# The Behavioral and Neural Effects of Language on Motion Perception

Jolien C. Francken[1], Peter Kok[1], Peter Hagoort[1,2], and Floris P. de Lange[1]

## Abstract

■ Perception does not function as an isolated module but is tightly linked with other cognitive functions. Several studies have demonstrated an influence of language on motion perception, but it remains debated at which level of processing this modulation takes place. Some studies argue for an interaction in perceptual areas, but it is also possible that the interaction is mediated by "language areas" that integrate linguistic and visual information. Here, we investigated whether language–perception interactions were specific to the language-dominant left hemisphere by comparing the effects of language on visual material presented in the right (RVF) and left visual fields (LVF). Furthermore, we determined the neural locus of the interaction using fMRI. Participants performed a visual motion detection task. On each trial, the visual motion stimulus was presented in either the LVF or in the RVF, preceded by a centrally presented word (e.g., "rise"). The word could be congruent, incongruent, or neutral with regard to the direction of the visual motion stimulus that was presented subsequently. Participants were faster and more accurate when the direction implied by the motion word was congruent with the direction of the visual motion stimulus. Interestingly, the speed benefit was present only for motion stimuli that were presented in the RVF. We observed a neural counterpart of the behavioral facilitation effects in the left middle temporal gyrus, an area involved in semantic processing of verbal material. Together, our results suggest that semantic information about motion retrieved in language regions may automatically modulate perceptual decisions about motion. ■

## INTRODUCTION

Perception is influenced by a host of top–down factors, such as attention, expectation, and task set (Gilbert & Li, 2013). It has been hotly debated whether language also influences perception. Recent studies observed an influence of language on the perception of color (Regier & Kay, 2009; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009; Gilbert, Regier, Kay, & Ivry, 2006), faces (Anderson, Siegel, Bliss-Moreau, & Barrett, 2011; Landau, Aziz-Zadeh, & Ivry, 2010; Aziz-Zadeh et al., 2008), objects (Lupyan & Ward, 2013; Hirschfeld, Zwitserlood, & Dobel, 2011; Stanfield & Zwaan, 2001), and motion (Pavan, Skujevskis, & Baggio, 2013; Dils & Boroditsky, 2010; Meteyard, Bahrami, & Vigliocco, 2007). Although evidence for an interaction between language and perception has been forthcoming, it remains unclear at which level of processing this interaction takes place.

Some studies have suggested that language interacts with perception by modulating sensory processing and by showing that language leads to changes in speed and sensitivity of perceptual decisions (Lupyan & Spivey,

2010; Barsalou, 2008; Meteyard et al., 2007) and that language modulates neural activity in sensory cortex at an early stage during a perceptual task (Hirschfeld et al., 2011; Mo, Xu, Kay, & Tan, 2011; Thierry et al., 2009). Alternatively, language–perception interactions could take place in "language areas" by biasing the perceptual decision at the semantic level (Tan et al., 2008). Lexical semantic selection is mediated by the middle temporal gyrus of the left hemisphere (Indefrey & Levelt, 2000, 2004), and this region has been shown to integrate semantic information from different modalities (Noppeney, Josephs, Hocking, Price, & Friston, 2008; Schneider, Debener, Oostenveld, & Engel, 2008; Beauchamp, Lee, Argall, & Martin, 2004). Therefore, it is conceivable that lexical semantic processes may bias the translation of sensory evidence into perceptual decisions.

One factor that may influence whether language modulates perception is the hemisphere that is processing the sensory information. Several studies found a stronger effect of language on perception when visual stimuli are presented in the right visual field (RVF; Mo et al., 2011; Zhou et al., 2010; Gilbert, Regier, Kay, & Ivry, 2008; Drivonikou et al., 2007; Gilbert et al., 2006). Because both RVF stimuli and lexical items are processed by the left hemisphere, these findings are in line with an

[1]Radboud University Nijmegen, [2]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

interplay between perceptual and language processes, but they do not elucidate the processing stage at which this interaction occurs.

In the current study, we aimed to characterize the behavioral effects of motion language on motion perception and to determine the neural locus of these effects. To this end, we measured behavioral performance and neural activity using fMRI while participants were engaged in a motion detection task.

We presented participants with a visual motion stimulus in either the left visual field (LVF) or in the RVF. The motion stimulus was preceded by a word (e.g., "rise"), which was briefly flashed at the center of the visual field. The word had no predictive relation with the direction of the visual motion stimulus, and participants were told that they could ignore the word. Importantly, the word could be congruent, incongruent, or neutral with respect to the subsequent visual motion stimulus. This allowed us to probe whether and where semantic linguistic stimuli influence motion perception, as a function of the hemisphere that processes the sensory information.

## METHODS

### Participants

The experiment consisted of a behavioral and a neuroimaging (fMRI) part. Twenty-two participants (5 men, 17 women; age range = 18–31 years) were included in the behavioral study, and 25 participants (6 men, 19 women; age range = 18–28 years) engaged in the fMRI study. All participants were right-handed, had normal or corrected-to-normal vision, were native Dutch speakers, and had no reading problems. Compensation was €8 for participation in the behavioral study and €25 for participation in the fMRI study. The study was approved by the regional ethics committee, and a written informed consent was obtained from the participants according to the Declaration of Helsinki. Three participants were excluded from the fMRI study. One participant had excessive head movement during scanning (>5 mm), and two participants could not maintain vigilance during the experiment.

### Stimuli

Stimuli were generated using the Psychophysics Toolbox (Brainard, 1997) within MATLAB (MathWorks, Natick, MA) and displayed on a Samsung SyncMaster 940BF monitor (60 Hz refresh rate, 1280 × 1024 resolution) in the behavioral experiment and on a rear-projection screen using an EIKI projector (60 Hz refresh rate, 1024 × 768 resolution) in the fMRI experiment. To ensure constant viewing position and angle in the behavioral experiment, we used a chin and forehead rest to restrain head position. Both words and visual motion stimuli were presented in white (220 cd/m$^2$ behavioral experiment;

126 cd/m$^2$ fMRI experiment) on a light gray background (38 cd/m$^2$ behavioral experiment; 33 cd/m$^2$ fMRI experiment).

Twenty-five verbs describing each direction of motion (upward and downward) and 25 neutral verbs matched for lexical frequency (taken from the CELEX database), number of letters, number of syllables, and concreteness (all $p > .10$) were used in the experiment (Table 1). The visual random-dot motion (RDM) stimuli consisted of white dots (density = 2.4 dots/deg; speed = 14.0 deg/sec) that were plotted within a circular aperture (radius 11.0 deg) that was presented in either the lower left or lower right quadrant of the screen. During random motion trials, all dots were replotted in a random location every monitor refresh, leading to no coherent movement on the screen. During trials with coherent motion, a certain percentage (see below) of the dots was chosen on every frame to be replotted in the coherent direction on the next frame.

The percentage of the dots moving coherently in one direction (upward for half of the participants, downward for the other half, see below) was estimated for each participant using a Bayesian adaptive staircase procedure (Watson & Pelli, 1983). The staircase procedure was done jointly for LVF and RVF stimuli. This was done to yield comparable task difficulty and performance for all participants. During the training phase, participants first practiced the motion detection task in three blocks with fixed coherence levels (80%, 40%, and 20%, respectively). The coherence levels of the two subsequent training blocks were adjusted on the basis of performance in the previous block. The coherence level after the fifth training block was taken as the starting point for the adaptive staircase procedure in the threshold estimation block. Threshold for detection was defined as the percentage of coherent motion for which the staircase procedure predicted 75% accuracy. The coherence level was fixed during each block of trials, but was updated after each block with the same Bayesian staircase procedure to accommodate potential practice and fatigue effects over the course of the experiment.

### Procedure

Direction of motion was counterbalanced across participants, that is, half of the participants were presented with upward and the other half with downward motion stimuli. A central fixation cross (width = 0.3 degrees) was presented throughout the trial, except when a word was presented. Each trial started with a centrally presented word (duration = 100 msec), which could either be a motion word or a neutral word, and which was followed by a 200-msec ISI (see Figure 1). Presentation of the words was fully randomized within each block of the experiment. We instructed participants to ignore the word and maintain fixation. Next, a visual RDM stimulus was presented (duration = 200 msec) in either the LVF or in the RVF. Participants had to indicate, as quickly and accurately as possible, whether the RDM contained coherent motion,

**Table 1.** Dutch Word Lists (with English Translation) for Upward, Downward, and Neutral Words

| Up (Dutch) | English Translation | Down (Dutch) | English Translation | Neutral (Dutch) | English Translation |
|---|---|---|---|---|---|
| bestijgen | mount | afdalen | descend | aanraken | touch |
| heffen | lift | afglijden | slide down | beheren | manage |
| klauteren | clamber | afzakken | come down | bivakkeren | lodge |
| klimmen | climb | bezinken | settle down | boenen | polish |
| lanceren | launch | bukken | stoop | dichtnaaien | sew up |
| omhooggaan | go up | dalen | descend | fatsoeneren | model |
| omhoogkomen | come up | druipen | drip | filmen | film |
| opgaan | go up | duiken | dive | happen | bite |
| opgooien | throw up | gieten | pour | imiteren | imitate |
| ophijsen | pull up | instorten | collapse | kamperen | camp out |
| ophogen | raise | inzinken | break down | liplezen | lip-reading |
| opklimmen | climb | kieperen | tumble | markeren | mark |
| opkrikken | jack up | neerdalen | go down | meubileren | furnish |
| oplaten | launch | neergaan | go down | printen | print |
| oprijzen | rise | neerhalen | take down | ratelen | rattle |
| opstaan | stand up | neerkletteren | crash | rommelen | rumble |
| opstijgen | ascend | neerkomen | fall upon | rondvragen | ask |
| opstuwen | drive | neerploffen | plump down | scheren | shave |
| optillen | lift | neerstorten | crash | smullen | feast |
| opvliegen | fly up | neervallen | fall down | spieken | copy |
| rijzen | rise | storten | fall | troosten | comfort |
| stapelen | pile up | tuimelen | tumble | uitslapen | sleep late |
| stijgen | rise | verlagen | lower | verstoren | disturb |
| verrijzen | arise | zakken | drop | wassen | wash |
| zwellen | swell | zinken | sink | wegen | weigh |

Words are ordered alphabetically.
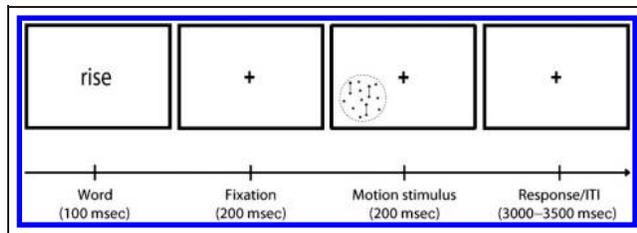


**Figure 1.** Task design. A congruent, incongruent, or neutral word is displayed before every motion detection trial. The visual motion stimulus is presented either in the left or right lower visual field. The dots move upward or randomly for half of the participants and downward or randomly for the other half. ITI = intertrial interval.

while fixating at the central fixation cross. The brief presentation time of the RDM stimulus (200 msec) served to minimize the chance of eye movements to the stimulus, as saccade latencies are in the order of ~200 msec (Carpenter, 1988). Participants were instructed to respond as quickly and accurately as possible by pressing a button with either the left or right index finger in the behavioral experiment and with either their right index or right middle finger in the fMRI experiment. We provided the participants with trial-by-trial feedback only during the training phase, by means of a green or red fixation cross for correct and incorrect responses, respectively. The intertrial interval was 3000–3500 msec for the behavioral experiment and 3500–5500 for the fMRI experiment. The behavioral experiment consisted of eight blocks of 75 trials (600 trials in total), and the fMRI experiment consisted of 10 blocks of 45 trials in two runs (450 trials in total). Summary feedback

(percentage correct) was provided to the participant during the break after each block. A training phase preceded the experiment to familiarize the participants with the task and assess their individual motion coherence threshold at which they performed at 75% correct. There was a resting period of 30 sec after every block in the fMRI experiment and a longer resting period between the sessions.

In the fMRI experiment, we also acquired two additional localizer tasks. In the motion localizer, we presented the same motion stimuli that we used in the experiment (see Stimuli). The motion coherence level was fixed to 80%, and the duration of a trial was 12 sec. There were 10 blocks of seven trials each, presented in pseudorandom order: upward, downward, and random motion in either the LVF or the RVF and a fixation condition. The participant's task was to press a button when the fixation cross turned from white to orange to help them fixate at the center of the screen. In the language localizer, we presented the same word lists that we used in the experiment (see Stimuli). Participants were presented with 10 blocks of five trials. Each trial consisted of 300 msec presentations of 25 words alternating with 300 msec fixation (15 sec per trial). Within a trial, all words were from the same category (upward, downward, neutral, consonant letter strings, and an additional fixation condition). Participants were instructed to monitor occasional word repetitions (1-back task, occurring on average three times per trial). We chose a 1-back task to make sure that participants would attentively read the words. For both localizer tasks, the intertrial interval was 1 sec. The order of the fMRI sessions was as follows: (1) short training of the task; (2) thresholding procedure; (3) experimental session 1; experimental session 2; language localizer; motion localizer; anatomical T1.

## Behavioral Analysis

We calculated congruency effects for four behavioral measures: RT, percentage correct and signal detection-theoretic measures $d'$ and $c$ (Macmillan & Creelman, 2005). $d'$ is a measure of a participant's stimulus discriminability, also known as perceptual sensitivity, and was calculated as follows:

$$d' = z(H) - z(F)$$

where $H$ denotes the hit rate, $F$ is the false alarm rate, and the $z$ transformation converts these measures to a $z$ score (i.e., to standard deviation units). This measure is independent of any potential biases induced by the motion direction suggested by the word. This bias can be analyzed separately by estimating $c$, the internal response criterion of the participant, which was calculated as follows:

$$c = -\frac{1}{2}[z(H) + z(F)].$$

A negative criterion arises when the false alarm rate exceeds the hit rate and therefore indicates liberal performance in reporting coherent motion during trials that contain no coherent motion in the current experimental setting, whereas a positive criterion denotes conservative reporting. Trials were labeled as congruent when the motion described by the word matched the direction of visual motion, for example, "rise" followed by a stimulus with upward moving dots. When the motion described by the word and the direction of visual motion did not match, the trial was labeled incongruent. Neutral words were used as a control condition. Trials with RTs that were >3 SD longer/shorter than the individual subject mean RT were excluded from the analyses (in total 2.0%).

Each of the four behavioral measures was subjected to a repeated-measures ANOVA, including factors Congruency (congruent, incongruent), Visual field (LVF, RVF), and Experiment (behavioral experiment, fMRI experiment).

## fMRI Acquisition

Images were acquired on a 1.5-T Avanto MRI system (Siemens, Erlangen, Germany). Whole-brain $T_2$*-weighted gradient-echo echo-planar images (repetition time = 2000 msec, echo time = 40 msec, 33 ascending slices, voxel size = 3 × 3 × 3 mm, flip angle = 80°, field of view = 192 mm) were acquired using a 32-channel head coil. A high-resolution anatomical image was collected using a $T_1$-weighted magnetization prepared rapid gradient-echo sequence (repetition time = 2730 msec, echo time = 2.95 msec, voxel size = 1 × 1 × 1 mm).

## fMRI Data Analysis

Analysis was performed using SPM8 (www.fil.ion.ucl.ac.uk/spm, Wellcome Trust Centre for Neuroimaging, London, UK). The first four volumes of each run were discarded to allow for scanner equilibration. Preprocessing consisted of realignment through rigid body registration to correct for head motion, slice timing correction to the onset of the first slice, coregistration of the functional and anatomical images, and normalization to a standard $T_1$ template centered in MNI space by using linear and nonlinear parameters and resampling at an isotropic voxel size of 2 mm. Normalized images were smoothed with a Gaussian kernel with a FWHM of 8 mm. A high-pass filter (cutoff = 128 sec) was applied to remove low-frequency signals, such as scanner drift. The ensuing preprocessed fMRI time series were analyzed on a subject-by-subject basis using an event-related approach in the context of the general linear model. Regressors for the first-level analysis were obtained by convolving the unit impulse time series for each condition with the canonical hemodynamic response function. We modeled the 12 different conditions of the experiment [word type (3) × motion

type (2) × visual field (2)] separately for each of the two sessions. Because "motion type" was varied between participants (half of the participants were presented "upward" and "random" motion and the other half "downward" and "random" motion), we collapsed the conditions over participants to obtain congruent, incongruent, and neutral conditions for both "coherent" and "random" motion stimuli for both visual fields. We assessed the effects of congruency between language and perception for the trials that contained coherent motion. Resting periods were modeled as a regressor of no interest. We included six nuisance regressors related to head motion: three regressors related to translation and three regressors related to rotation of the head. For the localizers, we used the same procedure. Both localizers used a block design. The motion localizer had seven conditions and block duration of 12 sec. The language localizer had five conditions and block duration of 15 sec.

### Statistical Analysis

We used a priori functional information on the basis of the results from the localizers to constrain our search space (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006). In particular, we isolated the regions that were involved in semantic language processing (language localizer) and visual motion processing (motion localizer). These corresponded to the left middle temporal gyrus (lMTG, language localizer) and bilateral hMT+/V5 (motion localizer). Specifically, we obtained the anatomical location of the left MTG by contrasting the three word conditions (up, down, neutral words) with the random consonant letter strings condition (MNI coordinates: [−54,−34,4]). We obtained the anatomical location of the right hMT+/V5 ROI by contrasting visual motion stimulation in the LVF > RVF (MNI coordinates: [40,−78,4]) and the left hMT+/V5 with the reverse contrast (MNI coordinates: [−40,−82,8]). We defined search volumes comprising spheres of 10 mm around these regions and corrected our results for multiple comparisons using a family-wise error rate threshold of $p < .05$ within this search volume (Worsley, 1996). We computed the mean activity over the voxels in each ROI for the different conditions. Finally, to verify the language–perceptual interactions that have previously been reported in parietal cortex (Sadaghiani, Hesselmann, & Kleinschmidt, 2009; Tan et al., 2008), we performed an additional ROI analysis with peak coordinates from Sadaghiani et al. (MNI coordinates: [45,−45,39] and [−42,−54,45]) and Tan et al. (2008; MNI coordinates: [−61,−32,27]) following the procedure described for the other ROI analyses. Additional whole-brain statistical inference was performed using a cluster-level statistical test to assess clusters of significant activation (Friston, Holmes, Poline, Price, & Frith, 1996). We used a corrected cluster threshold of $p < .05$, on the basis of an auxiliary voxel threshold of $p < .001$ at the whole-brain level.
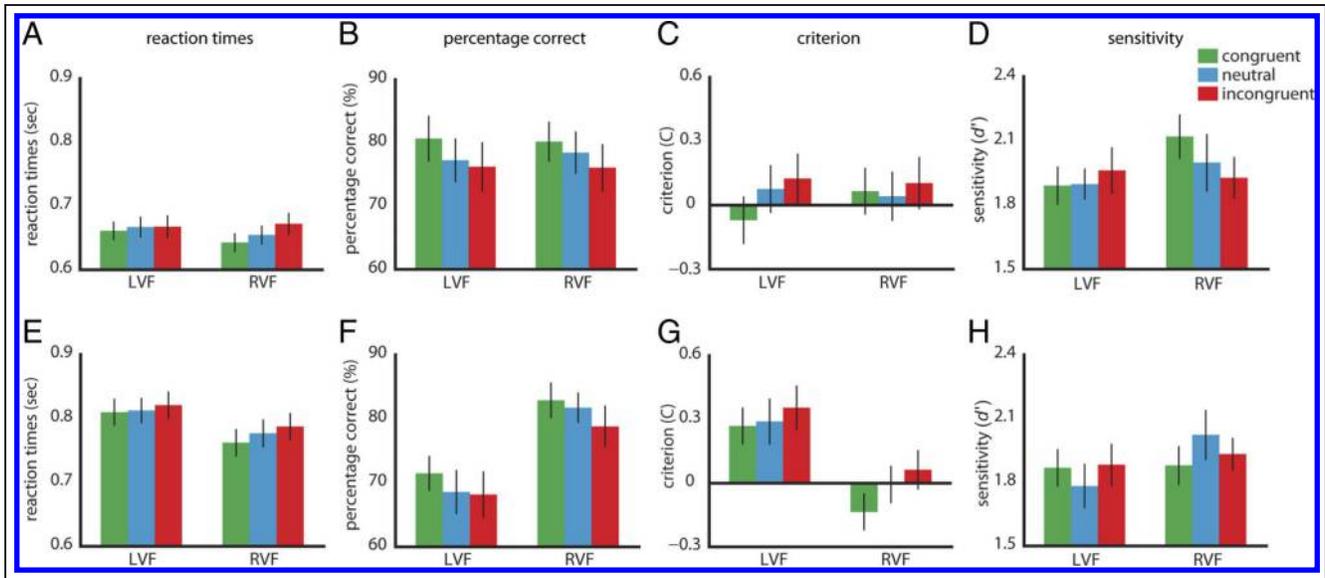
## RESULTS

### Behavioral Effects of Language on Motion Perception

Here, we report the combined behavioral data from the behavioral and fMRI experiment. Participants responded faster to the motion stimuli when they were preceded by a congruent motion word than by an incongruent word (congruency: $F(1, 42) = 10.914, p = .002$). Crucially, this congruency effect was modulated by visual field ($F(1, 42) = 4.915, p = .032$) (see Figure 2A). Motion stimuli that were preceded by congruent motion words were responded to faster when presented in the RVF (congruent: RT = 702 msec; incongruent: RT = 730 msec; ΔRT = 28 msec, $F(1, 42) = 23.588, p < .001$), but not in the LVF (congruent: RT = 735 msec; incongruent: RT = 744 msec; ΔRT = 9 msec, $F(1, 42) = 1.241, p = .27$). The RT effects did not differ between the two experiments (Congruency × Experiment: $F(1, 42) < 0.001, p = .98$; Visual field × Congruency × Experiment: $F(1, 42) = 0.260, p = .61$) indicating that the congruency effect was larger for RVF than for LVF in both studies. There was also a general RVF advantage for RTs (Visual field: $F(1, 42) = 10.552, p = .002$), which was larger for the fMRI experiment than the behavioral experiment (Visual field × Experiment: $F(1, 42) = 5.292, p = .026$).

Participants' task performance was individually thresholded using an adaptive staircasing procedure (see Methods) to ensure overall approximately 75% correct performance. On average, participants answered 79% of trials correctly (±4.2%, mean ± SD) at a motion coherence level of 19% (±8.5%, mean ± SD). Accuracy was significantly higher for congruent compared with incongruent trials for both visual fields (main effect of Congruency: $F(1, 42) = 8.848, p = .005$; LVF: congruent: 76.1%; incongruent: 72.2%; Δ = 3.9%, $F(1, 42) = 6.954, p = .012$; RVF: congruent: 81.5%; incongruent: 77.4%; Δ = 4.1%, $F(1, 42) = 4.717, p = .036$). There was no significant interaction between Congruency and Visual field ($F(1, 42) = 0.010, p = .92$) (see Figure 2B, F). The effects were similar in the two experiments (Congruency × Experiment: $F(1, 42) = 0.049, p = .83$; Visual field × Congruency × Experiment: $F(1, 42) = 0.075, p = .79$). Accuracy was higher in the RVF than in the LVF in the imaging experiment (Visual field × Experiment: $F(1, 42) = 3.006, p = .090$).

Participants exhibited a more liberal decision criterion when the motion word and visual motion stimulus were congruent than when they were incongruent for both visual fields (main effect of Congruency: $F(1, 42) = 11.104, p = .002$; LVF: congruent: $C = 0.10$; incongruent: $C = 0.24$; Δ$C = 0.14$, $F(1, 42) = 9.804, p = .003$; RVF: congruent: $C = -0.03$; incongruent: $C = 0.08$; Δ$C = 0.11$, $F(1, 42) = 6.020, p = .018$). No significant interaction between Congruency and Visual field was present ($F(1, 42) = 0.201, p = .66$) (see Figure 2C, G). Only for criterion, there was a significant difference in the lateralization of the congruency effects between the experiments

**Figure 2.** Behavioral results. (A–D) Behavioral study. (A) Mean RTs (in sec) for visual motion stimuli that were presented in the LVF (left bars) or RVF (right bars) and which were preceded by a congruent (green), neutral (blue), or incongruent (red) word ($n = 22$). (B) Percentage correct. Other conventions as in A. (C) Decision criterion. (D) Sensitivity ($d'$). (E–H) fMRI study. Conventions as in A–D.

(Visual field × Congruency × Experiment: $F(1, 42) = 6.887, p = .012$), which is caused by the fact that the more liberal criterion for congruent stimuli is stronger in the LVF during the behavioral experiment but stronger in the RVF during the imaging experiment. Participants were more conservative in their perceptual decisions in the LVF than in the RVF in the fMRI experiment (Visual field × Experiment: $F(1, 42) = 4.725, p = .035$).

Sensitivity for motion detection was neither different for congruent compared with incongruent trials in the LVF nor in the RVF (main effect of congruency: $F(1, 42) = 0.058, p = .81$; LVF: congruent: $d' = 1.88$; incongruent: $d' = 1.92$; $\Delta d' = -0.04$ $F(1, 42) = 0.314, p = .58$; RVF: congruent: $d' = 2.00$; incongruent: $d' = 1.93$; $\Delta d' = 0.07$, $F(1, 42) = 1.018, p = .32$), and there was no significant interaction between Congruency and Visual field ($F(1, 42) = 1.457, p = .23$) (see Figure 2D). There was no difference in sensitivity effects between the experiments (Congruency × Experiment: $F(1, 42) = 0.725, p = .40$; Visual field × Congruency × Experiment: $F(1, 42) = 2.65, p = .11$).

We included a neutral (no motion) words condition to aid the interpretation of the congruency effects. The neutral condition showed behavior that was intermediate between the congruent and incongruent conditions for RT, accuracy, and criterion, suggesting that the motion words could incur either a cost or benefit, depending on the congruency with the upcoming motion stimulus (RT: congruent > neutral LVF: $T_{43} = -0.77, p = .45$; RVF: $T_{43} = -2.63, p = .012$; neutral > incongruent LVF: $T_{43} = -0.75, p = .46$; RVF: $T_{43} = -2.71, p = .010$; accuracy: congruent > neutral LVF: $T_{43} = 2.24, p = .031$; RVF: $T_{43} = 1.04, p = .30$; neutral > incongruent LVF: $T_{43} = 0.51, p = .62$; RVF: $T_{43} = 1.88, p = .067$; criterion: congruent > neutral LVF: $T_{43} = -1.94, p = .059$; RVF:

$T_{43} = -1.21, p = .23$; neutral > incongruent LVF: $T_{43} = -1.73, p = .091$; RVF: $T_{43} = -1.62, p = .11$; sensitivity: congruent > neutral LVF: $T_{43} = 0.63, p = .53$ RVF: $T_{43} = -0.14, p = .89$; neutral > incongruent LVF: $T_{43} = -1.12, p = .27$; RVF: $T_{43} = 0.96, p = .34$).

## Neural Effects of Language on Motion Perception

As expected, motion stimuli in the LVF were associated with increased activity in the right hMT/V5+, whereas motion stimuli in the RVF led to stronger responses in the left hMT/V5+ (difference between ipsilateral and contralateral visual stimuli, lhMT+/V5: $T_{21} = 8.39, p < .001$; rhMT+/V5: $T_{21} = 8.76, p = .001$; see Figure 3B, D). However, hMT+/V5 was not modulated by the congruence between the motion word and the visual motion stimulus, not even at liberal statistical thresholds ($p > .05$ uncorrected). An effect of language on motion perception was observed however in the lMTG (MNI coordinates: $[-58, -34, -6]$), where we found a significant increase in activation for the congruent compared with the incongruent condition (see Figure 3A and C, $T_{21} = 4.17, p = .029$). The size of the congruency effect was not different for LVF compared with RVF stimuli in lMTG. Finally, there was a borderline significantly larger activation for the congruent than the incongruent condition in left anterior IPS ($T_{21} = 3.61, p = .050$).

We also carried out a whole-brain analysis to identify potential other regions that are modulated by the congruency between the motion word and motion stimulus. No other brain regions showed a significant difference in activation for the incongruent condition relative to the congruent condition, nor a significant interaction between congruency and visual field.

## DISCUSSION

We investigated the effects of motion language on motion perception in a combined behavioral and fMRI study. We found that when motion words were congruent with the direction of the visual motion stimulus, participants were faster, more accurate, and more liberal in detecting visual motion. Interestingly, the speed benefit was present only for visual stimuli that were presented in the RVF and thus processed in the left (language dominant) hemisphere. We observed a potential neural counterpart to these behavioral facilitatory effects in the lMTG, an area involved in lexical knowledge. This suggests that semantic categorization may be an integral part of the perceptual decision process and lMTG is a neural locus where language and perception interact.

Previous work already suggested an effect of motion words on motion perception. Meteyard et al. (2007) investigated whether a stream of auditorily presented motion words affected the detection of motion in centrally presented visual stimuli. They showed that, when motion stimuli were paired with congruent motion words, motion sensitivity ($d'$) was improved and decision criterion was more liberal. Despite the substantial differences in design (e.g., trial-by-trial presentation of words vs. blocked presentation, visual presentation vs. auditory presentation), we partly replicate and extend these findings by showing modulations of accuracy, criterion, and RTs. Interestingly, a variation of the Meteyard et al. (2007) study by Pavan et al. (2013) showed a double dissociation between discrimination sensitivity ($d'$) and RTs depending on whether motion coherence was above or at threshold. With suprathreshold motion, responses were faster for congruent stimuli, but sensitivity was equal across conditions. When the motion was at threshold however, sensitivity was higher for congruent stimuli, but responses were equally fast across conditions. Thus, differences in motion coherence level might explain the absence of sensitivity effects in our study and the lack of RT effects in the study of Meteyard et al. Another determinant of the nature of language–perception interactions might be the degree of temporal overlap between linguistic and perceptual information. In our study, the two events were separated by 300 msec, which might result in integration at a later stage in the decision process.

Interestingly, the RT effects were dependent on the visual field in which the motion stimuli were presented: only for motion stimuli that were presented in the RVF (which are processed by the language-dominant left hemisphere), we observed faster RTs when the motion stimuli were preceded by congruent, compared with incongruent, motion words. This lateralization of a language–perception interaction has been observed for other types of visual stimuli (e.g., color, objects; Mo et al., 2011; Zhou et al., 2010;
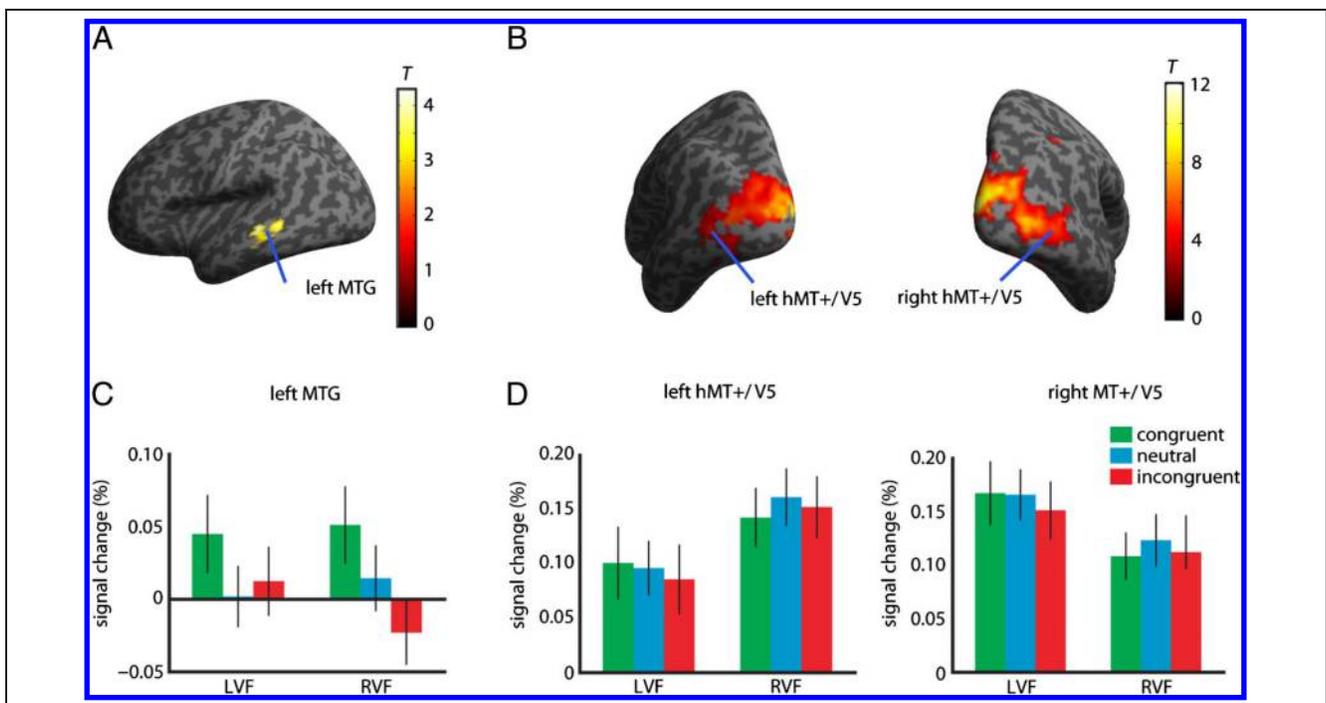


**Figure 3.** fMRI results. (A) Activation for congruent > incongruent conditions plotted on an inflated render MNI brain. The only significant modulation because of congruency is localized in lMTG ($n = 22$). (B) Activation in the motion localizer for motion > fixation plotted an inflated render MNI brain. (C) Within the lMTG blob ($p < .001$ uncorrected) the percentage signal change for the congruent (green), neutral (blue), and incongruent (red) conditions is plotted for both the LVF (left) and RVF (right). (D) For both hMT+/V5 ROIs, the percentage signal change for the congruent (green), neutral (blue), and incongruent (red) conditions is plotted. There is no modulation of either left or right hMT+/V5 by congruency, but there is a clear activation difference in both ROIs between stimuli presented in the LVF and RVF.

Regier & Kay, 2009; Gilbert et al., 2006, 2008; Drivonikou et al., 2007). The lateralization effect we find in our study supports the hypothesis that language changes perception in a specific way, that is, by a process in which word meaning is matched with the outcome of a semantic categorization of visual stimuli (e.g., "rise" matches with visual motion categorized as moving "upwards"). This appears fundamentally different from more general priming or response conflict effects that do not depend on stimulus hemifield, such as those observed in, for example, Stroop paradigms (Leung, Skudlarski, Gatenby, Peterson, & Gore, 2000). Related, the results are unlikely to be caused by attentional cueing, as the word cue had no probabilistic relationship with the following stimulus (direction of movement of visual motion). Furthermore, it is difficult to see why attentional cueing would only be present for stimuli that are presented in the RVF.

With our fMRI study, we aimed to elucidate which neural regions were sensitive to the congruency between the motion words and visual stimuli. Such a congruency effect was observed in the lMTG, although the congruency effect was not significantly stronger for motion presented in the RVF (as was the case for the behavioral congruency effect). The lMTG is part of the mostly left-lateralized language network and is known to be involved in both lexical retrieval including word semantics and multisensory processing and integration (Menenti, Gierhan, Segaert, & Hagoort, 2011; Hagoort, Baggio, & Willems, 2009; Noppeney et al., 2008; Schneider et al., 2008; Beauchamp et al., 2004). Similar to our finding that the lMTG shows increased activity for congruent compared with incongruent conditions, Schneider et al. (2008) showed a crossmodal priming effect in response to semantically congruent stimuli in the lMTG, using EEG. They suggest that the enhanced gamma-band power for congruent compared with incongruent conditions may reflect a crossmodal semantic matching process that is triggered by the expectation of an upcoming event (i.e., a congruent stimulus). This crossmodal matching process may also occur when making perceptual decisions, if the perceptual decision is translated into a lexical concept.

In an ROI-based post hoc test with peak coordinates from Sadaghiani et al. (2009), a cluster in left anterior IPS was also sensitive to the difference between congruent and incongruent linguistic and perceptual information, in line with previous studies (Sadaghiani et al., 2009; Tan et al., 2008).

Surprisingly, we did not find any interaction effects in motion-sensitive visual cortical area hMT+/V5. This is in contrast to earlier studies that have found neural activity modulations by linguistic stimuli during perceptual tasks that occurred early in time and was localized in sensory areas (Hirschfeld et al., 2011; Mo et al., 2011; Thierry et al., 2009). One potential reason for this discrepancy could be the fact that participants were instructed to ignore the motion words, which may have attenuated processing of the verbal material.

How do these behavioral and neural results inform the central question: at which level of processing does the interaction between language and perception occur? We conjectured two levels at which this interaction could occur. First, motion words could induce an "automatic prediction" about visual motion, thereby automatically recruiting the relevant sensory areas. Alternatively, but still in line with the sensory level hypothesis, motion words *themselves* may recruit the motion-sensitive visual cortex, as advocated by the embodied language hypothesis. This hypothesis claims that words describing motion are partly represented in the corresponding perceptual areas that process the actual visual stimuli the words describe (Barsalou, 2008). However, in our study we did not find evidence for engagement of hMT+/V5 or nearby sensory areas in the interaction between motion words and motion perception. Thus, our data do not support strong versions of embodiment according to which motion words automatically and necessarily activate visual motion areas. Second, the interaction between language and perception could occur at a higher level of language processing. The visual motion stimuli might be conceptually categorized ("up," "down"), as the participants are required to make a categorical perceptual decision. So although it is not necessary to perform the task, linguistic representations may be automatically activated (Tan et al., 2008). If the activated motion word meaning matches the subsequent semantic representation activated by the visual motion stimulus, this then leads to more activity in lMTG (Schneider et al., 2008), as well as improved behavioral performance. Klemfuss, Prinzmetal, and Ivry (2012) support this interpretation of the linguistic effects on perception by showing that the language effects may be postperceptual rather than directly influencing early perceptual processing. In a visual search experiment, they demonstrate that the disruption of visual search by automatically activated irrelevant linguistic information is the result of an interaction at a response selection stage of processing. Thus, semantic categorization may be an integral part of the perceptual decision process. This hypothesis is in line with both the behavioral data (showing RT and criterion effects) and the fMRI data (showing postperceptual integration effects of the semantic and visual information in the lMTG).

In the current study, motion words influenced motion perception despite the fact that the words had no predictive value for the upcoming stimulus and participants were instructed to ignore them. This suggests that the influence of language on perception is an automatic rather than a strategic process. However, the experimental effects were modest and "local" (i.e., only visible when the linguistic and visual stimuli were processed in the same hemisphere) compared with other studies, which suggests that a stronger context may be necessary for more robust and widespread language–perceptual interactions. For instance, Lupyan and Ward (2013) found that the presentation of a valid verbal cue before an invisible image of an object changed object detection performance relative to an uninformative cue. This suggests that attended and predictive language can exert a strong influence on perception.

Furthermore, when the linguistic context is stronger, that is, when stimuli are sentences or narratives describing motion, studies have found activation of motion processing areas more proximal to MT+ (Wallentin et al., 2011; Saygin, McCullough, Alac, & Emmorey, 2010).

The unattended nature of the motion words in our study (as a consequence of the task difficulty of the motion detection task and the task instructions) may be an explanation for the "local" effects of motion words on motion perception, in terms of neural activation and RTs: Motion words influenced RTs only for stimuli presented in the RVF. In these trials, the linguistic and visual material was processed within the same (left) hemisphere. Given that attention is often thought to have a "broadcasting" effect (Dehaene, Sergent, & Changeux, 2003; Dehaene & Naccache, 2001), it is an interesting question whether attention to the words would result in congruency effects on RTs also for visual material presented to the LVF and possibly to a more extended network of areas in the parietal and pFC that are involved in the "broadcasting" of information (Dehaene & Changeux, 2011). This hypothesis would provide an alternative explanation for the often reported, but debated, observation that language exerts stronger effects on RVF than on LVF stimuli. This asymmetry is thought to be related to the left lateralization of the language system (Klemfuss et al., 2012; Regier & Kay, 2009; Gilbert et al., 2006), but importantly, the crucial factor could be the degree to which the linguistic information is attended and thus broadcasted. Therefore, when the motion words are attended, we expect larger and potentially bilateral effects. This prediction could be tested in future experiments.

In conclusion, this study provides insight into the behavioral and neural effects of language on perception. We show that language affects motion perception, with stronger effects for motion stimuli that are processed in the language-dominant left hemisphere. These interactions are neurally mediated by "language areas" rather than perceptual areas, suggesting that these may form integral part of the network involved in perceptual decisions about visual motion stimuli.

Reprint requests should be sent to Jolien C. Francken, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, P.O. Box 9101, 6500 HB, Nijmegen, Netherlands, or via e-mail: j.francken@donders.ru.nl.

# REFERENCES

Anderson, E., Siegel, E. H., Bliss-Moreau, E., & Barrett, L. F. (2011). The visual impact of gossip. *Science, 332,* 1446–1448.

Aziz-Zadeh, L., Fiebach, C. J., Naranayan, S., Feldman, J., Dodge, E., & Ivry, R. B. (2008). Modulation of the FFA and PPA by language related to faces and places. *Society for Neuroscience, 3,* 229–238.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41,* 809–823.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Carpenter, R. H. S. (1988). *Movement of the eyes* (2nd ed.). London: Pion Limited.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70,* 200–227.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79,* 1–37.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 8520–8525.

Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 16396–16400.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 1097–1102.

Friston, K. J., Holmes, A., Poline, J. B., Price, C. J., & Frith, C. D. (1996). Detecting activations in PET and fMRI: Levels of inference and power. *Neuroimage, 4,* 223–235.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *Neuroimage, 30,* 1077–1087.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 489–494.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language, 105,* 91–98.

Gilbert, C. D., & Li, W. (2013). Top–down influences on visual processing. *Nature Reviews Neuroscience, 14,* 350–363.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 819–836). Cambridge, MA: MIT Press.

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing—MEG dissociates early perceptual and late N400 effects. *Brain and Language, 116,* 91–96.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 845–865). Cambridge, MA: MIT Press.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92,* 101–144.

Klemfuss, N., Prinzmetal, W., & Ivry, R. B. (2012). How does language change perception: A cautionary note. *Frontiers in Psychology, 3,* 78.

Landau, A. N., Aziz-Zadeh, L., & Ivry, R. B. (2010). The influence of language on perception: Listening to sentences about faces affects the perception of faces. *Journal of Neuroscience, 30,* 15254–15261.

Leung, H. C., Skudlarski, P., Gatenby, J. C., Peterson, B. S., & Gore, J. C. (2000). An event-related functional MRI study of the Stroop color word interference task. *Cerebral Cortex, 10,* 552–560.

Lupyan, G., & Spivey, M. J. (2010). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS One, 5,* e11452.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, U.S.A., 110,* 14196–14201.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science, 22,* 1173–1182.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science, 18,* 1007–1013.

Mo, L., Xu, G., Kay, P., & Tan, L. H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences, U.S.A., 108,* 14026–14030.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex, 18,* 598–609.

Pavan, A., Skujevskis, M., & Baggio, G. (2013). Motion words selectively modulate direction discrimination sensitivity for threshold motion. *Frontiers in Human Neuroscience, 7,* 134.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences, 13,* 439–446.

Sadaghiani, S., Hesselmann, G., & Kleinschmidt, A. (2009). Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. *Journal of Neuroscience, 29,* 13410–13417.

Saygin, A. P., McCullough, S., Alac, M., & Emmorey, K. (2010). Modulation of BOLD response in motion-sensitive lateral temporal cortex by real and fictive motion sentences. *Journal of Cognitive Neuroscience, 22,* 2480–2490.

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage, 42,* 1244–1254.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153–156.

Tan, L. H., Chan, A. H., Kay, P., Khong, P. L., Yip, L. K., & Luke, K. K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 4004–4009.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 4567–4570.

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). BOLD response to motion verbs in left posterior middle temporal gyrus during story comprehension. *Brain and Language, 119,* 221–225.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120.

Worsley, K. J. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping, 4,* 58–73.

Zhou, K., Mo, L., Kay, P., Kwok, V. P., Ip, T. N., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 9974–9978.

**This article has been cited by:**

1. Jason Samaha, Bastien Boutonnet, Bradley R. Postle, Gary Lupyan. 2018. Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. *Scientific Reports* **8**:1. . [Crossref]

2. Akihiko Gobara, Yuki Yamada, Kayo Miura. 2016. Crossmodal Modulation of Spatial Localization by Mimetic Words. *i-Perception* **7**:6, 204166951668424. [Crossref]

3. Jolien C. Francken, Erik L. Meijs, Peter Hagoort, Simon van Gaal, Floris P. de Lange. 2016. Exploring the automaticity of language-perception interactions: Effects of attention and awareness. *Scientific Reports* **5**:1. . [Crossref]

4. Irina Simanova, Jolien C. Francken, Floris P. de Lange, Harold Bekkering. 2016. Linguistic priors shape categorical perception. *Language, Cognition and Neuroscience* **31**:1, 159-165. [Crossref]

5. Joseph Fruchter, Tal Linzen, Masha Westerlund, Alec Marantz. 2015. Lexical Preactivation in Basic Linguistic Phrases. *Journal of Cognitive Neuroscience* **27**:10, 1912-1935. [Abstract] [Full Text] [PDF] [PDF Plus]

6. Jolien C. Francken, Erik L. Meijs, Odile M. Ridderinkhof, Peter Hagoort, Floris P. de Lange, Simon van Gaal. 2015. Manipulating word awareness dissociates feed-forward from feedback models of language-perception interactions. *Neuroscience of Consciousness* **2015**:1, niv003. [Crossref]