INVITATION
On Monday June 29, 2015, I will publicly defend my doctoral dissertation in the aula of the Raboud University, Comeniuslaan 2, Nijmegen. The defense begins at 14:30 sharp. You are most welcome for this special occasion, and for the subsequent reception.

Ana Todorović
todorovic.ana@gmail.com


Predictive adaptation in early auditory processing

Ana Todorović
Predictive adaptation
in early auditory processing

Ana Todorović
The work described in this thesis was carried out at the Donders Institute for Brain Cognition and Behaviour, Radboud University Nijmegen, with financial support from the Netherlands Organization for Scientific Research (NWO VENI-grant 451-09-001) awarded to Dr. Floris de Lange.

ISBN
978-94-6284-021-8

Cover page
Old radio tuning ©iStockphoto.com/CurvaBezier

Lay-out
Promotie In Zicht, Arnhem

Print
Ipskamp Drukkers, Enschede

Copyright © Ana Todorović, 2015
Predictive adaptation
in early auditory processing

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op maandag 29 juni 2015
om 14.30 uur precies

door

Ana Todorović
geboren op 9 februari 1979
te Belgrade (Servië)
Promotor
Prof. dr. P. Hagoort

Copromotor
Dr. F.P. de Lange

Manuscriptcommissie
Prof. dr. O. Jensen
Prof. dr. E. Schröger (Universität Leipzig, Duitsland)
Prof. dr. I. Winkler (Research Centre for Natural Sciences, Hungarian Academy of Sciences, Hongarije)
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>Expectation and attention share a mechanism of forming sensory templates</td>
<td>69</td>
</tr>
<tr>
<td>6</td>
<td>Discussion</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Addendum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nederlandse samenvatting</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>Acknowledgments</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>List of publications</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>About the author</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>Donders Series</td>
<td>131</td>
</tr>
</tbody>
</table>
Introduction
Imagine yourself at the ocean on a clear day. You throw a pebble into the calm water. It splashes, then creates ripples. It is captivating, and you instantly want to know more: how big is the initial splash? How far do the ripples reach? How long before the water is calm again? What does this tell you about the properties of the water? This is a difficult question: the ocean is deep and full of life, and you only have a view of its surface. You first experiment by throwing pebbles of different shapes and sizes. But then you notice: even if you throw identical pebbles, the effect on the water will vary. You might throw two in quick succession, and remark that the second one, falling into the place where the first one landed a moment ago, will not create such strong ripples. Or you might throw two side by side, and watch their ripples cancel each other out. As a wave arrives, your pebble hits the water at an angle, sometimes dampening the ripples, sometimes spreading them further. Or the tide might come in, and as the ocean moves to meet it, your pebble has less of a way to fall. Suddenly, it starts raining, thousands of droplets creating tiny ripples of their own. Just as you’re about to go, you chuck a final pebble sideways, to watch it skip on the water. It hits the surface four times and sinks. Moments later, a faint, fifth set of ripples emerges just where it would have landed next. What just happened?

Sensory neuroscientists also cast pebbles of a sort. We isolate people from the constant bombardment of sensory stimulation by putting them in a quiet lab. Then we send a stimulus their way, and we observe how it perturbs their brain activity. Most often we measure from a distance, taking note of the simultaneous activity of hundreds of thousands of neurons working together. If we play a brief tone, we know it will excite the auditory cortex, located bilaterally in the temporal lobes. The frequency of the tone will determine where precisely in the auditory cortex the largest perturbation will be, because certain neurons prefer certain frequencies. We know that louder tones, like larger pebbles, create more activity. However, unlike the ocean which passively receives anything we throw at it, the brain is in active interaction with the environment. The perception of a stable world is not fuelled by stable neural activity. Any stimulus, be it a sound, an image, a touch, can elicit a weaker or stronger response depending on various factors: whether it is attended or ignored, whether it is repeated or novel, whether it is predictable or unpredictable. In other words, the extent of the perturbation that a stimulus will bring depends on the state of the brain as it arrives. The state of the brain, conversely, depends on the history of its stimulation. It prepares for events by selectively changing its sensitivity to input that is in line with its predictions. Like the fictional fifth set of ripples, the sensory brain rises to stimuli as they arrive, meeting them with a faint reflection of the activity it predicts they will elicit.
The aim of my thesis is to discuss how previous stimulation and voluntary attention change the neural response to auditory stimuli. In all my experiments, I use sequences of tones to create statistical regularities in stimulation. People can use these regularities to predict which sound is coming next. This is a common feat of cognition and doesn’t seem to require much effort. When we listen to a melody, we have a feeling for which sound should come next, and when we listen to speech, we can often make a reasonable guess how a sentence might end. The sequences that are described in this thesis are much simpler: they consist of two tones linked by a conditional rule. To the person listening to the tones, some aspect of the second tone becomes predictable once they hear the first tone. The second tone can then conform to this prediction or violate it. In addition to prediction, selective attention also changes sensitivity to incoming stimulation, but (in the sense it is used here), it is more closely linked to volition. In order to test how prediction and attention come together to guide perception, I give tasks to draw attention to or away from the (predictable and unpredictable) tones I am analyzing. Then I drop the pebble – I play the second tone – and observe how the auditory cortex responds to it. From this, I draw inferences about how the brain adapts to the environment by preparing for incoming stimulation.

Neuronal suppression

A central topic of this thesis is neuronal suppression: why is the response of the sensory system to a stimulus sometimes reduced? A robust way to get the auditory cortex to decrease its activity is to repeatedly expose it to the same tone. This phenomenon is called repetition suppression (or, in some of the literature, stimulus-specific adaptation), and it has been observed across a wide range of species (Wark et al., 2007), in cortical (Rosburg et al., 2004) and subcortical areas (Ulanovsky et al., 2003a; Dean et al., 2005; Ayala and Malmierca, 2012), in the presence of awareness (Yee and White, 2001; Rosburg et al., 2004) or in its absence (Dehaene et al., 2001). Regardless of its ubiquitousness, the neural mechanisms underlying repetition suppression remain unresolved (Grill-Spector et al., 2006). One possibility is that neurons become fatigued, and cannot sustain the same firing rate over multiple tone presentations. Another possibility is that the stimulus representation sharpens over time, so that only stimulus-specific neurons become excited by the repeated presentations while other neurons return to baseline activity. Both of these accounts involve stimulus-driven mechanisms, where local neuronal connections become adjusted due to the fact that neurons just responded to a stimulus with certain features. However, repetition suppression studies share another important element: in them, tones get repeated many times over the course of an experiment. Therefore,
a crucial difference between the first instance when a tone was played and its second instance, is that the appearance of the second tone becomes highly predictable over time. A number of studies show that predictable (or expected) stimuli involve a reduced neural response. For example, if a tone precedes one of two gratings, in such a way that the a low tone is mostly coupled to a clockwise-oriented grating while a high tone mostly precedes the counter-clockwise one, then the level of visual activity will be smaller if the orientation of the grating conforms to the learned association than if it violates it (Kok et al., 2012a). If an object flashes on the screen at a predictable location or moment, it excites the visual cortex less than if it flashes slightly out of sync with the prediction (Alink et al., 2010). If the first letter in a word involves positioning the lips in a visually salient way, the remainder of the word can be predicted more accurately and is consequently processed using less auditory activity (Arnal et al., 2011). If people repeatedly press a button to play a tone, it leads to less auditory activity than if the same sequence is played back to them when they cannot predict the onset times with precision (Aliu et al., 2009). Might stimulus predictability play a role in repetition suppression? This is the question I tackle in Chapter 2.

The content and timing of predictive responses

Predictive processing is often explored using models of predictive coding (Rao and Ballard, 1999; Friston et al., 2005; Den Oud et al., 2012). The common interpretation of these models is that the brain, using feedback connections from higher order regions, prepares and holds in place a template of expected stimulation. Once the actual stimulus arrives, a comparison process ensues. If the input does not match the template, an error response is generated, scaled to the amount of the mismatch. The template needs to be adjusted based on the new input, and the error response, together with the adjustment process, cumulatively lead to an increase in net neural activity. If the input matches the template, the converse happens: there is a decrease in net neural activity (relative to a situation with an unspecific expectation). The rationale is that the sensory system does not need to invest processing resources into a stimulus that is already represented by the system (Clark, 2013). The process of adjusting templates through comparing with actual input explains why a tone of an expected pitch elicits less activity than a tone of an unexpected pitch. It does not, however, explain why the same stimulus leads to less activity when it arrives at an expected moment, than when its moment of onset is unexpected (Aliu et al., 2009; Alink et al., 2010). In fact, in Chapter 2 I will show that a tone does not even need to be played to elicit a predictive response: unexpected tone omissions lead to more auditory activity than expected ones. The omission response is strongest precisely at the moment when the tone would have elicited the most activity had it been played.
This suggests that the template of the predicted stimulation exists not as a stationary pattern of activation, but as a possibility of activation - perhaps present as a change in local synaptic weights, which becomes fully activated at a particular moment. Based on these observations, I explored the role of predictive timing in repetition suppression. Will a repeated tone undergo less suppression if its onset time becomes less certain? This is what I set out to test in Chapter 3.

**Mismatch negativity**

Over the past few decades, the role of regularities in auditory processing was championed by mismatch negativity studies (Näätänen et al., 2007; Winkler et al., 2009). The initial setup of these studies involved a sequence where tones are played at regular intervals. The tones would have one of two frequencies: one of them, called the standard, would be presented often, while the other, the deviant, would be rare. A comparison of their evoked potentials reveals that the deviant tone leads to a greater negative deflection at some point between 100 and 300 milliseconds after its onset. The size of the difference between the standard and the deviant depends both on their distance in pitch and their relative likelihoods (Näätänen, 1990), with more distant and rarer deviants leading to greater mismatch negativity. This phenomenon is explained by a memory trace that forms with exposure to the stimuli, where the automatically detected difference between the incoming stimulus and the memory trace is reflected in a stronger neural response.

A long standing debate in the literature is to what extent repetition suppression plays a role in mismatch negativity (Näätänen et al., 2005). Namely, when a frequent standard is compared to a rare deviant, the difference is not only in how predictable they are, but also in how many times each of them was previously processed by the sensory system. Therefore, mismatch negativity might partly reflect stimulus-driven fatigue, which attenuates neural activity to the more frequent standard tones. Mismatch negativity is however known to be present when complex rules are violated. For example, it arises from the mismatch of a learned combination of stimulus features (Gomes et al., 1997), conditional probabilities (Saarinen et al., 1992), global sequences (Horvath et al., 2001), and even grammatical rules (Pulvermüller and Shtyrov, 2003). Although these studies conclusively show the presence of genuine regularity detection, they do not address the question of whether repetition suppression plays into the mismatch negativity effect in simple sequences where standards and deviants differ in base rate. The outcome of experiments investigating this problem is that there is some partial truth to it: both repetition suppression and rule confirmation attenuate neural activity, in slightly overlapping time intervals.
(Näätänen et al., 2005; Nelken and Ulanovsky, 2007). To sidestep this problem, mismatch negativity researchers devised clever control conditions. For example, the rare deviant tone might be compared to itself embedded in a different sequence, where it is just as rare, but presented with many other tones of equal likelihood (Schröger and Wolff, 1996). While this type of comparison does away with repetition suppression when comparing tones of different likelihoods, it does not answer the bigger question of whether there is an interplay in suppression related to stimulus repetition and suppression related to stimulus predictability, when two tones differ in their base rate. Do repetition suppression and expectation suppression enhance each other? Do they influence the same components during tone processing? Do they rely on same or different mechanisms? These are the questions I address in Chapter 4.

**Prediction meets attention**

Prediction and attention both guide sensory processing. When the appearance of a stimulus is predictable, people are able to react to it more quickly (Nattkemper and Prinz, 1997). The same is true when a stimulus is attended (Posner et al., 1980). Outside the laboratory, where we move our attention through a world that is partly stationary and partly volatile, it is reasonable to assume that prediction and attention are intertwined, each one supporting the function of the other. Both are believed to selectively change neural sensitivity to incoming stimulation. This would suggest that they should involve similar effects on neural activity in early sensory regions. Contrary to this intuition, when attention is held constant, prediction leads to an attenuated sensory response (Summerfield et al., 2008), but when prediction is held constant, attention boosts sensory activity (Maunsell and Treue, 2006). Teasing apart prediction from attention, and then describing their unique and joint contributions, has been recognized as highly relevant but also highly intricate (Summerfield and Egner, 2009).

A number of researchers have set out to explore this, but the answer is proving to be more complex than previously assumed. Predictive effects have been found to reverse when attention is present (Kok et al., 2012b), to exist only in the presence of attention (Larsson and Smith, 2012; Stokes et al., 2014), or to exist only in its absence (Jones et al., 2013). One potential source of this confusion is that both prediction and attention tend to be conceptualized fairly generally (e.g. an entire stimulus display might be considered predictable or attended), when, in fact, both prediction and attention can refer to various aspects of the sensorium. For example, if the stimulus location is attended while its features are predicted, the comparison of prediction and attention might entail different neural mechanisms than if the location is predicted but the features attended. Additionally, attention is often conceptualized as task relevance, which involves it being linked to an action (or withholding an action) in experimental
paradigms, while in manipulations of prediction such a link is frequently absent. In this thesis I attempted to tackle the prediction-attention relationship with a design where both prediction and attention relate to the same aspect of stimulus processing (tone frequency), where the efforts of both prediction and attention can be confirmed or violated, and where all stimuli require a response, regardless of their status as predicted/unpredicted or attended/unattended. If both prediction and attention lead to forming a stimulus-specific template - prediction by using the learned associations between tones, attention by volitional preparation for a given tone - how will the auditory brain respond when prediction meets attention? Chapter 5 aims to answer that question.

**Top-down and bottom up, side by side**

A rapidly growing body of studies suggests that the division of sensory processing into an early perceptual and later cognitive stage is obsolete. As we think about the world using familiar concepts to guide us in our interpretations, so does the sensory system use familiar patterns to initially compare input against. While there is a short time window after stimulus onset when it is possible to claim that only a feedforward sweep exists without recurrent processing, processing related to past stimulation affects how the sensory system receives new input. The history of experienced events imprints itself on the sensory system, reshaping its response profile to new input. In such a scenario, the functional distinction between bottom-up and top-down processing becomes blurred.

In Chapter 6, I offer some concluding remarks about predictive processing. Overall, I argue both for the ubiquitousness and for the complexity of selective neural responsivity to regularities in the environment. I focus on the phenomenon of neural adaptation, and I argue that current research supports the idea that sensory areas encode the statistics of the stimulus environment. In doing this, I review different literatures that show predictive adaptation, i.e. the phenomenon that the neural response to repeated stimulation is more strongly suppressed if the stimulus repetition is predictable. I also discuss the neural bases of forming a neural template of predictable stimulation, and I touch upon the issue of hierarchical predictive processing (which refers to encoding probabilities of varying complexity), and discuss whether a common neural mechanism should be expected based on a successful common computational mechanism.
Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study

This chapter is based on:
The Journal of Neuroscience 31, 9118-23.
Abstract

Repetition suppression, the phenomenon that the second presentation of a stimulus attenuates neural activity, is typically viewed as an automatic consequence of repeated stimulus presentation. However, a recent neuro-imaging study has suggested that repetition suppression may be driven by top-down expectations. Here we examined whether and when repetition suppression can be modulated by top-down expectation. Participants listened to auditory stimuli in blocks where tone repetitions were either expected or unexpected, while we recorded ongoing neural activity using magnetoencephalography. We found robust repetition suppression in the auditory cortex for repeated tones. Interestingly, this reduction was significantly larger for expected than unexpected repetitions, both in terms of evoked activity and gamma-band synchrony. These findings indicate a role of top-down expectation in generating repetition suppression and are in line with predictive coding models of perception, in which the difference between expected and actual input is propagated from lower to higher cortical areas.
Introduction

When a stimulus is repeated, the neural activity evoked by its second appearance is reduced. This phenomenon, known as repetition suppression (RS), is robust and has been observed across a range of stimulus properties, time scales, sensory modalities and brain areas (Grill-Spector et al., 2006). Thereby, RS has become an indispensable tool in cognitive neuroscience to characterize the nature of neural representations.

In spite of its wide use, the underlying neural mechanisms of RS are still not well understood. On the one hand, it has been hypothesized that RS is the automatic result of changes in the responsivity of relevant neurons, due to stimulus-induced adaptation of the neuronal pool involved (Grill-Spector et al., 2006). Implicit in this view is that RS is an automatic effect, triggered by the repeated sensory stimulation. In contrast to this view, a recent study suggests that RS may rather be a consequence of top-down perceptual expectations (Summerfield et al., 2008). In this study, participants viewed pairs of face stimuli, which could be repeated. Crucially, the probability of a repetition could be either high or low. While the fusiform face area response was strongly reduced when repetitions were expected, it was only moderately reduced when the repetition was unexpected (i.e. RS was attenuated). This suggested the possibility that RS reflects a relative reduction in perceptual “prediction error” when processing an expected stimulus.

Although this explanation of RS in terms of perceptual expectation is appealing, a recent study that attempted to replicate the results of Summerfield et al. using electrophysiological recordings in inferotemporal cortex of macaques found no such effect (Kaliukhovich and Vogels, 2010). Although they observed robust RS, this was not modulated by expectation. While there are some important differences between the two studies, both in terms of task requirements (subjects in Kaliukhovich and Vogels’ study passively fixated while subjects in Summerfield’s study monitored the stimuli for occasional target stimuli) and neural activity measures (Kaliukhovich and Vogels measured electrophysiological markers of brain activity, while Summerfield et al. measured haemodynamic markers of neural activity), this negative finding potentially casts doubt on the account of RS in terms of perceptual expectations.

In the current study we quantified electrophysiological correlates of auditory RS in humans, in order to assess whether RS is modulated by top-down expectation. We measured neural activity over auditory cortex using magnetoencephalography (MEG) while participants listened to repeated tones, which could be either expected or unexpected, and responded to rare deviant tones. In short, we found strong RS for repeated tones in the auditory cortex, which was markedly reduced when the
repetition was unexpected. Thereby, our results provide empirical support for a role of top-down expectation in RS (Baldeweg, 2006; Summerfield et al., 2008).

Materials and Methods

Participants
Sixteen healthy participants (10 female, age 23 ± 3 years, mean ± SD) participated in the experiment upon signing an informed consent form in accordance with the Declaration of Helsinki. All participants had normal hearing and no history of neurological or psychiatric disorders. The study was approved by the regional ethics committee.

Stimuli and experimental design
The experimental stimuli consisted of brief auditory tones (frequency 1000 Hz, duration 5 ms, ~70dB SPL), which were presented binaurally via MEG-compatible air tubes. Stimuli were presented using a PC running Presentation software (Neuro-behavioral systems, Albany, USA).

Each trial started with the presentation of a small central fixation cross on the screen for 2-4 s. Then, an auditory tone was presented, which was either repeated after 500 ms (repetition trial) or not (non-repetition trial). This was followed by an additional period during which the fixation cross was presented (0.5-1 s), and a short period in which the participants could freely move their eyes and blink (1.5-2 s), resulting in a 4-6 seconds intertrial interval (defined as the interval between the last tone of the current trial and the first tone of the next trial). Participants were instructed to listen to the tones and press a button with their right hand if they heard a deviant tone (frequency 1200 Hz). Each block consisted of 90% standard tones and 10% deviant tones.

We manipulated the expectation of repetition by creating different types of blocks. Each block consisted of 86 trials. In some blocks, tone repetitions occurred frequently (75%) and were therefore expected, whereas in other blocks they occurred rarely (25%) and were therefore unexpected. Given that unexpected repetitions were relatively rare (as a logical consequence of our experimental design), we doubled the amount of blocks of trials within this context in order to generate a sufficient amount of trials for statistical analysis. In total, there were two blocks with expected repetitions, and four with unexpected repetitions. We also manipulated the temporal position of the task-relevant rare deviant tones, which could be on either the first or the second temporal position of a tone pair throughout the whole block. Since this manipulation did not induce overall differences in processing the standard tones, we collapsed over
this factor. The participants were informed in advance of each block on whether tone repetitions would be frequent or infrequent within the block, as well as the temporal position of the deviant tones. The experiment lasted one hour, and was preceded by an 8 minute practice session consisting of 60 trials.

**MEG measurements**

Ongoing brain activity was recorded using a whole-head MEG with 275 axial gradiometers (VSM/CTF Systems, Port Coquitlam, British Columbia, Canada) in a magnetically shielded room. Head localization was monitored continuously during the experiment using coils that were placed at the cardinal points of the head (nasion, left and right ear canal). As an aid for eye blink and heartbeat artifact rejection, an electro-oculogram (EOG) was recorded from the supraorbital and infraorbital ridge of the left eye, and an electrocardiogram (ECG) was recorded, both using 10-mm diameter Ag–AgCl surface electrodes.

**MEG data analysis**

The data were analyzed using the FieldTrip toolbox developed at Donders Institute for Brain, Cognition and Behaviour (Oostenveld et al., 2011b) using Matlab 7 (MathWorks, Natick, MA, USA). Data analysis was performed only on the trials consisting of standard tones, and only on trials in which a response was correctly withheld. Data epochs of interest were checked for artifacts using a semiautomatic routine that helped detecting and rejecting trials containing muscle artifacts and jumps in the MEG signal caused by the SQUID electronics. Subsequently, independent component analysis (Bell and Sejnowski, 1995) was used to partial out any remaining variance attributable to eye blinks and heartbeat artifacts (Jung et al., 2000). Finally, the data were visually inspected and any remaining trials with artifacts were removed manually.

**Event-related fields (ERFs)**

Prior to calculating ERFs, data was low-pass filtered using a two-pass Butterworth filter with a filter order of 6, and a frequency cutoff of 40 Hz. ERFs were baseline corrected using an interval of 500 to 400 ms prior to the occurrence of the first tone. A planar gradient transform was then calculated (Bastiaansen and Knosche, 2000). This simplifies the interpretation of the sensor-level data because it typically places the maximal signal above the source (Hamalainen, 1993). In order to avoid differences in the amount of noise when comparing blocks with different numbers of trials, we matched by randomly selecting a subsample of trials from the conditions with more trials.
Time-frequency analysis
We calculated time-frequency representations (TFRs) using a Fourier transform approach applied to short sliding time windows. Prior to the Fourier transform, one or more tapers were multiplied to each time window and the resulting power estimates were averaged across tapers. The power values were calculated for the horizontal and vertical component of the planar gradient and then summed. We then took the median of the planar gradient power estimates for all trials within a condition. For the frequencies 5-35 Hz, we used a single Hanning taper and applied an adaptive time window (T) of four cycles for each frequency (∆T = 4/ƒ) which resulted in an adaptive smoothing of ∆ƒ = 1/∆T. In the higher frequency bands (35-140 Hz) we used a fixed taper length of 200 ms with a ∆f = 20 Hz frequency smoothing (Percival and Walden, 1993). Percentage of change in power was calculated with respect to a baseline window, which was centered around 500 to 400 ms before the presentation of the first tone, and had equal window length as the time windows of interest. Based on the average spectral activity profile, we restricted our data analyses to the theta/alpha band (5-12 Hz) as well as the gamma-band (50-100 Hz) frequencies.

Source localization
Sources of evoked activity were identified using a time-domain beam-forming approach on the axial sensor data [Linearly-Constrained Minimum Variance (LCMV)]. We looked at average activity elicited by the standard tone, between 50-150 ms post-stimulus. We created a realistic single-shell head model for 15 of 16 participants of which we had acquired structural MRI images, using the brain surface from their individual segmented MRIs (Nolte, 2003). The brain volume of each participant was discretized to a grid with a 1 cm resolution and the lead field matrix was calculated for each grid point according to the head position in the system and the forward model. A spatial filter was then constructed for each grid point using the covariance and the lead field matrices. Source strength was calculated in the activation period, and normalized to unit strength for each participant. Individual source estimations were overlaid on the corresponding anatomical MRI, after which the anatomical and functional data were spatially normalized using SPM8 (Statistical Parametric Mapping; http://www.fil.ion.ucl.ac.uk/spm) to the MNI template (Montreal Neurological Institute (MNI), Montreal, Quebec, Canada).

Identification of auditory activation
We performed all statistical analyses on the average activity of 20 sensors (10 over the left hemisphere, 10 over the right hemisphere) that showed maximal auditory activation when averaged across all trial types, conditions, and tone presentations. Loci of auditory activation were defined by identifying the 10 left and right hemisphere sensors that showed maximal activity in the 50-150 ms period following the tone
presentation (Figure 1A). The evoked (Figure 1B) and oscillatory (Figure 1C) activity of this set of sensors constituted the measure of auditory activation that served as the dependent variable for all subsequent analyses.

Statistical analysis
Evoked and oscillatory auditory activity of different conditions were statistically compared using nonparametric cluster-based permutation t-tests (Maris and Oostenveld, 2007). This type of test controls the type I error rate in the context of multiple comparisons by identifying clusters of significant differences over space, time and/or frequency instead of performing a separate test on each sensor, sample and frequency pair. For all analyses, we averaged over the spatial (channel) dimension, on the basis of independent localization of the ten left and ten right channels that showed most robust auditory tone-related activity (see Figure 1). Therefore, our statistical analysis considered one-dimensional (temporal, for the analysis of evoked activity differences) or two-dimensional (spectro-temporal, for the analysis of oscillatory activity differences) clusters. All cluster-level statistics, defined as the sum of t-values within each cluster, were evaluated under the permutation distribution of the maximum (minimum) cluster-level statistic. This permutation distribution was approximated by drawing 5000 random permutations of the observed data. The obtained p values represent the probability under the null hypothesis (no difference between the conditions) of observing a maximum (minimum) cluster-level statistic that is larger (smaller) than the observed cluster-level statistics. We used this method to assess whether there were significant temporal (ERF) or spectro-temporal (TFR) clusters of differential activity.

Results
Behavioral results
The participants’ task was to press a button whenever a deviant tone was presented. Participants correctly responded to virtually all (96.0 ± 0.05%, mean ± SD) of the deviant tones and correctly refrained from responding to virtually all (99.6% ± 0.06%, mean ± SD) of the standard tones. Subjects responded faster to the deviant tone when it was the second tone (710 ms) compared to when it was the first tone (916 ms) of the pair (F(1,15) = 34.8, p < 0.001). Expectation of stimulus repetition did not affect response time (F<1, p>0.10).
Figure 1  Localization of auditory activation

A) Topographic representation of average MEG channel activation for all tones, in the 50-150 ms interval after tone onset. The ten maximally activated channels in each hemisphere are highlighted.  

B) Average auditory evoked field for all tones in selected channels that are highlighted in panel A.  

C) Time-frequency representation of average activity for all tones in selected channels.  

D) Source localization of activity to the first tone, in the 50-150 ms interval after tone onset. The power of the source reconstruction was thresholded at half-maximum.
Neural activity elicited by the auditory stimuli
Auditory tones elicited strong neural activity over bilateral temporal cortex (Fig 1A), which was maximal between 50-150 ms post-stimulus (Fig 1B). A time-frequency representation of the power in the signal showed that the auditory stimulus elicited an increase in low frequency power related to the phase-locked evoked response, as well as an increase in oscillatory activity in the gamma band (60-90 Hz, Fig 1C). Source localization of neural activity between 50-150 ms indicated a bilateral source

Figure 2 Effect of expectation on auditory repetition
A) Auditory evoked fields for expected (in blue) and unexpected (in red) tone repetitions. B) Time-frequency representations for expected (left panel) and unexpected (middle panel) tone repetitions, as well as their difference (right panel). Black rectangular boxes indicate temporal (evoked activity) or spectro-temporal (oscillatory activity) clusters of significant differences between conditions of expectation.
distribution along the superior temporal sulcus (Fig 1D), in the vicinity of the primary auditory cortex (Rademacher et al., 2001).

Expectation of repetition reduces auditory activity for repeated tones

We compared neural responses to expected and unexpected tone repetitions in the sensors that showed strongest auditory activity (Fig 1B). While there were no differences in neural activity elicited by the first tone as a function of repetition expectation,

**Figure 3** Effect of expectation on auditory omission

A) Auditory evoked fields for expected (in blue) and unexpected (in red) tone omissions. B) Time-frequency representations for expected (left panel) and unexpected (middle panel) tone omissions, as well as their difference (right panel). Black rectangular boxes indicate temporal (evoked activity) or spectro-temporal (oscillatory activity) clusters of significant differences between conditions of expectation.
(all p>0.10), repetition expectation strongly modulated the activation elicited by the second tone (Fig 2A). When the repetition was expected, the auditory stimulus resulted in lower evoked activity compared to unexpected repetitions (100-500 ms post-stimulus, p<0.001). Analysis of the time-frequency representations showed similar results, showing a significant spectro-temporal cluster of larger power for unexpected repetitions in the low frequencies (0-350 ms post-stimulus, frequency range 5-9 Hz, p<0.001), as well as in the gamma band (200-300 ms post-stimulus, frequency range 80-95 Hz, p<0.05, Fig 2B).

**Expectation of repetition increases auditory activity for omitted tones**

Interestingly, similar effects of expectation were observed when the second tone was omitted (Figure 3). When subjects expected a tone repetition, a tone omission resulted in a stronger evoked field (100-150 ms post-omission, p<0.05, Figure 3A). Similarly, analysis of oscillatory activity showed larger gamma band power for omitted tones when subjects expected a tone repetition (200-400 ms post-omission, frequency range 60-75 Hz, p<0.05, Figure 3B).

![Figure 4](image.png)

**Figure 4.** Correlation between expectation effects on tone repetition and tone omission

The activity difference between expected and unexpected repetitions (averaged over the 100-500 ms window following the second tone, see Figure 2A) is plotted (on the x-axis), against the activity difference between expected and unexpected omissions (averaged over the same temporal window, on the y-axis).
Since we observed both larger activity for unexpected repetitions (Figure 2A) and unexpected omissions (Figure 2B), we wondered whether these expectation-mediated activity differences were related. For this, we correlated the activity difference between expected and unexpected repetitions (averaged over the 100-500 ms window following the second tone, see Figure 2A) with the activity difference between expected and unexpected omissions (averaged over the same temporal window). Indeed, we found that individual differences in the amount of activity difference between expected and unexpected repetitions were correlated with the amount of activity difference between expected and unexpected omissions ($r=0.43$, $p<0.05$), suggesting that these two phenomena may be mediated by the same neural mechanisms.

**Discussion**

In this study, we examined whether the reduced neural activity for repeated events (repetition suppression, RS) was modulated by expectation of repetition. We observed that the expectation of repetition of auditory events strongly increased repetition suppression in the auditory cortex: the more expected a repeated tone was, the more its evoked response was suppressed. This effect was visible both in early evoked activity 100 ms post-stimulus and by a change in gamma band synchrony 200-300 ms after stimulus onset. The effect of expectation was also present in the absence of a physical stimulus: unexpected omissions resulted in stronger evoked activity and gamma band synchrony over auditory cortex. Finally, individual differences in the amount of activity increase for unexpected repetitions were correlated with the amount of activity increase for unexpected omissions, suggestive of a common neural mechanism for both phenomena.

Our findings are in favor of a top-down account of RS that has been previously suggested (Baldeweg, 2006; Summerfield et al., 2008). In particular, predictive coding models (Rao and Ballard, 1999; Lee and Mumford, 2003; Friston, 2005, 2009) posit that top-down expectations (which are derived from the statistical regularities in the world) help to suppress expected input, thereby constituting an efficient neural coding scheme (Olshausen and Field, 1996, 2004; Friston, 2005, 2009). In this view, feedforward stimulus-evoked activity reflects the mismatch between top-down expectation and sensory input, i.e. prediction error. In our paradigm, the (temporally unpredictable) occurrence of the first tone may set up an expectation about the occurrence of the second tone (which has a fixed temporal lag with respect to the first). This expectation is dependent on the observed statistical regularities within a block: when repetitions are more often observed, the prediction of the occurrence of the second tone will be stronger. Hereafter, the occurrence of the second tone is
associated with reduced prediction error, and hence attenuated neural activity, while the omission of the second tone is associated with increased prediction error, and hence increased neural activity (see den Ouden et al., 2009 for similar results). While our findings support this framework, an alternative (and not mutually exclusive) interpretation of the unexpected omission-induced activity is that it may be a reflection of the prediction signal itself. Indeed, a recent study suggests that the population response to expected and unexpected events may best be explained by a combination of prediction- and prediction-error related responses (Egner et al., 2010). Generally, our findings are well in line with earlier work that has shown that probability can have large effects on early cortical processing in the auditory cortex, both in non-human (Ulanovsky et al., 2003b) and human primates (Haenschel et al., 2000b; Haenschel et al., 2005; Weiland et al., 2008; Valentini et al., 2011a).

A potential limitation of our experimental design is that there was a difference in the total amount of tones between the expectation conditions: blocks where repetitions were expected had overall a larger number of tones than blocks in which repetitions were unexpected, which could potentially lead to generally larger adaptation effects. This effect should however be equally (or potentially even more strongly) present for the first tone of the tone pair. Our results however showed that auditory activity elicited by the first tone of the tone pair was indistinguishable between conditions, while there were large and robust differences between conditions related to the second tone. Also, inter-subject variability in the observed neural activity increase for unexpected repetitions were correlated with variability in neural activity increase for unexpected omissions. While this is in line with both these effects stemming from one neural mechanism (increased prediction error), these findings do not seem consistent with a low-level adaptation account.

In our study, we manipulated the expectation of tone repetition, while the pitch of the tones was kept constant. Therefore, it could be argued that our study jointly manipulated the expectation of the occurrence of an event in general (information-bound surprise), along with the expectation of a particular stimulus (stimulus-bound surprise). In this regard, our study differs from the RS design of Summerfield et al., in which expectations were induced about the identity of a face (stimulus-bound surprise), but not about whether a face stimulus would be presented in general (information-bound surprise). However, the fact that our expectation modulation was early (100 ms) and confined to sensory regions (auditory cortex) may suggest that it reflects sensory (stimulus-bound) predictions, rather than more general (information-bound) predictions about event occurrence. At a later processing stage, both forms of surprise are associated with a more wide-spread and later (~300 ms) increase in activity, which has been localized in a fronto-parietal network (McCarthy et al., 1997; Mars et al.,
2008; Bekinschtein et al., 2009). A follow-up study that specifically modulates the predictability of a particular tone, while keeping constant the probability of tone occurrence, could help to further dissociate these effects.

The observed expectancy modulation on repetition effects is in line with earlier human work that has shown increased neural activity in primary sensory cortex following both surprising presence and absence of sensory stimulation (Garrido et al., 2007; den Ouden et al., 2009b), which is also manifest in the gamma-band (Gurtubay et al., 2006; Wyart and Tallon-Baudry, 2008). Interestingly, recent neurophysiological work has indicated that gamma-band oscillations are particularly strong in superficial layers of the cortical column (Lakatos et al., 2005; Maier et al., 2010; Maier et al., 2011). These superficial layers have dense “forward” connections to higher-order areas (Thomson and Bannister, 2003b). Therefore, it is tempting to view the observed larger gamma-band activity for unexpected presence and absence of sensory stimulation in the auditory cortex as a “prediction error” response that is fed forward from early auditory cortex to higher order regions (Rao and Ballard, 1999; Friston, 2005). In this sense, the gamma-band activity we observed may signal a “prediction error” response. This response should therefore not be confused with the expectancy state itself, which is expressed in the temporal structure of activity patterns before the appearance of stimuli (Engel et al., 2001; van Ede et al., 2011).

While we observed a strong modulation of RS by expectation, this does not preclude that RS is also partly driven by automatic stimulus-driven mechanisms such as fatigue (Miller and Desimone, 1994; Grill-Spector and Malach, 2001) or sharpening (Desimone, 1996). In fact, a recent study that aimed to replicate the fMRI findings by Summerfield et al. found robust RS without any modulation of RS by expectation (Kaliukhovich and Vogels, 2010). This study measured neural activity (local field potentials and spike rates) in the inferior temporal (IT) cortex while monkeys observed complex visual stimuli (fractals and natural stimuli). The researchers observed robust RS, but complete absence of a RS modulation by expectation. While the lack of modulation by RS in their study is puzzling, it should be noted that there are marked differences between their study and the current study. While they measured neural activity from a higher-order area (inferotemporal cortex), we analyzed responses in early sensory cortex. It is well possible that expectation-related modulations are more pervasive in early sensory than later stages of cortical processing (Rao and Ballard, 1999). Another important aspect may be the presence or absence of selective attention. Previous studies have observed that RS is markedly increased by selective attention (Eger et al., 2004; Murray and Wojciulik, 2004a; Yi and Chun, 2005b). While the monkeys in the study by Kaliukhovich and Vogels (2011) passively fixated the screen, the subjects in our study (as well as in the study by Summerfield et al.) where required to monitor the
stimuli for occasional targets to which they had to respond. Therefore, the increased attentional state may have enabled the occurrence of expectation effects. This notion remains speculative however, and could be a fruitful topic of future research. Of note, while we find a strong modulation of expectation on RS, it is highly plausible that this is not the only mechanism by which RS can occur. Indeed, RS may partly reflect intrinsic cellular properties of the system (Zucker, 1989; Farley et al., 2010).

In conclusion, we provide evidence for a top-down mediation of repetition suppression (RS) by expectation in early auditory cortex. These findings are of importance for studies that use RS as a tool to probe the functional representation of neuronal populations, since activity reductions due to repetition may be related to the predictive relationship between the first and second stimulus, rather than the repetition of a particular representation.
Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band

This chapter is based on:
Abstract

The neural response to a stimulus is influenced by endogenous factors such as expectation and attention. Current research suggests that expectation and attention exert their effects in opposite directions, where expectation decreases neural activity in sensory areas, while attention increases it. However, expectation and attention are usually studied either in isolation or confounded with each other. A recent study suggests that expectation and attention may act jointly on sensory processing, by increasing the neural response to expected events when they are attended, but decreasing it when they are unattended. Here we test this hypothesis in an auditory temporal cueing paradigm using magnetoencephalography in humans. In our study participants attended to, or away from, tones that could arrive at expected or unexpected moments. We found a decrease in auditory beta band synchrony to expected (versus unexpected) tones if they were unattended, but no difference if they were attended. Modulations in beta power were already evident prior to the expected onset times of the tones. These findings suggest that expectation and attention jointly modulate sensory processing.
**Introduction**

In a world where the senses are continuously stimulated, perception acts to optimize information processing by prioritizing based on behavioral goals and expectations. This prioritization can encompass spatial locations or object features, but the brain can also prioritize temporal windows. There may be two distinct classes of processes guiding temporal selection: *temporal expectation*, the ability to extract temporal regularities from the environment, and *temporal attention*, the state of anticipating relevant future events. Researchers who focus on temporal expectation typically find that temporally predictable events lead to *less* neural activity than temporally unpredictable events (Lange, 2009; Alink et al., 2010; Bendixen et al., 2012; Schwartze et al., 2013). Conversely, researchers who focus on temporal attention typically find that attended events lead to *more* neural activity than unattended events (Correa et al., 2006; Schroeder and Lakatos, 2009). This would suggest that expectation and attention act as opposing forces in sensory processing. However, expectation and attention are largely studied either in isolation (Todorovic and de Lange, 2012; Wacongne et al., 2012; SanMiguel et al., 2013a) or conflated (Posner et al., 1980). For example, temporal cueing studies often contrast expected, attended events with unexpected, unattended ones (Miniussi et al., 1999; Lampar and Lange, 2011), making it difficult to tease apart the contributions of expectation and attention (Summerfield and Egner, 2009; Lange, 2013).

Several studies have looked into the interplay of expectation and attention on sensory processing, but the picture that emerges is ambiguous. On the one hand, it has been proposed that attention is necessary for expectations to affect sensory processing (Larsson and Smith, 2012). On the other hand, mismatch negativity studies, which compare rare deviant events to frequent standards, suggest that attention is not necessary for expectations to form (Näätänen, 1990); although see (Woldorff et al., 1991). Finally, a spatial cueing study where expectation and attention were orthogonally manipulated observed an interaction in the form of expectation suppression for unattended events, but expectation enhancement for attended events (Kok et al., 2012b). This result is consistent with the suggestion that expectation interacts with attention in a synergistic manner (Feldman and Friston, 2010; Friston, 2010), with increased activity to attended events and attenuated activity to ignored events. Here we test whether this interaction can be generalized to auditory temporal expectation.

We orthogonally manipulated expectation and attention in an auditory temporal cueing paradigm while recording neural activity using magnetoencephalography (MEG). Participants listened to pairs of tones which were separated by a predictable or unpredictable temporal interval (leading them to form a focal or distributed temporal
expectation), while performing a task on either the first or the second tone in each pair. We found evidence for interacting effects of expectation and attention, with a decrease in auditory beta power when having focal (versus distributed) expectation, but only if attention was drawn away from the tones. With attention, there was no effect of temporal expectation. This suggests that expectation and attention jointly act to guide sensory processing, with expectation potentially facilitating the filtering out of temporally predictable, irrelevant events.

**Methods**

**Participants.** Twenty five healthy participants (17 female, age 23.7 ± 7.8 years, mean ± SD) took part in the experiment upon signing an informed consent form in accordance with the Declaration of Helsinki. All participants had normal hearing and no history of neurological or psychiatric disorders. Sixteen of these participants also took part in another auditory expectation study on the same day (reported in Todorovic et al., 2011). The dataset of one participant was not analyzed due to excessive artifacts (>30% of the trials). The study was approved by the regional ethics committee (Committee on Research Involving Human Subjects, Region Arnhem-Nijmegen, The Netherlands).

**Stimuli and experimental design.** The experimental stimuli consisted of brief pure tones (frequency 1000 Hz or 1200 Hz, duration 5 ms, ~70 dB SPL), which were presented binaurally via MEG-compatible air tubes. Stimuli were presented using a PC running Presentation software (Neurobehavioral Systems).

Each trial started with the presentation of a central fixation cross (2–4 s). A standard tone (1000 Hz) was then presented twice in quick succession, with one of five possible inter-stimulus-intervals (ISI) between the two presentations (250, 375, 500, 625, and 750 ms). The fixation cross remained present for an additional period (0.5–1 s), followed by a short period when it was removed, in which the participants could freely move their eyes and blink (1.5–2 s). This resulted in a 4–6 s intertrial interval, defined as the interval between the last tone of the current trial and the first tone of the next trial. Participants were instructed to listen to the tones and press a button with their right index finger if they heard a deviant tone (1200 Hz). Each block consisted of 91% trials with standard tones and 9% trials with a deviant tone.

We manipulated the temporal expectation of the second tone by varying the relative frequencies of the different inter-stimulus intervals per block (Fig. 1). Recent research has shown that listeners are sensitive to the distribution of tone frequencies, with
narrower distributions leading to a stronger expectation that the following tone will fall close to the mean (Garrido et al., 2013). We adapted this paradigm to the temporal domain, by creating blocks with narrow or wide temporal distributions. In blocks with a wide temporal distribution, the second tone appeared at all five ISIs with similar rates (19%, 19%, 24%, 19%, 19%). This led to a state of distributed temporal expectation, as the second tone could be roughly equally expected to appear after any of the five possible intervals. Conversely, in blocks with a narrow temporal distribution, the second tone appeared at the middle ISI frequently in comparison with the surrounding four ISIs (75%, 75%, 70%, 75%, 75%). This led to a state of focal temporal expectation, as the second tone was most likely to appear after the duration of the middle temporal interval. We were interested in comparing neural activity elicited by tones at this middle ISI in relation to temporal expectation. The tone pairs separated by the remaining four ISIs were not analyzed due to a small number of trials. In sum, each of the analyzed trials contains two appearances of a 1000 Hz tone, separated by 500 ms, where the temporal predictability of the second tone varied per experimental block. Given that tones at the middle ISI were also comparatively rare in the condition of distributed expectation (as a logical consequence of our experimental design), we doubled the amount of blocks within this context to generate a sufficient amount of trials for statistical analysis. In total, there were four blocks with distributed temporal expectation (88 tones at the middle ISI and 254 tones at the remaining ISIs), and two blocks with focal temporal expectation (120 tones at the middle ISI and 52 tones at the remaining ISIs). When the trials with deviant tones and artifacts were removed, an average of 76 trials with distributed temporal expectation and 112 trials with focal expectation were analyzed per subject. Block order was counterbalanced across subjects.

In addition to temporal expectation, we also manipulated temporal attention, in an orthogonal fashion. We presented the task-relevant deviant tones on either the first or the second position in the tone pairs, but this position was held constant per block. In other words, in every block participants were informed whether to expect targets to appear only on the first or only on the second tone position. This manipulation thus drew their attention to the moment the first (or second) tone was displayed, and made the moment of presentation of the other tone task-irrelevant.

Prior to the experiment, there was an 8 min practice session which contained 60 trials. The distribution of trial and block types in the practice session was identical to the distribution in the experiment. The experiment lasted one hour.

**MEG measurements.** Brain activity was recorded using a whole-head MEG with 275 axial gradiometers (VSM/CTF Systems) in a magnetically shielded room. Head localization was monitored continuously during the experiment using coils placed at the nasion...
and left and right ear canals. As an aid for eye blink and heartbeat artifact rejection, an electro-oculogram (EOG) and electrocardiogram (ECG) were recorded, using 10-mm-diameter Ag–AgCl surface electrodes. For the EOG, these electrodes were placed on the supraorbital and infraorbital ridge of the left eye. For the ECG, these electrodes were placed above the clavicle on the left side, and on the abdomen on the right side, ~15cm below the rib cage.

**MEG data analysis.** The data were analyzed in Matlab (MathWorks) using the FieldTrip toolbox developed at The Donders Institute for Brain, Cognition, and Behavior (Oostenveld et al., 2011a)). Data analysis was performed on the trials consisting of standard tones only where the response was correctly withheld, and which were temporally separated by 500 ms. Data epochs of interest were checked for artifacts caused by muscle activity and SQUID jumps. We used semiautomatic routines to discarded contaminated trials. Subsequently, independent component analysis (Bell and Sejnowski, 1995) was used to remove any remaining variance attributable to eye blinks and heartbeat artifacts (Jung et al., 2000). Finally, the data were visually inspected and any remaining trials with artifacts were removed manually.

**Figure 1 Task structure**

All trials contained two tones. The first tone is shown here appearing at time point zero. In blocks with a distributed temporal expectation (left panels), the second tone appeared at a similar rate at five possible ISIs after the first tone, while in blocks with a focal expectation (right panels) the middle ISI was most frequent. Attention was either on the first tone (top panels) or the second tone (bottom panels). Numbers above the bars represent the percent of trials per block in which standard tones appeared at their respective time points. Numbers on the x-axis denote the duration of the ISIs. Only tone pairs with an ISI of 500 ms were analyzed.
**Time–frequency analysis.** We calculated time–frequency representations (TFRs) using a Fourier transform approach applied to short sliding time windows. For the low frequencies (5-35 Hz) the length of the time window $T$ was frequency dependent, and consisted of three oscillation cycles for each frequency ($\Delta T = 3/f$). We applied a Hanning taper to control for spectral leakage. Effectively, this led to an adaptive frequency resolution of $\Delta f = 1/T = f/3$. The time windows were advanced in steps of 25 ms. For the high frequencies (40-120 Hz), we used a fixed window length of 200 ms with a $\Delta f = 20$ Hz frequency smoothing, using multitapers (Percival and Walden, 1993). Power estimates were first calculated separately for the horizontal and vertical component of the planar gradient and then summed (Bastiaansen and Knoche, 2000). This simplifies the interpretation of the sensor-level data because it places the maximal signal above the source (Hämäläinen et al., 1993). Raw planar gradient power estimates were converted to decibel (dB), by log-transforming with base 10 and then multiplying by 10. We then subtracted the power in the baseline window, centered at 500 to 400 ms before the presentation of the first tone. The TFR was calculated and baseline corrected per trial, a procedure that is equivalent to taking a relative baseline of raw power. We then selected the median of the planar gradient power estimates across trials within each condition. Selecting the median, rather than the mean, has the benefit of the data being less affected by noise and outliers. This is especially relevant for power estimations which are always positive, i.e. noise and outliers will sum rather than average out if the mean is used. The obtained TFRs consist of both evoked and oscillatory activity, and do not allow for a decisive separation of the two. We additionally used a multitaper approach (Mitra and Pesaran, 1999) to visualize the beta power in the pre- and post-stimulus time windows where we found clusters of significant differences between conditions, by estimating the spectral energy at the peak frequency of observed differences (23 Hz) with a 9 Hz frequency smoothing bandwidth. Furthermore, we conducted a within-subjects trial-by-trial analysis, in which we separately calculated TFRs for the theta (5-7 Hz) and beta (13-32 Hz) bands, in two time windows corresponding to pre- and post-stimulus activity. These calculations were done separately (beta using an early, pre-stimulus window, theta using a post-stimulus window) in order to prevent temporal uncertainty in the power estimates from influencing the results of the correlation analysis. The range of frequencies within the beta band was chosen to match the range where we found an interaction between temporal prediction and attention in the between-subjects analysis. We previously observed that differences in low frequency power across expectation conditions correspond to effects obtained with ERFs (Todorovic et al., 2011), when using an equally long tone of the same frequency as used in this experiment. This is a consequence of the latency of the evoked response to auditory tones, which peaks about 100 ms post-stimulus and decreases over the next 100 ms (Hämäläinen et al., 1993), thus corresponding to a single 5 Hz cycle. The post-stimulus theta band thus contains the spectral energy that is produced by the evoked responses.
Source localization. We acquired structural MRI scans of 23 out of 24 participants in our study using a 1.5 T Siemens Magnetom Sonata system (Erlangen, Germany). Three of those scans were of insufficient quality for source reconstruction. We thus performed source reconstruction on 20 participants using a subject specific realistic head model extracted from the individual segmented MRIs. For the remaining four participants we used a head model created from a template MRI. We identified sources of activity using a frequency-domain beamformer, Dynamic Imaging of Coherent Sources (DICS). The brain volume of each participant was discretized to a grid with a 1 cm resolution and the lead field matrix was calculated for each grid point using a single shell volume conduction model based on the inner surface of the skull. For each grid position we computed an adaptive spatial filter to estimate oscillatory power in the entire brain (Gross et al., 2001; Liljestrom et al., 2005), using the lead field matrix at that location and a cross-spectral density matrix defined between all pairs of MEG sensors. We used a ‘common filter’ approach, where for a given contrast of interest the spatial filter was computed from a cross-spectral density matrix that had been estimated from the data of the baseline and post-stimulus period combined. We computed oscillatory activity in several frequency bands and time windows. For each of these power estimations, the baseline window was equally long as the post-stimulus window of interest, beginning at 500 ms prior to the onset of the first tone in a trial. First, in order to examine the average sensory representation that the tones elicited, we estimated theta power (5-7 Hz) in the 300 ms after the first tone in all trials by means of the Fourier transform of the Hanning-tapered signal. Second, in order to examine the spatial specificity of the experimental effect that we obtained at the sensor level, we calculated beta power in a pre-stimulus window 320 to 100 ms prior to the second tone, and in a window related to stimulus processing, from 20 ms prior to 130 ms after the onset of the second tone. These two time windows correspond to the first two clusters obtained in the statistical tests. We then made an averaged source reconstruction over these two time windows in order to get an estimation of the source of the experimental effect. Here we applied a multitaper approach (Mitra and Pesaran, 1999) to obtain a spectral estimate at 23 Hz with a 9 Hz frequency smoothing bandwidth, resulting in a power estimate from 14 to 32 Hz. We compared different conditions by subtracting the power values at each grid point. The individual subjects’ source reconstructions were coregistered to their anatomical MRIs, and the anatomical and functional data were subsequently spatially normalized to the International Consortium for Brain Mapping template (Montreal Neurological Institute, Montreal, Quebec, Canada; h...
Statistical analysis. Oscillatory activity in different conditions was statistically compared using nonparametric cluster-based permutation tests (Maris and Oostenveld, 2007). This type of test controls the type I error rate in the context of multiple comparisons by identifying clusters of significant differences over space, time, and/or frequency instead of performing a separate test for all pairs of sensors, samples, and frequency bins. All cluster-level statistics, defined as the sum of $t$ values within each cluster, were evaluated under the permutation distribution of the maximum (minimum) cluster level statistic. This permutation distribution was approximated by drawing 5000 random permutations of the observed data. The obtained $p$ values represent the probability under the null hypothesis (full exchangeability, i.e. no difference between the conditions) of observing a maximum cluster-level statistic that is larger than the observed cluster-level statistics, and a minimum cluster-level statistic that is smaller than the observed cluster-level statistics.

In order to examine auditory activity, we averaged over the spatial (channel) dimension on the basis of independent localization of the 10 left and 10 right channels that showed the most robust tone-related activity (Fig. 2A,B). We estimated the sensors based on the average activity of all the tones in our experiment, irrespective of experimental condition. We then assessed whether there were significant spectro-temporal clusters of differential activity between the experimental conditions, within a 1-second period after the first tone, in either low (5-35 Hz in steps of 1 Hz) or high (40-120 Hz in steps of 3 Hz) frequencies. For assessing the interaction between expectation and attention, we applied cluster-based permutation tests that evaluated whether there was a difference between the two levels of attention (task-relevant versus task-irrelevant) with respect to the expectation effect (distributed versus focal). To qualify the nature of the interaction, we constrained the analyses to the time points and the frequency bands of three clusters where we observed a significant interaction. We conducted T-tests for these post-hoc analyses. The $p$-values reported here are part of a post-hoc analysis of interaction effect identified in three time-frequency windows (using a method that controls the false rate across time and frequency). Because these time-frequency windows were identified on the basis of the interaction effect in the data, the reported $p$-values of this post-hoc analysis can be used as an index that allows us to explore the nature of the interaction effect (of which the existence was demonstrated in an analysis that controls the false alarm rate). Furthermore, to determine whether the effect which was observed before the onset of the second tone, had an influence on subsequent stimulus processing of this tone, we conducted a trial-by-trial analysis in which we correlated pre-stimulus beta power (13-32 Hz) with post-stimulus theta power (5-7 Hz) for each individual subject. We performed a T-test to assess whether the obtained correlation coefficients differed significantly from zero at the group level.
Results

Behavioral results
Participants were instructed to press a button whenever they heard a deviant tone. They did this task with high accuracy, correctly responding to 95.6% of the deviants (SD = 5.3), and correctly withholding a response to 98.8% of the standards (SD = 0.5). It is important to note that deviant tones in this experiment were rare, and that we only analyzed reaction times to deviant tones occurring at the middle ISI (a total of 14 trials). The behavioural results on the deviants should therefore not be taken as a strong indicator of the presence or absence of experimental effects on the standard tones, which were the subject of our MEG analyses. Participants were faster to respond to the deviants when they occurred on the second tone as opposed to the first tone (first: RT=844 ms, second: RT=635 ms; F(1,23)=46.3, p<0.01). Reaction time did not depend on whether temporal expectation was focal or distributed (F(1,23)=3.51, p=0.56), nor was there an interaction with the locus of attention (first versus second tone: F(1,23)=2.55, p=0.12).

Neural activity modulations elicited by auditory processing
Tones reliably elicited bilateral activity over temporal sensors. We calculated tone-related activity for all the tones regardless of experimental condition, and compared their spectral power to baseline activity. In the low frequencies (5-35 Hz), neural activity to the average of all tones was maximal 50-150 ms following tone onset, while in the high frequencies (50-90 Hz) this activity was maximal after 100-200 ms. From these time windows, we selected 10 left and 10 right channels which showed strongest activity (Fig 2A,B). The spectral signature of low frequency neural activity over all trial types in these channels (Fig. 2C) was predominantly visible as a power change in the theta frequency band (5-7 Hz), which spread to the alpha (8-12 Hz), and beta bands (15-25 Hz). Changes in gamma band power (50-90 Hz) following tone presentation were evident as well. In order to confirm that low frequency power changes involved auditory activity, we also performed source reconstruction of theta-band activity of these tones, which showed a focal increase in power over temporal cortices (Fig 2D), consistent with a source in the auditory cortex.

Interaction effects on beta-band oscillatory activity between temporal expectation and attention
We first investigated whether there were overall differences in oscillatory activity for distributed vs. focal expectation, as well as for attention to the first or to the second tone. Somewhat surprisingly, neither of these factors modulated the spectral signature of auditory processing in isolation (expectation, low frequencies: p=0.164, high frequencies, p=0.290; attention, low frequencies: p=0.164, high frequencies:
Temporal expectation and temporal attention interact

However, an interaction between temporal expectation and temporal attention was evident in the beta band ($p=0.019$), beginning prior to the onset of the second tone and lasting throughout the trial epoch. This interaction was most prominent in three clusters: before the onset of the second tone (360-100 ms pre-stimulus, frequencies: 13-32 Hz, Figure 3C – left panel), during early tone processing (20 ms prior to tone onset to 130 ms after tone onset, frequencies: 15-29 Hz, Figure 3C – right panel), and during late tone processing (350-500 ms after the onset of the second tone, frequencies: 15-35 Hz). We found no interaction between our experimental factors on oscillatory activity in the high frequency range ($p=0.177$). We further investigated the nature of the low frequency interaction by comparing average beta power in these three time windows between conditions of distributed and temporal expectation, separately for each of the two attention conditions. We found that the effect of temporal expectation on beta power depended on whether a tone was

Figure 2  Localization of auditory activation

A) Topographic representation of MEG channel activation for all tones at 5–35 Hz in the interval of maximal activity after tone onset. The maximally activated channels in each hemisphere are highlighted. B) Same as A for frequencies from 50 to 90 Hz. C) Time-frequency representation for all analyzed trials, averaged over the selected sensors. D) Source reconstruction of low frequency power in the theta band (5-7 Hz) shows bilateral peaks of activation in superior temporal cortices.
attended or not. When the first tone was attended (Fig 3A), there was less beta-band activity when expectation was focal than when expectation was distributed. This was true both prior to the onset of the second tone \(t(23) = 2.91, p = 0.007\) as well as after its onset \(t(23) = 2.76, p = 0.011\). This expectation effect was absent when the second tone was attended (pre-stimulus: \(t(23) = -0.96, p = 0.348\); post-stimulus: \(t(23) = -1.42, p = 0.17\)), with even a hint of more beta-band activity when expectation was focal than when it was distributed (Fig 3B). During late tone processing we found no difference in expectation conditions (unattended: \(t(23) = 1.75, p = 0.09\), attended: \(t(23) = -1.54, p = 0.136\)).

To gain more specificity with respect to the source of the interaction between temporal expectation and temporal attention, we used a beamformer approach to localize beta power in the time windows corresponding to the pre-stimulus and post-stimulus effects (i.e. the first two clusters). We averaged this source-level activity over the pre-stimulus and post-stimulus time windows. We compared the temporal expectation effects by calculating the difference in average power between brain activity related to distributed vs. focal expectation separately for each attention condition, and then subtracted the expectation effect when the first tone was attended, from the expectation effect when the second tone was attended. This procedure is equivalent to the sensor-level analysis, and the interaction we observed at the sensor level should manifest as increase in beta power in the source plot. We found that the interaction between temporal expectation and attention was most prominent over left superior temporal cortex (Fig 3D), which is in line with a sensory effect. There was an additional positivity in the right frontal cortex, a source which has often been described in mismatch negativity studies (Alho, 1995). Notably, the negativity in the source plots, spanning large areas of the occipital and right parietal cortex, is opposite in direction from our sensor level effect, indicating that these areas also potentially involve a beta modulation, but that this modulation is opposite in sign compared to the significant interaction we observed in the sensors above auditory cortex.

**Beta activity preceding a tone is predictive of activity later elicited by that tone**

Finally, we were interested to know whether activity in the beta band prior to the onset of the second tone, which we found to be jointly modulated by temporal expectation and temporal attention, is predictive of the neural activity that was elicited by the second tone. To investigate the relationship with pre-stimulus activity, we correlated pre-stimulus beta power (which we found to be related to expectation and attention), with theta power (which contains the energy of the evoked response) following the onset of the second tone. To control for factors which might lead to correlated power in all frequency bands within a trial (e.g. head movement, noise,
Temporal expectation and temporal attention interact

Chapter 3

Figure 3 Time-frequency representations of stimulus-related activity in selected sensors as a function of expectation and attention

A) When participants were attending to the first tone in each tone pair, there was more beta band activity in trials where temporal expectation was distributed (left panel) than when it was focal (middle). Rectangles represent spectrotemporal clusters where a significant interaction of expectation and attention was found (right). B) When participants were attending to the second tone in each tone pair, there was no difference in beta band activity in trials where temporal expectation was distributed (left panel) than when it was focal (middle). Rectangles represent spectrotemporal clusters where a significant interaction of expectation and attention was found (right). C) Interaction between expectation and attention during pre-stimulus activity (the leftmost cluster in A and B) and post-stimulus activity (the middle cluster). D) Source reconstruction of the beta-band interaction between temporal expectation and attention.
power fluctuation in lower frequencies, etc.) we removed all common variance between pre-stimulus beta power (360-100 ms pre-tone 2, 13-32 Hz) and pre-stimulus theta power (360-100 ms pre-tone 2, 5-7 Hz) prior to the correlation with post-stimulus theta power (0-300 ms post-tone 2, 5-7 Hz). We found that the magnitude of pre-stimulus beta power was positively correlated with the magnitude of the post-stimulus activity after correcting for instantaneous correlations (median $r=0.12$, $t(23)=8.67$, $p<0.001$). In other words, a stronger beta power prior to tone onset was associated with a stronger neural response to the tone after its onset.

**Discussion**

In this study, participants listened to pairs of tones while performing a task either on the first or on the second tone, where the expectation of the onset time of the second tone was either distributed or focal. We observed that temporal expectation and temporal attention jointly modulated beta band activity in an interacting fashion. We found that expectation led to a decrease in beta power, but only in the absence of attention (Fig 3A). With attention, an expectation effect was not significantly present (Fig 3B). This interaction was most pronounced in left superior temporal cortex, suggestive of a localization in auditory cortex. Additionally, we found that higher beta power prior to tone onset led to higher neural activity (theta power) evoked by the tones. Our results suggest that expectation guides temporal attention: expectation attenuates the response to events happening at unattended moments, but it does not affect the response during attended time windows.

Temporal expectation and attention guide perception by highlighting moments in time when sensory processing becomes particularly efficient. For example, stimuli are recognized more quickly if their appearance is temporally predictable (Lange and Roder, 2006; Bendixen et al., 2009; Schroeder and Lakatos, 2009). While temporal processing involves several specialized cortical circuits, such as the basal ganglia or the cerebellum (Schwartz et al., 2012), processing of time is partly also local to sensory cortices (Merchant et al., 2013). This local processing may rely on oscillatory processes that entrain to rhythms present in external stimulation (Nobre et al., 2007; Schroeder and Lakatos, 2009; Stefanics et al., 2010; Giraud and Poeppel, 2012). Sensory beta band activity is known to be modulated by integrative cortical processes such as attention or decision making (Donner and Siegel, 2011; Siegel et al., 2012), and beta oscillations are suggested to play a role in conveying predictions about future sensory input (Arnal and Giraud, 2012). For example, beta band activity has been linked to novelty processing in the auditory system in humans (Haenschel et al., 2000a), with increased power to unexpected tones in a mismatch negativity (MMN) paradigm. Additionally,
Temporal expectation and temporal attention interact

Chapter 3

Beta power is predictive of auditory repetition suppression (Hong et al., 2008), a phenomenon that has strong links with stimulus expectation (Summerfield et al., 2008; Todorovic et al., 2011). In studies of temporal attention, low frequency entrainment to externally generated rhythms (Schroeder and Lakatos, 2009) can become coupled to modulations in beta power (Cravo et al., 2011; Arnal et al., 2014) which varies as a function of stimulus anticipation (Jones et al., 2010; Saleh et al., 2010; van Ede et al., 2010; van Ede et al., 2011). Interestingly, our results showed a similar pattern of beta power modulations, but in the absence of an external rhythm. Although the task did not involve estimating durations, participants could make use of the temporal structure of the sounds in order to enhance their detection performance. Each of the trials contained only two tones, where the temporal uncertainty of the second tone was derived based on the distribution of inter-stimulus intervals across trials within each block, while the inter-trial interval varied randomly. While listening to the tones, beta band activity increased or decreased as a joint function of expectation and attention. How did this beta modulation, which is typical for rhythmic stimulation, arise in a situation where there is no rhythm? One possibility is that assessments of interval durations rely on oscillatory processes, and that already one tone may be sufficient to entrain activity, perhaps via a phase reset, in the presence of a task that benefits from accurate time assessment. The modulation in beta power may also reflect an endogenous change in neural synchrony related to expectation and attention. Finally, this modulation could reflect entrainment to activity related to omitted tones. In the trials we analyzed, there were two shorter and two longer intervals when the second tone could have potentially occurred, but did not. Unexpected auditory omissions have previously been shown to result in activity in the auditory cortex (Raij et al., 1997; SanMiguel et al., 2013b). This raises the intriguing possibility that not only a rhythmic stimulus train, but also a rhythmic expectation of a stimulus, could underlie oscillatory activity related to temporal processing. However, neither of these hypotheses is sufficient to explain why the experimental effects were particularly prominent in the blocks where participants performed a task on the first tone in the pairs. In these blocks, once the trial started with the presentation of the first tone, there was no longer anticipation that a target tone might appear on the second tone. A more accurate assessment of the interval between the two tones would thus not be beneficial for performing the task. In fact, this would have led to a main effect of temporal attention, which we did not observe. Similarly, we have previously observed that activity elicited by tone omissions scales with the (im)probability of the omission (Todorovic et al., 2011). As the absence of tones at early ISIs was more probable when temporal expectation was focal than when it was distributed, beta band modulations during tone omissions would also be expected to be more prominent in the distributed relative to the focal condition, which would have been reflected in a main effect of stimulus expectation. However, we did not find such
an effect either. Instead, we found an interacting effect of expectation and attention. It is possible that the manipulations of attention and expectation in the study were weak, and thus did not lead to main effects: the task could be solved without much attention thereby precluding a strong attentional modulation in our subjects. At the same time, the most likely tone was on the middle ISI in both types of expectation conditions, precluding a strong expectation modulation. However, the presence of the interaction indicates that both temporal expectation and temporal attention played a role in tone processing. An alternative possibility is that the main effects are averaged out because expectation differently modulates auditory activity in the presence of attention than in its absence, by filtering out potentially distracting information if it is predictable but preserving information if it is task-relevant.

Auditory processing appears to be especially sensitive to temporal information (Kosem et al., 2014). Temporal processing is inextricably linked to both stimulus expectation (reflected in the adjustment of neural activity to temporal regularities in the environment) and stimulus attention (reflected in the anticipation of a stimulus in order to assess it). In many situations, these two top-down factors cannot be well separated. For example, rhythmic stimulation prompts temporal attention around the occurrence of stimuli, but later beats are more temporally predictable than early ones, based on the stimulus history. Expectation and attention are also often conflated in research, especially in Posner-type tasks where stimuli appear in validly cued temporal windows more often than in invalidly cued ones. However, when attention is held constant, temporal expectation is found to reduce auditory activity (Costa-Faidella et al., 2011; Geiser et al., 2012), and when expectation is held constant, temporal attention increases it (Lange and Roder, 2006). This raises the question whether expectation and attention represent opposing forces, with the potential to cancel each other out. A recent review of predictive auditory processing suggests that unexpected events may lead to increased neural activity when unattended, but decreased activity when attended (Bendixen et al., 2012). Evidence from the visual domain supports this suggestion, as a spatial attention study found increased V1 activity to unexpected gratings if they were unattended, but reduced activity if they were attended (Kok et al., 2012b). We hypothesized that joint contributions of temporal expectation and temporal attention represent a general principle of sensory processing: when it is possible to form an expectation, an attended stimulus may undergo enhanced processing, but an unattended stimulus will be suppressed. We indeed found an interaction between expectation and attention. This interaction involved a modulation of beta power, and was already evident prior to the occurrence of the tones. However, once attention was directed to the time window of interest, expectation did not modulate auditory beta power. This might also indicate that temporal attention is relatively imprecise in comparison with the inter-stimulus intervals in our study,
Temporal expectation and temporal attention interact leading to prioritized processing around the entire time when the second tone may occur but without the capacity for further temporal fine-tuning based on expectation. Furthermore, the onset of the second tone overlaps in time with late processing of the first tone. The observed activity differences could thus also be partly related to late processing of the first tone, rather than to a preparatory process related to the second tone. Namely, additional attention might have been allocated to late processing of the first tone when the temporal prediction of the second tone was more precise, because the first tone is more informative about the onset of the second tone in this condition. However, in this case, the expected result would be a difference between focal and distributed processing when the second tone is task-relevant, whereas we found a difference only when the first tone was task-relevant.

It has been suggested that sensory correlates of timing involve activity generated by the motor system (Schubotz et al., 2000). Even though the cortical sources of the beta band activity that we observed were predominantly located over auditory cortex (Figure 3D), it cannot be excluded that they may also be partly generated in the motor system. However, none of the analyzed trials contains an overt motor response, and we did not observe reaction time differences across the different expectation blocks, which would have signaled a difference in motor preparation. Additionally, pre-stimulus modulations of beta power tend to be opposite in direction of post-stimulus motor activity (Pfurtscheller et al., 2003). In our study, neural activity elicited by the tones is indexed by theta activity (which, rather than being a true oscillation, represents a frequency description of the evoked field). Instead of a negative relationship, we found a positive one: within subjects, on trials where beta power was higher prior to the onset of the second tone, neural activity elicited by the tone was also higher. This effect was however not consistent enough to translate into a theta modulation by expectation or attention across subjects. Interestingly, a modulation of beta power in the opposite direction of the one we observed in our sensor level analyses was visible over parietal cortex, which is similar to results of a recent study where participants judged whether a tone was delayed relative to the beat consisting of previous tones (Arnal et al., 2014). This study found that higher beta power in auditory cortex indicated more precise temporal judgments. At the same time, lower beta power in motor cortex also indexed better task performance. Our study confirmed that temporal expectation and temporal attention lead to modulations of beta power in auditory cortex. We additionally show that the joint effect of expectation and attention is an interacting one, with expectation decreasing beta power in the absence of attention, but not modulating it in its presence.

Predictive coding models posit that unexpected events lead to more neural activity than expected ones (Rao and Ballard, 1999; Lee and Mumford, 2003). These models
suggest that the brain acts as a probabilistic inference machine (Friston, 2010), continuously forming predictions about future input. Numerous recent studies have provided evidence for this account, in auditory processing (Martikainen et al., 2005; Lange, 2009; Todorovic and de Lange, 2012; Wacongne et al., 2012; Nazimek et al., 2013; Schwartze et al., 2013), as well as visual processing (Den Ouden et al., 2009a; Fischer et al., 2012; Kok et al., 2012a) and somatosensory processing (van Ede et al., 2010; Valentini et al., 2011b). In our paradigm, the occurrence of the first tone was temporally unpredictable, but once it was perceived, an expectation could be formed about the onset time of the second tone. We observed an attenuation of beta power preceding and following tones with a more predictable onset time. Lower beta power to the occurrence of the second tone when expectation was focal (i.e. more predictable) could thus be indicative of reduced prediction error. Interestingly, we only evidenced a decrease in beta power prior to focally expected tones when attention was drawn away from the tones. Once attention was already directed towards the time window when the second tone was about to appear, temporal expectation did not additionally modulate neural activity. This finding would not directly follow from predictive coding models, but might be in line with the suggestion that the role of expectation is to monitor contingencies in the world and to direct attention towards unexpected events (Clark, 2013; Vetter and Newen, 2014).

At first sight, the predictive coding account, which posits a decrease in neural activity to expected events, and the attentional cueing account, which posits an increase in neural activity to expected/attended events, appear to be at odds with each other. Predictive coding models (Rao and Ballard, 1999) assume the existence of two types of neural units: those that form a sensory representation of the stimulus (representational neurons), and those that compute the mismatch between the incoming representation and the predicted one (error neurons). The larger response to surprising events stems from larger error neuron activity to the mismatch, which leads to a larger adjustment of the sensory representation. It has been shown, however, that predictive coding models can also successfully simulate attentional enhancement in Posner-type paradigms, by modulating the gain of the representation units (Spratling, 2008a, 2010) or error units (Feldman and Friston, 2010) in the presence of attention. This has led to the suggestion that the role of attention is to weight the precision of prediction errors (Feldman and Friston, 2010). However, given that the Posner paradigm compares expected, attended events to unexpected, unattended ones, applying it to these models does not provide insight into the potentially separable contributions of expectation and attention. By unconfounding these factors, our study lends credibility to the idea that expectation and attention interact in sensory processing.
In summary, we provide evidence that expectation might facilitate the withdrawing of resources from events by preparing to attenuate neural activity prior to their onset, but at the same time does not modulate pre-stimulus processing during attended time windows. This finding is of importance to studies on sensory processing which manipulate expectation and attention, whether jointly or in isolation.

Acknowledgements: We would like to thank Peter Kok from the Donders Institute for Brain, Cognition and Behaviour, for his helpful comments.
Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields

This chapter is based on:
Abstract

Repetition of a stimulus, as well as valid expectation that a stimulus will occur, both attenuate the neural response to it. These effects, repetition suppression and expectation suppression, are typically confounded in paradigms in which the non-repeated stimulus is also relatively rare (e.g., in oddball blocks of mismatch negativity paradigms, or repetition suppression paradigms with multiple repetitions before an alternation). However, recent hierarchical models of sensory processing inspire the hypothesis that the two might be separable in time, with repetition suppression occurring earlier, as a consequence of local transition probabilities, and suppression by expectation occurring later, as a consequence of learnt statistical regularities. Here we test this hypothesis in an auditory experiment by orthogonally manipulating stimulus repetition and stimulus expectation, and measuring the neural response over time in human subjects, using magnetoencephalography (MEG). We found that stimulus repetition (but not stimulus expectation) attenuates the early auditory response (40-60 ms), while stimulus expectation (but not stimulus repetition) attenuates the subsequent, intermediate stage of auditory processing (100-200 ms). These findings are well in line with hierarchical predictive coding models, which posit sequential stages of prediction error resolution, contingent on the level at which the hypothesis is generated.
Introduction

Repeating a stimulus leads to reduced neural activity, a phenomenon known as repetition suppression (RS). RS is robustly present in many signatures of neural activity. It has been observed in different sensory modalities, in humans and animals, awake or asleep, and it is visible in single-unit recordings, EEG/MEG data, as well as hemodynamic signals obtained with fMRI (for a review, see Grill-Spector et al., 2006). Similarly, valid expectation of a stimulus also leads to reduced neural activity (Summerfield et al., 2008; Alink et al., 2010; Den Ouden et al., 2010; Todorovic et al., 2011; Kok et al., 2012a), a phenomenon that we will refer to as expectation suppression (ES).

Given the repetitive and predictable nature of RS paradigms, it has been suggested that neuronal suppression to repeated stimuli may in fact be partly driven by ES (Summerfield et al., 2008). RS appears particularly present when stimulus repetitions are expected, and attenuated when they are infrequent (Summerfield et al., 2008; Todorovic et al., 2011). Conversely, other studies have found that RS can exist in the absence of ES (Kaliukhovich and Vogels, 2011), suggesting potentially separable neural signatures of these effects. Crucially, in many paradigms it is difficult to tease RS and ES apart. For example, in some mismatch negativity (MMN) paradigms, recurrent standard tones are interspersed with sporadic deviant tones, precluding a conclusive distinction between the effects of repetition and expectation on the sensory response (Friston, 2005; Garrido et al., 2007; Kiebel et al., 2008; Wacongne et al., 2012). Other MMN paradigms (Schröger and Wolff, 1996; Grimm et al., 2011) did demonstrate ES when controlling for RS, but they did not directly test if and when these factors interact on the sensory response. Similarly, repetition suppression paradigms often involve series of predictable tone repetitions (Rosburg et al., 2004), rendering ES a viable alternative explanation for the observed RS.

Recent theoretical models of sensory processing inspire the hypothesis that RS and ES may be manifestations of prediction errors on different time scales (Friston, 2005; Kiebel et al., 2008; Garrido et al., 2009; Wacongne et al., 2012). According to these models, the cortex generates predictions at different, hierarchically nested, levels. A first, low-level expectation based on local transition probabilities may generate larger neuronal responses when a stimulus is different from its directly preceding neighbour (Hosoya et al., 2005; Wacongne et al., 2011). This may generate RS, independently from ES, during the first stages of auditory processing. Higher-order expectations, based on more complex statistical regularities within the environment, may then shape subsequent neural processing to the sound, giving rise to ES. Here we empirically test these predictions, by orthogonally manipulating stimulus repetition and stimulus expectation in an auditory cue-target paradigm and measuring the neural response to
the auditory stimulus using magnetoencephalography (MEG). We presented participants with pairs of tones, consisting either of a pitch repetition or of a pitch alternation (allowing us to investigate RS), which could either be expected or unexpected (allowing us to investigate ES). In line with the models described above, we observed RS (but no ES) in early (40-60 ms following tone onset) auditory processing, while we observed ES (but no RS) during an intermediate (100-200 ms) stage of auditory processing. This provides compelling support for a hierarchy of predictions in auditory processing.

Materials and Methods

Participants. Twenty four healthy participants took part in the experiment. Two participants were excluded because of a technical error with the equipment, and two participants were excluded because of excessive measurement noise. The analyses were performed on the remaining 20 participants (9 female, age 24.5 ± 4.47 years, mean ± SD). All participants had normal hearing and no history of neurological or psychiatric disorders. The study was approved by the regional ethics committee, and a written informed consent was obtained from the participants according to the Declaration of Helsinki.

Stimuli and experimental design. Participants were presented with pairs of pure tones (20 ms, ~75dB). Tone pitch could be either low (1046 Hz), medium (1174 Hz) or high (1318 Hz). The tones were presented using a PC running Presentation software (Neurobehavioral Systems), binaurally via MEG-compatible air tubes.

Each trial started with the presentation of a small central fixation cross on the screen for 2–4 s). On each trial, a pair of tones, separated by 500 ms, was presented. The two tones in a pair could be identical (repetition trials) or different ( alternation trials), allowing us to investigate the neural consequences of repetition. Orthogonally to this, the first tone predicted the repetition or alternation of the second tone with a probability of 0.75. Therefore, the second tone could be validly predicted by the first tone (expected trials) or not (unexpected trials), allowing us to investigate the neural consequence of stimulus expectation. There were also omission trials, where the first tone was followed by a silence, which could either be expected or unexpected (Table 1). The tone pair (or single tone) was followed by a short period during which the fixation cross remained on the screen (0.5–1 s), and a rest period in which the participants could freely move their eyes and blink (1.5–2 s), resulting in a 4–6 s interval between the last tone (or tone omission) of the current trial and the first tone of the following trial. The first tone of each pair was always unpredictable in onset time and
Once the first tone was heard, a prediction could be formed about the onset and pitch of the second tone.

**Table 1** Transitional probability diagram

<table>
<thead>
<tr>
<th>First tone</th>
<th>Second tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral (p = 0.33)</td>
<td>Expected (p = 0.75)</td>
</tr>
<tr>
<td></td>
<td>Unexpected (p = 0.25)</td>
</tr>
<tr>
<td>Stimulus 1 (N = 200)</td>
<td>Repetition (stimulus 1) (N = 150)</td>
</tr>
<tr>
<td>Stimulus 2 (N = 200)</td>
<td>Omission (no stimulus) (N = 150)</td>
</tr>
<tr>
<td>Stimulus 3 (N = 200)</td>
<td>Alternation (stimulus 2) (N = 150)</td>
</tr>
</tbody>
</table>

The pitch of the first tone could be one of three stimuli with equal probability. The second tone followed the first tone with a probability of 0.75 (expected trials) or 0.25 (unexpected trials), and was of the same pitch (repetition trials), of a different pitch (alternation trials) or omitted (omission trials).

Alternation trials always began with high tones, while repetition and omission trials could begin either with a low or medium tone (counterbalanced across subjects). The statistical regularities between the first tone and the frequency/omission of the second tone were counterbalanced across participants. To control for overall differences in auditory processing due to pitch, we balanced pitch across trials in such a way that there was an equal number of trials beginning with low, medium and high tones, and an equal number of trials ending with low and medium tones and tone omissions. The experiment consisted of 600 trials, which were randomly intermixed and presented in four sessions of 12 minutes. The experiment contained 50 trials of each type of ‘unexpected’ tone pair, and 150 trials of each type of ‘expected’ tone pair. Before the actual experiment, we provided participants with 8 minutes of 72 practice trials, which contained the same tone contingencies as in the experiment. During the practice, participants were also explicitly notified about the tone contingencies.

On 25% of trials in each condition (‘catch trials’), participants were asked which of the three tones they heard on the first position of the tone pair. Participants received feedback by means of a brief color change of the fixation cross. This task kept participants focused, while ensuring that attention was similar between the different conditions. Participants answered by pressing one of three buttons with their right hand. To prevent eventual response preparation, the stimulus-response mapping was withheld until the response screen on catch trials.
MEG measurements. Ongoing brain activity was recorded using a whole-head MEG with 275 axial gradiometers (VSM/CTF Systems) in a magnetically shielded room. Head localization was monitored continuously during the experiment using coils that were placed at the cardinal points of the head (nasion and left and right ear canals). As an aid for eye blink and heartbeat artifact rejection, an electro-oculogram (EOG) was recorded from the supraorbital and infraorbital ridge of the left eye, and an electrocardiogram (ECG) was recorded as well, both using 10-mm-diameter Ag–AgCl surface electrodes.

MEG data analysis. The data were prepared for analyses using the FieldTrip toolbox developed at Donders Institute for Brain, Cognition, and Behavior (Oostenveld et al., 2011a) in Matlab 7 (MathWorks). Epochs of interest were checked for artifacts using a semiautomatic routine that helped detecting and rejecting trials containing muscle artifacts and jumps in the MEG signal caused by the SQUID electronics. Subsequently, independent component analysis (Bell and Sejnowski, 1995) was used to partial out any remaining variance attributable to eye blinks and heartbeat artifacts (Jung et al., 2000). The data were then visually inspected and any remaining trials with artifacts were removed manually. For the purpose of trial-by-trial analyses, the data from the head localization coils was used to partial out variance in the signal related to head motion (Stolk et al., 2013). Finally, to remove any differences in data analysis that may arise from having an unequal number of trials per condition, a random sample of trials, equal to the number of trials in the smallest condition, was taken from each condition before averaging. This resulted in an average of 46.15 trials (SD = 2.44) per condition. Data were low-pass filtered using a two-pass Butterworth filter with a filter order of 6 and a frequency cutoff of 40 Hz. We calculated a planar gradient transform in order to position the maximal signal above the sources of neural activation (Hämäläinen et al., 1993), which simplifies the interpretation of the sensor-level data (Bastiaansen and Knosche, 2000). ERPs were then baseline corrected by subtracting the mean signal in the 100 to 0 ms interval before the occurrence of each tone.

Loci of auditory activation were defined by identifying the 10 left and right hemisphere sensors that showed maximal activity in the 50–150 ms period following the tone presentation (Figure 1A). The average evoked activity of this set of sensors over time (Figure 1B) constituted the measure of auditory activation that served as the dependent variable for all subsequent analyses.

Statistical analyses. Statistical comparisons were performed using SPSS v16. Our factorial design allowed us to independently assess the neural consequences of repetition (repetition suppression: RS), and expectation (expectation suppression: ES), as well as their interaction. To test for specific differences in different stages of auditory
processing, we focused on three pre-specified time windows of tone processing: 40-60 ms (early), 100-200 ms (intermediate), and 200-500 ms post-stimulus (late). We based the choice of these three time windows on previously found effects of repetition (Rosburg, 2004; White and Yee, 2006) and expectation (Grimm et al., 2011; Todorovic et al., 2011) in the auditory domain. More specifically, for the early time window, we chose to focus on the P50 component, which peaks around 50 ms after tone onset. Since it is a brief component, we averaged over a small time window around it (+/-10 ms). For intermediate processing, we focussed on processing in the N1/ P2 complex. The beginning of the intermediate time window corresponds to the beginning of the MMN and ends with its peak. The remaining activity, in the late time window, includes all processing after the peak of the MMN (Näätänen, 1990). In each of these time windows, we investigated the effects of RS and ES on auditory activity, as well as their interaction, using a repeated measures ANOVA with the factors Repetition (repeated, alternating) and Expectation (expected, unexpected).

In order to assess whether effects of expectation were stable or developed across the experiment, we also examined trial-by-trial auditory activity. The activity corresponding to each tone was binned into blocks of 10 trials. For each participant and each condition, we estimated the slope and intercept of auditory activity as a function of time. The intercepts and slopes of the regression lines in the different conditions were compared using a paired samples t-test.

**Results**

**Behavioral results**

On a proportion of trials, participants were asked about the identity of the first tone of the tone pair. Participants performed this task with high accuracy (90% ± 8%, mean ± SD), demonstrating that they successfully attended to and processed the auditory stimuli. There were no differences in accuracy depending on whether the first tone was part of an expected or unexpected tone pair (t(19)=0.94, p=0.35). Tones that were part of an alternation tone pair were more accurately recognized than tones in a repetition pair (t(19) = 3.2, p=0.004).

**Auditory event-related activity**

Tones elicited strong neural activity over bilateral temporal sensors (Figure 1A). The second tone of the tone pairs elicited a smaller evoked field than the first tone (Figure 1B). As would be expected, there were no differences in evoked activity between the different conditions for the first tone of the tone pairs, in any of the three temporal windows of auditory processing (p>0.10 in all cases). The subsequent sections will
focus on activity differences elicited by the second tone in different stimulus pairs, as a function of the tone being a repetition/alternation and/or expected/unexpected.

Selective repetition suppression (RS) during early auditory processing
During early auditory processing (40-60 ms following stimulus onset), there was reduced neural activity for repeated, compared to non-repeated tones in the auditory evoked fields, i.e. repetition suppression (RS: F(1,19)=4.7, p=0.042), while there was no significant difference between expected and unexpected tones, i.e. no expectation
Repetition suppression precedes expectation suppression

Chapter 4

Figure 2  Effect of pitch repetition on neuronal suppression

A) Auditory evoked fields for repeated (green) and alternating (black) tones. Temporal windows are denoted on the x-axis. Significant effects are in gray. Dashed line represents tone onset. B) Average topography over time for repeated tones (top), alternating tones (middle) and their difference (bottom). Dots represent analyzed channels.
Selective expectation suppression (ES) during intermediate auditory processing

In the intermediate auditory processing interval (100-200 ms following stimulus onset), there was reduced neural activity for expected, compared to unexpected tones in the auditory evoked fields, i.e. expectation suppression (ES: F(1,19)=23.2, p<0.001, Figure 3). Interestingly, during this interval there was no difference between auditory evoked responses to repeated and non-repeated tones (RS: F(1,19)=0.12, p=0.73). Also, ES was not different in magnitude for repeated vs. non-repeated tones during intermediate processing, as indicated by the absence of an interaction (F(1,19)=2.76, p=0.11). Again, the topographical distribution of the middle-latency ES effect was similar to the topographical distribution of the activity elicited by the tones (Figure 3B).

Interacting effects of repetition and expectation during late auditory processing

In the late auditory processing interval (200-500 ms after stimulus onset), we observed both RS (F(1,19)=11.1, p=0.004,) and ES (F(1,19)=31.3, p<0.001). Also, there was a trend of ES being stronger for non-repeated, compared to repeated trials (F(1,19)=3.1, p=0.094). In this temporal window, ES was also evident following tone omissions: there was more activity following unexpected tone omissions than after expected omissions (t(19)=2.88, p=0.009). Topographical distributions of late RS and ES effects show slightly broader activity distributions, however with clear peaks over temporal sensors, overlying the topographical distribution of tone-induced neural activity (Figures 2B, and 3B).

In sum, the data show distinct, non-interacting contributions of RS and ES on neuronal suppression during early (RS) and intermediate (ES) tone processing, and overlapping (and somewhat interacting) effects during late tone processing.

We next examined the stability of ES over the course of the experiment. On the one hand, ES may develop slowly, given that the statistical regularities may be learned and strengthened over time. On the other hand, subjects received training before the start of the experiment, and the statistical regularities were of low complexity (i.e. first-order statistical regularities). To establish the time course of ES throughout the experiment, we looked at how neural activity elicited by the auditory tones evolved...
Repetition suppression precedes expectation suppression

Figure 3  Effect of pitch expectation on neuronal suppression

A) Auditory evoked fields for expected (blue) and unexpected (red) tones. Temporal windows are denoted on the x-axis. Significant effects are in gray. Dashed line represents tone onset.

B) Average topography over time for expected tones (top), unexpected tones (middle) and their difference (bottom). Dots represent analyzed channels.
over time, for expected and unexpected tones. For this analysis, we restricted the time of interest to 100-200 ms post-stimulus, i.e. the intermediate stage where ES was selectively present. We then compared the intercepts and slopes of regression lines fitted through data corresponding to expected and unexpected tones, plotted (in bins of ten trials) against time (Figure 4). We found that the difference between unexpected and expected tones was present and stable throughout the experiment, as indicated by the larger intercept for unexpected than expected tones ($t(19)=2.43, p=0.02$) and lack of difference in the slopes of these regression lines ($t(19)=0.71, p=0.47$).

**Discussion**

We presented participants with pairs of tones, of which the pitch of the second tone was either a repetition or not, and was either expected or unexpected. This allowed us to independently evaluate the effects of tone repetition (repetition suppression, RS) and tone expectation (expectation suppression, ES) on neural activity in auditory evoked fields. Our results suggest that while repetition and expectation both attenuate neural activity, the time course of their attenuation is distinct: repetition leads to an early (40-60 ms) reduction in neural activity, whereas valid expectation reduces neural activity during an intermediate stage of auditory processing (100-200 ms). After 200 ms, there were interactive effects of both factors, with larger ES for non-repeated than for repeated tones. The topographies of the effects suggest that

---

**Figure 4** Stability of expectation suppression throughout the experiment

Average evoked fields 100-200 ms after tone onset, for expected trials (blue) and unexpected trials (red). Each point represents an average of ten successive trials.
ES and RS are occurring within the same cortical regions that are processing the auditory stimuli (rather than being related to e.g. general arousal). Below we will discuss and interpret these findings within recent neuronal models of deviance detection in auditory processing (Wacongne et al., 2012).

The early attenuation to repeated tones, i.e. suppression of auditory activity in the 40-60 ms interval for the second tone relative to the first tone, is well in line with a comprehensively researched auditory phenomenon called P50 suppression (Yee and White, 2001; Rosburg et al., 2004). This phenomenon is often viewed as automatic inhibitory gating or filtering (Freedman et al., 1987). Our results show stronger activity suppression for repeated than non-repeated sounds, suggesting that the sensory filtering is sensitive to the pitch of the immediately preceding tone. This stimulus adaptation for identical tones could be the result of neuronal fatigue or sharpening of sensory representations (Grill-Spector et al., 2006). Note that they are however also compatible with hierarchical neuronal models of auditory processing (Garrido et al., 2007; Kiebel et al., 2008; Wacongne et al., 2011) in which “predictions” at the earliest stage are based on local transition probabilities only.

Within the context of these hierarchical models, the early (local) auditory prediction is followed by higher-order auditory predictions that are based on contextual regularities in the environment. Indeed, we observed that tones that were expected (on the basis of the tone transition probability structure) were marked by suppressed auditory activity in the subsequent 100-200 ms interval. Neuronal suppression at this latency has also been researched in considerable detail, and its mechanisms are hotly debated (Näätänen et al., 2005; Wacongne et al., 2011). A large body of literature has documented that when an unexpected deviant sound is introduced within a sequence of repeated frequent sounds, larger neural activity is observed within this time window: the mismatch negativity (MMN) (Näätänen et al., 2007). The MMN can be robustly observed to deviants in pitch, duration, tone onset or amplitude. This has led researchers to suggest that the MMN is driven by differences in tone features between the standard and deviant stimulus (Näätänen et al., 2007). However, as argued earlier, expectation and repetition can be confounded in oddball blocks of a number of MMN paradigms, where unexpected tones are at the same time non-repeated tones. Although other studies have provided compelling support for the MMN as resulting from an expectation violation (Schröger and Wolff, 1996), it is not clear from these studies whether the MMN is driven purely by ES, or additionally (and potentially, in interaction with) by RS. Our results provide unequivocal support for the auditory modulation to reflect ES, and not RS, in the 100-200 ms time window. ES was highly significant in this temporal window (p<0.001), while there was no significant influence of repetition on the response amplitude (p=0.73). This rules out a simple adaptation
account for the MMN (May and Tiitinen, 2010). This fits with recent theoretical frameworks that have reinterpreted the MMN as a prediction error response (Baldeweg, 2006; Garrido et al., 2009). The observed cascade of prediction error responses (early RS, followed by ES) is well in line with hierarchical predictive coding models (Wacongne et al., 2012) that posit distinct stages of auditory novelty detection, based on temporal integration mechanisms that operate on intervals of increasing length (Kiebel et al., 2008).

A recent EEG study found that even early tone processing may be influenced by expectation of tone pitch (Grimm et al., 2011), while we only see RS in this time window. There are several potential reasons for this apparent discrepancy. Firstly, the Grimm et al. (2011) study manipulated expectation by making a particular stimulus overall less likely (i.e., a base rate manipulation), whereas in our design stimuli were expected or unexpected due to their transition probabilities. Second, stimulus probabilities in the Grimm et al. (2011) study were blocked, allowing for expectation of an upcoming stimulus to form and strengthen over a period of several minutes. In contrast, in our study expectations could only be formed after hearing the first tone of each tone pair, as the prediction was contingent on the (unpredictable) first stimulus in the trial. Therefore, expectations in the study by Grimm et al. could be more automatic, which could explain why the signature of their violation may also be evident earlier in time (Wacongne et al., 2011). Finally, the Grimm et al. (2011) study had a shorter stimulus onset asynchrony than ours (293 ms vs. 500 ms). It is possible that decay times are shorter for early effects of expectation on tone processing, and that they are therefore not evident in our study.

It has been put forward that both RS and ES are influenced by, and partly depend on, the attentional state of the observer (Murray and Wojciulik, 2004b; Yi and Chun, 2005a; Larsson and Smith, 2012). The finding that RS modulates neural activity at an early, potentially pre-attentive (Kho et al., 2003; Rosburg et al., 2004; White and Yee, 2006) temporal window, whereas ES is visible in a later window where attentional effects are more prevalent (Näätänen, 1990) is consistent with the notion that RS is less dependent on attentional state than ES (Larsson and Smith, 2012). Nevertheless, although ES may depend on attentional state of the participant, the ES observed in our experiment is not likely to be a simple reflection of attentional differences between conditions for several reasons. Firstly, all analyses focused on the second tone of each tone pair, while subjects attended to the first tone of the pair. Therefore, the tones for which we observed RS and ES were task irrelevant, and as such tone contingencies did not help participants, making differences in selective attention unlikely. It is nevertheless possible that the violation of a statistical regularity generates a bottom-up attentional signal, and that ES represents attentional enhancement of
Repetition suppression precedes expectation suppression

The current results fit in a growing body of literature that show how sensory responses are subject to different modulatory effects over the time course of neural processing (Lee et al., 1998). This temporally ordered neuronal suppression can be explained by hierarchical predictive coding models (Rao and Ballard, 1999; Friston, 2005; Kiebel et al., 2008; Friston, 2009; Wacongne et al., 2011). These models posit two functionally distinct subpopulations of neurons, which encode the conditional expectations of perceptual causes, and the prediction error, respectively. Predictions at different levels of the processing hierarchy try to explain away prediction error on preceding levels, thus silencing their error neurons. Concurrently, neurons encoding sensory causes rapidly converge on the (correctly) predicted causes, yielding a relatively sharp population response. Further empirical support for this scheme is provided by a recent study that showed that expected visual stimuli lead to attenuated neural activity, yet they improve the fidelity (sharpness) of the sensory representation in the primary visual cortex (Kok et al., 2012a). In line with the idea that the formation and updating of predictions may be fundamentally related to the coding scheme of the brain, ES has been observed in numerous studies employing different sensory modalities, levels of processing and tasks. In vision, suppressed neural activity to expected, simple visual stimuli was found in area V1 (den Ouden et al., 2008; Alink et al., 2010; Kok et al., 2011), and to faces in the fusiform face area and houses in the parahippocampal place area (Den Ouden et al., 2010; Egner et al., 2010). Higher-order speech areas also show less neural activity when a syllable can be predicted from a visual cue (Arnal et al., 2011). Importantly, this may have implications for the interpretation of some experimental tasks in cognitive neuroscience, where frequency and expectancy are co-manipulated. Repetitive aspects of experimental designs may lead to expectations regarding different features of the stimulus, which, unless controlled for, may be an important factor in studies that aim to investigate stimulus-related phenomena such as repetition suppression.
Expectation and attention share a mechanism of forming sensory templates
Abstract

The sensory response to a stimulus is only partly driven by its physical properties. If the pattern of stimulation contains regularities which make it predictable (i.e. expected), neural activity will be attenuated. On the other hand, if the stimulation is particularly relevant to the observer (and therefore attended), neural activity will increase. How expectation and attention conspire to guide sensory processing is still an open question, with some studies giving precedence to expectation, others to attention, and yet others suggesting that the two are independent processes. Here we asked how human oscillatory activity recorded by magnetoencephalography differs between expected tones and tones that violate expectations, in three situations: when participants volitionally prepared for the tone that they subsequently observed, when participants prepared for a tone that was different from the one they observed, and when no preparation for a tone was present. We found evidence for interacting effects of expectation and attention in auditory areas: while performing the auditory task, unexpected tones led to more gamma band activity than expected tones, but only in the absence of volitional preparation to these tones. When participants already prepared for the tones, expectation did not modulate gamma power. Interestingly, in the absence of any preparation, increased gamma power to unexpected tones was observed. These findings suggest that expectation and attention involve overlapping mechanisms for forming sensory templates of impending stimuli.
Introduction

The sensory stimulation we process represents a complex mixture of information. Some of this information varies over time in a reliable fashion, while some of it is more volatile. Recent research provides an abundance of evidence that the sensory brain processes predictable stimulation using less neural activity (Summerfield et al., 2008; Den Ouden et al., 2009a; Alink et al., 2010; Todorovic et al., 2011; Kok et al., 2012a; Andics et al., 2013), either by ‘explaining away’ predicted events (Rao and Ballard, 1999) or by sharpening their neural signature (Lee and Mumford, 2003; Kok et al., 2012a). In both cases, the sensory response to the stimulus becomes shaped by top-down information, which, in turn, becomes shaped by the pattern of regularities in the environment (Friston, 2005). The top-down information can be thought of as a stimulus template: a prediction of the neural activity that the incoming stimulus will elicit. The heightened neural response to unpredictable stimulation is considered to be a consequence of invalidating the brain’s hypothesis of what information to expect next (Rao and Ballard, 1999; Bastos et al., 2012; Den Ouden et al., 2012). The strongest evidence that the brain forms a template of the incoming stimulus comes from studies where a stimulus was expected but omitted: not only is it possible to reconstruct the sources of activity to omitted stimuli as arising from primary sensory areas (SanMiguel et al., 2013b), but the pattern of neural activity upon omission depends on the identity of the omitted stimulus (Kok et al., 2014).

To date, little is known about the automaticity of the process of forming stimulus templates. On the one hand, paying attention to environmental regularities might sharpen predictions (Feldman and Friston, 2010; Jiang et al., 2013), leading to a larger difference between expected and unexpected stimuli (compared to the absence of attention). On the other hand, expectation might be a complementary mechanism to attention, whose main role is to track regularities in the environment in its absence, as suggested in a recent review (Vetter and Newen, 2014). The presence of attention would thus lead to a smaller difference between expected and unexpected stimuli. Notably, even though stimulus predictability can be parametrically manipulated through study design, endogenous attention is more difficult to manipulate with precision. Study participants might notice the regularities in stimulation and actively make use of them during an experiment. In some studies, participants are informed about the existing regularities in advance (Egner et al., 2010; Rahnev et al., 2011; Todorovic et al., 2011). In other studies, their attention is drawn away to an unrelated task while recording neural activity to a stream of ignored stimuli (Naatanen et al., 1993). In both cases, the amount of attention to a particular stimulus could potentially fluctuate in an unconstrained fashion throughout the course of the experiment. Any observed ‘prediction’ effects might thus be the combined outcome of automatically processing regularities, and active deployment of attention to expected stimuli.
Here, we employed a paradigm where participants volitionally prepared a stimulus template prior to tone presentation in order to perform a feature detection task. We then compared how tones of expected and unexpected frequencies were processed in a situation where participants prepared for the presented tone vs. when they prepared for a different tone. Additionally, we intermittently diverted participants’ attention to a visual task, while still presenting (expected and unexpected) tones. We found evidence for a neural activity modulation by expectation: an increased response to unexpected tones in the gamma band whenever participants did not prepare for the target tone. The strength of this effect did not depend on whether they prepared for a different tone, or for a stimulus in a different sensory modality. When they prepared for the target tone, however, expectation did not play a role in the amount of neural activity elicited. These findings indicate that expectation has effects on neural processing in the absence of attention, but has a complex interaction with attention based on feature specificity, when stimuli are task-relevant.

**Methods and results**

*Participants.* 27 healthy participants took part in the study. Four of these were not included in the analyses: three due to electronic errors during recording and one due to excessive motion artifacts (in > 30% of the trials). The neural activity of the remaining 23 participants was analyzed (14 female, mean age 23.27, SD 1.83). All participants had normal hearing and normal or corrected to normal vision, as well as no history of neurological or psychiatric disorders. The study was approved by the regional ethics committee.

*Stimuli and experimental design.* The auditory stimuli consisted of three tones (low, medium and high: 1000 Hz, 2000 Hz, and 4000 Hz). Each tone was presented for 100 ms, with 10 ms rise and fall times, via MEG-compatible ear tubes at ~70 dB SPL. Participants were seated 70 cm from the screen while listening to the tones. Simultaneously with the tones, one of three different gratings (Gabor stimuli) was presented on the screen. The gratings were tilted by 15, 75 and 135 degrees relative to a horizontal orientation. They were presented at 15 degrees of visual angle (1 degree in the center with no grating, surrounded by a ring of 6.5 degrees), with a spatial frequency of 1 cycle per degree and a constant 80% contrast.

Each trial was preceded by a preparation cue (Table 1). The identity of the cue ('1', '2', or '3') indicated which target tone frequency (low, medium or high) the participants had to look out for. At the end of the trial, participants had to indicate whether the final tone matched or mismatched the tone frequency they were preparing for. After this
cued, which was presented for 300 ms, a fixation cross appeared. Next, with the fixation cross still on the screen, a pair of tones (and concurrent ignored gratings) was presented. The inter-stimulus interval between the cue and the first tone, and the first tone and the second tone, was 1 second. The first tone never had the feature that matched the cue, and was thus never the target. The second one contained the target frequency in 50% of the trials. If the tone matched the target frequency indicated by the preparation cue (‘prepared’ trials), the participants were instructed to press a button indicating YES with the index or middle finger of their right hand. If the tone did not match this frequency (‘unprepared’ trials), they were instructed to press NO. In order to respond, they waited an additional 600 ms from the onset of the second tone, for a response screen to appear. The response screen randomly displayed ‘YES/NO’ or ‘NO/YES’, indicating whether the left or right button is mapped to their intended response. The purpose of the response delay and randomization of response mapping was to ensure that the effects of expectation and preparation on stimulus processing were not conflated with motor preparation. Participants were given feedback on their performance on each trial via a change of color of the fixation cross, which was on the screen at all times except during cue and stimulus presentation. Late responses (>1s) also led to a change of color in the fixation cross. The cue indicating the beginning of the next trial was presented at a random point between 1200 and 1700 ms after the fixation cross had changed color.

Additionally to this manipulation of preparation for target stimulus features (‘prepared’ trials, ‘unprepared’ trials), and orthogonally to it, we manipulated stimulus expectation. The frequency of the first tone in each trial was predictive of the frequency of the second tone with 75% accuracy (expected trials) and 25% violations (unexpected trials). The first tone, thus, elicited an expectation about the identity of the second tone. For example, if participants heard the medium tone first, they could expect to hear the high tone subsequently, but sometimes the low tone would be presented instead. There was an equal number of tone presentations of the low, medium and high tones in the experiment, as well as on each position (first or second tone) in the trials. Concurrently to the auditory stimuli, visual gratings were presented. Just like the auditory stimuli, the orientation of the first grating also predicted the orientation of the second grating with 75% accuracy. The tones and gratings were mutually not predictive of each other. Participants switched between doing the task on the auditory and the visual stimuli in a block-wise fashion, thus differentiating between situation where the auditory stimuli were task-relevant and should be attended, or task-irrelevant and could be unattended. During the visual blocks, the cue at the beginning of each trial was a letter (A, B or C). Cue type thus served as an additional reminder of the type of block participants were in.
In sum, when participants were performing the task on the auditory modality, they knew which tone frequency to look for after presentation of the preparation cue, but not how likely it was to be displayed. Once the expectation cue was presented, participants could form a prediction about the likelihood of the target tone’s identity. This led to four types of trials in the attended, task-relevant modality: prepared-expected and prepared-unexpected, where the target tone was correctly prepared for, and unprepared-expected and unprepared-unexpected, where participants were unprepared for the target tone. Additionally, when the task was performed on the visual modality, the stimuli in the auditory modality could be ignored-expected or ignored-unexpected.

The experiment lasted 1.5 hours, starting outside the MEG with a short introduction followed by two blocks of 48 practice trials, one in each modality. The participants then continued the experiment in the MEG, with up to 12 blocks (6 visual, 6 auditory) of 48 trials. This led to a total of 288 trials in the auditory modality when it was attended (108 prepared expected, 108 unprepared expected, 36 prepared unexpected, 36 unprepared unexpected) and 288 trials when it was unattended (216 ignored-expected, 72 ignored-unexpected). Participants could take brief breaks after each block. These blocks were interspersed with 8 additional localizer blocks of 4 minutes, where a stream of only auditory or only visual stimuli would be presented every 750 ms. The task in the localizer blocks was to press a button for each tone or grating repetition (~10%). The purpose of this task was to ensure that participants were paying attention.

### Table 1 Trial structure

<table>
<thead>
<tr>
<th>Preparation cue</th>
<th>Expectation cue</th>
<th>Expected target (p=0.75)</th>
<th>Unexpected target (p=0.25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1?</td>
<td>2</td>
<td>3 Unprepared</td>
<td>1 Prepared</td>
</tr>
<tr>
<td>2?</td>
<td>3</td>
<td>1 Unprepared</td>
<td>2 Prepared</td>
</tr>
<tr>
<td>3?</td>
<td>1</td>
<td>2 Unprepared</td>
<td>3 Prepared</td>
</tr>
<tr>
<td>1?</td>
<td>3</td>
<td>1 Prepared</td>
<td>2 Unprepared</td>
</tr>
<tr>
<td>2?</td>
<td>1</td>
<td>2 Prepared</td>
<td>3 Unprepared</td>
</tr>
<tr>
<td>3?</td>
<td>2</td>
<td>3 Prepared</td>
<td>1 Unprepared</td>
</tr>
</tbody>
</table>

A target tone was preceded by a preparation cue and an expectation cue. The preparation cue indicated which tone frequency to look out for. It was followed by two tones. The first tone (the expectation cue) contained information on whether the correct target is likely to follow. If the target indicated by the preparation cue was presented, it was a ‘prepared’ trial. If a different target was presented, it was an ‘unprepared’ trial. When the target was likely to appear based on the expectation cue, it was an ‘expected’ trial. When it was unlikely, it was an ‘unexpected’ trial.
MEG measurements and data analysis. Brain activity was recorded using a whole-head MEG with 275 axial gradiometers (VSM/CTF Systems) in a magnetically shielded room. Head localization was monitored continuously during the experiment using coils placed at the nasion and left and right ear canals. An electro-oculogram (EOG) was recorded for purposes of eye blink artifact rejection, using 10-mm-diameter Ag–AgCl surface electrodes. The electrodes were placed on the supraorbital and infraorbital ridge of the left eye.

The data were analyzed in Matlab (MathWorks) using the FieldTrip toolbox developed at The Donders Institute for Brain, Cognition, and Behavior (Oostenveld et al., 2011a). Data epochs of interest were checked for artifacts caused by muscle activity and SQUID jumps using semiautomatic routines. Independent component analysis (Bell and Sejnowski, 1995) was then applied to remove variance attributable to eye blink artifacts (Jung et al., 2000). Finally, the data were visually inspected and any remaining outlier trials were removed manually.

Time–frequency analysis. We calculated time–frequency representations (TFRs) using a Fourier transform approach applied to short sliding time windows. For the low frequencies (5-35 Hz) the length of the time window \( T \) was frequency-dependent, and consisting of four oscillation cycles per frequency (\( \Delta T = 4/f \)). A Hanning taper was applied to control for spectral leakage, resulting in an adaptive frequency resolution of \( \Delta f = 1/T = f/4 \). The time windows were advanced in steps of 50 ms. For the high frequencies (36-100 Hz in steps of 2 Hz), we used a fixed window length of 200 ms with a \( \Delta f = 20 \) Hz frequency smoothing, using multitapers (Percival and Walden, 1993).

Power estimates were first calculated separately for the horizontal and vertical component of the planar gradient and then summed (Bastiaansen and Knosche, 2000). This simplifies the interpretation of the sensor-level data because it places the maximal signal above the source (Hämäläinen et al., 1993). Raw planar gradient power estimates were converted to decibel (dB), by log-transforming with base 10 and then multiplying by 10. The power in the baseline window, centered at 500 to 400 ms before the presentation of the second tone was then subtracted. This procedure (log-transforming and subtracting a baseline) is equivalent to taking a relative baseline of raw power. We then selected the median of the planar gradient power estimates across trials within each condition. Selecting the median, rather than the mean, has the benefit of the data being less affected by noise and outliers. This is especially

\( \frac{\text{Power}}{\text{Baseline}} \)

Preparation abolishes expectation suppression
relevant for power estimations which are always positive, i.e. noise and outliers will sum rather than average out if the mean is used.

Statistical analyses. For the behavioural analyses, we used a 2×2 repeated measures ANOVA to assess whether participants had a different proportion of correct trials across the different levels of stimulus preparation and stimulus expectation in the task-relevant, attended modality. For the MEG analyses, we compared power in different frequency bands across conditions using nonparametric cluster-based permutation tests (Maris and Oostenveld, 2007). This type of test controls the type I error rate in the context of multiple comparisons by identifying clusters of significant differences over space, time, and/or frequency instead of performing a separate test for all pairs of sensors, samples, and frequency bins across conditions. All cluster-level statistics, defined as the sum of t values within each cluster, were evaluated under the permutation distribution of the maximum (minimum) cluster level statistic. This permutation distribution was approximated by drawing 5000 random permutations of the observed data. The obtained p values represent the probability under the null hypothesis (full exchangeability, i.e. no difference between the conditions) of observing a maximum cluster-level statistic that is larger than the observed cluster-level statistics, and a minimum cluster-level statistic that is smaller than the observed cluster-level statistics.

In order to examine auditory activity, we averaged over the spatial (channel) dimension on the basis of independent localization of the 10 left and 10 right channels that showed the most robust tone-related activity in the localizer trials. Localizer trials were presented in separate blocks that were interspersed with the main task. While participants had to pay attention to the tones in order to identify tone repetitions, neither attention nor expectation varied as a consequence of experimental manipulations in the localizer blocks. We took the median of all non-repeated tones in these blocks, and compared this oscillatory profile to baseline oscillatory activity. We then identified the sensors showing strongest auditory activity in these trials, which we then used for analyses of tone related activity in the experimental blocks. In the experimental blocks, we limited our statistical analyses to average power in the frequency bands where power changes were most prominent after tone onset, when averaged over the selected sensors. In the low frequencies, these were in the theta (5-7 Hz) and alpha band (8-12 Hz). In the high frequencies, gamma band activity peaked at 56 Hz. In order to get a more accurate representation of gamma activity, we averaged over a window of 10 Hz around this peak (i.e. 46-66 Hz). We then assessed whether there were temporal clusters of differential activity between the experimental conditions within a 600 ms period starting from the onset of the target tone in each trial and ending with the moment when the participants were expected to respond,
separately in each of these frequency bands. For evaluating interactions, we applied the same tests to a difference of differences between conditions.

Results

Behavioural results
We collected data on task performance from 19 of the 23 participants (behavioral data of 4 participants could not be recorded due to technical error). Participants performed the auditory task with 85% accuracy. This level of accuracy implies, on the one hand, that they were able to discriminate between the tones, but on the other hand, that the task was challenging and required attention. The participants’ performance was similar across all conditions, regardless of whether the cued tones were expected or unexpected, whether they matched the preparation cue or not, or how expectation and preparation were combined (all F<1, NS). Responses to target tones were also unaffected by whether the concurrent ignored visual grating was expected or not (t(18)=1.25, p=0.226). We did not analyze reaction times, in view of the 600 ms delay between stimulus presentation and the response screen (i.e., the task used a delayed response design).

Neural activity modulations elicited by auditory processing
We isolated sensors where auditory activity was maximal based on activity in localizer trials. In these localizer trials only tones were presented, without concurrent visual stimulation, while participants judged whether any two subsequent tones in the stream were identical. Auditory activity during localizer trials was maximal 100-300 ms after tone onset. In order to assess which sensors best captured auditory activity, we selected 10 left and 10 right channels which showed strongest power changes relative to baseline in the low frequencies (5-7 Hz; Figure 1A). This frequency band was chosen because it contains the energy of the evoked response. The selected sensors indicate bilateral activity over temporal cortex with a slightly stronger activity over the right hemisphere, which is typical of auditory MEG recordings (Shaw et al, 2013).

We then used the average power in these sensors to estimate auditory activity while participants were performing the audio-visual task (Figure 1B). In the trials where participants performed the auditory task, target tones elicited an increase in low frequency activity after tone onset in the theta (5-7 Hz) and alpha ranges (8-12 Hz), followed by an alpha desynchronization. Additionally, there was an increase in gamma synchronization in a time window of 100-300 ms after tone onset, which was maximal at 56 Hz.
Auditory activity is more prominent for attended tones

Auditory activity was more prominent during the auditory than during the visual task: when the tones were attended and task-relevant, relative to when they were not, theta power was stronger. This difference began at tone onset and lasted 250 ms ($p=0.007$). A second, shorter window was significant in the alpha band, lasting just 50 ms from tone onset ($p=0.049$), most probably originating from spectral leakage from the lower frequencies. Lower theta power to ignored tones was coupled with a more broadband gamma response, but we did not observe a significant effect of attention to modality on high frequency power ($p>0.1$).

Expectation and target preparation interact in the gamma band

We next investigated how preparation for a stimulus influenced neural activity. We compared the time-frequency representations of all trials where the participants prepared for the frequency of the tone that was subsequently presented, with all trials where they prepared for a different target tone. Both prepared and unprepared trials could be expected or unexpected, but we collapsed over stimulus expectation in this comparison in order to assess the main effect of preparation. We found that preparation for a tone influenced auditory activity in the alpha band: when the target tone was prepared for, i.e. when it matched the preparation cue, alpha desynchronization...
The effect of attention to modality on tone processing

Time-frequency representations for attended (left panel) and unattended (middle panel) auditory stimuli, as well as their difference (right panel). Black rectangular boxes in the low frequencies represent clusters of significant differences between the two attention conditions.

Figure 2  Effect of attention to modality on tone processing

was stronger than when the target did not match this cue \((p=0.018, 300-450\ \text{ms post-stimulus})\). This effect was accompanied by marginal high frequency effect: gamma band activity was stronger when the tone was not prepared, compared to the situation where it was prepared \((p=0.050)\). This was most prominent late during tone processing \((550-600\ \text{ms after tone onset})\).

We next compared how expected and unexpected tones were processed. In this comparison, we collapsed over trials with tones which the participants prepared for, and those they did not prepare for. Here, we found no difference in processing of expected and unexpected tones \((p>0.1 \text{ in all comparisons})\). However, in the gamma band, we found marginal evidence of an interaction of tone preparation and tone expectation: the difference between processing expected and unexpected tones depended on whether these tones matched the preparation cue or not \((p=0.051; 300-350\ \text{ms after tone onset})\). To qualify the interaction further, we collapsed across this time window and compared expected tones to unexpected tones, separately in trials where the target matched the preparation cue and trials where it did not match. When the subjects were preparing for the correct target tone, expectation did not play a role in how much gamma band activity was elicited by this tone \((p>0.1)\). However, when subjects were preparing for a different tone than the one that was...
Figure 3  Main effects of stimulus preparation (top) and stimulus expectation (bottom)

Top: Time-frequency representations for prepared (left panel) and unprepared (middle panel) auditory stimuli, as well as their difference, averaged over the different conditions of expectation. Black rectangular boxes in the alpha and gamma bands represent clusters of significant differences between the two preparation conditions. Bottom: Time-frequency representations for expected (left panel) and unexpected (middle panel) auditory stimuli, as well as their difference, averaged over the different conditions of preparation.
played, expectation influenced tone processing. Here, unexpected tones led to more gamma band activity than expected tones ($p=0.013$).

**Expectation modulates gamma band activity in the absence of preparation**

We next investigated the effect of expectation on tone processing when attention was drawn away, not only from a particular tone frequency which participants were preparing for, but also from the auditory towards the visual modality. In these blocks, participants prepared for visual stimuli, i.e. the task involved no preparation for any of the tones. The pairs of tones displayed in each trial (concurrently with the gratings) had a predictive relationship (such that the frequency of the first tone predicted the frequency of the second tone), regardless of whether subjects performed a task on the tones or a task on the gratings. We compared oscillatory activity in auditory sensors as a function of expectation when participants were performing the visual task. While there were no differences in tone processing in the low frequencies ($p>0.1$), we again found a modulation of expectation in the gamma band. In the absence of attention, unexpected tones led to more gamma power than expected tones ($p=0.021$; 350-450 ms after tone onset).

**Gamma modulation by expectation is abolished by tone preparation**

The previous analyses indicated that there is no difference in processing expected and unexpected tones when participants prepared for a particular tone feature, but that a difference is present when they prepared for a different feature, or when the auditory modality was unattended. We next investigated whether the size of this prediction effect depends on the experimental manipulations. If prediction depends on stimulus identity, then there should be no difference in the size of the prediction effect when participants prepared for the wrong tone feature (i.e. a different tone from the one that was presented), compared to when the auditory modality was unattended (i.e. when no tone preparation was present). Conversely, if prediction depends on overall level of attention, then the prediction effect when the auditory modality was unattended should be stronger than when it was attended but the particular tone feature was not. To formally test whether there is a reliable increase in the prediction effect in these conditions (prepared for the tone, unprepared for the tone, no preparation), we performed a one-way ANOVA on the gamma band difference between expected and unexpected tones. We collapsed over the time windows where we observed the prediction effects. In the attended modality, where the tones were task-relevant, this was 300-350 ms after tone onset (which is when we observed an interaction between prediction and preparation). When the auditory modality was ignored, we observed a prediction effect 350-450 ms after tone onset. We found evidence for the presence of a prediction effect of varying strength in these three
Figure 4  Effect of expectation on tone processing during a visual task

Time-frequency representations for expected (left panel) and unexpected (middle panel) auditory stimuli, as well as their difference (right panel). The black rectangular box in the gamma band represents a cluster of significant differences between the two expectation conditions when the auditory modality was ignored.

conditions (F(2,66)=4.33, p=0.017), but this effect did not stem from overall level of attention. Namely, post-hoc tests showed that the effect originated from a reliable difference in the prediction effect between the situation where participants prepared for the tone (where no difference between expected and unexpected tones was evident) and the other two conditions where preparation either involved a different tone (p=0.015) or where there was no tone preparation (p=0.035). There was, however, no difference in the size of the prediction effect when the auditory modality was ignored (no preparation) and when it was attended but the particular target tone frequency was not prepared for (p=0.805). This indicates that the difference in processing expected and unexpected tones hinges on the presence or absence of feature-based attention.
Discussion

In this study, we investigated the effect of expectation on auditory processing in three different situations: when participants prepared for the tone that was subsequently displayed, when participants prepared for a different tone, and when they focused their attention to a task in the visual modality, thus not preparing for any particular tone. In all three situations, participants heard tones which were either expected or unexpected, based on a preceding cue. We found that attention to the auditory modality (as opposed to the visual modality) increased the response to tones. Once the auditory modality was attended, the low frequency oscillatory activity to tones was further modulated by tone preparation, while the high frequency oscillations depended on tone preparation and tone expectation. Specifically, alpha power desynchronized more after correctly prepared (relative to incorrectly prepared) tones.
were played. In addition to this, correctly prepared tones elicited less gamma power. Finally, gamma power was lower following expected relative to unexpected tones, but only when they were unprepared. When they were prepared, i.e. when attention was focused on the particular tone frequency that was subsequently presented, expectation did not further modulate gamma power. This interaction was, however, only consistent enough to reach marginal significance. The ‘expectation effect’, a gamma increase to unexpected relative to expected tones, was also evident when participants were performing the visual task while the auditory modality was ignored. When comparing the magnitude of expectation effects across conditions, we found that, relative to the presence of feature based attention (where it was abolished), it was equally increased if tones were attended but unprepared, and if they were ignored.

It is well established that expectation of a stimulus leads to a reduction in the neural response to the stimulus (Alink et al., 2010; Den Ouden et al., 2010; Meyer and Olson, 2011; Kok et al., 2012a; Wacongne et al., 2012). This effect, coined expectation suppression, is evident for stimuli that are predictable based on the frequency of their occurrence (Schroger, 1998; Näätänen et al., 2007), but also on contextual cues (Den Ouden et al., 2010; Todorovic and de Lange, 2012), or more global statistical regularities (Summerfield et al., 2008; Wacongne et al., 2011; Garrido et al., 2013). This suggests that a change occurs in the cortical response in the presence of expectation of a stimulus, such that reactivity is lessened when the expected stimulus arrives, but not when other stimuli arrive. For such a change to occur, a stimulus-specific template must be stored in response to regularities in the stimulus environment (Wacongne et al., 2011). The incoming stimulus is then compared to this template, and if it matches the template, its sensory representation either becomes suppressed (Rao and Ballard, 1999), or sharpened (Lee and Mumford, 2003; Kok et al., 2012a), leading to an overall reduction in neural activity. However, preparation for a stimulus can also occur volitionally, in the absence of environmental regularities. Here, we asked how such volitional preparation for a stimulus might influence expectation suppression. On the one hand, preparation might enhance the effect of expectation, by increasing the difference between expected and unexpected stimuli through a sharper prediction that leads to more error when expectations are violated (Feldman and Friston, 2010). On the other hand, expectation suppression might operate only in the absence of volitional preparation, as a system for tracking surprising events while attention is deployed elsewhere (Bendixen et al., 2012; Clark, 2013; Vetter and Newen, 2014).

Our results are more in line with this second hypothesis: gamma band activity following tone presentations was not modulated by expectation when participants were preparing for the tone that got presented, but, when they were preparing for a different tone, unexpected tones led to an increase in gamma power, relative to expected ones.
Expectation and attention are both top-down processes that influence early sensory activity. Ostensibly, they exert their influences in opposite directions, with expectation reducing the neural response (Summerfield et al., 2008; Bastos et al., 2012; Den Ouden et al., 2012) and attention increasing it (Hillyard et al., 1998; McAdams and Maunsell, 1999; Treue and Martínez Trujillo, 1999). However, not only are expectation and attention often confounded in research (Posner et al., 1980), they also refer to a variety of different processes. Expectation refers to an automatic sensitivity to statistical regularities (Summerfield and Egner, 2009), but also to influences of local context (Rao and Ballard, 1999), or the presence of hypotheses of what a person might observe (Wacongne et al., 2012; Kok and de Lange, 2014). Moreover, some studies compare responses to expected stimuli with stimuli that violate expectations (Egner et al., 2010; Kok et al., 2012a; Todorovic and de Lange, 2012), while others compare expected stimuli with those for which no specific expectation exists (Hsu et al., 2014; Stokes et al., in press). Additionally, while models of predictive coding emphasize predicting the identity of an incoming stimulus (Rao and Ballard, 1999; Bastos et al., 2012) and error correction based on top-down influences from higher order cortical regions (Lee and Mumford, 2003; Bastos et al., 2012; Wacongne et al., 2012), expectation suppression has also been found in the absence of an incoming stimulus (Todorovic et al., 2011; SanMiguel et al., 2013b; Kok et al., 2014), as well as in situations that do not involve a unique higher order sensory region, such as expectations of time (Costa-Faidella et al., 2011; Lange, 2011), space (Kok et al., 2012b), or even stimulus intensity (Dean et al., 2005). Attention, similarly, involves a change in early sensory processing, but, just like expectation, can be allocated to sensory modalities (Spence and Driver, 1997), objects (Roelfsema et al., 1998), features (Maunsell and Treue, 2006), locations (Yantis and Serences, 2003), or time (Coull and Nobre, 1998). Expectation and attention, thus, both recruit a broad set of processes that become evident via attenuating or increasing sensory activity to a stimulus with its specific set of features.

In this study, an expected tone was always one that was likely to appear after a cue, while an unexpected tone was unlikely. Expectation, thus, refers to a stimulus feature, tone frequency. Attention, however, was manipulated in two ways: as stimulus preparation, which refers to keeping attention on the auditory modality but focusing it on or away from tones with a specific feature, and as a task relevant modality, which refers performing a task on auditory vs. visual stimuli. We observed a difference in how oscillatory activity was modulated by our two attentional modulations: by preparation for tones and by attention to the auditory modality. When participants focused on the auditory task (relative to the visual task), there was an increase in auditory low frequency power, most likely reflecting an increase in neural activity recruited to the displayed tones. On the other hand, within the auditory modality, there was a difference in processing of tones with an attended feature (which
participants prepared for) and of tones with an unattended feature (which they did not prepare for). Tones in all conditions were followed by an alpha suppression, but this suppression was more prominent for prepared than unprepared tones, and accompanied by suppression in the gamma band. In the visual modality, alpha suppression has been linked to attentional engagement (Jensen and Mazaheri, 2010). It is known that assessing similarity is cognitively less demanding than assessing dissimilarity (Buckingham et al., 2012). Correctly prepared tones represented a match between the cue and the target while incorrectly prepared tones were a mismatch, making it likely that the difference in alpha desynchronization, alongside a gamma desynchronization, represents a marker of this attentional difference.

Regardless of the differences in the oscillatory profiles of these two attentional effects, the effect of expectation, when present, manifested as an increase in gamma power to unexpected tones. Gamma power is thought to be linked to feedforward activity (Thomson and Bannister, 2003a; Buffalo et al., 2011; Spaak et al., 2012) in superficial cortical layers (Maier et al., 2010; Maier et al., 2011). Importantly, the effect of expectation (of a feature) hinged on the presence or absence of attention to that feature: expectation suppression was abolished with valid tone preparation. On the other hand, once attention was not on the expected or unexpected tone feature, we observed increased gamma power to unexpected tones. Although we did not observe a difference in expectation suppression between the situations where attention was on the auditory modality but on a different feature from the observed one, or on a different sensory modality, this should not be taken as evidence that expectation suppression in general is not modulated by attention to different sensory modalities. However, the lack of expectation suppression in the presence of attention to the feature that was subsequently observed indicates that tone preparation and expectation might share a mechanism of forming and maintaining a stimulus template. Once this template is in place due to tone preparation, tone expectation does not modulate it further. Tone expectation, however, leads to automatic formation of stimulus templates when attention is on a different template or different modality.

It has been argued that expectation and attention involve independent processes (Summerfield and Egner, 2009; Kok et al., 2012a; Jiang et al., 2013). The landscape of studies manipulating expectation and attention however offers ambiguous results. Several recent studies found expectation suppression in the absence of attention, but no modulation in its presence (Jones et al., 2013; Todorovic et al., in revision), one study found the converse, expectation suppression with attention but no such effect in its absence (Larsson and Smith, 2012), one found a cross-over interaction, with expectation enhancement in the presence of attention and expectation suppression in its absence (Kok et al., 2012b), and one found an enhancement for expected and
attended stimuli, but no expectation effect for unattended ones (Hsu et al., 2014). It is crucial to note, however, that expectation and attention in these studies did not always refer to the same top-down functions. In one study, expectation referred to temporal regularities, while attention was focused on stimulus identity (Todorovic et al., in revision). In another study, expectation referred to stimulus identity but attention referred to stimulus location (Jones et al., 2013). In a third study, both expectation and attention manipulated location, but attention additionally involved stimulus discrimination (Kok et al., 2012b). In a fourth study, expectation referred to the probability of stimulus repetition, while attention referred to the presence or absence of a concurrent working memory task (Larsson and Smith, 2012). In a fifth study, expectation referred to the step size between features of two subsequent stimuli, while attention referred to a combination of stimulus features, position and intensity (Hsu et al., 2014). In sum, even within studies, expectation and attention are not always focused on the same aspect of the sensory stimulation. Here we ensured that expectation and attention within the auditory modality refer to the same process: the formation of a feature-specific template of a stimulus. The attentional manipulation was in the form of a volitional preparation for the stimulus, while the expectation manipulation relied on an automatic template formed on the basis of statistical regularities throughout the experiment. We found that expectation modulates gamma power in the absence of preparation (i.e. feature based attention), but not in its presence. Our results are in line with prediction and attention relying on similar processes in feature space, which might lead to congruent or conflicting hypotheses. Namely, both the preparation and expectation cue involved preparing a stimulus template, but these templates were not always identical. On prepared-expected trials and unprepared-unexpected trials, these templates were congruent, as both prepared participants for the same stimulus. In the remaining trials, the templates were mutually incongruent, i.e. participants prepared for one stimulus but expected another. The interaction between preparation and expectation was mostly driven by a strong increase in gamma power to the unprepared-unexpected trials. Prediction error thus increased only when there was a single, sharp template which was violated by the appearance of a stimulus. If the expectation template was violated, but another, preparation template was concurrently present, such an error response was absent. In conclusion, expectation and attention refer to a variety of cognitive processes that modulate early sensory processing. We found that expectation modulates sensory processing in the absence, but not in the presence of feature based attention. This finding suggests that attention and expectation involve partially shared mechanisms.

Acknowledgements: We would like to thank Christopher Summerfield from the University of Oxford, for his contribution to developing the research paradigm, as well as his helpful comments during data analysis and interpretation.
Discussion

6
Selective changes in the firing rates of neurons are crucial for efficient information transmission. An identical stimulus can lead to a different neural response under different conditions. One such change in the neural response is adaptation: the reduction of neural activity to repeated or prolonged stimulation (Figure 1A). Adaptation has been extensively studied in the sensory domain, where it has been shown to be robustly present across many species (Grill-Spector et al., 1999; Ulanovsky et al., 2003a; Dean et al., 2005; Wark et al., 2007; Ayala and Malmierca, 2012). This reduction in responsivity is sometimes assumed to predominantly stem from changes in gating potential (Clague et al., 1997; Stemmler and Koch, 1999), which affects the refractory period of the neuron, leading to neural fatigue. The reduction of a neuron’s firing rate due to repeated stimulation is thus sometimes referred to as 
gating or refractoriness. As it occurs in the presence of repetitive stimulation, it is also called "repetition suppression." However, some forms of visual adaptation, such as colour afterimages, last longer than one would expect from changes in refractoriness (Thompson and Burr, 2009). Additionally, in auditory processing, neurons in the non-lemniscal region of the inferior colliculus (IC) of the rat show selective reductions for repeated tone frequencies, but if the frequency of the tone is changed, these neurons will recover their activity (Perez-Gonzalez et al., 2005; Antunes et al., 2010; Ayala and Malmierca, 2012). This recovery of spike rates would not be present if the neurons had stopped firing due to fatigue (Figure 1B). It has also been argued that repetition suppression to visual scenes in the human parahippocampal place area (PPA) contains an early, bottom-up phase and a distinct, late, top-down phase (Epstein et al., 2008). It is therefore increasingly understood that the reduction in the neural response in the presence of repeated stimulation is multifaceted (Krekelberg et al., 2006).

Neural adaptation influences sensory processing
Adaptation selectively influences perception. As a stimulus reaches the sensory system, it introduces a local change in sensitivity which modifies the response to subsequent stimuli. This effect is most readily available to our conscious experience in the visual domain, where adaptation can lead to afterimages. If we observe a bright red light and then look at a white surface, the result will be a greenish-blue afterimage. This effect seems directly related to neural fatigue: perception of colour relies on the differences in the output of three cones sensitive to different parts of the visible spectrum, and habituation of one of these cones tips this balance, resulting in a novel percept based on the disproportional response of the cones maximally sensitive to the shorter wavelengths of blue and green. In the auditory modality, an 'afterimage' can form by adapting to a complex tone with a spectral gap, resulting in an illusory tone that matches the frequency of that gap (Zwicker, 1964). The visual system can also adapt to the gradient of stimulation over time. Repeatedly observing a patch of light...
with gradually increasing luminance results in an afterimage which appears to gradually dim (Anstis, 1967), a finding that is mirrored in auditory adaptation to ramped sound volume (Reinhardt-Rutland and Anstis, 1982). After prolonged observation of a curved line, a straight line will appear curved in the opposite direction (Gibson, 1933), and after observing a moving stimulus, a stationary stimulus will appear to move in the opposite direction (Tootell et al., 1995; Anstis et al., 1998). A small sphere feels even smaller if a large sphere was grasped before (Kappers and Bergmann Tiest, 2013), and...
after touching a curved, convex surface, a flat surface appears concave to the touch (Vogels et al., 1996). The amount of perceptual modification by adaptation is a function of the difference between the current and the preceding stimulus. If a visual grating is observed for a period of time, the ability to judge the tilt of subsequent, similarly oriented lines will decrease, but this effect will systematically weaken for lines tilted further away from the first grating (Ware and Mitchell, 1974), while the amount of neural activity to such orientations will increase (Fang et al., 2005). Similarly, discrimination of oddball tones (in a task where frequent standard tone volumes are interspersed with rare tones of a deviant volume, or oddballs) improves for oddballs that are more distant from the standard tone volume (Simpson et al., 2014). These findings point to selective, stimulus-specific changes in neural responsivity during perception of repetitive stimulation.

The study of neural adaptation is of practical value to neuroscience. If the reduction in neural activity to repeated stimulation is greatest for those neurons that encode a stimulus (or stimulus feature), then the difference in the neural signature between its first and second presentation should be a good indicator of the location where that stimulus is encoded. For this reason, fMRI adaptation paradigms are believed to serve well for localizing both simple and complex neural processing (see Figure 1C for an example). Such paradigms have been used for uncovering the structures involved in, among other functions, phonological processing (Gold et al., 2005), processing of gaze orientation (Jenkins et al., 2006), numbers (Piazza et al., 2007; Roggeman et al., 2011), syntax (Devauchelle et al., 2008), written words (Dehaene and Cohen, 2011) and object identity (Andresen et al., 2009). However, if the reduction in neural activity to repeated stimulation is multifaceted, then it is important to tease apart the different contributions to adapted responses, in order to confirm (or disconfirm) that adaptation can be used in a straightforward manner for localizing different cognitive processes.

The purpose of this thesis is to discuss the influence of top-down processing on neural adaptation. One such influence is stimulus expectation, which I operationalize as the likelihood of a stimulus (feature) to occur under different experimental conditions. Another such influence is attention to a (likely or unlikely) stimulus.

Neural adaptation depends on stimulus novelty
An interesting feature of sensory processing is that repeated stimulation does not always lead to a decrease in the neural response. A repeated, clearly visible image will result in response attenuation in the fusiform gyri, but a degraded image can result in an increased response, if the object is recognized by the observer upon repetition (Dolan et al., 1997). Repetition of an unfamiliar face or symbol can lead to an increased response in right fusiform cortex, while repetition of a familiar one causes decreased
responsivity in the same area (Henson et al., 2000). Observing a stream of scenes where some of them repeat, leads to a complex modulation of PPA activity: initially, the BOLD amplitude lowers upon the first repetition of a scene, but, relative to the first repetition, subsequent repetitions involve a gradually increasing response, followed by a gradually decreasing response (Müller et al., 2013) (Fig 2A). In terms of cortical location, repetition enhancement is argued to be as widespread a phenomenon as repetition suppression, at least when measured with fMRI (Segaert et al., 2013) (Figure 2B).

Population activity measures are often inaccurate reflections of single neuron activity, as stimulus repetitions might involve additional processing, potentially by different neurons, not present at the time of the initial stimulus display. A recent single neuron study however shows that even individual neurons can increase their spike rate to stimulus repetitions. Although truncated, rapidly alternating images lead to a stronger response in monkey inferotemporal neurons if familiar than if novel, a finding that was replicated with human visual evoked responses (Meyer et al., 2014). Overall, the response

![Figure 2](image-url)

**Figure 2** Neural enhancement in the presence of stimulus repetitions

A) In a rapid visual sequence of noisy images of places, where some images were presented once (novel) and some were presented a total of 10 times during the experiment (rep1, 2… rep9), PPA shows a BOLD response that attenuates for the first stimulus repetition, but then increases until the fifth repetition, after which it decreases. Adapted from (Müller et al., 2013). B) Locations where fMRI BOLD repetition enhancement effects were observed across 29 different studies. Adapted from (Segaert et al., 2013).
to a repeated stimulus can increase in the situation when the stimulus is novel, degraded, or brief, where recognition of the stimulus increases with successive presentations. Such findings accentuate the notion that stimulus repetition can lead to a gain increase or decrease, depending on the state of the sensory system into which the repeated stimulus arrives (Cheadle et al., 2014). This state is largely driven by whether the stimulus was recently perceived, but it can also be influenced by conceptual experience.

**Contributions of this thesis**

The early sensory response to a repeated stimulus is modulated by its likelihood

The first question I ask in this thesis is whether neural adaptation to a repeated stimulus might be modulated by stimulus predictability. This question is important because stimulus repetitions and stimulus likelihood are typically confounded in research on neural adaptation. In such experimental paradigms, where stimuli are always repeated after they are presented for the first time, the difference between repeated and novel stimuli overlaps with the difference between likely and unlikely stimuli. In Chapter 2, I used a two-tone paradigm in which the second tone would be played often, in ‘expected’ blocks, or rarely, in ‘unexpected’ blocks (Figure 3C). I could then compare how the same stimulus, preceded by itself, is processed under conditions where it is likely vs. unlikely to occur (Todorovic et al., 2011). A similar study was conducted not long before, where it was shown that pairs of same faces led to a weaker BOLD response in the fusiform face area (FFA) than pairs of different faces, but that this difference was greater when repetitions were more likely (Summerfield et al., 2008) (Figure 3A). The suboptimal time resolution of fMRI however made it difficult to conclude with certainty whether the effect stemmed from an early sensory difference or a late attentional difference in processing likely and unlikely repeated trials. The first motivation for conducting the study in Chapter 2 was to gain an insight into the latency of this effect, which we call expectation suppression. MEG allowed us to assess this, and to conclude that effects of stimulus predictability on sensory processing are visible already 100 milliseconds after tone onset – a latency which has previously been implicated in ‘pure’ repetition suppression (Rosburg et al., 2004).

Interestingly, an electrophysiological macaque study using a similar paradigm found prominent repetition suppression to images of fractals and natural objects in the inferior temporal cortex, but the strength of this suppression was not modulated by expectation (Kaliukhovich and Vogels, 2011). There were a number of differences between this study and the Summerfield (2008) study, such as the neuroimaging method, the studied species and the presence vs. absence of a task. Another important
difference was in stimulus type: Summerfield and colleagues used faces as their stimuli, whereas Kaliukhovich and Vogels (2011) used objects and fractals (Figure 3B). One difference in these stimulus types is the amount of experience we have with faces vs. the amount of experience monkeys have with objects and fractals. A somewhat less appreciated difference is that faces resemble each other, while objects do not. If we average all the face stimuli in an experiment, the outcome also looks like a face, while the average of objects or fractals will not converge upon a known object or fractal. For this reason, in an experimental setup such as the one used in the Summerfield et al. (2008) and Kaliukhovich and Vogels (2011) studies, it might be easier to form a sensory template that represents a neural prediction of a face than a prediction of an object. Additionally, in both these studies, blocks with more repetitions contained a smaller number of mutually dissimilar stimuli, whereas blocks with more alternations contained a larger number. In other words, the amount of variance in physical stimulation differed per block, which could have also potentially influenced the results. For example, in blocks with more variance it might be more difficult for the stimulus-specific changes in neural sensitivity to converge towards a stable stimulus template (and, as stated above, this template might only form in the presence of mutually similar stimuli). This could mean that the differences in base rates of various stimulus features were driving the expectation suppression, for example in a fatigue-based manner, as opposed to genuine probabilistic computations.

In Chapter 2 we removed the different types of stimuli entirely and thus also the potentially confounding factors of the total amount and variation of sensory templates, by using only a single tone, always identical, throughout the experiment. In other words, the predicted tone was identical to the observed tone, but what varied was the likelihood that it would be presented. Therefore, if the same stimulus were to elicit a stronger neural response in one of the experimental conditions, we could conclude with certainty that the difference can be ascribed to the likelihood of the stimulus’ appearance. This is indeed what we observed: this one, repeated stimulus, led to decisively stronger neural activity when its repetition was statistically unlikely than when it was likely. This finding is of importance not only to studies of repetition suppression, but also to studies of predictive processing in general. Most such studies investigate the situation where the occurrence of an unexpected stimulus involves an expectation of a different stimulus (expect stimulus A and observe stimulus B), whereas the occurrence of an expected stimulus involves only an expectation of that stimulus (expect and observe stimulus A) (Den Ouden et al., 2009a; Kok et al., 2012a; Wacongne et al., 2012). A stronger neural response to unexpected stimuli could, therefore, stem from the sum of two concurrently present sensory patterns: one arising from the actual stimulus, and one related to the predicted stimulus. This study provided evidence for the idea that the strength, or sharpness of the prediction of a single
Figure 3  Stronger attenuation of the neural response to repeated stimuli if predictable

A) PPA BOLD response to face repetitions and face alternations. Neural activity to repeated faces is suppressed relative to alternating faces, but more strongly in blocks where repetitions are frequent (REP blocks) than infrequent (ALT blocks). B) Strong repetition suppression (red and yellow vs. light and dark blue lines) but absent expectation suppression (red and dark blue vs. yellow and light blue lines) in spike rates of macaque IT cortex to frequent repetitions of fractals (left) and natural images (right). C) Bottom: Time-frequency representation of a tone played twice (at time point 0 and repeated at time point 0.5). If expected, the same tone leads to less auditory activity than if unexpected. Top: evoked fields to tone omissions (at time point 0.5). Unexpected omissions lead to more auditory activity than expected ones. Rectangular box represents cluster where significant differences in the conditions were found.
stimulus can vary probabilistically, and that this variation influences the amount of neural activity that the incoming stimulus will elicit. This finding mirrors sound level adaptation studies, where an identical tone is presented at different volumes, where the most frequently presented volume leads to an attenuated neural response (Dean et al., 2005; Wen et al., 2009; Rabinowitz et al., 2011).

The content and nature of stimulus templates
The expectation-based change in the sensory system that leads to a neural response of modified strength, can be defined as a sensory template, or a template of expected stimulation. The incoming stimulus, it can be assumed, meets this template of expected stimulation during sensory processing. A useful question to ask is how this template is formed. One suggestion is that the brain keeps opposing predictions online, which become confirmed or disconfirmed during sensory perception, in a process similar to biased competition (Den Ouden et al., 2012; Wacongne et al., 2012). This view is compatible with the idea that overt stimulus recognition can also play a role in neuronal suppression due to stimulus expectation. While biased competition among predictions is a plausible solution in a situation with competing hypotheses, it is less obvious how to apply it to situations where there is only one template but the same stimulus (repetition) has different probabilities of occurrence under different conditions, as in Chapter 2.

One possibility is that regularities in stimulation result in a local change in synaptic weights, which could be entirely dependent on the physical characteristics of the history of stimulation and blind to the identity of the perceived objects. In such a case, expectation suppression would be prominent for objects that are mutually similar, like upright faces (Summerfield et al., 2008) or inverted ones (Grotheer et al., 2014), but absent for series of stimuli that are mutually physically dissimilar, such as natural objects (Kaliukhovich and Vogels, 2011). Another possibility is that conceptual knowledge also plays a role in expectation suppression. Two studies that investigated the modulations of repetition suppression by expectation point to the conclusion that conceptual knowledge, or experience with stimulus categories, is crucial for expectation suppression to arise. In the first study, human LOC activity to repeated vs. alternating pictures of objects was modulated by expectation (Mayrhofer et al., 2014). In the second study, stronger repetition suppression was observed for expected than unexpected letter repetitions in the human letter form area, but not when the images contained scrambled versions of the same letters (Grotheer and Kovacs, 2014). A third study, where monkeys were exposed to long training sessions of probabilistically paired images of objects, also found that unexpected object transitions led to stronger IT activity than expected ones (Meyer and Olson, 2011), further strengthening the evidence for the role of experience in expectation suppression.
A related question is whether it is possible to observe variations in the stimulus template based on varying stimulus expectations. To this end, in Chapter 2 we also compared situations where no tone was played, but where the expectation that it would be played, differed. We found that unexpected tone omissions led to a stronger auditory response than expected ones, at a latency that matched the auditory N1 peak. We replicated this finding in Chapter 4. Further studies have since confirmed that stimulus omissions arise in sensory areas that correspond to where the actual stimuli are processed (SanMiguel et al., 2013b), and that these templates consist of a pattern of neural activity that is comparable to responses to actual stimuli (Kok et al., 2014). This suggests the possibility that stimulus templates become activated at the moment of expected stimulus input, and that their strength varies probabilistically. In other words, it appears that the sensory brain rises to meet stimuli with an image of the input it is likely to receive, at the moment the input is expected to arrive. This opens the possibility, already much explored in auditory processing (Winkler et al., 2012; Schröger et al., 2014), that predictions of stimulus timing are co-coded with predictions of stimulus identity. The strength of this template (and not just the identity of the predicted stimulus) further modulates the sensory response to the incoming stimulus.

Adaptation to stimulus identity is dissociable from adaptation to stimulus likelihood

Sensitivity to base rate differences has been extensively researched in the auditory domain, both in comparisons of frequent tones to infrequent ones, and in comparisons of frequent tone volumes to infrequent ones. The standard finding in these studies is that the less predictable stimulus elicits more neural activity. Adaptation to frequent, relative to rare tone frequencies (i.e. adaptation to tone repetitions relative to tone alternations), is evident in subcortical processing (Ulanovsky et al., 2003a) where some neurons in the auditory pathway cease to fire in the presence of repetitive tones. Such paradigms are also standard in mismatch negativity (MMN) research aimed at investigating cortical processing (Näätänen et al., 2007). On the one hand, this type of neural adaptation has been forwarded as evidence that feedback loops involved in predictive processing might also span subcortical areas (Chandrasekaran et al., 2014). The recovery of the spike rate of adapted neurons when presented with infrequent tones is indeed compatible with the view that tone alternations represent a violation of the prediction that the same tone will continue to be repeated (Perez-Gonzalez et al., 2005). This has led to the suggestion that the role of base rate adaptation is to encode stimulus expectation (Antunes and Malmierca, 2014). On the other hand, when feedback connections are disrupted by cooling the auditory cortex, neurons in the medial geniculate body (Antunes and Malmierca, 2014) and inferior colliculus will respond less (Anderson and Malmierca, 2013), but the difference in their probability of
firing to a frequent vs. infrequent tone will be maintained. This has led to the opposing suggestion, that adaptation to base rates of stimulation arises locally, and independently at various stages of sensory processing (Escera and Malmierca, 2014).

An important feature of such paradigms is that tone repetitions are frequent, whereas tone alternations are rare. In other words, stimulus repetition (i.e. tone identity) is conflated with stimulus likelihood, both of which are known to lead to a suppressed neural response. This makes it difficult to fully tease apart the relative contributions of neuronal fatigue and stimulus expectation. In Chapter 4, we applied a two-tone MEG paradigm where we aimed to dissociate effects of stimulus frequency and stimulus expectation. We used a design that where tone repetitions and tone alternations were presented equally often. In fact, all the different tone frequencies in the experiment were presented equally often, both on the first and on the second position of the tone pairs. What differed was the conditional probability that a given tone frequency will be presented just after another given tone frequency. Based on such a conditional rule, both tone repetitions and tone alternations could be either expected or unexpected. We found that the evoked fields corresponding to the second tone were always robustly suppressed relative to the first tone regardless of its status as a repetition or alternation, or as expected or unexpected, but within this strong suppression, isolated latencies of the early auditory response were dedicated to first processing tone identity (repeated vs. alternating) and then tone expectation (expected vs. unexpected) (Todorovic and de Lange, 2012). This result speaks to the multifaceted nature of neural adaptation, and additionally suggests that paradigms where likelihood is manipulated using conditional rules instead of base rates might give a somewhat cleaner insight into mismatch negativity.

Attention determines the extent of expectation suppression

Attention is a well-known endogenous factor that selectively increases the neural response to incoming stimuli. Many paradigms, such as the Posner task (Posner et al., 1980), involve an attentional cue that study participants can use to prepare for an incoming stimulus. In this task the cue is often probabilistic (e.g. 80% valid), meaning that prediction and attention might jointly guide sensory processing and behavioural responses. In the case of expectation suppression, attention has also been shown to play a role. While likely face repetitions presented in the visual periphery (during an easy task which is foveally presented) lead to a smaller BOLD response than unlikely repetitions, this effect disappears in the situation of high foveal attentional load (Larsson and Smith, 2012). This result parallels the finding that pure repetition suppression to peripheral houses diminishes in the PPA under similar conditions (Yi et al., 2004). Attention, thus, introduces further complexity into this already intricate balance of adaptation effects related to stimulus repetition, stimulus alternation,
stimulus recognition and stimulus expectation. I explored the effect of attention in **Chapters 3 and 5**. The results I obtained reinforced the belief, now published in a number of studies, that attention has the potential to change the selective sensitivity to expected information (Spratling, 2008b; Feldman and Friston, 2010; Kok et al., 2012b; Larsson and Smith, 2012; Jones et al., 2013). These findings have however not been uniform, ranging from the claim of necessity of attention for predictions to form (Larsson and Smith, 2012), through an interacting effect on prediction (Kok et al., 2012b), to the its lack of effect on predictions (Naatanen et al., 1993). Both these chapters argue that in the context of repetitive stimulation, where the first tone cues the onset of the second tone (which is statistically likely or unlikely to contain a certain property), the presence of attention abolishes expectation suppression. This suggests that neural adaptation stemming from expectation might not be as general a mechanism as assumed, and that it in fact might have the role of scanning the unattended environment for unusual (attention-worthy) features.

In terms of stimulus templates, the study in **Chapter 5** suggests that both expectation and attention have to reinforce the same template in order for unexpected input to lead to an increased neural response. Here, neural activity remained unperturbed by violations of expectation if attention was concurrently focussed on a different stimulus. If both expectation and attention were focussed on the same stimulus, but a different stimulus appeared, then gamma power following it increased. However, prediction is notoriously difficult to dissociate from attention (Summerfield and Egner, 2009), and, as I discuss in **Chapter 5**, most studies so far have manipulated different features of the stimulus with these two experimental factors. The design of **Chapter 3** also suffers from this limitation. In that chapter, I manipulate attention by using different tone frequencies, and prediction by using different distributions of tone onsets of a single frequency (Todorovic et al., in revision). In the context of neural adaptation, however, it is safe to assume that the endogenous factor of attention can reshape the profile of sensitivity to likely vs. unlikely information. In fact, **Chapter 3** suggests that the complex interaction of prediction and attention can begin as soon as the (prediction and attention) cue is presented, at least in the situation of temporal uncertainty of the stimulus onset.

**Functional properties of neural adaptation**

Neural adaptation is, therefore, more than just a fatigue response to the physical properties of repeated stimulation. Adaptation allows for flexibility in the rate of neuronal discharge when representing identical information over successive presentations. It is a multifaceted reduction in the neural response that is partly driven by stimulus
features, partly by stimulus likelihood, and partly by experience with the stimulus type. The modulation of adaptation by stimulus likelihood suggests the possibility that adaptation might be a functional property of the nervous system. One appealing suggestion is that the purpose of adaptation is to increase coding efficiency (Barlow, 1961; Simoncelli and Olshausen, 2001; Clifford et al., 2007; Thompson and Burr, 2009). Instead of consistently representing identical stimulation with a fixed firing rate, the sensory system might rescale its activity so as to adapt to the range and distribution of currently available information (Barlow, 1961; Clifford et al., 2007; Wen et al., 2009; Rabinowitz et al., 2011). The outcome would be that neural sensitivity to information varies with the context this information is presented in.

In addition to neural adaptation to stimulus intensity, an important source of efficient encoding relates to differential sensitivity to environmental statistics. The sensory environment is rich with regularities, which can be efficiently used to reduce the amount of processing required for efficient information transmission. Predictive coding models suggest that likely stimulation undergoes diminished processing relative to unlikely stimulation (Rao and Ballard, 1999; Lee and Mumford, 2003; Friston, 2005; Spratling, 2010; Bastos et al., 2012), through a process of comparing current input with a prediction of what input that brain area is likely to receive, fed back from higher order brain areas. The output of this comparison, in which the amount of mismatch between the prediction and input is computed, is fed forward to the next stage of sensory processing, leading to an iterative adjustment of the representation based on both the prediction and the input. Predicted input is then either suppressed (Rao and Ballard, 1999), or sharpened at the expense of unpredicted output (Lee and Mumford, 2003). Such a computation results in a higher population response to novel or unlikely stimulation, whereas likely, redundant aspects of the environment involve less processing. This principle, of comparing predicted to actual input, has been proposed to apply equally well to automatic processing of stimulus likelihoods stemming from the correlations in local statistics in the visual field (Rao and Ballard, 1999), as to likelihoods stemming from different base rates of stimulation over time (Chandrasekaran et al., 2014), stimulus distributions (Garrido et al., 2013), conditional rules linking two stimuli (Den Ouden et al., 2009a; Kok et al., 2012b; Todorovic and de Lange, 2012) or longer stimulus sequences (Alink et al., 2010), conscious, conceptual expectations (Koster-Hale and Saxe, 2013), as well as the formation of stable percepts in noisy environments (Lee and Mumford, 2003). It is therefore not unusual that predictive coding has also been suggested to underlie some portion of neural adaptation related to stimulus repetition (Summerfield et al., 2008; Todorovic et al., 2011).
Predictive coding therefore assumes that the strength of repetition suppression will depend on how likely a stimulus repetition is, a suggestion that the experiments in this thesis agree with, but only in the situation where the stimuli were not task-relevant. However, the likelihood of observing a stimulus repetition can vary based on simpler or more complex probabilistic rules. A repetition can be made likely through a difference in base rates of two stimuli (Nelken, 2004), or based on a conditional probability rule between (same vs. different) neighbouring stimuli (Todorovic and de Lange, 2012), or based on a global rule, where it is the expected outcome of a series of preceding stimuli (Wacongne et al., 2011) - or based on a combination thereof. Just like adaptation to any repeated stimulus is a complex, multifaceted process, so is the portion of adaptation stemming from stimulus likelihood. As the probabilistic structure of stimulation grows in complexity, the influence of conceptual experience, overt stimulus recognition, ability to integrate information over time, and working memory capacity may play an increasingly larger role in sensitivity to repetition likelihoods. In a hierarchical framework, simpler probabilities would be processed earlier and fed forward, while more complex probabilities would be processed later and fed back, allowing for the entire sequence of brain areas involved in sensory processing to display sensitivity to all types of stimulus likelihoods. However, on top of this hierarchical complexity of stimulus likelihoods, sensitivity to different stimulus types or features also varies across the brain, with more complex objects, such as faces, being processed at a later stage of sensory processing. These later stages may or may not be better equipped to assess more complex probabilistic structures of stimulation relative to early sensory processing.

Conclusions

Neural adaptation is a ubiquitous property of sensory processing. The response to a single stimulus varies with the neural context into which that stimulus arrives when it meets the sensory system, and this context is shaped by the history of stimulation. The sensitivity of the sensory cortex to an incoming stimulus is largely modulated by whether the same stimulus was just observed, but is also dependent on more sophisticated properties of sensory processing, such as the distribution of previous stimulus events, local transitional rules between stimuli, their temporal characteristics, global regularities, as well as endogenous states such as recognition of an object, expectation and attention. It is important to recognize and balance these factors in designs that use neural adaptation as a means of localizing different types of information processing.
References


Barlow HB (1961) Possible principles underlying the transformation of sensory messages. MIT press.


Hsu YF, Hamalainen JA, Waszak F (2014) Both attention and prediction are necessary for adaptive neuronal tuning in sensory processing. Front Hum Neurosci 8:152.


References


Todorovic A, Schoffelen JM, Van Ede F, Maris E, de Lange FP (in revision) Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band.
References


Nederlandse samenvatting

Grote delen van het brein hebben als functie om sensorische stimuli uit de omgeving te verwerken. Deze stimuli worden gedetecteerd door onze zintuigen - de ogen, oren, huid, etc., en daarna omgezet in elektrische signalen. Het is misschien verrassend, maar de intensiteit van deze elektrische signalen is niet een simpele functie van de fysische eigenschappen van de stimulus. Bijvoorbeeld, wanneer we onze aandacht richten op een bepaald aspect uit de omgeving, wordt de neurale representatie ervan versterkt, zelfs wanneer de stimulus zelf constant blijft. Aan de andere kant, zelfs wanneer we onze aandacht ergens op richten, zal een stimulus die voor een tweede keer getoond wordt een zwakkere neurale representatie hebben. Kortom, de interne staat van het brein hangt af van de voorafgaande stimulatie en van aandacht, en speelt een rol bij het vormen van een neurale respons voor elke stimulus.

Een ander aspect van de omgeving dat invloed heeft op de neurale respons is voor-spelbaarheid. Veel zaken in onze sensorische omgeving hebben een bepaalde regelmaat, welke gebruikt kan worden om sensorische informatie efficiënter te verwerken. In mijn proefschrift, heb ik onderzocht hoe het brein omgaat met deze regelmaat tijdens de verwerking van informatie in de vroege sensorische cortex. Wanneer een stimulus voorspelbaar is, wordt aangenomen dat het brein informatie over deze voorspelbaarheid kan onthouden en later weer kan gebruiken om nieuwe stimuli efficiënter te verwerken, dus met minder neurale activiteit. Echter, wat gebeurt er wanneer deze voorspelbaarheid samenvalt met aandacht? Zorgt het richten van aandacht op een voorspelbare stimulus voor een sterkere (omdat we met aandacht een hogere precisie hebben), of zwakkere verandering van neurale activiteit (omdat de stimulus voorspelt wordt door het brein)? Is de respons voor repetitieven stimuli verminderd omdat we de herhalingen kunnen voorspellen, of omdat neuronen vermoeid raken door continu dezelfde informatie te verwerken? Is de voorspelbaarheid van de identiteit van een stimulus hetzelfde als de voorspelbaarheid van het tijdstip van verschijnen, of behandelt het brein deze twee vormen van voorspelbaarheid als even behulpzaam?

De eerste vraag die ik in dit proefschrift heb gesteld is of neurale adaptatie aan een herhaalde stimulus beïnvloed wordt door de voorspelbaarheid van de stimulus. Deze vraag is belangrijk, omdat stimulus herhalingen en de verwachting van de stimulus zelf, vaak door elkaar heen gebruikt worden in voorgaande studies naar neurale adaptatie. In zulke experimentele paradigma’s, waar stimuli altijd worden herhaald nadat ze voor de eerste keer gepresenteerd zijn, overlapt het verschil tussen herhaalde en nieuwe stimuli met het verschil tussen waarschijnlijke en onwaarschijnlijke stimuli. In hoofdstuk 2 heb ik een paradigm gebruikt met twee tonen, waarbij de tweede toon vaak (‘verwachte’ blokken), of niet vaak (‘onverwachte’ blokken) werd afgespeeld.
Hierdoor kon ik vergelijken hoe een stimulus, voorafgegaan door eenzelfde stimulus, verwerkt werd onder condities waarin het wel of niet waarschijnlijk was dat er een stimulus zou worden afgespeeld. Het resultaat liet zien dat herhaalde tonen leiden tot een suppressie van neurale activiteit, maar dat deze suppressie veel minder is wanneer de herhaling onverwacht was.

Een andere vraag die ik gesteld heb is, of het mogelijk is om de relatieve bijdragen van neurale vermoeidheid en stimulus verwachting op neurale suppressie uit elkaar te halen. In hoofdstuk 4 heb ik gebruik gemaakt van een MEG paradigma met twee tonen, om de effecten van stimulus frequentie en stimulus verwachting te scheiden. Ik heb het experiment zo ontworpen dat een toon even vaak herhaald werd als dat hij gevolgd werd door een andere toon. Op basis van een conditionele regel, waren de herhalingen en veranderingen ofwel verwacht of onverwacht. Het resultaat van dit experiment was dat de sterkte van het magnetisch veld veroorzaakt door de tweede toon altijd kleiner was dan dat van de eerste toon, ongeacht of het een herhaling of verandering was, en of het verwacht of onverwacht was. Maar binnen deze sterke suppressie, lieten geïsoleerde latenties van de vroege auditieve respons zien dat eerst de identiteit van de toon wordt verwerkt (herhaling of verandering) en dan de toon verwachting (verwacht of onverwacht). De resultaat ondersteund de veelzijdige achtergrond van neurale adaptatie, en suggereert daarnaast dat paradigma’s waarin de verwachting van een stimulus wordt gemanipuleerd door conditionele regels een schoner inzicht geven in de mismatch negativiteit.

Aandacht is een bekende endogene factor welke selectief de neurale respons versterkt voor binnenkomende stimuli. Verschillende paradigma’s, zoals de Posner-taak, gebruiken een aandachtscue die proefpersonen kunnen gebruiken om zich voor te bereiden op een stimulus. In deze taak is de cue vaak probabilistisch (bijvoorbeeld in 80% van de keren juist), wat betekent dat predictie en aandacht samen de verwerking van sensorische informatie en gedrag beïnvloeden. Aandacht introduceert nog meer complexiteit in de al reeds ingewikkelde balans van adaptatie effecten gerelateerd aan stimulus repetities, stimulus veranderingen, stimulus recognitie en stimulus verwachting. Het effect van aandacht heb ik onderzocht in hoofdstukken 3 en 5. De resultaten die ik verkregen heb ondersteunen het geloof, reeds gepubliceerd in meerdere studies, dat aandacht de potentie heeft om de selectieve gevoeligheid voor verwachte informatie te veranderen. Deze uitkomsten wat betreft aandacht zijn in de literatuur echter niet uniform, variërend van de claim dat aandacht nodig is om predicties te vormen, via een interactie effect met predictie, tot het hebben van geen effect op predicties. Deze beide hoofdstukken laten zien dat in de context van repetitieve stimulatie, waar de eerste toon hint naar de start van een tweede toon (waarvan het wel of niet waarschijnlijk is dat deze een bepaalde eigenschap heeft), de aanwezigheid...
van aandacht zorgt voor een afwezigheid van de suppressie door verwachting. Dit suggereert dat neurale adaptatie veroorzaakt door verwachting geen algemeen mechanisme is zoals wordt aangenomen, en dat het mogelijk de rol heeft om de omgeving waar geen aandacht op wordt gericht te scannen op ongewone (aandachtswaardige) eigenschappen.

Kortom, neurale adaptatie is een veelvoorkomende eigenschap van de verwerking van sensorische informatie. De respons op een enkele stimulus varieert met de neurale context waarin de stimulus arriveert in het sensorische systeem, en deze context wordt gevormd door de geschiedenis van stimulatie. De sensitiviteit van de sensorische cortex voor een inkomende stimulus is grotendeels afhankelijk van het feit of eenzelfde stimulus zojuist was geobserveerd, maar is ook afhankelijk van ingewikkeldere eigenschappen van sensorische informatieverwerking, zoals de distributie van voorafgaande stimuli, lokale transitieregels tussen stimuli, de temporele eigenschappen, globale regelmaat, alsook endogene functies als herkenning van een object, verwachting en aandacht.
Acknowledgments

Acquiring a PhD signals a position of privilege. The privilege to spend time learning, exploring, pushing personal boundaries. As with any other privilege, this one doesn’t come just with hard work and a bit of luck, but also with a lifetime’s worth of support.

There is nobody as responsible for my interest in science as my parents. My mom, who nurtured confidence, who made up bedtime stories about girls who were curious and daring and loved nature. My dad, who himself is mesmerized with the beauty of human cognition. Even my sister, who used to ask me questions such as ‘how come wood is wooden, but metal things are metal?’ back when she was small and I was in high school. They always supported me when I wanted to spend just a few more years learning new things.

Floris, working with you has been an inspiration. I was completely in awe of your scientific skills back when I started. Five years on - I still am. I try to mimic your teaching approach, your friendly, inquisitive attitude towards other scientists, and your research integrity. Thank you for choosing one of the most interesting topics out there for your research, and thank you for choosing me to take part in it.

Predattors! It’s a difficult cross to bear, being the coolest group at the DCCN, but you carry that burden valiantly. Peter, you have been a true friend and a big source of support. I’m looking forward to writing that review paper with you, exchanging ideas and disagreeing on how prediction works. Elexa, Claudia, I have the best paranymphs ever. I’m glad this defense is giving us an excuse to spend more time together. Erik, my auditory MEG buddy, it was great to have someone who understood the impact of those DSO electronic errors on my day, and someone to get excited with about gamma to omitted stimuli. Jolien, I had a blast organizing that retreat with you. Yvonne, you made me feel welcome at Berkeley, my time there would have been so different if it wasn’t for you. Pim, you got me started on support vector machines, for which I am grateful. I hope our group proves to be a good place for you. Anke Marit, Loek (and Jolien, Joost, Rasim), those train rides would have never been the same without you. Matthias, I enjoyed the predatt dinners at your place, and your stories about being chased by monkeys on mountaintops. Marius, thanks for challenging my cooking skills. Sasha, you made me feel welcome on my first day at work. It was great to have an ex-Yugoslav enemy around. Christian, it was great to spend time with you at that workshop in Cambridge. I hope that tea you bought was as special as you hoped. Erik M, did you ever find out what the fox says?

Freek, Arjen, Jan-Mathijs, Eric, Robert, it was so much fun to work side by side, and have our names side by side in papers. I would never have become so immersed in the world of electrophysiology if it wasn’t for you. Heidi, you gave me the chance to transmit that knowledge further. Working with you was such a pleasure. Paul, this
MRI-challenged person thanks you for recording all those anatomical scans, to supplement my MEG data.

Jörn, roomie, it was great to share an office with you! You were always laid-back, always interested in having a chat, yet there were always multiple Matlabs open, computing something in the background. Sometimes I would trip on your helmet, or drums, or on the wires of the many external disks on the floor. But that’s OK, because you could frequently give me advice on my analysis pipeline, which more than made up for it.

Esther, Anne, thank you for taking the time to help me keep me practicing my Dutch. We were probably not the best of students, but you were awesome teachers.

Tildie, you are truly the backbone of the DCCN. From the very first days, you were personally interested, and were asking how you could help make the transition to PhD life smoother. You took an interest in my child, and in my health, and you made me feel appreciated. Your personal touch is everywhere in the institute, and it means so much.

André, there are no words. I love you.
List of publications

Peer-reviewed publications


In preparation

Solberg ØH, **Todorović** A, Lüttke CS, McQueen JM, de Lange FP. (in preparation). Predicting speech: how semantic context and visual cues modulate audiovisual speech processing.

About the author

Ana Todorović was born on February 9th 1979 in Belgrade, Serbia. There she completed a Bachelor’s degree in psychology, and a Master’s degree in psychology of individual differences. She then moved to Amsterdam, to complete a Research Master focusing on cognitive psychology and research methods. Under the supervision of Hans Phaf, she worked on the effects of emotional priming on attention to different frequency bands within the visual spectrum. In 2009, she became a PhD candidate at the Radboud University in Nijmegen. Under the supervision of Floris de Lange, she investigated effects of stimulus predictability on early auditory processing. She is interested in how changes in neural activity relate to changes in perception, and hopes to continue researching this subject in the future.
Donders Graduate School for Cognitive Neuroscience Series

32. Van Dijk, J.P. (2010). On the Number of Motor Units. Radboud University Nijmegen, Nijmegen, the Netherlands.
34. Kok, F. (2010). Word order and verb inflection in agrammatic sentence production. Radboud University Nijmegen, Nijmegen, the Netherlands.
38. Grootens, K.P. (2010). Cognitive dysfunction and effects of antipsychotics in schizophrenia and borderline personality disorder. Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.
44. Timmer, N.M. (2011). The interaction of heparan sulfate proteoglycans with the amyloid protein. Radboud University Nijmegen, Nijmegen, the Netherlands.


57. van der Linden, M.H. (2011). Experience-based cortical plasticity in object category representation. Radboud University Nijmegen, Nijmegen, the Netherlands.


61. Van Leeuwen, T.M. (2011). ‘How one can see what is not there’. Neural mechanisms of grapheme-colour synaesthesia. Radboud University Nijmegen, Nijmegen, the Netherlands.


64. Voorsmans, N. (2011). Neuromuscular features of Ehlers-Danlos syndrome and Marfan syndrome; expanding the phenotype of inherited connective tissue disorders and investigating the role of the extracellular matrix in muscle. Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


78. van Barneveld, D.C.P.B.M. (2012). Integration of exteroceptive and interoceptive cues in spatial localization. Radboud University Nijmegen, Nijmegen, the Netherlands.
86. Verhagen, L. (2012). How to grasp a ripe tomato. Utrecht University, Utrecht, the Netherlands.
89. Xiang, HD. (2012). The language networks of the brain. Radboud University Nijmegen, Nijmegen, the Netherlands.


122. Lagro, J. (2013). Cardiovascular and cerebrovascular physiological measurements in clinical practice and prognostics in geriatric patients. Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.


INVITATION
On Monday June 29, 2015, I will publicly defend my doctoral dissertation in the aula of the Raboud University, Comeniuslaan 2, Nijmegen. The defense begins at 14:30 sharp. You are most welcome for this special occasion, and for the subsequent reception.

Ana Todorović
todorovic.ana@gmail.com


Predictive adaptation in early auditory processing

Ana Todorović

Predictive adaptation in early auditory processing