHODS

In the present work, we proposed to decode the intensity of perceived pain from both pre- and post-stimulus brain activities (sampled by EEG or fMRI) in laser-evoked pain experiments.

Experiments
EEG Experiments
EEG data were collected from 96 healthy participants (51 females) aged 21.6 ± 1.7 years (mean ± SD). All participants gave their written informed consent and the experimental procedures were approved by the local ethics committee. Details of experimental design and recordings have been published previously (Hu et al., 2014).

In brief, nociceptive-specific radiant-heat stimuli were generated by laser and a total of 40 pulses, 10 for each of the four stimulus energies (E1: 2.5 J; E2: 3.0 J; E3: 3.5 J; E4, 4.0 J), were delivered in a pseudorandom order. The inter-stimulus interval varied between 10 and 15 s. After each stimulus, subjects were instructed to rate the intensity of the painful sensation elicited by the laser pulse, using a visual analog scale (VAS) ranging from 0 to 10 (0 corresponds to "no pain," "<5" corresponds to "heat pain," "≥5" corresponds to "acute pain," and "10" corresponds to "pain as bad as it could be"; Jensen and Karoly, 1992). EEG data were continuously recorded using 64 Ag-AgCl scalp electrodes placed according to the International 10–20 system (Brain

Products GmbH; Munich, Germany; pass-band: 0.01–100 Hz; sampling rate: 1000 Hz), using the nose as reference. Electrode impedances were kept below 10 kΩ. Electro-oculographic (EOG) signals were simultaneously recorded using surface electrodes to monitor ocular movements and eye blinks.

Functional MRI Experiments

Functional MRI data were collected from 32 healthy participants (20 females) aged 22.1 $\pm$ 2.0 years (mean $\pm$ SD). All participants gave their written informed consent and the experimental procedures were approved by the local ethics committee. The current dataset follows a similar experimental design as was adopted in the EEG dataset, with the exception that inter-stimulus interval was longer (varied between 27 and 33 s) due to the low temporal resolution of fMRI recording. Functional MRI data were acquired using a Siemens 3.0 Tesla Trio scanner with a standard head coil. A whole-brain gradient-echo, echo-planar-imaging sequence was used for functional scanning with a repetition time (TR) of 1500 ms (29 ms echo time, 25 5.0 mm-thick slices with 0.5 mm inter-slice gaps, 3 $\times$ 3 mm in-plane resolution, field of view 192 $\times$ 192 mm, matrix 64 $\times$ 64; flip angle 90°). A high-resolution T1-weighted structural image (1 mm^3 isotropic voxel MPRAGE) was acquired after functional imaging.

Methods

The proposed pain decoding pipeline is shown in **Figure 1**. The pipeline consists of 3 steps: (1) pre-processing (not shown in **Figures 1, 2** feature extraction and selection; and (3) pain prediction. Firstly, pre-processing is aimed to remove noise and artifacts from raw EEG and fMRI recordings. Secondly, a subset of pain-related features is selected from high-dimensional neuroimaging data (time-frequency EEG data or whole-brain fMRI data) in both pre- and post-stimulus periods to form discriminative temporal-spectral EEG patterns and spatial fMRI patterns. Thirdly, a prediction model is established to describe the relationship between the level of pain perception and identified EEG or fMRI patterns in both pre- and post-stimulus periods. Two machine learning methods, PLSR and SVM, were used in step 2 and step 3, respectively. Although, EEG and fMRI have different pre-processing steps, they share similar methods in the steps of feature selection and prediction.

Pre-Processing

For EEG data, five subjects were excluded from the dataset since they did not have variable painful sensation in response to different stimulus energies. EEG data were preprocessed using EEGLAB (Delorme and Makeig, 2004) and underwent standard pre-processing. Continuous data were filtered (1–100 Hz) and segmented into epochs (–500 to 0 ms and 0 to 1000 ms for pre- and post-stimulus, respectively) and baseline-corrected using pre-stimulus interval. An infomax independent component analysis (ICA; Delorme and Makeig, 2004) was used to correct trials contaminated by eye blinks and movements.

For fMRI data, two subjects were excluded from the dataset since they did not have variable painful sensation

in response to different stimulus energies. The preprocessing routine was conducted using SPM8 (Wellcome Trust Center for Neuroimaging, London, UK). Images were slice-timing corrected, head motion corrected, normalized to the Montreal Neurological Institute (MNI) space (voxel size = 3$\times$ 3 $\times$ 3 mm) by mapping T1-weighted structural image to MNI template (Ashburner and Friston, 2005), and spatially smoothed using a Gaussian kernel of 8 mm full width at half maximum (FWHM = 8 mm). A high-pass filter was applied (cut-off frequency = 1/128 Hz) to the BOLD time-series to remove low-frequency drifts. BOLD responses were modeled as a series of events using a stick function and ratings were included as a parametric modulator of each stimulus, which were then convolved with a canonical hemodynamic response function (HRF). Group-level statistical analyses were carried out using a random effects analysis with one-sample t-test as implemented in SPM8. Brain regions activated by laser stimuli were illustrated in **Figure 3**.

Feature Extraction and Selection

EEG spectral power in the time-frequency domain and BOLD-fMRI strength are used as features to predict pain levels. For EEG, short-time Fourier transform (STFT) with a fixed 200 ms Hanning window (Zhang et al., 2012) was applied to single-trial data at electrode C4 (LEP has maximal responses at the contralateral site of somatosensory area) (Valentini et al., 2012) to obtain their time-frequency distributions. Pre-stimulus (–500 to 0 ms) and post-stimulus (0 to 1000 ms) EEG spectrograms were extracted as pre- and post-stimulus features (i.e., each feature represents the power at a time-frequency pixel in the spectrogram) for further analysis. For fMRI, the whole-brain scan at stimulus onset containing pre-stimulus brain patterns immediately before stimulus onset (onset scan), and the whole-brain scan corresponding to the maximum BOLD response to nociceptive pain (peak scan, i.e., 4th scan after stimulus onset, **Figure 3B**), were extracted as pre- and post-stimulus features (i.e., each feature represents a voxel at stimulus onset or response peak in the scans) for further analysis. Since in both EEG and fMRI experiments painful stimuli were delivered in 4 different energy levels, both EEG and fMRI features as well as subjective pain ratings were normalized by removing the mean values of ratings within each energy group to minimize the influence of stimulus energy on the assessment of their trial-to-trial relationship.

For each subject, a linear model is used to describe the relationship between the level of pain perception and pain-related neuroimaging features, which include pre-stimulus features (X_m^{pre}, $m = 1, ..., M$, where m denotes the index of pre-stimulus features and M is the total number of pre-stimulus features) and post-stimulus features (X_n^{post}, $n = 1, ..., N$, where n denotes the index of post-stimulus features and N is the total number of post-stimulus features). The linear function linking the reported intensity of pain, Y, and EEG or fMRI features of one trial reads:

$$Y = a_0 + \sum_{m}^{M} a_m^{pre} X_m^{pre} + \sum_{n}^{N} a_n^{post} X_n^{post} + \varepsilon, \qquad (1)$$

FIGURE 1 | Overview of the proposed pain decoding method (pre-processing steps are not shown). (1) *Feature extraction*. Single-trial pre- and post-stimulus EEG spectrogram at C4 electrode (**left**) and whole-brain fMRI BOLD signals of onset and peak scans (**right**) were extracted from preprocessed data as features. (2) *Feature selection*. Extracted features and perceived intensity of pain were fitted in a PLSR model. Features that were significantly predictive of pain perception were identified from group analysis and form a set of patterns (temporal-spectral EEG patetrns and spatial fMRI patterns). (3) *Prediction*. Selected brain patterns were fed to SVC and SVR models for pain classification and regression.

where a_m^{pre} and a_n^{post}, respectively, denote the model coefficients for X_m^{pre} and X_n^{post}, a_0 denotes the intercept, and ε denotes the model residual.

Following, a subset of features that are most predictive of pain perception was selected. Here, the predictive power of

each feature was defined according to its corresponding model coefficient (a_m^{pre} or a_n^{post}) in Equation (1). That is, those features with a corresponding model coefficient significantly different from 0 across subjects were regarded as regions with predictive power. To achieve this, PLSR [implemented by Nonlinear

FIGURE 2 | (A) Identified pain related EEG time-frequency patterns. PLSR coefficients indicating the relationship between EEG spectrogram and the intensity of pain. Five time-frequency clusters (C4-nose) significantly modulated the intensity of pain. **(B)** Group-level average of EEG spectrogram (C4-nose). Time-frequency power within corresponding five time-frequency clusters were selected for pain prediction. **(C)** The performance to predict subjective intensity of pain based on "Pre+Post" and "Post" EEG features. Classification accuracies to discriminate low and high pain trials were 83.5 ± 6.8% ("Pre+Post") and 78.2 ± 9.1% ("Post"), respectively ($p <$ 0.0001, paired t-test). Significant difference was observed in classification sensitivity ($p <$ 0.05, paired t-test), but not in specificity. Prediction errors to continuous predict pain intensity were 1.15 ± 0.32 ("Pre+Post") and 1.27 ± 0.38 ("Post"), respectively ($p <$ 0.0001, paired t-test). Error bars represent SD across subjects. **(D)** The performance to predict subjective intensity of pain based on individual EEG patterns. "LEP" provided strongest and most significant prediction performance.

Iterative Partial Least Squares algorithm (NIPALS); Wold et al., 2001] was applied to estimate the model coefficients in Equation (1). PLSR was applied here as it can solve the problems of high dimensionality and multicollinearity, which are typical in neuroimaging data. Statistical significance of the estimated model coefficients across subjects were assessed with a point-by-point one-sample t-test against zero, combined with nonparametric permutation testing (see PLSR analysis in Tu et al., 2016 for details of this method). The statistical result defines a number of pain-related patterns (temporal-spectral patterns for EEG and spatial patterns for fMRI), which consist of features that are most predictive of pain intensity and share similar temporal-spectral characteristics (for EEG) or spatial characteristics (for fMRI) across subjects. More precisely, for EEG data, these patterns are neighboring time-frequency pixels having power values that are significantly correlated with pain perception, while for fMRI data, these patterns are neighboring voxels having BOLD strengths that are significantly correlated with pain perception.

Pain Prediction
In this step, pain prediction models were trained to decode single-trial intensity of pain perception from identified pain-related patterns (time-frequency patterns for EEG and spatial patterns for fMRI). Two types of pain decoding models were trained in the current study: (1) classification, which qualitatively predicts the intensity of pain by classifying trials into two levels

(low pain: VAS < 5; high pain: VAS ≥ 5); and (2) regression, which quantitatively predicts the intensity of pain as a continuous value (0–10). Linear support vector classification (SVC) and support vector regression (SVR) model were, respectively, adopted as the classification and regression model (Pereira et al., 2009).

A leave-one-out-cross-validation (LOOCV) strategy was adopted to evaluate the performance of the pain decoding model (SVC and SVR) for each subject (Huang et al., 2013). For each iteration in LOOCV, one trial was selected as the test sample and fed to the SVC/SVR model trained with remaining samples, and the iterations were repeated for every trial. To quantify the performance of SVC, classification accuracy, sensitivity, and specificity were calculated. Sensitivity and specificity are defined as:

$$Sensitivity = \frac{\sum TP}{\sum TP + \sum FN}, Specificity = \frac{\sum TN}{\sum TN + \sum FP}, (2)$$

where $\sum TP$ and $\sum FN$ denote the number of true positive and false negative respectively, $\sum TN$ and $\sum FP$ denote the number of true negative and false positive respectively. Here positive is defined as high pain trials and negative is defined as low pain trials.

To quantify the performance of SVR, we used mean absolute error (MAE), which is defined as:

$$MAE = \frac{1}{P} \sum_{p=1}^{P} \left| R_p - \hat{R}_p \right|, \tag{3}$$

where R_p and $\hat{R}_p$ denote, respectively, the actual and predicted intensity of pain perception for trial p, and P is the number of trials of each subject. The above steps were repeated for each subject, and the performance measures were assessed at group level (e.g., whether SVC yielded significantly above-chance classification accuracy).

Since one of our focuses is to investigate whether the combination of pain-related information in both pre- and post-stimulus information can improve pain decoding performance, we further compared the prediction performance using two sets of patterns: (1) both pre- and post-stimulus pain-related patterns (the proposed method); (2) post-stimulus pain-related patterns only (the conventional method). Last, we also evaluated the individual contribution of each pain-related brain pattern to pain decoding, yielding a ranked contribution of these pain-related patterns.

## RESULTS

EEG Results

Psychophysics
Ninety-six subjects overall had an average subjective pain intensity of 5.74 ± 1.03 (mean $\pm$ SD). Five subjects were excluded for the following analyses since they did not have variable sensation in response to different stimulus energies. Nociceptive-specific laser stimuli of four energies (E1–E4) elicited clear pinprick sensation in the remaining 91 subjects (E1: 3.81 ± 1.41; E2: 4.86 ± 1.34; E3: 6.63 ± 1.07; E4: 7.75 ± 1.03).

Pain-Related Time-Frequency Patterns
Five time-frequency clusters were identified to significantly modulate the perceived pain intensity (**Figures 2A,B**). In the pre-stimulus interval, a cluster in the alpha band ("Pre-ABO": –221–31 ms, 8–15 Hz; $p < 0.001$) and a cluster in gamma band ("Pre-GBO": –180–85 ms, 74–87 Hz; $p = 0.001$) negatively modulated the perceived intensity of a subsequent stimulus. In the post-stimulus interval, three significant clusters were observed: the low-frequency "LEP" (74–470 ms, 1–22 Hz; $p < 0.001$), the low-frequency ABO ("Post-ABO": 637–935 ms, 8–20 Hz; $p < 0.001$), and the high-frequency GBO ("Post-GBO": 127–377 ms, 62–100 Hz; $p < 0.001$). It was confirmed that the magnitude of "LEP" and "Post-GBO" positively correlated with perceived intensity of pain, while the magnitude of "Post-ABO" negatively correlated with perceived intensity of pain.

Predicting Pain from Pre- and Post-stimulus Time-Frequency Patterns
We trained and tested a linear SVM on two different sets of patterns: (1) post-stimulus time-frequency patterns ("Post") including "LEP," "Post-ABO," and "Post-GBO"; (2) both pre- and post-stimulus time-frequency patterns ("Pre+Post")
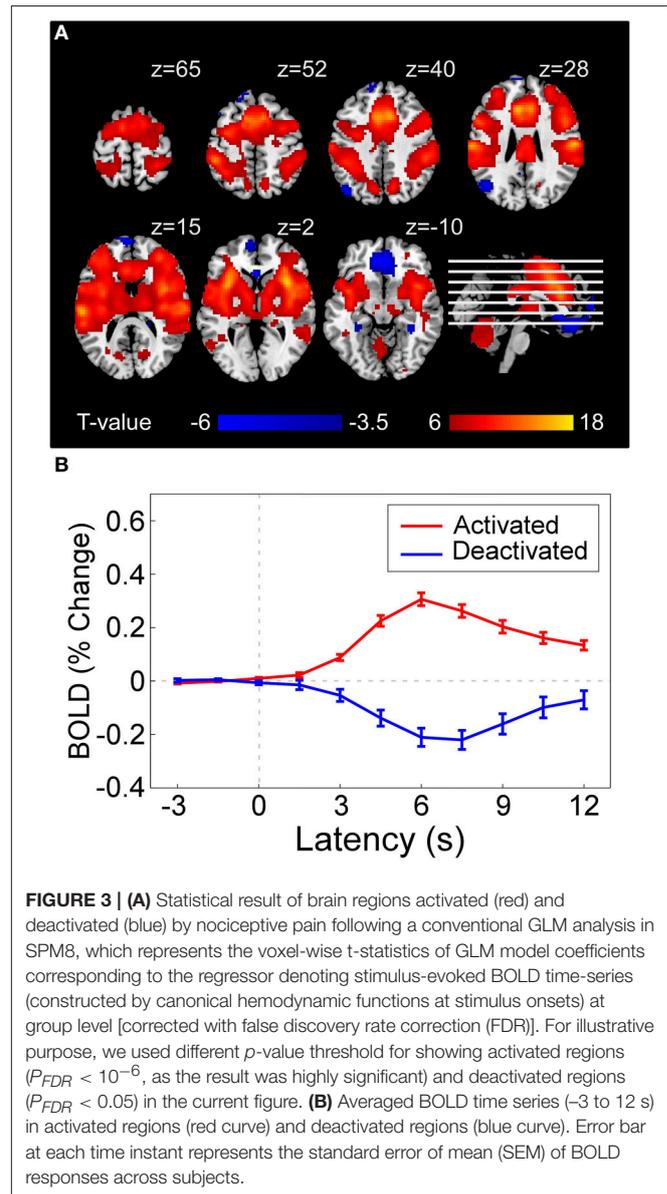
FIGURE 3 | (A) Statistical result of brain regions activated (red) and deactivated (blue) by nociceptive pain following a conventional GLM analysis in SPM8, which represents the voxel-wise t-statistics of GLM model coefficients corresponding to the regressor denoting stimulus-evoked BOLD time-series (constructed by canonical hemodynamic functions at stimulus onsets) at group level [corrected with false discovery rate correction (FDR)]. For illustrative purpose, we used different p-value threshold for showing activated regions ($P_{FDR} < 10^{-6}$, as the result was highly significant) and deactivated regions ($P_{FDR} < 0.05$) in the current figure. **(B)** Averaged BOLD time series (–3 to 12 s) in activated regions (red curve) and deactivated regions (blue curve). Error bar at each time instant represents the standard error of mean (SEM) of BOLD responses across subjects.

including "Pre-ABO," "Pre-GBO," "LEP," "Post-ABO," and "Post-GBO" (**Figure 2C**). For classification accuracy (mean $\pm$ SD), "Pre+Post" provided significantly higher accuracy than "Post" ("Pre+Post": $83.5 \pm 6.8\%$; "Post": $78.2 \pm 9.1\%$; $p < 0.0001$, paired t-test). Significant difference was observed in classification sensitivity ("Pre+Post": $79.2 \pm 14.6\%$; "Post": $77.0 \pm 17.3\%$; $p = 0.04$, paired t-test), but no significant difference in specificity was observed ("Pre+Post": $72.2 \pm 14.2\%$; "Post": $72.0 \pm 17.3\%$; $p = 0.91$). For prediction error (mean $\pm$ SD), "Pre+Post" also provided significantly lower MAE than "Post" ("Pre+Post": 1.15 ± 0.32; "Post": 1.27 ± 0.38; $p < 0.0001$, paired t-test).

Predicting Pain from Individual Time-Frequency Patterns
We attempted to predict the perception of pain from individual time-frequency EEG patterns, with the aim to rank their

respective predictive power. Therefore, results could offer us an understanding of to what extent the extracted patterns contribute to pain prediction. It should be noticed that pre-stimulus brain patterns can only contribute to the fluctuation of perceived pain perception within identical stimuli, as the brain could not forecast the energy of forthcoming nociceptive stimuli which were randomized across trials. Therefore, we used "Pre-ABO" and "Pre-GBO" to predict the normalized intensity of pain perception, while "LEP," "Post-ABO," and "Post-GBO" to predict the perceived intensity of pain perception. Thus, we could propose a rank order of pain-related brain patterns (**Figure 2D**).

All five time-frequency patterns obtained significant prediction results. The most predictive pattern in the pre-stimulus period was "Pre-ABO" in terms of classification accuracy and prediction error ($55.3 \pm 6.6\%$ and 1.79 ± 0.56) and "LEP" was the most predictive pattern in the post-stimulus period ($77.0 \pm 9.3\%$ and 1.30 ± 0.36). Other patterns also provided above-chance performance ("Pre-GBO": $55.0 \pm 7.6\%$ and 1.86 ± 0.55; "Post-ABO": $57.5 \pm 6.9\%$ and 1.72 ± 0.51; "Post-GBO": $56.9 \pm 4.6\%$ and 1.75 ± 0.55).

Functional MRI Results
Psychophysics
Thirty-two subjects overall had an average subjective pain intensity of 4.82 ± 1.55 (mean $\pm$ SD). Nociceptive-specific laser stimuli of four energies (E1–E4) elicited clear pinprick sensation in 30 subjects (E1: 2.92 ± 1.53; E2: 3.84 ± 1.69; E3: 5.68 ± 1.62; E4: 6.91 ± 1.54). Two subjects were excluded for the following analyses since they did not have variable sensation in response to different stimulus energies.

Laser-Evoked BOLD Responses
Single-subject fMRI data were analyzed on a voxel-by-voxel basis, using a general linear model (GLM) approach (Frackowiak et al., 2004), to assess the laser-evoked BOLD activations/deactivations. **Figure 3A** shows that laser stimuli elicited activations within various brain regions, including anterior/middle cingulate cortex (ACC and MCC), supplementary motor area (SMA), primary, and secondary somatosensory cortex (S1 and S2), insula (INS), and thalamus, while deactivations in rectus and DLPFC. Group-level BOLD responses in positive and negative activated regions were illustrated in **Figure 3B**. The peak response of positive activation was located around 6 s (4th scan) after stimulus, while was located around 7.5 s (5th scan) for deactivation.

Pain-Related fMRI Patterns
Post-stimulus evoked BOLD responses in several brain regions showed the capability of significantly modulating the pain perception (**Figure 4A**). These regions include INS, ACC, MCC, S1, SMA, and S2 in the "pain matrix" (Legrain et al., 2011), which can positively modulate pain perception, rectus in default mode network (DMN), and dorsal lateral prefrontal cortex (DLPFC), which can negatively modulate pain perception. Because of the intrinsic delay of the hemodynamic response, the fMRI signal sampled at stimulus onset reflects the brain activity preceding the arrival of the sensory input to the nervous system. At the stimulus

onset time, we found a positive correlation between subsequent normalized pain intensity and BOLD in S1, DLPFC, MCC, SMA, and ACC, and a negative correlation in angular, amygdala and precuneus (**Figure 4A**).

Predicting Pain from Pre- and Post-Stimulus fMRI Patterns
Similar to EEG analysis, we trained and tested linear SVM on two sets of patterns: (1) post-stimulus fMRI patterns ("Post") including identified patterns at the peak scan; (2) both pre- and post-stimulus fMRI patterns ("Pre+Post") including identified patterns at both onset and peak scans. For classification accuracy, "Pre+Post" provided significantly higher accuracy than "Post" ("Pre+Post": $75.0 \pm 10.5\%$; "Post": $72.5 \pm 11.0\%$; $p = 0.0018$, paired t-test). No significant difference was observed in classification sensitivity ("Pre+Post": $63.1 \pm 31.7\%$; "Post": $58.9 \pm 35.0\%$; $p = 0.12$, paired t-test) and specificity ("Pre+Post": $57.1 \pm 37.0\%$; "Post": $54.4 \pm 39.5\%$; $p = 0.24$, paired t-test). For prediction error, "Pre+Post" also provided significantly lower error than "Post" ("Pre+Post": 1.66 ± 0.47; "Post": 1.76 ± 0.47; $p < 0.0035$, paired t-test) (**Figure 4B**).

## Predicting Pain from Individual fMRI Patterns
We further predicted the perception of pain from individual fMRI spatial patterns. Similarly, pre-stimulus fMRI patterns were assessed to predict the normalized intensity of pain while post-stimulus fMRI patterns were assessed to predict the perceived intensity of pain (please refer to "*Predicting Pain from Individual Time-frequency Patterns*" for the reason). All regions achieved significant above-chance prediction accuracy ($p < 0.05$). The most predictive pre-stimulus fMRI patterns were S1 ($56.0 \pm 8.4\%$ and 1.95 ± 0.34) (in terms of accuracy) and DLPFC ($55.3 \pm 7.7\%$ and 1.92 ± 0.39) (in terms of prediction error), while the most predictive post-stimulus fMRI pattern was insula ($72.4 \pm 10.6\%$ and MAE: 1.89 ± 0.62) (**Figure 4C**).

DISCUSSION

In the present work, we proposed a novel pain decoding method which uses both post-stimulus evoked brain activity and pre-stimulus brain activity as features to enhance the prediction performance compared to conventional methods based on post-stimulus evoked brain activities only. Our analysis led to two main findings.

First and foremost, our results demonstrated that by further incorporating pain-related information in pre-stimulus brain activities into the conventional pain prediction model solely based on post-stimulus evoked brain activities, the prediction performance can be significantly improved. The present work highlights the significance of pre-stimulus brain activities in encoding pain perception in the brain, and indicates the bias between actual pain perception and predicted pain perception may also be contributed from pre-stimulus brain activities.

Second, the individual predictive power of pain-related neural features is investigated and ranked, which offers us a better understanding of the predictive capacity of pain-associated

FIGURE 4 | (A) Post-stimulus and pre-stimulus fMRI patterns defined by PLSR coefficients. Positively and negatively predictive patterns are shown in red and blue, respectively (*p* < 0.05, cluster-level permutation test). **(B)** The performance to predict subjective intensity of pain based on "Pre+Post" and "Post" fMRI patterns. Classification accuracies to discriminate low and high pain trials were 75.0 ± 10.5% ("Pre+Post") and 72.5 ± 11.0% ("Post"), respectively (*p* = 0.0018, paired *t*-test). No significant difference was observed in classification sensitivity and specificity. Prediction errors (quantified as MAE) to continuously predict pain intensity were 1.66 ± 0.47 ("Pre+Post") and 1.76 ± 0.47 ("Post"), respectively (*p* < 0.0035, paired *t*-test). Error bars represent SEM across subjects. **(C)** Ranked contribution of pain-related patterns. Error bars represent SEM across subjects. ACC, Anterior Cingulate Cortex; AMYG, Amygdala; DLPFC, Dorsal Lateral Prefrontal Cortex; INS, Insula; MCC, Middle Cingulate Cortex; PREC, Precuneus; SMA, Supplementary Motor Area; S1, Primary Somatosensory Cortex; S2, Secondary Somatosensory Cortex.

brain patterns. The combined predictive power of these neural features is also obtained. Although, most of identified neural features provided above-chance prediction individually, they could not be able to yield higher predictive power when being used with other features, which implies that these regions may not provide completely independent and complimentary pain-related information.

Significance of Pre-stimulus Brain Activities in Pain Decoding

Conventional pain prediction approaches only rely on the relationship between post-stimulus evoked brain activity and pain perception, but they seldom consider the predictive power of pre-stimulus ongoing activities, which have been shown to be correlated with pain (Brodersen et al., 2012; Huang et al., 2013; Wager et al., 2013). In the present work, we demonstrated that, a prediction model, which describes the joint contribution of post-stimulus evoked brain activities and pre-stimulus ongoing brain activities to pain, can provide significantly higher prediction performance.

Actually, pre-stimulus brain oscillations have been repeatedly shown to be predictive of forthcoming sensory perception and they play an important role in the brain mechanisms underlying perception (Linkenkaer-Hansen et al., 2004; Hanslmayr et al., 2007; Van Dijk et al., 2008; Zhang and Ding, 2010; Lange et al., 2012; De Lange et al., 2013). It has been reported that

the fluctuation of ongoing brain activities is able to capture the ongoing brain state and reflect various cognitive terms such as vigilance, attention, and expectation (Buzsaki, 2009). Such ongoing variation in brain state, as captured by ongoing brain activities, has been shown to be able to bias various sensory perceptions. As for pain perception, literature has shown that pain does not only reflect the neural processing of nociceptive information, but is also influenced by various psychosocial contexts and psycho-physiological factors (i.e., brain states; Wiech et al., 2008). In our previous work (Tu et al., 2016), we reported that pre-stimulus alpha and gamma oscillations sampled by EEG and BOLD activities in sensorimotor resting state network and DMN were implicated in top-down modulation of pain, and consequently modulated the perception of subsequent painful stimuli. These findings advanced our understanding of the neural mechanisms of pain, and inspired us to further utilize pre-stimulus information for pain decoding.

Predictive Neural Patterns

For EEG data, when decoding pain perception from pre-stimulus activity, both "Pre-ABO" and "Pre-GBO" afforded significant accuracies. "Pre-ABO" has been interpreted as a measure of altered excitability of neuronal ensembles in primary sensory cortex, while "Pre-GBO" modulates long-range communication between distributed neuronal assemblies. Thereby they offered

complementary information in terms of classification accuracy. When decoding pain-related information embedded in post-stimulus EEG activities, "LEP" had the highest accuracies, indicating its strongest contribution to pain prediction [in terms of mean accuracy (%) and variability (t-value)]. Not surprisingly, no significant difference of prediction accuracy was observed between "LEP" and "Post" ("LEP": 77.0 ± 9.3% and 1.30 ± 0.36; "Post": 78.2 ± 9.1% and 1.27 ± 0.38; $p = 0.72$ and $p = 0.54$ for classification and regression, respectively). Although "Post-ABO" and "Post-GBO" enabled significant accuracies when predicting pain perception individually, they did not provide additional information when being considered along with "LEP." It may due to the predictive information provided by "Post-ABO" and "Post-GBO" is also contained in "LEP." Therefore, it is possible to remove "Post-ABO" and "Post-GBO" and develop a more concise pain prediction model in clinical practice.

For fMRI data, we found the activities at DLPFC are predictive of pain perception, no matter whether they are measured before or after stimulus onset. But the degrees of importance of pre-stimulus DLPFC activities and post-stimulus DLPFC activities are largely different. Pre-stimulus DLPFC activities provide the highest prediction performance among all pre-stimulus patterns (measured by MAE), showing DLPFC is one of the most important regions executing cognitive pain modulation (Wiech et al., 2008). One the other hand, post-stimulus DLPFC activities cannot offer as high prediction performance as "pain matrix" (insula, ACC, MCC, S1, and SMA) does (Legrain et al., 2011), indicating that cognitive modulation is less important after stimulus.

Machine Learning Classifiers for Brain Decoding

Machine learning has gained popularity in the community of brain science and engineering recently for it allows for decoding stimuli, mental states, behaviors, and other variables of interest from neuroimaging data (Pereira et al., 2009). Various machine learning classifiers have been applied to brain decoding, including Logistic Regression (LR; Ryali et al., 2010), linear SVM (Ryali et al., 2010), Gaussian Naïve Bayes (GNB; Huang et al., 2013) and Fisher's Linear Discriminant Analysis (LDA) (Davatzikos et al., 2005). In the present work, we used PLSR to select discriminative features from EEG and fMRI data and applied linear SVC and SVR, both of which are extensions of the classical SVM, to classify and continuously predict subjective pain perception from EEG and fMRI features. PLSR and SVM are both popular machine learning methods and they are gradually used in many applications of brain decoding.

Since there are generally more predictors than experimental trials or subjects, it is often advantageous to reduce the number of predictors by selecting an informative subset. Wager and his colleagues used LASSO-PCR, which is combination of two dimension reduction methods (LASSO and PCA), to extract features and predict pain perception (Wager et al., 2013). LASSO is based on sparsity-enhancing L_1 regularization on regression coefficients and it can shrink small regression coefficients to

zero to realize dimension reduction. But, when dealing with strongly correlated predictors (e.g., adjacent fMRI voxels), LASSO arbitrarily selects one variable from a group of highly-correlated variables, which degrades the interpretability of the prediction model (Cecchi et al., 2012). Therefore, in (Wager et al., 2013), the authors first used PCA to reduce the number of predictors and then used LASSO on the orthogonal principle components (PCs) rather than the original fMRI data. However, since PCA is an unsupervised method and the PCs are obtained according to the variance of the data solely, it cannot guarantee that the classes can be well-separated in the space defined by reduced dimensions. Here we used PLSR to decode pain related brain patterns, because it is a supervised dimension reduction method that can exploit class information to ensure that high-dimensional data can be amped into a low-dimensional space where different classes are well-separated. PLSR is still a linear method but the relationship between pain intensity and brain signals could be non-linear (Wager et al., 2005; Loggia et al., 2012). There are more sophisticated methods that can explore nonlinear relationships between brain responses and behavior variables, which is in accord with intrinsic nonlinear neuro-dynamics of the brain (Tu et al., 2015).

More recently, deep machine learning algorithms which can model the data with multiple processing layers, have been applied for brain decoding and neuroscience discovery (Plis et al., 2014). What differentiates them from other classifiers is the automatic feature learning from data which largely contributes to improvements in accuracy. Deep models such as deep belief networks (DBNs) and restricted Boltzmann machine (RBM), separate linear factors from functional brain imaging data by fitting a probability distribution model to the data, has been used for fMRI classification (Schmah et al., 2008) and for identifying intrinsic networks (Hjelm et al., 2014). It is potentially a suitable solution for pain decoding model, and more advanced feature selection and machine learning techniques will be used to build a more powerful pain decoding model in our future study.

Further Developments for Clinical Uses

In the present study, we proposed a novel pain decoding model incorporating both pre-stimulus brain activities and post-stimulus brain activities, and adopted machine learning classifiers to effectively predict pain perception in single-trials. Such decoding model and prediction strategy could be executed rapidly, reliably, and automatically, thus satisfying most requirements of various basic and clinical applications.

Pre-stimulus brain activities could be a great indicator of subject's ongoing cognitive states and they include much useful information for decoding within-subject variability. However, our decoding model does not take inter-subject variability of pain perception and brain responses into consideration. In future, we would like to apply the decoding model on cross-subject prediction which is more favored because it does not need any training on new individuals.

EEG and fMRI are the commonly used techniques for pain assessment in clinical applications. Particularly, EEG is more

favored because it is cheap and easy to operate. For pain-related clinical study, our proposed novel pain-related brain patterns hold great potential to help diagnose nociceptive system deficit as well as to predict subjective pain perception (e.g., to monitor the effect of analgesic drug or the recovery of nociceptive system for non-communicative patients). Moreover, compared with our previous work on pain prediction (Huang et al., 2013), which rely on EEG data from a high-density EEG cap, we obtain desirable pain prediction accuracy from EEG at only one electrode so that the preparation period is significantly reduced, which makes the proposed EEG-based prediction method more suitable for both clinicians and patients.

REFERENCES

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Babiloni, C., Brancucci, A., Percio, C. D., Capotosto, P., Arendt-Nielsen, L., Chen, A. C., et al. (2006). Anticipatory electroencephalography alpha rhythm predicts subjective perception of pain intensity. *J. Pain* 7, 709–717. doi: 10.1016/j.jpain.2006.03.005

Boly, M., Balteau, E., Schnakers, C., Degueldre, C., Moonen, G., Luxen, A., et al. (2007). Baseline brain activity fluctuations predict somatosensory perception in humans. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12187–12192. doi: 10.1073/pnas.0611404104

Brodersen, K. H., Wiech, K., Lomakina, E. I., Lin, C.-S., Buhmann, J. M., Bingel, U., et al. (2012). Decoding the perception of pain from fMRI using multivariate pattern analysis. *Neuroimage* 63, 1162–1170. doi: 10.1016/j.neuroimage.2012.08.035

Brown, J. E., Chatterjee, N., Younger, J., and Mackey, S. (2011). Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. *PLoS ONE* 6:e24124. doi: 10.1371/journal.pone.0024124

Buffum, M. D., Hutt, E., Chang, V. T., Craine, M. H., and Snow, A. L. (2007). Cognitive impairment and pain management: review of issues and challenges. *J. Rehabil. Res. Dev.* 44, 315. doi: 10.1682/JRRD.2006.06.0064

Buzsaki, G. (2009). *Rhythms of the Brain*. New York, NY: Oxford University Press.

Cecchi, G. A., Huang, L., Hashmi, J. A., Baliki, M., Centeno, M. V., Rish, I., et al. (2012). Predictive dynamics of human pain perception. *PLoS Comput. Biol.* 8:e1002719. doi: 10.1371/journal.pcbi.1002719

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Cruccu, G., Gronseth, G., Alksne, J., Argoff, C., Brainin, M., Burchiel, K., et al. (2008). AAN-EFNS guidelines on trigeminal neuralgia management. *Euro. J. Neurol.* 15, 1013–1028. doi: 10.1111/j.1468-1331.2008.02185.x

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D., Acharyya, M., Loughead, J., et al. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage* 28, 663–668. doi: 10.1016/j.neuroimage.2005.08.009

De Lange, F. P., Rahnev, D. A., Donner, T. H., and Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *J. Neurosci.* 33, 1400–1410. doi: 10.1523/JNEUROSCI.1094-12.2013

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Frackowiak, R. S., Friston, K. J., Frith, C. D., Dolan, R. J., and Mazziotta, J. (2004). *Human Brain Function*. San Diego, CA: Academic Press.

Haanpää, M., Attal, N., Backonja, M., Baron, R., Bennett, M., Bouhassira, D., et al. (2011). NeuPSIG guidelines on neuropathic pain assessment. *Pain* 152, 14–27. doi: 10.1016/j.pain.2010.07.031

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., and Bäuml, K.-H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *Neuroimage* 37, 1465–1473. doi: 10.1016/j.neuroimage.2007.07.011

Herr, K. A., Spratt, K., Mobily, P. R., and Richardson, G. (2004). Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin. J. Pain* 20, 207–219. doi: 10.1097/00002508-200407000-00002

Hjelm, R. D., Calhoun, V. D., Salakhutdinov, R., Allen, E. A., Adali, T., and Plis, S. M. (2014). Restricted Boltzmann machines for neuroimaging: an application in identifying intrinsic networks. *Neuroimage* 96, 245–260. doi: 10.1016/j.neuroimage.2014.03.048

Hu, L., Xiao, P., Zhang, Z., Mouraux, A., and Iannetti, G. (2014). Single-trial time–frequency analysis of electrocortical signals: baseline correction and beyond. *Neuroimage* 84, 876–887. doi: 10.1016/j.neuroimage.2013.09.055

Huang, G., Xiao, P., Hung, Y., Iannetti, G. D., Zhang, Z., and Hu, L. (2013). A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. *Neuroimage* 81, 283–293. doi: 10.1016/j.neuroimage.2013.05.017

Jensen, M. P., and Karoly, P. (1992). "Self-report scales and procedures for assessing pain in adults," in *Handbook of Pain Assessment,* eds D. C. Turk and R. Melzack (New York, NY: Guilford Press), 131–151.

Lange, J., Halacz, J., Van Dijk, H., Kahlbrock, N., and Schnitzler, A. (2012). Fluctuations of prestimulus oscillatory power predict subjective perception of tactile simultaneity. *Cereb. Cortex* 22, 2564–2574. doi: 10.1093/cercor/bhr329

Legrain, V., Iannetti, G. D., Plaghki, L., and Mouraux, A. (2011). The pain matrix reloaded: a salience detection system for the body. *Prog. Neurobiol.* 93, 111–124. doi: 10.1016/j.pneurobio.2010.10.005

Linkenkaer-Hansen, K., Nikulin, V. V., Palva, S., Ilmoniemi, R. J., and Palva, J. M. (2004). Prestimulus oscillations enhance psychophysical performance in humans. *J. Neurosci.* 24, 10186–10190. doi: 10.1523/JNEUROSCI.2584-04.2004

Loggia, M. L., Edwards, R. R., Kim, J., Vangel, M. G., Wasan, A. D., Gollub, R. L., et al. (2012). Disentangling linear and nonlinear brain responses to evoked deep tissue pain. *Pain* 153, 2140–2151. doi: 10.1016/j.pain.2012.07.014

Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45, S199–S209. doi: 10.1016/j.neuroimage.2008.11.007

Plis, S. M., Hjelm, D. R., Salakhutdinov, R., Allen, E. A., Bockholt, H. J., Long, J. D., et al. (2014). Deep learning for neuroimaging: a validation study. *Front. Neurosci.* 8:229. doi: 10.3389/fnins.2014.00229

Ploner, M., Lee, M. C., Wiech, K., Bingel, U., and Tracey, I. (2010). Prestimulus functional connectivity determines pain perception in humans. *Proc. Natl. Acad. Sci. U.S.A.* 107, 355–360. doi: 10.1073/pnas.0906186106

Quevedo, A. S., and Coghill, R. C. (2007). Attentional modulation of spatial integration of pain: evidence for dynamic spatial tuning. *J. Neurosci.* 27, 11635–11640. doi: 10.1523/JNEUROSCI.3356-07.2007

Ryali, S., Supekar, K., Abrams, D. A., and Menon, V. (2010). Sparse logistic regression for whole-brain classification of fMRI data. *Neuroimage* 51, 752–764. doi: 10.1016/j.neuroimage.2010.02.040

Schmah, T., Hinton, G. E., Small, S. L., Strother, S., and Zemel, R. S. (2008). "Generative versus discriminative training of RBMs for classification of fMRI images," in *Advances in Neural Information Processing Systems (NIPS)*. (Vancouver, BC).

Schnakers, C., and Zasler, N. D. (2007). Pain assessment and management in disorders of consciousness. *Curr. Opin. Neurol.* 20, 620–626. doi: 10.1097/WCO.0b013e3282f169d9

Schulz, E., Zherdin, A., Tiemann, L., Plant, C., and Ploner, M. (2011). Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. *Cereb. Cortex* 22, 1118–1123. doi: 10.1093/cercor/bhr186

Terkelsen, A. J., Andersen, O. K., Mølgaard, H., Hansen, J., and Jensen, T. (2004). Mental stress inhibits pain perception and heart rate variability but not a nociceptive withdrawal reflex. *Acta Physiol. Scand.* 180, 405–414. doi: 10.1111/j.1365-201X.2004.01263.x

Tu, Y., Tan, A., Fu, Z., Hung, Y. S., Hu, L., and Zhang, Z. (2015). "Supervised nonlinear dimension reduction of functional magnetic resonance imaging data using Sliced Inverse Regression," in *Engineering in Medicine and Biology Society (EMBC)* (Milan).

Tu, Y., Zhang, Z., Tan, A., Peng, W., Hung, Y. S., Moayedi, M., et al. (2016). Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming nociceptive stimuli. *Hum. Brain Mapp.* 37, 501–514. doi: 10.1002/hbm.23048

Valentini, E., Hu, L., Chakrabarti, B., Hu, Y., Aglioti, S. M., and Iannetti, G. D. (2012). The primary somatosensory cortex largely contributes to the early part of the cortical response elicited by nociceptive stimuli. *Neuroimage* 59, 1571–1581. doi: 10.1016/j.neuroimage.2011.08.069

Van Dijk, H., Schoffelen, J.-M., Oostenveld, R., and Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *J. Neurosci.* 28, 1816–1823. doi: 10.1523/JNEUROSCI.1853-07.2008

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., and Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *New Engl. J. Med.* 368, 1388–1397. doi: 10.1056/NEJMoa1204471

Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. (2005). Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. *Neuroimage* 25, 206–218. doi: 10.1016/j.neuroimage.2004.11.008

Wiech, K., Ploner, M., and Tracey, I. (2008). Neurocognitive aspects of pain perception. *Trends Cogn. Sci. (Regul. Ed).* 12, 306–313. doi: 10.1016/j.tics.2008.05.005

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics Intell. Lab. Syst.* 58, 109–130. doi: 10.1016/S0169-7439(01)00155-1

Wong, D. L., and Baker, C. M. (1988). Pain in children: comparison of assessment scales. *Pediatr. Nurs.* 14, 9–17.

Zhang, Y., and Ding, M. (2010). Detection of a weak somatosensory stimulus: role of the prestimulus mu rhythm and its top–down modulation. *J. Cogn. Neurosci.* 22, 307–322. doi: 10.1162/jocn.2009.21247

Zhang, Z., Hu, L., Hung, Y., Mouraux, A., and Iannetti, G. (2012). Gamma-band oscillations in the primary somatosensory cortex—a direct and obligatory correlate of subjective pain intensity. *J. Neurosci.* 32, 7429–7438. doi: 10.1523/JNEUROSCI.5877-11.2012

Advantages of publishing in Frontiers

OPEN ACCESS

Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW

Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION

Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS

No limit to article distribution and re-use

TRANSPARENT

Editors and reviewers acknowledged by name on published articles

SUPPORT

By our Swiss-based editorial team

IMPACT METRICS

Advanced metrics track your article's impact

GLOBAL SPREAD

5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK

Our network increases readership for your article

Find us on