Decomposing neural responses to melodic surprise in musicians and non-musicians: evidence for a hierarchy of predictions in the auditory system

Quiroga-Martinez D.R.[a], Hansen N.C.[b], Højlund A.[c], Pearce M.[a,d], Brattico E.[a,e], 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 of Functionally Integrative Neuroscience, Aarhus University, Denmark
d) School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
e) Department of Educational Sciences, Psychology and Communication, University of Bari Aldo Moro, Italy

Corresponding author:
David Ricardo Quiroga-Martinez
…..
Address: Nørrebrogade 44 - Building 1A - 8000 Aarhus C, Denmark
Email: dquiroga@clin.au.dk

Abstract

Neural responses to auditory surprise are typically studied with highly unexpected, disruptive sounds. Consequently, little is known about auditory prediction in everyday contexts that are characterized by fine-grained, non-disruptive fluctuations of auditory surprise. To address this issue, we used IDyOM, a computational model of auditory expectation, to obtain continuous surprise estimates for a set of newly composed melodies. Our main goal was to assess whether the neural correlates of non-disruptive surprising sounds in a musical context are affected by musical expertise. Using magnetoencephalography (MEG), auditory responses were recorded from musicians and non-musicians while they listened to the melodies. Consistent with a previous study, the amplitude of the N1m component increased with higher levels of computationally estimated surprise. This effect, however, was not different between the two groups. Further analyses offered an explanation for this finding: Pitch interval size itself, rather than probabilistic prediction, was responsible for the modulation of the N1m, thus pointing to low-level sensory adaptation as the underlying mechanism. In turn, the formation of auditory regularities and proper probabilistic prediction were reflected in later components: the mismatch negativity (MMNm) and the P3am, respectively. Overall, our findings reveal a hierarchy of expectations in the auditory system and highlight the need to properly account for sensory adaptation in research addressing statistical learning.

Keywords: Surprise, IDyOM, music, prediction, hierarchy

1. Introduction

Surprising sounds in auditory sequences generate neural prediction error responses (den Ouden, Kok, & de Lange, 2012). These are thought to reflect the degree to which internal predictive models are updated by novel information (Friston, 2005; Friston, Rosch, Parr, Price, & Bowman, 2017; Lieder, Stephan, Daunizeau, Garrido, & Friston, 2013). However, most research on auditory surprise has employed very simple and repetitive stimuli, occasionally disrupted by highly unexpected deviant sounds (Heilbron & Chait, 2018). Consequently, little is known about prediction in everyday auditory environments that are characterized by fine-grained, non-disruptive changes in auditory surprise.

One potential way to address this issue is to employ computational modeling to produce continuous estimates of auditory surprise in realistic stimuli. This approach was adopted by Omigie and colleagues (2013), who used Information Dynamics of Music (IDyOM) (Pearce, 2005, 2018), a variable-order Markov model of auditory expectation, to estimate note-by-note surprise in a set of melodies. Using electroencephalography (EEG), they found that the amplitude of the N1 component of the event related potential (ERP) became larger as the estimated surprise of the tones increased. This suggests that it is possible to record neural responses to subtle changes in auditory expectedness in more realistic settings.

Importantly, IDyOM incorporates both a short-term (*stm*) component, which derives dynamic expectations from the statistics of the current stimulus, and a long-term (*ltm*) component that derives schematic expectations from a training corpus. The latter simulates the knowledge of auditory signals that a listener acquires during her life-span and therefore could be a good model of auditory enculturation (Morrison, Demorest, & Pearce, 2018). Interestingly, behavioral studies have shown that IDyOM's surprise estimates are more strongly associated with expectedness ratings in musicians compared to non-musicians (Hansen & Pearce, 2014; Hansen, Vuust, & Pearce, 2016). This has been interpreted as an expertise-related enhancement in the accuracy of internal predictive models, thus providing support to IDyOM as a model of auditory enculturation. However, it remains unknown whether similar signatures of long-term statistical learning can also be observed in the neural responses to non-disruptive auditory surprise, as modeled with IDyOM.

In the present work, we used magnetoencephalography (MEG) to address this question by recording magnetic correlates of neural activity while musicians and non-musicians listened passively to melodic sequences. Following the results of Omigie et al., (2013), we expected larger magnetic N1 (N1m)[1] amplitudes with increasing levels of estimated surprise. Crucially, the association between surprise and neural responses was expected to be stronger in musically trained participants, which would provide evidence for a modulation of neural activity by auditory enculturation.

In order to gain a deeper understanding of the nature of the expectations reflected in the N1m, we conducted two sets of exploratory analyses. As will be seen, these analyses were crucial for the interpretation of the results from the comparison between musicians and non-musicians. First, we aimed to disentangle the contribution of individual components of the computational model by assessing the explanatory power of different IDyOM configurations. We contrasted configurations including short-term or long-term components; configurations predicting representations of pitch interval (the pitch distance between consecutive tones) and scale degree (the pitch interval between a tone and another

[1] Henceforth, an "m" will be added when referring to the magnetic counterpart of ERP components. When the "m" is omitted, we refer to the components in a more general sense, encompassing both their EEG and MEG manifestations.

tone perceived as the tonal center of the context); and configurations with different combinations of these factors (see section 3.2 for further details). These comparisons aimed to reveal, for example, the extent to which participants' expectations relied on long-term schematic knowledge relative to short-term knowledge, or on pitch interval representations relative to scale-degree representations.

Second, we compared—in sensor and source space—the N1m modulation with the magnetic counterpart of the mismatch negativity (MMNm), which is a well-studied brain response to the violation of auditory regularities (Bendixen, SanMiguel, & Schröger, 2012; Garrido, Kilner, Stephan, & Friston, 2009; Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen, Paavilainen, Rinne, & Alho, 2007). Since our dataset was recorded employing a novel MMN experimental paradigm with realistic non-repetitive melodies as stimuli (Quiroga-Martinez et al., 2019b, 2019a), it provided a valuable opportunity to compare these responses in the same subjects. This comparison aimed to determine whether the N1m modulation could be better interpreted as an MMNm, something that was not clear from the results in Omigie et al., (2013). This is interesting because, although these components have similar scalp topography and latency, the N1 is thought to reflect stimulus-specific adaptation (SSA)—a process whereby neurons become less responsive to repeated sensory stimulation—(May & Tiitinen, 2010; Risto Näätänen & Picton, 1987; Ulanovsky, Las, & Nelken, 2003), whereas the MMN has been proposed to reflect the violation of auditory predictive models (Bendixen et al., 2012; Garrido et al., 2009). Therefore, since the type of expectations modeled by IDyOM are much closer to the ones that would give rise to the MMN—or similar responses such as the Early Right Anterior Negativity (ERAN) (Koelsch, Gunter, Friederici, & Schröger, 2000)—it would be surprising to see such an early sensory component as the N1m, but not the MMNm, being modulated. This distinction is crucial when considering musical expertise, as one would expect probabilistic prediction, rather than SSA, to be modulated by accurate knowledge of the statistical regularities of a musical style.

Overall, this study sought to unveil the nature of the expectations reflected in neural responses to non-disruptive surprise, as well as the effect of expertise on them. Anticipating the results, the lack of differences between musicians and non-musicians, the fact that pitch-interval models show the best performance, and the clear dissociation between the N1m and the MMNm, all point to a rather surprising set of conclusions: that SSA, instead of probabilistic prediction, underlies the modulation of the N1m; that pitch-interval size alone better explains this effect; and that auditory regularities and higher-order probabilistic predictions are reflected in later components such as the MMNm and the P3am.

2. Methods

The data, code and materials necessary to reproduce these experiments and results are openly available at: https://osf.io/my6te/; DOI: 10.17605/OSF.IO/MY6TE

2.1. Participants

Twenty-six musicians and 24 non-musicians took part in the experiment (see Table 1 for demographics). All participants were neurologically healthy, right-handed and did not possess absolute pitch. Musical expertise was assessed with the musical training subscale of the Goldsmiths Musical Sophistication Index (GMSI) (Müllensiefen, Gingras, Musil, & Stewart, 2014). Musical skills were measured with the rhythm and melody sections of the Musical Ear Test (MET) (Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010). GMSI ($t = 17.9$, $p < .001$) and MET total scores ($t = 5.2$, $p < .001$) were significantly higher for musicians than for non-musicians. Participants were recruited

through an online database and agreed to take part in the experiment voluntarily. All participants gave informed consent and were paid 300 Danish kroner (approximately 40 euro) as compensation. Two musicians (not included in the reported demographics) were excluded from the analysis due to strong artefacts caused by dental implants. The study was approved by the Regional Ethics Committee (De Videnskabsetiske Komitéer for Region Midtjylland in Denmark) and conducted in accordance with the Helsinki Declaration.

Table 1. Participants' demographic and musical expertise information in the two experiments. M = musicians, NM = nonmusicians, Beh = behavioral.

Experiment	Group	Age (mean)	Age (SD)	Female	Male	GMSI (mean)	GMSI (SD)	MET Mel (mean)	MET Mel (SD)	MET Rhy (mean)	MET Rhy (SD)	MET Total (mean)	MET Total (SD)
MEG	NM (n=24)	26.54	3.4	13	11	10.17	3.48	33.17	5.39	35.79	5.33	69.12	9.44
	M (n=26)	24.15	2.89	10	16	35.85	6.35	41.5	4.43	40.77	4.55	82.27	8.35
Beh	NM (n=21)	21.9	5.18	16	5	12.9	5.77	-	-	-	-	-	-
	M (n=24)	22.75	4.42	14	10	35.42	5.69	-	-	-	-	-	-

2.2. Stimuli

The stimuli corresponded to a set of six novel melodies composed following the rules of classical Western tonal music. These melodies were used in a recent MMN experiment (high entropy condition in Quiroga-Martinez et al., 2019b). Each melody was 32 notes long and lasted eight seconds. Major- and minor-mode versions of the melodies were transposed to six different keys (C, C#, D, D#, E, F) and were presented pseudorandomly one after the other so that no melody was repeated before all melodies were played. The major and minor versions of each melody were repeated twelve times, in randomly selected transpositions. The stimuli were delivered in three blocks lasting around seven minutes each and were part of a longer experimental session that included other conditions addressing questions beyond the scope of this study. Data from these conditions have been (Quiroga-Martinez et al., 2019a, 2019b) and will be published elsewhere. Individual tones were created using a piano sample and had a duration of 250 ms. The pitch range of the melodies, as presented during the session, spanned 31 semitones from B3 ($F_0 \approx 247$ Hz) to F6 ($F_0 \approx 1397$ Hz). Since the experiment was designed as a multifeature MMN paradigm, several types of deviants were inserted in the melodies. Relevant for this study are pitch (or mistuning) deviants, which consisted of a 50-cents (quarter-tone) pitch rise with respect to the standard tones. See Quiroga-Martinez et al., (2019b) for further details.

2.3. A computational model of auditory expectation

To obtain continuous measures of auditory surprise we used IDyOM (Pearce, 2005, 2018), a variable-order Markov model of expectation that quantifies surprise as information content (IC):

$$IC_p = -log_2\, p$$

Here *p* is the conditional probability of the current sound event, given previous events in the sequence and the long-term training of the model. This formulation implies that the lower the probability of the event, the larger its IC value and thus the surprise it generates in the listener. IDyOM estimates continuation probabilities by keeping track of the number of times a given pattern of events has occurred. The model takes into account patterns of variable length ("*n*-grams") whose probabilities are

combined through a smoothing process to produce the output values (see Pearce, 2005 for details). An advantage of IDyOM is that it can simulate different types of expectations. Specifically, it has a short-term (*stm*) submodel that generates dynamic expectations derived from the current auditory sequence and a long-term (*ltm*) submodel that simulates life-long schematic expectations derived from a large training corpus. In Omigie et al., (2013), a configuration that combined both the *ltm* and *stm* submodels was used. This configuration, known as *both+*, simulates a listener who employs short-term and long-term expectations to predict pitch continuations, but who also updates the *ltm* submodel with the knowledge gathered from the current stimulus (as indicated by the "+" symbol). Here, the corpus used to train the *ltm* submodel was the same as in Omigie et al., (2013) and corresponded to a collection of hymns and folk songs belonging to the Western tonal tradition.

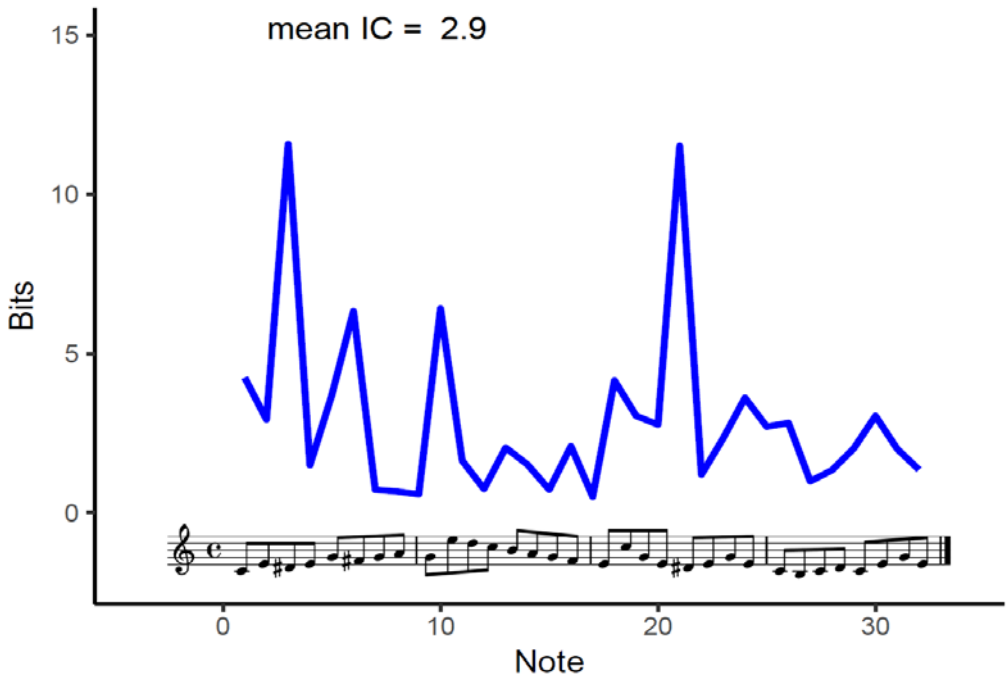

Figure 1. Example of one of the melodies used in the experiment and its estimated information content (IC, blue line). Values were obtained employing the reference model (*Both+* with scale degree and pitch-interval viewpoints combined). Note that out-of-key accidental tones (preceded by sharps, i.e. #) that belong to the 12-tone chromatic set (corresponding to all keys on the piano) but fall outside the 7-tone diatonic set (corresponding to the white keys on the piano in the key of C-major) typically result in higher information content (IC) than notes belonging to the diatonic set. For this experiment, out-of-tune deviant tones outside the chromatic set were also introduced (not displayed here). See Quiroga-Martinez et al. (Quiroga-Martinez et al., 2019b) for details in this regard, and the HE condition in the supplementary file 1 of the same study for the full stimulus set and corresponding IC estimates.

IDyOM can use representations of different features of the auditory input, known as *viewpoints*, to make its probabilistic predictions. Following Omigie et al., (2013) and other behavioral work (Agres, Abdallah, & Pearce, 2018; Hansen & Pearce, 2014; Hansen et al., 2016), we used a mixture of scale degree ("*cpint*") and pitch interval ("*cpintfref*") viewpoints to predict pitch ("*cpitch*") continuations in the melodies. This is our reference model. The scale degree viewpoint assigns a category to each tone in a musical scale with reference to its tonal center, irrespective of its absolute pitch height. Scale degrees are hierarchically organized in Western tonal music so that, if context is kept constant, more prominent degrees (e.g., the tonic) are more frequent and are perceived as more expected than less prominent degrees (e.g., the leading tone) (Krumhansl, 1990). Transitions between scale degrees are idiomatic so that, for example, a leading tone would most often be followed by the tonic (Huron, 2006,

p. 160). The pitch-interval viewpoint, on the other hand, quantifies the distance in semitones between consecutive tones. Idiomatic preferences pertaining to pitch interval also exist in Western tonal music such as the preponderance of small pitch intervals and note repetitions (Huron, 2006, p. 74).

To assess the contribution of different components to model performance, we compared different configurations of IDyOM. As a first step, we compared the reference configuration (*both+* with scale degree and pitch interval viewpoints combined) with a model that did not derive long-term knowledge from the current sequence (i.e., *both*). This aimed to reveal the extent to which listeners updated their long-term knowledge based on the current melodies. In a second step, the parameters were manipulated along two dimensions. First, we created submodels with either an *stm* or *ltm* component only, the comparison of which indicated the extent to which listeners relied on long-term or short-term expectations only. Second, we created submodels with either a scale degree or a pitch interval viewpoint only. Comparing these models allowed us to assess the extent to which participants' expectations were primarily based on one of these musical features compared to the other. Orthogonal manipulations of these dimensions gave rise to the configurations shown in Figure 4 and table 3. IC values based on each of these configurations were obtained for every tone in the melodies. The values were binned into ten quantiles corresponding to increasing IC levels across the entire stimulus set and used for the analyses of the MEG data. Note that an *stm* model with scale degree and interval viewpoints was not included, as the range of IC values was not sufficiently large to avoid having the same value repeated in different quantiles.

2.4. 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 60 dB above threshold. Participants were instructed to watch a silent movie, 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 adjust 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.

2.5. MEG recording and preprocessing

Magnetic correlates of brain activity were recorded at a sampling rate of 1000 Hz using an Elekta Neuromag MEG TRIUX system with 306 channels (204 planar gradiometers and 102 magnetometers). Participants' head position was continuously 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 & Simola, 2006) using Elekta's MaxFilter software (Version 2.2.15). This procedure included movement compensation in all but two non-musicians, for whom 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 P50m 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, 2013). Visual inspection was used as a quality check.

The ensuing analysis steps were conducted with the Fieldtrip toolbox (version r9093) 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. Each epoch was assigned to a category according to the IC quantile of the corresponding sound, for a given IDyOM model. For each participant and quantile, ERFs were computed by averaging the responses for all the tones belonging to the quantile. The four initial tones of each melody were excluded from the analyses to avoid neural activity related to the transition between melodies (e.g., effects of key change). Tones preceded by a deviant were also excluded to avoid carryover effects from the deviant response. Between 250 and 350 epochs were averaged per IC quantile.

2.6. Sensor-level analyses

The statistical analyses were performed on the magnetometers because this allowed us to properly look at the polarity of the magnetic fields, which is fundamental to disentangle different components such as the P50m, N1m and P2m. To assess whether ERF amplitude increased with information content, we performed a dependent-samples regression (*ft_statfun_depsamplesregrT* function in Fieldtrip) for musicians and non-musicians separately in a mass-univariate analysis. This type of regression employs the mean of the participant-wise coefficients to assess whether the association is greater than zero, thus giving a group-level *t*-statistic as output. To account for multiple comparisons, we used cluster-based permutations with $p = 0.05$ as cluster-forming threshold, t-*maxsum* as statistic, and 10,000 iterations. A time window between 0 and 300 ms after onset was selected, to avoid substantial overlap with the neural activity of the next tone (starting at 250 ms), but at the same time explore possible unexpected effects (e.g., in the P50m or the P2m components). To assess whether the association between neural activity and surprise was different for musicians and non-musicians, in a first-level analysis we obtained *t*-values reflecting the strength association between IC quantile and neural activity for each participant, at each time point and sensor (*ft_statfun_indepsamplespregrT* function), which were then compared between groups in a second-level analysis with a two-sided independent-samples *t*-test (*ft_statfun_indepsamplesT* function). Cluster-based permutations were used to correct for multiple comparisons as indicated above.

To assess the relative evidence for the absence or the presence of an IC quantile-by-expertise interaction, we applied Bayesian estimation to mean N1m amplitudes. These amplitudes corresponded to the mean activity ± 20 ms around the peak, extracted from the average of the 4 channels in each hemisphere that showed the strongest effect of IC quantile in the mass-univariate analyses (left channels: 0231, 0241, 1611, 1621; right channels: 1341, 1331, 2411, 2421). Since the ERFs have opposite polarities in different hemispheres, we multiplied left-hemisphere mean amplitudes by -1, so that they had the same (negative-going) polarity found in right-hemisphere amplitudes.

Several mixed-effects models were estimated using the *brm* function from the *brms* package (Bürkner, 2017) in R (R Core Team, 2019) and compared in an incremental way, adding one term at the time until reaching a full model with all main effects and interactions (Table 2). Participant random effects were included for intercept and slopes. The prior for the effect of IC quantile was taken from the reported results in Omigie et al (2013), which showed a difference of around 1.5 μV between quantiles 1 and 10 for control participants. Standardizing this difference with the variance around 100 ms gives

an effect size of about 0.96. This effect was extrapolated to magnetometer data using the variance taken from a previous dataset collected with the same MEG scanner, which resulted in a difference of about 30 fT between quantiles 1 and 10. Therefore, a plausible slope would be 30/9 = 3.33 fT/quantile, which we rounded to 3.5 fT/quantile. With this rough estimate, we set a conservative Gaussian prior centered at 0 with SD = 3.5 fT/quantile. In other words, we regard effects similar or smaller than 3.5 fT/quantile as most likely, and effects larger than twice this value as very unlikely. For the IC quantile-by-expertise interaction we set a similar prior centered at zero with SD = 1.75 fT/quantile, which means that we regard modulations equal to or smaller than half of the main effect to be most likely. The same prior was set for other interaction terms involving IC quantile. Regarding the main effects of expertise and hemisphere, Gaussian priors centered at zero with SD = 10 fT were set, which are also conservative and represent rather small effect sizes. Finally, for the intercept of the model as well as the standard deviation and random effects, weakly informative priors were set, corresponding to a uniform distribution between 0 and 100 fT. Models were built in an incremental way, with *m0* having an intercept only (null model), *m1* adding a term for IC quantile, *m2* adding a term for hemisphere, and so on (Table 2). The full model (*m7*) included all main effects and interactions. Comparisons between adjacent models (i.e., models with and without a particular factor) were performed by estimating Bayes factors (BF) and Widely Applicable Information Criteria (WAIC) (Wagenmakers et al., 2018; Watanabe, 2010). Of particular interest is the comparison between *m3* and *m4*, as it assessed the evidence for the IC quantile-by-expertise interaction.

2.6.1. Comparisons between IDyOM configurations

To compare the different IDyOM configurations shown in Table 3, several metrics were used. First, we took the peak r^2-value across the whole sensor array for each model, resulting from the average of participant-wise first-level regression analyses (*ft_statsfun_indepsamplesregrT* function). Second, we obtained maximum likelihood estimates of linear mixed effects models (function *lmer*, package *lme4*, Bates, Mächler, Bolker, & Walker, 2015) of N1m mean amplitudes (instead of Bayesian estimates for speed of processing), including IC quantile and hemisphere as predictors, and random intercepts and slopes for participants. r^2-values for the effect of IC quantile were also obtained.

2.6.2. Comparison with the MMNm

The preprocessing steps for the measurement of the MMNm were the same as for the N1m, with the difference that, instead of surprise levels, an ERF was computed for only two conditions: standards and deviants. The MMNm is the difference between these two conditions, which was assessed here for magnetometers through paired-samples *t*-tests in mass-univariate analyses (note that statistical analyses for the MMNm had previously been conducted for gradiometers in Quiroga-Martinez et al. 2019a, 2019b). The deviants analyzed here correspond to out-of-tune tones (i.e., tones with pitches outside the musical tuning system that cannot normally be played on a piano). Note that other deviants were also present in the experiment, but we focused on mistunings as they are disruptive with regard to pitch, which is the feature modeled with IDyOM in this study. Independent-samples *t*-tests on MMNm difference waves were used to assess the effect of expertise. For the comparison between the two components, the N1m was calculated as the difference between IC quantile 10 and IC quantile 1. In this case, the model with the best performance was used (*ltm* with a pitch-interval viewpoint), to increase the signal-to-noise ratio. Two-sided paired-samples *t*-tests were used to compare the resulting MMNm and N1m difference waves. Cluster-based permutations were employed as multiple comparisons corrections in all analyses, as described above. Since, based on visual inspection, the peak latency of the N1m and the MMNm seemed to be different—which might give misleading results in the

permutation tests—we additionally performed two-sided paired-samples *t*-tests on peak latencies and mean amplitudes separately, estimated as described above.

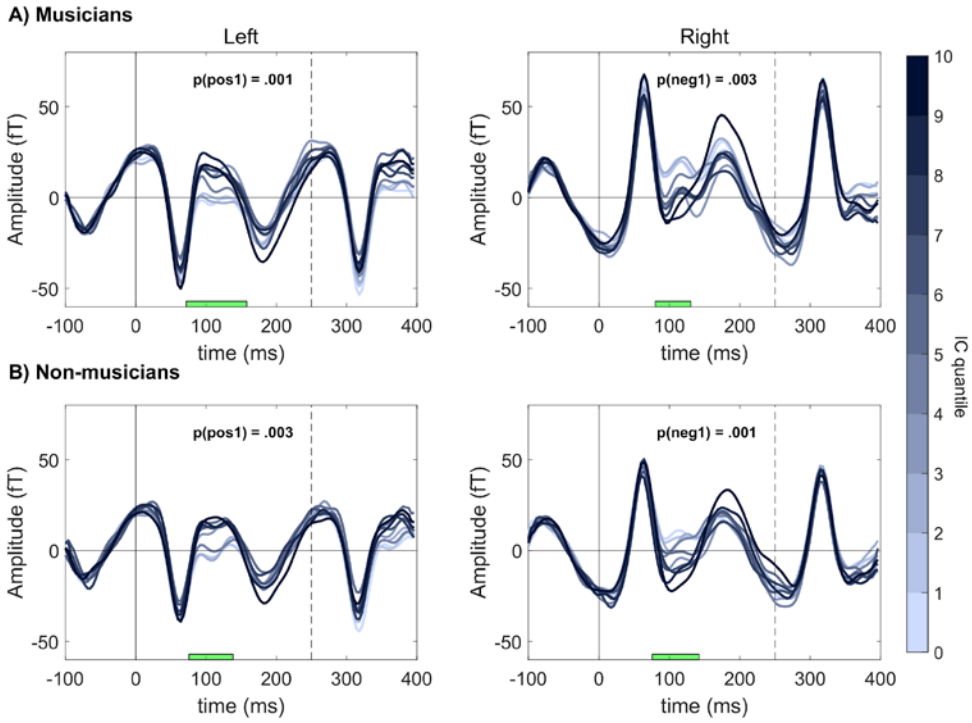

Figure 2. Effect of surprise (IC quantile) on the amplitude of the evoked response for both groups and hemispheres. For descriptive purposes, the time points when positive ("pos") and negative ("neg") clusters were significant are indicated with green horizontal bars. Note that this is not an accurate estimate of the true extent of the effects (Sassenhagen & Draschkow, 2019). Displayed activity corresponds to the average of the four temporal magnetometers in each hemisphere with the strongest effect (left channels: 0231, 0241, 1611, 1621; right channels: 1341, 1331, 2411, 2421). Vertical dashed lines indicate the onset of the next tone.

2.7. Source reconstruction

For source reconstruction we employed the multiple sparse priors method (Friston et al., 2008) in SPM12 (version 7478). Individual anatomical magnetic resonance images (MRI) were available for 20 musicians and 20 non-musicians only. These are the participants included in the analysis. In the case of two excluded musicians and one excluded non-musician, the images were corrupted by artefacts. The remaining excluded participants did not attend the MRI session. Two brain images were acquired with a magnetization-prepared two rapid gradient echo (MP2RAGE) sequence (Marques et al., 2010) in a Siemens Magnetom Skyra 3T scanner. These images were combined and motion-corrected to form unified brain volumes, which were subsequently segmented, projected into MNI coordinates, and automatically coregistered with the MEG sensor array employing digitized head shapes and preauricular and nasion landmarks. Coregistration outputs were visually inspected. Lead-fields were constructed using a single-shell BEM model with 20.484 dipoles (fine grid). A volume of the inverse solution was created for each participant between 75 and 125 ms for the N1m (difference between quantile 10 and quantile 1 for the *ltm* pitch-interval model), and between 175 and 225 ms for the MMNm (difference between deviants and standards). These time windows were chosen based on the peak amplitude of each component and are warranted by the output of statistical tests. The volumes for each component were submitted to a two-sided one-sample *t*-test to reveal the sources consistently identified across all participants. To assess possible differences between the N1m and the MMNm, the volumes were also

compared in a two-sided paired-samples *t*-test. The error rate of voxel-wise multiple comparisons was corrected with random field theory, with a cluster-level alpha threshold of 0.05 (Worsley, 2007).

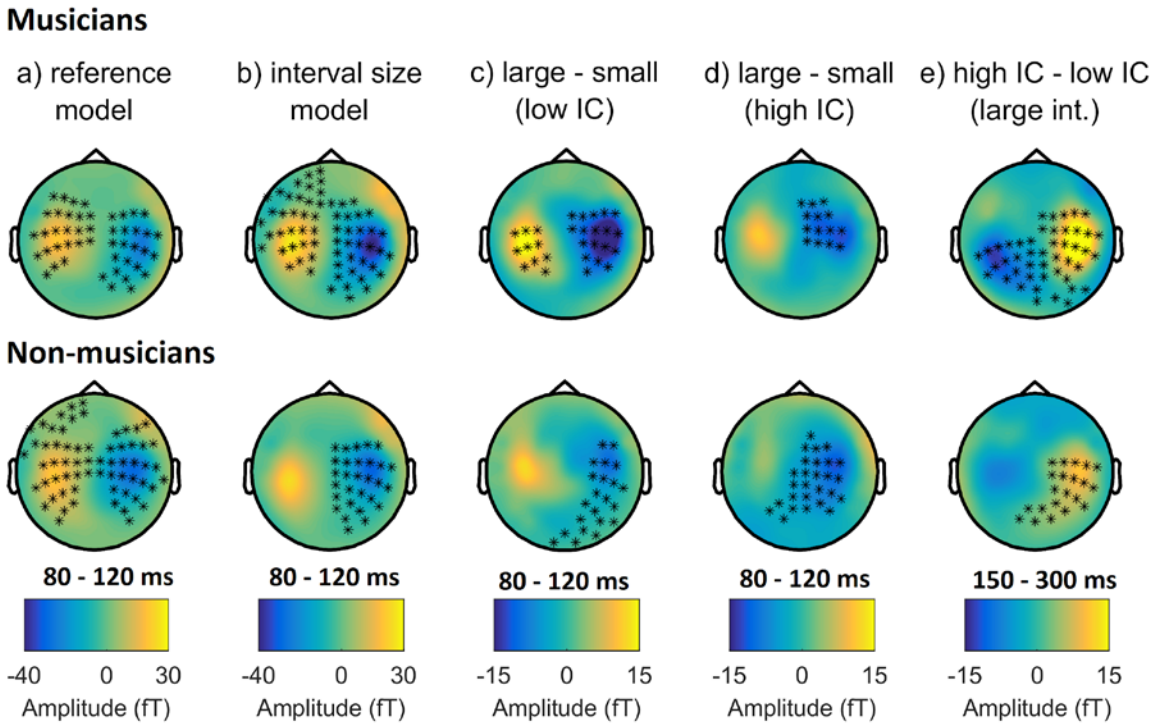

Figure 3. Topographic maps of the difference between: **a)** surprising (quantile 10) and unsurprising (quantile 1) tones for the reference model (*both+* with scale degree and pitch interval viewpoints combined); **b)** tones following large (≥ 5 semitones) and small (1 semitone) intervals; tones following large (≥ 3 semitones) and small (≤ 2 semitones) intervals with either **c)** low-IC or **d)** high-IC; and **e)** high-IC and low-IC tones following large intervals. For **a**, **b**, **c,** and **d**, activity between 80 and 120 ms is displayed, corresponding to a modulation of the N1m component; whereas for **e**, activity between 150 and 300 ms is displayed, potentially corresponding to a P3am component. Note the change in polarity between the two time windows. Stars mark the channels where regression analyses (**a**, **b**) or pairwise contrasts (**c**, **d**, **e**) were significant. Note that this is not an accurate estimate of the true extent of the effects (Sassenhagen & Draschkow, 2019).

3. Results

3.1. Modulation of the N1m and expertise effects

Dependent-samples regressions revealed a significant association between the amplitude of the evoked responses and IC quantile (Figure 2) for both musicians and non-musicians around 100 ms after sound onset. This association was largest at temporal magnetometers, was negative in the right hemisphere and was positive in the left hemisphere (Figure 3a). This reflects the inversion of the recorded magnetic field given the hemisphere-dependent orientation of the source with respect to the sensor array. No significant differences between musicians and non-musicians were detected in the second-level analysis (smallest *p*-value = .2).

Table 2. Widely Applicable Information Criteria (WAIC) and their standard error (SE), for each of the estimated mixed-effects models. For each comparison between adjacent models (i.e. "Model" vs "Null"), Bayes factors in favor of the alternative (BF_{10}) and the null (BF_{01}) are also reported. The predictors included in each model are indicated as follows: Int = Intercept, quant = IC quantile, hem = hemisphere, exp = expertise. Interactions are marked with colons (":"). Note that the models also included participant-wise random effects for the intercept, the effects of IC quantile and hemisphere, and the quantile-by-hemisphere interaction. Comparisons with moderate or strong evidence for either the null or the alternative are highlighted in bold and marked with an asterisk "*".

Model	Predictors	Null	WAIC	WAIC (SE)	BF_{10}	BF_{01}
m0	Int	NA	8945.24	56.57	NA	NA
m1	**Int + quant**	***m0***	**8839.13**	**55.83**	**> 1000***	**< .001**
m2	**Int + quant + hem**	***m1***	**8360.36**	**65.14**	**> 1000***	**< .001**
m3	Int + quant + hem + exp	*m2*	8360.44	65.12	0.34	2.93
m4	**Int + quant + hem + exp + quant:exp**	***m3***	**8361.24**	**65.06**	**0.24**	**4.21***
m5	**Int + quant + hem + exp + quant:exp + quant:hem**	***m4***	**8362.08**	**64.69**	**<.001**	**434.07 ***
m6	Int + quant + hem + exp + quant:exp + quant:hem + hem:exp	*m5*	8362.46	64.71	0.89	1.12
m7	Int + quant + hem + exp + quant:exp + quant:hem + hem:exp + quant:hem:exp	*m6*	8363.41	64.67	0.43	2.33

3.2. Bayesian analyses of N1m amplitudes

Results from the Bayesian analyses are shown in Table 2 and Figure 4. WAIC values indicate that adding IC quantile (*m1*) and hemisphere (*m2*) significantly improves predictive power, but adding the main effect of expertise (*m3*) or any interaction term does not improve the predictive power any further. Estimations of model weights (Figure 4c) also suggest *m2* as the winning model. A similar picture can be inferred from Bayes factors. The *m1* model—with a term for IC quantile—is much more likely than an intercept-only model (*m0*), and the *m2* model—including hemisphere—is much more likely than the *m1* model. This strong effect of hemisphere might be a carryover from the P50m component, which was larger in right-hemisphere channels. No other comparisons support alternative models. Instead, in a few cases there is evidence for null models. Of paramount interest here is the comparison between *m3* and *m4*, which suggests that a model with no interaction between IC quantile and expertise is about 4.2 times more likely than a model with it (Table 2). We regard this as moderate evidence for the null hypothesis. Similarly, a model with no IC quantile-by-hemisphere interaction (*m4*) is about 434 times more likely than a model with it (*m5*). The remaining BF suggest the absence of an effect of expertise, the absence of a hemisphere-by-expertise interaction, and the absence of a three-way-interaction between IC quantile, hemisphere and expertise. In these cases, however, the BF are inconclusive. Finally, a look at the parameters shows that only the estimates for the main effects of IC quantile and hemisphere are credibly different from 0 across all models compared (supplementary file 1).

3.3. Comparison of IDyOM configurations

The performance of IDyOM configurations in terms of predicting the N1m amplitude from IC estimates is reported in Table 3. Even though the reference model (*Both+*, with scale degree and interval viewpoints combined) performed well, models that included a long-term component and an interval-only viewpoint performed better. This was reflected in larger r^2-values and steeper slopes. Models with

only scale-degree viewpoints and/or short-term components performed poorly. The best model employed a long-term component and an interval-only viewpoint. Given that in Western tonal music (and therefore in our training corpus), smaller intervals are more common than larger ones, it could be the case that the effects observed are caused by the size of the interval itself rather than its IC, which would explain why long-term interval-only models had the best fit. For this reason, in further exploratory analyses, four categories of tones were averaged for each participant, comprising small-interval (≤ 2 semitones) transitions with either low IC (small/low-IC, ≤ 4) or high IC (small/high-IC, > 4); and large-interval (≥ 3 semitones) transitions with either low IC (large/low-IC) or high IC (large/high-IC). Cluster-based, permutation-corrected paired-samples *t*-tests conducted in a 50-300 ms time window revealed significant differences between large/high-IC and large/low-IC for musicians and non-musicians (Figure 6). Crucially, these differences were found in later time windows that excluded earlier N1m latencies. Moreover, the direction of this effect was the opposite of that of the N1m modulation. This resulted in the polarity shown in Figure 3e, which would presumably correspond to a positive deflection in an EEG recording and could be interpreted as the magnetic counterpart of the P3a (P3am) (see section 4.2). Regarding small intervals, a significant difference was found between small/high-IC and small/low-IC around 200 ms after sound onset, only for musicians in the right hemisphere. The direction of this effect was opposite to the one found for the comparison between large/high-IC and large/low-IC (Figure 6). Note, however, that the *p*-value was close to the significance threshold in this case. Finally, in contrast to these analyses, when large and small intervals were compared for low-IC and high-IC tones separately, significant differences in the N1m time window (50-150 ms) in the expected direction were found in both groups (Figures 3c and 3d).

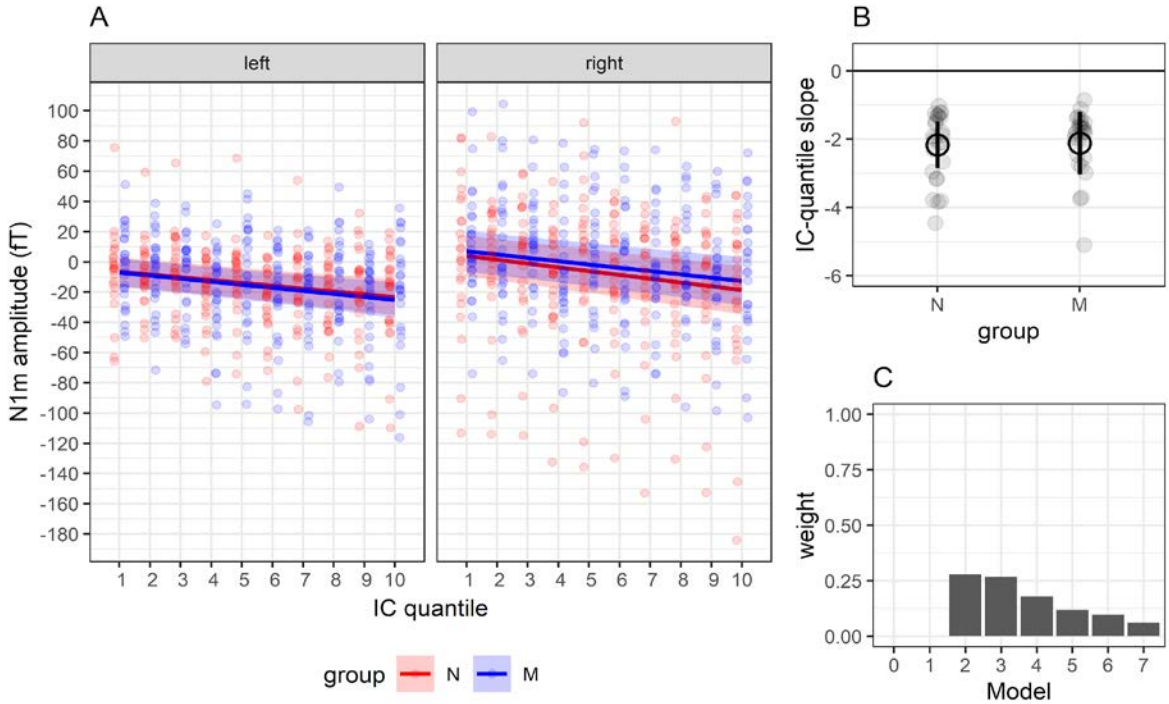

Figure 4. A) N1m amplitude as a function of surprise (IC quantile). Bayesian estimates of the IC quantile slopes (colored lines) and corresponding 95% credible intervals (shaded areas) are also displayed. Note that mean amplitudes in the left hemisphere were multiplied by -1 in order for them to have the same polarity as those in the right hemisphere. B) Bayesian estimates of the individual- (small gray dots) and group-level (larger circles) association between IC quantile and N1m amplitude. Error bars represent 95% credible intervals. C) Posterior model weights resulting from the incremental Bayesian model comparisons. N = non-musicians, M = musicians. See Table 2 for a description of each model.

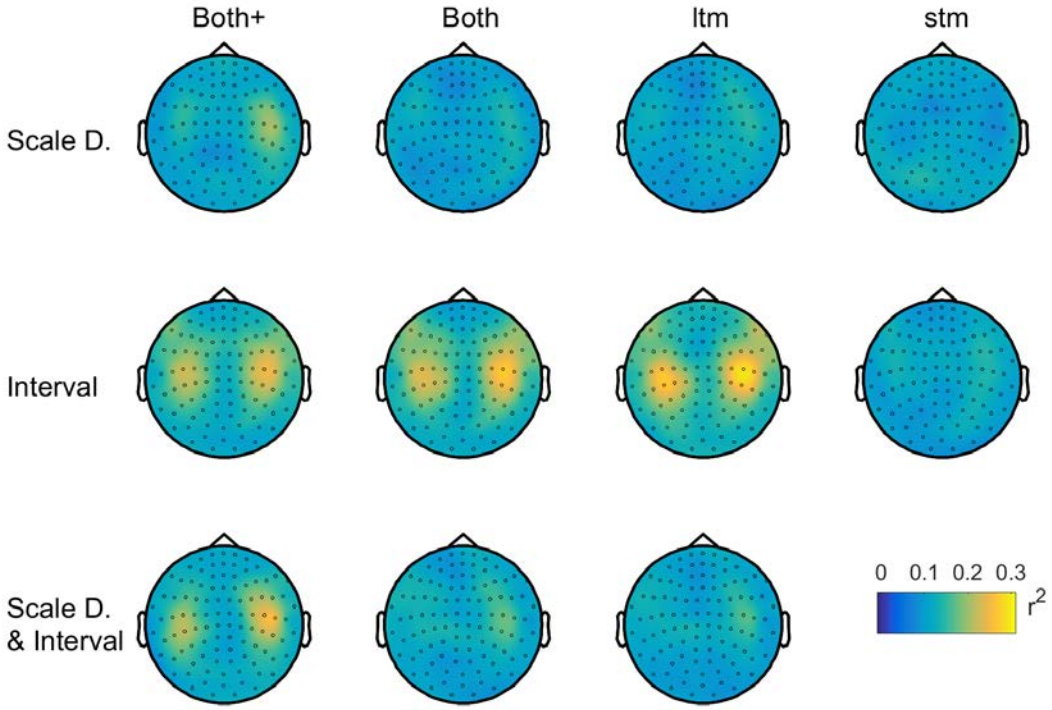

Figure 5. Grand-averaged topographic maps of the variance (r^2) in neural activity explained by each IDyOM configuration for each participant (i.e., first-level intra-subject variance). Configurations included scale degree and pitch interval viewpoints, and their combination, as well as short-term (*stm*) and long-term (*ltm*) submodels, and their combination (*Both* and *Both+*).

In addition, we performed a within-subjects regression analysis using absolute interval size (with no IDyOM modeling) as predictor. We grouped and averaged the tones into five categories comprising intervals with either 1, 2, 3, 4 or ≥5 semitones. The number of categories could not be increased as there were few instances of large intervals. The analysis showed that interval size was strongly associated with the amplitude of the N1m in musicians and non-musicians (Figures 7 and 3b; maximum grand-averaged intra-subject $r^2 = 0.52$). The variance explained by this model was higher than the one explained by any of the IDyOM configurations (Table 3). However, note that part of this outcome might be due to including five instead of ten categories, which already reduced the variance in the data. Interestingly, all categories showed a monotonic amplitude increase, except for 1-semitone intervals, whose amplitude was larger than expected (Figure 7).

3.4. Comparison between N1m and MMNm

Mass-univariate analyses revealed that the ERFs of the MMNm and N1m were substantially different (Figure 8). The latency of the MMNm was significantly longer ($t = 35.68$, $p < .001$) and its amplitude was significantly larger ($t = 8.37$, $p < .001$) than those of the N1m, in the mean and peak amplitude analyses. Moreover, expertise effects were found for the MMNm ($p < .001$), thus reproducing in the magnetometer space the differences previously reported for combined gradiometers (Quiroga-Martinez et al., 2019a). Regarding source analyses, one-sample *t*-tests suggested that the main generators of both the MMNm and the N1m were located in the surroundings of primary auditory cortex, as expected (Figure 9). The peak activation for the MMNm was located in the anterior part of the primary auditory cortex (A1) in the right hemisphere (56, 0, 2), and in the posterior portion of A1 in the left hemisphere (-48,-16, -4). For the N1m, the peak was located in the posterior part of A1 in both hemispheres (right: 48, -18, 6; left: -46,-18, -4). The contrast between the two components did not yield significant results.

Table 3. Performance of different IDyOM configurations in predicting the N1m amplitude from estimates of information content (IC), as measured by the variance explained at the group (mixed-model) and individual (intra-subject) level. The slope of the association between surprise (IC quantile) and N1m amplitude and corresponding confidence interval (CI), and the correlation between estimated surprise and the N1m amplitude averaged (or collapsed) across all subjects are also reported. The latter provides measures comparable to the correlations reported in Omigie et al. (2013).

Viewpoint	Submodel	r^2 (mixed model)	max r^2 (intra-subject)	Slope	CI (2.5%)	CI (97.5%)	r (collapsed mean amplitude)
interval	Ltm	0.15	0.44	-2.91	-3.5	-2.32	-0.89
interval	Both	0.14	0.41	-2.77	-3.32	-2.22	-0.86
interval	both+	0.12	0.37	-2.47	-3.01	-1.94	-0.79
scale degree & interval	both+	0.1	0.34	-2.16	-2.64	-1.69	-0.85
scale degree	both+	0.05	0.26	-1.52	-1.96	-1.08	-0.75
interval	Stm	0.04	0.21	-1.3	-1.77	-0.83	-0.68
scale degree & interval	Both	0.03	0.28	-1.22	-1.68	-0.75	-0.56
scale degree & interval	ltm	0.02	0.26	-1.07	-1.55	-0.6	-0.44
scale degree	ltm	0.02	0.19	-0.92	-1.38	-0.47	-0.41
scale degree	Both	0.02	0.19	-0.81	-1.25	-0.38	-0.43
scale degree	Stm	0	0.19	0.47	0	0.93	0.16

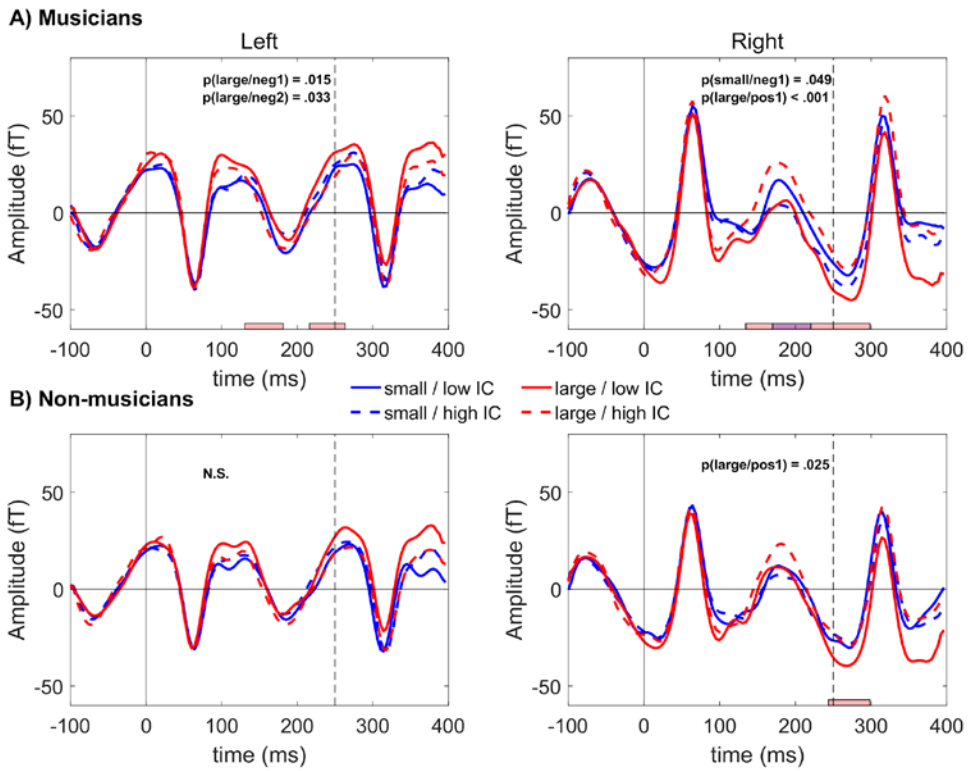

Figure 6. Event-related fields for tones with high (> 4 bits) or low (≤ 4 bits) surprise (i.e., information content) following small (≤ 2 semitones) or large (≥ 3 semitones) intervals. The reference model (*Both+* with scale degree and pitch-interval viewpoints combined) was used to obtain the estimates. For descriptive purposes, horizontal colored lines indicate positive ("pos") or negative ("neg") clusters associated with differences between conditions. Note that they are not an accurate estimate of the true extent of the effects (Sassenhagen & Draschkow, 2019). The color of these lines indicates whether the contrast between high-IC (dashed lines) and low-IC (solid lines) tones was made for large (red) or small (blue) intervals. Vertical dashed lines indicate tone onsets.

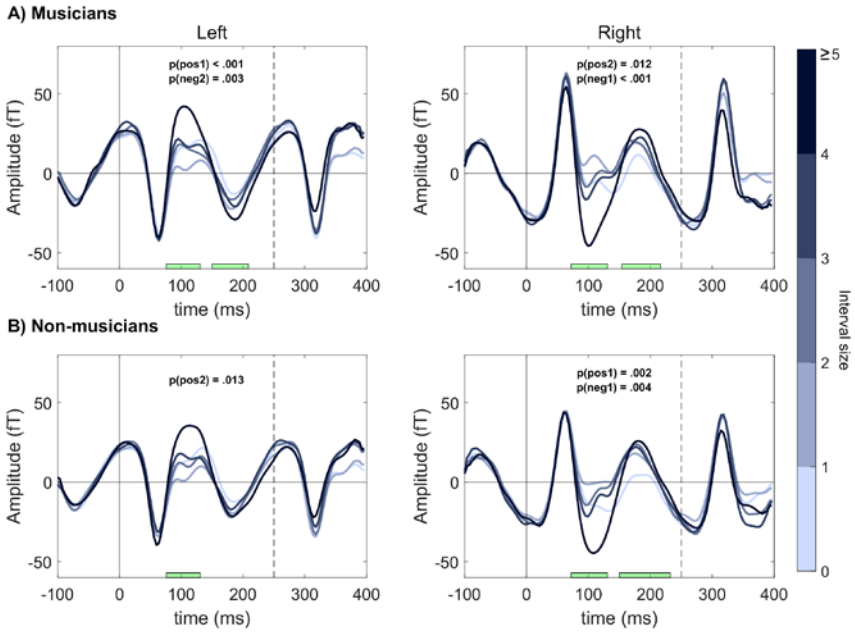

Figure 7. Neural responses to pitch intervals of different size (in semitones) in both groups and hemispheres. For descriptive purposes, the time points when positive ("pos") and negative ("neg") clusters were significant are indicated with green horizontal bars. Note that they are not an accurate estimate of the true extent of the effects (Sassenhagen & Draschkow, 2019). Displayed activity corresponds to the average of the four temporal magnetometers in each hemisphere with the strongest effect (left channels: 0231, 0241, 1611, 1621; right channels: 1341, 1331, 2411, 2421). Vertical dashed lines indicate the onset of the next tone.

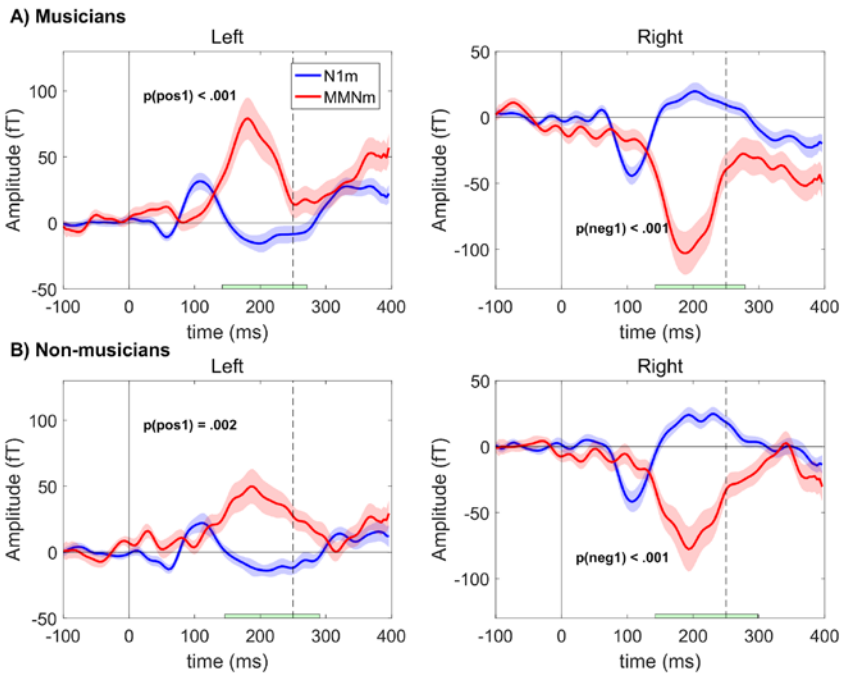

Figure 8. Event-related fields of the N1m modulation (i.e. quantile 10 - quantile 1 of the *ltm* interval-only model) and the MMNm (deviant - standard). For descriptive purposes, green horizontal lines indicate the times when differences between components were significant. Note that this is not an accurate estimate of the true extent of the effects (Sassenhagen & Draschkow, 2019). Shaded areas depict 95% confidence intervals. Displayed activity corresponds to the average of the four temporal magnetometers in each hemisphere with the strongest N1m modulation (left channels: 0231, 0241, 1611, 1621; right channels: 1341, 1331, 2411, 2421). Dashed vertical lines indicate tone onsets.

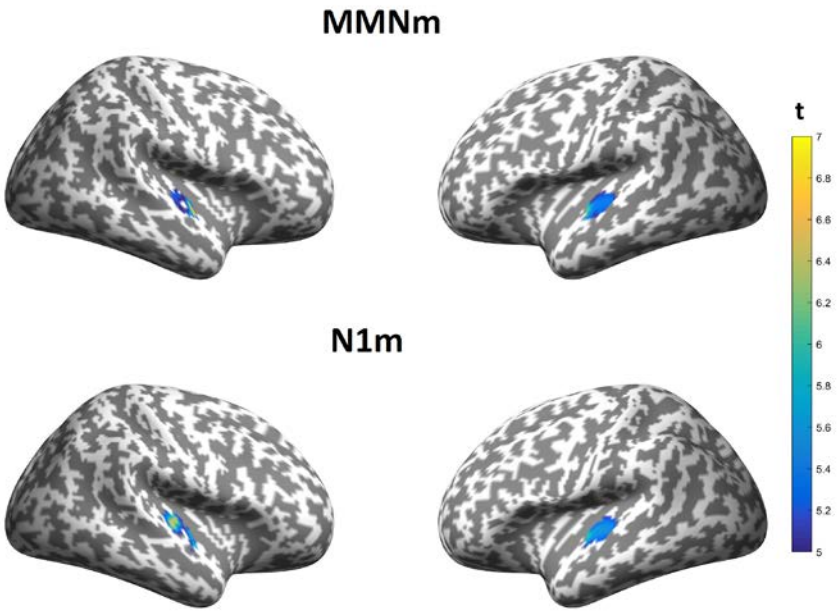

Figure 9. Neural generators of the MMNm and the N1m. Color maps depict *t*-statistics from one-sample *t*-tests thresholded at $p = 0.05$, after multiple-comparisons correction. Participants from both groups were included in the tests, as no statistically significant differences were detected between musicians and non-musicians.

4. Discussion

In this study, we aimed to determine whether the modulation of neural activity by subtle changes in auditory surprise—as estimated with a computational model of expectations (IDyOM)—is affected by musical expertise. Our results showed that surprise estimates were associated with the amplitude of the N1m in both groups, thus replicating the findings by Omigie and colleagues (2013). However, contrary to our expectations, non-significant null results and Bayesian analyses indicated no differences between musicians and non-musicians in the strength of this association.

4.1. Interval size and sensory adaptation better explain N1m responses

Further exploratory analyses provided clues about the nature of the expectations reflected in the N1m modulation, and an explanation for the lack of differences between the groups. Comparisons of IDyOM configurations indicated that models predicting pitch interval based on long-term knowledge performed best, and even better than the reference model used by Omigie and colleagues (2013) (*Both+* with pitch interval and scale degree viewpoints combined). In contrast, models predicting scale degree based on short-term expectations performed the worst. Given that in Western tonal music (which is the style that IDyOM was configured to model in the present experiment), smaller intervals are much more common and therefore overall less surprising than larger intervals, it is possible that the modulation of the N1m could be explained by a much simpler factor: interval size.

In additional analyses, we showed that if high-IC (i.e., surprising) tones are contrasted with low-IC (i.e., less surprising) tones when interval size is kept constant, the differences previously observed in the N1m are not detected anymore. Conversely, if tones following a large interval are contrasted with those following a small interval while surprise is kept constant, then the differences still persist. Furthermore,

N1m amplitude was shown to significantly increase with larger interval sizes. All of this indicates that interval size, rather than probabilistic prediction, is the most likely factor behind the reported effect.

This explanation is compelling when one considers that the amplitude of the N1 is modulated by stimulus specific adaptation or SSA (May & Tiitinen, 2010; May, Westö, & Tiitinen, 2015; Näätänen & Picton, 1987; Pérez-González & Malmierca, 2014). SSA occurs when neural populations that respond to the spectral content of the stimulus become less responsive with subsequent presentations of the same or a similar stimulus, likely due to synaptic depression (May et al., 2015; Ulanovsky, 2004; Ulanovsky, Las, & Nelken, 2003; Yarden & Nelken, 2017). Thus, tonotopic neurons responding to a tone become adapted, and therefore the neural activity generated by an equal or spectrally similar successive tone would be attenuated. However, if the successive tone is sufficiently different, non-adapted neurons would be engaged, thus producing more robust neural responses. Therefore, while the melodies used here were deliberately composed to avoid pitch repetitions, it could be that our results are driven by increased neural responses to sounds farther apart in their spectral content, which is the case for tone transitions with larger pitch intervals. It has to be noted that, although SSA has almost exclusively been studied with pure tones—as opposed to the complex tones used here—some studies have demonstrated SSA for complex sounds (Nelken, Yaron, Polterovich, & Hershenhoren, 2013) and even frequency selectivity for broadband sounds (Rauschecker, Tian, & Hauser, 1995). Furthermore, pitch distance might not be the only relevant variable affecting SSA, as 1-semitone intervals were shown to have a larger amplitude than expected. This points to future research efforts in which the spectral difference between consecutive tones and its effect on N1 amplitudes are carefully measured and tested.

The results in Omigie et al., (2013) are in agreement with the explanation above. When comparing different IDyOM configurations, they found that those including a pitch-interval viewpoint performed the best. Note, though, that scale degree viewpoints had a good performance as well. This could be explained by the fact that scale degree and interval size are also correlated, since scale degree transitions are dominated by stepwise motion (Huron, 2006, p. 160). In our study, scale degree was also associated with neural activity, although to a lesser degree than in Omigie et al., (2013). This might have to do with potential differences in interval distributions between the two stimulus sets. Moreover, Hansen and Pearce (2014) found that long-term models were the best at predicting expectedness ratings. This is consistent with the *ltm*/pitch-interval configuration yielding the best performance in our analyses. This might be due to a comprehensive and accurate estimation of pitch intervals in a Western tonal context—where interval size and interval probability are confounded—presumably arising from the larger amount of data present in the long-term training corpus, as compared to the stimuli themselves.

The current explanation is also consistent with early research showing that the amplitude of the N1 is larger when a pure tone is farther apart from preceding pure tones in the frequency continuum (Butler, 1968; Näätänen et al., 1988; Picton, Woods, & Proulx, 1978). Thus, our results generalize this effect to the spectral content of complex tones in more realistic and complex auditory sequences. Moreover, speech research showing modulations of the N1m component by the acoustic properties of phonemes provides further support to our proposal (Manca & Grimaldi, 2016; Shestakova, Brattico, Soloviev, Klucharev, & Huotilainen, 2004). On the same line, in a previous EEG study reporting an effect similar to the one found here, manipulations of surprise were also confounded with interval size, which is consistent with our data (Koelsch & Jentschke, 2010). Finally and most importantly, the proposed confound would explain why no differences between the groups were found in our study, since the long-term expectations captured by IDyOM rely on culture-specific knowledge of higher-order statistical dependencies between tones, which are arguably different from the early and rather unspecific sensory processes reflected in the N1m.

4.2. Hierarchical auditory predictive processing

The current results are not necessarily at odds with IDyOM being a good model of auditory expectation, for which there is otherwise solid supporting behavioral evidence (Agres et al., 2018; Hansen & Pearce, 2014; Hansen et al., 2016; Morgan, Fogel, Nair, & Patel, 2019; Pearce, 2018). Rather, they indicate that probabilistic prediction is not reflected in the N1, but instead might engage later processing stages. Interestingly, in the analyses keeping interval size constant, IC modulated the amplitude of a later component in the case of large intervals. This effect could be interpreted as a P3, a component associated with the orientation of attention and the engagement of higher-order cognitive processes (Masson & Bidet-Caulet, 2019; Polich & Criado, 2006; Squires, Squires, & Hillyard, 1975). Notably, in the first EEG study using IDyOM as a model of expectedness, surprising tones generated a similar positive late response that was larger for unexpected than expected sounds (Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010).

Modulations of the P3 have been found in global-local MMN paradigms where local expectations about tones and global expectations about patterns are generated and violated independently (Bekinschtein et al., 2009; Wacongne et al., 2011). In this case, P3 responses are observed for global deviants. Thus, it might be that the higher-order musical expectations reflected in IDyOM's estimates are of a similar kind to the global patterns in the aforementioned paradigm. Note, however, that in Bekinschtein et al., (2009), responses to global deviants were observed only under conscious perception, which differs from the passive listening condition used in our study. Therefore, the late response found here may correspond to the magnetic counterpart of an early subcomponent of the P3, the P3a (or P3am), which is thought to reflect the orientation of attention to violations of the context, in contrast to the later P3b subcomponent, which is taken to reflect proper conscious perception (Polich & Criado, 2006; Squires et al., 1975).

In addition to the late effect of IC in the case of large intervals, we found similar differences for small intervals in musicians in the right hemisphere. However, the direction of this effect was opposite to that of the putative P3am, the reason for which is difficult to infer from our data. It might be the case that this effect corresponds to an emergent MMNm hidden by attention-related P3am responses in the case of large intervals, which are arguably more salient than small intervals. In any case, this effect should be interpreted with caution since the corresponding *p*-value was close to the significance threshold.

While these findings come from exploratory analyses and need to be replicated in a controlled experiment, they speak to a predictive processing hierarchy encompassing different representational levels and time scales. Thus, the N1m would be restricted to tonotopic predictions about the spectral content of sounds mainly driven by the immediate past. In contrast, later components, such as the P3am, may reflect higher-order stylistic predictions about categories derived from the initial sensory parsing—e.g., pitch—and would index the encoding of patterns spanning longer temporal scales. This is consistent with empirical evidence suggesting hierarchical processing along the auditory system (Escera & Malmierca, 2014; Griffiths & Warren, 2002; Parras et al., 2017; Rauschecker et al., 1995; Wacongne et al., 2011), and with theories of predictive processing in the brain (Clark, 2016; Friston, 2005; Friston, Rosch, et al., 2017; Vuust & Witek, 2014).

4.3. The distinction between MMN and N1 revisited

The relationship between N1 and MMN has been long debated (Jaaskelainen et al., 2004; May & Tiitinen, 2010; Risto Näätänen, Jacobsen, & Winkler, 2005; Risto Näätänen & Picton, 1987). While

their scalp topographies and latencies are similar, and both components respond to changes in the auditory signal, they can be dissociated in specific cases. For example, an unexpected tone repetition (Tervaniemi, Saarinen, Paavilainen, Danilova, & Näätänen, 1994), tone omission (Bendixen et al., 2012), or intensity decrease (Risto Näätänen et al., 2007) would attenuate the N1 but nonetheless elicit a robust MMN. This has led to the conclusion that the N1 is mainly modulated by SSA, whereas the MMN reflects the breach of a memory trace (Näätänen et al., 2007; Näätänen & Picton, 1987) or a probabilistic predictive model (Bendixen et al., 2012; Garrido et al., 2009; Lieder et al., 2013; Wacongne, Changeux, & Dehaene, 2012). This is consistent with the two components reflecting different levels of hierarchical processing.

Here we were able to dissociate the N1m and the MMNm in the same subjects and in the same stimulus sequence. Consistent with hierarchical processing, our results indicate that the N1m effect is much smaller and happens earlier than the MMNm. Importantly, the sounds that gave rise to the MMNm in our experiment (i.e., quarter-tone mistunings) violated the musical tuning system, which prescribes the pitches and pitch intervals that can be expected (Brattico, Tervaniemi, Näätänen, & Peretz, 2006). This entails that the listener needs to infer abstract regularities and relationships between tones (such as: “the distance between consecutive pitches cannot be smaller than a semitone”), which goes beyond the sensory parsing reflected in the N1m. This is consistent with the fact that the MMNm, but not the N1m, was modulated by expertise, given that learning with precision the pitch heights and pitch intervals allowed in a musical system (i.e., learning to be “in tune”) is an essential musical skill.

Source reconstructions revealed overlapping generators for the N1m and MMNm in primary auditory cortex. This indicates that the two hierarchical processing stages are performed by the same or contiguous neural populations. Note that it has been previously argued that the neural generators of the MMN are slightly anterior to those of the N1 (e.g., Rosburg, Haueisen, & Kreitschmann-Andermahr, 2004; Sams, Kaukoranta, Hämäläinen, & Näätanen, 1991). These studies, however, typically rely on equivalent current dipoles for source estimation, whereas here we used distributed sources. Note, though, that peak activity in the right hemisphere was more anterior for the MMNm than the N1m, which would be consistent with the literature and with hierarchical processing stages. Methods such as intracranial recordings (e.g. Omigie et al., 2019) and dynamic causal modeling (Moran, Pinotsis, & Friston, 2013) could potentially disentangle the generators and the dynamics of the processes reflected in the two components, in the case of musical stimuli.

When considered together, our results present a coherent picture in which three stages of processing can be identified. In an initial stage, tonotopic neurons in A1 adapt to incoming sensory input thus modulating N1m amplitude. In a second stage, auditory objects are formed and low-level regularities are established, giving rise to the MMNm when violated. Finally, in a third stage, higher-order predictions arising from knowledge of the musical context are deployed. These are indexed by late components such as the P3am.

4.4. Implications, limitations and future directions

One limitation of our study is that the explanatory power of surprise estimates could not be assessed at the single-trial level due to the inherent noisiness of neurophysiological data. Nonetheless, this is an interesting future research direction that could benefit from techniques such as intracranial recordings and multivariate pattern analysis (e.g. Demarchi, Sanchez, & Weisz, 2019). Moreover, while here we found compelling evidence for the dissociation between interval size and estimated surprise, a more controlled study is needed to properly demonstrate this claim and confirm SSA as the underlying

mechanism. Similarly, although we employed more realistic stimuli than many previous studies on auditory predictive processing, the melodies used here are still far from being truly musical. Future experiments could address, for example, how introducing rhythm, expressive timing and dynamics and concurrent melodic lines affects predictive processing at the sensory levels reflected in the N1.

Despite these shortcomings, our findings have clear consequences for research on music and auditory predictive processing. We have shown that it is crucial to account for interval size and SSA when addressing statistical learning, even in the case of complex sounds. Regarding music perception, our results provide neural evidence that early sensory mechanisms might be fundamental for the perceptual processing of melodies and coexist with higher-order probabilistic predictions arising from knowledge of musical styles. Similarly, the evidence presented makes a contribution to the ongoing debate regarding the distinction between N1 and MMN by further suggesting that these reflect different hierarchical processing stages.

Two interesting research avenues can be derived from this study. First, we have shown that interval size can be a good predictor of N1m amplitude. However, a mechanistic link between this metric and neural responses is yet to be made. This would imply creating and refining a detailed computational model that links the acoustic properties of successive sounds with the corresponding neural activity. Second, our study shows how early sensory processing and higher-order probabilistic prediction can be disentangled. Therefore, experimental designs similar to ours might be a good way to test theories of cortical function such as predictive coding (Friston, 2005; Rao & Ballard, 1999) and active inference (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017; Friston, Rosch, et al., 2017), which have hierarchical processing at their core.

Conclusion

In this study, we aimed to determine whether the modulation of N1m amplitude by auditory surprise was different between musicians and non-musicians. Using a computational model of auditory expectation, we showed no differences between the groups in the otherwise clear association between neural responses and estimated surprise. Further exploratory analyses suggested that interval size and stimulus-specific adaptation, rather than probabilistic prediction, underlie the observed effects. This offers an explanation to why no effect of expertise was found, since we would expect higher-order probabilistic prediction, instead of early sensory processing, to be affected by the accuracy of musical knowledge. Interestingly, our results also suggest that auditory regularities and probabilistic prediction are reflected in later processing stages indexed by the MMNm and the P3am. Overall, our findings reveal a hierarchy of expectations in the auditory system, including early sensory processing, the formation of low-level regularities and higher-order probabilistic prediction. Therefore, our study constitutes an advance towards the understanding of hierarchical predictive processing in complex and realistic auditory contexts.

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 Varankaite, Giulio Carraturo, Riccardo Proietti, and Claudia Iorio for assistance during MEG recordings. We are indebted as well to the group of Italian trainees fromIISS 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), which did not have any influence on the scientific content of this article.

Declaration of interests

None

Open practices

Materials and data for the study are available at https://osf.io/my6te/; DOI: 10.17605/OSF.IO/MY6TE

CRediT authorship contribution statement

David R. Quiroga-Martinez: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Visualization, Investigation. Niels C. Hansen: Conceptualization, Methodology, Writing - original draft. Andreas Højlund: Conceptualization, Methodology, Software, Writing - original draft, Supervision. Marcus T. Pearce: Software, Formal analysis, Writing - original draft. Elvira Brattico: Conceptualization, Supervision, Writing - original draft. Peter Vuust: Conceptualization, Methodology, Supervision, Writing - original draft.

References

Agres, K., Abdallah, S., & Pearce, M. (2018). Information-Theoretic Properties of Auditory Sequences Dynamically Influence Expectation and Memory. *Cognitive Science*, *42*(1), 43–76. https://doi.org/10.1111/cogs.12477

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences*, *106*(5), 1672–1677. https://doi.org/10.1073/pnas.0809667106

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

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

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Butler, R. A. (1968). Effect of Changes in Stimulus Frequency and Intensity on Habituation of the Human Vertex Potential. *The Journal of the Acoustical Society of America*, *44*(4), 945–950. https://doi.org/10.1121/1.1911233

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford ; New York: Oxford University Press.

Demarchi, G., Sanchez, G., & Weisz, N. (2019). Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nature Communications*, *10*(1), 3440. https://doi.org/10.1038/s41467-019-11440-1

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

Escera, C., & Malmierca, M. S. (2014). The auditory novelty system: An attempt to integrate human and animal research. *Psychophysiology*, *51*(2), 111–123. https://doi.org/10.1111/psyp.12156

Friston, K. J. (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. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Computation*, *29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Friston, K. J., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., … Mattout, J. (2008). Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, *39*(3), 1104–1120. https://doi.org/10.1016/j.neuroimage.2007.09.048

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2017). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, *77*, 388–402. https://doi.org/10.1016/j.neubiorev.2017.04.009

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*(7), 348–353. https://doi.org/10.1016/S0166-2236(02)02191-4

Hansen, N. C., & Pearce, M. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*. 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), e0163584. https://doi.org/10.1371/journal.pone.0163584

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Huron, D. B. (2006). Sweet anticipation: Music and the psychology of expectation. Cambridge, Mass: MIT Press.

Jaaskelainen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levanen, S., … Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences*, *101*(17), 6809–6814. https://doi.org/10.1073/pnas.0303760101

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

Koelsch, S., & Jentschke, S. (2010). Differences in Electric Brain Responses to Melodies and Chords. *Journal of Cognitive Neuroscience*, *22*(10), 2251–2262. https://doi.org/10.1162/jocn.2009.21338

Krumhansl, C. L. (1990). Cognitive foundations of musical pitch. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=144048

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. https://doi.org/10.1371/journal.pcbi.1003288

Manca, A. D., & Grimaldi, M. (2016). Vowels and Consonants in the Brain: Evidence from Magnetoencephalographic Studies on the N1m in Normal-Hearing Listeners. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01413

Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage*, *49*(2), 1271–1281. https://doi.org/10.1016/j.neuroimage.2009.10.002

Masson, R., & Bidet-Caulet, A. (2019). Fronto-central P3a to distracting sounds: An index of their arousing properties. *NeuroImage*, *185*, 164–180. https://doi.org/10.1016/j.neuroimage.2018.10.041

May, P. J. C., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. https://doi.org/10.1111/j.1469-

8986.2009.00856.x
May, P. J. C., Westö, J., & Tiitinen, H. (2015). Computational modelling suggests that temporal integration results from synaptic adaptation in auditory cortex. *European Journal of Neuroscience*, *41*(5), 615–630. https://doi.org/10.1111/ejn.12820
Moran, R. J., Pinotsis, D. A., & Friston, K. J. (2013). Neural masses and fields in dynamic causal modeling. *Frontiers in Computational Neuroscience*, *7*. https://doi.org/10.3389/fncom.2013.00057
Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and Gestalt-like principles predict melodic expectations. *Cognition*, *189*, 23–34. https://doi.org/10.1016/j.cognition.2018.12.015
Morrison, S. J., Demorest, S. M., & Pearce, M. (2018). Cultural Distance: A Computational Approach to Exploring Cultural Influences on Music Cognition (Vol. 1; M. H. Thaut & D. A. Hodges, Eds.). https://doi.org/10.1093/oxfordhb/9780198804123.013.3
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*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642
Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, *69*(6), 523–531. https://doi.org/10.1016/0013-4694(88)90164-2
Näätänen, Risto, Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9
Näätänen, Risto, Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x
Näätänen, Risto, Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026
Näätänen, Risto, & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x
Nelken, I., Yaron, A., Polterovich, A., & Hershenhoren, I. (2013). Stimulus-Specific Adaptation Beyond Pure Tones. In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), *Basic Aspects of Hearing* (Vol. 787, pp. 411–418). https://doi.org/10.1007/978-1-4614-1590-9_45
Omigie, D., Pearce, M., Lehongre, K., Hasboun, D., Navarro, V., Adam, C., & Samson, S. (2019). Intracranial Recordings and Computational Modeling of Music Reveal the Time Course of Prediction Error Signaling in Frontal and Temporal Cortices. *Journal of Cognitive Neuroscience*, *31*(6), 855–873. https://doi.org/10.1162/jocn_a_01388
Omigie, D., Pearce, M., 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*, 1–9. https://doi.org/10.1155/2011/156869
Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, *8*(1), 2148. https://doi.org/10.1038/s41467-017-02038-6
Pearce, M. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. City University London.
Pearce, M. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. https://doi.org/10.1111/nyas.13654
Pearce, M., 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
Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: An overview. *Frontiers in Integrative Neuroscience*, *8*. https://doi.org/10.3389/fnint.2014.00019
Picton, T. W., Woods, D. L., & Proulx, G. B. (1978). Human auditory sustained potentials. I. The nature of the response. *Electroencephalography and Clinical Neurophysiology*, *45*(2), 186–197. https://doi.org/10.1016/0013-4694(78)90003-2
Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*(2), 172–185. https://doi.org/10.1016/j.ijpsycho.2005.12.012
Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M., Brattico, E., & Vuust, P. (2019a). Musical prediction error responses similarly reduced by predictive uncertainty in musicians and non-musicians [Preprint]. https://doi.org/10.1101/754333
Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M., Brattico, E., & Vuust, P. (2019b). Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. *Cortex*, *120*, 181–200. https://doi.org/10.1016/j.cortex.2019.06.010
R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.R-project.org
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
Rauschecker, J., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, *268*(5207), 111–114. https://doi.org/10.1126/science.7701330
Rosburg, T., Haueisen, J., & Kreitschmann-Andermahr, I. (2004). The dipole location shift within the auditory evoked neuromagnetic field components N100m and mismatch negativity (MMNm). *Clinical Neurophysiology*, *115*(4), 906–913. https://doi.org/10.1016/j.clinph.2003.11.039
Sams, M., Kaukoranta, E., Hämäläinen, M., & Näätanen, R. (1991). Cortical Activity Elicited by Changes in Auditory Stimuli: Different Sources for the Magnetic N1OOm and Mismatch Responses. *Psychophysiology*, *28*(1), 21–29. https://doi.org/10.1111/j.1469-8986.1991.tb03382.x
Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, e13335. https://doi.org/10.1111/psyp.13335
Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Cognitive Brain Research*, *21*(3), 342–350. https://doi.org/10.1016/j.cogbrainres.2004.06.011
Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1
Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology*, *51*(7), 1759–1768. https://doi.org/10.1088/0031-9155/51/7/008
Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology*, *38*(2–3), 157–167. https://doi.org/10.1016/0301-0511(94)90036-1
Ulanovsky, N. (2004). Multiple Time Scales of Adaptation in Auditory Cortex Neurons. *Journal of Neuroscience*, *24*(46), 10440–10453. https://doi.org/10.1523/JNEUROSCI.1905-04.2004
Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032
Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01111
Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wacongne, C., Labyt, E., Wassenhove, V. van, 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*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

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

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, *11*, 3571-3594.

Worsley, K. (2007). Random Field Theory. In K. J. Friston, J. Ashburner, S. Kiebel, T. E. Nichols, & W. Penny (Eds.), *Statistical Parametric Mapping* (pp. 232–236). https://doi.org/10.1016/B978-012372560-8/50018-8

Yarden, T. S., & Nelken, I. (2017). Stimulus-specific adaptation in a recurrent network model of primary auditory cortex. *PLOS Computational Biology*, *13*(3), e1005437. https://doi.org/10.1371/journal.pcbi.1005437