

Title: The cognitive resource and foreknowledge dependence of auditory perceptual inference.

Authors: Jade D. Frost<sup>1</sup>, Katherine Haasnoot<sup>1</sup>, Kelly McDonnell<sup>1</sup>, István Winkler<sup>2</sup>, Juanita Todd<sup>1</sup>

Affiliations:

1. School of Psychology, University of Newcastle, University Drive, Callaghan, NSW, Australia, 2308.
2. Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, MTA, Budapest, Hungary

Corresponding author:

Jade D. Frost  
School of Psychology, University of Newcastle  
University Drive, Callaghan, NSW, Australia, 2308.  
Email: jade.frost@newcastle.edu.au  
Phone: +61 2 49 21 7455

## Abstract

Auditory perceptual inference engages learning of complex statistical information about the environment. Inferences assist us to simplify perception highlighting what can be predicted on the basis of prior learning (through the formation of internal “prediction” models) and what might be new, potentially necessitating an investment of resources to remodel predictions. In the present study, we tested the hypothesis that sound sequences with multiple levels of predictability may rely on cognitive resources and be cognitively penetrable to a greater extent than was previously shown by studies presenting simpler sound sequences. Auditory-evoked potentials (AEPs) were recorded from 117 participants. All participants heard the exact same sound sequence but under different conditions: 51 while watching a DVD movie and 66 while performing a cognitively demanding task. Participants were asked to ignore the sounds and focus their attention on the movie/task. However, prior to commencing the experiment we manipulated what participants knew about the sound sequence by providing explicit sequence information to 15 and 34 of the participants in the DVD and cognitive-task conditions, respectively, and no information to the others. The results demonstrated that although local pattern violations elicited distinctive AEP responses (namely, mismatch negativity), the way the amplitude of this response was modulated by sequence learning over time was dependent upon both task and explicit sequence knowledge. The implications are discussed with reference to how the division of available attention resources between the primary task and concurrent sound impacts what is learned.

### Keywords:

Auditory inference, predictive processing, mismatch negativity, attention, learning.

## 1.0 Introduction

We are incredibly adept at recognising patterns in sound and using these to infer the likely nature of subsequent experience. In fact, when tasked with detecting the emergence of sound patterning, we perform similarly to ideal Bayesian observers – detecting patterns at the earliest possible mathematical point at which the identity of the next sound can be extrapolated from prior experience (Barascud, Pearce, Griffiths, Friston, & Chait, 2016). Pattern learning is so essential to auditory system function that some elements of this process are automatic, in that they continue in altered states of consciousness like light sleep (Sabri, Labelle, Gosselin, & Campbell, 2003) and coma (Fischer, Morlet, & Giard, 2000). Here we present evidence showing that in contrast long timescale auditory pattern learning is dependent on the availability of cognitive resources.

Pattern learning enables the brain to modify its response to sounds in a way that conserves limited attentional resources. Regularities within sound sequences facilitate the emergence of internal models, stored in memory, that are reflected in modified neural responses (see Winkler, 2007 for review & Friston, 2005 for a formal account). By learning transition statistics, we can anticipate the next state of the world based on present input. Evoked brain responses to events that conform to anticipated states are progressively suppressed. While confirmatory experience strengthens the model, contradictory experience elicits larger responses signalling a departure from model predictions. These deviations prompt model updating, moving expectations in the direction of the properties exhibited by the deviant event (Winkler, Karmos, & Näätänen, 1996). The process is highly dynamic; such that if the pattern resumes then the prior model is maintained, but if the deviation continues a new model is formed (Baldeweg, 2006). These models enable a sensory relevance filter that contributes to our ability to parse the auditory scene into perceptual objects (i.e. physical source or temporal sound patterns; Winkler & Schröger, 2015), and maintain goal directed activity when the environment is predictable, but keeps us open and alerted to potentially important change (Näätänen, Kujala, & Winkler, 2011).

The Hierarchical Gaussian Filter model of learning (Mathys, Daunizeau, Friston, & Stephan, 2011) suggests that information about regularity in the environment can be extracted on multiple timescales simultaneously (for evidence from auditory pattern detection, see Sussman & Winkler, 2001). For example, imagine you are sitting in an airport lounge in the presence of a regular low pitch beeping noise which occasionally changes to a higher pitch noise, signalling that a nearby door is opening as security staff enter the area. Once you have established the meaning of these sounds, and established that no action is required, you can ignore both. You will have learned that the environment is characterised by two states, a low and a high-pitched beep. Over time you will have further established that there is a higher tendency toward the low-pitched state. As these tendencies remain constant over a longer period you will glean that volatility in this environment is low. In this setting, the internal model formed provides a reliable account of sound regularities as well as the applicable learning rates so that the energy engaged to monitor this environment will be minimal, and you can easily direct your attention toward some task such as watching the news on a nearby screen. However, if there is a sudden change in tendencies or states, such as a continuous stream of high-pitched beeps when the door is held open for a larger group of security staff to enter, your attention may be drawn to this alteration in the environment as it cannot be accounted for by the predictions based on your set of models. This change requires an update to your volatility estimates as now your experience indicates that tendencies can change significantly, and it also triggers an adjustment in learning rates to accommodate the new information. A deeper discussion of the underlying concepts of hierarchical learning can be found elsewhere (Mathys et al., 2011; Mathys et al., 2014).

A rare change in a regular acoustic pattern elicits a component of the auditory evoked potential (AEP) called mismatch negativity (MMN), with the name derived from the trigger being a mismatch between model and actual experience (Näätänen, 1992). MMN is characterised as a negative potential 100-250 ms post-stimulus recorded from scalp electrodes placed over fronto-central sites. MMN is an excellent tool with which to study internal model formation and updating, and the extraction of volatility estimates. The latter contributes to precision in the model, a value inversely related to volatility, and influential in determining the amplitude of MMN response to deviations from model predictions (Friston, 2005; Lieder, Stephan, Daunizeau, Garrido, & Friston, 2013). What is less clear however is whether MMN can be used to study higher levels of hierarchical learning. Some studies indicate that the demands of hierarchical learning would exceed the capacity of this largely automatic system as indicated in comments such as, "...the auditory prediction underlying the MMN may rely on several recent stimuli, but it only uses a limited time window and is blind to the global overall rule or pattern followed by the stimuli" (p20759, Wacongne et al., 2011). Other studies show that the time window can be reasonably long at 10-15 s (Böttcher-Gandor & Ullsperger, 1992) and even beyond 30 s with a reminder stimulus (Winkler et al., 2002). Furthermore, the number of stimuli preceding the deviant which may affect the response is approximately 10 s by some accounts (Rubin, Ulanovsky, Nelkin, & Tishby, 2016). However, the present study concerns learning influences over very long time courses, even exceeding those that are consistent with the idea that rules more global than those based on stimulus transitions can indeed modulate the MMN response (Bouwer, Werner, Knetemann, & Honing, 2016; Ladinig, Honing, Hádén, & Winkler, 2009; Sussman & Winkler, 2001).

The present study expands work indicating that although MMN elicitation follows local probabilistic rules (Costa-Faidella, Grimm, Slabu, Díaz- Santaella, & Escera, 2011; DiMittag, Takegata, & Winkler, 2016), the amplitude of this signal can indeed expose sensitivity to much longer-timescale information present in the environment in a way that might conform to the predictions of the Hierarchical Gaussian Filter (e.g., Fitzgerald, Provost, & Todd, 2018; Mullens et al., 2014; Todd, Provost, & Cooper, 2011). Sound sequences used to expose this sensitivity contain regular patterning on multiple timescales (the "multi-timescale paradigm"; Todd et al., 2011). Prior to sequence exposure participants are told that the brain responses being measured are automatically elicited and that they should try to ignore the sounds and focus attention on a movie with subtitles<sup>1</sup>. The sequences comprise two sounds and at any given time one of these sounds is highly common and the other rare. The rare sound will elicit MMN as it differs significantly in physical properties from the common sound, and therefore deviates from the internal model that reflects the high tendency toward the common tone properties. The common and rare tones alternate probabilities at regular rates, and these rates can be faster (e.g., every 0.8 minutes or 160 tone blocks) or slower (e.g., every 2.4 minutes or 480 tone blocks). Theoretically, the auditory system could learn not only about the tendencies in state, but also about this regularity in the volatilities in these tendencies. If this were the case, an impact on learning rates should be evident at key points of surprise in the sequence. This is precisely what happens.

MMN amplitude is generally expected to increment in size as estimates of precision in a model accumulate (Friston, 2005; Wacongne et al., 2011; Winkler, 2007). However, in the multi-timescale paradigm an asymmetry has been revealed in the rate with which precision accumulates when MMN amplitude is assessed at earlier versus later periods within sequence blocks (Todd et al., 2014). Precision accumulates rapidly in blocks consistent with

---

<sup>1</sup> Note, if sounds are deliberately attended MMN is typically overlapped by attention based components like N2b (Näätänen, 1992)

the common-rare tone configuration at the sequence onset, with MMN reaching its maximum amplitude in the early period and remaining stable in the later period. In contrast, MMN amplitude to deviants in blocks where tone probabilities are reversed compared to the configuration at the sequence onset is significantly smaller early in blocks than in the later period. In other words, precision accumulates more slowly in blocks that produce the first volatility violation when the former rare sound becomes common. Further evidence that this reflects learning over longer timescales comes from the observation that these differential learning rates actually invert if the rate of block alternation suddenly changes (Mullens et al., 2014). The inversion in learning rates has been attributed to the change in block alternation rate creating a higher-order model prediction-error based on estimates of block length (whether changing from shorter to longer or longer to shorter). Furthermore, these differential learning rates are abolished if the longer and shorter blocks are presented pseudo-randomly with no regular pattern (Todd, Petherbridge, Speirs, Provost, & Paton, 2017). In summary, both the sudden change in tendencies, and the sudden change in block length induce significant impacts on local level learning rates measurable in MMN amplitude.

Learning block lengths within multi-timescale sequences requires extraction of patterning that is unfolding over many minutes and far exceeds prior estimates of the memory upon which MMN is thought to be based (Böttcher-Gandor & Ullsperger, 1992; see Todd et al., 2014 for discussion). The strong effects of first learning exposed in these sequences similarly expose modulations that endure for long periods and appear to be evident not only in the first two blocks of a sequence but seem to be reactivated in later blocks (Frost et al., 2016). First impressions (also referred to as primacy effects or anchoring) are generally considered a cognitive phenomenon referring to the tendency to place higher weight on information learned early, enabling it to have an undue influence over beliefs, choices, and behaviour (Tversky & Kahneman, 1974). In the present paper we explore whether powerful first impressions and long timescale learning are more susceptible to the availability of cognitive resources in the listener.

Cognitive resources could be important for building and utilising internal models of large-scale sound organization in several ways. A cognitive resource that could be critical to these phenomena is attention. Although the sounds were never the focus of attention in previous studies, attention has been shown to impact perceptual inference processes underlying MMN by influencing how information is parsed into perceptual “objects” within an auditory scene (Sussman, 2007). We have elsewhere suggested that the first-impression bias may emerge through an initial difference in allocated “information value” to the two sounds – the initial standard as predictable and redundant, and therefore holding limited information value, and the initial deviant as unpredictable and thus having comparatively higher information value (e.g., Todd, Provost, Whitson, Cooper, & Heathcote, 2013). These categorical distinctions may require a certain level of attention to be encoded. Secondly, sequence knowledge can alter the way sounds are grouped together, which can in turn impact AEPs (Sussman, Winkler, Houtilainen, Ritter, & Naatanen, 2002). It is therefore conceivable that knowledge could interact with attention-based mechanisms to alter learning rates in these sequences. Finally, dependence on cognitive resources is indirectly implied by the brain areas proposed to be engaged during encoding of sensory input from dynamic environmental states. Learning over multiple timescales is proposed to follow a rostro-caudal gradient with sensory areas sensitive to event probabilities on a local-shorter timescale. This requires the engagement of progressively higher-order areas, and ultimately the frontal cortices, to extract patterns over the periods involved in the multi-timescale sequence (Kiebel, Daunizeau, & Friston 2008). The availability of frontal lobe resources, central to cognitive function, could therefore be important in extracting longer term information.

In this paper, we report on key manipulations of the conditions under which the multi-timescale sequence was heard. We demonstrate how learning rates in the multi-timescale sequence are altered both by reducing available cognitive resources, and by removing the linear emergence of structure through manipulating sequence knowledge. The former is demonstrated by presenting the sequences to participants while they are engaged in a concurrent task of high or low cognitive demanding. The latter is achieved by informing participants about the sequence composition and large-scale structure or omitting specific sequence foreknowledge completely.

## 2.0 Material and Methods

### 2.1 Participants

Participants were 117 adults aged 18-40 years (31 Male). All were recruited from the University of Newcastle, either as students through an online research participation site, or from the general community through advertisement. Students were offered course credit for participation and community volunteers were remunerated with a modest sum. All participants were screened for the presence of exclusion criteria which included personal mental health issues or history of psychosis in first-degree relatives, history of epilepsy or brain injury, the presence of hearing loss or excessive substance use. All participants provided written informed consent, and all procedures were conducted in accordance with those approved by Human Research Ethics committee of the University of Newcastle, Australia.

### 2.2 Sound Sequences

All participants were presented with the same sound sequences in the same order depicted in Figure 1. Sequences were identical to those used in prior published studies (e.g., Todd et al., 2013) and each comprised 1920 pure tones presented at 1000 Hz and 75 dB SPL. Within sequences 960 tones were 30 ms in duration (5-ms Hanning window rise/fall, 20-ms pedestal) and 960 were 60 ms in duration (5-ms Hanning window rise fall, 50-ms pedestal). The tones were arranged into blocks in which one sound was common and the other rare ( $p=0.875$  and  $p=0.125$ , respectively). Common occurrences are hereafter referred to as “standards” and rare occurrences as “deviants” in line with conventional nomenclature in MMN literature. Blocks began with a minimum of five instances of the standard tone and all instances of the deviant were followed by a minimum of three standard tones. Slowly changing sequences contained four blocks of 480 tones and faster changing sequences contained 12 blocks of 160 tones. Tones were presented over stereo headphones (Sennheiser HD280pro head) at a constant stimulus onset asynchrony of 300 ms. Participants heard the slow sequence first followed by a 40 sec silent break before the fast sequence commenced. Both sequences began with the 30 ms sound as a standard and 60 ms sound as deviant (greyed blocks of Figure 1).

### 2.3 Tasks

Participants in this study completed one of three tasks while the sound sequences were presented. For all tasks, stimuli were presented on a video monitor (60 cm x 34 cm) at a viewing distance of approximately 150 cm. In the *Movie Condition*, participants watched a self-selected DVD with the sound muted and subtitles displayed. In the *N-Back Condition*, participants were asked to monitor the identity of a series of visual stimuli and respond (by way of button-press on a game console response pad) only when the current stimulus was the same as that presented two trials before (hereafter referred to as 2-back task). Thirteen uppercase letters (S, W, P, V, D, B, R, X, E, C, J, K & L) subtended a vertical visual angle of

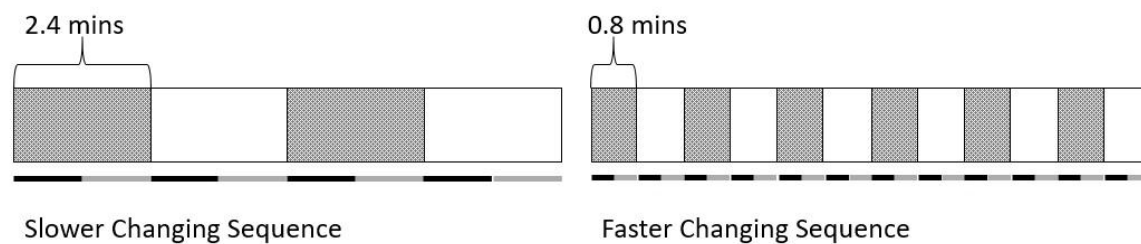


Figure 1. Graphic depiction of sound sequences. The sequence on the left changes more slowly (every 480 tones) and comprises four 2.4 min blocks. The sequence on the right changes more rapidly (every 160 tones) and comprises twelve 0.8 min blocks. Data extracted from recordings in early block periods (black bars under sequence blocks) was analysed separately from data from later in block periods (grey bars under sequence blocks).

1.0° and a horizontal visual angle of 1.2° (based on Owen et al., 2005). The letters were white in colour and presented at the centre of a black screen. A white fixation cross (vertical & horizontal visual angles of 5° & 4°, respectively) was presented at the centre of the screen between successive letters. Letters were presented for 500 ms at an average stimulus onset-to-onset interval of 2 s with onsets falling between 1.5 - 2.5 s on any given trial (uniform distribution). Scripted instructions were read prior to 10 practice trials followed by two test blocks of 600 trials separated by a 40 s break. The onset of the task preceded the tones and ran for the full duration of the auditory sequence resulting in the presentation of 158–160 targets (2-back matches) and 408–423 non-targets (2-back mismatches). The 2-back task would place a high demand on visual selective attention and working memory. The task has been found to robustly activate dorsolateral prefrontal cortex regions associated with higher-level cognitive processes (Owen et al., 2005) and has no significant impact on MMN amplitude to deviations from simple regularities (e.g., the auditory oddball paradigm with no dynamic changes in the make-up of the sequence; see e.g., Winkler et al., 2003).

In the *Visual Inspection Time (VIT) Condition* participants were asked to select (by way of button press) whether the left or right line of a visual stimulus was longer (based on Badcock, Williams, Anderson, & Jablensky, 2004). A centrally positioned fixation cue (a small white plus sign measuring 6.6 mm with vertical and horizontal visual angles of 5° and 4°, respectively) preceded all trials. Trials commenced when this cue was replaced by a figure consisting of two vertical lines (one 15 mm & the other 30mm long) joined at the top by a horizontal line of approximately 18 mm. A flash mask replaced this figure after a visual inspection time of 150 ms, and consisted of two vertical lines 35 mm in length, shaped as lightning bolts. The mask was presented for 375 ms and trial to trial onset time was 1250 ms. On each trial the participant was required to indicate which line was longer by pressing a left or right button on a game console response pad (equal probability across trials). Scripted instructions were given before 10 practice trials followed by two test blocks of 435 trials separated by a 40 s break. Participants were told to emphasize both accuracy and speed when responding. The onset of the task preceded the tones and ran for the full duration of each of the auditory sequences. The VIT was included as a task that places high demand on visual selective attention but without the high working-memory load of the 2-back task. The combination of 2-back and VIT tasks enabled us to determine whether learning sound sequence structures on shorter vs. longer timescales is impacted by shared access to both working memory and attentional load, or could be accounted for by reduced availability of attention resources alone for each task type, respectively.

## 2.4 Instruction

Prior to commencing the primary task participants were given either minimal (*uninformed condition*) or explicit (*informed condition*) information about the sound sequences. Minimal instruction involved informing the participant that they would hear a sequence of sound over headphones and we would measure the brain's response to those sounds. They were told that the brain responses we were interested in occurred automatically and were best recorded if they ignored the sounds and focused on the task. Explicit instruction involved presenting participants with an image similar to Figure 1 and explaining that they would hear a sequence that contained a short and long sound that would alternate as common and rare at different rates in different sequences. With both versions of the instructions, participants were asked to sit as still as possible and that there would be a short break in sound after about 10 minutes, during which they could stretch and move if desired.

## 2.5 Self-reported Awareness

A two-item post-experiment questionnaire was employed for all participants except those in the Movie-Uninformed condition. This questionnaire was designed to assess awareness of sound sequence structure and took approximately 5 mins to complete. Item one required a yes-no response to the question, "*Were you aware of any patterning in the sound sequences?*". If the participant answered "yes" they were requested to "*briefly describe what you noticed about the sounds*".

## 2.6 Procedure

Participants were allocated to one of six groups with no significant age or gender differences between groups (3 tasks  $\times$  2 information). Three groups were not informed about sequence structure (*Movie-Uninformed*, *N-Back-Uninformed*, *VIT-Uninformed*) and three were informed (*Movie-Informed*, *N-Back-Informed*, *VIT-Informed*; see participant group statistics in Table 1 & 2).

Different groups were tested at different time points as part of individual studies, but no individual was allowed to participate in more than one study. Data for the Movie-Uninformed group have been reported elsewhere (Frost et al., 2016) and are included here for comparison to the new manipulations. Participants were pseudo-randomly appointed to conditions by the experimenter and were not told that different groups had different levels of instruction.

Once written informed consent was obtained, the participant was fitted with a cap comprising electrodes required for obtaining vertical and horizontal electro-oculograms, a reference electrode placed on the nose, and two placed on the left and right mastoid bones. Impedances were reduced to below 5 k $\Omega$  before the recording commenced. The 2-back Informed and VIT Informed group data was obtained using a 64 channel Neuroscan quick cap with Ag/AgCl electrodes, while all other data was collected prior to this using a 32 channel Neuroscan quick cap with tin electrodes. All electroencephalographic (EEG) data was acquired using Synamps 2 Neuroscan amplifier in continuous mode with a notch filter at 50Hz, a bandpass filter of 0.1-70 Hz, and a fixed gain of 2010.

Once caps were fitted, participants were provided with the group-appropriate level of information and were given instruction on the group-appropriate task. Those completing the 2-back and VIT tasks were provided with practice trials with no concurrent sound before the main condition commenced. Headphones were then fitted and the task commenced just prior to the onset of the sound sequences.

## 2.7 Data Analysis



Task performance on the 2-back was assessed by examining hit rates and false alarm rates. Task performance on the VIT was assessed by examining the proportion of correct trials. Self-reported awareness was quantified as number reporting “yes” responses and was thereafter qualitatively assessed for clarity of response.

EEG data was analysed offline using Neuroscan software version 4.5. Data were first subjected to visual inspection to identify and reject any large artifacts in the recording and to identify any consistently bad electrode channels. The continuous EEG was divided into epochs extending from 50 ms before to 300 ms after sound onset. Epochs were baseline corrected to the prestimulus period and epochs with voltage exceeding  $\pm 70\mu\text{V}$  were rejected. Remaining epochs were averaged together separately for each sequence according to stimulus type, local probability, and period of block (the latter marked by black vs. grey bars at the bottom of the figure). Averaging therefore resulted in eight averaged AEPs to 30ms sounds and eight for the 60 ms sounds: a first-half and a second-half standard response for the slow sequence and the fast sequence, and a first-half and a second-half deviant response for the slow and the fast sequence. Eight difference waveforms were then computed for each participant by subtracting the response to each sound as a standard from that to the same sound as a deviant, for the same period, of the same type of sequence. For example, the slow 30 ms first-half period difference waveform involved subtraction of the averaged first-half 30 ms standard response in slow blocks from the averaged first-half 30 ms deviant response in slow blocks (see *Figure 1. Supplementary Material* for standard/deviant ERPs contributing to difference waveforms). There was a maximum of approximately 60 deviants in each block-half average. The average number of trials contributing to the mean deviant AEP waveforms was between 51 and 59 across all groups and conditions. When a participant had less than 45 trials in an average, data were individually inspected to confirm that the average was not excessively noisy and provided a sensible index of the response. The resultant AEPs were re-referenced to the average activity of the left and right mastoid electrodes. This process is recommended to improve signal to noise ratio given that many auditory components (including MMN) invert polarity below the mastoid region (Kujala, Tervaniemi, & Schröger, 2007).

The impact of task and instruction was explored using the MMN amplitudes extracted from difference waveforms at the F4 site where effects have been reported previously (see *Figure 2. Supplementary Material* for MMN amplitude plotted by multiple sites, demonstrating that MMN generally reached maximum amplitude at F4). MMN amplitude was quantified by an automated process in custom software that identified the most negative point between 50-250ms post-stimulus and the voltage was averaged from a window extending 10 ms on either side of the peak in each individual mean AEP response. MMN amplitude measures were entered into repeated measures ANOVAs to determine whether task and/or instruction impacted the differential learning rates previously observed in multi-timescale sequences (e.g. Frost et al., 2016; Todd et al., 2014). The data set was large and complex so to improve interpretability we engaged in some pre-analyses to facilitate data reduction. Namely, we assessed whether the nature of the cognitive task mattered to the results. Since the analyses revealed the same patterns of MMN amplitude modulation in participants performing the 2-back and VIT task conditions, data from these two conditions were grouped together in the results section to create *Cognitive Task-Uninformed* ( $n = 32$ ) and *Cognitive Task-Informed* ( $n = 34$ ) groups. As the electrodes used for recording the Cognitive Task-Informed group were different from those used for the other groups (i.e., replacement tin electrode caps were no longer available from suppliers), it could be theoretically problematic to compare data obtained from the Cognitive Task-Informed group with those obtained from the other groups. However, in light of the overall comparability of

morphology, latency and amplitude of evoked potentials in this group, we have included them in the omnibus comparison.

The data analysis was conducted to test the core hypothesis that the two manipulations would alter learning in the sequence and therefore impact the primacy bias pattern of MMN modulation (i.e., the sequence by deviant by half interaction). To assess whether the task and/or instruction manipulations impacted MMN amplitude modulations, the first analysis performed was a mixed model ANOVA with two between groups factors (Instruction – Yes/No  $\times$  Task – Movie/Cognitive Task) and within-subjects factors of sequence (Slower/Faster Changing), deviant (60/30ms) and half period (first-/second-half). Significant interactions involving the between group variables were then followed up to assess the MMN modulation patterns. Where within-group interactions were present, paired t-tests were used to assess significant changes in MMN amplitude across block half periods. A Bonferroni correction for multiple comparisons was applied such that  $\alpha$  level was adjusted to accommodate the four tests (two for the 60ms sound first to second half comparison, separately for slower and faster changing sequences and two for the corresponding comparisons for the 30ms tone: therefore  $\alpha = 0.05/4 = 0.0125$ ).

### 3.0 Results

#### 3.1 Task Performance

The hit rates and false alarm rates for 2-back task performance and the proportion of correct trials on the VIT are reported in Table 1. Hit rates on the 2-back task were high in general and false-alarm rates low indicating that participants were engaged in the primary task. There were three cases of low hit rates on the 2-back (41%, 44% and 46%) but in each case false alarm rates were also low (7%, 6% and 3%, respectively) indicating non-random responses and that these participants probably found the task difficult. Results were analysed with and without these participants and did not significantly differ. Therefore, the three participants are included in analyses below. There were also three cases of low correct trials (38%, 44% and 54%) on the VIT task that were excluded from the analysis. The hit rates and false alarm rates for the 2-back task did not differ between the informed and uninformed groups, but hit rates were significantly higher on the VIT task for the uninformed than the informed group ( $t_{30} = 2.175$ ,  $p < .05$ ). This group difference disappeared if the three outliers were removed and the mean hit rate for the informed group increased to 88%, compatible with the mean of 90% in the uninformed group. Behavioural performance was also extracted separately for the first and second half of the task trials. As there were no significant changes in performance over time these results are not presented here.

#### 3.2. Self-reported Awareness

In the Cognitive-Uninformed condition 22/32 (69%) participants (14 from the 2-back and 8 from the VIT task) indicated that they were aware of patterning within the sequences. In the Cognitive-Informed condition 30/34 (88%) participants (12 from N-Back task and 18 from the VIT task) responded “yes” to indicate an awareness of patterning within the sequence. In both cases the descriptions of what was heard ranged from inaccurate to vague. For example, one participant from the Cognitive-Uninformed group commented “*The sounds had a pattern of low frequency clicks followed by high frequency beep. Sometimes it would switch with high frequency beeps followed by one low frequency beep*”. Another from the Cognitive-Informed group stated “*The sounds were long and then short with changing tempos and they progressively got higher in pitch*” and another “*Two pitches; one high, other lower like a clock. Continuous beeping with other sound interrupting at various intervals*”. Interestingly, in the Movie-Informed condition all participants reported awareness of

Group	N-Back		N-Back		VIT	VIT
	Uninformed		Informed		Uninformed	Informed
	(n = 17)		(n = 18)		(n = 15)	(n = 16)
Measure	Hit Rate	False-Alarm	Hit Rate	False Alarm	Correct	Correct
	(%)	Rate (%)	(%)	Rate (%)	Trials (%)	Trials (%)
Mean	76	3	74	4	90	80
Range	41-92	1-7	44-94	1-17	76-98	38-98

Table 1. Performance measures on cognitive tasks.

patterning but only three could provide any reasonable description with most unable to clearly articulate sequence structures (e.g. *“I didn't notice the pattern too much. I could not tell how much time was going by either. Overall I would say that the sounds were noticeable but I was not fully aware”*). From these outcomes we assume that most participants performing the cognitive tasks, whether informed or not, had some level of awareness of regular patterning within the sequences even if they could not describe this with a high level of detail or accuracy.

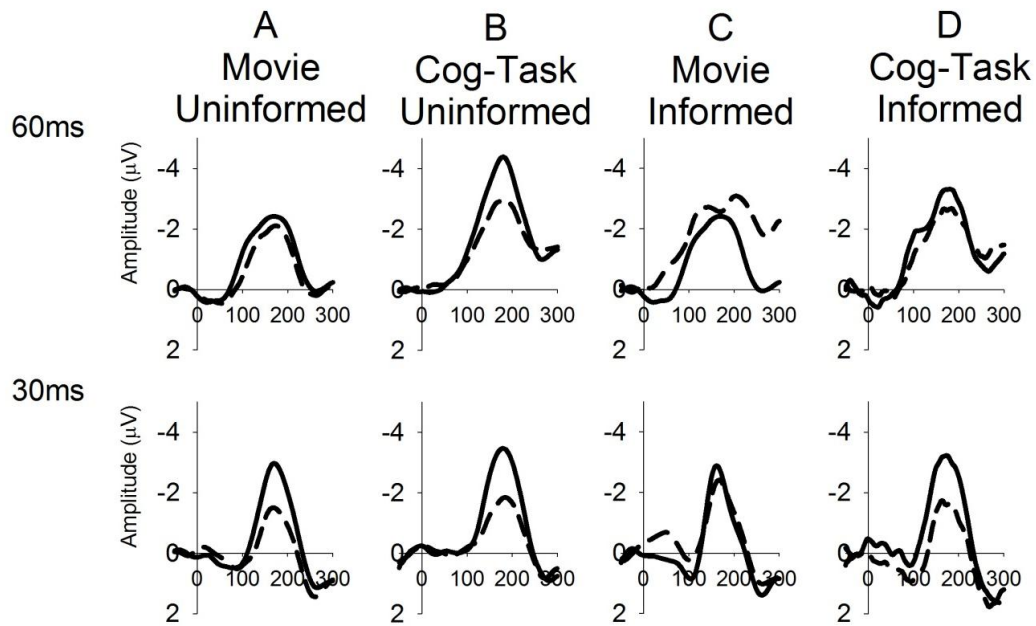
### 3.3 Omnibus Analysis of Task and Instruction

The difference waveforms produced to the two deviant tones for both sequences for each of the four groupings are presented in Figure 2 and the group-mean MMN amplitudes in Table 2. The typical primacy-bias pattern of MMN modulation is evident in the Movie-Uninformed group and there are visible differences in the modulation patterns in the other three groups. The mixed model ANOVA across all groups revealed main effects of sequence (larger in slower changing,  $F(1,110) = 49.89$ ,  $p < .001$ ,  $\eta^2 = 0.31$ ), deviant (larger overall for the 60ms tone,  $F(1,110) = 22.95$ ,  $p < .001$ ,  $\eta^2 = 0.17$ ) and half (larger in general by the end of blocks) but also confirmed a significant interaction between these three factors (interaction between sequence, deviant, and half,  $F(1,110) = 5.26$ ,  $p < .05$ ,  $\eta^2 = 0.5$ ). This interaction was not significantly modified by the influence of task collapsed over instruction, nor instruction collapsed over task. However, a significant five-way interaction between sequence, deviant, half, task, and instruction ( $F(1,110) = 6.10$ ,  $p < .05$ ,  $\eta^2 = 0.05$ ) confirmed that these factors did impact the typical primacy bias MMN modulation pattern but in different ways. In light of this result the results for each group are analysed separately to assess modulation of MMN amplitude.

### 3.3 Impact of Task and Instruction

The MMN amplitudes in the Movie-Uninformed group (Figure 2A) display main effects of sequence (slower changing larger than faster changing ( $F(1,35) = 16.67$ ,  $p < .001$ ,  $\eta^2 = 0.32$ ) and half (larger by 2<sup>nd</sup> half ( $F(1,35) = 11.16$ ,  $p < .005$ ,  $\eta^2 = 0.24$ ) but these were modified by the three way interaction between sequence, deviant, and half ( $F(1,35) = 8.53$ ,  $p < .01$ ,  $\eta^2 = 0.20$ ), a pattern found in many previous studies employing the multi-timescale paradigm (e.g., Frost, Winkler, Provost, & Todd, 2016; Todd et al., 2011; Todd et al., 2013; Todd et al., 2014.). Paired t-tests confirmed that the MMN amplitude was significantly

## 1. Slow Changing Sequence



## 2. Fast Changing Sequence

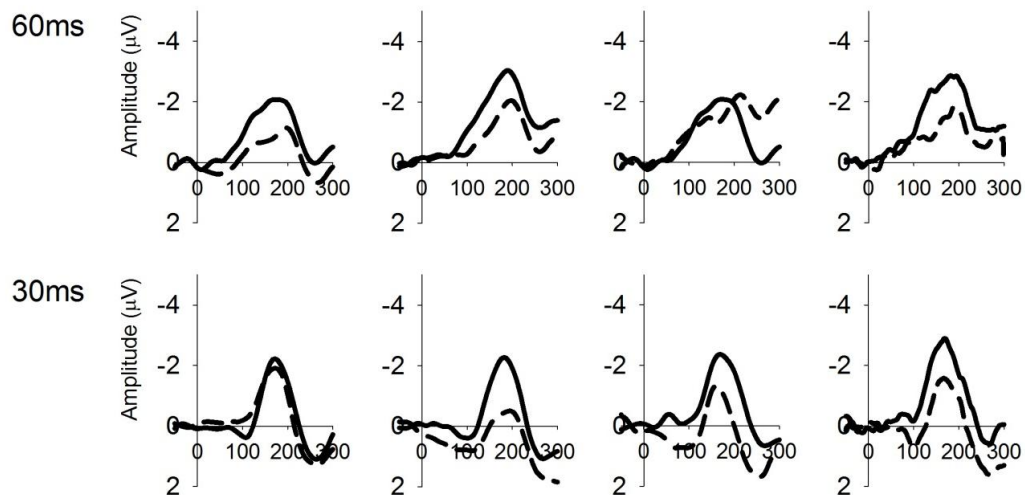


Figure 2. Group averaged deviant-minus-standard difference waveforms for first 60ms (top row) and second 30ms (bottom row) deviant type in each sequence for the first-half period within blocks (early; dotted lines) and the second-half period within blocks (late; filled lines).

smaller at early relative to late periods in the slow blocks for the 30ms tone ( $t_{35}=3.15$ ,  $p<.001$ ) but not the 60ms tone, and that these differential learning rates are inverted in the subsequent sequence – that is, significantly smaller MMN in the early relative to late periods of the fast blocks for the 60 ms tone ( $t_{35}=2.83$ ,  $p<.01$ ) but not the 30ms tone.

	Movie	Cognitive	Movie-	Cognitive
	Uninformed	Task	Informed	Task
		Uninformed		Informed
	( <i>n</i> = 36)	( <i>n</i> = 32)	( <i>n</i> = 15)	( <i>n</i> = 34)
Slower Changing Sequence				
60 First-Half	-3.22 (.62)	-3.71 (.46)	-4.13 (.42)	-3.72 (.41)
60 Second-Half	-3.58 (.45)	-4.90 (.37)	-3.38 (.37)	-3.96 (.42)
30 First-Half	-2.10 (.47)	-2.87 (.32)	-2.60 (.31)	-2.59 (.30)
30 Second-Half	-3.11 (.59)	-3.57 (.37)	-3.11 (.51)	-4.27 (.42)
Faster Changing Sequence				
60 First-Half	-1.70 (.37)	-2.48 (.32)	-3.19 (.32)	-2.04 (.01)
60 Second-Half	-2.86 (.41)	-3.94 (.31)	-3.42 (.38)	-4.11 (.36)
30 First-Half	-2.42 (.31)	-1.26 (.24)	-1.95 (.24)	-1.99 (.23)
30 Second-Half	-2.19 (.51)	-2.71 (.33)	-3.27 (.33)	-3.50 (.42)

Table 2. Group mean amplitude and standard error (parentheses) for the amplitudes extracted from deviant-minus-standard waveforms at F4. Standard errors of each mean are presented in brackets.

In contrast, differential MMN amplitude modulations for the two tones are absent for Cognitive Task-Uninformed group, who only demonstrate main effects on MMN amplitude. As is clearly evident in Figure 2B, MMN amplitudes for the Cognitive Task-Uninformed group are larger in the slow than fast changing blocks (main effect of sequence,  $F(1,31) = 42.30$ ,  $p < .001$ ,  $\eta^2 = 0.58$ ), larger later within blocks (main effect of block half,  $F(1,31) = 33.59$ ,  $p < .001$ ,  $\eta^2 = 0.52$ ), and larger for the 60 ms than the 30 ms deviant (main effect of deviant,  $F(1,31) = 21.71$ ,  $p < .001$ ,  $\eta^2 = 0.41$ ). That is, we only see the local effects expected on the basis with longer exposure and higher stability of the stimulus configuration. The results are therefore consistent with high cognitive task demand disrupting the typical differential learning rates seen in using the multi-timescale sequence under low cognitive demand.

The same repeated measures ANOVA applied to the data for the Movie-Informed group (Figure 2C) indicated a main effect of deviant only ( $F(1,14) = 6.10$ ,  $p < .05$ ,  $\eta^2 = 0.30$ ) with MMN amplitude larger for the 60ms deviant tone overall. Task instruction appears to have disrupted differential learning rates typically seen in the multi-timescale sequences. In

fact they eliminated the significant impact of sequence, block length and block-half in these data consistent with comparatively expedited learning rates in informed participants.

Finally, difference waveforms produced by the Cognitive Task-Informed group, who heard sequences while performing a cognitively demanding task, are presented in Figure 2D (see Table 2. for group mean amplitudes). Here the same repeated measures ANOVA revealed MMN to be larger overall in the slower than faster changing blocks (main effect of sequence,  $F(1,30) = 10.07$ ,  $p < .005$ ,  $\eta^2 = 0.25$ ) and larger later within blocks (main effect of block half,  $F(1,30) = 35.08$ ,  $p < .001$ ,  $\eta^2 = 0.54$ ), but these effects were modified by a significant three-way interaction between these and the deviant (30/60 ms) factors ( $F(1,30) = 6.10$ ,  $p < .05$ ,  $\eta^2 = 0.17$ ). The three-way interaction was reflected in MMN reaching its maximum amplitude in the first-half period for the 60 ms deviant in the slowly changing blocks, but being significantly smaller in the first than in the second-half periods for the 30 ms deviant ( $t_{30} = 3.91$ ,  $p < .001$ ). However, in the faster changing blocks the MMN amplitude was smaller initially and increased significantly over time within blocks for both the 60 ms ( $t_{30} = 5.74$ ,  $p < .001$ ) and the 30 ms sounds ( $t_{30} = 3.27$ ,  $p < .005$ ). Thus the three-way interaction pattern typical of the primacy bias was present for the first but not for the second sequence.

#### 4.0 Discussion

The results of the present study show that the pattern of MMN amplitude modulation observed in prior studies of the multi-timescale sequence can be altered by manipulating the availability of cognitive resources and/or prior knowledge of the sequence structure. As reviewed in the introduction, prior studies have exposed an asymmetry in the speed with which precision accumulates in blocks of the multi-timescale sequence. This asymmetry often inverts between the first and the second of the two sequences. The pattern is exemplified here by the Movie-Uninformed group for whom there is a three-way interaction between sequence, deviant and block-half. In the first sequence heard (the slower changing sequence) precision accumulates fast for the 60ms tone (the first deviant) revealed in MMN that reaches maximum amplitude in the first half of blocks. In contrast, MMN to the 30ms tone starts significantly smaller in the early portion of blocks and increases by the second half. This pattern of MMN amplitude modulation remains present if this same sequence structure with compatible block lengths is repeated (Frost et al, 2015) but is broken, often even reversing (e.g. Todd et al., 2014a), if the subsequent sequence breaks the expected block length pattern. The reversal has been attributed to a higher order prediction violation. This explanation assumes that the auditory system learns the regularity in block length and is “surprised” by the first block of the faster changing sequence being shorter than expected. The higher precision associated with the block of the original duration then drops. Mullens et al (2014) have shown this drop in precision for a block that violates block-length expectations will occur both when slower blocks violate a faster block pattern and vice versa (as per the results presented here). In the present paper we test the hypothesis that these effects will only be seen if there are sufficient cognitive resources available to support longer-term learning about the sequence.

When naïve listeners are asked to perform a concurrent cognitively demanding task, in this case a 2-back or VIT task, the modulation patterns in MMN amplitude conform to those we would expect based on local stability effects within the sequences. MMN amplitude in the Cognitive Task Uninformed group simply increases in size as the evidence about local stability increases over the block duration both in slower and faster changing sequences. The amplitude accumulates further in the slower sequence blocks than in the faster sequence blocks, resulting in a larger MMN overall in the slower changing sequence, and this pattern is the same for both tones. This result was not dependent on the type of cognitive task performed, and so appears to be equivalent provided attention is effectively diverted to the

primary visual task. The overall larger MMN for the 60 ms tone could be a function of its novelty at the beginning of the sequence, indicating the presence of a simpler form of order-effect remaining in the data – that is, the perception of potentially higher information value for the first deviant tone. It is noteworthy that others have observed longer deviants among short duration tones to elicit larger deviant-minus-standard difference than the reverse (e.g., Catts, Shelley, Ward, & Liebert, 1995; Jaramillo, Alku, & Paavilainen, 1999; Jaramillo, Paavilainen, & Näätänen, 2000; Näätänen, Paavilainen, & Reinikainen, 1989; Peter, McArthur, & Thompson, 2010), which could be due to energy accumulation differences (see Nordby, Hammerborg, Roth, Hugdahl, 1994; Todd & Michie, 2000). However, we note that the MMNs reach the same amplitude eventually in the Movie-Uninformed condition in the present study so we suspect it is better explained by the novelty at initial exposure.

The cognitive tasks employed could interrupt two of the learning mechanisms thought to be responsible for the primacy effect. Firstly, the participants in this condition are required to selectively attend to visual input when the sequences commence. Although this is also true for the Movie-Uninformed group (i.e., they are asked to focus on the movie), it is conceivable that all representations can be built from samples (e.g., by Bayesian chunk learning; for compatible evidence, see e.g., Barascud et al., 2016; Orbán, Fiser, Aslin, & Lengyel, 2008), which is possible while attending to the movie. Therefore, the consequences of attending to a different channel may be more pronounced for the cognitive task group who are required to make responses to what they see, and indeed sustain that effort across the duration of the task. Task performance measures indicate that participants were actively engaged in the concurrent task. This engagement clearly does not interfere with learning about the local tone probabilities. MMN is always elicited for the rare sound within the multi-timescale sequence blocks, implying that internal models are formed and dynamically updated. The increase in MMN amplitude across the block halves also implies that precision-weighting is intact with precision accumulating with progressive confirmatory evidence. This is consistent with previous studies showing that MMN continues to be elicited in conventional oddball sequences if the participant is completing a demanding concurrent task (e.g., Winkler et al., 2003). What is missing however is the learning modulations that have been attributed to: (1) assumptions about stability in the tone tendencies; and (2) the formation of predictions about higher-order structure (block length). Both of these aspects require acquisition and retention of information over a longer timescale compared to the local rule and would be attributed to a “higher” level within the hierarchical Gaussian filter account of learning under uncertainty (Mathys et al., 2011). The results from the Cognitive-Task Uninformed group resemble those obtained when the same sound blocks from Figure 1 are presented in a pseudorandom sequence with no higher-order temporal regularity in block length (Todd et al., 2018); providing further evidence consistent with the notion that the participants in this condition did not detect or encode this higher-order structure. In summary, the MMN amplitude modulations within the Cognitive Task-Uniformed group are consistent with the proposition that learning these longer timescale higher order attributes of the sequence demanding more processing resources and therefore missing in this group.

The provision of information about the sequence structure enables very different expectations in the listener. The uninformed listener can only accumulate information about the sequence in a linear sequential fashion lending itself to order effects. The informed listener on the other hand, is provided with key information that could reduce or eliminate surprise points in the sequence. Critically, the listener knows that the tone probabilities will change and that the rate at which they will change will alter later in the experiment. In the Movie-Informed group, this manipulation removes the significant sequence and block half effects from the data, consistent with comparatively higher precision weighting for the internal models throughout. Once again, MMN is always elicited to the locally rare sound so

the formation of internal models and their dynamic updating is present. This is consistent with several previous studies showing that MMN is not eliminated by foreknowledge of the occurrence of deviants (e.g., Rinne, Antila, & Winkler, 2001; Horváth, Sussman, Winkler, & Schröger, 2011). However, the differential modulations (learning rates) for the two tones seen in the Movie-Uninformed participants are absent. Provision of information may therefore allow the semantic knowledge to alter the assumptions about stability that are credited with the primacy effect – specifically, the knowledge that the tendencies will change could alter the assumptions about their stability, and the knowledge that block length will change could alter the assumptions about stability in the volatility rates. Rather than eliciting surprise, the changes within the sequence may in fact act as confirmation of the prior information. Importantly, these impacts would once again be attributed to “higher” levels within the hierarchical Gaussian filter account of learning under uncertainty. The overall higher amplitude for the 60 ms MMN is likely to be explained by the same factors discussed for the Cognitive Task-Uniformed group above.

Finally, we also explored what would happen if informed listeners heard the sequences while performing a concurrent demanding cognitive task. In the Cognitive Task-Informed group a differential modulation of the MMN amplitudes to the two tones is present. In the slower changing sequence the differential modulation resembles that seen in the primacy effect – slower accumulation of MMN amplitude for the 30 ms deviants. By the second sequence (the faster changing sequence), differential learning rates are not present and the MMN amplitudes simply increase over time within blocks for both tones similar to that in the Cognitive Task-Uniformed group. The cognitive demands associated with the task overtime seem to have prevented this group from being able to use evidence to confirm prior knowledge throughout the stimulus sequence in the manner we have suggested for the Movie-Informed group. One way to account for the effects is to propose that participants may have initially divided their attention between the visual task input and the tones at the onset of the sequence. In the opening moments of the sequence a brief period of allocation of attention to the sounds might have been sufficient to consolidate high precision in the first impression formed (see a possible model referred above). This high precision may in turn explain the differential learning rates to the two tones in the slower changing sequence blocks. However, the cognitive demands associated with sustained attention indicate that the visual input overtime may have exceeded the capacity to divide attention between the two sensory modalities, and as a result, the system displayed the simple stability effects for the subsequent sequence later in the experiment.

Based on the results of the present study we can conclude that the primacy effects on MMN amplitude modulation seen in previous multi-timescale studies are not an inevitable consequence of the sequence structures. If these effects were attributable to bottom-up stimulus driven impacts of the sequence composition, then they should not be susceptible to manipulation of cognitive resources and/or prior knowledge. Instead these data are consistent with models of perceptual inference that allow for top-down modulations of evoked potentials based on longer-term information gleaned from the acoustic environment encountered in the experiment. It is our contention that the manipulation of cognitive resources and prior knowledge may impact these perceptual learning processes by altering attention-based augmentations of learning.

Attention impacts many AEP components (Näätänen, 1992), and although MMN elicitation does not require attention to be directed to sounds, MMN amplitude can be modulated by attention and by available attention resources (see Sussman, 2007 for discussion). Here we extend this literature by showing that modulation of MMN amplitudes proposed to be reliant on strong first-learning effects and the learning of high-order structure can be abolished by depletion of attention-based cognitive resources.



Both the primacy effect, and its elimination here, are reminiscent of learning effects modelled by the Adaptive Resonance Theory (ART) neural network and its biased variant (Carpenter & Gaddam, 2011). ART neural networks serve as models of human cognitive information processing and implement a pre-processing step called complement coding. Complement coding is an example of opponent processing that takes place within the nervous system, where introducing antagonistic states enable systems to act on relative rather than absolute quantities that are tolerant to some variability in actual values (Hurvich & Jameson, 1957). Many concepts employed in ART are analogous to those in predictive coding where bottom-up input is compared to a top-down learned expectation of a “critical feature pattern”. A parameter called vigilance determines the current state with low vigilance favouring abstract prototype learning and high vigilance more specific exemplar learning. Vigilance rises after prediction errors and as vigilance rises, more attention is paid to the fit between top-down and bottom-up information. The learning rules implemented in ART created errors in which focusing on features that were critical to categorical decisions early in learning lead the system to pay ‘too much’ attention (prioritise) these features. This differential attention distorted memory representations later which in turn led to errors when the input changed. This issue could be overcome if a bias was introduced into the system that ensured that attention would be drawn to previously unattended information after prediction errors occurred.

Similar to the concept of a complement coding feature of ART, we have previously argued (e.g., Todd et al., 2013) that the first-impression bias involves an initial categorical-like allocation of one sound to a relatively redundant status (the standard which is predictable and not signalling anything important in the environment), and the other as potentially higher in information value (the deviant whose occurrence cannot be predicted and its rarity means we do not yet know what it might signal). The primacy effect can in fact be prevented for a while when this categorical or binary distinction cannot be made at sequence onset, but the effect returns if it can be deduced from longer term exposure to the sequence (Mullens et al., 2014). We couple this assertion of a binary value state with the supposition that the brain by default assumes low volatility in the environment (e.g., Mullens et al., 2014; Todd et al., 2017). This assumption is then thought causal in delaying the sensitivity to change, a delay or error that within ART is noted as an attentional flaw prioritising early learning and inhibiting a shift in attention to alternatives. If attention is required to encode categorical attributions, or to fix the strong first-impression of value, then this may explain how directing attention away from sounds by a demanding concurrent cognitive tasks interferes with primacy effects. Conceptualising the impact of explicit sequence knowledge within this framework suggests that it could prevent a categorical distinction between sounds at sequence onset by educating the listener that distinctions between rare and common are not fixed during the sequence.

We acknowledge certain limitations of the present study. It is focused on the MMN component of responses to rare pattern deviations although of course MMN is only one component in a chain of events. It may be pertinent, for example to ask whether the manipulations we report here would impact subsequent components such as the P3a and the reorienting negativity that are considered part of an orienting response and reorienting to task, respectively. It is clear from Figure 2 that our rapid presentation rate allows for too short epochs to see any impact on reorienting negativity (peaking ~500 ms after deviance) and not optimal for extraction of P3a which while visible at the end of our epoch but is curtailed. We do have two other data sets, which are better suited to P3a analysis, because in them, the deviants elicited an earlier MMN in response to frequency (Todd, Heathcote, Whitson et al., 2014) or spatial location deviation (Fitzgerald, Provost and Todd, 2018) than the MMN to the duration deviation used in the current study. The order effects on MMN evident in these papers were not reflected in changes in the subsequent P3a. In the current data, there is

certainly a generalised trend for P3a to be larger to the 30ms than the 60ms tone in Figure 2, however this global difference was not the focus of the present study with precision in internal models best reflected in the MMN component.

A second limitation of the current study methodology is the analysis of AEP averages with fewer trials (<100) than what is usually collected to capture the change in the AEPs over time within a block. However, it has become evident across our studies of the primacy bias phenomenon that averaging all responses within the blocks would inevitably collapse some of the differential learning effects: Averaging across whole blocks instead of halves in the multi-timescale sequence typically impacts the 60ms tone MMN only (an effect discussed at length in Todd et al., 2014), presumably owed to its novelty at sequence onset as noted above. Indeed, we have shown previously that MMN is larger overall for the first deviant irrespective of whether it is of shorter or longer duration (Mullens et al., 2014). Whole-block averaging masks some learning effects because the smaller MMN at the onset of blocks reduces MMN average amplitude for the 30ms tone in the slower changing sequence and the 60ms tone in the faster changing sequence. From Figure 2 it is clear that this differential effect of sequence is present for the Movie-Uniformed group, is not present for the Movie-Informed group who show no effect of sequence, and is not present for either Cognitive-Task group who show sequence effects for both tones overall. Therefore, although averaging over whole blocks provides a large number of trials for each average, it is less informative overall about the potential underlying mechanisms.

## 5.0 Conclusions

The results reported here provide a clear demonstration that the way in which perceptual inference processes filter sound information is resource dependent and subject to prior knowledge about how the input is likely to be structured. Both observations are based on differences in the dynamic modulation of MMN amplitude in sequences that contain sound patterning on multiple timescales. Importantly, filters that are based on more local timescale information remain largely impervious to manipulations of resource and knowledge, with effects confined to modulations that are based on information that must be extracted over longer timescales. The outcomes add to literature suggesting that the term “preattentive” is a misnomer when applied to MMN, whose elicitation clearly is preattentive (based on observations during sleep and coma), but whose amplitude clearly is not (see also Sussman, 2007). The data conform to models of cognition that suppose that information can be learned about regularities over many timescales simultaneously, and that learning rates will be shifted by regularity violations, with the impact of violations at a given level being contingent upon prevailing knowledge at higher levels (Mathys et al., 2011). Combined, these observations are consistent with the assertion that cognition becomes increasingly important to perception as we move from local to more global inferences that may aid in simplifying the processing of the sound input.

## Acknowledgements

JDF acknowledges receipt of an Australian Postgraduate Award scholarship. We thank Gavin Cooper for his assistance in programming these experiments and to Alex Provost for his support in training authors JDF, KH and KM in data acquisition and processing. This research was supported by funds provided by the National Health and Medical Research Council of Australia (APP1002995) and the Hungarian Academy of Sciences' Lendület project (LP2012-36/2012 awarded to IW).

Declarations of interest: none

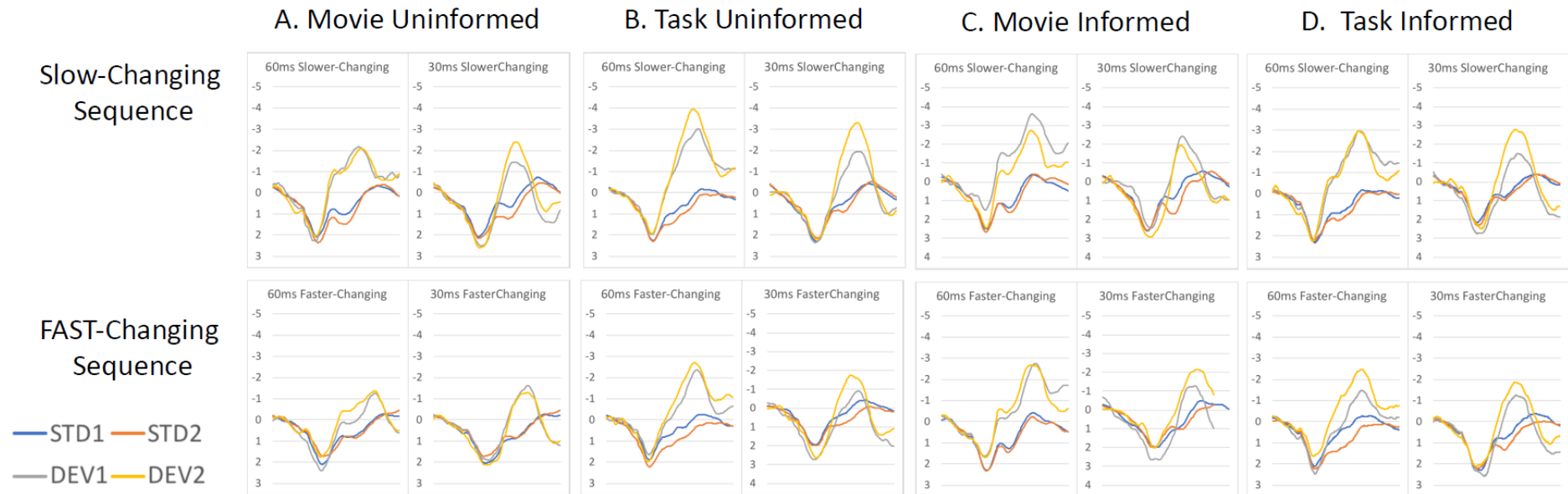
## References

- Badcock, J.C., Williams, R.J., Anderson, M. & Jablensky A. (2004). Speed of processing and individual differences in IQ in schizophrenia: General or specific cognitive deficits? *Cognitive Neuropsychiatry*, 9, 233-47. doi: <https://doi.org/10.1080/13546800344000228>
- Baldeweg, T. (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, 10, 93-94. doi: <http://dx.doi.org/10.1016/j.tics.2006.01.010>
- Barascud, N., Pearce, M. T., Griffiths, T. D., Friston K. J. & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, 113, E616-E625. doi: [10.1073/pnas.1508523113](https://doi.org/10.1073/pnas.1508523113)
- Böttcher-Gandor, C. & Ullsperger, P. (1992). Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval. *Psychophysiology*, 29, 546-550. doi: <https://doi.org/10.1111/j.1469-8986.1992.tb02028.x>
- Bouwer FL, Werner CM, Knetemann M, Honing H. Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia*. 2016 May;85:80-90. doi: [10.1016/j.neuropsychologia.2016.02.018](https://doi.org/10.1016/j.neuropsychologia.2016.02.018)
- Catts, S. V., Shelley, A. M., Ward, P. B., & Liebert, B. (1995). Brain potential evidence for an auditory sensory memory deficit in schizophrenia. *The American journal of psychiatry*, 152, 213. doi: [10.1176/ajp.152.2.213](https://doi.org/10.1176/ajp.152.2.213)
- Carpenter, G. A. & Gaddam, S. C. (2010). Biased ART: A neural architecture that shifts attention toward previously disregarded features following an incorrect prediction. *Neural Networks*, 23, 435-451. doi: [10.1016/j.neunet.2009.07.025](https://doi.org/10.1016/j.neunet.2009.07.025)
- Costa- Faidella, J., Grimm, S., Slabu, L., Díaz- Santaella, F., & Escera, C. (2011). Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. *Psychophysiology*, 48(6), 774-783. doi: [10.1111/j.1469-8986.2010.01144.x](https://doi.org/10.1111/j.1469-8986.2010.01144.x)
- Fischer, C., Morlet, D. & Giard, M. (2000). Mismatch negativity and N100 in comatose patients. *Audiology & Neuro Otology*, 5, 192-197. doi: [10.1159/000013880](https://doi.org/10.1159/000013880)
- Fitzgerald, K., Provost, A., & Todd, J. (2018). First- impression bias effects on mismatch negativity to auditory spatial deviants. *Psychophysiology*, 55(4). doi: [0.1111/psyp.13013](https://doi.org/10.1111/psyp.13013)
- Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456), 815-836. doi: [10.1098/rstb.2005.1622](https://doi.org/10.1098/rstb.2005.1622)
- Frost, J. D., Winkler, I. Provost, A. & Todd, J. (2016). Surprising sequential effects on MMN. *Biological Psychology*, 116, 47-56. doi: [10.1016/j.biopsycho.2015.10.005](https://doi.org/10.1016/j.biopsycho.2015.10.005)
- Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, 64, 384-390. doi: <http://dx.doi.org/10.1037/h0041403>
- Horváth, J., Sussman, E., Winkler, I., & Schröger, E. (2011). Preventing distraction: assessing stimulus-specific and general effects of the predictive cueing of deviant auditory events. *Biological Psychology*, 87, 35-48. doi: [10.1016/j.biopsycho.2011.01.011](https://doi.org/10.1016/j.biopsycho.2011.01.011)
- Jaramillo, M., Alku, P., & Paavilainen, P. (1999). An event- related potential (ERP) study of duration changes in speech and non- speech sounds. *NeuroReport*, 10, 3301. doi: <https://doi.org/10.1016/j.dcn.2016.04.001>
- Jaramillo, M., Paavilainen, P., & Näätänen, R. (2000). Mismatch negativity and behavioural discrimination in humans as a function of the magnitude of change in sound duration. *Neuroscience Letters*, 290, 101-104. doi: [10.1016/S0304-3940\(00\)01344-6](https://doi.org/10.1016/S0304-3940(00)01344-6)
- Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput Biol*, 4(11), doi: <https://doi.org/10.1371/journal.pcbi.1000209>

- Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biological Psychology*, 74(1), 1-19. doi: <https://doi.org/10.1016/j.biopsycho.2006.06.001>
- Ladinig, O., Honing, H., Háden, G., & Winkler, I. (2009). Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training. *Music Perception*, 26 (4), 377-386. doi: <http://dx.doi.org/10.1525/mp.2009.26.4.377>
- Lieder, F., Stephan, K. E., Daunizeau, J., Garrido, M. I., & Friston, K. J. (2013). A neurocomputational model of the mismatch negativity. *PLoS computational biology*, 9 (11), e1003288. doi: <https://doi.org/10.1371/annotation/ca4c3cdf-9573-4a93-9542-3a62cddb8396>
- Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 5, 39. doi: [10.3389/fnhum.2011.00039](https://doi.org/10.3389/fnhum.2011.00039)
- Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J. & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, 8, 825. doi: <https://doi.org/10.3389/fnhum.2014.00825>
- Mittag, M., Takegata, R., & Winkler, I. (2016). Transitional probabilities are prioritized over stimulus/pattern probabilities in auditory deviance detection: Memory basis for predictive sound processing. *The Journal of Neuroscience*, 36(37), 9572–9579. doi: [10.1523/JNEUROSCI.1041-16.2016](https://doi.org/10.1523/JNEUROSCI.1041-16.2016)
- Mullens, D., Woodley, J., Whitson, L., Provost, A., Heathcote, A., Winkler, I & Todd, J. (2014). Altering the primacy bias-How does a prior task affect mismatch negativity? *Psychophysiology*, 51, 437-45. doi: [10.1111/psyp.12190](https://doi.org/10.1111/psyp.12190)
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ, Lawrence Erlbaum Associates.
- Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, 48(1), 4-22. doi: [10.1111/j.1469-8986.2010.01114.x](https://doi.org/10.1111/j.1469-8986.2010.01114.x)
- Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neuroscience letters*, 107, 347-352. doi: [https://doi.org/10.1016/0304-3940\(89\)90844-6](https://doi.org/10.1016/0304-3940(89)90844-6)
- Nordby, H., Hammerborg, D., Roth, W.T. & Hugdahl, K. (1994). ERPs to infrequent omissions and inclusions of stimulus elements. *Psychophysiology*, 31(6), 544-52. doi: <https://doi.org/10.1111/j.1469-8986.1994.tb02347.x>
- Orbán, G., Fiser, J., Aslin, R.N. & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceeding of the National Academy of Sciences USA*, 105 (7), 2745-2750. doi: <https://doi.org/10.1073/pnas.0708424105>
- Owen, A. M., McMillan, K. M., Laird, A. R. & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25, 46-59. doi: [10.1002/hbm.20131](https://doi.org/10.1002/hbm.20131)
- Peter, V., McArthur, G., & Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neuroscience letters*, 482(1), 71-75. doi: [10.1016/j.neulet.2010.07.010](https://doi.org/10.1016/j.neulet.2010.07.010)
- Rinne, T., Antila, S., & Winkler, I. (2001). Mismatch negativity is unaffected by top down predictive information. *NeuroReport*, 12, 2209-2213. doi: [http://dx.doi.org/10.1097/00001756-200107200-00033](https://doi.org/10.1097/00001756-200107200-00033)
- Rubin, J., Ulanovsky, N., Nelken, I. & Tishby, N. (2016). The Representation of prediction error in auditory cortex. *PLoS Comput Biol* 12(8), e1005058. doi: <https://doi.org/10.1371/journal.pcbi.1005058>

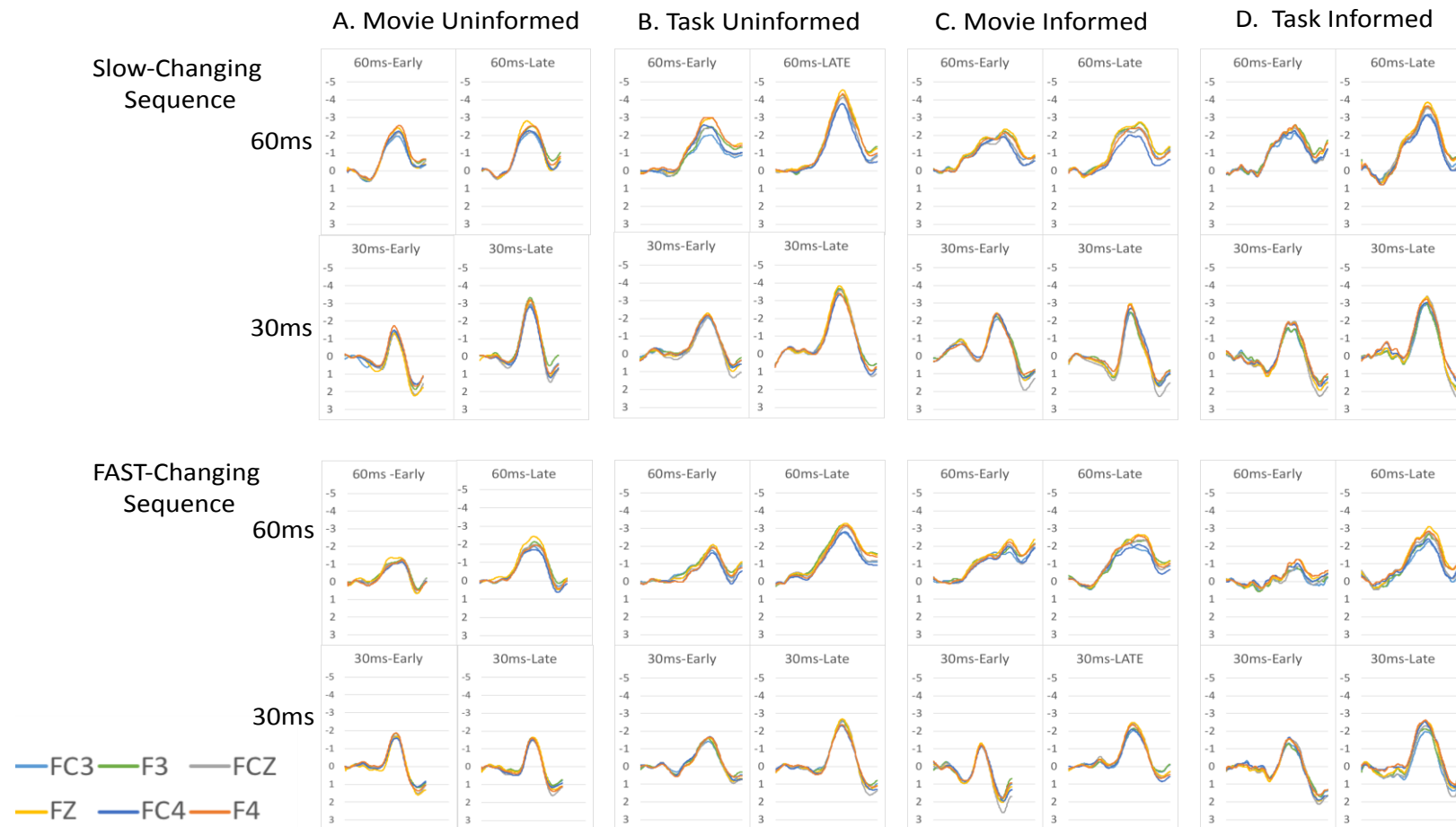
- Sabri, M., Labelle, S., Gosselin, A., and Campbell, K. B. (2003). Effects of sleep onset on the mismatch negativity (MMN) to frequency deviants using a rapid rate of presentation. *Cognitive Brain Research*, 17, 164-76. doi: <http://dx.doi.org/10.1037/h0087343>
- Sussman, E. (2007). A New View on the MMN and Attention Debate: The Role of Context in Processing Auditory Events. *Journal of Psychophysiology*, 21, 164-75. doi: <https://doi.org/10.1027/0269-8803.21.34.164>
- Sussman, E., Winkler, I., Houtilainen, M., Ritter, W., and Naatanen, R. (2002) Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, 13, 393-405. doi: [https://doi.org/10.1016/S0926-6410\(01\)00131-8](https://doi.org/10.1016/S0926-6410(01)00131-8)
- Sussman, E., & Winkler, I. (2001). Dynamic sensory updating in the auditory system. *Cognitive Brain Research*, 12, 431-439. doi: [http://dx.doi.org/10.1016/S0926-6410\(01\)00067-2](http://dx.doi.org/10.1016/S0926-6410(01)00067-2)
- Todd, J., Heathcote, A., Mullens, D., Whitson, L. R., Provost, A., & Winkler, I. (2014a). What controls gain in gain control? Mismatch negativity (MMN), priors and system biases. *Brain Topogr*, 27(4), 578-589. doi: [10.1007/s10548-013-0344-4](https://doi.org/10.1007/s10548-013-0344-4)
- Todd, J., Heathcote, A., Whitson, L. R., Mullens, D., Provost, A., & Winkler, I. (2014b). Mismatch negativity (MMN) to pitch change is susceptible to order-dependent bias. *Front Neurosci*, 8, 180. doi: [10.3389/fnins.2014.00180](https://doi.org/10.3389/fnins.2014.00180)
- Todd, J. & Michie, P. T. (2000). Do perceived loudness cues contribute to duration mismatch negativity (MMN)? *Neuroreport*, 11(17), 3771-3774. doi: [10.1097/00001756-200011270-00035](https://doi.org/10.1097/00001756-200011270-00035)
- Todd, J., Petherbridge, A., Speirs, B., Provost, A., Paton, B. (2018). Time as context: The influence of hierarchical patterning on sensory inference. *Schizophrenia Research*, 191, 123-131. doi: [10.1016/j.schres.2017.03.033](https://doi.org/10.1016/j.schres.2017.03.033)
- Todd, J., Provost, A. L., & Cooper, G. J. (2011). Lasting first impressions: A conservative bias in automatic filters of the acoustic environment. *Neuropsychologia*, 49(12), 3399-3405. doi: [10.1016/j.neuropsychologia.2011.08.016](https://doi.org/10.1016/j.neuropsychologia.2011.08.016)
- Todd, J., Provost, A., Whitson, L. R., Cooper, G., & Heathcote, A. (2013). Not so primitive: context-sensitive meta-learning about unattended sound sequences. *Journal of Neurophysiology*, 109(1), 99-105. doi: [10.1152/jn.00581.2012](https://doi.org/10.1152/jn.00581.2012)
- Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131. doi: <http://dx.doi.org/10.1126/science.185.4157.1124>
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache L., & Dehaene S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), 20754-20759. doi: <https://doi.org/10.1073/pnas.1117807108>
- Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, 21(3-4), 147-163. doi: <https://doi.org/10.1027/0269-8803.21.34.147>
- Winkler, I., Karmos, G. & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, 742, 239-252. doi: [https://doi.org/10.1016/S0006-8993\(96\)01008-6](https://doi.org/10.1016/S0006-8993(96)01008-6)
- Winkler, I., Korzyukov, O., Gumenyuk, V., Cowan, N., Linkenkaer-Hansen, K., Alho, K., Ilmoniemi, R.J. & Näätänen, R. (2002). Temporary and longer retention of acoustic information. *Psychophysiology*, 39, 530-534. doi: <https://doi.org/10.1017/S0048577201393186>
- Winkler, I., & Schröger, E. (2015). Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain and Language*, 148, 1-22. doi: <https://doi.org/10.1016/j.bandl.2015.05.003>
- Winkler, I., Sussman, E., Tervaniemi, M., Ritter, W., Horváth, J., & Näätänen, R. (2003). Pre-attentive auditory context effects. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 57-77. doi: [10.3758/CABN.3.1.57](https://doi.org/10.3758/CABN.3.1.57)

## Supplementary Material



*Figure 1.* Group averaged deviant and standard ERPs by condition in the A. Movie Uninformed, B. Task Uninformed, C. Movie Informed, and D. Task Informed for each deviant type in slower-changing (top row) and faster-changing (bottom-row) sequence for the first-half (early; labelled as 1 on figure legend) and the second-half period (late; labelled as 2 on figure legend) within blocks.





*Figure 2.* Group averaged deviant-minus-standard difference waveforms for each deviant type in each sequence for the first-half period within blocks (early) and the second-half period within blocks (late). Sites FC3, F3, FCZ, FZ, FC4 and F4 are plotted to demonstrate both that MMN generally reached maximum amplitude at F4 and that the results at F4 in manuscript are representative of any experimental effects on the data.