Precision-weighting of musical prediction error: Converging neurophysiological and behavioral evidence

Quiroga-Martinez, D.R.[a], Hansen N.C.[b], Højlund A.[c], Pearce M.[a,d], Brattico E.[a], Vuust P.[a]

a) Center for Music in the Brain, Aarhus University & The Royal Academy of music, Denmark
b) The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
c) Center for Functionally Integrative Neuroscience, Aarhus University, Denmark
d) School of Electronic Engineering and Computer Science, Queen Mary University of London, England

Corresponding author:
David Ricardo Quiroga-Martinez
…..
Address: Nørrebrogade 44 - Building 10G, 5th floor - 8000 Aarhus C., Denmark
Phone: +45 52526481
Email: dquiroga@clin.au.dk

Abstract

Theories of predictive processing propose that prediction error responses are modulated by *precision*. While there is some evidence for this phenomenon in the visual and, to some extent, the auditory domain, little is known about whether and how it happens in the complex auditory contexts of daily life. Here, we looked at the precision-weighting of prediction error in a common, structurally rich and more ecologically valid context such as music. We created musical tone sequences with different degrees of entropy to manipulate the precision of participants' auditory expectations. Magnetoencephalography (MEG) was used to measure the magnetic counterpart of the mismatch negativity (MMNm) as a neural marker of prediction error in a multi-feature paradigm. Pitch, slide, intensity, and timbre deviants were included. We compared high-entropy stimuli, consisting of a set of non-repetitive melodies, with low-entropy stimuli consisting of a simpler and more predictable pitch pattern. Entropy was estimated with an information-theoretic model of auditory expectation. We found a strong reduction in pitch and slide MMNm amplitudes in the high-entropy as compared to the low-entropy context. No significant differences were found for intensity and timbre MMNm amplitudes. Furthermore, in a separate behavioral experiment investigating deviance detection, similar decreases were found for accuracy and confidence measures in response to more subtle increases in stimulus entropy. Our results indicate that precision modulates musical prediction error and suggest that this effect is specific to features that depend on the manipulated auditory dimension—pitch information, in this case. We thus provide evidence consistent with a precision-weighting mechanism in the auditory domain.

Keywords: Precision, Prediction error, Music, Mismatch Negativity, Multi-feature, IDyOM

1. Introduction

Prediction is increasingly considered a core principle for brain function. Several theories propose that the brain creates predictive models of the world that it uses to anticipate and explain incoming sensory information during perception (Bar, 2009; Clark, 2016; Friston, 2005; Hohwy, 2013; Rao & Ballard, 1999). The difference between the predicted and the actual sensory input generates prediction error and its minimization through inference, learning, and action is considered fundamental for survival (den Ouden, Kok, & de Lange, 2012; Friston, 2010). A related but less well-understood phenomenon is *precision*, which can be defined as the certainty or specificity of predictions. Precision is regarded as an estimate of the reliability of sensory signals and has been proposed to modulate the gain of neural activity (Clark, 2013; Feldman & Friston, 2010; Hohwy, 2012). More concretely, the strength of prediction error responses is proposed to be up-weighted in contexts with high precision and down-weighted as contexts become more imprecise. Such a precision-weighting mechanism would ensure that only the more reliable sensory signals drive learning and behavior. So far, the evidence in support of this mechanism is scarce. Most of it has come from the visual domain and has centered on selective attention as a means to optimize precision (Feldman & Friston, 2010; Jiang, Summerfield, & Egner, 2013; Kok, Rahnev, Jehee, Lau, & De Lange, 2012). In the auditory domain, some studies have similarly addressed precision in terms of attention (Auksztulewicz & Friston, 2015; Chennu et al., 2013; Garrido, Rowe, Halász, & Mattingley, 2017; Schröger, Marzecová, & Sanmiguel, 2015) and only a handful have started to show how the statistical properties of the stimuli themselves can modulate predictive precision (Garrido, Sahani, & Dolan, 2013; Heilbron & Chait, 2017; Hsu, Le Bars, Hamalainen, & Waszak, 2015; Sohoglu & Chait, 2016). Since previous work has employed very simple and artificial auditory stimuli, the present study aims to determine whether and how precision modulates auditory prediction error in a common, structurally rich and more ecologically valid context such as music listening.

Music perception provides a useful model of auditory prediction. There is substantial evidence that listeners constantly generate expectations about different features in a musical piece, based on the statistical regularities of the context and long-term knowledge of a musical style (Huron, 2006; Pearce, 2018). The violation of these expectations generates prediction error responses that can be measured with neuroimaging methods (e.g. Carrus, Pearce, & Bhattacharya, 2013; Koelsch, Gunter, Friederici, & Schröger, 2000; Vuust et al., 2005). Importantly, precision has been suggested to modulate musical prediction error (Hansen, Dietz, & Vuust, 2017; Ross & Hansen, 2016; Vuust, Witek, Dietz, & Kringelbach, 2018). For instance, relevant to temporal features in music, Vuust and colleagues (2018) have proposed a model, the Predictive Coding of Rhythmic Incongruity (PCRI), that explains how prediction error and precision interact during rhythm perception. According to this framework, the meter acts as a mental model of the temporal structure of music, and the actual rhythms—i.e., sound onsets—violate or confirm this model to different degrees. When rhythms are more unpredictable, the model's predictions become less precise and deviations become less salient. The PCRI framework has been used to explain the inverted U-shaped relationship between rhythmic complexity, and groove and pleasure (Witek, Clarke, Wallentin, Kringelbach, & Vuust, 2014). Moreover, a similar interaction between precision and prediction error has been proposed to underlie more general music-related reward (Hansen et al., 2017). Thus, precision is thought to play an important role, not only in music perception, but also in the aesthetic and emotional dimensions of musical experience. Interestingly, there is already behavioral evidence for a precision-weighting effect in music listening. Two studies have shown that listeners estimate the precision of musical expectations and that unexpected tones are more salient in contexts with low as compared to high uncertainty (Hansen & Pearce, 2014; Hansen, Vuust, & Pearce, 2016). Nevertheless, how precision affects musical prediction error at the neural level remains unknown.

To address this question, we manipulated the precision of participants' expectations by creating auditory sequences with different degrees of entropy, as quantified by Information Dynamics of Music (IDyOM), a variable-order Markov model of auditory expectation (Pearce, 2005, 2018). IDyOM generates expectations at each point of an auditory sequence in the form of a probability distribution (P) of all possible continuations at that particular moment. These probabilities are conditional on the preceding context and the previous long-term exposure of the model. The uncertainty of the expectations is quantified in terms of entropy:

$$E(P) = -\sum_{p \in P} p \, log_2 p$$

Since the probabilities of all possible continuations (p) sum to one, entropy is minimal when only one event has a very high probability and is maximal when all possible events are equally likely. IDyOM's entropy estimates have been shown to correlate with behavioral measures of uncertainty (Hansen & Pearce, 2014; Hansen et al., 2016). Once the next event in the sequence appears, IDyOM estimates its unexpectedness in terms of information content (IC):

$$IC_p = -log_2 p$$

Thus, unexpected events have high IC. In this study, mean entropy and mean IC values are used as measures of how the statistical properties of the stimuli drive listeners' predictive precision (see section 2.1.2.1. for further details). Our primary measure is entropy, as it directly estimates precision, given that precision is the inverse of uncertainty. Thus, contexts with low entropy would generate precise expectations and contexts with high entropy would generate imprecise expectations. Our secondary measure is mean IC, since stimuli with decreasing levels of precision—and thus increasing levels of entropy—would tend to yield higher levels of unexpectedness and IC in the long run, at least for models generating accurate predictions. Note, however, that IC and entropy can be dissociated in some situations. For example, it is possible to have very precise expectations which are nonetheless constantly violated. This could happen in the case of a listener that has precise long-term knowledge of a musical style but applies this knowledge to an unfamiliar musical style. Consequently, here we consider both measures to obtain a more accurate picture of the listener's predictive model in relation to the stimuli.

As a marker of prediction error, we recorded the mismatch negativity (MMNm) (Näätänen, Gaillard, & Mantysalo, 1978), which is a well-known neural response to regularity violations in a stimulus sequence (Näätänen, Paavilainen, Rinne, & Alho, 2007). It is generated by neuronal populations located in the primary or non-primary auditory cortex, with additional generators in the inferior frontal cortex (Deouell, 2007). Recent work suggests that the MMN reflects the violation and resulting update of neural predictive models (Bendixen, SanMiguel, & Schröger, 2012; Garrido, Kilner, Stephan, & Friston, 2009; Lieder, Stephan, Daunizeau, Garrido, & Friston, 2013). Some studies already hint at a precision modulation of the MMN. In these studies, repetitive standard patterns are compared with stimuli in which the formation of regularities is prevented by the introduction of random sequences of standard tones (Hsu et al., 2015; Jacobsen & Schröger, 2001; Maess, Jacobsen, Schröger, & Friederici, 2007). The results show no MMN for the latter. In other words, a very imprecise predictive model seems to lead to highly reduced prediction error responses. Perhaps for this reason, most MMN studies employ very simple and repetitive stimuli, which favor the precision of the predictive model and the strength of the recorded signal, but do not tell us much about predictive processing in the rich and complex auditory environments of daily life. This is the case even for musical MMN paradigms that aim at making auditory stimuli more real-sounding (e.g., Tervaniemi, Huotilainen, & Brattico, 2014; Vuust et al., 2011; but see Brattico, Tervaniemi, Näätänen & Peretz, 2006 for an example of more ecologically valid stimuli). To assess properly the effect of precision on musical prediction error, an intermediate type of stimulus is needed; one that is more complex than current paradigms, but at the same time less complex and more real sounding than a random succession of tones. Observing subtle modulations of the MMN in this type of auditory context would provide stronger evidence for the precision-weighting mechanism in a more ecologically valid setting.

In the present study, we conducted separate neurophysiological and behavioral experiments to determine whether and how musical prediction error is modulated by precision. In the neurophysiological experiment, we used magnetoencephalography (MEG) to record MMNm responses to pitch, intensity, timbre and slide deviants in high-entropy (HE) and low-entropy (LE) contexts. The LE context consisted of an adapted version of the simple and repetitive musical multi-feature paradigm (Vuust et al., 2011) (see section 2.1.2. for more details). In contrast, the HE context consisted of a set of novel non-repetitive melodies. In line with a precision-weighting mechanism, we expected reduced—but still present—MMNm amplitudes for the HE as compared to the LE context. Moreover, since the paradigm allows the study of several musical features at the same time, in further analyses we assessed the possibility that different features were affected by stimulus entropy in different ways. This was of particular interest given that we only manipulated entropy in the pitch dimension, while entropy for features such as timbre or intensity was kept constant. In the behavioral experiment, we asked participants to detect pitch deviants introduced in several tone sequences, and to report the confidence of their responses. This experiment had two aims. The first was to assess behaviorally the putative precision-weighting of the MMNm. The second was to determine whether listeners are sensitive to fine-grained modulations of stimulus entropy. For this reason, we employed stimuli with five degrees of entropy, which included a subset of the HE/LE stimuli used in the MEG experiment and three additional conditions with intermediate entropy levels. We expected lower accuracy and confidence ratings as the entropy of the sequences increased.

2. Materials and Methods

2.1. MEG experiment

2.1.1. Participants

Twenty-four non-musicians (13 women, mean age 26.5, range 19-34) took part in the MEG experiment. They were right-handed and reported normal hearing and being neurologically healthy. Musical expertise was measured with the musical training subscale of the Goldsmiths Musical Sophistication Index (Gold-MSI) questionnaire (Müllensiefen, Gingras, Musil, & Stewart, 2014) for which the mean score was 10.17 (SD = ± 3.5). All scores lay within the 26th percentile of the norm for the subscale. Moreover, participants' musical competence was assessed with the Musical Ear Test (MET) (Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010) which yielded a mean total score of 69.12 (SD = ± 9.44). This score falls within normal values for this population (Wallentin et al., 2010). Participants were recruited through an online database for experiment participation, agreed to take part voluntarily, gave their informed consent and received 300 Danish kroner (approximately 40 Euros) as compensation. The study was conducted in accordance with the Helsinki declaration and was approved by the regional ethics committee (De Videnskabsetiske Komitéer for Region Midtjylland in Denmark).

2.1.2. Stimuli

In the MEG experiment, high-entropy and low-entropy stimuli were included (Figure 1a). For the LE condition, we adapted the musical multi-feature paradigm (Vuust et al., 2011; Vuust, Brattico, Seppänen, Näätänen, & Tervaniemi, 2012; Vuust, Liikala, Näätänen, Brattico, & Brattico, 2016), which consists of a four-note repeating pattern (low-high-medium-high pitch) employing the notes from a major or minor chord. This pattern is used across musical styles and has become known as the Alberti bass (Fuller, 2001). The HE condition consisted of a set of six novel melodies which contained almost no repetitions of melodic patterns. All the tones in the melodies were isochronous, in order to make both conditions directly comparable with each other and other MMN paradigms (the full set of stimuli is shown in supplementary materials 1). Individual tone sequences—i.e., single melodies or Alberti bass sequences—in both conditions were 32-notes long, lasted eight seconds and were pseudo-randomly transposed to twelve keys comprising the major and melodic-minor modes of C, C#, D, D#, E, and F. After transposition, the pitch alphabet—i.e. the number of distinct pitch categories—of the HE condition spanned up to thirty-one semitones from B3 ($F_0 \approx 247\,Hz$) to F6 ($F_0 \approx 1397\,Hz$). To minimize acoustic and pitch-alphabet confounds, we made sure that LE sequences spanned approximately the same range as HE sequences by transposing half of them to the octave from C4 to C5, and the other half to the octave from C5 to C6 (supplementary materials 2).
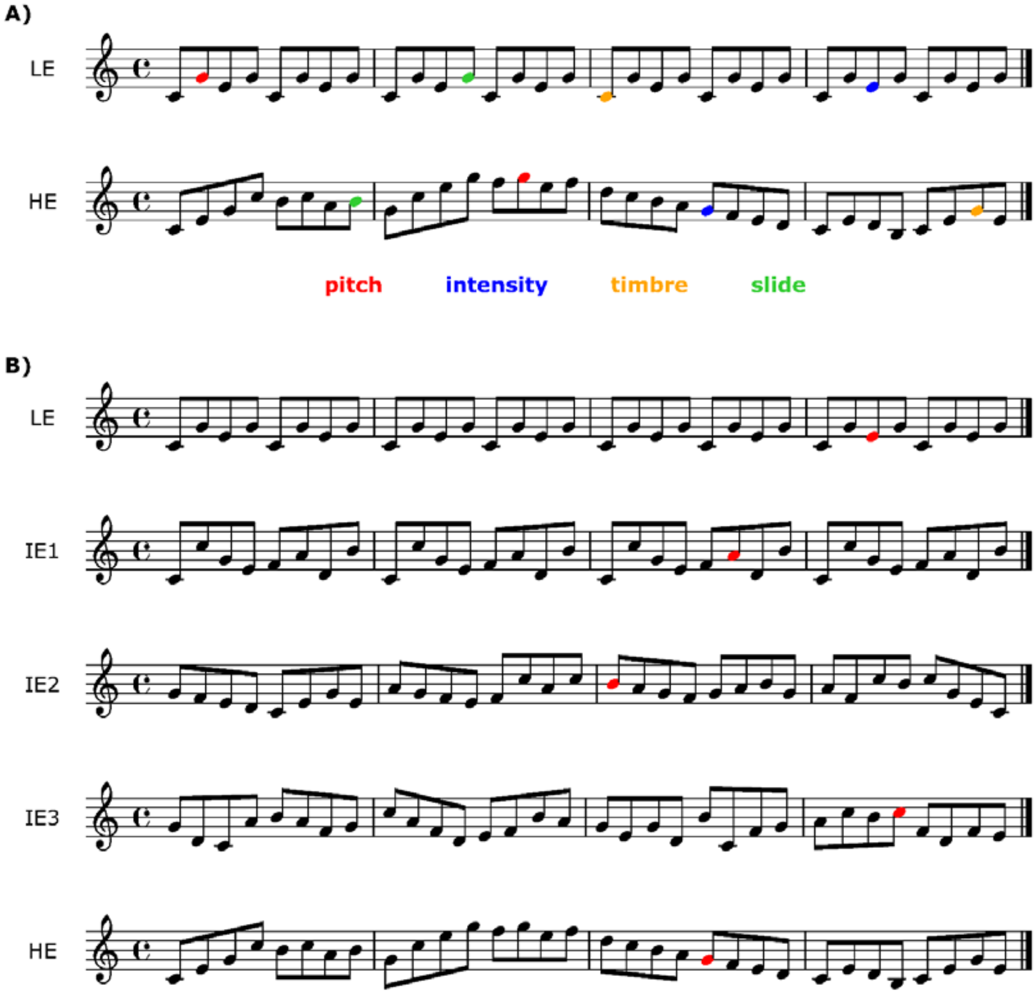
Figure 1. Examples of the sequences used in A) the MEG experiment and B) the behavioral experiment. LE = low entropy, IE = intermediate entropy, HE = high entropy. Two of the conditions in the behavioral experiment (LE, HE) correspond to the conditions in the MEG experiment. Colored tones represent deviants. Note that LE and IE1 consist exclusively of the patterns presented here, whereas IE2, IE3 and HE consist of several different sequences/melodies, which are shown in supplementary materials 1. Also note that the positions of the deviants were randomized as described in the text and the ones presented here only constitute an example.

Stimuli were presented using a pool of piano tones made with the "Warm Grand" sample from the Halion sampler in Cubase (Steinberg Media Technology, version 8). Each tone had a duration of 250 ms, was peak-amplitude normalized and had 3-ms-long fade-in and fade-out to prevent clicking. There were no gaps between tones. Pitch deviants were created by mistuning the original tones 50 cents up. The slide deviant was a continuous pitch glide which spanned the whole duration of the tone, going from two semitones below towards the pitch of the corresponding standard tone. For the intensity deviant, sound level was decreased by 20 dB. The timbre deviant was made with a telephone receiver effect consisting of a band pass filter between 1 and 4 kHz. All deviants were created with Audition (Adobe Systems Incorporated, version 8).

Each of the two conditions was presented in a separate group of three consecutive blocks. Within each block, the tone sequences were played one after the other without pause between them. The order of HE sequences was pseudorandom so that any sequence of twelve consecutive melodies contained no more than one major and minor version of each. No melody was played twice in a row. Transpositions in both conditions were pseudorandomized in the same way. At the beginning of each block, a sequence with no deviants was added in order to ensure a certain level of auditory regularity at the outset. The duration of the pause between blocks was not fixed but usually took around one minute.

Deviants were introduced as follows. Each 32-note sequence was divided into eight groups of four notes (Figure 1a). In half of the sequences, deviants occurred in groups 1, 3, 5 and 7. In the other half, they occurred in groups 2, 4, 6 and 8. This was done because we also included a combined condition where HE and LE sequences were played simultaneously, thereby creating two-part musical excerpts. Thus, the position of the deviants was distributed across streams to counterbalance the effects of key-changes between parts. The purpose of this condition was to assess the predictive processing of simultaneous musical streams, which is beyond the scope of this article. The corresponding results will be reported elsewhere. Within each four-note group, only one deviant could occur randomly in any of the four positions with equal probability. There was one deviant per feature in each sequence and their order of appearance was pseudorandom. There were 144 sequences in each condition and the same number of deviants per feature. Given that each deviant type occurred once per thirty-two notes, its overall zeroth-order probability was $1/32 \approx 0.031$. In the session we also included another group of three consecutive blocks in which Alberti bass sequences were played in a low pitch range. This condition served as a control for the combined condition and therefore is not the focus of this article either. The order of this and the HE and LE conditions was counterbalanced across participants. These conditions always came after the three blocks of the combined condition.

2.1.2.1. Quantitative estimates with IDyOM

Mean IC and mean entropy were quantified with IDyOM using a model that combined short-term probabilities inferred from the sequences themselves with long-term probabilities learned from a corpus of Western tonal hymns and folksongs (datasets 1, 2, and 9 from Table 4.1 in Pearce 2005, comprising 50,867 notes**)**. This corpus has been extensively used in prior research thus aiding comparison with existing results. The model simulates a listener that generates predictions based on life-long knowledge of Western tonal music, but who is also capable of learning the local structure of the current stimuli. IDyOM can use different parameters of the musical surface—known as viewpoints—to derive its probabilistic predictions (Conklin & Witten, 1995; Pearce, 2005). In the previous literature, IDyOM has often been configured to use a viewpoint combining tonal scale degree and pitch-interval (e.g. Carrus et al., 2013; Hansen & Pearce, 2014; Omigie, Pearce, Williamson, & Stewart, 2013; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010) in order to capture both melodic and tonal structure (see Pearce, 2005 for more details). This viewpoint was used here to obtain note-by-note IC and entropy values, which were then averaged for each sequence and condition. Crucially, IDyOM uses the pitch alphabet of the training corpus for its predictions, which in this case was larger than the alphabets of our stimuli. This makes the model insensitive to differences in pitch alphabet between conditions, which have been identified as an important source of uncertainty (Auksztulewicz et al., 2017; Barascud, Pearce, Griffiths, Friston, & Chait, 2016). For this reason, we adjusted the distributions to include only the probabilities of the tones present in each condition, which were then renormalized to sum to one. The IDyOM analysis confirmed higher mean entropy and IC for HE than LE stimuli (Table 1). The IC and entropy profiles of all the sequences are shown in supplementary materials 1.

Experiment	Condition	Entropy		Information content	
		Mean	SD	Mean	SD
MEG	LE	1.22	0.29	1.37	1.89
	HE	2.26	0.25	2.79	2.43
Behavioral	LE	1.33	0.37	1.60	0.97
	IE1	2.01	0.48	2.59	2.64
	IE2	2.10	0.41	2.75	1.62
	IE3	2.11	0.54	4.70	2.79
	HE	2.41	0.61	2.48	2.14

Table 1. Mean and standard deviation for entropy and information content values measured in bits for the stimuli included in the MEG and behavioral experiments.

2.1.3. Procedure

At the beginning of the session, participants received oral and written information and gave their consent. Then they filled out the Gold-MSI questionnaire and completed the MET. Once participants had put on MEG-compatible clothing, electrodes and coils were attached to their skin and their heads were digitized. During the MEG recording, they were sitting upright in the MEG device looking at a screen. Before presenting the musical stimuli, their auditory threshold was measured through a staircase procedure and the sound level was set at 60dB above threshold. Participants were instructed to watch a silent movie of their choice, ignore the sounds and move as little as possible. They were told there would be music playing in the background interrupted by short pauses so that they could take a break and readjust their posture. Sounds were presented through isolated MEG-compatible ear tubes (Etymotic ER•30). The MEG recording lasted approximately 90 minutes and the whole experimental session took between 2.5 and 3 hours including consent, musical expertise tests, preparation, instructions, breaks, and debriefing.

2.1.4. MEG recording and analyses

Magnetic correlates of brain activity were recorded using an Elekta Neuromag MEG TRIUX system with 306 channels (204 planar gradiometers and 102 magnetometers) and a sampling rate of 1000 Hz. Participants' head position was monitored with four coils (cHPI) attached to the forehead and the mastoids. Offline, the signals coming from inside the skull were isolated with the temporal extension of the signal source separation (tSSS) technique (Taulu, Kajola, & Simola, 2004) using Elekta's MaxFilter software (Version 2.2.15). This procedure included movement compensation in all but two participants, in which the continuous head position information was not reliable due to suboptimal placement of the coils. However, in these cases the presence of reliable auditory event-related fields (ERFs) was successfully verified by visually inspecting the amplitude and polarity of the M50 component. Eye-blink and heartbeat artifacts were corrected with the aid of electrocardiography (ECG) and electrooculography (EOG) recordings, and independent component analysis as implemented by a semi-automatic routine

(FastICA algorithm and functions find_bads_eog and find_bads_ecg in the software MNE-Python) (Gramfort et al., 2013). Visual inspection was used as a quality check.

The ensuing analysis steps were conducted with the Fieldtrip toolbox (version r9093) implemented in Matlab (Oostenveld, Fries, Maris, & Schoffelen, 2011). Epochs comprising a time window of 400 ms after sound onset were extracted and baseline-corrected, with a pre-stimulus baseline of 100 ms. The epochs were then low-pass filtered with a cut-off frequency of 35 Hz and down-sampled to a resolution of 256 Hz. For each participant, ERFs were computed by averaging the responses for all deviants for each feature and averaging a selection of an equal number of standards. These were selected by finding, for each single deviant, a standard tone that was not preceded by a deviant and was in the same position of the same HE or LE sequence—although not necessarily the same transposition—in a different trial. This reduced additional noise caused by fluctuations in melodic expectations. After averaging, planar gradiometers were combined by computing root mean square values. Finally, a new baseline correction was applied and MMNm difference waves were computed by subtracting the ERFs of standards from the ERFs of deviants.

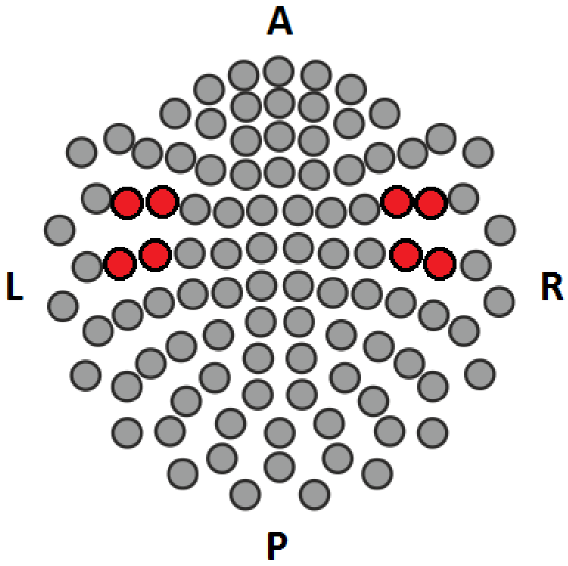

Figure 2. MEG sensor array. Channels with the largest M50 responses selected for the secondary mean gradient amplitude analyses are shown in red. The channels on the right hemisphere were used for the plots shown here.

The statistical analyses were performed on the data from the combined gradiometers, as these sensors measure activity directly above the neural sources and have a better signal-to-noise ratio (Haumann, Parkkonen, Kliuchko, Vuust & Brattico, 2016). Magnetometers were used for the inspection of the polarity of the components. For the primary statistical analyses, we used cluster-based permutation tests, which control the family-wise error rate by computing only one test statistic for any relevant contrast, based on a clustering in space and time of the sensor activity. The statistic is then compared to the probability of equal or more extreme values in a null distribution built from a number of random permutations of the data, yielding a Monte Carlo estimate of significance (Maris & Oostenveld, 2007). In our case, the calculation was based on one-sided paired *t*-tests, as the expected effects were unidirectional. The sample-level significance threshold was .05. The chosen statistic was the maximal sum of clustered positive *T*-values (maxsum). The cluster-level null distribution consisted of 10,000 random permutations of the condition labels relevant for each comparison of interest. The tests were restricted to a time window between 100 and 250 ms after sound onset as this covers the typical latency of the MMNm (Näätänen et al., 2007). To check the presence of MMNm responses, the ERFs of deviants and standards were contrasted for each feature and condition. To evaluate the effect of stimulus entropy, the MMNm difference waves in the HE and LE conditions were contrasted for each feature. Since separate tests were performed for each feature, a Bonferroni correction for multiple comparisons was applied by multiplying Monte Carlo estimates by four.

In addition to the permutation tests, secondary analyses on mean gradient amplitudes (MGA) were run, in order to determine if MMNm amplitudes for different features were affected differently by stimulus entropy. MGAs were computed for each participant, feature and condition, by averaging 50 ms around the peak amplitude, which was defined as the highest local maxima of the MMNm difference wave between 100 and 250 ms. This procedure was restricted to the average of the four combined gradiometers in each hemisphere with the largest M50 responses in the grand average (Figure 2). The differences in MGA between the two conditions were computed and submitted to a two-way repeated-measures ANOVA with feature and hemisphere as factors.

2.2. Behavioral experiment

In this experiment, a deviance detection task was used to confirm behaviorally the hypothesized neurophysiological effects. Moreover, since the MEG experiment included only two highly contrasting conditions in order to observe clear differences in the neural signal, in the behavioral experiment we took the opportunity to assess a more fine-grained precision modulation of prediction error.

2.2.1. Participants

Twenty-one non-musicians (16 women, mean age 21.9, range 18-36) participated in the experiment. Musical expertise was measured with the Gold-MSI musical training subscale which yielded a score of 12.9 (SD = ±5.77). All values lay within the 42nd percentile of the norm for this subscale. Participants were recruited through an online database for experiment participation, agreed to take part voluntarily, gave their informed consent and received 100 Danish kroner (approximately 13.5 Euros) as compensation. Two subjects had previously participated in the MEG experiment.

2.2.2. Experimental design

Participants were presented with 32-note sequences with different levels of entropy and asked to decide after each one if a mistuning was present in the sequence or not, and how certain they were about their response on a scale from 1 (not certain at all) to 7 (completely certain). Five conditions with different degrees of entropy were included (Figure 1b; the full stimulus set is shown in supplementary materials 1). As in the MEG experiment, there was an LE condition consisting of an Alberti bass sequence. Similarly, there was an HE condition corresponding to a subset of five of the six melodies used in the MEG experiment. Three intermediate conditions (IE) were added in order to test for more fine-grained effects of entropy (Figure 1b). The alphabet of these conditions was restricted to a C-major scale spanning eight tones from C4 to C5. Based on previous research showing an effect of pitch alphabet on the uncertainty of auditory sequences (Auksztulewicz et al., 2017; Barascud et al., 2016), we conjectured that these sequences would have higher mean entropy than LE sequences, which spanned only three pitch categories, and lower mean entropy than HE sequences, which spanned up to fifteen pitch categories. For the three intermediate conditions, entropy was manipulated by changing their repetitiveness. Thus, the least uncertain of the three (IE1) was an eight-note pattern repeated four times. Since each pitch of this pattern was completely determined by its predecessor, entropy was low but still higher than for LE sequences, given the larger pitch alphabet. The middle condition (IE2), which consisted of five melodies, relaxed the constraint for exact repetition leading to reduced precision over the IE1 condition. Finally, the most uncertain of the three conditions (IE3) consisted of random orderings of the eight tones, with equal probability and without playing any of them twice in a row. These sequences were generated individually for each participant. Since sequential constraints are minimal in this condition, it was expected to have higher entropy levels than IE1 and IE2, but lower levels than HE, given its smaller pitch alphabet.

The conjectured order of entropy levels (LE < IE1 < IE2 < IE3 < HE) was confirmed by IDyOM's mean entropy values (Table 1), estimated as described in section 2.1.2.1. Mean IC values followed a similar pattern (LE < HE < IE1 < IE2 < IE3), with the only exception that HE had lower IC than all the intermediate conditions in this case. This might reflect the fact that HE melodies tended to have smaller pitch intervals (mostly 1- or 2-semitone steps) which are more expected than larger intervals in Western tonal music (Huron, 2006) (supplementary materials 2). The entropy values presented here were not meant to be used as predictors in the analyses. Rather, they were used as approximate estimates to order the conditions and assess the possibility of a linear trend in the data. Note that, since IE3 (random) sequences were unique for each subject, in this case entropy and IC values were estimated only for one of the participants, as a representative sample. These are the ten sequences shown in supplementary materials 1.

To simplify the stimuli and make them comparable between conditions, only sequences in the C-major key were included. Target sequences were created by randomly choosing a tone from the second half of each sequence and mistuning it 25 cents up. The deviations were smaller than in the MEG experiment to avoid ceiling effects observed during piloting. Only pitch deviants were included since this feature showed the strongest reduction in amplitude between conditions in the neurophysiological data (see section 3.1.2). The creation of the pool of standard and deviant tones followed the same procedure as in the MEG experiment. Ten targets and ten foil sequences were presented for each condition in a random order. Note that the number of possible sequences or melodies differed between conditions which meant that they were repeated a different number of times. Thus, LE and IE1 consisted of only one sequence, and for this reason they were repeated ten times as targets and ten times as foils. In contrast, IE2 and HE consisted of five different sequences, which entailed that they were repeated twice as targets and twice as foils. Finally, since there were ten unique IE3 sequences for each participant, they were played only once as foils and once as targets. There were four practice trials at the beginning of the session. The complete procedure lasted approximately 25 minutes.

2.2.3. Statistical analyses

A *d'*-score was computed for each participant and condition. In the few cases where a participant scored 100% of hits or 0% of false alarms, values were adjusted to 95% or 5%, respectively. This prevented the *z* scores in the *d'* estimation from reaching infinity. Statistical analyses were performed using the software *R* (R Core Team, 2017). In order to assess the effects of stimulus entropy, different mixed models were fitted using the *lme* function from the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018). The models were compared through likelihood ratio testing. All of them allowed random intercepts for the participants. For the analyses of the confidence scores, ordinal logistic regression was used in the form of a cumulative link mixed model (Christensen, 2015a), as implemented by the function *clmm* from the *ordinal* package (Christensen, 2015b). Log-odds ("logit") were used as the link function. The estimated parameters, random effects and model comparisons were the same as in the analysis of *d'*-scores, with the difference that in these models there was an intercept for each of the six cut-points between response categories in the dependent variable (see Table 3). Finally, post-hoc, Bonferroni-corrected pairwise comparisons between conditions were conducted for confidence scores with the functions *lsmeans* (from the *lsmeans* package, Lenth, 2016) and *contrast* (built-in to R).

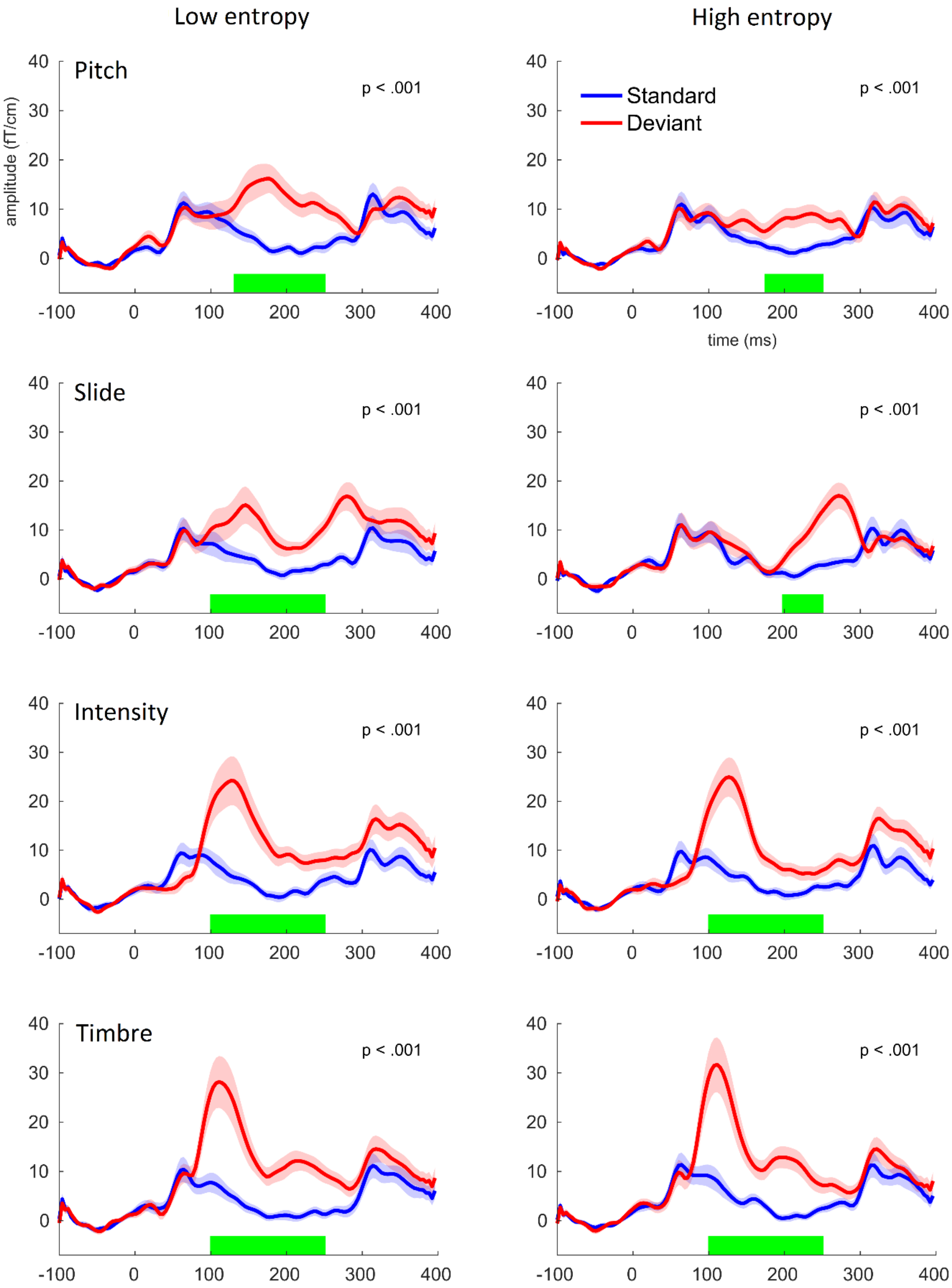

Figure 3. Standard and deviant ERFs for each feature and condition. These correspond to an average of four temporal channels in the right hemisphere (see Figure 2). Shaded areas indicate standard error. Green bars indicate the most prominent cluster from the permutation test.

3. Results

3.1. MEG experiment

3.1.1. Presence of the MMNm

Significant differences were found between standard and deviant ERFs in the 100-250 ms post-stimulus time window for all features in both conditions (Figure 3). In all cases, the Monte Carlo estimate was smaller than .001. The differences were present bilaterally, were largest over right temporal gradiometers, and showed a polarity—as observed in the magnetometers—consistent with previous reports of the MMNm (e.g. Bonetti, Haumann, Vuust, Kliuchko, & Brattico, 2017) (Figure 4).

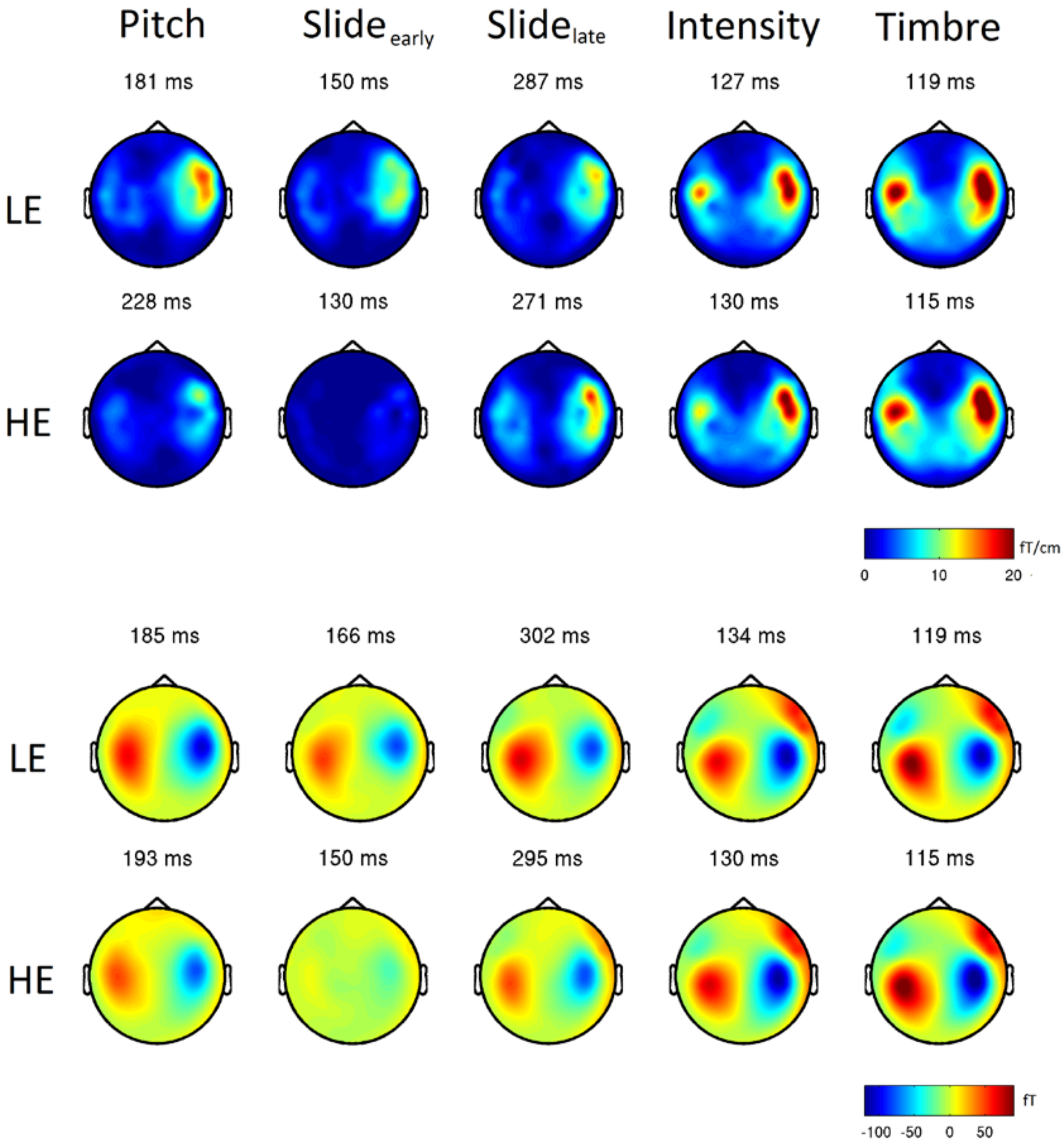

Figure 4. Topographic maps of the MMNm (difference between standards and deviants) for gradiometers (top) and magnetometers (bottom). Peak latencies are shown above each plot. The displayed activity corresponds to an average of 50 ms around the peak. Slide MMNm topographies are shown for activity around the peak in both early (100-250 ms) and late (250-350 ms) time windows. LE = low entropy, HE = high entropy.

3.1.2. Low-entropy vs. high-entropy stimuli

A strong amplitude reduction in the MMNm difference waves was found for HE as compared to LE stimuli, for pitch and slide deviants bilaterally (Figure 5). This reduction was maximal at temporal gradiometers. No significant differences were found for intensity or timbre. The analysis of the mean gradient amplitude (MGA) differences between HE and LE contexts revealed a significant main effect of feature, $F_{(3,69)} = 4.58, p = .006, \eta_G^2 = 0.05$, but not of hemisphere, $F_{(1,23)} = 0.19, p = .67$, $\eta_G^2 = .001$, and an interaction between feature and hemisphere, $F_{(3,69)} = 4.22, p = .008, \eta_G^2 = .03$. A one-way ANOVA for each hemisphere revealed a significant effect of feature in the right, $F_{(3,69)} = 7.23, p < .001, \eta_G^2 = .11$, but not in the left hemisphere, $F_{(3,69)} = 1.13$, $p = .34, \eta_G^2 = .03$. Bonferroni-corrected pairwise comparisons for right-hemisphere MGAs revealed significant differences between pitch and timbre, $p < .001, z = 3.95$, pitch and intensity, $p = .04, z = 2.74$, and slide and timbre $p = .002, z = 3.592$. A trend towards a difference was observed for slide and intensity, $p = .09, z = 2.40$. No significant differences were found between timbre and intensity, $p = 1, z = -1.20$ or pitch and slide, $p = 1, z = 0.36$.

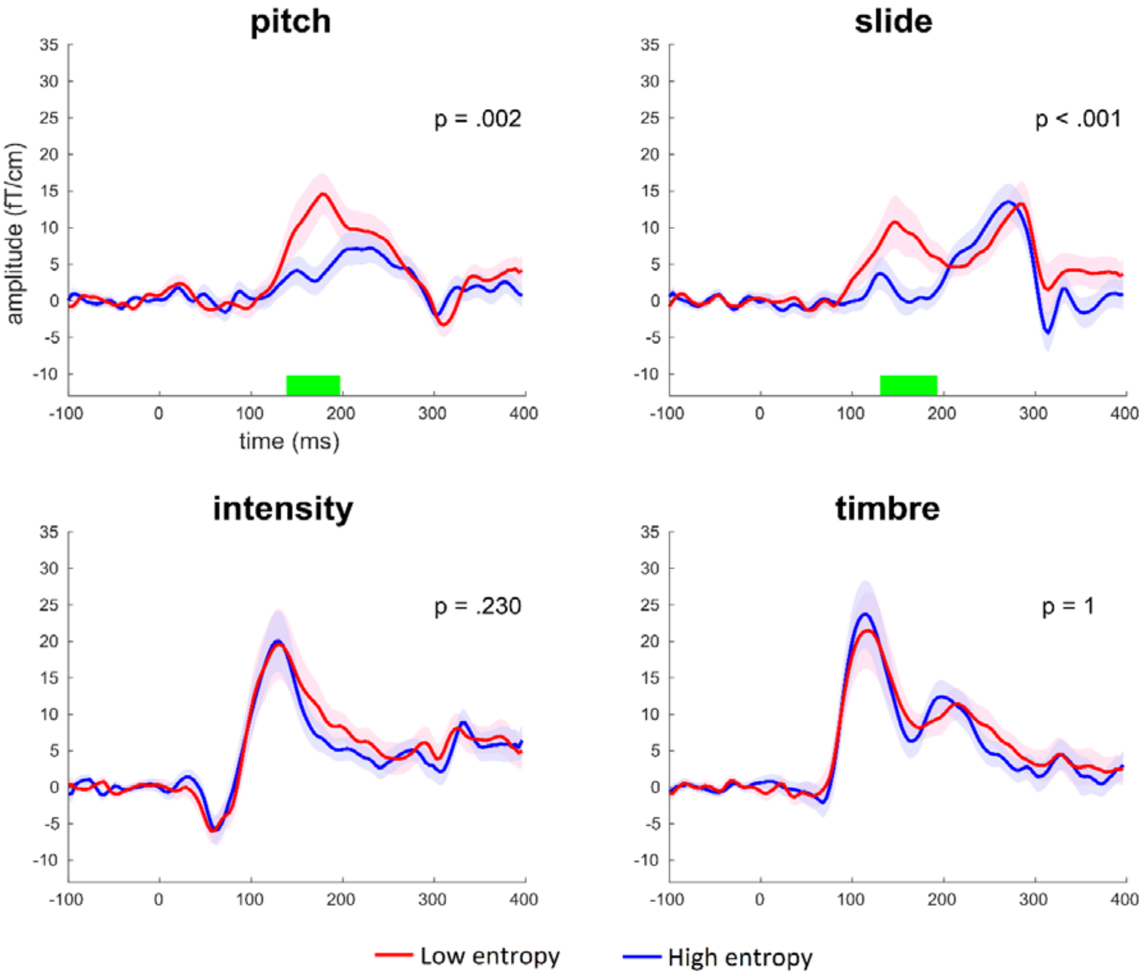

Figure 5. MMNm difference waves according to the two conditions. The ERFs correspond to an average of four right temporal gradiometers (see Figure 2). Shaded areas depict standard error. Green bars indicate the most prominent cluster from the permutation tests revealing a significant difference in MMNm amplitude between conditions.

Interestingly, the slide MMNm displayed an unusual shape (Figures 3 and 5), which motivated us to perform a further post hoc analysis. Specifically, the MMNm was sustained beyond 250 ms and peaked around 280 ms. The polarity of this late response, as revealed by the magnetometers, was the same as that of the MMNm for the other features (Figure 4). Furthermore, stimulus entropy seemed to affect the earlier portion of the ERF more than the later part. A follow-up permutation test in the 250-350 ms time-window for the slide MMNm revealed no significant differences between HE and LE contexts (*p* = .16). Furthermore, a two-sided paired-samples *t*-test comparing the MGA differences in the early (100-250 ms) and late (250-350 ms) time windows revealed significantly higher differences between conditions for the former (M= 5.34, SD = 9.45) than for the latter (M = 0.66, SD = 7.48); *t* = 2.98, *p* = .006.

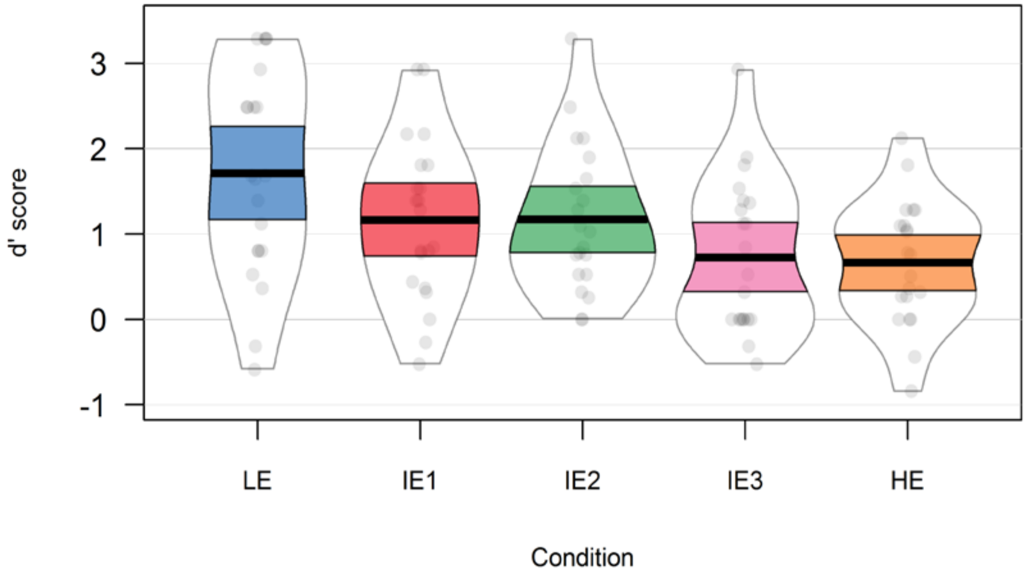

Figure 6. Pirate plot of *d'*-scores. Colored boxes depict 95% confidence intervals.

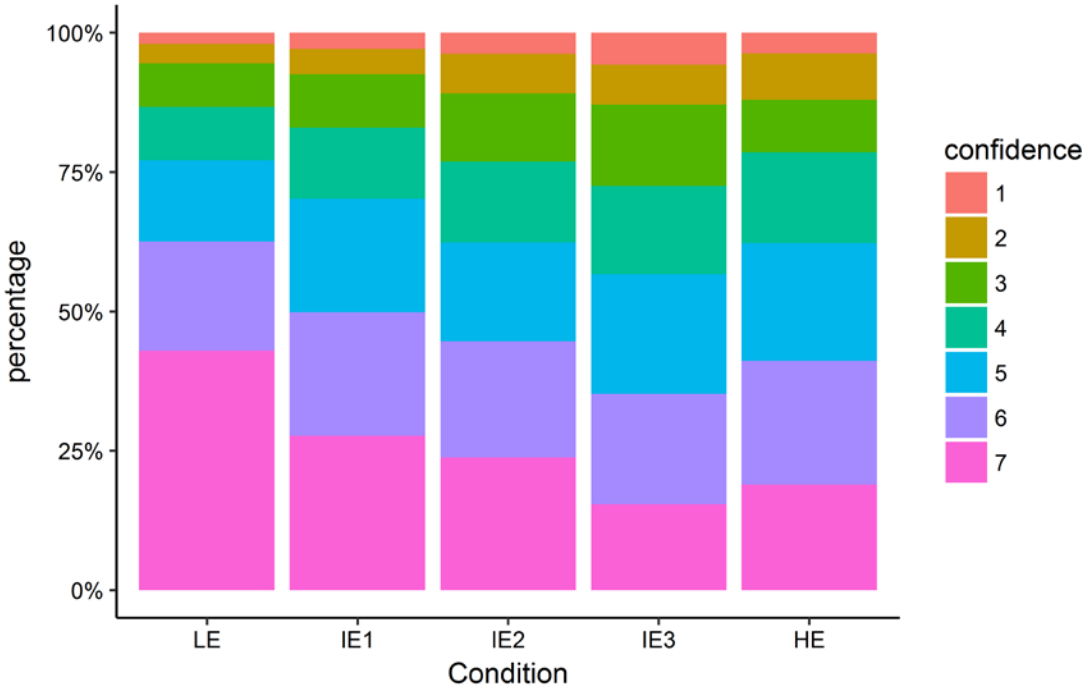

Figure 7. Percentage of responses for each category of the confidence ratings.

3.2. Behavioral experiment

For the analysis of *d'*-scores, a null model was compared with a linear model in which the conditions were ordered according to entropy values (see Section 2.2.2). The results revealed a significant effect of entropy ($\chi^2 = 33.55$, $p < .001$). To assess the validity of this model, we compared it with a free-curve model in which the conditions were treated as an unordered categorical factor. The test revealed no significant differences ($\chi^2 = 4.69$, $p = .20$), which suggests that the linear model is indeed a good fit for the data. Note, however, that by using linear modeling in this way it is assumed that entropy decreases by the same amount between consecutive conditions. While this might not necessarily be true, we regard this approach as adequately justified according to the hypothesized decreasing trend. In sum, *d'*-scores could be best explained with a linear model that included participant as random effect and a fixed effect of entropy, with conditions ordered according to the estimated values (Figure 6; parameters are reported in Table 2). A visual inspection revealed that the residuals of this model were normally distributed.

Table 2. Parameters for the mixed effects model of *d'*-scores.

Parameter	β	95% C.I.	t
Intercept (β_0)	1.85	[1.44, 2.26]	8.87
Condition (β_1)	-0.25	[-0.33, -0.17]	-6.36
Random intercept participant ($\mathcal{E}$)	Standard deviation = 0.73		

$$\text{Model: } \boldsymbol{D_c = \beta_0 + \beta_1 c + \mathcal{E}}$$

Table 3. Model parameters (A) and cut-points (B) for the cumulative link mixed model for confidence ratings. The parameters are expressed as proportional cumulative log-odds.

A)

Parameter	β_C	95% C.I.	z
Condition (IE1)	-0.70	[-0.88, -0.53]	-7.86
Condition (IE2)	-1.04	[-1.22, -0.97]	-11.64
Condition (IE3)	-1.46	[-1.64, -1.28]	-16.28
Condition (HE)	-1.14	[-1.32, -0.97]	-12.85
Random intercept ($\mathcal{E}$)	Standard deviation = 1.32		

B)

Cut-point	θ_r	95% C.I.	z
1|2	-5.14	[-5.59, -4.69]	-22.39
2|3	-3.76	[-4.19, -3.34]	-17.45
3|4	-2.63	[-3.05, -2.22]	-12.48
4|5	-1.71	[-2.12, -1.30]	-8.17
5|6	-0.67	[-1.07, -0.26]	-3.21
6|7	0.55	[0.15, 0.96]	2.67

$$\text{Model: } \boldsymbol{\log\left(\frac{P(Y_c \leq r)}{1 - P(Y_c \leq r)}\right) = \theta_r - \beta_c - \mathcal{E}}$$

Following the same procedure for the confidence scores, it was found that the unordered model explained the data better than the null model ($\chi^2 = 225.99$, $p < .001$). However, unlike for the *d'*-scores, the unordered model did explain the data better than the linear model ($\chi^2 = 84.53$, $p < .001$). This means that no strict linear trend can be inferred from the analysis. However, post-hoc pairwise comparisons revealed that, in most cases, significantly lower confidence ratings were present in conditions with lower entropy (Table 4). One of the exceptions was the comparison between IE2 and HE, which yielded no significant differences. The other exception was the higher confidence ratings for HE than IE3, since the opposite was found for *d'*-scores. In sum, the confidence scores could be best explained by a model of proportional odds that included participant as a random effect, intercepts for each of the cut-points of the response variable, and a fixed effect of entropy as an unordered categorical factor (Figure 7; parameters are reported in Table 3).

Table 4. Bonferroni-corrected pairwise comparisons for the model of confidence ratings.

Comparison	z-value	p
LE - IE1	7.86	<.001
LE - IE2	11.64	<.001
LE - IE3	16.28	<.001
LE - HE	12.85	<.001
IE1 - IE2	3.99	<.001
IE1 - IE3	8.92	<.001
IE1 - HE	5.21	<.001
IE2 - IE3	4.95	<.001
IE2 - HE	1.19	1
IE3 - HE	-3.80	.002

4. Discussion

In the present study, we provide neurophysiological and behavioral evidence consistent with a precision modulation of prediction error for pitch-related musical features. We found decreased MMNm amplitudes for pitch and slide deviants in high-entropy (HE) as compared to low-entropy (LE) contexts. This modulation was paralleled by accuracy and confidence scores in a behavioral deviance detection task, which both decreased with higher entropy levels. In line with recent theories of predictive processing (Clark, 2013; Feldman & Friston, 2010; Hohwy, 2012) and models of musical expectations (Hansen et al., 2017; Ross & Hansen, 2016; Vuust et al., 2018), this indicates a down-weighting of prediction error due to the reduction in the precision of predictive models.

Our results are consistent with empirical research already showing reduced auditory prediction error responses in uncertain contexts. Garrido and colleagues (2013) found less pronounced neural responses to outlier tones, when these were embedded in rapid tone sequences whose frequencies followed broad (i.e., imprecise) as compared to narrow (i.e., precise) normal distributions. Similarly, Sohoglu and Chait (2016) found that the appearance of a new stream in an auditory scene was more salient and generated a larger neural response when the context was temporally regular than when it was not. Furthermore, some electrophysiological studies have found increased sustained tonic activity—as opposed to phasic responses such as the MMN—in low- as compared to high-entropy contexts (Auksztulewicz et al., 2017; Barascud et al., 2016), although the opposite pattern has been found in fMRI experiments (Nastase, Iacovella, & Hasson, 2014; Overath et al., 2007). Closer to assessing the precision-weighting mechanism in a musical context, recent research shows that the entropy of short rhythmic sequences modulates MMN responses (Lumaca, Haumman, Brattico, Grube, & Vuust, in press). However, in this study rhythms were presented as repeated short patterns, which makes them less akin to actual musical stimuli than our HE sequences. Together with studies showing an absence of MMN responses in random contexts (Hsu et al., 2015; Jacobsen & Schröger, 2001; Maess, et al., 2007), these observations yield some support for a precision-weighting mechanism in the auditory domain. However, we are the first to show that the mechanism applies to music—a common, highly structured, and more ecologically valid type of auditory stimulus.

The uncertainty of the stimuli was quantitatively estimated with IDyOM. This method confirmed higher mean entropy and mean IC for HE as compared to LE sequences. This can be clearly observed in individual tone sequences: for LE, IC drops after the first occurrence of the pattern whereas HE sequences tend to have higher IC levels throughout (supplementary materials 1). The higher entropy of HE stimuli is mainly caused by three factors. First, HE sequences are much less repetitive, which may make upcoming tones harder to predict from previous tones in the local context. Second, HE sequences spanned a larger pitch alphabet, which entails that the relative frequency of occurrence of single tones generally became lower. While this factor most likely played a role in some conditions of the behavioral experiment, its influence on the MEG results is not as clear, given that LE sequences were transposed to cover approximately the same range as HE sequences (supplementary materials 2). Nevertheless, it is still possible that the local restriction of the pitch alphabet to only three pitch categories within each 32-note sequence generated the difference in entropy levels. Finally, the alphabet of possible pitch intervals (i.e., the distances between consecutive pitches in semitones) was definitely larger in the HE condition. This might have also contributed to the increase in entropy. Note, however, that, as shown in supplementary materials 3, in both cases most of the occurrences were concentrated in only a few pitch intervals. The factors just mentioned are in line with research showing that both the alphabet size and the repetitiveness of auditory stimuli affect their uncertainty (Auksztulewicz et al., 2017; Barascud et al., 2016). Further research is needed to disentangle exactly in what way these factors contribute to the precision modulation of musical prediction error.

Interestingly, the observed MMNm amplitudes were more affected by entropy for pitch and slide as compared to timbre and intensity deviants. This is supported by the lack of significant differences for the latter two features in the permutation tests. We attribute this to the fact that we manipulated entropy in the pitch dimension, while other dimensions were restricted to the same two standard and deviant values (piano timbre vs. telephone receiver; high intensity vs. low intensity) across all conditions. The fact that this effect was found only in right temporal gradiometers might be due to the MMNm signal being largest in these sites, which in turn is consistent with the rightward asymmetry for music processing (e.g. Brattico et al. 2006; Koelsch et al., 2000; Zatorre, Belin, & Penhune, 2002). Consequently, it seems that the statistical properties of the stimuli particularly affected deviants that depend on pitch information, which points to a feature-specific precision-weighting of the MMN. This interpretation is consistent with MMN recordings in multi-feature (Näätänen, Pakarinen, Rinne, & Takegata, 2004; Vuust et al., 2011) and no-standard (Kliuchko, Heinonen-Guzejev, Vuust, Tervaniemi, & Brattico, 2016; Pakarinen, Huotilainen, & Näätänen, 2010) paradigms in which auditory regularities are created for a specific feature even though sounds constantly change in other features.

Furthermore, the suggested feature-selectivity is particularly interesting in the case of the slide MMNm, which had an unusual shape that extended and peaked beyond 250 ms. This shape can be attributed to the fact that, unlike in previous experiments, the pitch glide spanned the whole duration of the tone and thus the MMNm amplitude seemed to mirror the increasing magnitude of the continuous deviation (taking the initial pitch of the deviant tone as the reference). The fact that only the earlier portion of the response was different between conditions indicates that the first part of the deviant was harder to detect in HE sequences. This, in turn, might reflect the coexistence of two violations. Since the slide deviant started two semitones below its corresponding standard, we propose that the first section of the ERF is a pitch MMNm, while the second corresponds to a proper slide MMNm. Thus, in the LE block, where there were much more precise pitch expectations, slide deviants were heard first as a "wrong" pitch and afterwards as a pitch glide. In contrast, for the HE sequences, the sense of a "wrong" pitch would be weaker but the gliding quality of the tone would be equally surprising. If this account is correct, the fact that the first (pitch) but not the second (slide proper) part is reduced for HE stimuli fits nicely with the idea of a feature-specific precision-weighting mechanism. In any case, the differences between features discussed above have to be taken with caution since they constitute a non-hypothesized finding. Moreover, we did not change precision in dimensions other than pitch and in the behavioral experiment we did not include deviants other than mistunings. These two manipulations provide an interesting path of enquiry for the future, with the potential to properly elucidate the feature selectivity suggested by our data.

The results from the behavioral experiment showed that listeners are sensitive even to fine-grained changes in the entropy of auditory sequences, which provides stronger support for the precision-weighting mechanism. Apart from finding lower *d'*-scores and confidence ratings for HE compared to LE sequences—thus confirming the MEG results—we found a decreasing trend for the five conditions which indicated lower *d'*-scores as the entropy of the sequences increased. A similar pattern was observed in the pairwise comparisons of confidence scores which showed that participants tended to be more confident in their judgements in the less uncertain conditions. Moreover, there was a difference between the two outcome measures: confidence ratings for HE sequences were higher than ratings for IE3 sequences, whereas the opposite was true for *d'*-scores. A possible explanation for this is that HE sequences (i.e. melodies) follow the properties of Western tonal music more so than the (random) IE3 sequences. Thus, we may speculate that confidence reflects the agreement of the stimuli with participants' long-term knowledge of Western tonal music, which is more consistent with the lower IC values found for HE as compared to the IE3 sequences. Actual deviance detection, on the other hand, might depend more on the uncertainty of the signal, which is more consistent with the entropy estimations.

When behavioral and MEG results are considered together, a coherent picture emerges in which precision modulates prediction error responses. However, the relationship between our two experiments needs to be taken with caution since the behavioral task required the participants to actively detect deviations, while in the MEG session they listened to the sounds passively while watching a silent movie. Thus, there were additional higher-order processes involved in the former which means that the differences in *d'*-scores and confidence ratings cannot be ascribed exclusively to the processes reflected in the MMN. Further research involving active tasks and neurophysiological recordings is needed to assess the contribution of different components and processing stages to the precision-weighting effect.

The work presented here has some limitations. For example, in the MEG experiment, we compared two types of stimuli that differ in several aspects. As mentioned before, factors such as repetitiveness, and the pitch and pitch-interval alphabets might have all played a role in the differences found, and even though all of them shape the entropy of the sequences, it is not possible to disentangle their individual contributions from the present dataset. It has to be noted, however, that the influence of pitch alphabet was controlled in the behavioral experiment in the three intermediate conditions, which nevertheless showed a decreasing trend in *d'*-scores and confidence ratings. Another aspect to consider is the repetition rate of individual sequences. In the LE condition of the MEG experiment, individual Alberti bass sequences were repeated every two trials, whereas individual melodies in the HE condition were repeated every twelve trials. This might have created a stronger long-term memory representation for the LE sequences which might be responsible at least for part of the effect that we found. Something similar might have happened in the behavioral experiment since individual sequences were repeated a different number of times for different conditions. However, this possible confound cannot be regarded as the only explanation for our results since we found differences for conditions with equal repetition rates (e.g. LE and IE1). Furthermore, this interpretation is compatible with the precision-weighting explanation if one regards these long-term representations as very precise expectations. Finally, even though acoustic confounds were minimized in the MEG experiment by using the same pitch alphabet in both conditions, it is still possible that the different zeroth-order probabilities of individual pitches created an acoustic difference between them. For example, since pitch discrimination is more difficult at very low or very high frequency ranges (Sek & Moore, 1995), a difference between conditions in the occurrence of high or low pitches might as well explain our results. However, this explanation is unlikely since in both conditions the range of frequencies was similarly covered (supplementary materials 2) and even though there was a small tendency for LE sequences to cover higher tones, their pitches were not higher than 5 kHz, which is the frequency range for which discrimination significantly decreases.

Importantly, our results constitute an advance in the multi-feature experimental paradigms used to study prediction error responses through the MMN, particularly in areas related to music perception. We demonstrate here that it is possible to obtain reliable MMN signals with more complex and musically realistic stimuli in a multi-feature paradigm. We have done so by exploiting the possibilities of abstract-feature MMN responses (Paavilainen, 2013), which arise, not from the violation of an exact sensory representation established through the repetition of a stimulus, but from the breach of an abstract regularity established in a constantly changing auditory stream. Thus, even though tones in the HE condition constantly changed, loudness, timbre, the tuning system and the use of steady pitches—as opposed to pitch glides—were kept constant, thus providing regularities for which violations generated MMNm responses. In consequence, our paradigm opens up the possibility of asking more detailed questions related to a wider range of aspects of complex auditory sequence processing. For example, one could use the MMN to assess whether cognitive capacities are related to the effect of stimulus complexity on neural responses at the individual level. Or one could aim to determine whether sound features specific to a musical style are processed differently by people with different musical backgrounds or belonging to different cultures. Therefore, the work presented here moves us one step closer to understanding auditory predictive processing in the rich environments of daily life.

5. Conclusion

In this study, we provide converging neurophysiological and behavioral evidence consistent with precision-weighting of prediction error in the auditory domain. Our results show that prediction error responses—as indexed by the MMNm, accuracy scores, and confidence ratings—are reduced in auditory contexts with higher levels of entropy and suggest that this reduction is constrained to features that depend on the auditory dimension whose uncertainty is manipulated. Thus, in line with recent theories of predictive processing, we provide further support to precision-weighted prediction error as a fundamental principle for brain function and open new possibilities for the study and understanding of predictive processing in the complex auditory contexts of real life.

Acknowledgments

We wish to thank the project initiation group, namely Christopher Bailey, Torben Lund and Dora Grauballe, for their help with setting up the experiments. We also thank Nader Sedghi, Massimo Lumaca, Giulia Donati, Ulrika Varankaité, Giulio Caraturro and Claudia Iorio for assistance during MEG recordings. We are indebted as well to the group of Italian trainees from IISS Simone-Morea, Conversano, who helped with the behavioral experiment. Finally, we thank Hella Kastbjerg for checking the English language of this manuscript. The Center for Music in the Brain is funded by the Danish National Research Foundation (DNRF 117).

References

Auksztulewicz, R., Barascud, N., Cooray, G., Nobre, A. C., Chait, M., & Friston, K. (2017). The Cumulative Effects of Predictability on Synaptic Gain in the Auditory Processing Stream. *The Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0291-17.2017

Auksztulewicz, R., & Friston, K. (2015). Attentional enhancement of auditory mismatch responses: A DCM/MEG study. *Cerebral Cortex*, *25*(11), 4273–4283. https://doi.org/10.1093/cercor/bhu323

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1181–2. https://doi.org/10.1098/rstb.2008.0321

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, *113*(5), E616–E625. https://doi.org/10.1073/pnas.1508523113

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Bonetti, L., Haumann, N. T., Vuust, P., Kliuchko, M., & Brattico, E. (2017). Risk of depression enhances auditory Pitch discrimination in the brain as indexed by the mismatch negativity. *Clinical Neurophysiology*, *128*(10), 1923–1936. https://doi.org/10.1016/j.clinph.2017.07.004

Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, *1117*(1), 162–174. https://doi.org/10.1016/j.brainres.2006.08.023

Carrus, E., Pearce, M. T., & Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex*, *49*(8), 2186–200. https://doi.org/10.1016/j.cortex.2012.08.024

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibanez, A., … Bekinschtein, T. A. (2013). Expectation and Attention in Hierarchical Auditory Prediction. *Journal of Neuroscience*, *33*(27), 11194–11205. https://doi.org/10.1523/JNEUROSCI.0114-13.2013

Christensen, R. (2015a). Analysis of ordinal data with cumulative link models. Retrieved from cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf.

Christensen, R. (2015b). Package ‘ ordinal .’ *Cran*, 1–61.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing Uncertainty*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, (24), 51–73.

Deouell, L. Y. (2007). The frontal generator of the mismatch negativity revisited. *Journal of Psychophysiology*. https://doi.org/10.1027/0269-8803.21.34.188

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*(APR). https://doi.org/10.3389/fpsyg.2012.00096

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press UK.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00548

Feldman, H., & Friston, K. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00215

Fuller, D. (2001). Alberti bass. *Grove Music Online*. https://doi.org/10.1093/gmo/9781561592630.article.00447

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2008.11.029

Garrido, M. I., Rowe, E. G., Halász, V., & Mattingley, J. B. (2017). Bayesian Mapping Reveals That Attention Boosts Neural Responses to Predicted and Unpredicted Stimuli. *Cerebral Cortex*, 1–12. https://doi.org/10.1093/cercor/bhx087

Garrido, M. I., Sahani, M., & Dolan, R. J. (2013). Outlier Responses Reflect Sensitivity to Statistical Structure in the Human Brain. *PLoS Computational Biology*, *9*(3). https://doi.org/10.1371/journal.pcbi.1002999

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, (7 DEC). https://doi.org/10.3389/fnins.2013.00267

Hansen, N. C., Dietz, M., & Vuust, P. (2017). Commentary: Predictions and the brain: how musical sounds become rewarding. *Frontiers in Human Neuroscience*. https://doi.org/doi: 10.3389/fnhum.2017.00168

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*(SEP). https://doi.org/10.3389/fpsyg.2014.01052

Hansen, N. C., Vuust, P., & Pearce, M. (2016). “If you have to ask, you’ll never know”: Effects of specialised stylistic expertise on predictive processing of music. *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0163584

Haumann, N. T., Parkkonen, L., Kliuchko, M., Vuust, P., & Brattico, E. (2016). Comparing the Performance of Popular MEG/EEG Artifact Correction Methods in an Evoked-Response Study. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2016/7489108

Heilbron, M., & Chait, M. (2017). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2017.07.061

Hsu, Y.-F., Le Bars, S., Hamalainen, J. A., & Waszak, F. (2015). Distinctive Representation of Mispredicted and Unpredicted Prediction Errors in Human Electroencephalography. *Journal of Neuroscience*, *35*(43), 14653–14660. https://doi.org/10.1523/JNEUROSCI.2204-15.2015

Huron, D. (2006). *Sweet anticipation: Music and the Psychology of Expectation*. *Exposure* (Vol. 443). https://doi.org/10.1525/mp.2007.24.5.511

Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*(4), 723–727. https://doi.org/10.1017/S0048577201000993

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain. *Journal of Neuroscience*, *33*(47), 18438–18447. https://doi.org/10.1523/JNEUROSCI.3308-13.2013

Kliuchko, M., Heinonen-Guzejev, M., Vuust, P., Tervaniemi, M., & Brattico, E. (2016). A window into the brain mechanisms associated with noise sensitivity. *Scientific Reports*, *6*. https://doi.org/10.1038/srep39236

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain Indices of Music Processing: “Nonmusicians” are Musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541.

https://doi.org/10.1162/089892900562183

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, *22*(9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v069.i01

Lieder, F., Stephan, K. E., Daunizeau, J., Garrido, M. I., & Friston, K. (2013). A Neurocomputational Model of the Mismatch Negativity. *PLoS Computational Biology*, *9*(11). https://doi.org/10.1371/journal.pcbi.1003288

Lumaca, M., Haumman, N., Brattico, E., Grube, M.,Vuust, P. (in press). An information theoretic characterization of neural predictive coding in rhythm perception.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Maess, B., Jacobsen, T., Schröger, E., & Friederici, A. D. (2007). Localizing pre-attentive auditory memory-based comparison: Magnetic mismatch negativity to pitch change. *NeuroImage*, *37*(2), 561–571. https://doi.org/10.1016/j.neuroimage.2007.05.040

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0089642

Nastase, S., Iacovella, V., & Hasson, U. (2014). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Human Brain Mapping*. https://doi.org/10.1002/hbm.22238

Näätänen, R., Gaillard, A. W. K., & Mantysalo, S. (1978). Early Selective Attention Effect on Evoked Potential Reinterpreted. *Acta Psychologica*.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, *115*(1), 140–144. https://doi.org/10.1016/j.clinph.2003.04.001

Omigie, D., Pearce, M. T., Williamson, V. J., & Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia*, *51*(9), 1749–1762. https://doi.org/10.1016/j.neuropsychologia.2013.05.010

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Overath, T., Cusack, R., Kumar, S., Von Kriegstein, K., Warren, J. D., Grube, M., … Griffiths, T. D. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biology*, *5*(11), 2723–2732. https://doi.org/10.1371/journal.pbio.0050288

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2013.03.015

Pakarinen, S., Huotilainen, M., & Näätänen, R. (2010). The mismatch negativity (MMN) with no standard stimulus. *Clinical Neurophysiology*, *121*(7), 1043–1050. https://doi.org/10.1016/j.clinph.2010.02.009

Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. City University of London, UK.

Pearce, M. T. (2018). Statistical Learning and Probabilistic Prediction in Music Cognition. *Annals of the New York Academy of Sciences*, *1423*, 378–395. https://doi.org/10.1111/nyas.13654

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, *50*(1), 302–313. https://doi.org/10.1016/j.neuroimage.2009.12.019

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). nlme: Linear and Nonlinear Mixed Effects Models.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, *24*(2), 560–564. https://doi.org/10.1016/j.neuroimage.2004.08.039

Ross, S.*, & Hansen, N. C.* (2016). Dissociating Prediction Failure: Considerations from Music Perception. *Journal of Neuroscience*, *36*(11), 3103–3105. https://doi.org/10.1523/JNEUROSCI.0053-16.2016

Sek, A., & Moore, B. C. J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.411968

Schröger, E., Marzecová, A., & Sanmiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

Sohoglu, E., & Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *ELife*, *5*(Se). https://doi.org/10.7554/eLife.19113

Taulu, S., Kajola, M., & Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topography*, *16*(4), 269–275. https://doi.org/10.1023/B:BRAT.0000032864.93890.f9

Tervaniemi, M., Huotilainen, M., & Brattico, E. (2014). Melodic multi-feature paradigm reveals auditory profiles in music-sound encoding. *Frontiers in Human Neuroscience*, *8*(July), 496. https://doi.org/10.3389/fnhum.2014.00496

Vuust, P., Brattico, E., Glerean, E., Seppänen, M., Pakarinen, S., Tervaniemi, M., & Näätänen, R. (2011). New fast mismatch negativity paradigm for determining the neural prerequisites for musical ability. *Cortex*, *47*(9), 1091–1098. https://doi.org/10.1016/j.cortex.2011.04.026

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia*, *50*(7), 1432–1443. https://doi.org/10.1016/j.neuropsychologia.2012.02.028

Vuust, P., Liikala, L., Näätänen, R., Brattico, P., & Brattico, E. (2016). Comprehensive auditory discrimination profiles recorded with a fast parametric musical multi-feature mismatch negativity paradigm. *Clinical Neurophysiology*, *127*(4), 2065–2077. https://doi.org/10.1016/j.clinph.2015.11.009

Vuust, P., Witek, M., Dietz, M., & Kringelbach, M. L. (2018). Now You Hear It: A novel predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*, 1–11. https://doi.org/10.1111/nyas.13622

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188–196. https://doi.org/10.1016/j.lindif.2010.02.004

Witek, M. A. G. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0094446

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(00)01816-7