Turning visual shapes into sounds: Early stages of reading acquisition revealed in the ventral occipitotemporal cortex

M. Perrone-Bertolotti, J.R. Vidal, L. de Palma, C. M. Hamamé, T. Ossandon, P. Kahane, L. Minotti, O. Bertrand, J.-P. Lachaux

Abstract

The exact role of the left ventral occipitotemporal cortex (VOTC) during the initial stages of reading acquisition is a hotly debated issue, especially regarding the comparative effect of learning on early stimulus-dependent vs. later task-dependent processes. We show that this controversy can be solved with high-temporal resolution intracerebral EEG recordings of the VOTC. We measured High-Frequency Activity (50–150 Hz) as a proxy of population-level spiking activity while participants learned Japanese Katakana symbols, and found that learning primarily affects top-down/task-dependent neural processing, after a few minutes only. In contrast, adaptation of early bottom-up/stimulus-dependent processing takes several days to adapt and provides the basis for fluent reading. Such evidence that two consecutive stages of neural processing, stimulus- and task-dependent are differentially affected by learning, can reconcile seemingly opposite hypotheses on the role of the VOTC during reading acquisition.

Introduction

What makes written characters so special and different from other visual shapes? How does the brain learn to associate systematic sounds to written graphemes (a process called grapho-phonological conversion at the basis of reading acquisition)? The cortical mechanisms underlying the transformation of unknown visual shapes into text (which can be read), through learning, are not well understood. Some of these studies have suggested that it might rely on enhanced functional connectivity between cortical regions supporting phonological on one hand and visuo-orthographic processing on the other hand (Hashimoto and Sakai, 2004, see also Blomert, 2011), but others have claimed that local neural plasticity mechanisms in the latter, in the visual cortex, might be sufficient (Mei et al., 2012). One key question is the exact role in the learning process of the ventral occipitotemporal cortex (VOTC), which is known to be highly involved in visuo-orthographic analysis (Dehaene et al., 2002) in general and very active while learning to read, as shown by several studies using artificial language training (e.g., Xue et al., 2006), associative learning (e.g., Song et al., 2012) or letter-sound crossmodal matching tasks (e.g., Hashimoto et al., 2004).

We chose to investigate the role of the VOTC while learning to read by studying neural activity in that region in western adult readers learning to read Japanese Katakana symbols (one of the three Japanese syllabic writing systems). Although similar to roman letters in terms of basic visual properties Katakana symbols do not spontaneously evoke a sound-form in western readers unfamiliar with that alphabet (in contrast with other line drawings of familiar objects or geometrical shapes, which often do evoke their name), but they can be learned in a few minutes and therefore be ‘read’. Comparing the brain response to Katakana before and after learning can potentially shed light on the neural mechanisms which underlie the transformation of meaningless shapes into text.

In this paper, we test a novel hypothesis based on a recent study by Bastin et al. (2013) introducing a clear temporal distinction between two components of the response to visual stimuli in high-level visual areas along the ventral visual pathway. Using intracranial EEG recordings of the parahippocampal area, Bastin et al. (2013) found that the initial component (<300 ms) of the neural response to visual stimuli is stimulus-dependent, that is, highly tuned to the category of the visual stimulus (i.e. present when the stimulus is a face, absent when it is a place) but independent on task-instruction. That early response component is immediately followed by a more sustained component (>300 ms) in which amplitude and duration vary according to task specifications (i.e. what kind of information must be extracted from the stimulus to guide action selection). That study suggested two consecutive information processing stages: an early recognition of the

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.12.027

Please cite this article as: Perrone-Bertolotti, M., et al., Turning visual shapes into sounds: Early stages of reading acquisition revealed in the ventral occipitotemporal cortex, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.12.027
stimulus by neurons specialized for that particular type of visual object or scene (either genetically or through long-term plasticity), followed by a more detailed and prolonged analysis of the stimulus in the same cortical region to extract information according to task instructions. Although Bastin’s study used scene perception, we reasoned that when symbols should also elicit an early stimulus-dependent and late task-dependent components in high-order visual areas. We thus hypothesized that during the initial stage of reading acquisition, the emergence of grapho-phonological conversion for newly learned symbols should mostly affect the later (the extraction of visual information for action selection, i.e. pronunciation) and not the former, because neural specialization for the new symbols should not occur over the course of a few minutes (given the limits of neural plasticity).

This hypothesis directly relates to an ongoing debate regarding the neural basis of reading acquisition: two influential theories predict seemingly opposite changes at the neural level, with a major disagreement regarding the role of a high-level visual region with neurons highly tuned to line configurations in the left VOTC (Dehaene and Cohen, 2011; Price and Devlin, 2011). The Neuronal Recycling Hypothesis (“NRH”, Dehaene and Cohen, 2011) claims that during reading acquisition, VOTC neurons are recycled into detectors with optimal selectivity for “being-learned” letters and word-forms, turning the VOTC into a “Word-Form Area”. Learning to read unfamiliar symbols such as Katakana would therefore involve a progressive tuning of VOTC neurons adjusting their selectivity to those characters, even in a few hours (Brem et al., 2010) provided that readers explicitly associate visual and speech forms (Dehaene and Cohen, 2011). Following Bastin’s nomenclature, the NRH would therefore predict that reading acquisition primarily affects the early stimulus-dependent response in the VOTC, shaped by neural tuning. The second theory, the Interactive Account (“IAT”), proposes that VOTC neurons are not tuned to specific word-forms (Price and Devlin, 2011), but that their activity is sustained during reading by an integration process combining the bottom-up flow of visual inputs and top-down influences – called “predictions” – from high-level language regions (such as Broca’s Area, e.g. Yvert et al., 2012): the strong BOLD activity measured in VOTC in response to orthographic stimuli would be due to the maintenance of neural firing locally by long-range interactions with the reading network, as long as the visual stimulus has not been associated with a phonological form (Song et al., 2012), and not by a simple tuning of those neurons, turned into rapid detectors of those letters. One prediction of the IAT is that the integration process cannot take place for unknown shapes, and is maximal when such association has not yet become fully automatic and take time to be retrieved. Therefore, the duration of VOT neuron responses to Katakana should be short before learning, then reach a maximum longer than for overlearned, roman characters (Price and Devlin, 2011). Following Bastin (2013), again, the IAT would mostly predict a change in the late, task-dependent response, contingent upon the pronunciation of the visual symbol (the task).

One complicating factor of that debate is that both theories are based on fMRI data, which lack the sufficient temporal resolution to distinguish between the stimulus- and task-dependent components of the neural response (Logothetis, 2008). We reasoned that iEEG recordings in the VOTC could resolve the apparent contradiction between NRH and IAT thanks to its high temporal precision. We recorded iEEG signals in the VOTC of four epileptic patients learning Katakana and used High-Frequency Activity (50–150 Hz) as a proxy of population spiking activity (Lachaux et al., 2012; Manning et al., 2009), and a strong correlate of the BOLD signal (Kim et al., 2004; Lachaux et al., 2007; Logothetis et al., 2001).

We showed that both roman characters and Katakana symbols (learned and unlearned) generate an early stimulus-dependent response in the posterior VOTC (peak before 300 ms), which was not affected by learning over the course of 3 h. In contrast, Katakana elicited a subsequent and prolonged task-dependent response (up to 800 ms) when participants had learned to read them, which took a few minutes only. We interpret that component as the neural trace of an action selection mechanism which associates learned Katakana’s (as well as roman characters) with their phonemic expression, and which instantiates the earliest stage of reading acquisition. Those results confirmed our hypothesis and resolve the apparent contradiction between the NRH and the IAT: both theories are compatible, but the IAT primarily describes the short-term effects of reading acquisition on the task-dependent response component in the VOTC while the NRH accounts for long-term plasticity effects changing neural specialization and the stimulus-dependent response component.

Materials and methods

Participants

Intracranial recordings were obtained in four neurosurgical patients with intractable epilepsy (3 females, right-handed, mean age: 32 ± 6 years; see Table 1 for demographic information) at the Epilepsy Department of the Grenoble Neurological Hospital (Grenoble, France). Eleven to fifteen semi-rigid, multi-lead electrodes were stereotactically implanted in each patient (stereotactic EEG–SEEG, Kahane et al., 2003). The SEEG electrodes used have a diameter of 0.8 mm and, depending on the target structure, consist of 10 to 15 contact leads 2 mm wide and 1.5 mm apart (DIXI Medical Instruments). Selection of sites to implant was entirely based on clinical purposes, with no reference to the present experimental protocol. The electrodes recorded in this study did not exhibit pathological waveforms. The 4 participants were native French speakers, with no prior knowledge of the Japanese Katakana writing system, and all had normal verbal and performance IQ in French (see Table 1). All patients provided written informed consent, and the experimental procedures were approved by the Institutional Review Board and by the National French Science Ethical Committee.

Patients were selected according to our working hypothesis, and the fact that the ongoing debate between the two theories on reading acquisition, the NRH and the IAT, explicitly address the functional role of a specific cortical region, the ventral occipitotemporal cortex (VOTC), defined by a preferential response to general letter-strings and/or word-forms. This is why; out of the 100 + cortical sites which were recorded in each patient (see Supplementary Fig. 1), we selected only sites with such response properties, for each patient. This selection was based on visual functional localizers (allowing for the delineation of a functional region with preferential response to word-like items, see below: Stimuli and tasks section and also Vidal et al., 2010). It is worth reminding that electrodes accessing that particular region are extremely rare (about once every fifteen patients on average in our center), because of the diversity of the implantations across patients (imposed by clinical considerations). Those sites only were analyzed in the context of Katakana learning (with the exception of one site in Broca, to test for connectivity, see Table 1).

Intracranial recordings

Intracranial recordings were conducted using a video-SEEG monitoring system (Micromed), which allowed the simultaneous data recording from 128 depth-EEG electrode sites. The data were bandpass filtered online from 0.1 to 200 Hz and sampled at 512 Hz in all patients. At the time of acquisition the data are recorded using a reference electrode located in white matter, and each electrode trace is subsequently re-referenced with respect to its direct neighbor (bipolar derivations). This bipolar montage has a number of advantages over common referencing. It helps eliminate signal artifacts common to adjacent electrode contacts (such as the 50 Hz mains artifact or distant physiological artifacts) and achieves a high local specificity by canceling out effects of distant sources that spread equally to both adjacent sites through volume conduction. The spatial resolution achieved by the bipolar SEEG is on the order of 3 mm (Jerbi et al., 2009; Kahane et al., 2003; Lachaux et al., 2003).
Both spatial resolution and spatial sampling achieved with SEEG differ slightly from that obtained with subdural grid electrocorticography (Jerbi et al., 2009).

### Data analysis and statistical analysis

#### Time-frequency analysis, gamma power, and envelope computations

The frequency band of interest, between 50 Hz and 150 Hz, was defined from preliminary time–frequency (TF) analysis of the SEEG data using wavelets (Tallon-Baudry et al., 1997), performed with in-house software package for electrophysiological signal analysis (ELAN) developed at INSERM U1028, Lyon, France (Aguea et al., 2011), and from previous studies by our group (Jerbi et al., 2009).

Raw data were transformed into High-Frequency Amplitude (HFA) time-series with the following procedure (Ossandon et al., 2011; Perrone-Bertolotti et al., 2012): step 1) continuous SEEG Signals were first bandpass filtered in multiple successive 10 Hz wide frequency bands (e.g., 10 bands from [50–60 Hz] to [140–150 Hz]) using a zero-phase shift no causal finite impulse filter with 0.5 Hz roll-off. Step 2) next, for each bandpass filtered signal we computed the envelope using standard Hilbert transform (Le Van Quyen et al., 2001). The obtained envelope is down sampled to a sampling rate of 64 Hz (i.e., one time sample every 15,625 ms). Step 3) for each band this envelope signal (i.e., time-varying amplitude) was divided by its mean across the entire recording session and multiplied by 100. This yields instantaneous envelope values expressed in percentage (%) of the mean. Step 4) the envelope signals computed for each consecutive frequency bands (the ten bands of 10 Hz intervals between 50 and 150 Hz) were averaged together to provide one single time series (gamma-band activity, GBA) across the entire session. By construction, the mean value of that time series across the recording session is equal to 100. Finally amplitude-time series was epochs into data segments centered on each stimulus, between −200 and +1000 ms relative to stimulus onset, and then averaged together for each stimulus category. Note that computing the Hilbert envelopes in 10 Hz sub-bands and normalizing them individually before averaging over the broadband interval allows us to account for a bias toward the lower frequencies of the interval that would otherwise occur due to the 1/f drop-off in amplitude.

Statistical analyses were performed on high-frequency activity time series, HFA, computed as above. To test for significant increases or decreases compared to baseline activity, we used paired-sample Wilcoxon signed rank test, followed by false discovery rate (FDR) correction across all time samples. This allowed for quantitative definition of the onset time and duration of activation of specific recording sites. To compare the activity from paired conditions we used the non-parametrical Kruskal–Wallis test followed by FDR correction across all time samples.

Finally, the anatomical display of all significant HFA modulations was obtained by pooling data from all subjects and mapping them onto the standard Montreal Neurological Institute (MNI) single-subject brain based on the localization of each electrode. The precise anatomical location of the electrodes (and their MNI coordinates) was obtained by aligning the pre-implantation and the post-implantation (showing the electrodes in place) structural MRIs of each patient using the NUTMEG toolbox (Dalal et al., 2004), SPM8 (http://www.fil.ion.ucl.ac.uk/spm) and custom Matlab routines.

#### Visual functional localizers

This first visual localizer (called in short VISL, and analyzed extensively in Vidal et al., 2010) allowed us to delineate a functional region with preferential response to word-like items during a visual oddball task with no specific language demand. Visual stimuli were gray-scale pictures (with the same average luminance) drawn from nine categories: houses; faces; animals, scenes, objects, pronounceable pseudo-words (e.g., “boutele”); 6–9 letters long), unpronounceable consonant-strings (e.g. “bestvila”; 6–9 letters long), fruits and scrambled images. No stimulus was presented more than once. Participants were instructed to press a button each time a picture of a fruit (target) appeared on screen. Stimuli were presented during 200 ms every 1000–1200 ms in series of 5 pictures interleaved by 3-s pause periods during which participants could freely blink. Participants reported the detection of a target through a right-hand button press and were given feedback of their performance after each fruit item. A 2-s delay was placed after each button press before presenting the follow-up stimulus in order to avoid mixing signals related to motor action with signals from stimulus presentation. Each visual category was presented in a randomized order 50 times during the whole experiment; each item was presented only once. We compared HFA for each of the categories of items and selected the most selective/specific site to visual word-like items for each patient in the VOTC. Further, we compared neuronal activity in response to word-like items (pseudo-words and consonant strings). Visual stimulation was delivered via the Presentation stimulus delivery software (Neurobehavioral systems Inc.).

#### Learning pronunciation task

The pronunciation learning (PROL, see Fig. 1) task was performed in 4 sessions (except for P1 and P4, who left the epilepsy unit before session 4). The delay between sessions was variable: the first three sessions were performed the same day (1 h between sessions 1 and 2, 4 h between sessions 1 and 3) and the last session was performed four days later. The experimental paradigm and stimuli were the same in all sessions. Participants were instructed to covertly pronounce all visual items that they could pronounce. Stimuli were: 5 syllables (in alphabetic writing system, called in short SY condition, SU, KE, RI, TA, RA) and 10 Katakana Japanese scripts (in syllabic writing system; called in short KATA condition).

The pronunciation learning task was performed in 4 sessions (except for P1 and P4, who left the epilepsy unit before session 4). The delay between sessions was variable: the first three sessions were performed the same day (1 h between sessions 1 and 2, 4 h between sessions 1 and 3) and the last session was performed four days later. The experimental paradigm and stimuli were the same in all sessions. Participants were instructed to covertly pronounce all visual items that they could pronounce. Stimuli were: 5 syllables (in alphabetic writing system, called in short SY condition, SU, KE, RI, TA, RA) and 10 Katakana Japanese scripts (in syllabic writing system; called in short KATA condition).
were instructed to learn by themselves, at their own pace, the association between the five upper Katakana items (half of the Katakana items, called KL) and their corresponding phonological syllabic form until they felt confident they could read all of them if shown in the same context as session S1. All participants took less than ten minutes to do so. From session 2 the participants were thus able to pronounce five syllables and five KL items but not the other five KATA items (called KU — unlearned-condition since session 2).

Results

Stimulus-dependent responses to written symbols in the VOTC

Our analysis strategy was intended to test two claims: a) that written symbols elicit a dual-component response in the VOTC, with an early stimulus-dependent response followed by a late task-dependent response, b) that learning to read does not affect initially the stimulus-dependent response. In general, we call “stimulus-dependent” a response component which depends solely on the nature of the visual stimulus and not on the task that is explicitly or implicitly (i.e. automatic grapho-phonological conversion) associated by the participant; it is produced by neurons tuned to the physical features of that stimulus. In contrast, we call “task-dependent” a response component which appears only if participants process the stimulus further to associate it with an overt or covert behavioral response (i.e. mentally pronounce a written symbols). In this particular context, we were interested in stimulus-dependent responses tuned to orthographic stimuli.

The visual localizer (VISU) was designed to identify such responses. In the VISU task, all stimuli were associated with the same behavioral response (no response unless it’s a fruit). Therefore, stimulus-specific responses to letter strings were easy to identify: they should be observed to strings which can be read (pseudo-words) as well as to strings which cannot (consonant strings), but not to other stimulus categories (faces, landscapes, animals…). Fig. 2 shows four sites with such a response (P2b, P3, P4, P2c, Fig. 2, left panel). All sites were characterized by a stronger HFA [50–150 Hz] energy increase after orthographic stimuli than after stimuli of other categories such as house, animal, scenes (Kruskal–Wallis, corrected p < 0.05). In each patient, selective sites were located in the left VOTC (Fig. 2 and Table 1). All three sites were within the commonly accepted frame of the visual word form area (P1, P2a, P2b), as defined by Dehaene and Cohen (2011).

In those three posterior VOTC sites, the comparison between pseudo-words and consonant strings (Fig. 2, middle panel) revealed a response component with an initial peak between 200 and 300 ms which was identical for all strings, whether they could be pronounced or not. The responses to consonant strings did not last beyond 500 ms, which then marked the end of the stimulus-specific component. The responses to pseudo-words lasted longer (Kruskal–Wallis, corrected p < 0.05, see figure) but since those strings could be pronounced, they might be attributed to automatic grapho-phonological conversion (a form of covert behavioral response).

Responses selective to letter strings were also observed in three more anterior VOTC sites (Fig. 2, left and middle panel). However, those responses did not qualify as stimulus-specific (to letter strings) because they were not systematically observed in response to all consonant strings (Fig. 2, left panel). The possibility that their initial portion (≤400 ms) might correspond to a stimulus-dependent response tuned to pronounceable word-forms is presented in the Discussion section.

In short, the analysis of the VISU task revealed stimulus-dependent responses specific to written symbols in three posterior VOTC sites, characterized by an early peak between 200 and 300 ms.

KATA: Syllables vs. symbols and effect of learning

The stimulus-dependent peak in the posterior VOTC was observed in the same neural populations in response to both Katakana and roman characters, before learning already (Fig. 2, right panel, top three rows, SY1 vs. KATA), but it is stronger for Syllables (Kruskal–Wallis, corrected p < 0.05, Wilcoxon signed rank test, FDR corrected for multiple comparisons).
Fig. 3. Learning effect (before vs after) for each item category during the PROL-task. The figure shows the HFA [50–150 Hz] increase, averaged across trials (± SEM), for the first (1) vs last (3) session of learning, for syllables (SY), learned (KL) and unlearned kana (KU) in each of recorded sites (P1–P4). Colored lines under the curves indicate for each condition, time samples for which HFA was significantly above pre-stimulus baseline level (p < 0.05, Wilcoxon signed rank test, FDR corrected for multiple comparisons); black lines above the curves indicate time samples at which HFA was significantly different between first and last session of each of categories of stimuli (p < 0.05, Kruskal–Wallis, FDR corrected for multiple comparisons).
p < 0.05; Fig. 2; right panel). That difference was significant from the very start of the response, as soon as 150–200 ms after stimulus onset. This observation contradicts previous suggestion that VOTC neurons might respond equally well to the subject’s native alphabet and to unknown alphabets (Xue and Poldrack, 2007). Since Katakana and syllabic items are made of similar sets of basic line-junction features, it rather suggests that the VOTC is indeed tuned to familiar orthographic systems (Baker et al., 2007; Dehaene et al., 2010; Pegado et al., 2011), as predicted by the NRH.

The comparison of the response to not-to-be-learned KATA (KU) across sessions revealed virtually no effect of repetition (Fig. 3, right panel, KU1 vs KU3). The global response remained unchanged, despite the fact that participants had seen each KU stimuli at least 30 times when reaching the end of session 3. The effect of learning was only observed for to-be-learned KATA (KL), when comparing the response before and after learning (Fig. 3, middle panel, KL1 vs KL3). In all six VOTC sites, including anterior sites with no stimulus-specific response (to letter strings), we observed a massive and sustained effect of learning after 300 ms, with a latency that came after the peak of the stimulus-dependent response in posterior VOTC sites (Kruskal–Wallis, corrected p < 0.05), consistent with the timing of what we called “task-dependent response component”, and in line with the fact that KL3 is now associated with a covert behavioral response (phonological) while KL1 was not. Visual inspection of posterior VOTC sites in P1 and P2a (same figure, top rows) shows an early effect of learning, before 300 ms, but it is always after the initial rise and peak which, as we now know from VISU, are strictly stimulus-dependent (the initial peak of the KL1 response, in particular, the rising part of the response remains unchanged in those two patients). We therefore interpret those early effects as part of an additional component contingent upon the covert pronunciation of the symbols.

Anterior VOTC sites had a similar dynamics. The early phase of the response, before 400 ms, which was mostly observed for pronounceable chains of characters (Fig. 2, pseudo-words vs consonant, middle rows) showed a massive and sustained effect of learning after 300 ms, with a latency that came after the peak of the stimulus-dependent response in posterior VOTC sites (Kruskal–Wallis, corrected p < 0.05), consistent with the timing of what we called “task-dependent response component”, and in line with the fact that KL3 is now associated with a covert behavioral response (phonological) while KL1 was not. Visual inspection of posterior VOTC sites in P1 and P2a (same figure, top rows) shows an early effect of learning, before 300 ms, but it is always after the initial rise and peak which, as we now know from VISU, are strictly stimulus-dependent (the initial peak of the KL1 response, in particular, the rising part of the response remains unchanged in those two patients). We therefore interpret those early effects as part of an additional component contingent upon the covert pronunciation of the symbols.

Fig. 4. Ventral occipitotemporal cortex (E’8, in red) and inferior frontal gyrus (Q’6, in green) activity during learning in P3. Figure shows the anatomical localization of E’8 and Q’6 sites, and the HFA[50–150 Hz] averaged (± SEM) across trials for syllables (SY1), katakana-learned (KL) and unlearned (KU) before (session 1) and four days later (session 4). Shaded areas indicate time windows during which HFA for E’8 and Q’6 were significantly correlated. The colored lines under the curves indicate time samples when the conditions were significantly different from the baseline (p < 0.05, Wilcoxon signed rank test, FDR corrected for multiple comparisons).

Please cite this article as: Perrone-Bertolotti, M., et al., Turning visual shapes into sounds: Early stages of reading acquisition revealed in the ventral occipitotemporal cortex, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.12.027
suggests that neurons within the VOTC are optimally tuned to familiar orthographic stimuli. iEEG data from our group and also from other groups are consistent with that claim (see also Hamamé et al., 2013; Thesen et al., 2012; Vidal et al., 2012). Indeed, as shown in Fig. 2 (Panel B), neural response in the posterior VOTC regions was stronger to roman characters (SY items) than to Katakana items around 200 ms, despite extremely similar visual features. Such specificity might indeed arise from neural recycling of neural groups with high-sensitivity to the line configurations found in letters, which would adapt their sensitivity to match the specific forms of learned graphemes. This is consistent with the fMRI-based claim by Dehaene and Cohen (2011), that “3.6 hours are sufficient to enhance the response to letter strings relative to false fonts”; although the distinction between a strong transient response and weaker sustained neural response is invisible in fMRI (Lachaux et al., 2007). Future experiments are thus needed to test whether the specific increase of the early stimulus-dependent response component can be induced in that VOTC neural population, over the long-term, for non-orthographic stimuli with no systematic speech-form associated (such as unlearned Katakana), after extensive familiarization with those stimuli. This would provide evidence, against the NRH, that this specific neural group can be tuned to non-orthographic stimuli. In this regard, it is interesting that in one participant (P2), the very early response to syllables (SY, before 200 ms) increased between session 1 and session 3. This might indicate that neural plasticity can still occur for overlearned stimuli. An important question for future work is whether that early effect of session repetition points to an overlooked function of the VOTC during character recognition.

Second, the IAT of the VOTC (Price and Devlin, 2011) suggesting a prolonged task-dependent response component reflects the interactions between VOTC and higher-order language areas encoding phonology or semantics (Kherif et al., 2011; Song et al., 2012). The theory predicts that Katakana characters which pronunciation is being learned should trigger a strong and sustained interaction between VOTC and phonological regions, until the patterns of activity in neural populations supporting visual and phonological decoding of the sensory input are consistent with each other (what the theory calls “minimizing the prediction error”). The interaction should not occur for characters that the participant cannot read yet, and be shorter for overlearned ones. The late task-dependent response component we observed, which was correlated between VOTC and Broca’s Area, was indeed longer for recently learned Katakana than for roman characters and absent for unlearned Katakana. In addition, high-resolution iEEG recordings also revealed that the latency of the task-dependent response decreases with learning (Fig. 5), suggesting that the interaction with phonological regions starts earlier as the grapho-phonological conversion becomes more systematic.

Interestingly, the IAT and the NRH have been considered as two opposite interpretations of the VOTC and its role in reading. Yet, regarding the early stages of reading acquisition in that region, we observe that both theories make correct predictions, but on separate neural processes: on the late task-dependent response (IAT) and on the early stimulus-dependent responses (NRH) respectively. The enhancement of the task-dependent response occurs almost immediately, as soon as the participant can associate a phonological form with the visual shape, while a much slower tuning of VOTC neurons starts taking place, which will eventually cause a stronger stimulus-dependent response to learned characters, after extensive learning (as seen in the response to syllables).

Turning visual shapes into text

Our study shows that the main immediate effect of learning a grapheme-phoneme association is to generate a sustained neural response in the VOTC and higher-order areas encoding phonology, such as Broca’s Area. It should be emphasized that in our data, the VOTC was continuously active while retrieving the correct pronunciation of...
a written symbols. This contradicts a scenario where the VOTC would analyze the visual in “one shot” and pass that information to other brain regions which would then retrieve the phonological form: it clearly demonstrates that grapho-phonological conversion is indeed a sustained process involving the VOTC continuously.

However, it is also remarkable that the effect of learning is not simultaneous across the VOTC: it is earlier in the posterior sites (Fig. 3, P2a, b, c). We can therefore exclude that, for katakana, the newly learned reading process associating grapheme with phoneme is a global one, involving as a whole a network of cortical populations that would add to the purely visual response. Even in the VOTC, that effect is heterogeneous, and in fact, simultaneous recordings from the VOTC and Broca show that the rule probably extends to the whole reading network: the differential reaction to learned and unlearned written symbols is later as we move up in the hierarchy of cortical areas.

The IAT provides a very explicit interpretation of the sustained response in the VOTC, but it can also be understood in the more general context of perceptual decision making (Donner et al., 2009; Heekeren et al., 2004, 2008), if we consider that our task is to associate a visual input with a motor program (subtending mental pronunciation of the Katakana). Several models have been proposed to account for the selection of a motor process in response to a sensory stimulus that must be interpreted (Shadlen and Newsome, 2001; Gold and Shadlen, 2007), and most of them propose that sensory regions accumulate evidence until a perceptual decision is taken regarding what is on screen (Heekeren et al., 2004, 2008). This model is consistent with single-unit data recorded while monkeys must perceive global motion direction in random moving dot displays (Shadlen and Newsome, 2001).

The accumulation process stops when activity of neurons encoding a specific direction reaches above a critical threshold, which takes longer when the motion is ambiguous. In our task, one might argue that the sensory interpretation of newly learned Katakana also involves an evidence accumulation process before critical VOTC neurons reach a decision threshold. In a sequential model, this decision would then be propagated to phonological areas to guide the selection of a motor articulatory program. Thus, if the first step, within the more posterior groups of neurons, is not globally fulfilled, the next steps cannot be achieved, as observed for the unlearned Katakana items. This interpretation is supported by recent results showing disruption in naming during cortical stimulation in the posterior sites of the VOTC (Thesen et al., 2012). This is not the case, since we found that the sustained response components in the VOTC and Broca’s Area overlap in time, suggesting rather two simultaneous evidence accumulation processes at sensory and premotor levels. One possibility, therefore, is that both processes constrain each other in a bidirectional way, in line with the IAT. Indeed, top-down information could help to rapidly conclude the competition between different groups of low-level selective neurons and allow neural patterns refinement within the VOTC (Wang, 2008).

What is increasingly clear is that processing of orthographic stimuli in the VOTC itself is not one single homogeneous process, but rather a collection of processes of increasing complexity along a posterior–anterior gradient, as suggested by previous fMRI and iEEG studies already (Vinkier et al., 2007; Thesen et al., 2012). We observed that anterior sites lack the initial stimulus-specific analysis component and get into play later, for the actual task-at-hand (grapho-phonological conversion in our case). One possibility is that the anterior VOTC contains a true word-form area (calling a word-form a chain of symbols which can be pronounced, i.e. pseudo-words and not consonant strings); the possibility remains that the stronger response to word-forms in anterior VOTC sites (VISU) is in fact a stimulus-dependent response of neurons tuned to detect word-forms (and not letter strings in general, as considered so far) based on their visual features. We cannot exclude that possibility: if grapho-phonological conversion occurs each time it is possible, it is in principle very difficult to distinguish between a) a response to word-forms that in fact corresponds to their grapho-phonological conversion and b) a response that differentiates, on the basis of physical features alone, chains of symbols which can be converted into phonological forms and those which cannot. However, we have previously reported that grapho-phonological conversion, at least the production of a full-blown phonological form reaching the auditory cortex, is not systematic (Perrone-Bertolotti et al., 2012). Further, the KATA task is designed to transform meaningless shapes into symbols that can be converted into phonemes, and we observed that the grapho-phonological conversion, which starts occurring after the learning session for KL2 and KL3, generates a response, in anterior VOTC sites, which is always later (after 400 ms, Fig. 3, middle panel, P3, P4, P2c) than the initial response to word forms (before 400 ms, see pseudo-words in VISU and SY in KATA, Figs. 2 and 3). The neural effect of an association between a symbol and a phoneme is thus always later than the neural response that differentiates familiar word forms and unfamiliar ones or non-word forms (SY vs learned Katakana, KL2 to KL4 or pseudo-words vs consonant strings in the VISU). Therefore, word-forms would elicit in anterior VOTC sites, both a stimulus-dependent response (before 400 ms), and a task-dependent response (contingent upon grapho-phonological conversion, after 400 ms). It does not contradict our main claim, as it would always be the task-dependent response which is enhanced while learning to read. But there would be neurons tuned to pronounceable word-forms in the anterior VOTC, in addition to neurons tuned to written symbols in general in the posterior VOTC, in line with previous fMRI-based suggestions that selectivity to words might increase along a posterior–anterior gradient (Vinkier et al., 2007). Still, we show that none of them adjust their selectivity to respond to new learned characters (Katakana) over the short term.

As we see, the exact role of the anterior VOTC is not clearly understood yet. Recently, fMRI studies have shown that the anterior VOTC participates in semantic information processing (Mechelli et al., 2005) and that its activity is modulated by reading experience (Olulade et al., 2013). We confirm the later, even for very short term experience, and extend the former claim to non-semantic processes. This region could then serve as a relay between posterior VOTC and phonological areas, when newly learned items start being pronounceable or for old, but unfamiliar orthographic forms. Actually, van der Mark et al. (2011) showed significant difference, between dyslexic and control children, in the functional connectivity of the anterior VOTC (where patients P2 and P3 were recorded from). Anterior VOTC activity was found...
to be more connected to the insula and middle/superior temporal cortex activity in dyslexic than in control children. These regions are especially involved during overt and covert syllable production (Heim et al., 2002) and during phonological processing (Vigneau et al., 2005); therefore, we can hypothesize that the more anterior VOTC should be more active during non-fluent reading, as in the first stage of reading acquisition or in dyslexic people (van der Mark et al., 2011), that is, before grapho-phonological conversion becomes systematic and sustained by the par-allel activity of the VOTC and associated phonological regions (such as Broca’s area); and, when automatic grapho-phonological conversion fails. This possibility will be explored in future research.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.12.027.

Q13 Uncited reference

Cohen and Dehaene, 2004

References


Price, CJ, Devlin, JT, 2011. The interactive account of ventral occipitotemporal contribu-


