Journal of Experimental Psychology:
Human Perception and Performance
2017, Vol. 43, No. 3, 499–508

© 2017 American Psychological Association
0096-1523/17/$12.00 http://dx.doi.org/10.1037/xhp0000313

Spoken Words Can Make the Invisible Visible—Testing the Involvement of Low-Level Visual Representations in Spoken Word Processing

Markus Ostarek
Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands and International Max Planck Research School for Language Sciences, Nijmegen, the Netherlands

Falk Huettig
Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands and Radboud University

The notion that processing spoken (object) words involves activation of category-specific representations in visual cortex is a key prediction of modality-specific theories of representation that contrasts with theories assuming dedicated conceptual representational systems abstracted away from sensorimotor systems. In the present study, we investigated whether participants can detect otherwise invisible pictures of objects when they are presented with the corresponding spoken word shortly before the picture appears. Our results showed facilitated detection for congruent ("bottle" → picture of a bottle) versus incongruent ("bottle" → picture of a banana) trials. A second experiment investigated the time-course of the effect by manipulating the timing of picture presentation relative to word onset and revealed that it arises as soon as 200–400 ms after word onset and decays at 600 ms after word onset. Together, these data strongly suggest that spoken words can rapidly activate low-level category-specific visual representations that affect the mere detection of a stimulus, that is, what we see. More generally, our findings fit best with the notion that spoken words activate modality-specific visual representations that are low level enough to provide information related to a given token and at the same time abstract enough to be relevant not only for previously seen tokens but also for generalizing to novel exemplars one has never seen before.

Public Significance Statement

Understanding the cognitive and neural underpinnings of language comprehension is a key problem in the field of cognitive science. The present study advances our insight into these mechanisms by demonstrating that spoken words referring to concrete object words activate visual processes involved in the earliest stages of conscious perception. Our data suggest that basic visual processes are recruited for language comprehension and support the view that conceptual knowledge at least partly relies on sensory brain systems.

Keywords: language processing, semantic processing, spoken word comprehension, modality-specific representations

In the last decade, much of the research on the mechanisms underlying language comprehension has focused on the role of modality-specific representations that have traditionally been implicated with vision, hearing, planning and executing movements, and so forth. A substantial amount of evidence is consistent with the idea that sensorimotor systems are a component in the process chain related to conceptual access in language comprehension (Barsalou, 2008). Numerous studies have found modality-specific activity (Binder & Desai, 2011; Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008; Martin, Wiggs, Ungerleider, & Haxby, 1996), such that conceptual processing involves activation of the visual cortex (Lewis & Poeppel, 2014), auditory regions in superior posterior and middle temporal gyri (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008), gustatory regions (Simmons, Martin, & Barsalou, 2005), primary olfactory cortex (González et al., 2006), or motor regions (Pulvermüller, Shtyrov, & Ilmoniemi, 2005; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014). Based on these findings it has been proposed that perceptual simulation is a prime mechanism for language comprehension (e.g., Barsalou, 2008; Pulvermüller, 2005), or at least an important part of it. Several recent review papers, however, have pointed out (rightly, in our view) that theoretical progress has been limited and some of the fundamental questions remain unclear (Binder & Desai, 2011;

This article was published Online First January 12, 2017.

Markus Ostarek, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands and International Max Planck Research School for Language Sciences, Nijmegen, the Netherlands; Falk Huettig, Max Planck Institute for Psycholinguistics and Donders Institute for Brain, Cognition, and Behavior, Radboud University.

We would like to thank Gary Lupyan for his helpful comments, Laura Geurts and Jessica Aguilar Diaz for their help during data acquisition, and Antje Meyer for fruitful discussions in early stages of the project.

Correspondence concerning this article should be addressed to Markus Ostarek, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands. E-mail: markus.ostarek@mpi.nl

Hauk & Tschentscher, 2013; Kiefer & Pulvermüller, 2012; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Willems & Casasanto, 2011).

One such question is whether processing of (object) words involves activating category-specific low-level visual representations in visual cortex. This is a key prediction of modality-specific theories of representation, which contrasts with theories assuming dedicated conceptual representational systems abstracted away from sensorimotor systems. Some experimental evidence is consistent with such a prediction. A functional magnetic resonance imaging study found increased activation in Brodmann area (BA) 19 (a visual association area) for visual compared with abstract sentences (Desai, Binder, Conant, & Seidenberg, 2010). Similarly, overtly and covertly producing words related to visually salient concepts activated BA 18 (Hwang, Palmer, Basho, Zadra, & Müller, 2009). Moreover, a recent magnetoencephalography (MEG) study benefiting from excellent temporal resolution found a correlation between the imageability of object words and activation in BA 19 as rapid as ca. 200 ms after word onset (Lewis & Poeppel, 2014). While intriguing, these findings, however, do not tell us much about the nature of the representations that were accessed. To be useful for conceptual processing, visual representations need to be specific enough to provide information related to a given concept and at the same time abstract enough to be relevant not only for previously seen tokens (a signature of episodic memory) but also for recognizing novel exemplars and grasping the concept's meaning more generally. Because of these requirements, it is not obvious that low-level sensory areas (as opposed to high-level areas that are based on more holistic representations) are useful for conceptual processing and can provide this sort of information.

In the present study, we directly tested whether low-level visual representations are involved in spoken word processing. Using continuous flash suppression (CFS) we show that spoken words activate behaviorally relevant low-level visual representations and pin down the time-course of this effect to the first hundreds of milliseconds after word onset.

Thus far, research into object knowledge representation has been largely restricted to high-level visual cortex. While it is uncontroversial that higher level visual areas are strongly implicated in conceptual processing related to objects (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009; Martin, 2007), it remains open what types of representations are involved. On the one hand, several recent studies found that activation patterns in these regions are dominated by strikingly low-level visual features (Andrews, Watson, Rice, & Hartley, 2015; Nasr, Echavarria, & Tootell, 2014; Watson, Hymers, Hartley, & Andrews, 2016). A recent study with macaque monkeys suggests that low-level information may be represented most explicitly in anterior regions such as inferior temporal lobe (IT) rather than visual cortex V4 and V1 (Hong, Yamins, Majaj, & DiCarlo, 2016), such that category-orthogonal features of pictures, such as position, size, pose, and so forth can be decoded much more reliably from IT than early visual cortex. On the other hand, there is evidence from studies with blind individuals suggesting that the properties of high-level visual areas are better described as multimodal (e.g., Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009; for review and discussion, see: Bi, Wang, & Caramazza, 2016), pointing to the possibility that visual information is transduced into conceptual information relatively early in the ventral stream. Though fascinating, the currently available data from research on high-level visual brain regions therefore cannot elucidate to what extent conceptual processing relies on low-level modality-specific visual representations.

A similar issue arises in the context of the available behavioral data. Many studies point to a tight link between conceptual processing of object concepts and visual perception. Words can enhance attention to the corresponding categories in a visual search task (Lupyan & Spivey, 2010b), have a stronger effect on visual categorization than equally informative environmental sounds (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012), and facilitate the perception of shape-matching versus mismatching objects (Zwaan, Stanfield, & Yaxley, 2002; but see Rommers, Meyer, & Huettig, 2013). However, what all of these studies have in common is that the picture targets are not only processed on the visual level, but also the conceptual level. Therefore it cannot be ruled out that the reported effects arose from congruency on a high-level system instead of the visual system. This problem cannot easily be circumvented, as even simple visual tasks, such as categorizing motion patterns as moving up or down (Meteyard, Bahrami, & Vigliocco, 2007) or detecting backward-masked stimuli (Lupyan & Spivey, 2010a) are still subject to semantic processing (Kouider & Dehaene, 2007) and may therefore involve high-level integration processes (Francken, Meijs, Hagoort, van Gaal, & de Lange, 2015; Francken, Meijs, Ridderinkhof, et al., 2015).

In order to find out whether object words activate low-level visual representations, it is necessary to isolate processes that are unequivocally part of *basic* visual processing and that do not tap into high-level systems. This can be achieved by means of a detection paradigm implemented in the binocular rivalry technique called continuous flash suppression (CFS). CFS is typically used in research on subconscious visual processing because of its capacity to render pictures invisible for relatively long periods of time (Tsuchiya & Koch, 2005). CFS disrupts processing of visual stimuli (Gayet, van der Stigchel, & Paffen, 2014; Kang, Blake, & Woodman, 2011; Stein, Hebart, & Sterzer, 2011), such that they elicit only weak activity in visual areas. At the same time it does not hinder other concurrent cognitive tasks, such that, for instance, auditory stimuli can be fully processed. Importantly, CSF can be used to study basic visual detection, which is generally accepted to be a basic visual process. A recent study (Stein, Thoma, & Sterzer, 2015) has scrutinized the factors underlying the detection of pictures in CFS and found that holistic representations played no integral role for this task, further supporting the low-level nature of the process. This makes CFS an ideal tool to study the potential involvement of low-level visual processes in spoken word processing, namely, by testing the effect of words on the mere detection of the presence of a stimulus.

Previous research indicates that it is possible for words to boost otherwise invisible pictures into awareness (Forder, Taylor, Mankin, Scott, & Franklin, 2016; Lupyan & Ward, 2013; Pinto, van Gaal, de Lange, Lamme, & Seth, 2015). However, these studies cannot be taken as evidence that spoken word comprehension involves rapid activation of low-level visual representations because the paradigms they used did not tap into the time-course typically associated with semantic processing (see Hauk & Tschentscher, 2013 for a discussion of the crucial importance of timing). For instance, Lupyan and Ward (2013) had at least 1 s and Forder et al. (2016) several seconds between word and picture presentation. These authors used such long delays to address a different research question, namely, the influence of expectations on perception, and thus used a set-up that encouraged the use of top-down expectations. Indeed, Pinto and colleagues' (2015) Experiment 3 demonstrated that with such long presentation delays the effect disappears when participants are told that the cue word is not predictive of the target picture. Thus, it is unclear from the above research whether spoken word processing involves the rapid activation of low-level visual representations that is predicted by theories that ascribe an important role to modality-specific representation during processing.

The Present Study

In the present study, we implemented the CFS technique in a priming experiment to test whether spoken (object) words activate low-level visual representations. In particular, we tested whether participants can detect otherwise invisible objects when they are near-simultaneously presented with a spoken word (e.g., the word "bottle") and a suppressed picture (e.g., the picture of a bottle). In contrast to the studies mentioned above, we tested for rapid activation of visual representations by minimizing the delay between word and target picture presentation. Our hypothesis was that spoken object words can activate category-specific low-level visual representations rapidly and in a short time window. Access to category-specific visual representations was operationalized as facilitated detection of nearly invisible pictures that were masked with CFS and only briefly presented (for a duration of 400 ms) shortly after word onset.

Experiment 1

Method

Participants. Twenty-four native Dutch speakers from the local Max Planck Institute (MPI) database took part in the study. All participants had normal hearing and normal or corrected-to-normal vision. Five participants had to be excluded (two because of a detection rate <5%, two because of false alarm rates >50%, one because of technical failure) and were replaced by new participants to reach a total of 24 participants. This number should yield adequate statistical power based on similar previous research (Lupyan & Ward, 2013; Pinto et al., 2015).

Stimuli, apparatus, and procedure. Participants were asked to wear custom-made prism glasses (prism diopter: 10Δ) as well as headphones and to put their head on a chinrest. A separator was placed centrally from the nose of the participant to the computer screen, such that each eye could only see the ipsilateral half of the screen. The stimuli were presented on a computer screen (resolution: 1900 × 720, refresh rate: 60 Hz) using Presentation Software (Version 16.2, www.neurobs.com) in 80 cm viewing distance which ensured optimal interocular fusion in this particular setup. The CFS masks were randomly selected for every trial from a set of 50 Mondrian-type images composed of 1,000 randomly superimposed rectangles of different colors and sizes, which were created in Matlab (similar to Hesselmann, Darcy, Ludwig, & Sterzer, 2016). The 16 target pictures were selected from the normed de Groot, Koelewijn, Huettig, and Olivers (2016) database, converted into grayscale and their edges blurred using a Gaussian filter with a radius of 3 pixels in Adobe Photoshop to facilitate suppression. The 16 spoken cue words (listed in the Appendix, mean length: 500 ms, range: 287–685 ms) were recorded by a female native Dutch speaker and cut into single files with Praat (Boersma, 2002).

At the beginning of each trial (Figures 1 and 2), a central fixation cross was displayed alongside a black rectangular frame (600 × 600 pixels) in both halves of the screen for 500 ms. Next, a spoken word was presented. In picture-present trials, 200 ms after word onset the suppressed picture appeared: the CFS masks filled the black frame on one side (random Mondrian-type colorful rectangular shapes changing at 12 Hz) and a gray-scale picture filled 350 × 350 pixels the other (on half of the trials this side stayed empty as there was no picture). In keeping with Lupyan and Ward (2013), the target pictures were presented to the right eye. The visual stimuli remained on the screen for 400 ms and were then replaced by the question "Was there a picture?" to which participants were instructed to respond immediately by pressing the left or right button. On picture-absent trials, the 350 × 350 pixel area that would otherwise be covered by an image that showed only an empty white square (everything else was identical to picture-present trials). Participants were informed that the words neither predicted whether a picture was presented on a given trial nor what picture appeared in picture-present trials. There were a total of 256 trials (corresponding to 8 repetitions of each of the 16 words and targets), including 128 picture-present trials of which 64 had congruent and 64 had incongruent prime words.

One prerequisite for this study was good control over the suppression strength such that, ideally, the detection rate of target

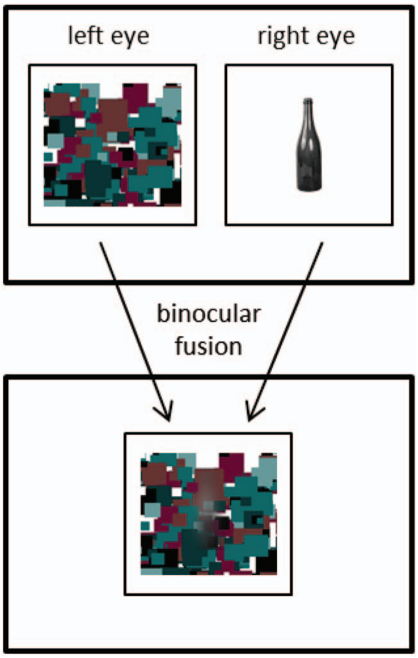

Figure 1. Binocular fusion. A picture was presented to the right eye, while the left eye saw CFS masks changing at ca. 10 Hz. See the online article for the color version of this figure.

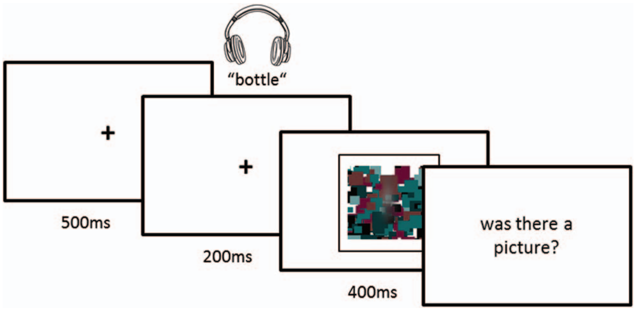

Figure 2. Trial structure. Note that the target picture was presented 200 ms after word onset, but the words extended toward target picture offset (mean word length: 500 ms). See the online article for the color version of this figure.

pictures would be about 50%. To that aim, we first conducted a short staircase procedure (96 trials) that was identical to the main study described above, except that no auditory cues were played and another set of pictures was used. In this staircase procedure, every hit resulted in a slight decrease and every miss in a slight increase of the target picture's contrast, such that the suppression strength usually converges roughly on 50% (as in Lupyan & Ward, 2013, Experiment 3). The final contrast level was then used in the main experiment to counter individual differences in visibility thresholds and avoid ceiling or floor effects.

Analysis. We implemented a design in which all words appeared equally often in picture-present (50%) and picture-absent trials (50%). In picture-present trials, all cue words were followed equally often by congruent (50%) and incongruent targets (50%). Thus, in half of the picture-present trials, the spoken word was congruent with the target picture ("bottle" → picture of a bottle), while on the other half it was incongruent ("bottle" → picture of a banana). We chose this design because it rules out any confounds due to differences in low-level features between conditions, as all picture stimuli are evenly distributed over the experimental conditions. Our main prediction that congruency would lead to facilitated detection was tested in three ways: First, we analyzed detection rates in a binomial mixed-effects model, including congruency (congruent vs. incongruent) as fixed effect and random intercepts and slopes for the effect of congruency by participant and target. Second, we analyzed d'-scores that take into account the false alarm rate each participant produced. This was achieved by calculating mean detection rates in the congruent and incongruent condition as wells as false alarm rates per subject, then calculating d'-scores, and finally comparing the resulting scores in the congruent versus incongruent condition in paired-samples t tests. Third, we analyzed reaction times (RTs) on all picture-present trials with hits using a linear mixed-effects model with the same fixed and random effects structure as the binomial model above. RTs slower than 2.5 s and further than 2.5 *SD* from the resulting mean per condition were removed from the analysis.

Results

Of primary interest was whether congruent word cues facilitate the detection of almost invisible pictures. Inspection of the hit rate means suggests that this is indeed so (see Figure 3); in congruent trials the mean detection rate was 46.7%, whereas in incongruent trials the mean detection rate was only 41.7% (false alarm rate was 15%). A d'-analysis revealed higher d'-scores, $t(23) = 3.008$, $p = .006$ in the congruent ($M = 1.15$, $SD = 0.63$) versus incongruent condition (M = 0.99, $SD = 0.60$; Figure 4). The binomial mixed-effects model further backs up this result by revealing a significant difference in hit rates between the congruent versus incongruent condition (coef. = −0.2724, $SE = 0.1168$, $z = -2.332$, $p = .0197$). Finally, a linear mixed-effects model showed that response latencies in correct hit trials were reliably shorter in the congruent versus incongruent condition (coef. = 46.30, $SE = 17.68$, $t = 2.619$, $p = .009$; Figure 5).

Discussion

As predicted, processing spoken words made it more likely for participants to detect congruent compared with incongruent suppressed pictures. This result conceptually replicates previous CFS studies that showed that words can be used as cues to affect perception (cf., Lupyan & Ward, 2013; Pinto et al., 2015) and a recent EEG study (Boutonnet & Lupyan, 2015) that demonstrated that words (e.g., *cat*) but not equally informative environmental sounds (a meowing cat) modulate the P1 component in a subsequent picture-matching task. Our results go further by demonstrating that spoken word comprehension results in rapid activation of low-level visual representations, as the spoken word and suppressed picture were presented near-simultaneously. We thus conclude that spoken words can activate category-specific visual representations that are low-level enough to influence a process as basic as the mere detection of a stimulus in line with modality-specific theories of conceptual representation. The timing of the spoken word and the following suppressed picture tells us that this effect is present roughly somewhere between 200 and 600 ms post word onset. To further pin down its exact time-course, we conducted a second experiment in which the timing of the suppressed picture relative to the prime word was systematically manipulated.

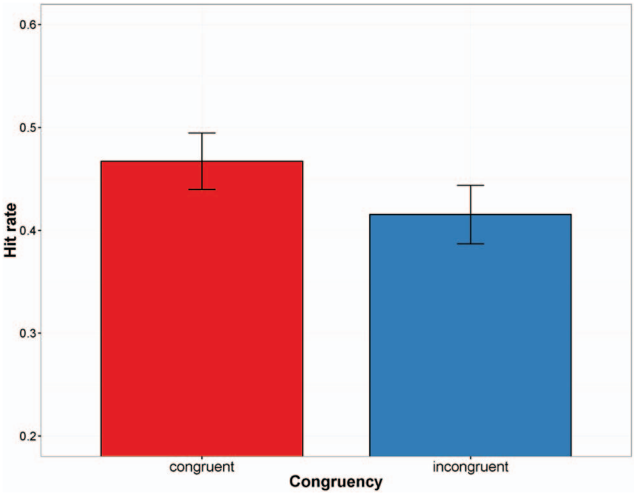

Figure 3. Experiment 1. Mean hit rates in the congruent and incongruent conditions. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

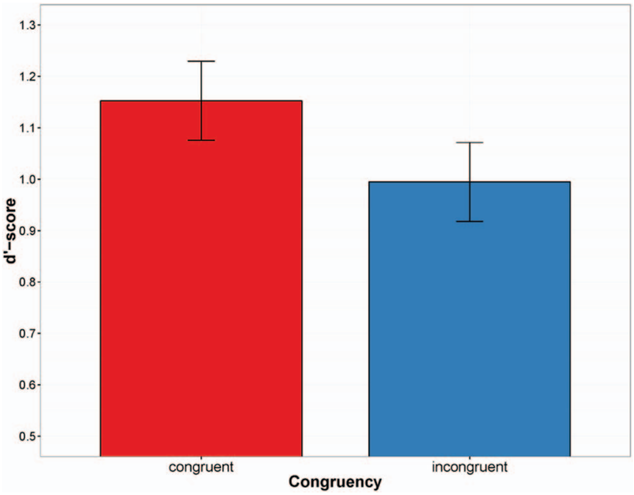

Figure 4. Experiment 1. Mean *d*′-scores in the congruent and incongruent conditions. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

Experiment 2

Experiment 2 was designed to replicate the effect we found in Experiment 1 and to identify its time-course. To that aim, we modified the paradigm slightly such that the suppressed pictures were again presented for 400 ms, but the timing of the picture relative to word onset was manipulated within subjects. We were mainly interested in two aspects: first, we wanted to find out how early visual representations are activated during word processing, as this should be informative about the underlying processing dynamics of language comprehension. We predicted early recruitment (within the first 400 ms), which would be consistent with the idea that the visual system contributes to conceptual processing of spoken words. Second, and related to that matter, we wanted to see how long-lasting the effect is. In the scenario we envisage, spoken words trigger a rapid and short-lived sweep of activation in visual cortex as part of information retrieval related to word meaning. From that the prediction follows that the effect is only present in the first hundreds of milliseconds post word onset and that it then vanishes. At first sight, this prediction may seem to contradict previous demonstrations of late effects of word cues on detection in CFS (Lupyan & Ward, 2013; Pinto et al., 2015). However, as discussed above, because of the differences in timing our experimental design likely taps mechanisms directly related to initial word processing, whereas these previous studies were designed to capture the effect of expectations on perception that may follow a different time-course.

Method

Participants. Thirty-eight new native Dutch speakers with normal hearing and vision from the MPI database participated in the experiment. Five of them had to be excluded from the analysis (3 because of a mean detection rate >90%, 2 because of false alarm rates >40%), leaving us with 33 participants. The increased number of participants was chosen to make sure the study is powerful enough, given that we introduce a new factor (timing) to the design.

Stimuli, apparatus, and design. We used the same stimuli, apparatus, and design as in Experiment 1 with the following exceptions: (a) we introduced the factor timing to the design, which had four levels corresponding to stimulus onset asynchrony (SOA) 1 (−200–200 ms; picture appears 200 ms before until 200 ms after word onset), SOA 2 (0–400 ms; picture appears at word onset until 400 ms after word onset), SOA 3 (200–600 ms; from 200 ms after word onset until 600 ms after word onset), and SOA 4 (600–1,000 ms; from 600 ms after word onset until 1,000 ms after word onset); and (b) to ensure sufficient statistical power we increased the number of trials to 384 and changed the ratio of picture-present versus picture-absent trials to 2:1, as we were primarily interested in detected pictures. This made it possible to repeat each picture twice in each timing condition per prime word. Again, before the main experiment, participants completed the staircase procedure described above to identify individual detection thresholds that were then used during the main experiment.

Analysis. We predicted that congruency between the prime word and the suppressed picture would facilitate detection at SOA 2 and 3, but not at SOA 1 and 4. Regarding SOA 1 (−200 ms), by the time that the word cue becomes informative the picture has already disappeared and it therefore should not have an effect. With respect to SOA 4 (600 ms), we propose that detection performance is not affected by the word primes because it falls outside of the window typically associated with online semantic processing. Category-specific activation in the visual cortex triggered by the prime word should therefore have decayed when the picture appears. To test these hypotheses, we used the same set of analyses as before, adapted for the current design as follows. First, the binomial mixed-effects model included congruency and timing as well as their interaction as fixed effects and random intercepts and slopes for both fixed effects per subjects and per items. To analyze in which time-windows there was an effect, we used the function lsmeans from the lsmeans package that

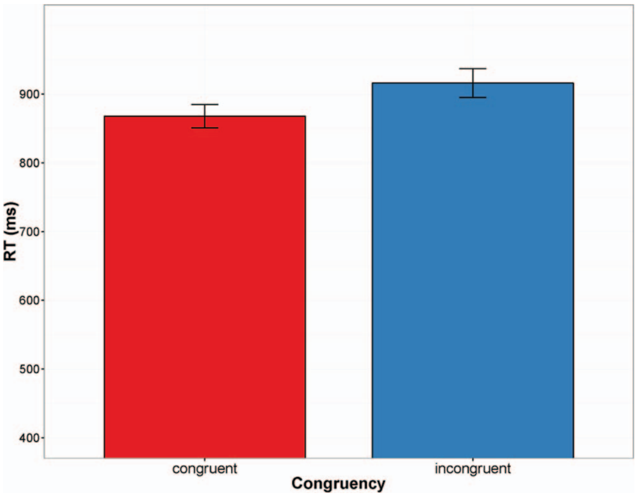

Figure 5. Experiment 1. Mean reaction times (RTs) in the congruent and incongruent conditions on all trials with hits after removing outliers (RTs >2.5 s and >2.5 *SD* from the mean). Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

computes least-squares means (also called predicted marginal means) for all factor combinations and the contrasts among them. Second, we calculated d'-scores and submitted them to a 2 × 4 repeated-measures analysis of variance (ANOVA) for which we predicted a significant interaction. Follow-up paired-samples t tests were used to determine the effect of congruency at all four SOAs. Finally, for the RT data we ran a linear mixed-effects model on all trials with detected pictures excluding responses slower than 2,500 ms, as well as responses further than 2.5 *SD* from the mean calculated per condition. The model had the same fixed and random effects structure as the binomial model above and all effects were analyzed in the same way.

Results and Discussion

Of primary interest was at what SOA congruency affected detection. Figures 6 and 8 suggest that detection was facilitated by congruent words at Timings 2 and 3, but that there was no difference between congruent and incongruent trials at the other two timings. The 2 × 4 ANOVA on d'-scores revealed a main effect of congruency ($F(1, 32) = 14.066$, $p = .001$, $\eta_p^2 = 0.305$) and an interaction between congruency and timing ($F(3, 96) = 3.681$, $p = .015$, $\eta_p^2 = 0.103$). Follow-up paired t tests revealed that, indeed, congruency had no effect on detection sensitivity at SOA 1 (mean d' difference: 0.048, $t(32) < 1$) and SOA 4 (mean difference: 0.045, $t(32) = 1.087$, $p > .2$), while there was a reliable effect at SOA 2 (mean difference: 0.174, $t(32) = 3.349$, $p = .002$) and SOA 3 (mean difference: 1.191, $t(32) = 4.874$, $p < .001$). In line with this, the binomial mixed-effects model on hit rates revealed a significant interaction between congruency and timing ($\chi^2(3) = 7.863$, $p = .048$), characterized by the pattern that hit rates were higher in the congruent versus incongruent condition at SOA 2 (coef. = 0.348, $SE = 0.12$, $p = .003$) and 3 (coef. = 0.415, $SE = 0.122$, $p < .001$), but not at SOA 1 (coef. = 0.081, $SE = 0.121$, $p > .5$) and 4 (coef. = 0.108, $SE = 1.21$, $p > .3$).[1] As a follow-up, we also conducted a logistic growth curve analysis (Mirman, 2016) with a second-order polynomial including fixed-effects of congruency on both time terms and participant and participant-by-condition random effects on the quadratic term (the full model did not converge, so the linear term was removed because it was not expected to capture key differences). As expected from the results above, it revealed a significant main effect of congruency (coef. = −0.178, $SE = 0.049$, $p < .001$) and a significant effect of congruency on the quadratic term (coef. = 0.218, $SE = 0.091$, $p = .016$). Figure 7 shows the model fit. In sum, the hit rate and d'-scores provide converging evidence that our design captured both the onset and offset of the detection facilitation effect (Figures 6, 7, and 8).

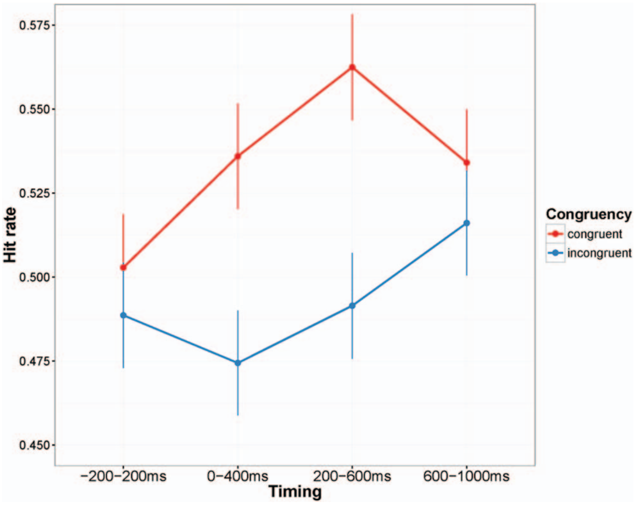

Figure 6. Experiment 2. Mean hit rates in all experimental conditions (congruent vs. incongruent at all four timing conditions). Error bars indicate the standard error. See the online article for the color version of this figure.

RT analyses revealed a main effect of congruency (coef. = 0.144, $SE = 0.063$, $t = 2.287$, $p = .022$), but no interaction with timing ($p > .1$). This result in combination with visual inspection suggests that RTs were influenced by congruency at the earliest and latest SOAs (Figure 9). The most likely explanation for this seems to be that at SOA 1 and 4 Congruency affected participants' confidence in their decisions, rather than perceptual processing. Given the clear results from the hit rate and d' data, we consider our results to provide strong evidence for a time-dependent effect, as described below.

General Discussion

The results of Experiment 2 replicate the effect we observed in Experiment 1 and also revealed the dynamic activation and deactivation of low-level visual representations over time during spoken word processing. In addition to our replication of the effect at Timing 3, where the nearly invisible picture is presented 200–600 ms after word onset, the novel effect at Timing 2 (0–400 ms) in combination with the absence of an effect at 200 ms before until 200 ms after word onset suggests that category-specific low-level visual representations were first accessed at 200–400 ms after word onset. Strikingly, for some of our words this is before the uniqueness point that disambiguates them from other possible words. This is reminiscent of Lewis and Poeppel's (2014) MEG study that found a correlation between imageability and occipital cortex activation (in BA 19) at 160–190 ms post onset, simultaneously or even before typical lexical effects (such as cohort and frequency effects) in temporal areas. A likely explanation for this even quicker activation compared with our study is that these authors only used monosyllabic words consisting of three to four phonemes that were simply shorter than the stimuli used here. Access to visual areas around the uniqueness point is consistent with a role for semantic processing. Furthermore, the absence of an effect at the latest time window (600–1,000 ms after word onset) suggests that the recruitment of low-level visual representations is limited to the first hundreds of milliseconds of spoken word processing. At around word offset, the effect has completely vanished, suggesting a rapid decay after initial rapid activation.

Thus, we observed activation of visual representations in precisely the time-window that is associated with semantic access. Given the behavioral relevance of the activated representations for the detection paradigm we used, we can conclude that spoken words (referring to concrete concepts) activate modality-specific visual representations that are low-level enough to provide information related to the perception of a given token and, at the same time, abstract enough to be relevant not only for previously seen tokens (a signature of episodic memory) but also for generalizing to novel exemplars one has never

[1] False alarm rates were similar across SOAs (8.2%–10.3%).

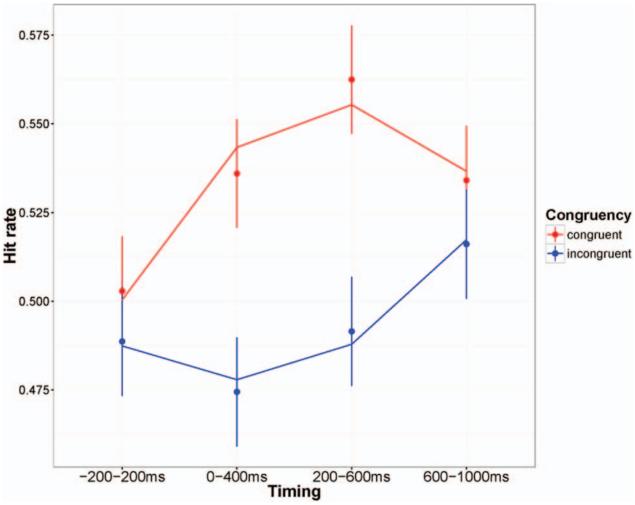

Figure 7. Experiment 2. Model fit of the logistic growth curve analysis depicted by plotting the observed data (dots correspond to the observed means) along with the predicted curve. Error bars indicate the standard error. See the online article for the color version of this figure.

seen before (a signature of semantic memory). A mechanism that could accomplish this is the activation of visual features that are diagnostic of the category the word refers to. Attention to a semantic category can shift tuning of neuron populations in occipital and temporal cortex toward that category (Çukur, Nishimoto, Huth, & Gallant, 2013). Similarly, our results suggest that words may rapidly trigger the activation of low-level visual representations and, as a result, enhance processing of visual items sharing some of these representations.

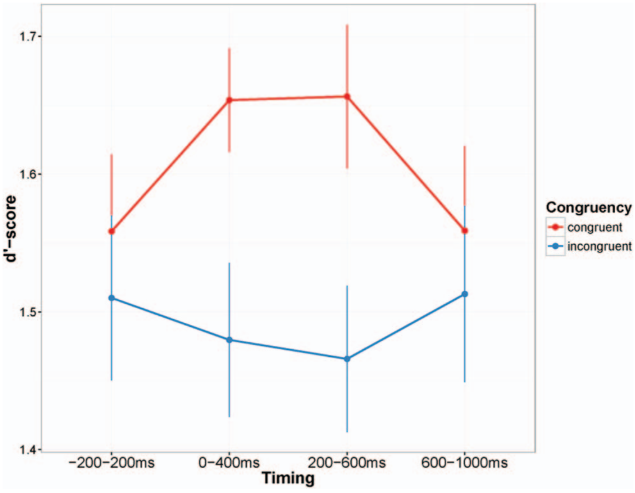

Figure 8. Experiment 2. Mean d'-scores in all experimental conditions (congruent vs. incongruent at all four timing conditions). Error bars indicate the standard error. See the online article for the color version of this figure.

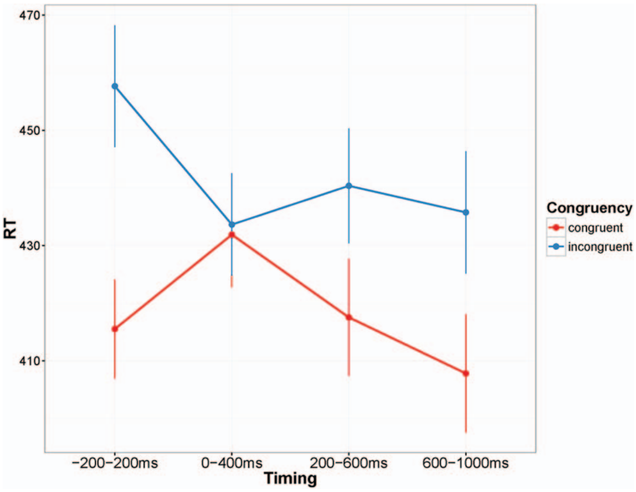

Figure 9. Experiment 2. Mean reaction times (RTs) in all experimental conditions (congruent vs. incongruent at all four timing conditions), calculated on trials with hits after removing outliers (RTs >2.5 s and >2.5 *SD* from the mean). Error bars indicate the standard error. See the online article for the color version of this figure.

Note that our present findings do not contradict the previous CFS studies reporting late cuing effects (e.g., Lupyan & Ward, 2013; Pinto et al., 2015). Their elegant studies add on to the increasing literature on how expectations can affect perception. We would like to point out that both their findings and our present results could be interpreted in a predictive coding framework and may rely on similar neural mechanisms (described in D'Esposito & Postle, 2015; Pratte & Tong, 2014; Scocchia, Valsecchi, & Triesch, 2014). Importantly, however, regarding our research question, the present study demonstrates the rapid involvement of low-level visual representations during *online* spoken word processing in a situation where the words were no reliable cues for the visual targets.

Future work could usefully investigate the underlying processing dynamics in terms of information exchange between sensorimotor systems, on the one hand, and high-level integration systems on the other. There is now good evidence that parts of the parietal lobe (especially the angular gyrus) and/or the anterior temporal lobe (ATL) are consistently involved in conceptual processing encompassing information from multiple modalities and it has been proposed that they could function as hubs that integrate information from distributed cortical sites (Binder & Desai, 2011; Binder et al., 2009; Patterson, Nestor, & Rogers, 2007; Xu, Lin, Han, He, & Bi, 2016). Strong evidence comes from studies showing modality-independent semantic impairments induced by ATL damage (e.g., Mummery et al., 2000) or by transcranial magnetic stimulation on ATL (Pobric, Jefferies, & Lambon Ralph, 2010a, 2010b), suggesting a functional role of ATL for conceptual processing. Using electrocorticography, Chen and colleagues (2016) recently managed to decode semantic information related to concepts from a wide array of categories from ATL activity patterns, supporting its category-general role for semantic processing. Similarly, Correia et al. (2014) were able to decode word meanings from ATL using fMRI.

Moreover, there are some clues as to what neural mechanisms could underlie communication between the different systems. ATL has been shown to engage in increased connectivity with modality-specific visual or auditory regions during semantic processing. When information from multiple modalities has to be combined, theta oscillations in ATL and distributed cortical networks become phaselocked as a function of the featural content required (van Ackeren, 2014, chapters 3 and 6; van Ackeren & Rueschemeyer, 2014; van Ackeren, Schneider, Müsch, & Rueschemeyer, 2014). Coutanche and Thompson-Schill (2015) found that the identity of expected objects participants were thinking about could be decoded from ATL patterns and that decoding accuracy was predicted by activation patterns in visual areas processing shape and color information related to the object, further pointing to the functional link between visual and high-level systems in conceptual processing. In the light of these findings, our data fit with the view that words trigger (potentially simultaneously) the activation of modality-specific as well as modality-independent systems that remain coupled for several hundreds of milliseconds to build a conceptual representation in concert.

References

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision, 15,* 3. http://dx.doi.org/10.1167/15.7.3

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645. http://dx.doi.org/10.1146/annurev.psych.59.103006.093639

Bi, Y., Wang, X., & Caramazza, A. (2016). Object domain and modality in the ventral visual pathway. *Trends in Cognitive Sciences, 20,* 282–290. http://dx.doi.org/10.1016/j.tics.2016.02.002

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15,* 527–536. http://dx.doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19,* 2767–2796. http://dx.doi.org/10.1093/cercor/bhp055

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience, 35,* 9329–9335. http://dx.doi.org/10.1523/JNEUROSCI.5111-14.2015

Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., . . . Lambon Ralph, M. A. (2016). The 'when' and 'where' of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 79,* 1–13. http://dx.doi.org/10.1016/j.cortex.2016.02.015

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: FMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience, 34,* 332–338. http://dx.doi.org/10.1523/JNEUROSCI.1302-13.2014

Coutanche, M. N., & Thompson-Schill, S. L. (2015). Creating concepts from converging features in human cortex. *Cerebral Cortex, 25,* 2584–2593.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience, 16,* 763–770. http://dx.doi.org/10.1038/nn.3381

de Groot, F., Koelewijn, T., Huettig, F., & Olivers, C. N. (2016). A stimulus set of words and pictures matched for visual and semantic similarity. *Journal of Cognitive Psychology, 28,* 1–15.

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M. S. (2010). Activation of sensory–motor areas in sentence comprehension. *Cerebral Cortex, 20,* 468–478.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology, 66,* 115–142. http://dx.doi.org/10.1146/annurev-psych-010814-015031

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition, 143,* 93–100. http://dx.doi.org/10.1016/j.cognition.2015.06.008

Forder, L., Taylor, O., Mankin, H., Scott, R. B., & Franklin, A. (2016). Colour terms affect detection of colour and colour-associated objects suppressed from visual awareness. *PLoS ONE, 11,* e0152212. http://dx.doi.org/10.1371/journal.pone.0152212

Francken, J. C., Meijs, E. L., Hagoort, P., van Gaal, S., & de Lange, F. P. (2015). Exploring the automaticity of language-perception interactions: Effects of attention and awareness. *Scientific Reports, 5,* 17725. http://dx.doi.org/10.1038/srep17725

Francken, J. C., Meijs, E. L., Ridderinkhof, O. M., Hagoort, P., de Lange, F. P., & van Gaal, S. (2015). Manipulating word awareness dissociates feed-forward from feedback models of language-perception interactions. *Neuroscience of Consciousness, 2015,* niv003.

Gayet, S., van der Stigchel, S., & Paffen, C. L. (2014). Breaking continuous flash suppression: Competing for consciousness on the presemantic battlefield. *Frontiers in Psychology, 5,* 460. http://dx.doi.org/10.3389/fpsyg.2014.00460

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage, 32,* 906–912. http://dx.doi.org/10.1016/j.neuroimage.2006.03.037

Hauk, O., & Tschentscher, N. (2013). The body of evidence: What can neuroscience tell us about embodied semantics? *Frontiers in Psychology, 4,* 50. http://dx.doi.org/10.3389/fpsyg.2013.00050

Hesselmann, G., Darcy, N., Ludwig, K., & Sterzer, P. (2016). Priming in a shape task but not in a category task under continuous flash suppression. *Journal of Vision, 16,* 17. http://dx.doi.org/10.1167/16.3.17

Hoenig, K., Sim, E. J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience, 20,* 1799–1814.

Hong, H., Yamins, D. L., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience, 19,* 613–622. http://dx.doi.org/10.1038/nn.4247

Hwang, K., Palmer, E. D., Basho, S., Zadra, J. R., & Müller, R. A. (2009). Category-specific activations during word generation reflect experiential sensorimotor modalities. *NeuroImage, 48,* 717–725. http://dx.doi.org/10.1016/j.neuroimage.2009.06.042

Kang, M. S., Blake, R., & Woodman, G. F. (2011). Semantic analysis does not occur in the absence of awareness induced by interocular suppression. *The Journal of Neuroscience, 31,* 13535–13545. http://dx.doi.org/10.1523/JNEUROSCI.1691-11.2011

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex, 48,* 805–825.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience, 28,* 12224–12230. http://dx.doi.org/10.1523/JNEUROSCI.3579-08.2008

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362,* 857–875.

Lewis, G., & Poeppel, D. (2014). The role of visual representations during the lexical access of spoken words. *Brain and Language, 134,* 1–10. http://dx.doi.org/10.1016/j.bandl.2014.03.008

Lupyan, G., & Spivey, M. J. (2010a). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS ONE, 5,* e11452. http://dx.doi.org/10.1371/journal.pone.0011452

Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics, 72,* 2236–2253. http://dx.doi.org/10.3758/BF03196698

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General, 141,* 170–186. http://dx.doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 14196–14201. http://dx.doi.org/10.1073/pnas.1303312110

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron, 63,* 397–405. http://dx.doi.org/10.1016/j.neuron.2009.07.012

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58,* 25–45. http://dx.doi.org/10.1146/annurev.psych.57.102904.190143

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. http://dx.doi.org/10.1038/379649a0

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science, 18,* 1007–1013. http://dx.doi.org/10.1111/j.1467-9280.2007.02016.x

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 48,* 788–804. http://dx.doi.org/10.1016/j.cortex.2010.11.002

Mirman, D. (2016). *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology, 47,* 36–45. http://dx.doi.org/10.1002/1531-8249(200001)47:1<36::AID-ANA8>3.0.CO;2-L

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *The Journal of Neuroscience, 34,* 6721–6735. http://dx.doi.org/10.1523/JNEUROSCI.4802-13.2014

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8,* 976–987. http://dx.doi.org/10.1038/nrn2277

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision, 15,* 13. http://dx.doi.org/10.1167/15.8.13

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010a). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current Biology, 20,* 964–968. http://dx.doi.org/10.1016/j.cub.2010.03.070

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010b). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia, 48,* 1336–1342. http://dx.doi.org/10.1016/j.neuropsychologia.2009.12.036

Pratte, M. S., & Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *Journal of Vision, 14,* 22. http://dx.doi.org/10.1167/14.3.22

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6,* 576–582. http://dx.doi.org/10.1038/nrn1706

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17,* 884–892. http://dx.doi.org/10.1162/0898929054021111

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science, 24,* 2218–2225. http://dx.doi.org/10.1177/0956797613490746

Scocchia, L., Valsecchi, M., & Triesch, J. (2014). Top-down influences on ambiguous perception: The role of stable and transient states of the observer. *Frontiers in Human Neuroscience, 8,* 979. http://dx.doi.org/10.3389/fnhum.2014.00979

Shtyrov, Y., Butorina, A., Nikolaeva, A., & Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 111,* E1918–E1923. http://dx.doi.org/10.1073/pnas.1323158111

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex, 15,* 1602–1608. http://dx.doi.org/10.1093/cercor/bhi038

Stein, T., Hebart, M. N., & Sterzer, P. (2011). Breaking continuous flash suppression: A new measure of unconscious processing during interocular suppression? *Frontiers in Human Neuroscience, 5,* 167. http://dx.doi.org/10.3389/fnhum.2011.00167

Stein, T., Thoma, V., & Sterzer, P. (2015). Priming of object detection under continuous flash suppression depends on attention but not on part-whole configuration. *Journal of Vision, 15,* 15. http://dx.doi.org/10.1167/15.3.15

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience, 8,* 1096–1101.

van Ackeren, M. J. (2014). *Oscillatory neuronal dynamics during lexical-semantic retrieval and integration* (Doctoral dissertation) University of York, York, United Kingdom.

van Ackeren, M. J., & Rueschemeyer, S. A. (2014). Cross-modal integration of lexical-semantic features during word processing: Evidence from oscillatory dynamics during EEG. *PLoS ONE, 9,* e101042. http://dx.doi.org/10.1371/journal.pone.0101042

van Ackeren, M. J., Schneider, T. R., Müsch, K., & Rueschemeyer, S. A. (2014). Oscillatory neuronal activity reflects lexical-semantic feature integration within and across sensory modalities in distributed cortical networks. *The Journal of Neuroscience, 34,* 14318–14323. http://dx.doi.org/10.1523/JNEUROSCI.0958-14.2014

Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *NeuroImage, 124,* 107–117. http://dx.doi.org/10.1016/j.neuroimage.2015.08.058

Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology, 2,* 116. http://dx.doi.org/10.3389/fpsyg.2011.00116

Xu, Y., Lin, Q., Han, Z., He, Y., & Bi, Y. (2016). Intrinsic functional network architecture of human semantic processing: Modules and hubs. *NeuroImage, 132,* 542–555. http://dx.doi.org/10.1016/j.neuroimage.2016.03.004

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13,* 168–171. http://dx.doi.org/10.1111/1467-9280.00430

(*Appendix follows*)

Appendix

List of Stimuli

Stimuli categories: banana, crane, palm tree, hark, ball, injection, pen, butterfly, brain, bottle, hat, cushion, rocket, flute, (a kind of) sweet.

Corresponding Dutch words: banaan, kraan, palm, hark, ball, spuitje, pen, vlinder, hersens, fles, hoed, kussen, raket, fluit, bonbon.

Received May 12, 2016
Revision received August 15, 2016
Accepted August 28, 2016 ■