Intact sensory processing but hampered conflict detection when stimulus input is task-irrelevant

Stijn A. **Nuiten**[1,2*], Andrés **Canales-Johnson**[1,2,3.4], Lola **Beerendonk**[1,2,3], Nutsa **Nanuashvili**[1,2], Johannes J. **Fahrenfort**[1,2], Tristan **Bekinschtein**[3,4,^] and Simon **van Gaal**[1,2,*,^]

[1] Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT, Amsterdam, The Netherlands

[2] Amsterdam Brain & Cognition, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, The Netherlands

[3] Department of Psychology, University of Cambridge, Downing Street, CB2 3EB Cambridge, United Kingdom

[4] Behavioural and Clinical Neuroscience Institute, Downing Street, CB2 3EB Cambridge, United Kingdom

[*] Corresponding authors

[^] Shared senior authorship

Abstract (now 150, max 150)

Conflict detection in sensory input is central to adaptive human behavior. Perhaps unsurprisingly, past research has shown that conflict may be detected even in the absence of conflict awareness, suggesting that conflict detection is a fully automatic process that does not require attention. Across six behavioral tasks, we manipulated task relevance and response overlap of potentially conflicting stimulus features to test the possibility of conflict processing in the absence of attention. Multivariate analyses on human electroencephalographic data revealed that neural signatures of conflict are only present when at least one feature of a conflicting stimulus is attended, regardless of whether that feature is part of the conflict. In contrast, neural signatures of basic sensory processes are present even when a stimulus is completely unattended. These data reveal an attentional bottleneck at the level of objects, suggesting that object-based attention is a prerequisite for cognitive control operations involved in conflict detection.

Introduction

Every day we are bombarded with sensory information from the environment and we often face the challenge of selecting the relevant information and ignoring irrelevant –potentially conflicting– information to maximize performance. These selection processes require much effort and our full attention, sometimes rendering us deceptively oblivious to irrelevant sensory input (e.g. chest-banging apes), as illustrated by the famous inattentional blindness phenomenon (Simons & Chabris, 1999). However, unattended events that are not relevant for the current task might still capture our attention or interfere with ongoing task performance, for example when they are inherently relevant to us (e.g. our own name). This is illustrated by another famous psychological phenomenon: the cocktail party effect (Cherry, 1953; Moray, 1959). Thus, under specific circumstances, task-irrelevant information may capture attentional resources and be subsequently processed with different degrees of depth.

It is currently a matter of debate which processes require top-down attention (Dehaene et al., 2006; Koch & Tsuchiya, 2007; Koelewijn et al., 2010; Lamme, 2003; Lamme & Roelfsema, 2000; Rousselet et al., 2004; VanRullen, 2007). It was long thought that only basic physical stimulus features or very salient stimuli are processed in the absence of attention (Treisman & Gelade, 1980), due to an "attentional bottleneck" at higher levels of analysis (Broadbent, 1958; Deutsch & Deutsch, 1963; Lachter et al., 2004; Wolfe & Horowitz, 2004). However, there is now solid evidence that several tasks may in fact still unfold in the (near) absence of attention, including perceptual integration (Fahrenfort et al., 2017), the processing of emotional valence (Sand & Wiens, 2011; Stefanics et al., 2012), semantical processing of written words (Schnuerch et al., 2016) and visual scene categorization. Although one should be cautious in claiming complete absence of attention (Lachter et al., 2004), these and other studies have pushed the boundaries of input processing that is task-irrelevant (without attention), and may even question the existence of an attentional bottleneck at all. Conceivably, the attentional bottleneck is only present at higher, more complex, levels of cognitive processing, like cognitive control functions.

Over the years, various theories have been proposed with regard to this attentional bottleneck among which the load theory of selective attention and cognitive control (Lavie et al., 2004), the multiple resources theory (Wickens, 2002), the hierarchical central executive bottleneck theory and formalizations thereof in a cortical network model for serial and parallel processing (Sigman & Dehaene, 2006; Zylberberg et al., 2010, 2011). These theories all hinge on the idea that resources for the processing of information are limited and that the brain therefore has to allocate resources to processes that are currently most relevant, via selective attention (Broadbent, 1958; Treisman, 1969). Resource (re-)allocation, and thus flexible behavior, is thought to be governed by an executive network, most prominently involving the prefrontal cortex (Goldman-Rakic, 1995, 1996). Information that is deemed task-irrelevant has fewer resources to its disposal and is therefore likely processed to a lesser extent. When more resources are necessary for processing the task-relevant information, e.g. under high perceptual load, processing of task-irrelevant information diminishes (Lavie et al., 2003, 2004). Yet even under high perceptual load, task-irrelevant features can be processed through object-based attention (Chen, 2012; Chen & Cave, 2006; Cosman & Vecera, 2012; Kahneman et al., 1992; O'Craven et al., 1999; Schoenfeld et al., 2014; Wegener et al., 2014). There is currently no consensus what type of information can be processed in parallel by the brain and which attentional mechanisms determine what information passes the attentional bottleneck. One unresolved issue is that most empirical work has investigated the bottleneck with regard to sensory features, however it is unknown if the bottleneck and the distribution of processing resources also takes place for more complex, cognitive processes. Here, we test whether such a high-level attentional bottleneck indeed exists in the human brain.

Specifically, we aim to test whether cognitive control operations, necessary to identify and resolve conflicting sensory input, are operational when that input is irrelevant for the task at hand (and hence unattended). Additionally, we test whether in the absence of task-relevance of the conflicting feature, object-based attention

might overcome any hampered control operations. Previous work has shown that the brain has dedicated networks for the detection and resolution of conflict, in which the medial frontal cortex (MFC) plays a pivotal role (Ridderinkhof et al., 2004). Conflict detection and subsequent behavioral adaptation is central to human cognitive control and, hence, it maybe not surprising that past research has shown that conflict detection can even occur unconsciously (Atas et al., 2016; D'Ostilio & Garraux, 2012a; Huber-Huber & Ansorge, 2018; van Gaal et al., 2008), suggesting that the brain may detect conflict fully automatically and that it may even occur without paying attention (e.g. Rahnev et al., 2012).

Conclusive evidence regarding the latter claim has, to our knowledge, not been provided and therefore the necessity of attention for cognitive control operations remains open for debate. Previous studies have shown that cognitive control processes are operational when to-be-ignored features from either a task-relevant or a task-irrelevant stimulus overlap with the behavioral response to be made to the primary task, causing interference in performance (Mao & Wang, 2008; Padrão et al., 2015; Zimmer et al., 2010). In these circumstances the interfering stimulus feature carries information related to the primary task and are therefore *de facto* not task-irrelevant. Consequently, it is currently unknown whether cognitive control operations are active for conflicting sensory input that is not related to the task at hand. Given the immense stream of sensory input we encounter in our daily lives, conflict between two (unattended) sources of perceptual information is inevitable.

Here, we investigated whether conflict between two features of an auditory stimulus (its content and its spatial location) would be detected by the brain under varying levels of task-relevance of these features. The main aspect of the task was as follows. We presented auditory spoken words ("left" and "right" in Dutch) through speakers located on the left and right side of the body. By presenting these stimuli through either the left or the right speaker, content-location conflict arises on specific trials (e.g. the word "left" from the right speaker) but not on others (e.g. the word "right" from the right speaker) (Buzzell et al., 2013; Canales-Johnson et al., 2020). A wealth of previous studies has revealed that conflict arises between task-relevant and task-irrelevant features of the stimulus in these type of tasks (similar to the Simon task and Stroop task, Egner & Hirsch, 2005; Hommel, 2011). Here, these potentially conflicting auditory stimuli were presented during six different behavioral tasks, divided over two separate experiments, multiple experimental sessions and different participant groups (both experiments N=24). In all tasks we focus on the processing of content-location conflict of the auditory stimulus. There were several critical differences between the behavioral tasks: (1) task relevance of a conflicting feature, (2) task relevance of a non-conflicting feature that was part of a conflicting stimulus and (3) whether the response mapped onto a conflicting feature (e.g. respond right when left/right were conflicting features). Note that in all tasks, only one feature could be task-relevant, and that all the other feature(s) had to be ignored. The systematic manipulation of task-relevance and the response mapping allowed us to explore the full landscape of possibilities of how varying levels of attention affect sensory and conflict processing. Electroencephalography (EEG) was recorded and multivariate analyses on the EEG data was used to extract any neural signatures of conflict detection (i.e. theta-band neural oscillations; Cavanagh & Frank, 2014; Cohen & Cavanagh, 2011) and sensory processing for any of its features.

Results

Experiment 1: Can the brain detect fully task-irrelevant conflict?

In the first experiment, 24 human participants performed two behavioral tasks (see **figure 1A**). In the auditory conflict task (from hereon: content discrimination task I) the feature "sound content" was task-relevant. Participants were instructed to respond according to the content of the auditory stimulus ("left" vs "right"), ignoring its spatial location that could conflict with the content response (presented from the left or right side of the participant). For the other behavioral task, participants performed a demanding visual random dot-motion (RDM) task in which they had to discriminate the direction of vertical motion (from hereon: vertical RDM task),

while being presented with the same auditory stimuli—all features of which being fully irrelevant for task performance. Behavioral responses on this visual task were orthogonal to the response tendencies potentially triggered by the auditory features, excluding any task or response related interference (see **figure 1B**). Under this manipulation, all auditory features are task-irrelevant and are orthogonal to the response-mapping. To maximize the possibility of observing conflict detection when conflicting features are task-irrelevant, and to explore the effect of task automatization on conflict processing, participants performed the tasks both before and after extensive training, which may increase the efficiency of cognitive control (see **figure 1C**) (van Gaal et al., 2008).

To investigate whether our experimental design was apt to induce conflict effects for task-relevant sensory input and to test whether conflict effects were still present when sensory input was task-irrelevant, we performed repeated measures (rm-)ANOVAs (2x2x2 factorial) on mean reaction times (RTs) and error rates (ERs) gathered during the EEG-recording sessions (session 1, "before training"; session 4, "after training"). This allowed us to include 1) task-relevance (yes/no), 2) training (before/after) and 3) congruency of auditory content with location of auditory source (congruent/incongruent). Note that congruency is always defined based on the relationship between two features of the auditorily presented stimuli, also when participants performed the visual task (and therefore the auditory features were task-irrelevant).

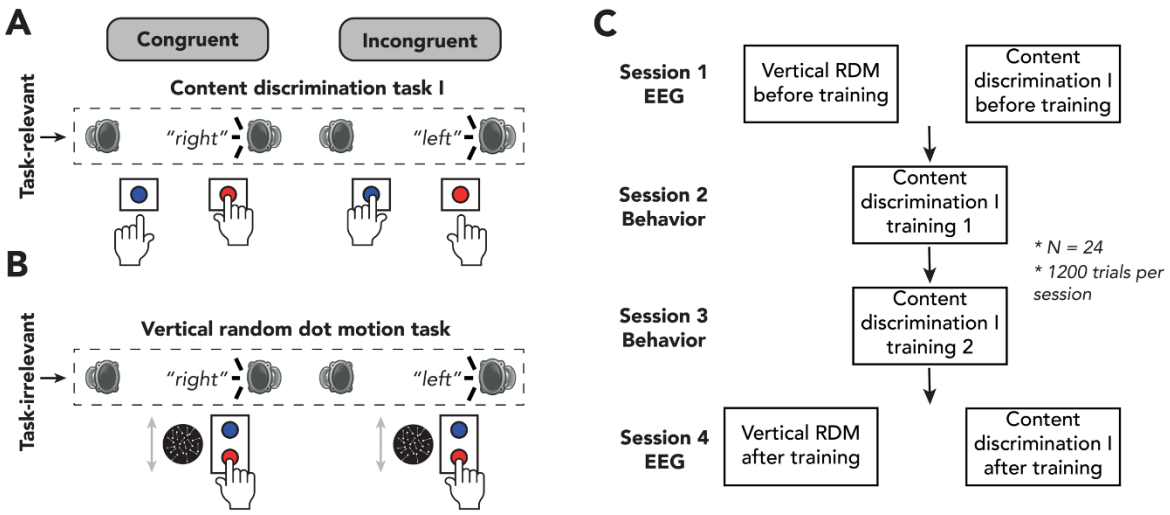


Figure 1. Experimental design of experiment 1. (A/B) Schematic representation of the experimental design for auditory content discrimination task I **(A)** and vertical RDM task **(B)**. In both tasks, the spoken words "left" and "right were presented through either a speaker located on the left or right side of the participant. Important to note is that auditory stimuli are only task-relevant in auditory content discrimination task I and not in the vertical RDM task. In this figure, sounds are only depicted as originating from the right, whereas in the experiment the sounds could also originate from the left speaker. **(A)** In content discrimination task I, participants were instructed to report the content ("left" or "right") of an auditory stimulus via a button press with their left or right hand, respectively, and to ignore the spatial location at which the auditory stimulus was presented. **(B)** During the vertical RDM task, participants were instructed to report the overall movement of dots (up or down) via a button press with their right hand, whilst still being presented with the auditory stimuli, which were therefore task-irrelevant. In both tasks, content of the auditory stimuli could be congruent or incongruent with its location of presentation (50% congruent/incongruent trials). **(C)** Overview of the sequence of the four experimental sessions of this study. Participants performed two EEG sessions during which they first performed the vertical RDM task followed by auditory content discrimination task I. Each session consisted of 1200 trials, divided over twelve blocks, allowing participants to rest in between blocks. In between experimental sessions, participants were trained on auditory content discrimination task I on two training sessions of 1 hour each.

Experiment 1: Conflicting information induces slower responses and decreased accuracy only for task-relevant sensory input

Detection of conflict is typically associated with behavioral slowing and increased error rates. Indeed, we observed that participants were slower and made more errors on incongruent trials as compared to congruent

trials (the conflict effect, RT: $F_{1,23}$=44.58, p<0.001, η_p^2=0.66; ER: $F_{1,23}$=9.13, p=0.006, η_p^2=0.284). This conflict effect was modulated by task-relevance of the auditory features (RT: $F_{1,23}$=152.76, p<0.001, η_p^2=0.87; ER: $F_{1,23}$=11.15, p=0.003, η_p^2=0.33) and post-hoc tests showed that the conflict effect was present when the auditory feature content was task-relevant ($RT_{cont(I)}$: $F_{1,23}$=285.00, p<0.001, η_p^2=0.93; $ER_{cont(I)}$: $F_{1,23}$=23.85, p<0.001, η_p^2=0.51**; figure 2A**, left panel), but not when they all auditory features were task-irrelevant (RT_{VRDM}: $F_{1,23}$=1.96, p=0.18, η_p^2=0.08, BF_{01}=5.71; ER_{VRDM}: $F_{1,23}$=0.26, p=0.62, η_p^2=0.01, BF_{01}=6.37; **figure 2A**, right panel). For content discrimination task I the conflict effect decreased in magnitude after extensive training, suggesting increased efficiency of within-trial conflict resolution mechanisms ($RT_{cont(I)}$: $F_{1,23}$=29.85, p<0.001, η_p^2=0.57; $ER_{cont(I)}$: $F_{1,23}$=9.76, p=0.005, η_p^2=0.30; data shown in supplementary **figure S1**). Over all trials, in content discrimination task I, mean RTs decreased after training, whereas ERs did not ($RT_{cont(I)}$: $F_{1,23}$=45.05, p<0.001, η_p^2=0.66; $ER_{cont(I)}$: $F_{1,23}$=1.77, p=0.20, η_p^2=0.07, BF_{01}=0.89, data not shown). All other interactions of current trial effects were not reliable (p>0.05). Data for all four runs of content discrimination task I are shown in the **supplementary information** (**figure S1**).

Experiment 1: Neural signatures of conflict detection only for task-relevant sensory features

The observation that task-irrelevant conflicting features had no effect on RTs and ERs, even after substantial training, whereas task-relevant conflicting features had, may not come as a surprise, because manual responses on the visual task (motion up/down with index and middle finger of right hand) were fully orthogonal to the potential conflicting nature of the auditory features (i.e. left/right). Further, content discrimination task I and the vertical RDM were independent tasks, requiring different cognitive processes. For example, mean RTs on the vertical RDM were on average 267ms longer than mean RTs for content discrimination task I (see **supplementary information**). However, caution is required in concluding that conflict detection is absent for task-irrelevant features based on these behavioral results alone, as neural and/or behavioral effects can sometimes be observed in isolation (one is observed but not the other) (Canales-Johnson et al., 2020; van Gaal et al., 2014). Therefore, in order to test whether unattended conflict is detected by the brain we turn to the multivariate analysis of our neural data.

Plausibly, the neural dynamics of conflict processing for task-irrelevant sensory input are different -in physical (across electrodes) and frequency space- from those related to the processing of conflict when sensory input is task-relevant. Therefore, we applied multivariate decoding techniques in the frequency domain to inspect whether and—if so—to what extent certain stimulus features were processed. These multivariate approaches have some advantages over traditional univariate approaches, for example, they are less sensitive to individual differences in spatial topography, because decoding accuracies are derived at a single participant level (Fahrenfort et al., 2018; Grootswagers et al., 2017; Haxby et al., 2001). Therefore, group statistics do not critically depend on the presence of effects in specific electrodes or clusters of electrodes. Further, although a wealth of studies have shown that conflict processing is related to an increase in power of theta-band neural oscillations (~4Hz-8Hz) after stimulus presentation (Cohen & Cavanagh, 2011; Jiang, Zhang, & van Gaal, 2015; Nigbur et al., 2012), it is unknown whether this is also the case for task-irrelevant conflict. By performing our MVPA analysis in frequency space, we could potentially find neural signatures in non-theta frequency-bands related to the processing of task-irrelevant conflict. However, due to the temporal and frequency space that has to be covered, strict multiple comparison corrections have to be performed (across time and frequency, see Methods). Therefore, we adopted an additional hypothesis-driven analysis, which also allowed us to obtain evidence for the absence of effects. Throughout this paper, we will discuss our neural data in the following order. First, the MVPA analyses in the frequency domain are presented for all critical features of the task (conflict, content, location, corrected for multiple comparisons, see **Methods**). Then, we report results from the additional hypothesis-driven analysis, where we extracted classifier accuracies from a predefined time-frequency ROI on which we performed (Bayesian) tests (see **Methods**).

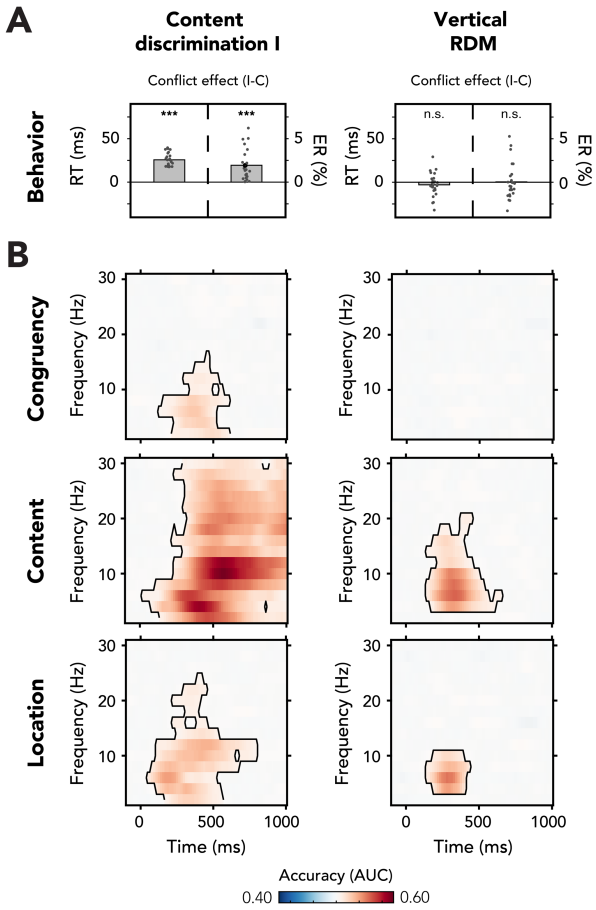


Figure 2. Behavioral and multivariate decoding results of experiment 1. (A/B) All results depicted here are from the merged data of both experimental sessions. The left column of plots shows the results for content discrimination task I, where auditory stimuli and conflicting features were task-relevant. The right column of plots shows the results for the vertical RDM, where neither the auditory stimulus nor its conflicting features were task-relevant. **(A)** The behavioral results are plotted as conflict effects (incongruent – congruent). Effects of conflict were present in content discrimination task I, with longer RTs (left bar) and increased error rates (right bar) for incongruent vs. congruent trials. For the vertical RDM task, no significant effects of conflict were found in behavior. Dots represent individual participants. **(B)** Multivariate classifier accuracies for different stimulus features. We trained classifiers on three stimulus features: auditory congruency, auditory content and auditory location. Classifier accuracies (AUC) are plotted across a time-frequency window of -100ms to 1000ms and 2Hz to 30Hz. Classifier accuracies are thresholded (cluster-base corrected, one-sided: $\bar{x}$>0.5, p<0.05) and significant clusters are outlined with a solid black line. Note that conflicting features of the auditory stimulus, content and location, could be decoded from neural data regardless of attention to the auditory stimulus. Information related to auditory congruency was present in a theta-band cluster, but only when the auditory stimulus was attended.

First, we trained a classifier on data from all EEG electrodes to distinguish between congruent vs. incongruent trials, for both content discrimination task I and the vertical RDM task (see **Methods**). Above-chance classification accuracies imply that relevant information about the decoded stimulus feature is present in the neural data, meaning that some processing of that feature occurred (Hebart & Baker, 2017). We performed our analysis on data from both EEG sessions separately to investigate effects of training. Accuracy scores of congruency decoding of the content discrimination task I did not differ significantly between sessions (p>0.05, two-sided, cluster-corrected; **supplementary figure S2**). Therefore, we subsequently conducted the decoding analysis on merged data from before and after the extensive behavioral training in the conflict-inducing content discrimination task I, thereby maximizing power to establish effects in our crucial comparisons. Thus, results reported here were obtained through the multivariate analyses of EEG-data of the two sessions combined. Decoding of congruency reveals that stimulus congruency was represented in neural data only when conflict was task-relevant (*p*<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-16Hz, peak frequency: 6Hz, time-range: 125ms-609ms, peak time: 359ms; **figure 2B, left panel**). The peak of the conflict effect was

observed in the theta-band (6Hz), which confirms a firm body of literature linking conflict detection to post-conflict modulations in theta-band oscillatory dynamics (Cavanagh & Frank, 2014; Cohen & Cavanagh, 2011; Nigbur et al., 2012). No significant cluster was found for the vertical RDM task (**figure 2B, right panel**). To qualify the absence of this effect, we followed up this hypothesis-free (with respect to frequency and time) MVPA-analysis with a hypothesis-driven analysis focused on the post-stimulus theta-band (2Hz-8Hz, 100ms-700ms). This more restricted analysis showed no significant effect (t_{23}=-0.24, p=.59, d=-0.05) and an additional Bayesian analysis revealed moderate evidence in favor of the null hypothesis (i.e. no effect of conflict on theta-band power) than the alternative hypothesis (BF_{01}=5.52).

Experiment 1: Stimulus features are processed in parallel, independent of task-relevance

Thus, cognitive control networks -or possible substitute networks- are seemingly not capable of detecting conflict when sensory features are task-irrelevant. However, the question is whether this observation is specific to the conflicting nature of the auditory stimuli, or whether the auditory stimuli are not processed whatsoever when attention is reallocated to the visually demanding task. To address this question, we trained classifiers on two other features of the auditory stimuli, i.e. location and content, to test whether these features were processed by the brain regardless of task-relevance. Indeed the content of auditory stimuli was processed both when the stimuli were task-relevant (p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-30 Hz, peak frequency: 10Hz, time-range: 16ms-1000ms, peak time: 578ms; **figure 2B, left panel**) and task-irrelevant (p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 4Hz-20Hz, peak frequency: 6Hz, time-range: 141ms-656ms, peak time: 344ms; **figure 2B, right panel**).

Similarly, the location of auditory stimuli could also be decoded from neural data for both content discrimination task I (p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-24Hz, peak frequency: 6Hz, time-range: 47ms-797ms, peak time: 188ms; **figure 2B, left panel**) and the vertical RDM task (p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 4Hz-10Hz, peak frequency: 6Hz, time-range: 141ms-438ms, peak time: 281ms; **figure 2B, right panel**). The above chance performance of the classifiers for the auditory stimulus features demonstrates that location and content information were processed, even when these features were task-irrelevant. Processing of task-irrelevant stimulus features was, however, more transient in time and more narrow-band in frequency as compared to processing of the same task-relevant features. Further, especially in content discrimination task I content decoding revealed a much broader frequency spectrum than any of the other comparisons. In the next experiment we show that this is related to the fact that this feature was response relevant and that this effect therefore partially reflects response preparation and response execution processes.

Moreover, the features content and location of the auditory stimulus could also be decoded from data obtained in the first session of the vertical RDM, when participants had not yet performed the task in which conflicting features were task-relevant (see **supplementary figure S3**). During this first session, the auditory stimuli had therefore not been coupled to any response-mappings whatsoever, thereby excluding potential effects of acquired task-relevance of these stimuli as a result of frequent exposure to the conflict task. Thus, task-related response-mappings belonging to the conflict task do not explain our observations of task-irrelevant feature processing in the vertical RDM task. Summarizing, we show that when conflicting features are truly and consistently task-irrelevant, they are no longer detected by –nor relevant to– the conflict monitoring system, but their individual features (content and location) are still processed.

In conclusion, in the first experiment we observed neural signatures of the processing of sensory stimulus features (i.e. location and content of auditory stimulus) regardless of task-relevance of these features, but a lack of integration of these features to form conflict when the auditory stimulus was fully task-irrelevant.

Experiment 2: Does detection of conflict depend on task-relevance of the stimulus or its individual features?

The experimental design of the first experiment rendered the auditory features to be located at the extreme ends of the scale of task-relevance, i.e. either the conflicting features were task-relevant and the conflicting features were consistently mapped to specific responses, or the conflicting features were task-irrelevant and the conflicting features were not mapped to responses. However, to further understand the relationship between the relevance of the conflicting features and the overlap with responses, we performed a second experiment containing four behavioral tasks. For this second experiment we recruited 24 new participants. We included two auditory conflicting tasks, similar to content discrimination task I. In one of the auditory tasks (from hereon: content discrimination task II, **figure 3A**) participants again had to respond according to the content of the auditory stimulus, whereas in the other auditory task (from hereon: location discrimination task, **figure 3B**) they were instructed to report from which side the auditory stimulus was presented (i.e. left or right speaker).

Furthermore, we included two new tasks in which the conflicting features (location and content) were not task-relevant and participants responded to a non-conflicting feature that was part of the conflicting stimulus (from hereon: volume oddball detection task, **figure 3C**) or the auditory stimulus was task-irrelevant but its features location and content overlapped with the responses to be given (from hereon: horizontal RDM task, **figure 3D**). The horizontal RDM task was similar to vertical RDM task of experiment 1, however the dots were now moving in a horizontal plane. Thus, participants were instructed to classify the overall movement of moving dots to either the left or the right. As this is a visual paradigm, the simultaneously presented auditory stimuli are fully task-irrelevant. However, both features of conflict, the content (i.e. 'left' and 'right') and the location (i.e. left and right speaker), of the auditory stimuli could potentially interfere with participants' responses on the visual task, thereby inducing a crossmodal Stroop-like type of conflict (Stroop, 1935).

In the volume oddball detection task, participants were presented with the same auditory stimuli as before, however occasionally (1/8 trials) they were presented at a lower volume. Participants were instructed to detect these volume oddballs by pressing the spacebar with their right hand as fast as possible. If they did not hear an oddball, they were instructed to withhold from responding. Participants performed 500 trials of the volume oddball detection task twice, at the very beginning of a session and at the end of a session (**figure 3E**). The rationale behind this was to see if effects in both runs differed because in the intermediate tasks sound location and sound content were relevant. In other words, during the first run sound content and sound location had never been related to any behavioral responses, whereas during the second run they might have acquired some intrinsic relevance through training on the other tasks, despite the fact that they were not task-relevant within the context of the task the participant is performing at that time. Theoretically, the selection of an object's feature (e.g. volume) may lead to the selection of all of its features (e.g. sound content, location), as suggested by theories of object-based attention (Chen, 2012), which may lead to conflict detection if this would be the case.

Again, EEG was recorded while participants performed these tasks, in order to see if auditory conflict was detected when the auditory stimulus, or its conflicting features (i.e. location and content) were task-irrelevant. In order to keep sensory input similar, moving dots (coherence: 0) were presented on the monitor during content discrimination task II, the location discrimination task and the volume oddball detection task, but these could be ignored.

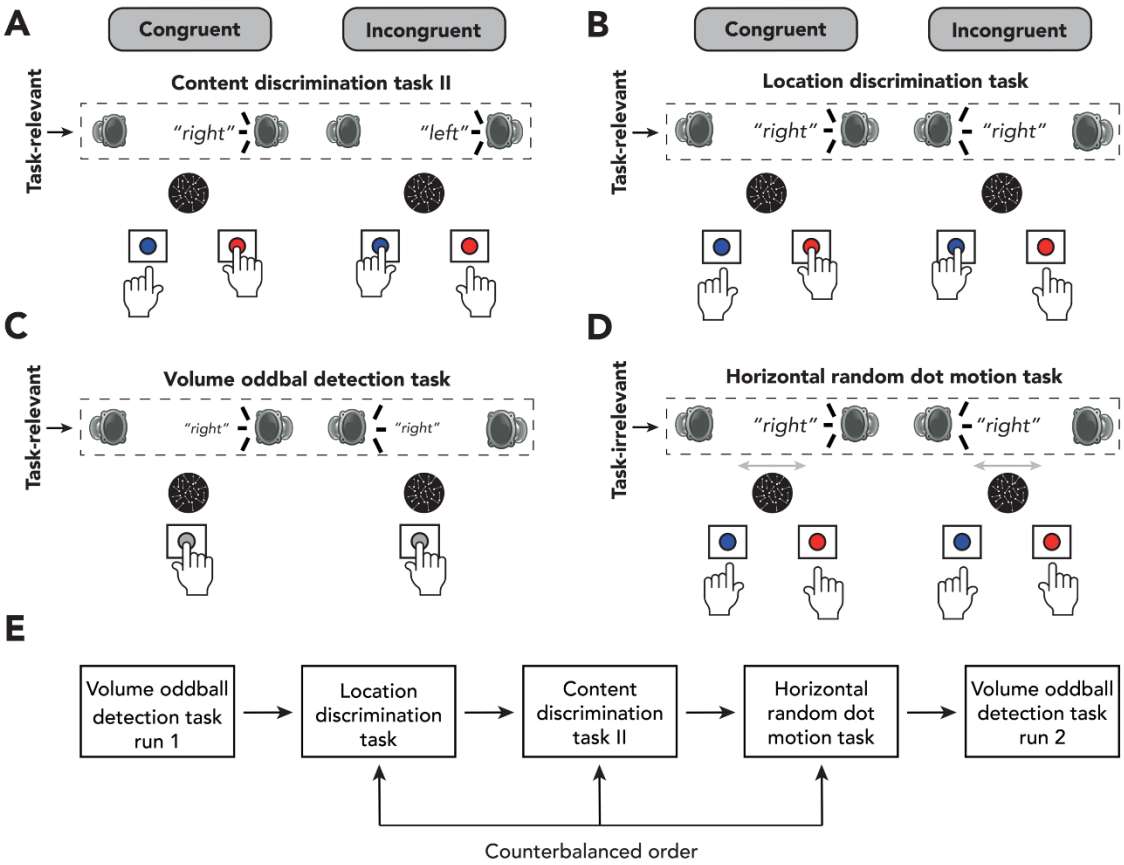


Figure 3. Experimental design of experiment 2. (A/B/C/D) Schematic representation of the experimental design for auditory content discrimination task II **(A)**, location discrimination task **(B)**, volume oddball task **(C)** and horizontal RDM task **(D)**. In all tasks, the spoken words "left" and "right were presented through either a speaker located on the left or right side of the participant. **(A)** In auditory content discrimination task II, participants were instructed to report the content ("left" or "right") of an auditory stimulus via a button press with their left or right hand, respectively, and to ignore the location of the auditory stimulus that was presented. **(B)** In the auditory location discrimination task, participants were instructed to report the location (left or right speaker) of an auditory stimulus via a button press with their left or right hand, respectively, and to ignore the content of the auditory stimulus that was presented. **(C)** During the volume oddball task, participants were instructed to detect auditory stimuli that were presented at a lower volume than ordinary stimuli (i.e. oddballs) by pressing the spacebar with their right hand. **(D)**. In the horizontal RDM, participants were instructed to report the overall movement of dots (left or right) via a button press with their left and right hands respectively, whilst still being presented with the auditory stimuli. In all four tasks, content of the auditory stimuli could be congruent or incongruent with its location of presentation (50% congruent/incongruent trials). **(E)** Order of behavioral tasks in experiment 2. Participants always started with the volume oddball task, followed by the location discrimination task, content discrimination task and horizontal RDM in randomized order and ended with another run of the volume oddball task.

Experiment 2: Behavioral effects of conflict only for task-relevant auditory sensory input

We will first discuss the behavioral results of the content discrimination, location discrimination and horizontal RDM tasks, as behavioral data for the volume oddball task is represented in d', rather than ER. Repeated measures ANOVAs (3x2 factorial) were performed on mean RTs and ERs from these three tasks, with the factors 1) task and 2) congruency of the auditory features. Again, congruency always relates to the combination of the auditory stimulus features sound content ("left" vs. "right") and sound location (left speaker vs. right speaker).

We observed that participants were slower and made more errors on incongruent trials as compared to congruent trials (RT: $F_{1,23}$=75.41, p<0.001, η_p^2=0.77; ER: $F_{1,23}$=68.00, p<0.001, η_p^2=0.75; **figure 4A**). This conflict effect was modulated by task (RT: $F_{1.55,35.71}$=22.80, p<0.001, η_p^2=0.50; ER: $F_{1.58,36.36}$=10.18, p<0.001, η_p^2=0.31) and post-hoc paired samples t-tests (incongruent – congruent) showed that conflict effects were present in the location discrimination task (RT_{loc}: t_{23}=5.03, p<0.001, d=1.03; ER_{loc}: t_{23}=6.25, p<0.001, d=1.28), content

discrimination task II ($RT_{cont(II)}$: t_{23}=8.95, p<0.001, d=1.83; $ER_{cont(II)}$: t_{23}=5.93, p<0.001, d=1.21), but not in the horizontal RDM task (RT_{HRDM}: t_{23}=1.44, p=0.164, d=0.29, BF_{01}=1.88; ER_{HRDM}: t_{23}=1.65, p=0.11, d=0.34, BF_{01}=1.44).

For the volume oddball task, d' scores were calculated as a measure of performance rather than percentage correct, due to the unbalanced number of targets (i.e. only 1/8 trials was an oddball) (Green & Swets, 1966). We will also test the effect of auditory congruency on reaction times which, by virtue of task-instruction, only covers oddball trials in which a correct response was made (i.e. hits). Repeated measures ANOVAs (2x2 factorial) were performed on RTs and d' values, with the factors 1) run number and 2) feature congruency.

On average participants had 40.54 hits (SD: 12.78) out of 61.83 oddball trials (SD: 8.57), per run of 500 trials. We observed no effects of auditory conflict on the RTs of these hit trials ($F_{1,23}$=2.78, p=0.11, η_p^2=0.10, BF_{01}=3.34; **figure 4A**). Furthermore, there was no difference in RT between the two runs ($F_{1,23}$=0.02, p=0.90, η_p^2=0.001, BF_{01}=5.21). There was also no interaction between run number and congruency in RT ($F_{1,23}$=0.09, p=0.77, η_p^2=0.004, BF_{01}=52.63). We did observe that sensitivity increased for incongruent (mean=2.66, SD=0.77) versus congruent trials (mean=2.11, SD 0.81), as measured in d' ($F_{1,23}$=45.62, p<0.001, η_p^2=0.67; **figure 4A**). Also, sensitivity was increased in the second run as compared to the first run ($F_{1,23}$=15.32, p<0.001, η_p^2=0.40, data not shown), however, there was no significant interaction between run number and auditory stimulus congruency ($F_{1,23}$=1.40, p=0.25, η_p^2=0.06, BF_{01}=3.13). These latter results thus indicate that task-irrelevant conflicting features resulted in behavioral measures of conflict detection in the volume oddball task.

Experiment 2: Detection of conflict occurs when any feature of a stimulus is task-relevant (even a non-conflicting feature), but never for a stimulus that is entirely task-irrelevant

We again trained multivariate classifiers on single-trial time-frequency data to test whether the auditory stimulus features (i.e. content, location and congruency) were processed when (I) the auditory conflicting features were task-relevant and overlapped with the response-mapping (content and location discrimination tasks), (II) the auditory conflicting features were task-irrelevant and another stimulus of the conflicting stimulus was task-relevant (volume oddball task), or when (III) the auditory conflicting features was task-irrelevant, but its conflicting features overlapped with the response-mapping in the task (horizontal RDM task) (see **Methods**).

Neural signatures of conflict processing in the theta-band were clearly present when one of the conflicting features was task-relevant, i.e. both when the content of the auditory stimulus was task-relevant (content discrimination task II: p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-20Hz, peak frequency: 4Hz, time-range: 266ms-906ms, peak time: 438ms; **figure 4B**) and when the location of the auditory stimulus was task-relevant (location discrimination task: p=0.002, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-6Hz, peak frequency: 4Hz, time-range: 328ms-813ms, peak time: 516ms; **figure 4B**). However, for both the volume oddball task and the horizontal RDM task, no significant clusters of above-chance classifier accuracy were found after multiple comparisons (**figure 4B**). This was surprising for the volume oddball task, because even though a non-conflicting feature was task-relevant, strong and significant behavioral measures of conflict detection were observed in this task (e.g. d' higher for incongruent than congruent trials, see **figure 4A**). Again, we followed up these hypothesis-free (with respect to frequency and time) MVPA analysis with a hypothesis driven analysis focused on the post-stimulus theta-band (2Hz-8 Hz, 100ms-700ms). Indeed, as expected, this more restricted analysis revealed significant above chance decoding in the volume oddball task for congruency (t_{23}=2.29, p=0.016, d=0.47, BF_{10}=3.66), in agreement with the behavioral results. Although significant, this conflict effect was clearly weaker than for the tasks where one of the conflicting features was task-relevant. When no feature of the conflict-containing stimulus was task-relevant, no effects of conflict were apparent (t_{23}=0.00, p=0.49, d=0.00, BF_{01}=4.65). Therefore, to qualify the differences between tasks, below we combine the data from all experiments and compare effect sizes across tasks (see **Figure 5**).

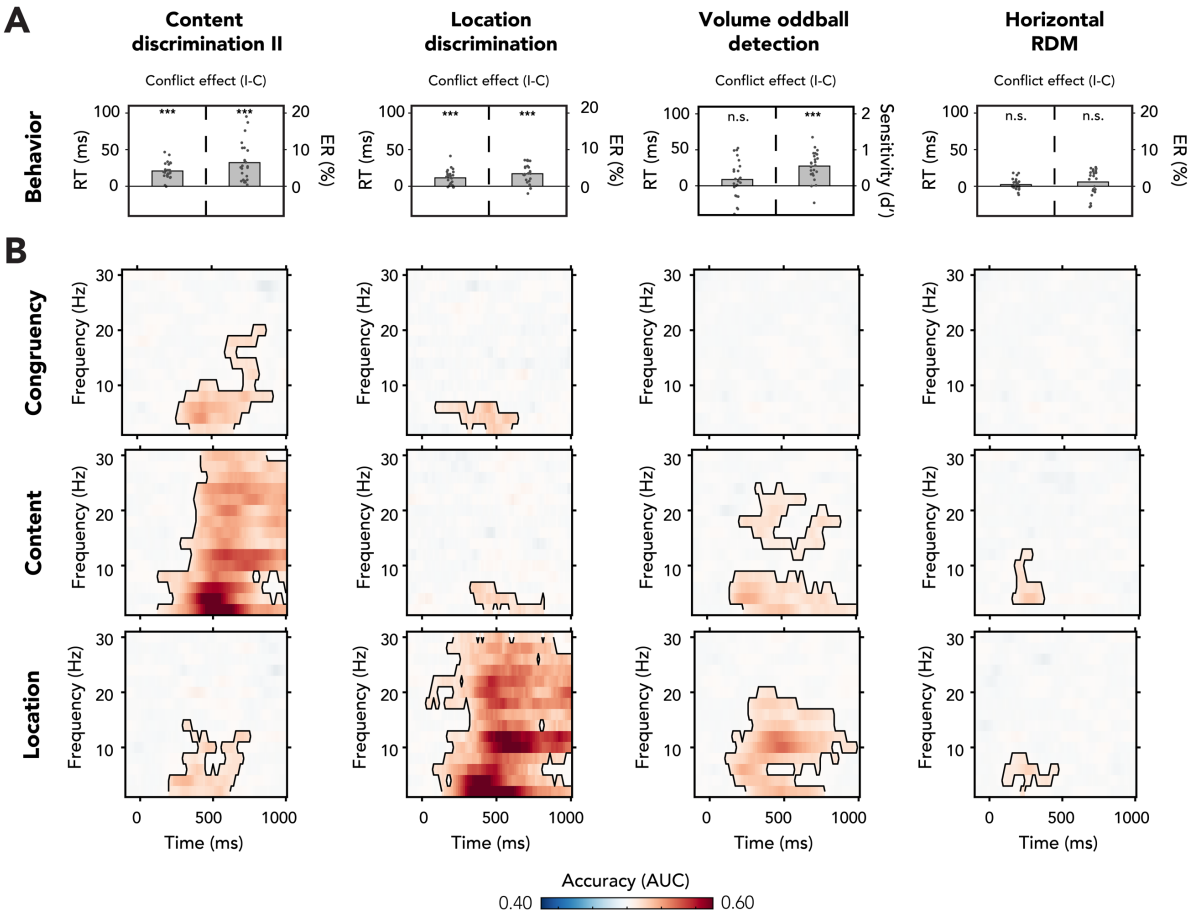


Figure 4. Behavioral and multivariate decoding results for experiment 2. (A/B) The four columns show data belonging to, from left to right, content discrimination task II, the location discrimination task, the volume oddball detection task and the horizontal RDM task. **(A)** Behavioral results are plotted as conflict effects (incongruent – congruent). Effects of conflict were present in all tasks where the auditory stimulus was task-relevant (content discrimination task II, location discrimination task and volume oddball). In both auditory discrimination tasks we observed longer RTs (left bar) and increased error rates (right bar) for incongruent vs. congruent trials. For the volume oddball we did not observe an effect in RT, but increased sensitivity (d') on incongruent vs. congruent trials. Dots represent individual participants. **(B)** Multivariate classifier accuracies for different stimulus features. We trained classifiers on three stimulus features: auditory congruency, auditory content and auditory location. Classifier accuracies (AUC) are plotted across a time-frequency window of -100ms to 1000ms and 2Hz to 30Hz. Classifier accuracies are thresholded (cluster-base corrected, one-sided: $\bar{x}$>0.5, p<0.05) and significant clusters are outlined with a solid black line. Note that the conflicting features of the auditory stimulus, i.e. content and location, could be decoded from neural data of regardless of attention to the auditory stimulus and task-relevance of its conflicting features. Clusters of above-chance decoding accuracy were only present for the two auditory tasks where both the auditory stimulus was attended and at least one of its conflicting features was task-relevant, namely content discrimination task II and the location discrimination task. Information related to auditory congruency was observed in theta- to alpha band clusters for these tasks. For the volume oddball task and the horizontal RDM no above-chance decoding clusters were observed. Note that the data shown for the volume oddball task was merged over both sessions.

Experiment 2: Sensory features processed irrespective of occurrence of conflict

Thus, we observed that conflict is not detected when the auditory stimulus itself is task-irrelevant (when one of its non-conflicting features is relevant we do observe conflict processing), irrespective of any overlap in response mappings with any of its features (i.e. sound content, location). To once more verify that this was specifically due to hampered functioning of the control network and not to disrupted sensory processing, we trained classifiers to distinguish trials based on sound location and content. We found neural signatures of the processing of sound content for all four tasks: content discrimination II (p<0.001, cluster-corrected; frequency-range: 2Hz-30Hz, peak frequency: 4Hz, time-range: 125ms-1000ms, peak time: 484ms; **figure 4B**), location discrimination (p<0.001, cluster-corrected; frequency-range: 2Hz-6Hz, peak frequency: 6Hz, time-range: 94ms-641ms, peak time: 453ms; **figure 4B**), and horizontal RDM content (p<0.01, cluster-corrected; frequency-range: 4Hz-12Hz, peak frequency: 4Hz, time-range: 156ms-344ms, peak time: 297ms; **figure 4**). For the volume oddball

task we observed two clusters, where content decoding was significantly above chance. A theta-band cluster (p<0.001, cluster-corrected; frequency-range: 2-8Hz, peak frequency: 4Hz, time-range: 156ms-1000ms, peak-time: 281ms; **figure 4**) and an alpha/beta-band cluster (p<0.001, cluster-corrected; frequency-range: 12Hz-24Hz, peak frequency: 18Hz, time-range: 219ms-891ms, peak time: 766ms; **figure 4B**).

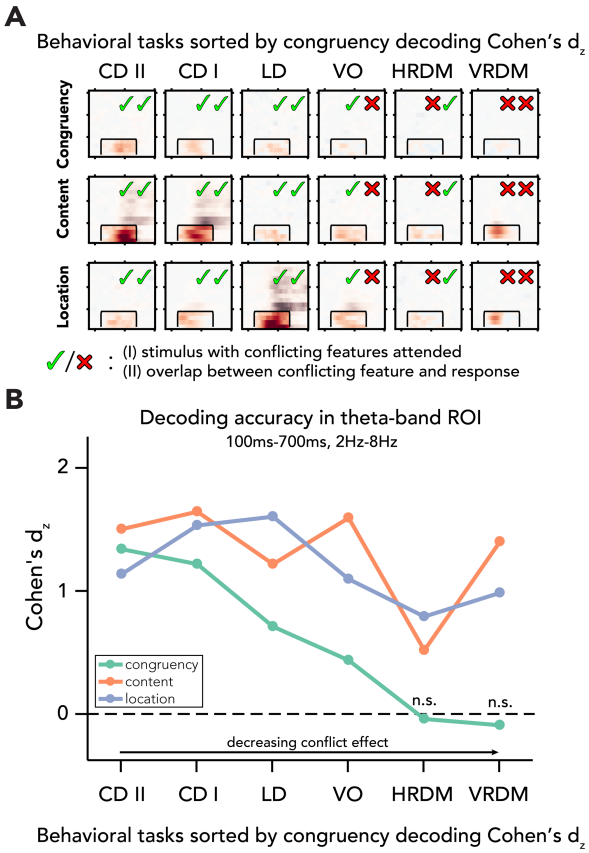
Furthermore, sound location could also be decoded from all tasks: content discrimination II (p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-14Hz, peak frequency: 4Hz, time-range: 203ms-750ms, peak time: 313ms; **figure 4B**), location discrimination (p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-30Hz, peak frequency: 2Hz, time-range: 31ms-1000ms, peak time: 453ms; **figure 4B**), volume oddball (p<0.001, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-20Hz, peak frequency: 10Hz, time-range: 109ms-984ms, peak time: 484ms; **figure 4B**), and horizontal RDM (p<0.01, one-sided: $\bar{x}$>0.5, cluster-corrected; frequency-range: 2Hz-8Hz, peak frequency: 6Hz, time-range: 94ms-469ms, peak time: 266ms; **figure 4B**). Two aspects of these results are worth highlighting. First, when participants responded to the location of the auditory stimulus, decoding revealed a broadband power spectrum, similar to when sound content was task-relevant in experiment 1. This overlap reveals that the broadband effect observed in Experiment 1 and 2 reflects response preparation and response execution processes. Second, the time-frequency (T-F) windows of significant sound content and location decoding in the volume oddball task was considerably more extended in both time and frequency space than observed in the two RDM tasks. This highlights that the presence of object-based attention in the volume oddball task, because one feature of the auditory stimulus was task-relevant, led to the rapid attentional selection and hence neural enhancement of the task-irrelevant features sound content and location this task only (Chen, 2012; O'Craven et al., 1999; Schoenfeld et al., 2014; Wegener et al., 2014).

In conclusion, in experiment 2 processing of conflicting stimulus features (i.e. location and content of auditory stimulus) was absent when none of the features of the auditory stimulus where task-relevant. Furthermore, we show that it is specifically important that the auditory stimulus is attended that matters whether conflicting features of the stimulus will be detected. First, in the volume oddball task neither sound location nor content were related to the responses in the primary task and there were still effects of conflict in both behavioral and neural data. This reveals that when the stimulus itself is attended (object-based attention), all of its features seem to be processed and integrated to elicit conflict. Second, in the horizontal RDM, sound location and content were related to the responses in the primary task, although the auditory stimulus was not task-relevant, and under this manipulation, effects of conflict in behavior and neural data were absent.

All experiments: Decreasing task-relevance hampers cognitive control operations, but not sensory processing

The neural data from all six different tasks over two experiments suggests that if sensory input is task-irrelevant, processing of sensory information is intact, while cognitive control operations are strongly deteriorated (**figures 2 and 4, 5A**). To quantify this observation, we calculated Cohen's d_z for all tasks and features (based on the preselected ROI), sorted the effect sizes based on the conflict effect (**figure 5A**) and plotted Cohen's d_z across tasks for all features (conflict, content and location, **figure 5B**). To be able to compare results from the different tasks we selected a time-frequency ROI that ranged from 100ms-700ms post-stimulus in the theta-band (frequency range: 2Hz-8Hz). For each feature and task, we extracted individual classifier AUC values and performed a 6x3 rm-ANOVA on these accuracies, with the factors task and stimulus feature. We found main effects for behavioral task and stimulus feature (task: $F_{1,2.17}$=20.27, p<0.001, η_p^2=0.47; stimulus feature: $F_{1,2}$=59.78, p<0.001, η_p^2=0.72). Crucially, the interaction between task and stimulus feature was also significant ($F_{1,3.14}$=26.30, p<0.001, η_p^2=0.53), showing that the accuracy of conflict decoding decreased more across tasks compared to content and location decoding. We next performed (Bayesian) t-tests (one-sided, $\bar{x}$>0.5) on ROI accuracies from every task/feature-combination to assess decoding performance for all stimulus features for the different task-relevance manipulations (see **supplementary table 1**). We observed that congruency

481 decoding accuracies were strongly influenced by task, showing a gradual decline as a factor of task-relevance,
482 whereas this was not the case for decoding accuracies of stimulus content and location (**figure 5B**). Note that
483 these results were robust and not dependent on the specific ROI selected, because using other ROI windows led
484 to similar patterns of results, i.e. decreased congruency decoding accuracy under task-irrelevant input but
485 relatively stable sensory feature decoding accuracy (**supplementary figure S4**).
486



487
488 **Figure 5. Processing of sensory and conflict features for different levels of task-relevance.** (**A**) Summary of the decoding results of all
489 behavioral tasks, sorted by congruency decoding effect size (Cohen's d_z) in a preselected time-frequency ROI. The data in these plots are
490 identical to the ones shown in **figures 2** and **4**. (**B**) To be able to compare feature decoding between tasks, we performed a hypothesis-
491 driven analysis for we extracted individual classifier accuracies from a pre-defined time-frequency ROI (100ms-700ms, 2Hz-8Hz) for every
492 behavioral task and stimulus feature. We performed a 6x3 rm-ANOVA on these accuracies, with the factors (I) behavioral task and (II)
493 stimulus feature. Behavioral task, stimulus feature and the interaction between them significantly modulated classifier accuracies. Next we
494 performed (Bayesian) one-sample t-tests (one-sided, $\bar{x}$>0.5) on every task/feature combination. Effect sizes for congruency decoding were
495 dependent on behavioral task (downward slope of the green line), whereas this was not the case, or less so, for the decoding of content and
496 location. Effect sizes are shown for all task/feature combinations. Behavioral tasks are shown on the x-axis and are sorted according to effect
497 size of congruency decoding. *Abbreviations:* **CD II**: content discrimination task II, CD I: content discrimination task I, **LD**: location
498 discrimination task, **VO**: volume oddball detection task, **HRDM**: horizontal RDM task, VRDM=vertical RDM task.

499

500 ## Discussion

501 Although it has been hypothesized for a long time that only basic physical properties of task-irrelevant sensory
502 input are processed (Treisman & Gelade, 1980), over the past few years an overabundance of processes have
503 been found to be preserved in the absence of attention (e.g. Fahrenfort et al., 2017; Fei-Fei et al., 2002; Peelen
504 et al., 2009; Sand & Wiens, 2011). Here we aimed to push the limits of the brains capacity to process unattended
505 information and addressed whether cognitive control networks could be recruited when conflicting features of
506 sensory input are task-irrelevant. Interestingly, similar cognitive control functions have been shown to occur
507 when stimuli are masked and hence conscious awareness is strongly reduced (Atas et al., 2016; D'Ostilio &
508 Garraux, 2012b; Huber-Huber & Ansorge, 2018; Jiang, Zhang, & van Gaal, 2015; Jiang et al., 2018; van Gaal et
509 al., 2008, 2010).

In two omnibus experiments with 6 different tasks, we presented potentially auditory-spatial conflicting stimulus features (e.g. the word "left" presented on the right side) to participants, whilst they were performing several behavioral tasks. These tasks manipulated whether the features sound content and location of the auditory stimulus were task-relevant and whether these features were mapped to specific overlapping responses to the primary task. We observed clear signals of conflict processing in behavior (i.e. longer reactions times, increased error rates, increased sensitivity) and brain activity (i.e. above chance decoding accuracy in the theta-band) when the conflicting features of the auditory stimulus were task-relevant, i.e. in the content and location discrimination tasks, and when another non-conflicting feature of the auditory stimulus was task-relevant, but the conflicting features content and location were not (volume oddball task). When none of the features of the auditory stimulus were task-relevant, i.e. in the vertical/horizontal RDM task, we did not observe any effects of conflict in behavior or neural measures, even when the task-irrelevant features mapped on the response scheme of the task (horizontal RDM task). The absence of conflict effects was supported by Bayesian analyses, showing reliable evidence in favor of the null hypothesis. Strikingly, the individual stimulus features, i.e. stimulus location/content, were always processed, regardless of their task-relevance. Note that this dissociation -hampered conflict processing yet intact perceptual processing- cannot be explained by a lack of statistical power, because decoding accuracy of stimulus location/content was comparable between behavioral tasks. These results highlight that relatively basic stimulus properties escape the attentional bottleneck, lending support to previous studies (e.g. Fahrenfort et al., 2017; Fei-Fei et al., 2002; Peelen et al., 2009; Sand & Wiens, 2011; Treisman & Gelade, 1980), but furthermore showcase that an attentional bottleneck for detecting conflict (integration of stimulus features) exists upstream in the hierarchy of cognitive processing. Below we link the observed results to the existing literature.

Object-based attention as a prerequisite for feature integration (leading to conflict)

So why are auditory content and location not integrated to form conflict when all of the auditory features are task-irrelevant? It has been suggested that the medial frontal cortex is crucial in monitoring the presence conflict, through the detection of coinciding inputs (Cohen, 2014). In our paradigm, it is thus crucial that information related to the auditory features content and location reaches the medial frontal cortex to be able to detect conflict. Previous studies have shown that task-irrelevant stimuli can still undergo elaborate processing (Fei-Fei et al., 2002; Peelen et al., 2009). Our decoding results show that task-irrelevant features are indeed processed by the brain (see **figure 5A**). Interestingly, conflict between two task-irrelevant features was only detected when another feature of the conflicting stimulus was the task-relevant stimulus (volume oddball task), but absent when none of the auditory features were task-relevant (RDM tasks). We argue that this difference is due to the fact that in the volume oddball task, the task-irrelevant conflicting features were selected through object-based attention. Theories of object-based attention have suggested that when one stimulus feature of an object is task-relevant and selected, attention "spreads" to all other features of the attended stimulus, even when these features are task-irrelevant (Chen, 2012; Chen & Cave, 2006; O'Craven et al., 1999; Wegener et al., 2014; Xu, 2010). In the volume oddball task a non-conflicting feature of the auditory stimulus (volume) was task-relevant, but this allowed for the selection of the other task-irrelevant features through object-based attention. Because of this selection, the available attentional resources can be exploited for the processing of these unattended features. When we inspect the time-frequency (T-F) windows of significant sound content and location decoding in the volume oddball task and compare this to the T-F windows observed in the RDM tasks, it is striking that the statistically reliable T-F windows are more extended in both time and frequency space in the volume oddball task compared to the RDM tasks (see **supplementary figure S5**). This may highlight that object-based attention led to the selection and neural enhancement of the task-irrelevant features only in the volume oddball task. In opposition, in the RDM tasks none of the features of the auditory stimulus was task-relevant and therefore object-based attention could not enhance the selection of the conflicting auditory features. When inspecting the T-F maps for the RDM tasks, the relatively fleeting temporal characteristics of the processing of the task-irrelevant stimulus features (sound content and location) might suggest that the

integration of these features may not be possible due to a lack of time, as proposed in the incremental grouping model of attention (Roelfsema, 2006; Roelfsema & Houtkamp, 2011). However, the time-window in which conflict was decodable when the auditory conflicting features were task-relevant coincides with the time-range in which these features could be decoded when the auditory conflicting features were task-irrelevant (see **figure 5A**). Therefore, it seems unlikely that the more temporally constrained processing of task-irrelevant stimulus features is the cause of hampered conflict detection.

Besides time being a factor, the processing of task-irrelevant features in the RDM tasks may have also been too constrained to (early) sensory cortices and therefore could not progress to integration networks, including the MFC, necessary for the detection of conflict. Speculatively, the processing of task-irrelevant auditory features was relatively superficial, due to the relatively few remaining resources (Lavie et al., 2004; Sigman & Dehaene, 2006; Zylberberg et al., 2010, 2011), and combined with a lack of object-based attention, this may have prevented the fast feedforward propagation of the information to the MFC. It has been hypothesized that unattended (sometimes referred to as "preconscious"; Dehaene et al., 2006; Dehaene & Changeux, 2011) stimuli are not propagated deeply in the brain but still allow for shallow recurrent interactions in sensory cortices. The poor spatial resolution of EEG-measurements and our experimental setup, however, do not allow to test these specific ideas regarding the involvement of spatially distinct cortices. Yet, previous work of our group suggests that task-irrelevant nonconscious information does not propagate to the frontal cortices, whereas task-relevant nonconscious information does. We demonstrated that masked task-irrelevant conflicting cues induced similar early processing in sensory cortices as compared to masked task-relevant cues, but prohibit activation of frontal cortices (van Gaal et al., 2008). These findings are not conclusive and so we believe that uncovering the role of task-relevance in processing of (nonconscious) information deserves more attention in future work (see also van Gaal et al., 2012 for a discussion on this issue).

Conflict between features of a task-irrelevant stimulus versus conflict between stimuli

Contrary to the current study, previous studies using a variety of conflict-inducing paradigms and attentional manipulations reported conflict effects in behavior and electrophysiological recordings induced by unattended stimuli or stimulus features (Mao & Wang, 2008; Padrão et al., 2015; Zimmer et al., 2010). However, our study deviates from those studies in several crucial aspects.

First, we explicitly separate task-relevant stimulus features that cause conflict and task-relevant features that do not, parsing the cognitive components that induce cognitive control in this context. Furthermore, in the RDM and volume oddball tasks we tested whether conflict between two task-irrelevant features could be detected by the brain. Specifically, we investigated if conflict between two task-irrelevant features would be detected in the presence or absence of object-based attention (volume oddball task vs. RDM tasks), also manipulating whether task-irrelevant conflicting features mapped onto the response (horizontal RDM) or not (vertical RDM). This approach is crucially different from previous studies that exclusively tested whether a task-irrelevant or unattended stimulus (feature) could interfere with processing of a task-relevant feature (Mao & Wang, 2008; Padrão et al., 2015; Zimmer et al., 2010). Under such conditions, at least one source contributing to the generation of conflict (i.e. the task-relevant stimulus) is fully attended and therefore one cannot claim that under those circumstance conflict detection occurs outside the scope of attention.

It can be argued that in our horizontal RDM task, the task-irrelevant auditory features (location and content) that mapped onto the response of the primary task could interfere with the processing of horizontal dot-motion, i.e. the task-relevant feature. This is in fact true, as we found effects of auditory content-dot motion and auditory location-dot motion conflict in RT and ER (see **supplementary figure S6**). This nicely highlights that a single feature of a task-irrelevant stimulus can interfere with the response to a task-relevant stimulus when there are overlapping feature-response mappings. This is different from two features of a task-irrelevant stimulus to produce inherent conflict (e.g. between auditory content and location), which is what we specifically

investigated by always testing the presence of auditory content-location conflict only. A similar argument might be made for our vertical RDM and volume oddball tasks, because in those cases the auditorily presented stimuli could potentially conflict with responses that were exclusively made with the right hand, e.g. the spoken word "left" or the sound from left location may conflict generally more with a right hand response (independent of the up/down classification or oddball detection) than the spoken word "right" or the sound from right location. In the vertical RDM task, where the auditorily presented stimuli were truly task-irrelevant, as both stimulus content and location did not affect mean RTs (see **supplementary figure S7**). In the volume oddball task sound content and location were task-irrelevant features, but these features were subject to object-based attention as they were part of the attended stimulus. In this behavioral task, the content of the auditory stimuli interfered with responses to the volume oddball task, resulting in longer RTs (see **supplementary figure S8**). Moreover, under this manipulation we did find behavioral and neural effects of conflict between two auditory features (see **figure 4**). The absence of conflict effects in the horizontal RDM and presence of such effects in the volume oddball task strengthens our conclusion that at least one feature of the stimulus containing the conflicting features should be task-relevant -and thus the presence of object-based attention- is what ultimately determines whether conflict can be detected. Summarizing, we show that the brain is not able to detect conflict that emerges between two features of a task-irrelevant stimulus (in the absence of object-based attention).

Lastly, in other studies conflicting stimuli were often task-irrelevant on one trial (e.g. because they were presented at an unattended location) but task-relevant on the next (e.g. because they were presented at the attended location) (e.g. Padrão et al., 2015; Zimmer et al., 2010). Such trial-by-trial fluctuations of task-relevance allow for across-trial modulations to confound any current trial effects (e.g. conflict-adaptation effect) and also induce a "stand-by attentional mode" where participants never truly disengage to be able to determine if a stimulus is task-relevant. We prevented such confounding effects in the present study, where the (potentially) conflicting features or the auditory stimulus itself were task-irrelevant on every single trial in the vertical RDM, horizontal RDM and volume oddball task.

Differences between response conflict and perceptual conflict cannot account for absence of conflict detection in task-irrelevant sensory input

One difference between the content and location discrimination tasks on the one hand and the volume oddball and RDM tasks on the other, was the task-relevance of the (conflicting) auditory features. Another major difference between these groups of tasks was, consequently, the origin of the conflict. When the auditory stimuli were task-relevant, the origin of conflict was found in the interference of a task-irrelevant feature on behavioral performance, whereas for the other tasks this was not the case. We argued that in the volume oddball and RDM tasks, salient auditory stimuli could be *intrinsically* conflicting. Intrinsic conflict is often referred to as perceptual conflict, as opposed to the aforementioned behavioral conflict (Kornblum, 1994). Although perceptual conflict effects are usually weaker than response conflict effects, both in behavior and electrophysiology (Frühholz et al., 2011; van Veen et al., 2001; Wang et al., 2014), this difference in the origin of the conflict is unlikely to explain why we did not observe effects of conflict under task-irrelevant sensory input, as opposed to earlier studies.

First, several neurophysiological studies have previously reported electrophysiological modulations by perceptual conflict over the MFC (Jiang, Zhang, & Van Gaal, 2015; Nigbur et al., 2012; van Veen et al., 2001; Wang et al., 2014; Zhao et al., 2015). Second, an earlier study using a task similar to ours (but including only task-relevant stimuli) showed effects of perceptual conflict, i.e. unrelated to response-mapping, in terms of behavioral and neural data (Buzzell et al., 2013). Third, the prefrontal monitoring system has previously been observed to respond when participants view other people making errors (Jääskeläinen et al., 2016; Van Schie et al., 2004), suggesting that cognitive control can be triggered without the need to make a response. Fourth, in our volume oddball task, where conflict was perceptual in nature as well, we did observe conflict effects, both in behavior and neural data. Thus, the absence of conflict-related effects for task-irrelevant sensory input (in the

visual RDM tasks) suggests that task-relevance of a stimulus is indeed necessary for the detection of conflict.

Inattentional deafness or genuine processing of stimulus features?

The lack of conflict effects in behavior and EEG when auditory conflict features were task-irrelevant (in the absence of object-based attention) might suggest a case of inattentional deafness, a phenomenon known to be induced by demanding visual tasks, which manifests itself in weakened early (~100ms) auditory evoked responses (Molloy et al., 2015). Interestingly, human speech seems to escape such load modulations and is still processed when unattended and task-irrelevant (Olguin et al., 2018; Röer et al., 2017; Zäske et al., 2016), potentially because of its inherent relevance, similar to (other) evolutionary relevant stimuli such as faces (Finkbeiner & Palermo, 2017; Lavie et al., 2003). Indeed, the results of our multivariate analyses demonstrate that spoken words are processed (at least to some extent) when they are task-irrelevant, as stimulus content (the words "left" and "right", middle row **figure 4B**) and stimulus location (whether the word was presented on the left or the right side", bottom row **figure 4B**) could be decoded from time-frequency data for all behavioral tasks. For all tasks, classification of stimulus content was most accurate in the theta-band which is in line with a previously proposed theoretical role for theta-oscillations in speech processing, namely that they track the acoustic-envelope of speech (Giraud & Poeppel, 2012). In the volume oddball detection task, we also observed an alpha/beta-band cluster of above-chance decoding (ranging from 12-24Hz, peak in 18Hz), which has previously been linked to single word intelligibility (Becker et al., 2013). After this initial processing stage, further processing of stimulus content is reflected in more durable broadband activity for the content discrimination tasks, possibly related to higher-order processes (e.g. semantic) and response preparation (**figure 4**, two left columns).

Similarly to processing of stimulus content, early processing of stimulus location was most strongly reflected in the delta- to theta-range for all tasks (**figure 4**), which may relate to the auditory N1 ERP component, an ERP signal that is modulated by stimulus location (Fuentemilla et al., 2006; Lewald & Getzmann, 2011; Salminen et al., 2015). We also observed above-chance decoding in the alpha-band for task-relevant auditory stimuli, convergent with the previously reported role of alpha-band oscillations in orienting and allocating audio-spatial attention (Weisz et al., 2014).

Thus, the characteristics of the early sensory processing of (task-irrelevant) auditory stimulus features in our study are in line with recent findings of auditory and speech processing. Moreover, our observations are in line with recent empirical findings that suggest a dominant role for late control operations, as opposed to early selection processes, in resolving conflict (Itthipuripat et al., 2019). Specifically, this work showed that in a Stroop-like paradigm both target and distractor information is analyzed fully, after which the conflicting input is resolved. Extending on this, we show intact initial processing of task-irrelevant sensory input, but hampered late control operations necessary to detect conflict, at least in the current set-up.

Automatization of conflict processing does not promote detection of task-irrelevant conflict

Previous studies investigating unconscious conflict through masking procedures concluded that conflict detection by the MFC may happen automatically and is still operational under strongly reduced levels of stimulus visibility (D'Ostilio & Garraux, 2012c; Jiang, Zhang, & van Gaal, 2015; van Gaal et al., 2008). Such automaticity can often be enhanced through training of the task. For example, training in a stop-signal paradigm in which stop-signals were rendered unconscious through masking, led to an increase in the strength of behavioral modulations of these stimuli (van Gaal et al., 2009).

In order to see whether enhancing such automaticity could hypothetically increase the likelihood of conflict detection, we included extensive training sessions in the first experiment and measurements of the volume oddball task before and after exposure to conflicting tasks in the second experiment. In the first experiment, we found no neural effects of conflict detection in the vertical RDM task, even when participants had been trained

on the auditory task for 3600 trials (see **supplementary figure S2**). Training did result in decreased conflict effects in the auditory content-discrimination task of experiment 1, indicating that our training procedure was successful. Further, participants improved within-trial conflict resolution (see **supplementary figure S1**), suggesting more efficient functioning of conflict resolution mechanisms. In experiment 2, participants performed the volume oddball task twice, once before and once after sound location and content had been mapped to responses. Again, we aimed to see if training on conflict tasks would enhance automaticity of conflict processing in a paradigm where the auditory conflicting features were task-irrelevant. We did not find any statistically reliable differences in behavioral conflict effects (d', RT, see **supplementary figure S9**) or accuracy of congruency decoding between the two runs (see **supplementary figure S10**). Therefore, it seems that the automaticity of conflict detection by the MFC and associated networks do not hold when the auditory stimulus is task-irrelevant (at least after the extent of training and exposure as studied here).

Conclusion

Summarizing, high-level cognitive processes that require the integration of conflict inducing stimulus features are strongly hampered when none of the stimulus features of the conflict inducing stimulus is task-relevant and hence object-based attention is absent. This work nicely extends previous findings of perceptual processing outside the scope of attention (Peelen et al., 2009; Sand & Wiens, 2011; Schnuerch et al., 2016; Tusche et al., 2013), but suggest crucial limitations of the brain's capacity to process task-irrelevant "complex" cognitive control-initiating stimuli, indicative of an attentional bottleneck to detect conflict at high levels of information analysis. In contrast, the processing of more basic physical features of sensory input appears to be less deteriorated when input is task-irrelevant (Lachter et al., 2004).

Materials and methods

Participants

We performed two separate experiments, each containing multiple behavioral tasks. For each of these experiments we recruited 24 healthy human participants from the University of Amsterdam. None of the participants were took part in both experiments. So, in total 48 participants (37 female) aged 18 to 30, participated in this experiment for monetary compensation or participant credits. All participants had normal or corrected-to-normal vision and had no history of head injury or physical and mental illness. This study was approved by the local ethics committee of the University of Amsterdam and written informed consent was obtained from all participants after explanation of the experimental protocol. We will describe the experimental design and procedures for the two experiments separately. Data analyses and statistics were similar across experiments and will be discussed in the same section.

Experiment 1: design and procedures

Participants performed two tasks in which conflicting auditory stimuli were either task-relevant or task-irrelevant. In both tasks, conflict was elicited through a paradigm adapted from (Buzzell et al., 2013), in which spatial information and content of auditory stimuli could interfere. In content discrimination task I participants had to respond to the auditory stimuli, whereas in vertical RDM task participants had to perform a demanding RDM-task, while the auditory conflicting stimuli were still presented (**figure 1A**). Participants performed both tasks on two experimental sessions of approximately 2.5 hours. In between these two experimental sessions, participants had two training sessions of one hour during which they only performed the task-relevant task (**figure 1B**). On each experimental session, participants first performed a shortened version of the RDM-task to determine the appropriate coherency of the moving dots (73%-77% correct), followed by the actual task-irrelevant auditory task and finally the task-relevant auditory task. Participants were seated in a darkened, sound isolated room, 50cm from a 69x39cm screen (frequency: 120Hz, resolution: 1920 x 1080, RGB: 128, 128, 128). Both tasks were programmed in MATLAB (R2012b, The MathWorks, Inc.), using functionalities from Psychtoolbox (Kleiner et al., 2007).

Experiment 1: behavioral tasks

Auditory content discrimination task I

In the task-relevant auditory conflict task, the spoken words "links" (i.e. "left" in Dutch) and "rechts" (i.e. "right" in Dutch) were presented through speakers located on both sides of the participant (**figure 1A**). Auditory stimuli were matched in duration and sample rate (44 kHz) and were recorded by the same male voice. By presenting these stimuli through either the left or the right speaker, content-location conflict arose on specific trials (e.g. the word "left" through the right speaker). Trials were classified accordingly, as either congruent (i.e. location and content are the same) or incongruent (i.e. location and content are different). Participants were instructed to respond as fast and accurate as possible, by pressing left ("a") or right ("l") on a keyboard located in front of the participants, according to the stimulus content, ignoring stimulus location. Responses had to be made with the left or right index finger, respectively. The task was divided in twelve blocks of 100 trials each, allowing participants to rest in between blocks. After stimulus presentation, participants had a 2s period in which they could respond. A variable inter-trial-interval between 850ms-1250ms was initiated directly after the response. If no response was made, the subsequent trial would start after the 2s response period. Congruent and incongruent trials occurred equally often (i.e. 50% of all trials), as expectancy of conflict has been shown to affect conflict processing (Soutschek et al., 2015). Due to an error in the script, there was a disbalance in the amount of trials coming from the left (70%) vs. right (30%) speaker location for the first 14 participants. However, the amount of congruent vs. incongruent and “left” vs. “right” trials were equally distributed. For the upcoming analyses, all trial classes were balanced in trial count.

Vertical random-dot motion task

In the task-irrelevant auditory task, participants performed an RDM-task in which they had to discriminate the motion (up versus down) of white dots (n=603) presented on a black circle (RGB: 0, 0, 0; ~14° visual angle; see **figure 1A**). Onset of the visual stimulus was paired with the presentation of the auditory conflicting stimulus. Participants were instructed to respond according to the direction of the dots, by pressing the "up" or "down" key on a keyboard with their right hand as fast and accurate as possible. Again, participants could respond in a 2s time-interval which was terminated after responses and followed by an inter-trial-interval of 850-1250ms. Task difficulty, in terms of dot-motion coherency (i.e. proportion of dots moving in the same direction) was titrated between blocks to 73-77% correct of all trials within that block. Similar to content discrimination task I, the vertical RDM was divided in twelve blocks containing 100 trials each, separated by short breaks. Again, congruent and incongruent trials, with respect to the auditory stimuli, occurred equally often.

Experiment 2: design and procedures

In the second experiment, we wanted to investigate whether it is task-irrelevance of the auditory stimulus itself or task-irrelevance of the auditory features (i.e. content and location) that determine whether prefrontal control processes are hampered. Participants performed two tasks in which auditory stimuli were fully task-relevant (location-discrimination and content-discrimination), one task in which the auditory stimulus was relevant but the features auditory location and content were not (volume oddball) and one task in which the auditory stimulus itself was task-irrelevant, but its features location and content could potentially interfere with behavior (horizontal RDM). Participants came to the lab for one session, lasting three hours. Each session started with the volume oddball task, followed by (in a counterbalanced order) the location discrimination, content discrimination and horizontal RDM tasks, and ended with a block of the volume oddball task again.

We included the location discrimination and content discrimination tasks both to replicate the results of experiment 1 and also to see if there would be differences in these results between the two tasks. Specifically, we investigated whether processing of auditory stimulus features -as indicated by multivariate classification accuracies- would differ between the two tasks. Participants were seated in a darkened, sound isolated room, 50cm from a 69x39cm screen (frequency: 120Hz, resolution: 1920 x 1080, RGB: 128, 128, 128). All four experiments were programmed in Python 2.7 using functionalities from PsychoPy (Peirce et al., 2019).

Experiment 2: behavioral tasks

Auditory content discrimination task (II)

The auditory content discrimination task from experiment 2 is a (near identical) replication of the auditory content discrimination task from experiment 1. Participants were fixating on a fixation mark in the center of the screen. Again, the spoken words "links" (i.e. "left" in Dutch) and "rechts" (i.e. "right" in Dutch) were presented through speakers located on both sides of the participant. Participants were instructed to respond according to the stimulus content by pressing left (“A”) or right (“L”) on a keyboard located in front of them, with their left and right index fingers respectively. Concurrently, on every trial, a black disk with randomly moving dots (coherence: 0) was presented to keep sensory input similar between tasks. After stimulus presentation, participants had an 800ms period in which they could respond. After a response, the response window would be terminated directly. A variable inter-trial-interval between 250ms-450ms was initiated directly after the response. If no response was made, the subsequent trial would start after the ITI. All stimulus features (i.e. sound content, location and congruency) were presented in a balanced manner (e.g. 50% congruent, 50% incongruent). The task was divided in six blocks of 100 trials each, allowing participants to rest in between blocks.

Auditory location discrimination task

The auditory location discrimination task was identical to the auditory content discrimination task II, with the exception that participants were now instructed to respond according to the location of the auditory stimulus.

Thus, participants had to press a left button (“A”) for sounds coming from a left speaker and right button (“L”) for sounds coming from a right speaker. Again, participants performed six blocks of 100 trials.

Volume oddball task

In the volume oddball task, the same auditory stimuli were presented. Again, on every trial, a black disk with randomly moving dots (coherence: 0) was presented to keep sensory input similar between tasks. Occasionally an auditory stimulus would be presented at a lower volume. The initial volume of the oddballs was set to 70% , but was staircased in between blocks to yield 83%-87% correct answers. If participants performance on the previous block was below or above this range the volume increased or decreased with 5% respectively. The odds of a trial being an oddball trial were 1/8 (drawn from a uniform distribution). Participants were instructed to detect these oddballs, by pressing the space-bar as fast as possible whenever they thought they heard a volume oddball. If they thought that the stimulus was presented at a normal volume, they were instructed to refrain from responding. The response interval was 800ms, which was terminated at response. A variable inter-trial-interval 150ms-350ms was initiated after this response interval. Participants performed two runs of this task, at the beginning of each session and at the end of each session. Each run contained five blocks of 100 trials each.

Horizontal RDM task

In the horizontal RDM task participants had to discriminate the motion (left versus right) of white dots (n=603) presented on a black circle (RGB: 0, 0, 0; ~14° visual angle). Onset of the visual stimulus was paired with the presentation of the auditory stimulus. Participants were instructed to respond according to the direction of the dots, by pressing a left key ("A"), or right key ("L") on a keyboard with left and right index fingers respectively as fast and accurate as possible. Participants could respond in an 800ms time-interval which was terminated after responses and followed by an inter-trial-interval of 250-450ms. Task difficulty, in terms of dot-motion coherency (i.e. proportion of dots moving in the same direction) was set to 0.3 in the first block. This value indicated an intermediate coherence, as every trial in that block could be the intermediate coherence, but also half (0.15) or twice this intermediate coherence (0.6). The intermediate coherence was titrated between blocks to 73-77% correct. If behavioral performance fell outside that range, 0.01 was added to or subtracted from the intermediate coherence. The horizontal RDM consisted of 10 blocks of 60 trials. All stimulus features (i.e. sound content, location and congruency) were presented in a balanced manner (e.g. 50% congruent, 50% incongruent).

Data analysis

We were primarily interested in the effects of congruency of the auditory stimuli on both behavioral and neural data. Therefore, we defined trial-congruency on the basis of these auditory stimuli, in all behavioral tasks of the two experiments. All behavioral analyses were programmed in MATLAB (R2017b, The MathWorks, Inc.).

Analysis of behavioral data

Missed trials and trials with an RT <100ms or >1500ms were excluded from behavioral analyses. In order to investigate whether current trial conflict effects were present both when conflict was task-relevant and task-irrelevant and to inspect if training of the auditory task affected such conflict effects, we performed rm-ANOVAs on RTs and d’-scores for the volume oddball task of experiment 2 and on RTs and ERs for all other tasks. For the tasks from experiment 1, we performed these ANOVAs with task-relevance, training (before vs. after) and current trial congruency as factors (2x2x2 factorial design). Additional post-hoc rm-ANOVAs, for content discrimination task I and vertical RDM task separately (2x2 factorial design), were used to inspect the origin of significant factors that were modulated by task-relevance. If the assumption of sphericity was violated, we applied a Greenhouse-Geisser correction.

For content discrimination task I, the location discrimination task and horizontal RDM task we performed a rm-

ANOVA with task and congruency as factors (3x2 factorial design). Additional post-hoc paired sample t-tests were performed per task to test the difference between incongruent and congruent trials.

In case of null findings, we performed a Bayesian analysis (rm-ANOVA or paired sample t-test) with identical parameters and settings on the same data, to test if there was actual support of the null hypothesis (JASP Team, 2018).

Analysis of EEG data

EEG data were analyzed using custom-made software written in MATLAB, with support from the toolboxes EEGLAB (Delorme & Makeig, 2004) and ADAM (Fahrenfort et al., 2018) .

Acquisition and preprocessing

EEG-data were recorded with a 64-channel BioSemi apparatus (BioSemi B.V., Amsterdam, The Netherlands), at 512 Hz. Vertical eye-movements were recorded with electrodes located above and below the left eye, horizontal eye-movements were recorded with electrodes located at the outer canthi of the left and the right eye. All EEG traces were re-referenced to the average of two electrodes located on the left and right earlobes (mastoidal reference for 1 participant in experiment 1). The data were not filtered, as filtering can induce artifacts in multivariate classification of neural data (van Driel et al., 2019). Also, no automatic or manual artefact rejection was performed, in order to preserve as many trials as possible and to refrain from arbitrary data selection. Epochs were created by taking data from -1s to 2s around onset of stimulus presentation.

Multivariate pattern analysis (decoding)

We applied a multivariate backwards decoding model to EEG data that was transformed to the time-frequency domain. We used multivariate analyses both because its higher sensitivity in comparison with univariate analyses and to inspect if and to what extent different stimulus features (i.e. location and content) were processed in both tasks, without having to preselect spatial or time-frequency ROIs. The ADAM-toolbox was used on epoched EEG data, that were transformed to time-frequency using default methods of the toolbox but with specific settings (epochs: -100ms to 1000ms, 2Hz-30Hz in linear steps of 2Hz) (Fahrenfort et al., 2018). Trials were classified according to current trial stimulus features (i.e. location and content) resulting in 4 trial types. As decoding algorithms are known to be time-consuming, epochs were resampled to 64Hz. Next, a decoding algorithm, using either stimulus location, stimulus content or congruency as stimulus class, was applied on the data according to a tenfold cross-validation scheme. Specifically, a linear discriminant analysis (LDA) was used to discriminate between stimulus classes (e.g. left versus right speaker location etc.). Classification accuracy was computed as the area under the curve (AUC), a measure derived from Signal Detection Theory (Green & Swets, 1966).

The multivariate classifiers were on different subsets of trials, depending on the behavioral task. For the auditory tasks (content discrimination I and II, location discrimination and volume oddball detection), only correct trials were included in the analysis as errors tend to elicit a similar, albeit not identical, neural response as cognitive conflict. Errors are more likely on incongruent trials and therefore error trials had to be excluded. For the volume oddball detection task we additionally excluded all oddball trials, thus only testing correct rejections, in order to prevent conflict arising between responses made exclusively with the right hand and sound content and location. For the two visual tasks (horizontal and vertical RDM) we trained the classifier on all trials.

Statistical comparisons

The AUC scores we obtained via multivariate analyses of our EEG data were tested per time-point and frequency with one-sided t-tests ($\bar{x}$>0.5) across participants against chance-level (50%). These t-test were corrected for multiple comparisons over time, using cluster-based permutation tests ($p<0.05$, 1000 iterations). This procedure

yields time clusters of significant above-chance classifier accuracy, indicative of information processing. Note that this procedure yields results that should be interpreted as fixed effects (Allefeld et al., 2016), but is nonetheless standard in the scientific community.

In addition to the cluster analysis, we performed a hypothesis-driven analyses on classifier accuracies that were extracted from a pre-defined time-frequency ROI. Specifically, for every task and every stimulus feature (i.e. congruency, content, location), we extracted mean ROI accuracies per subject. This ROI was selected visually, to cover a time- and frequency-range in which most significant clusters could be found. The ROI ranged from 100ms-700ms and 2Hz-8Hz. We then applied a 6x3 rm-ANOVA on these accuracies with factors task and stimulus feature. Next, one-sample t-tests (one-sided, $\bar{x}>0.5$) were performed on every task/feature combination to determine whether decoding accuracy of a specific feature within our pre-selected ROI was above chance during the various behavioral tasks. Additional Bayesian one-sample t-tests (one-sided, $\bar{x}>0.5$, Cauchy scale=0.71) were performed to inspect evidence in favor of the null-hypothesis that decoding accuracy was not above chance.

We performed the same analysis on different ROIs. The results of those analyses can be found in the **Supplementary information**.

Literature

Allefeld, C., Görgen, K., & Haynes, J. D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, *141*, 378–392. https://doi.org/10.1016/j.neuroimage.2016.07.040

Atas, A., Desender, K., Gevers, W., & Cleeremans, A. (2016). Dissociating perception from action during conscious and unconscious conflict adaptation. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000206

Becker, R., Pefkou, M., Michel, C. M., & Hervais-Adelman, A. G. (2013). Left temporal alpha-band activity reflects single word intelligibility. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00121

Broadbent, D. E. (1958). *Perception and Communication*. Oxford University Press.

Buzzell, G. A., Roberts, D. M., Baldwin, C. L., & McDonald, C. G. (2013). An electrophysiological correlate of conflict processing in an auditory spatial Stroop task: The effect of individual differences in navigational style. *International Journal of Psychophysiology*, *90*(2), 265–271. https://doi.org/10.1016/j.ijpsycho.2013.08.008

Canales-Johnson, A., Beerendonk, L., Blain, S., Kitaoka, S., Ezquerro-Nassar, A., Nuiten, S., Fahrenfort, J., Gaal, S. van, & Bekinschtein, T. A. (2020). Decreased alertness reconfigures cognitive control networks. *In Press*, 831727. https://doi.org/10.1101/831727

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics*, *74*(5), 784–802. https://doi.org/10.3758/s13414-012-0322-z

Chen, Z., & Cave, K. R. (2006). When does visual attention select all features of a distractor? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1452–1464. https://doi.org/10.1037/0096-1523.32.6.1452

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with 2 ears. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1907229

Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends in Neurosciences*, *37*(9), 480–490. https://doi.org/10.1016/j.tins.2014.06.004

Cohen, M. X., & Cavanagh, J. F. (2011). Single-Trial Regression Elucidates the Role of Prefrontal Theta Oscillations in Response Conflict. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00030

Cosman, J. D., & Vecera, S. P. (2012). Object-based attention overrides perceptual load to modulate visual distraction. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(3), 576–579. https://doi.org/10.1037/a0027406

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*. https://doi.org/10.1037/h0042712

D'Ostilio, K., & Garraux, G. (2012a). Brain mechanisms underlying automatic and unconscious control of motor action. *Frontiers in Human Neuroscience*, *6*(September), 1–5. https://doi.org/10.3389/fnhum.2012.00265

D'Ostilio, K., & Garraux, G. (2012b). Dissociation between unconscious motor response facilitation and conflict in medial frontal areas. *European Journal of Neuroscience*, *35*(2), 332–340. https://doi.org/10.1111/j.1460-9568.2011.07941.x

D'Ostilio, K., & Garraux, G. (2012c). Dissociation between unconscious motor response facilitation and conflict in medial frontal areas. *European Journal of Neuroscience*, *35*(2), 332–340. https://doi.org/10.1111/j.1460-9568.2011.07941.x

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–1790. https://doi.org/10.1038/nn1594

Fahrenfort, J. J., Van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA using the Amsterdam Decoding and Modeling toolbox (ADAM). *Frontiers in Neuroscience - Brain Imaging Methods*, *12*(July). https://doi.org/10.3389/fnins.2018.00368

Fahrenfort, J. J., van Leeuwen, J., Olivers, C. N. L., & Hogendoorn, H. (2017). Perceptual integration without conscious access. *Proceedings of the National Academy of Sciences*, *114*(14), 3744–3749. https://doi.org/10.1073/pnas.1617268114

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

Finkbeiner, M., & Palermo, R. (2017). The Role of Spatial Attention in Nonconscious Processing A Comparison of Face and Nonface Stimuli. *Psychological Science*, *20*(1), 42–51.

Frühholz, S., Godde, B., Finke, M., & Herrmann, M. (2011). Spatio-temporal brain dynamics in a combined stimulus-stimulus and stimulus-response conflict task. *NeuroImage*, *54*(1), 622–634. https://doi.org/10.1016/j.neuroimage.2010.07.071

Fuentemilla, L., Marco-Pallarés, J., & Grau, C. (2006). Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *NeuroImage*, *30*(3), 909–916. https://doi.org/10.1016/j.neuroimage.2005.10.036

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Goldman-Rakic. (1995). Architecture of the prefrontal cortex and the central executive. In *Structure and functions of the human prefrontal cortex* (pp. 71–83). New York Academy of Sciences.

Goldman-Rakic. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1445–1453. https://doi.org/10.1098/rstb.1996.0129

Green, D. M., & Swets, J. a. (1966). Signal detection theory and psychophysics. *Society*. https://doi.org/10.1901/jeab.1969.12-475

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science (New York, N.Y.)*, *293*, 2425–2430. https://doi.org/10.1126/science.1063736

Hebart, M. N., & Baker, C. I. (2017). Deconstructing multivariate decoding for the study of brain function. *NeuroImage*, *August*, 1–15. https://doi.org/10.1016/j.neuroimage.2017.08.005

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, *136*(2), 189–202. https://doi.org/10.1016/j.actpsy.2010.04.011

Huber-Huber, C., & Ansorge, U. (2018). Unconscious conflict adaptation without feature-repetitions and response time carry-over. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 169–175. https://doi.org/10.1037/xhp0000450

Itthipuripat, S., Deering, S., & Serences, J. T. (2019). When Conflict Cannot be Avoided: Relative Contributions of Early Selection and Frontal Executive Control in Mitigating Stroop Conflict. *Cerebral Cortex*, 1–12. https://doi.org/10.1093/cercor/bhz042

Jääskeläinen, I. P., Tapani, K., Lahnakoski, J. M., Manoach, D. S., Sams, M., Ahveninen, J., Halme, H.-L., Agam, Y., & Glerean, E. (2016). Neural mechanisms supporting evaluation of others' errors in real-life like conditions. *Scientific Reports*, *6*(1), 1–10. https://doi.org/10.1038/srep18714

JASP Team. (2018). JASP (Version 0.8.6.0). In *[Computer software].*

Jiang, J., Correa, C. M., Geerts, J., & van Gaal, S. (2018). The relationship between conflict awareness and behavioral and oscillatory signatures of immediate and delayed cognitive control. *NeuroImage*, *177*(May), 11–19. https://doi.org/10.1016/j.neuroimage.2018.05.007

Jiang, J., Zhang, Q., & van Gaal, S. (2015). Conflict awareness dissociates theta-band neural dynamics of the medial frontal and lateral frontal cortex during trial-by-trial cognitive control. *NeuroImage*, *116*, 102–111. https://doi.org/10.1016/j.neuroimage.2015.04.062

Jiang, J., Zhang, Q., & Van Gaal, S. (2015). EEG neural oscillatory dynamics reveal semantic and response conflict at difference levels of conflict awareness. *Scientific Reports*, *5*(July), 1–11. https://doi.org/10.1038/srep12008

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-O

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoobox-3? *Perception*.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. https://doi.org/10.1016/j.tics.2006.10.012

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–384. https://doi.org/10.1016/j.actpsy.2010.03.010

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research*, *56*(3), 130–135. https://doi.org/10.1007/BF00419699

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): Still no identification without attention. In *Psychological Review*. https://doi.org/10.1037/0033-295X.111.4.880

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18. https://doi.org/10.1016/S1364-6613(02)00013-X

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 1–9. https://doi.org/10.10162236

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, *133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological Science*, *14*(5), 510–515. https://doi.org/10.1111/1467-9280.03453

Lewald, J., & Getzmann, S. (2011). When and where of auditory spatial processing in cortex: A novel approach using electrotomography. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0025146

Mao, W., & Wang, Y. (2008). The active inhibition for the processing of visual irrelevant conflict information. *International Journal of Psychophysiology*, *67*(1), 47–53. https://doi.org/10.1016/j.ijpsycho.2007.10.003

Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattentional Deafness: Visual Load Leads to Time-Specific Suppression of Auditory Evoked Responses. *Journal of Neuroscience*, *35*(49), 16046–16054. https://doi.org/10.1523/JNEUROSCI.2931-15.2015

Moray, N. (1959). Attention in Dichotic Listening: Affective Cues and the Influence of Instructions. *Quarterly Journal of Experimental Psychology*, *11*(1). https://doi.org/10.1080/17470215908416289

Nigbur, R., Cohen, M. X., Ridderinkhof, K. R., & Stürmer, B. (2012). Theta Dynamics Reveal Domain-specific Control over Stimulus and Response Conflict. *Journal of Cognitive Neuroscience*, *24*(5), 1264–1274. https://doi.org/10.1162/jocn_a_00128

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). FMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–587. https://doi.org/10.1038/44134

Olguin, A., Bekinschtein, T. A., & Bozic, M. (2018). Neural encoding of attended continuous speech under different types of interference. *Journal of Cognitive Neuroscience*, *30*(11), 1606–1619. https://doi.org/10.1162/jocn_a_01303

Padrão, G., Rodriguez-Herreros, B., Pérez Zapata, L., & Rodriguez-Fornells, A. (2015). Exogenous capture of medial-frontal oscillatory mechanisms by unattended conflicting information. *Neuropsychologia*, *75*, 458–468. https://doi.org/10.1016/j.neuropsychologia.2015.07.004

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97. https://doi.org/10.1038/nature08103

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rahnev, D. A., Huang, E., & Lau, H. C. (2012). Subliminal stimuli in the near absence of attention influence top-down cognitive control. *Attention, Perception, and Psychophysics*, *74*(3), 521–532. https://doi.org/10.3758/s13414-011-0246-z

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447. https://doi.org/10.1126/science.1100301

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*(1), 203–227. https://doi.org/10.1146/annurev.neuro.29.051605.112939

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception & Psychophysics*, *73*(8), 2542–2572. https://doi.org/10.3758/s13414-011-0200-0

Röer, J. P., Körner, U., Buchner, A., & Bell, R. (2017). Semantic priming by irrelevant speech. *Psychonomic Bulletin and Review*, *24*(4), 1205–1210. https://doi.org/10.3758/s13423-016-1186-3

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, *8*(8), 363–370. https://doi.org/10.1016/j.tics.2004.06.003

Salminen, N. H., Takanen, M., Santala, O., Lamminsalo, J., Altoè, A., & Pulkki, V. (2015). Integrated processing of spatial cues in human auditory cortex. *Hearing Research*, *327*, 143–152. https://doi.org/10.1016/j.heares.2015.06.006

Sand, A., & Wiens, S. (2011). Processing of unattended, simple negative pictures resists perceptual load. *NeuroReport*, *22*(7), 348–352. https://doi.org/10.1097/WNR.0b013e3283463cb1

Schnuerch, R., Kreitz, C., Gibbons, H., & Memmert, D. (2016). Not quite so blind: Semantic processing despite inattentional blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 459–463. https://doi.org/10.1037/xhp0000205

Schoenfeld, M. A., Hopf, J.-M., Merkel, C., Heinze, H.-J., & Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience*, *17*(4), 619–624. https://doi.org/10.1038/nn.3656

Sigman, M., & Dehaene, S. (2006). Dynamics of the Central Bottleneck: Dual-Task and Task Uncertainty. *PLoS Biology*, *4*(7), e220. https://doi.org/10.1371/journal.pbio.0040220

Simons, D. J., & Chabris, C. F. (1999). Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p281059

Soutschek, A., Stelzel, C., Paschke, L., Walter, H., & Schubert, T. (2015). Dissociable Effects of Motivation and Expectancy on Conflict Processing: An fMRI Study. *J Cogn Neurosci*, *27*(2), 409–423. https://doi.org/10.1162/jocn_a_00712

Stefanics, G., Csukly, G., Komlósi, S., Czobor, P., & Czigler, I. (2012). Processing of unattended facial emotions: A visual mismatch negativity study. *NeuroImage*, *59*(3), 3042–3049. https://doi.org/10.1016/j.neuroimage.2011.10.041

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. https://doi.org/10.1037/h0027242

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tusche, A., Kahnt, T., Wisniewski, D., & Haynes, J. D. (2013). Automatic processing of political preferences in the human brain. *NeuroImage*, *72*, 174–182. https://doi.org/10.1016/j.neuroimage.2013.01.020

van Driel, J., Olivers, C. N. L., & Fahrenfort, J. J. (2019). *High-pass filtering artifacts in multivariate classification of neural time series data* [Preprint]. Neuroscience. https://doi.org/10.1101/530220

van Gaal, S., de Lange, F. P., & Cohen, M. X. (2012). The role of consciousness in cognitive control and decision making. *Frontiers in Human Neuroscience*, *6*(May), 1–15. https://doi.org/10.3389/fnhum.2012.00121

van Gaal, S., Lamme, V. A. F., Fahrenfort, J. J., & Ridderinkhof, K. R. (2010). Dissociable Brain Mechanisms Underlying the Conscious and Unconscious Control of Behavior. *J.Cogn Neurosci.*, *0898-929X (Linking)*, 91–105. https://doi.org/10.1162/jocn.2010.21431

van Gaal, S., Naccache, L., Meuwese, J. D. I., van Loon, A. M., Leighton, A. H., Cohen, L., & Dehaene, S. (2014). Can the meaning of multiple words be integrated unconsciously? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), 20130212–20130212. https://doi.org/10.1098/rstb.2013.0212

van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). Frontal Cortex Mediates Unconsciously Triggered Inhibitory Control. *Journal of Neuroscience*, *28*(32), 8053–8062. https://doi.org/10.1523/JNEUROSCI.1278-08.2008

van Gaal, S., Ridderinkhof, K. R., van den Wildenberg, W. P. M., & Lamme, V. A. F. (2009). Dissociating Consciousness From Inhibitory Control: Evidence for Unconsciously Triggered Response Inhibition in the Stop-Signal Task. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1129–1139. https://doi.org/10.1037/a0013551

Van Schie, H. T., Mars, R. B., Coles, M. G. H., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, *7*(5), 549–554. https://doi.org/10.1038/nn1239

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior Cingulate Cortex, Conflict Monitoring, and Levels of Processing. *NeuroImage*, *14*(6), 1302–1308. https://doi.org/10.1006/nimg.2001.0923

VanRullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology*, *3*(1–2), 167–176. https://doi.org/10.2478/v10053-008-0022-3

Wang, K., Li, Q., Zheng, Y., Wang, H., & Liu, X. (2014). Temporal and spectral profiles of stimulus-stimulus and stimulus-response conflict processing. *NeuroImage*, *89*, 280–288. https://doi.org/10.1016/j.neuroimage.2013.11.045

Wegener, D., Galashan, F. O., Aurich, M. K., & Kreiter, A. K. (2014). Attentional spreading to task-irrelevant object features: Experimental support and a 3-step model of attention for object-based selection and feature-based processing modulation. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00414

Weisz, N., Müller, N., Jatzev, S., & Bertrand, O. (2014). Oscillatory alpha modulations in right auditory regions reflect the validity of acoustic cues in an auditory spatial attention task. *Cerebral Cortex*, *24*(10), 2579–2590. https://doi.org/10.1093/cercor/bht113

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? In *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn1411

Xu, Y. (2010). The Neural Fate of Task-Irrelevant Features in Object-Based Processing. *Journal of Neuroscience*, *30*(42), 14020–14028. https://doi.org/10.1523/JNEUROSCI.3011-10.2010

Zäske, R., Perlich, M. C., & Schweinberger, S. R. (2016). To hear or not to hear: Voice processing under visual load. *Attention, Perception, and Psychophysics*, *78*(5), 1488–1495. https://doi.org/10.3758/s13414-016-1119-2

Zhao, J., Liang, W. K., Juan, C. H., Wang, L., Wang, S., & Zhu, Z. (2015). Dissociated stimulus and response conflict effect in the Stroop task: Evidence from evoked brain potentials and brain oscillations. *Biological Psychology*, *104*, 130–138. https://doi.org/10.1016/j.biopsycho.2014.12.001

Zimmer, U., Itthipanyanan, S., Grent-'T-Jong, T., & Woldorff, M. G. (2010). The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention. *European Journal of Neuroscience*, *31*(10), 1744–1754. https://doi.org/10.1111/j.1460-9568.2010.07229.x

Zylberberg, A., Dehaene, S., Roelfsema, P., & Sigman, M. (2011). The human Turing machine: A neural framework for mental programs. *Trends in Cognitive Sciences*, *15*, 293–300. https://doi.org/10.1016/j.tics.2011.05.007

Zylberberg, A., Fernández Slezak, D., Roelfsema, P. R., Dehaene, S., & Sigman, M. (2010). The Brain's Router: A Cortical Network Model of Serial Processing in the Primate Brain. *PLoS Computational Biology*, *6*(4). https://doi.org/10.1371/journal.pcbi.1000765

Supplementary information and figures

Descriptive statistics of behavioral results of experiment 1

For content discrimination task I, error rates (ERs) and reaction times (RTs), averaged over all four sessions, were 2.6% (*s*=2.7%) and 477.2ms (*s*=76.1ms), respectively. For the vertical RDM were mean ERs were 19.2% (*s*=6.6%) and mean RTs were 711.4ms (*s*=151.3ms). The mean ER of vertical RDM indicates that our staircasing-procedure, 75% correct, was effective.

Effects of training on relative conflict effect and adaptation in experiment 1

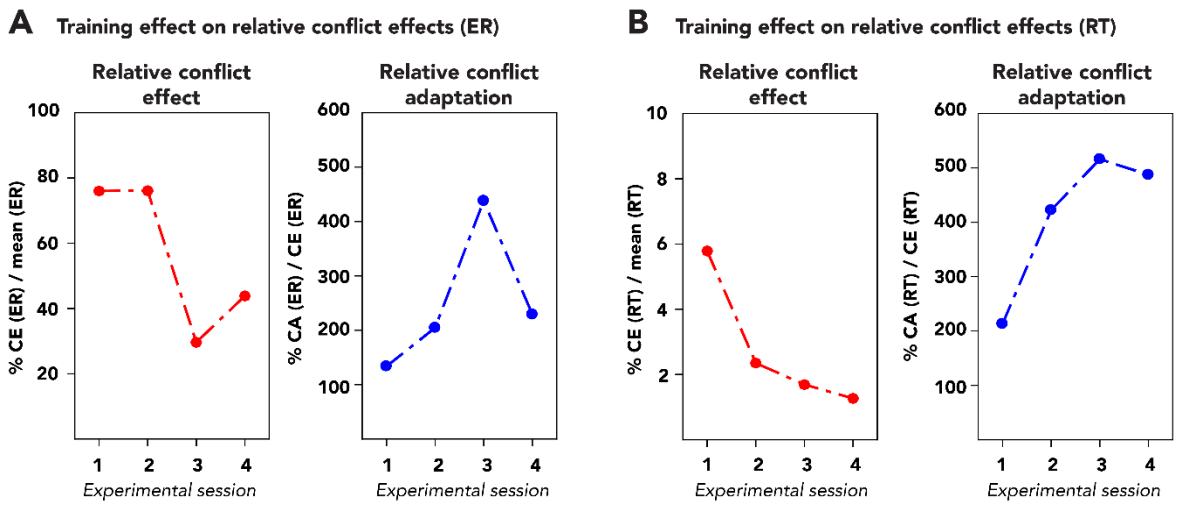


Figure S1. Effects of training on relative conflict effect and adaptation in terms of ER and RT. Relative measures of conflict effect and conflict adaptation provide better insight in the magnitude of effects, albeit unconventional measures. All values are grand averages, and statistics were not applied on them. **(A)** Training effects on relative conflict effect (left) and conflict adaptation (right) for ERs. Whereas relative conflict effect seems to decrease as a result of training, relative conflict adaptation appears to increase. This is suggestive of more efficient conflict resolution. **(B)** Training effects on relative conflict effect (left) and conflict adaptation (right) for RTs. Similar to ERs relative conflict effect decreases as a result of training, whereas relative conflict adaptation appears to increase. Again, this suggests more efficient conflict resolution.

Decoding performance of before and after training sessions of experiment 1

In our **Results** section we report the outcome of multivariate analyses on EEG-data of experiment 1 that was merged over the two experimental sessions. We merged these data to increase classifier performance. Before merging the data, we ran an identical analysis on the session-specific data. Crucially, no differences in congruency classifier performance were found between before and after training data for neither content discrimination task I nor vertical RDM task (p>0.05, cluster-corrected; see supplementary **figure S2**).

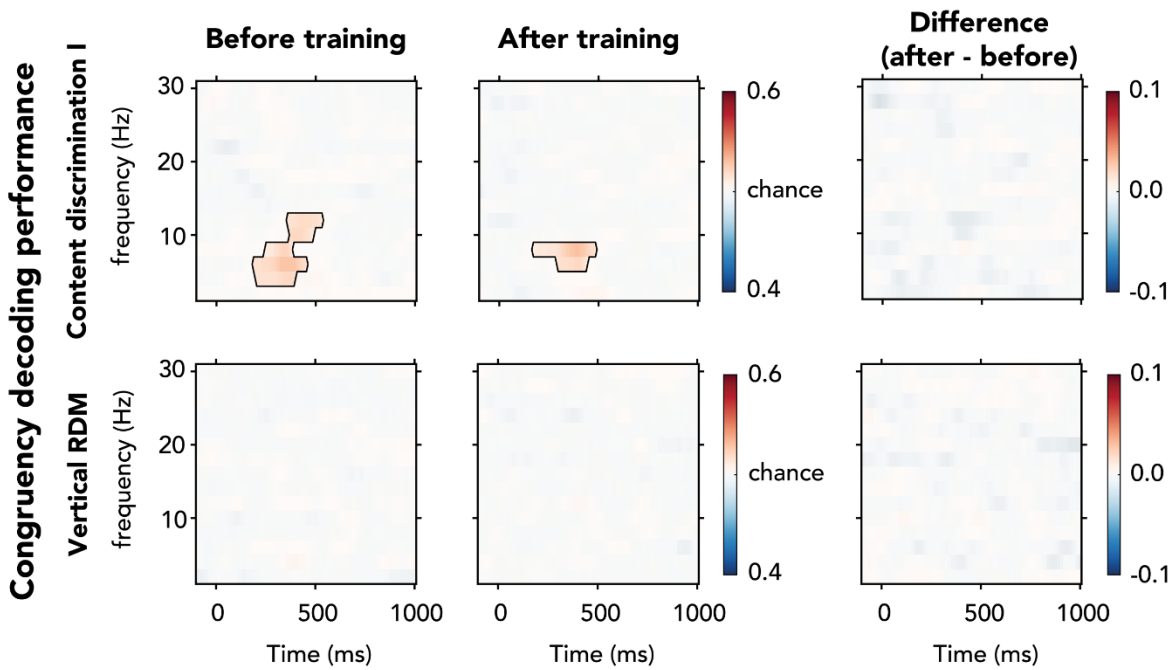


Figure S2. Between session difference (after training – before training) in classifier performance for congruency in experiment 1. Thresholded (cluster-base corrected, p<0.05) accuracies are depicted across the frequency-range (2-30Hz) and significant clusters are outlined with a solid black line. For both content discrimination task I and vertical RDM, classifier accuracy was not affected by extensive training of the task.

Excluding task-related effects of content discrimination task I on decoding in vertical RDM

In our experimental design, we excluded across-trial fluctuations of task-relevance of the auditory stimulus, by applying a blocked design. This ensured that the auditory stimulus in the vertical RDM task was task-irrelevant on *every single trial*. However, participants were extensively trained in the conflict-inducing content discrimination task I, and as such participants might have been prompted to detect the auditory stimuli in the second session of the vertical RDM task, due to their earlier relevance. This in turn might have confounded our decoding results, for which we merged data from both EEG sessions, before and after training in content discrimination task I. That is, the above chance decoding of stimulus content in the vertical RDM task might mainly be caused by some sort of lingering *background* process aiming to detect the previously task-relevant stimulus content of the auditory stimulus. However, stimulus content could already be decoded from data of the first session of the vertical RDM task, when participants were not yet exposed to content discrimination task I (see **supplementary figure S3**). Similarly, stimulus location could also be classified above-chance before the auditory stimuli gained any task-relevance (see **supplementary figure S3**). Thus, the significant decoding of stimulus content and location in the vertical RDM task indeed reflects perceptual processing of truly irrelevant stimulus features.

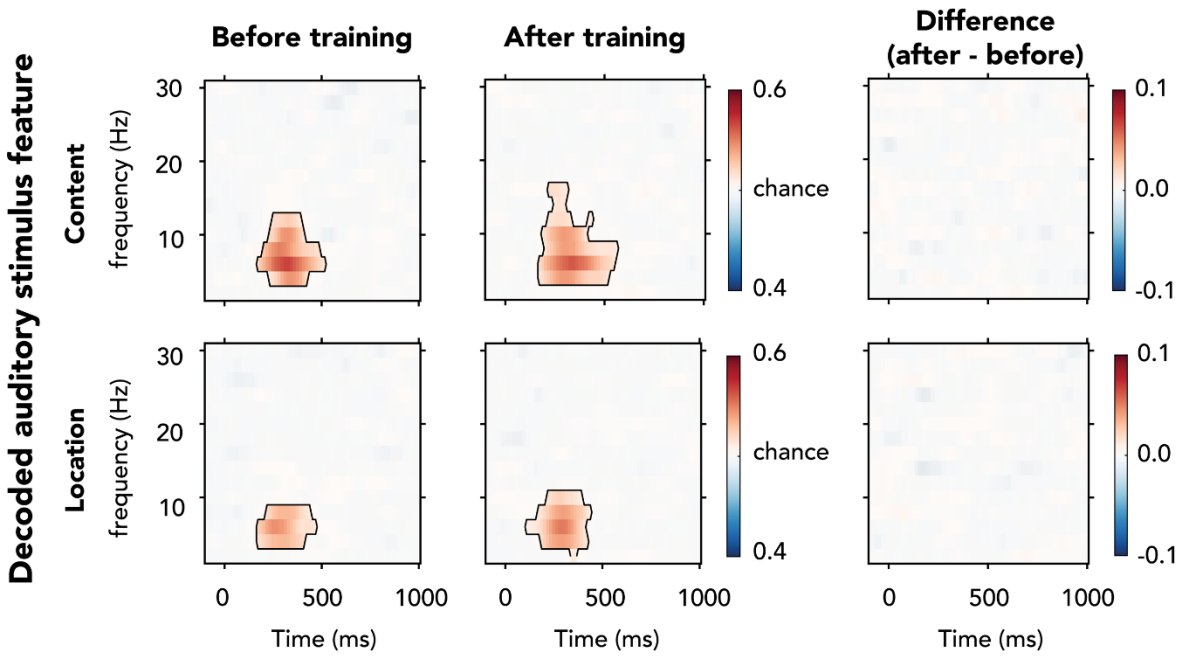


Figure S3. Classifier accuracies for auditory stimulus content and location in the vertical RDM before and training in the conflicting task. Thresholded (cluster-base corrected, p<0.05) accuracies are depicted across the frequency-range (2-30Hz) and significant clusters are outlined with a solid black line. Classifier accuracies of stimulus content (top row) and stimulus location (bottom row) in the vertical RDM before training (left column) and after training (middle column) in the conflicting task. Difference (after training-before training) is shown in the right column. During the first session of the vertical RDM, participants had not yet performed a single trial of the conflicting task. Therefore, the above-chance classification of stimulus content and location shown here is not confounded by any effects related to the conflicting task and thus shows perceptual processing of truly task-irrelevant stimulus features.

Decoding accuracies of stimulus features in theta-band for different levels of task-relevance

We extracted individual classifier accuracies from a theta-band ROI (100ms-700ms, 2Hz-8Hz) for all behavioral tasks and stimulus features. We tested these individual accuracies with (Bayesian) one-sample t-tests against chance ($\bar{x}$>0.5). All results are shown in supplementary **table 1**. With decreasing levels of task-relevance of sensory input, effects of congruency on theta-band power diminish while the effects of stimulus content and location are unaffected (see **figure 5**).

One Sample T-Test ($\bar{x}$>0.5)

	t	df	p	Cohen's d	BF_{10}
Experiment 1					
Content discrimination I - congruency	6.08	23	< .001	1.24	$1.17e^4$
Content discrimination I - content	8.14	23	< .001	1.66	$8.70e^5$
Content discrimination I - location	7.58	23	< .001	1.55	$2.83e^5$
Vertical RDM - congruency	-0.24	23	0.59	-0.05	5.52*
Vertical RDM - content	6.95	23	< .001	1.42	7.66e4
Vertical RDM - location	4.96	23	< .001	1.01	989.52
Experiment 2					
Content discrimination I - congruency	6.67	23	< .001	1.36	$4.19e^4$
Content discrimination II - content	7.43	23	< .001	1.52	$2.07e^5$
Content discrimination I - location	5.69	23	< .001	1.16	$5.00e^3$
Location discrimination - congruency	3.63	23	< .001	0.74	52.36
Location discrimination - content	6.06	23	< .001	1.24	$1.12e^4$
Location discrimination - location	7.95	23	< .001	1.62	$6.03e^5$
Volume oddball detection - congruency	2.29	23	< .05	0.47	3.66

One Sample T-Test ($\bar{x}$>0.5)

	t	df	p	Cohen's d	BF_{10}
Volume oddball detection - content	7.89	23	< .001	1.61	$5.31e^{5}$
Volume oddball detection - location	5.50	23	< .001	1.12	$3.27e^{3}$
Horizontal RDM - congruency	0.00	23	0.499	$2.80e^{-4}$	4.65*
Horizontal RDM - content	2.67	23	< 0.001	0.55	7.42
Horizontal RDM - location	4.04	23	< .001	0.82	126.17

Table 1. One-sample t-test results for decoding accuracies of every task and stimulus feature. Note: *BF_{01}

Comparison of decoding accuracies for different time-frequency ROIs

The reported results from our ROI-analyses could have been accidental, due to the fact that the ROI was chosen on the basis of previous results and visual inspection of the decoding results. In order to exclude this possibility, we performed the exact same analyses on data that were extracted from different ROIs. In total we used three different ROIs.

ROI 2: 200ms-700ms, 2Hz-8Hz (**figure S4A**); ROI 3: 200ms-700ms, 4Hz-8Hz (**figure S4B**); ROI 4: 200-500ms, 2Hz-8Hz (**figure S4C**).

<u>ROI 2: 200ms-700ms, 2Hz-8Hz</u>

We found main effects for behavioral task and stimulus feature (task: $F_{1,2.23}$=21.96, p<0.001, η_p^2=0.49; stimulus feature: $F_{1,2}$=57.70, p<0.001, η_p^2=0.72). The interaction between task and stimulus was also significant ($F_{1,3.04}$=28.03, p<0.001, η_p^2=0.55)

<u>ROI 3: 200ms-700ms, 4Hz-8Hz</u>

We found main effects for behavioral task and stimulus feature (task: $F_{1,2.21}$=15.39, p<0.001, η_p^2=0.040; stimulus feature: $F_{1,2}$=49.56, p<0.001, η_p^2=0.68). The interaction between task and stimulus was also significant ($F_{1,3.45}$=18.77, p<0.001, η_p^2=0.45)

<u>ROI 4: 200-500ms, 2Hz-8Hz</u>

We found main effects for behavioral task and stimulus feature (task: $F_{1,2.32}$=12.76, p<0.001, η_p^2=0.36; stimulus feature: $F_{1,2}$=39.63, p<0.001, η_p^2=0.63). The interaction between task and stimulus was also significant ($F_{1,2.79}$=18.03, p<0.001, η_p^2=0.44)

In **figure S4**, behavioral tasks on the x-axis are again sorted by magnitude of the effect size of congruency decoding. Note that in all figures, the HRDM task is the one with the lowest effects size in congruency. However, for all ROIs the same pattern as in **figure 5** is visible, namely that congruency decoding deteriorates to a point where accuracies are no longer significant, whereas for decoding of sensory stimulus features this is not the case.

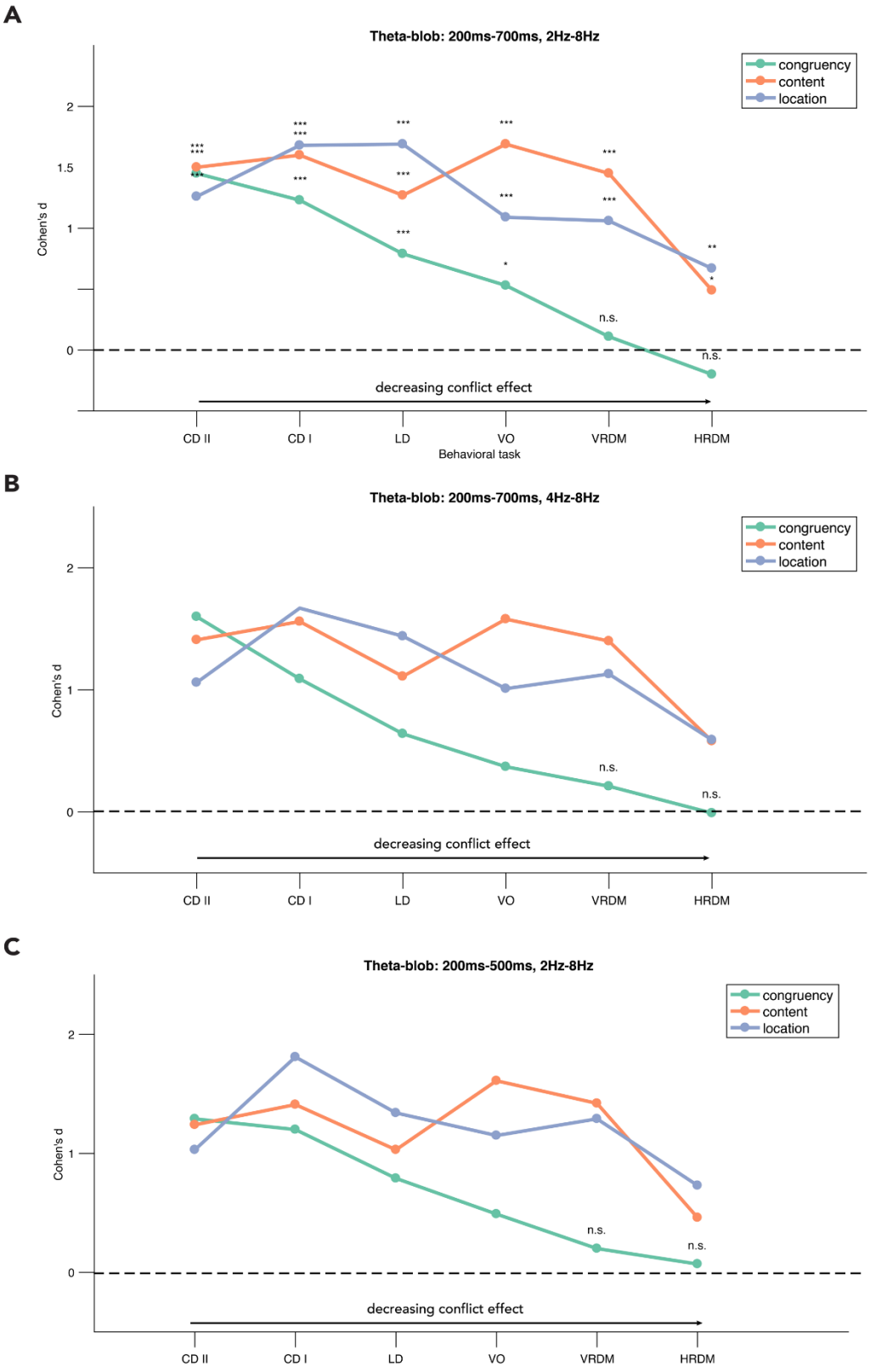


Figure S4. Effect sizes for ROI decoding accuracies for every task and feature. (A) ROI 2: 200ms-700ms, 2Hz-8Hz. **(B)** ROI 3: 200ms-700ms, 4Hz-8Hz, **(C)** ROI 4: 200ms-500ms, 2Hz-8Hz. The data is sorted (from left to right) in accordance with highest effect size on congruency decoding.

More extensive processing of conflicting features for volume oddball task as compared to horizontal RDM

When we compare the accuracies of sound content and location decoding between the volume oddball task and the horizontal RDM, we observe that processing of these features is more extensive in the volume oddball task. Specifically, accuracies of sound content decoding are higher for the volume oddball task in a delta-band cluster ($p=0.02$, two-sided: $\bar{x}>0.5$, cluster-corrected; frequency-range: 2Hz-4Hz, peak frequency: 2Hz, time-range: 250ms-563ms, peak time: 438ms; see **supplementary figure S5**). Decoding accuracies for sound location were significantly higher for the volume oddball task in an alpha/beta-band cluster ($p<0.001$, two-sided: $\bar{x}>0.5$, cluster-corrected; frequency-range: 8Hz-18Hz, peak frequency: 10Hz, time-range: 250ms-719ms, peak time: 516ms; see **supplementary figure S5**). These higher accuracies for the decoding of task-irrelevant features may highlight that object-based attention led to the selection and neural enhancement of the task-irrelevant features only in the volume oddball task.

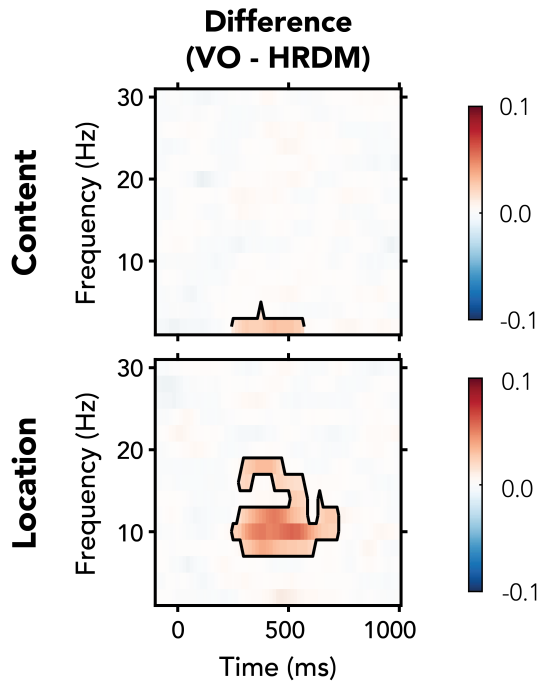


Figure S5. Difference in classifier accuracies for auditory stimulus content and location between the volume oddball task and the horizontal RDM task. Thresholded (cluster-base corrected, $p<0.05$) accuracy differences are depicted across the frequency-range (2-30Hz) and significant clusters are outlined with a solid black line. Decoding accuracies for sound content were significantly higher in a delta-band cluster for the volume oddball task versus the horizontal RDM task. Furthermore, decoding accuracies for sound location were significantly higher in an alpha/beta-band cluster for the volume oddball task versus the horizontal RDM task. This suggests that sound content and location were processed to a greater extend in the volume oddball task as compared to the horizontal RDM, possibly because these features were selected by object-based attention in the volume oddball task only.

Descriptive statistics of behavioral results of experiment 2

Reaction times in the location discrimination task were on average 338.2ms (s=112.7ms) and the overall error rate was 4.5% (s=3.2%). For content discrimination task II, reaction times were 364.0ms (s=127.4ms) and error rates were 5.5% (s=5.8%). For the horizontal RDM, reaction times were 362.5ms (s=116.5ms) and error rates were 27.7% (s=5.2%).

Behavioral conflict in the horizontal RDM task

We report no effect of auditory conflict on performance in the horizontal RDM task even though the conflicting auditory features (location and content) were overlapping in feature space with the relevant stimulus of the task (horizontally moving dots). Perhaps, however, the auditory features did individually interfere with task

performance. To test this we performed (Bayesian) one-sample t-tests (against 0) on the conflict effect for every conflict type (auditory location-dot direction, auditory content-dot direction, auditory location-auditory content). We found effects of auditory location-dot direction conflict on RTs and ERs (RT: t_{23}=2.12, p=0.045, d=0.43; ER: t_{23}=5.32, p<0.001, d=1.09, see **supplementary figure S6**). In addition we found effects of auditory content-dot direction on ERs but not RTs (RT: t_{23}=1.30,p=0.21,d=0.27, BF_{01}=2.20; ER: t_{23}=5.40, p<0.001, d=1.10, see **supplementary figure S6**). Lastly, for the pure auditory conflict between location and content we found no conflict effects on RTs and ERs (RT: t_{23}=1.72,p=0.10,d=0.35, BF_{01}=1.31; ER: t_{23}=1.65, p=0.11, d=0.34, BF_{01}=1.44).

In addition, we performed paired sample t-tests to directly test the difference in conflict effects between various sources of conflict. In RTs we did not find significant differences between auditory location-dot direction conflict and auditory content-dot direction conflict (t_{23}=1.11, p=0.28, d=0.23, BF_{01}=2.69), auditory location-dot direction conflict and auditory location-auditory content conflict (t_{23}=0.46, p=0.65, d=0.09, BF_{01}=4.23) and auditory content-dot direction conflict and auditory location-auditory content conflict (t_{23}=-1.05, p=0.31, d=-0.21, BF_{01}=2.85). For ERs, we found that the conflict effects for auditory location-dot direction an auditory content-dot direction conflict sources were higher than for auditory content-location conflict ($ER_{loc\text{-}dir}$: t_{23}3.08, p<0.01, d=0.63; $ER_{cont\text{-}dir}$: t_{23}=2.86, p<0.01, d=0.58). The conflict in ER did however not differ between the two sources of auditory-visual conflict (t_{23}=-0.87, p=0.39, d=-0.18, BF_{01}=3.30).

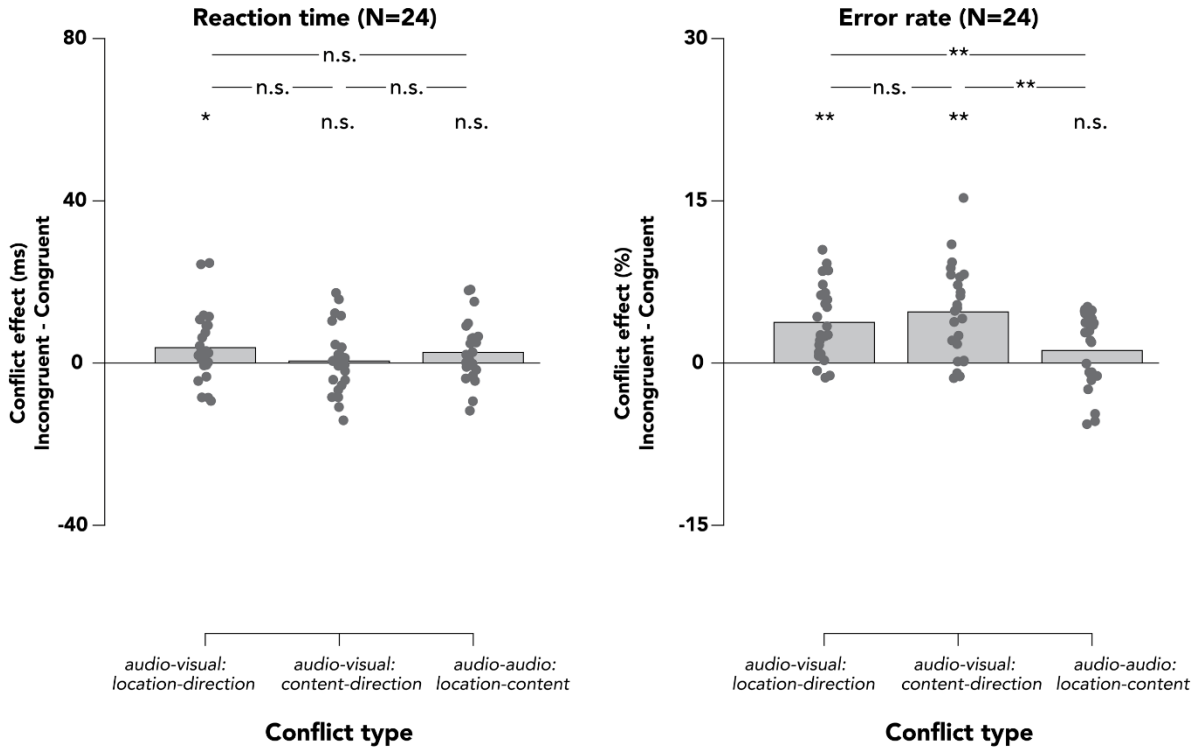


Figure S6. Effect of different sources of conflict on reaction times and error rates in horizontal RDM task. Reaction times were increased for auditory stimuli of which the content (first letter, L or R denoted in quotation marks) conflicted with responses made exclusively with the right hand. So when the word 'left' was presented, reaction times were longer than when the word 'right' was presented. Location of the auditory stimulus (first letter, L or R) did not have an effect on reaction times. The interaction between content and location (i.e. congruency) was not significant either.

All types of auditory interference in experiment 1

Although we report no effects of task-irrelevant auditory conflict in the vertical RDM, some dimensions of the auditory stimuli could still have interfered with performance on the RDM-task. For example, the spoken word "left" or an auditory stimulus presented from the left side of the participant could lead to conflict with right hand responses in the vertical RDM task. To test such potential interference, we performed a 2x2x2 factorial rm-ANOVA on mean RTs of vertical RDM task, with session (before/after training), stimulus content ("left"/"right"), and stimulus location (left/right) as factors. RTs in the vertical RDM task were unaffected both

by stimulus content ($F_{1,23}$=0.29, p=0.59, η_p^2=0.01, BF_{01}=6.42) and location ($F_{1,23}$=0.41, p=0.53, η_p^2=0.02, BF_{01}=6.33) (see **supplementary figure S7**, data of content discrimination task I plotted as control).

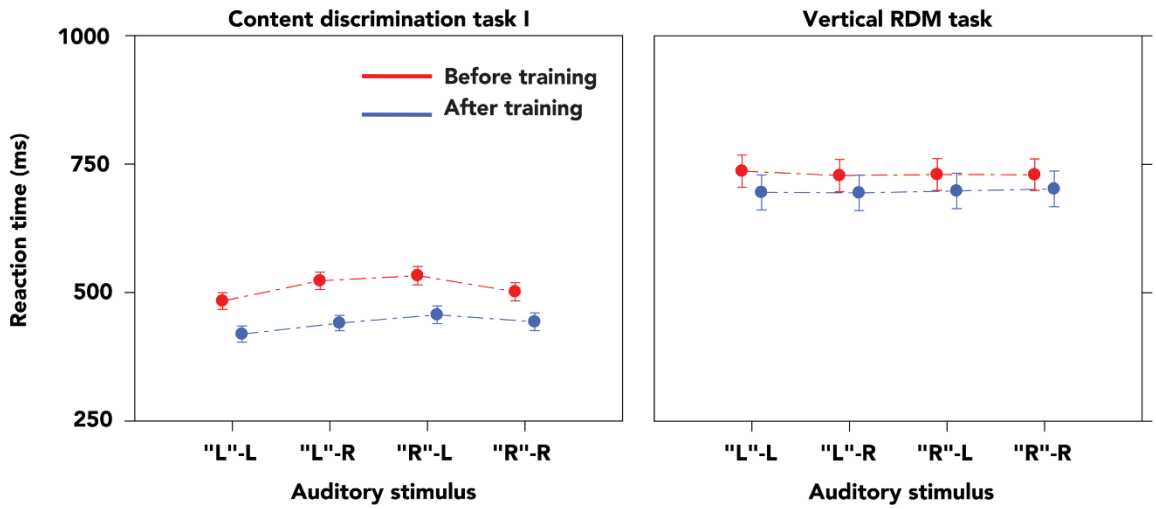


Figure S7. Effect of auditory stimulus on reaction times for both content discrimination task I and vertical RDM. In the content discrimination task, when the auditory stimulus was task-relevant, reaction times were affected by the specific auditory stimulus that was presented (left pane). Specifically, reaction times were increased for stimuli of which stimulus content (first letter, L or R denoted in quotation marks) and stimulus location (second letter, L or R) were incongruent. Contrary, reaction times in the vertical RDM task were unaffected by the specific auditory stimulus, excluding any possible conflict between stimulus content or location and responses made with exclusively the right hand (right pane).

Specific stimulus interference on volume oddball detection

Perhaps, specific auditory stimuli resulted in stronger or weaker interference on the oddball detection task. Specifically, participants were required to respond with their right hand (only for oddballs, else they should not respond) and the content and location of the auditory stimulus could have interfered with this response. We performed a 2x2 factorial rm-ANOVA on RTs of hit trials, with sound content and location as factors. We found that sound content significantly affected RTs, whereas sound location did not (content: $F_{1,23}$=26.16, p<0.001, η_p^2=0.53; location: $F_{1,23}$=1.15, p=0.29, η_p^2=0.05, BF_{01}=2.33, see **supplementary figure S8**). Specifically, participants responded faster on trials where the spoken word was 'right' as opposed to 'left'. There was no significant interaction effect, or congruency effect, on RTs ($F_{1,23}$=024, p=0.63, η_p^2=0.01, BF_{01}=2.74). We would like to point out that the interference effect of sound content on task performance (in RT) could not have contributed to our observations of above-chance decoding of congruency in the oddball task (of extracted cluster values) as our decoding analyses were performed on correct rejections and not on hit trials.

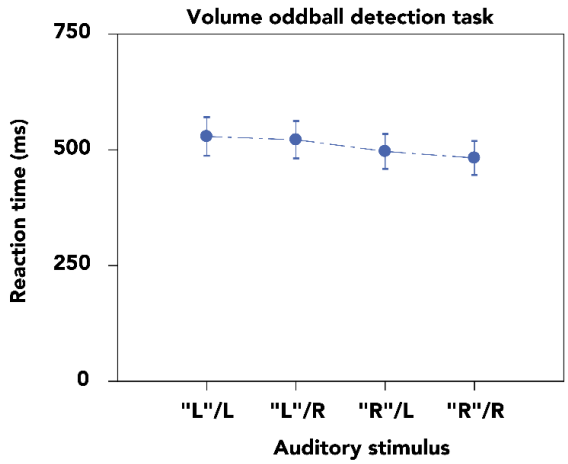


Figure S8. Effect of auditory stimulus on reaction times of hit trials the volume oddball task. Reaction times were increased for auditory stimuli of which the content (first letter, L or R denoted in quotation marks) conflicted with responses made exclusively with the right hand. So when the word 'left' was presented, reaction times were longer than when the word 'right' was presented. Location of the auditory

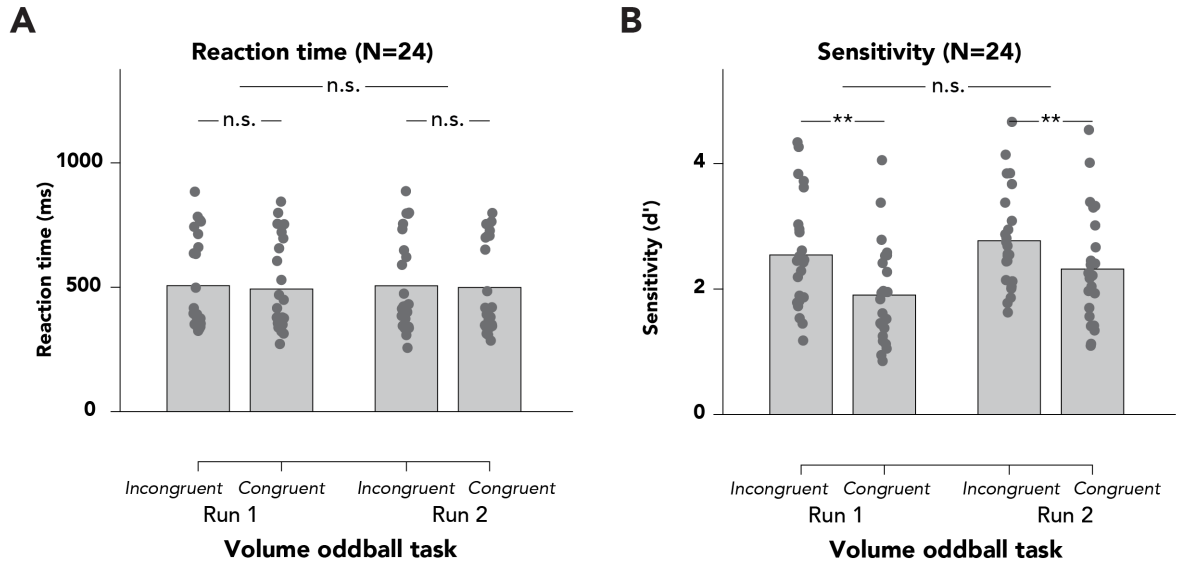
stimulus (first letter, L or R) did not have an effect on reaction times. The interaction between content and location (i.e. congruency) was not significant either.

Effects of exposure to conflict on behavior and decoding accuracies in volume oddball task

We performed a 2x2 rm-ANOVA on RTs and d' of the volume oddball task with factors been run number and congruency of the auditory stimulus. We found no effects of run number ($F_{1,23}$=0.02, p=0.90, η_p^2=0.00, BF_{01}=6.46) and congruency ($F_{1,23}$=2.78, p=0.11, η_p^2=0.11, BF_{01}=3.15) on RT (see **supplementary figure S9A**). There was also no interaction effect between run number and congruency in terms of RT ($F_{1,23}$=0.09, p=0.77, η_p^2=0.00, BF_{01}=16.82). We did find effects of run number ($F_{1,23}$=15.32, p<0.001, η_p^2=0.40) and congruency ($F_{1,23}$=45.62, p<0.001, η_p^2=0.67) on d', but the interaction between them was not significant ($F_{1,23}$=1.40, p=0,25, η_p^2=0.06, BF_{01}=0.52, see **supplementary figure S9B**). Specifically, d' was higher in the second run as compared to the first run and higher for incongruent than congruent trials. This conflict effect, however, was not modulated by run number.

Decoding analyses of the volume oddball detection task were performed on data that were merged for the both runs of the task, i.e. one run before and one run after exposure to conflict-inducing tasks. To investigate the possible effect of exposure to the conflict tasks, and thus the mapping of responses to auditory content and location, on decoding performance in the volume oddball task we performed the decoding analyses separately for the two runs (see **supplementary figure S10**). We then calculated the difference between the two runs. We found no significant clusters for auditory congruency, auditory location and auditory congruency (p>0.05, cluster-corrected). In addition we used the hypothesis-driven time-frequency ROI (2Hz-8Hz, 100ms-700ms) to extract individual congruency decoding classifier accuracies of the first and second run. Previously we reported above chance decoding (over merged session data) by using this procedure.

For the first run of the volume oddball task, there was no effect of conflict on power in the time-frequency ROI (t_{23}=-0.034, p=0.63, d=-0.07, BF_{01}=5.89). For the second run, we did observe above-chance congruency decoding in the ROI (t_{23}=1.90, p=0.04, d=0.39, BF_{10}=1.93). However, a (Bayesian) paired sample t-test on the difference between runs (session 2 – session 1) shows decoding is not different between runs (t_{23}=1.58, p=0.13, d=0.32, BF_{01}=1.56).



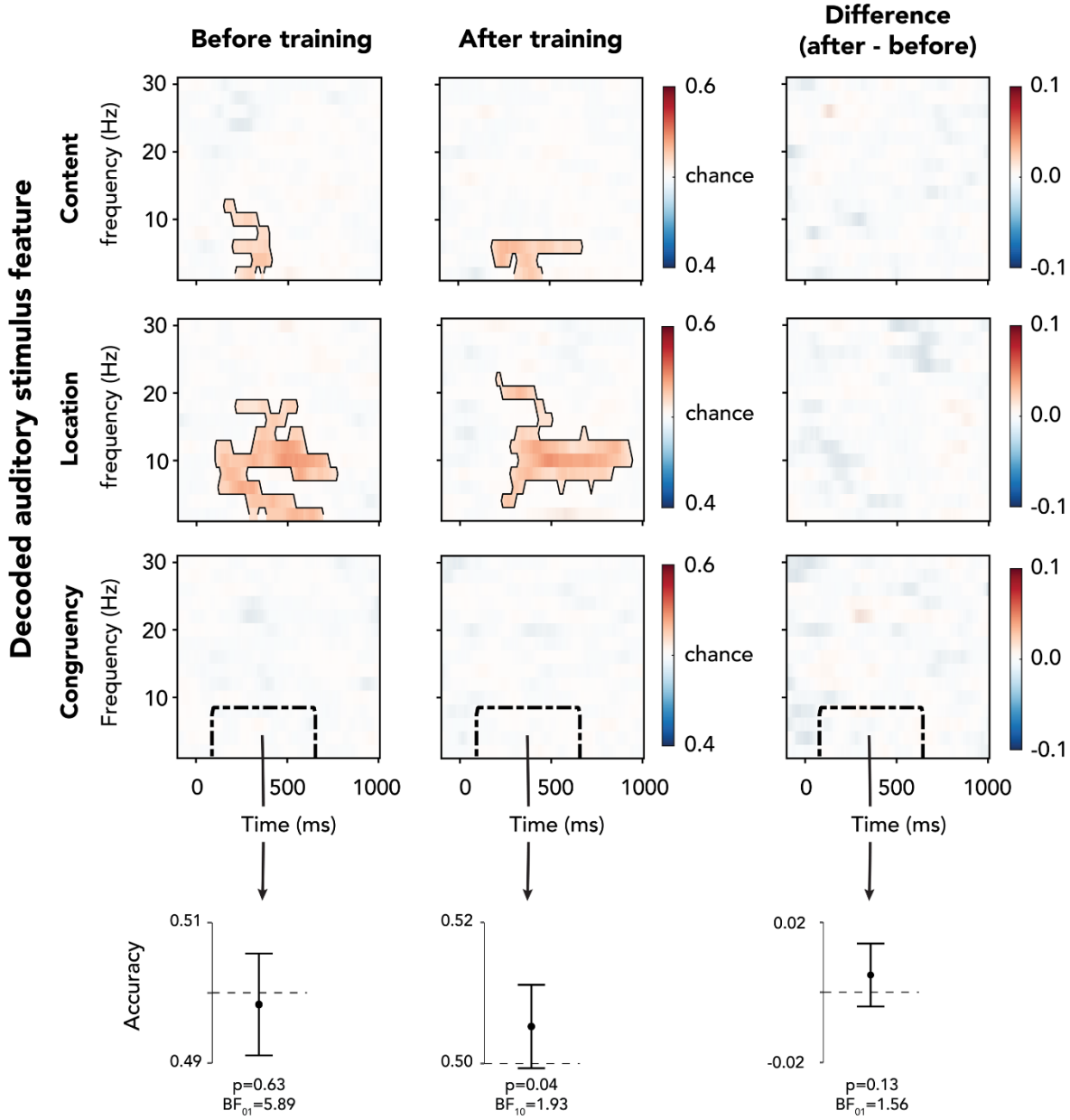


Figure S10. Classifier accuracies for auditory stimulus content, location and congruency in the volume oddball task before and after exposure to conflict inducing tasks. Thresholded (cluster-base corrected, p<0.05) accuracies are depicted across the frequency-range (2-30Hz) and significant clusters are outlined with a solid black line. Classifier accuracies of stimulus content (top row), stimulus location (middle row) and stimulus congruency (bottom row) in the volume oddball task before (left panel) and after (right panel) exposure to conflict inducing tasks. Inset plots in the bottom of the figure show congruency decoding accuracy scores that were extracted from the clusters of content discrimination task II and location discrimination task. During the first run of the volume oddball task, participants had not yet performed a single trial of a conflicting task. Therefore, the above-chance classification of stimulus content and location shown here is not confounded by any effects related to the conflicting task and thus shows perceptual processing of truly task-irrelevant stimulus features.