Temporal tuning of word and face selective cortex

Jason D. Yeatman[1] and **Anthony M. Norcia**[2]

[1]Institute for Learning and Brain Sciences (I-LABS) and Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington

[2]Department of Psychology, Stanford University, Stanford, California

Abstract

Sensitivity to temporal change places fundamental limits on object processing in the visual system. An emerging consensus from the behavioral and neuroimaging literature suggests that temporal resolution differs substantially for stimuli of different complexity and for brain areas at different levels of the cortical hierarchy. Here we used Steady-State Visual Evoked Potentials (SSVEPs) to directly measure three fundamental parameters that characterize the underlying neural response to text and face images: temporal resolution, peak temporal frequency and response latency. We presented full-screen images of text or a human face, alternated with a scrambled image, at temporal frequencies between 1 and 12 Hz. These images elicited a robust response at the first harmonic that showed differential tuning, scalp topography and delay for the text and face images. Face selective responses were maximal at 4 Hz, but text selective responses, by contrast were maximal at 1 Hz. The topography of the text image response was strongly left-lateralized at higher stimulation rates, while the response to the face image was slightly right-lateralized but nearly bilateral at all frequencies. Both text and face images elicited steady state activity at more than one apparent latency; we observed early (141ms–160ms) and late (>250ms) text and face selective responses. These differences in temporal tuning profiles are likely to reflect differences in the nature of the computations performed by word and face selective cortex. Despite the close proximity of word and face selective regions on the cortical surface, our measurements demonstrate substantial differences in the temporal dynamics of word-versus face-selective responses.

Keywords

face perception; word perception; Visual Evoked Potentials; temporal resolution; temporal tuning

Introduction

Neurons in visual cortex are tuned to a myriad of features of the visual stimulus ranging from simple image statistics, such as spatial frequency, orientation and disparity (Barlow, Blakemore, & Pettigrew, 1967; De Valois, Albrecht, & Thorell, 1982; Hubel & Wiesel,

Corresponding Author: Jason D. Yeatman, Institute for Learning & Brain Sciences (I-LABS), Department of Speech & Hearing Sciences, University of Washington, 1715 Columbia Road N, Portage Bay Building, Seattle, WA 98115, Phone: (206) 685-2924, jyeatman@uw.edu.

1962), to dynamic properties, such as stimulus duration and direction of motion (Hubel & Wiesel, 1965; Movshon, Thompson, & Tolhurst, 1978), to high-level features such as semantic similarity, and category membership (Grill-Spector & Weiner, 2014; Huth, Nishimoto, Vu, & Gallant, 2012; Kanwisher, McDermott, & Chun, 1997). Regions of visual cortex that are sensitive to particular visual categories, such as the fusiform face area (FFA) which responds selectively to faces and the visual word form area (VWFA) which responds selectively to words, are believed to perform computations that are critical for the perception of these stimulus classes (Cohen et al., 2002; Grill-Spector & Weiner, 2014; Kanwisher et al., 1997; Wandell, Rauschecker, & Yeatman, 2012). For example disruption of signals in the FFA through electrical stimulation impairs face perception (Jonas et al., 2012; Parvizi et al., 2012) and lesions in the vicinity of the VWFA impair the ability to rapidly recognize words (a condition known as pure alexia or word blindness (Dejerine, 1891; Gaillard et al., 2006).

Despite the striking sensitivity of these ventral occipitotemporal regions to category membership, low-level features of the visual stimulus still influence neural responses. Understanding the low-level stimulus features that drive responses in ventral occipitotemporal cortex has helped elucidate fundamental aspects of visual computation and perception. For example spatial tuning, one of the most extensively studied properties of neurons in visual cortex, has been fundamental for understanding differences in the computations performed by different visual regions and linking computation to perceptual function. Ventral stream regions that are important for the perception of objects, including words and faces, predominantly receive inputs from the foveal representations of early visual areas and consequently the responses of these regions are principally driven by stimuli in the center of the visual field (Hasson, Levy, Behrmann, Hendler, & Malach, 2002; Levy, Hasson, Avidan, Hendler, & Malach, 2001). This foveal bias is believed to underlie our poor perceptual performance for objects in the periphery. For example word recognition in the periphery is substantially slower and less accurate than would be predicted by visual acuity alone (Chung, Mansfield, & Legge, 1998).

The temporal properties of the visual system also impose fundamental limits on cortical computations, but have received far less attention than spatial properties. Temporal tuning properties of the visual system can be characterized by three fundamental parameters: (1) temporal resolution or temporal acuity, (i.e., the highest temporal frequency that elicits a response to a given visual feature); (2) the temporal frequency that elicits the maximal response to that feature and (3) the delay of the response with respect to the stimulus onset (latency). The fastest rate at which neurons can track changes in a stimulus is related to the integration time of the system: neurons that integrate over long time periods effectively low-pass filter their inputs and have low temporal acuity/resolution.

In the case of simple features such as luminance and contrast, temporal resolution is very high (Kelly, 1961a, 1961b), but for more complex features, and objects, temporal resolution is much lower (Battelli, Cavanagh, Martini, & Barton, 2003; Holcombe, 2009; McMains & Somers, 2004; Potter & Faulconer, 1975). A parallel temporal hierarchy has also been observed as one progresses from early visual cortex to extra-striate areas in the temporal lobe. Early PET measurements in striate cortex indicated that peak responses to reversing checkerboards occurred between 4–15 Hz and similar tuning was observed using fMRI

(Kwong et al., 1992; Thomas & Menon, 1998; Zhu et al., 1998) with a consensus that peak responses occur near 8 Hz (but see (Ozus et al., 2001), which reported that the peak response plateaus at 6 Hz). Temporal integration of more complex information present in natural object images was first reported to differ between early retinotopic cortex and higher-order occipito-temporal areas by (Mukamel, Harel, Hendler, & Malach, 2004). Using fMRI, they found that, while activation increased by 200% in early visual cortex for presentation rates between 1 Hz and 4 Hz, the increase was only 25% in occipitotemporal cortex. The difference was attributed to differences in integration time among areas that are at different stages of the visual hierarchy. Later work (McKeeff, Remus, & Tong, 2007) compared temporal tuning profiles over both retinotopic visual areas and occipito-temporal areas that were selectively responsive to face images (FFA) or house images (PPA). They found that maximal activation occurred around 18 Hz in early visual areas V1–V3, at ~ 9 Hz in V4, but at only 4–5 Hz in FFA and PPA. In a complementary study (Hasson, Yang, Vallines, Heeger, & Rubin, 2008), silent films were temporally scrambled by cutting them into time-segments of varying duration and randomizing the order of presentation. Activation in later visual areas was maximal for longer segments, suggesting that high-level areas integrate information over long time periods. (Gauthier, Eger, Hesselmann, Giraud, & Kleinschmidt, 2012) alternated a single face image with a single house image using rates between 1.2 and 10 Hz. They found a progressive decrease in the optimal frequency of presentation going from V1 to the lateral occipital complex (LOC) to FFA and PPA. What is clear from this collection of studies is that temporal response properties slow down at higher stages in the visual system, and that these response properties place fundamental constraints on perception. This suggests a simple hypothesis: responses to stimuli (or features) represented at similar levels of the visual hierarchy will have similar temporal dynamics.

Here we use evoked potential measures of temporal processing as a means to compare the temporal limits of word and face selective cortex. This choice is motivated by the fact that word and face selective regions are immediately adjacent (within a few millimeters) on the ventral surface (Dehaene et al., 2010; Wandell et al., 2012; Yeatman, Rauschecker, & Wandell, 2013). We therefor might expect these regions to share equivalent tuning properties even though the computations required to read a word (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Seidenberg & McClelland, 1989) are certainly very different from the computations required to recognize a face (Meyers, Borzello, Freiwald, & Tsao, 2015). In support of the hypothesis that there is a canonical temporal processing profile in adjacent category-selective regions, both words and faces produce a characteristic Event-related Potential (ERP) at comparable latencies (150–170 milliseconds) after the presentation of the visual stimuli (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Maurer, Brandeis, & McCandliss, 2005). Although the N150–170 for words and faces each have distinct scalp topographies (Rossion et al 2003), the temporal similarity between their ERP responses could be hypothesized to reflect consistent temporal tuning properties of neurons across ventral temporal cortex: if one makes the assumption that the ERP is equivalent to the impulse response of a linear system, then one would predict that the temporal tuning of faces and text should be very similar, given the similarity in the latency of the selective activity in the two tasks. An alternative hypothesis is that temporal response properties depend substantially on the

specific nature of the computations that the visual system performs on different categories of stimuli, such as words and faces.

The present study uses steady state visual evoked potentials (SSVEPs) to test the hypothesis that there are canonical temporal response properties for regions at the same level of the visual hierarchy (for a recent review of the SSVEP approach see (Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015)). Using the SSVEP, we assessed the temporal frequency tuning preference, the temporal resolution and the apparent latency of word and face selective cortex. Despite the similarity of the N170 response to words and faces, we find markedly distinct temporal properties for the two categories of stimuli.

Methods

Participants

11 adults (4 female) between the ages of 18 and 56 participated. They had normal visual acuity and were screened for neurological and cognitive impairments. Each participant provided written informed consent under a protocol that conformed to the tenets of the Declaration of Helsinki that was approved by the Institutional Review Board of Stanford University.

Stimuli

The text image comprised a block of common English words derived from the MCWord database (http://www.neuro.mcw.edu/mcword/). The face image comprised a black and white photograph of a cropped female head and face placed on a random texture background. Images extended 12 degrees in each direction from a fixation cross in the center of the screen. To provide a comparison stimulus with the same low- and mid-level image statistics, each image was scrambled using the algorithm developed by Portilla and Simoncelli (Portilla & Simoncelli, 2000), which is available at http://www.cns.nyu.edu/~lcv/texture/. The algorithm learns the joint distribution of filter locations, orientations and scales from the image (separate distributions were computed for the text and face images) and preserves this histogram in the synthesized, scrambled version. Stimuli are shown in Figure 1.

Intact and scrambled versions of the stimuli were presented in temporal alternation at rates of 7 frequencies: 1, 2, 3, 4, 6, 9 and 12 Hz. These frequencies were chosen based on prior SSVEP work on faces (Alonso-Prieto, Belle, Liu-Shuang, Norcia, & Rossion, 2013) because we expected: (a) the amplitude of the odd harmonic to drop close to the noise floor by 12 Hz, and (b) more rapid changes in amplitude as a function of frequency at low compare to high frequencies motivating a more dense sampling of lower frequencies (1 to 6 Hz). Observers were given a fixation mark in the center of the image and were instructed to hold their fixation on the mark and to refrain from blinking. The image sequences were presented for 12 sec, with the first and last seconds being excluded from the analysis of the SSVEP. Five trials were run for each temporal frequency and image type with the stimuli presented in random order.

EEG recording and SSVEP analysis

EEG was recorded over 128 channels at a sampling rate of 500 Hz using HydroCell SensorNets connected to an Electrical Geodesics NetAmp 300 running NetStation 4.3 software. Data analysis was performed offline using in house software after exporting the data and digital bandpass filtering between 0.3 and 200 Hz.

The Steady-State Visual Evoked Potential was extracted from the individual 10 second trials by first calculating a time average of five 2 sec consecutive bins of the original 10 sec trial record, yielding a spectrum resolution of 0.5 Hz. The Fourier coefficients at the first harmonics were then averaged coherently to determine the amplitude and phase of the response for each stimulus condition for each participant. Previous work (Liu-Shuang, Ales, Rossion, & Norcia, 2015) has shown that the first harmonic of the SSVEP to alternations between intact and scrambled natural images is generated predominantly by responses to the higher-order configural information in the natural image. An estimate of SSVEP delay (*d*) with respect to the stimulus was calculated from the slope of the phase versus frequency function (Lopes da Silva, van Rotterdam, Storm van Leeuwen, & Tielen, 1970)

$$d = 1/360 * \delta\phi/\delta f$$

where δ_ϕ is the change in phase over the frequency range in degrees and δf is the change in frequency in Hz. This relationship is derived from the properties of physical systems that are "causal", or those whose output can only come after rather than before the input. In such systems, the real and imaginary components are tied together via the Kramers-Kronig relationship – knowing the real component at a given frequency implies knowing the imaginary component and a related relationship discovered by Bode in 1937 that ties the shape of the system gain/amplitude *vs* frequency function to the slope of the phase *vs* frequency function (see (Bechoefer, 2011) for review).

Results

Word and face selective responses have different temporal tuning curves

We find that word and face selective responses each have a unique temporal tuning curve, preferred stimulus frequency and scalp topography (Figure 2). Word selective cortex shows a peak response to text presented a 1Hz and the amplitude of the response declines monotonically as a function of presentation frequency. The brain no longer tracks the change from scrambled to intact text at presentation frequencies above 9Hz. Face selective cortex shows a peak response to faces presented at 4Hz and the amplitude of the response declines for slower or faster presentation frequencies. For faces, the response is equivalent for 1Hz and 6Hz presentation rates. Both word and face selective regions show equivalent and minimal responses to stimuli presented at 9Hz. The left-lateralized scalp topography for words, goes from being nearly equal for the two hemispheres at 1 Hz, to being strongly left lateralized at 4 Hz. The right hemisphere word-response declines more rapidly as a function of presentation rate than the left hemisphere word-response (Figure 2). By contrast, the response to the face image is almost equal for both hemispheres (with a slight right

hemisphere preference) at all frequencies where it is measurable, and there is not a substantial change in lateralization at different presentation rates.

Latency topography demonstrates two distinct sources at two different times

By comparing SSVEP phase values across temporal frequencies we derived latency estimates for responses to face and word images (see equation 1). In a linear-time invariant (LTI) system, there is a linear relationship between the phase and frequency of a signal. This linear relationship indicates that all frequencies are delayed by the same constant amount (constant group delay). Consistent with the underlying model assumption of a LTI system, the phase *vs* frequency functions are linear for text and face stimuli. They differ, however, in slope, with the inferred delay differing by region and stimulus category. By mapping delay over the sensor array, it is apparent that both words and faces show two distinct latencies (Figure 3). This observation suggests the existence of at least two different underlying sources. In the occipito-temporal ROIs the shortest delay for text is 140.0 +/− 6.6 msec but is 159.4 +/−3.0 msec for the face image. A longer latency source is apparent over left occipito-temporal cortex for the text stimuli with a latency of 257.6 +/− 7.1 msec. For the face stimuli, longer latency activity is present over right anterior temporal cortex at a latency of 287.9 +/− 11.8 msec.

Discussion

By measuring both the amplitude and phase of the SSVEP as a function of temporal frequency, we derive a richer description of the dynamics of word and face processing than has been possible with traditional ERP measurements, PET or fMRI. From our measurements we determined that temporal acuity, peak response frequency and delay, each differ for text and face images. These differences in temporal tuning profiles might be surprising considering: (a) word and face selective ERPs have been described to have a similar time delay (Cao et al., 2014; Pegna, Khateb, Michel, & Landis, 2004; Rossion et al., 2003); (b) word and face selective regions are immediately adjacent on the ventral surface of the cortex (Dehaene et al., 2010; Wandell et al., 2012; Yeatman et al., 2013); (c) word and face selective regions have been hypothesized to share a common neuronal architecture (Dehaene & Cohen, 2007; Dehaene et al., 2010).

Differences in temporal tuning profiles reflect differences in the nature of the computations performed by word and face selective cortex. Despite the close spatial proximity of these regions, our measurements suggest that there must be substantial differences in either the neuronal architecture of, or the hierarchy of regions that feed signals into, word and face selective cortex. We find that temporal acuity for faces is substantially higher than for text -- the amplitude of the face-selective response at 4–6 Hz is several times higher than the text-selective response. Hence, regions that process faces are more sensitive to rapidly changing stimuli than regions that process text. This observation predicts that perceptual decisions will show markedly different time courses for words and faces.

Previous work has found that the differential SSVEP response to changing-identity faces *vs* constant-identity faces is maximal at 6 Hz (Alonso-Prieto, Belle, Liu-Shuang, Norcia, & Rossion, 2013). One interpretation of this peak frequency is that it is due to the linear

superposition of transient ERPs with a latency of 150–170ms. However, it is important to note the latency of ERP is influenced by two factors: (1) integration time, or the amount of time required for a brain region to process the incoming information and reach a maximal response, and (2) conduction delay, or the amount of time required for the signal to reach this brain region. Hence, the similar ERP latency for words and faces does not by itself indicate that temporal processing is equivalent in word and face selective cortex.

Here we find the best temporal frequencies for driving cortical responses are substantially lower for text (1 Hz) than for face (4 Hz) images. A direct tying of these peak frequencies to transient response latencies via the superposition model would predict latencies of 1000 msec for transient ERPs to words and 250 msec for face responses. These predicted latencies are clearly inconsistent with the common 150–170 msec ERP latency for both stimulus categories (Cao et al., 2014; Pegna, Khateb, Michel, & Landis, 2004; Rossion et al., 2003). This finding shows that under a different set of measurement conditions, the temporal aspects of the signal in word and face selective cortex can be substantially different despite previous reports noting similarities between the ERP waveform.

Finally, in addition to the mixture of fixed conduction delays and integration delays inherent in visual processing, the visual system is also manifestly non-linear and the conditions under which SSVEP measurements are made --- temporally dense stimuli --- are very different than the temporally sparse conditions used to measure ERP parameters. The presence of temporal non-linearities, such as adaptation, also make it difficult to make direct predictions in the absence of a full non-linear model of the system response. Here we used the first harmonic of the evoked response as a proxy measure and found the phase-frequency relationship to be linear and thus were able to calculate and aggregate delay measure for the two stimulus classes we used.

This is the first EEG study to use the Portilla-Simoncelli algorithm (Portilla & Simoncelli, 2000) to create the baseline condition against which the object-level response is compared. This algorithm preserves a set of higher-order, joint statistics that are lost when the phase of the power spectrum is scrambled. Our paradigm thus isolates responses (at the first harmonic) to text and face images that are higher-order than those driven by the power spectrum of the image. They are also higher-order than responses driven by the joint statistics encoded by the Portilla and Simoncelli algorithm. Previous work in macaque (Rust & Dicarlo, 2010) has found that responses in IT differ between intact and scrambled versions of the same image to a greater degree than do the responses in V4 when the Portilla-Simoncelli algorithm is used. A recent report using fMRI in humans (Freeman, Ziemba, Simoncelli, & Movshon, 2013; Movshon & Simoncelli, 2014) has contrasted responses to Portilla-Simoncelli scrambled textures and intact natural textures and found differential responses occurred only at and beyond area V4. Our approach may thus make the resulting SSVEP more selective to the intrinsic structure of orthography and faces than other approaches such as phase scrambling.

By mapping the temporal delay over the electrode array, we find evidence for multiple underlying sources on the basis of significantly different response delays. It is interesting to note that even these long latency sources continue to respond to steady state stimulation. In

the case of the text response, longer latency activity may reflect increasingly complex orthographic processing. In the case face-related activity, the long latency responses over right anterior temporal cortex may arise in the "extended" face network (Haxby, Hoffman, & Gobbini, 2000) that includes anterior infero-temporal cortex (Kriegeskorte, Formisano, Sorger, & Goebel, 2007). Consistent with this interpretation, intra-cranial recordings with similar stimuli have found SSVEP responses to face images in AIT (Liu-Shuang, Jonas, et al., 2015). Previous transient ERP studies have found a negativity around 250 msec for face stimuli (Schweinberger, Huddy, & Burton, 2004; Schweinberger, Pickering, Jentzsch, Burton, & Kaufmann, 2002) and for objects such as birds or cars after expertise training (Scott, Tanaka, Sheinberg, & Curran, 2006, 2008). These responses are sensitive to repetition and familiarity effects that are not seen in the N170 response. Our approach may be tapping a similar process, as both faces and text are highly overlearned stimuli in typical adults.

Conclusions

SSVEPs represent a promising approach for characterizing the temporal dynamics of high-level visual regions that are selective for text, faces and other important visual categories. Temporal tuning curves can be reliably estimated from relatively short stimulation paradigms, opening the possibility of studying changes in neural dynamics over the course of development (e.g., learning to read) and in the case of developmental disorders (e.g., dyslexia and prosopagnosia). Our measurements clearly demonstrate that the temporal dynamics of word- versus face-selective cortex differ substantially, laying the foundation for models that relate temporal processing to perception and behavior.

Acknowledgments

Literature Cited

Alonso-Prieto E, Belle GV, Liu-Shuang J, Norcia AM, Rossion B. The 6 Hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex. Neuropsychologia. 2013; 51(13):2863–2875. [PubMed: 24007879]

Apkarian P. Temporal frequency responsivity shows multiple maturational phases: state-dependent visual evoked potential luminance flicker fusion from birth to 9 months. Vis Neurosci. 1993; 10(6): 1007–1018. [PubMed: 8257658]

Barlow HB, Blakemore C, Pettigrew JD. The neural mechanism of binocular depth discrimination. J Physiol. 1967; 193(2):327–342. [PubMed: 6065881]

Battelli L, Cavanagh P, Martini P, Barton JJ. Bilateral deficits of transient visual attention in right parietal patients. Brain. 2003; 126(Pt 10):2164–2174. [PubMed: 12821511]

Bechoefer J. Kramers-Kronig, Bode, and the meaning of zero. Amer J Physics. 2011; 79(10):1053–1059.

Bentin S, Allison T, Puce A, Perez E, McCarthy G. Electrophysiological Studies of Face Perception in Humans. J Cogn Neurosci. 1996; 8(6):551–565. [PubMed: 20740065]

Bentin S, Mouchetant-Rostaing Y, Giard MH, Echallier JF, Pernier J. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. J Cogn Neurosci. 1999; 11(3):235–260. [PubMed: 10402254]

Cao X, Jiang B, Gaspar C, Li C. The overlap of neural selectivity between faces and words: evidences from the N170 adaptation effect. Exp Brain Res. 2014; 232(9):3015–3021. [PubMed: 24854017]

Chung ST, Mansfield JS, Legge GE. Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. Vision Res. 1998; 38(19):2949–2962. [PubMed: 9797990]

Cohen L, Lehericy S, Chochon F, Lemer C, Rivaud S, Dehaene S. Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. Brain. 2002; 125(Pt 5):1054–1069. [PubMed: 11960895]

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol Rev. 2001; 108(1):204–256. [PubMed: 11212628]

De Valois RL, Albrecht DG, Thorell LG. Spatial frequency selectivity of cells in macaque visual cortex. Vision Res. 1982; 22(5):545–559. [PubMed: 7112954]

Dehaene S, Cohen L. Cultural recycling of cortical maps. Neuron. 2007; 56(2):384–398. [PubMed: 17964253]

Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, et al. How learning to read changes the cortical networks for vision and language. Science. 2010; 330(6009):1359–1364. [PubMed: 21071632]

Dejerine J. Su un ca de cecite verbale avec agraphie, suivi d'autopsie. Memoires la Societe Biol. 1891; 3:197–201.

Freeman J, Ziemba C, Simoncelli EP, Movshon JA. Functionally partioning the ventral stream with controlled natural stimuli. Soc Neurosci Abstr. 2013 406.401 Online.

Gaillard R, Naccache L, Pinel P, Clemenceau S, Volle E, Hasboun D, et al. Direct intracranial, FMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. Neuron. 2006; 50(2):191–204. [PubMed: 16630832]

Gauthier B, Eger E, Hesselmann G, Giraud AL, Kleinschmidt A. Temporal tuning properties along the human ventral visual stream. J Neurosci. 2012; 32(41):14433–14441. [PubMed: 23055513]

Gentile F, Rossion B. Temporal frequency tuning of cortical face-sensitive areas for individual face perception. Neuroimage. 2014; 90:256–265. [PubMed: 24321556]

Grill-Spector K, Weiner KS. The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci. 2014; 15(8):536–548. [PubMed: 24962370]

Hasson U, Levy I, Behrmann M, Hendler T, Malach R. Eccentricity bias as an organizing principle for human high-order object areas. Neuron. 2002; 34(3):479–490. [PubMed: 11988177]

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. A hierarchy of temporal receptive windows in human cortex. J Neurosci. 2008; 28(10):2539–2550. [PubMed: 18322098]

Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends Cogn Sci. 2000; 4(6):223–233. [PubMed: 10827445]

Holcombe AO. Seeing slow and seeing fast: two limits on perception. Trends Cogn Sci. 2009; 13(5): 216–221. [PubMed: 19386535]

Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962; 160:106–154. [PubMed: 14449617]

Hubel DH, Wiesel TN. Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. J Neurophysiol. 1965; 28:229–289. [PubMed: 14283058]

Huth AG, Nishimoto S, Vu AT, Gallant JL. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron. 2012; 76(6):1210–1224. [PubMed: 23259955]

Jonas J, Descoins M, Koessler L, Colnat-Coulbois S, Sauvee M, Guye M, et al. Focal electrical intracerebral stimulation of a face-sensitive area causes transient prosopagnosia. Neuroscience. 2012; 222:281–288. [PubMed: 22813996]

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 1997; 17(11):4302–4311. [PubMed: 9151747]

Kelly DH. Flicker fusion and harmonic analysis. J Opt Soc Am. 1961a; 51:917–918. [PubMed: 13752373]

Kelly DH. Visual response to time-dependent stimuli. I. Amplitude sensitivity measurements. J Opt Soc Am. 1961b; 51:422–429. [PubMed: 13752375]

Kriegeskorte N, Formisano E, Sorger B, Goebel R. Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A. 2007; 104(51):20600–20605. [PubMed: 18077383]

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A. 1992; 89(12):5675–5679. [PubMed: 1608978]

Levy I, Hasson U, Avidan G, Hendler T, Malach R. Center-periphery organization of human object areas. Nat Neurosci. 2001; 4(5):533–539. [PubMed: 11319563]

Liu-Shuang J, Ales JM, Rossion B, Norcia AM. The effect of contrast polarity reversal on face detection: evidence of perceptual asymmetry from sweep VEP. Vision Res. 2015; 108:8–19. [PubMed: 25595857]

Liu-Shuang, J.; Jonas, J.; Ales, J.; Norcia, A.; Maillard, L.; Rossion, B. Relative sensitivity to low- vs. high-level visual properties in face-sensitive regions of the human ventral occipito-temporal cortex: evidence from intra-cerebral recording. Paper presented at the Vision Sciences Society; 2015.

Lochy A, Van Belle G, Rossion B. A robust index of lexical representation in the left occipito-temporal cortex as evidenced by EEG responses to fast periodic visual stimulation. Neuropsychologia. 2014; 66C:18–31. [PubMed: 25448857]

Lopes da Silva FH, van Rotterdam A, Storm van Leeuwen W, Tielen AM. Dynamic characteristics of visual evoked potentials in the dog. II. Beta frequency selectivity in evoked potentials and background activity. Electroencephalogr Clin Neurophysiol. 1970; 29(3):260–268. [PubMed: 4195648]

Maurer U, Brandeis D, McCandliss BD. Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. Behav Brain Funct. 2005; 1:13. [PubMed: 16091138]

McKeeff TJ, Remus DA, Tong F. Temporal limitations in object processing across the human ventral visual pathway. J Neurophysiol. 2007; 98(1):382–393. [PubMed: 17493920]

McMains SA, Somers DC. Multiple spotlights of attentional selection in human visual cortex. Neuron. 2004; 42(4):677–686. [PubMed: 15157427]

Meyers EM, Borzello M, Freiwald WA, Tsao D. Intelligent information loss: the coding of facial identity, head pose, and non-face information in the macaque face patch system. J Neurosci. 2015; 35(18):7069–7081. [PubMed: 25948258]

Movshon JA, Simoncelli EP. Representation of Naturalistic Image Structure in the Primate Visual Cortex. Cold Spring Harb Symp Quant Biol. 2014; 79:115–122. [PubMed: 25943766]

Movshon JA, Thompson ID, Tolhurst DJ. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. J Physiol. 1978; 283:101–120. [PubMed: 722570]

Mukamel R, Harel M, Hendler T, Malach R. Enhanced temporal non-linearities in human object-related occipito-temporal cortex. Cereb Cortex. 2004; 14(5):575–585. [PubMed: 15054073]

Norcia AM, Appelbaum LG, Ales JM, Cottereau BR, Rossion B. The steady-state visual evoked potential in vision research: A review. J Vis. 2015; 15(6):4. [PubMed: 26024451]

Ozus B, Liu HL, Chen L, Iyer MB, Fox PT, Gao JH. Rate dependence of human visual cortical response due to brief stimulation: an event-related fMRI study. Magn Reson Imaging. 2001; 19(1): 21–25. [PubMed: 11295342]

Parvizi J, Jacques C, Foster BL, Witthoft N, Rangarajan V, Weiner KS, et al. Electrical stimulation of human fusiform face-selective regions distorts face perception. J Neurosci. 2012; 32(43):14915–14920. [PubMed: 23100414]

Pegna AJ, Khateb A, Michel CM, Landis T. Visual recognition of faces, objects, and words using degraded stimuli: where and when it occurs. Hum Brain Mapp. 2004; 22(4):300–311. [PubMed: 15202108]

Portilla J, Simoncelli EP. A parametric texture model basedon joint statistics of complex wavelet coefficients. Internation Journal of Computer Vision. 2000; 40(1):49–71.

Potter MC, Faulconer BA. Time to understand pictures and words. Nature. 1975; 253(5491):437–438. [PubMed: 1110787]

Rossion B, Joyce CA, Cottrell GW, Tarr MJ. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. Neuroimage. 2003; 20(3):1609–1624. [PubMed: 14642472]

Rossion B. Understanding individual face discrimination by means of fast periodic visual stimulation. Exp Brain Res. 2014; 232(6):1599–1621. DOI: 10.1007/s00221-014-3934-9 [PubMed: 24728131]

Rust NC, Dicarlo JJ. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J Neurosci. 2010; 30(39):12978–12995. [PubMed: 20881116]

Schweinberger SR, Huddy V, Burton AM. N250r: a face-selective brain response to stimulus repetitions. Neuroreport. 2004; 15(9):1501–1505. [PubMed: 15194883]

Schweinberger SR, Pickering EC, Jentzsch I, Burton AM, Kaufmann JM. Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. Brain Res Cogn Brain Res. 2002; 14(3):398–409. [PubMed: 12421663]

Scott LS, Tanaka JW, Sheinberg DL, Curran T. A reevaluation of the electrophysiological correlates of expert object processing. J Cogn Neurosci. 2006; 18(9):1453–1465. [PubMed: 16989547]

Scott LS, Tanaka JW, Sheinberg DL, Curran T. The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. Brain Res. 2008; 1210:204–215. [PubMed: 18417106]

Seidenberg MS, McClelland JL. A distributed, developmental model of word recognition and naming. Psychol Rev. 1989; 96(4):523–568. [PubMed: 2798649]

Thomas CG, Menon RS. Amplitude response and stimulus presentation frequency response of human primary visual cortex using BOLD EPI at 4 T. Magn Reson Med. 1998; 40(2):203–209. [PubMed: 9702702]

Wandell BA, Rauschecker AM, Yeatman JD. Learning to see words. Annu Rev Psychol. 2012; 63:31–53. [PubMed: 21801018]

Yeatman JD, Rauschecker AM, Wandell BA. Anatomy of the visual word form area: adjacent cortical circuits and long-range white matter connections. Brain Lang. 2013; 125(2):146–155. [PubMed: 22632810]

Zhu XH, Kim SG, Andersen P, Ogawa S, Ugurbil K, Chen W. Simultaneous oxygenation and perfusion imaging study of functional activity in primary visual cortex at different visual stimulation frequency: quantitative correlation between BOLD and CBF changes. Magn Reson Med. 1998; 40(5):703–711. [PubMed: 9797153]
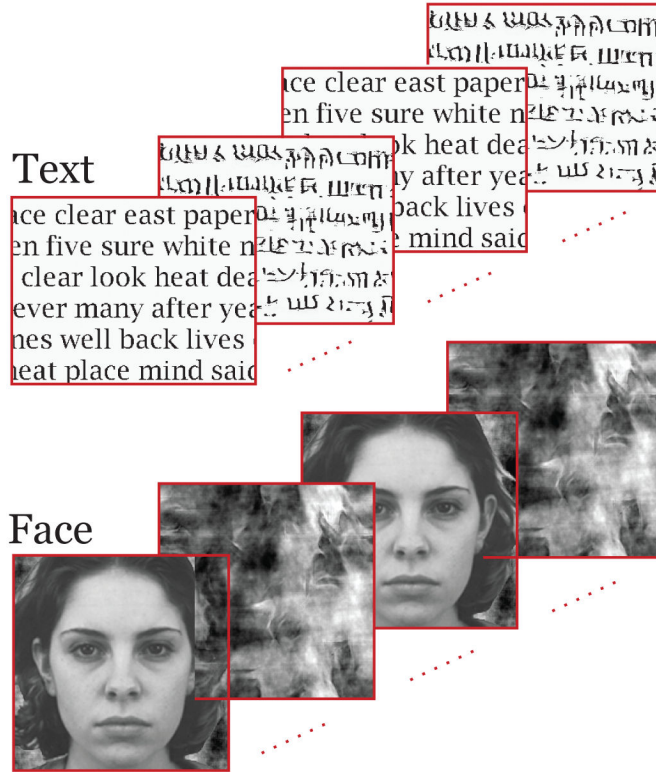
Figure 1. Stimuli
Image sequences consisted of periodic alternations of an intact image (text or face) with a scrambled image whose lower-level statistics were equated to the text or face image, respectively. Image sequences were presented at seven stimulus frequencies spanning 1 to 12 Hz.
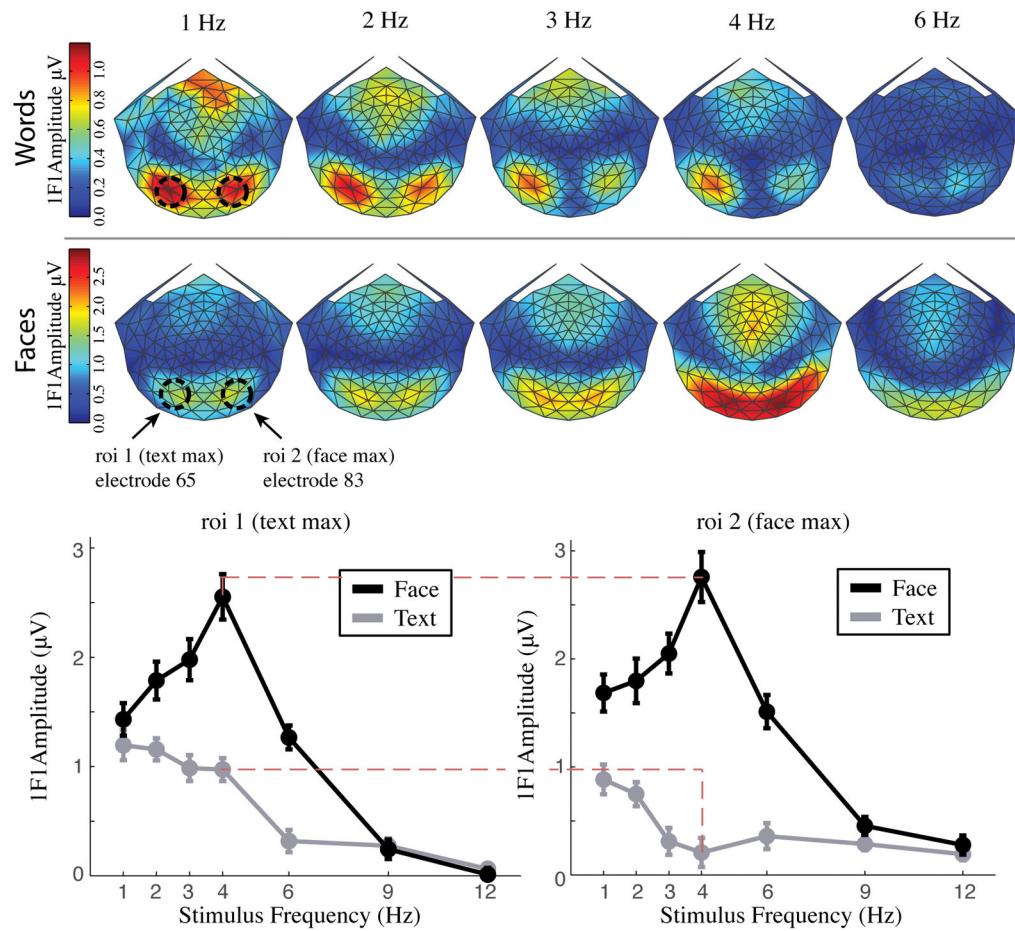
Figure 2. Temporal tuning of word and face selective cortex

Top panel: Scalp topography of the first harmonic response as a function of temporal frequency for text and face image sequences. Dashed lines indicate regions of interest (ROIs) identified on the basis of amplitude maxima in the group average maps. The maximum response to words was at electrode 65 and the maximum response to faces was at electrode 83. The results are very similar for electrode 90 (right hemisphere homologue of electrode 65). **Bottom panel:** Temporal frequency tuning functions for text (gray curves) and face (black curves) stimuli for left (ROI 1) and right (ROI 2) hemisphere ROIs. Face-image tuning functions peaked at 4 Hz, while text-image functions were maximal at 1 Hz. Face-imaging tuning function extends to higher temporal frequencies than do text-image functions.
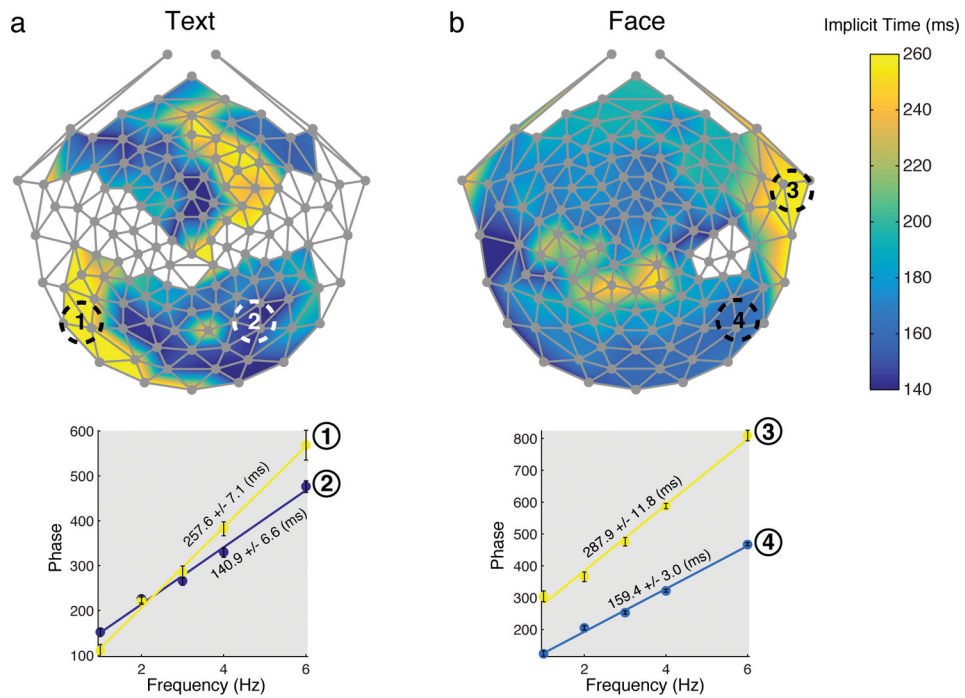
Figure 3. Implicit time for text and face responses
Top panel: Response latency in millisecondsis color coded with cool colors indicating shorter latencies and warm colors longer latencies (see color bar on right). The maps were thresholded to exclude channels where more than one data point was unreliable due to the lack of statistically significant SSVEP responses. Both (a) text and (b) face maps contain regions with more than one delay. **Bottom panel:** SSVEP phase vs frequency plots for text (left) and face (right images) for selected regions of interest indicated by dashed circles. Linear regression fits to the phase versus frequency function are indicated by the solid lines, with the corresponding estimates of latency+/− 1 s.e.m.