Cortical encoding of melodic expectations in human temporal cortex

Giovanni M. Di Liberto[1], Claire Pelofi[2,3,*], Roberta Bianco[4,*], Prachi Patel[5,6], Ashesh D. Mehta[7], Jose L. Herrero[7], Alain de Cheveigné[1,4], Shihab Shamma[1,8,**], Nima Mesgarani[5,6,**]

1 Laboratoire des Systèmes Perceptifs, UMR 8248, CNRS, France. Ecole Normale Supérieure, PSL University, France

2 Department of Psychology, New York University, New York, NY, USA

3 Institut de Neurosciences des Système, UMR S 1106, INSERM, Aix Marseille Université, France

4 UCL Ear Institute, London, United Kingdom

5 Department of Electrical Engineering, Columbia University, New York, NY, USA

6 Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, United States
7 Department of Neurosurgery, Zucker School of Medicine at Hofstra/Northwell and Feinstein Institute of Medical Research, Manhasset, NY 11030, United States

8 Institute for Systems Research, Electrical and Computer Engineering, University of Maryland, College Park, USA

** The authors contributed equally*

*** Co-senior authors*

Correspondence: diliberg@tcd.ie

Conflicts of interest: none declared.

Funding sources: This study was funded by an Advanced ERC grant (Neume). G.D.L. and A.d.C. were supported by the EU H2020-ICT grant 644732 (COCOHA). N.M., P.P., A.D.M., and J.L.H. were supported by NIMH MH114166-01. C.P. and R.B. were supported by the Agence Nationale de la Recherche (ANR-16-CE28-0012 RALP) and by a BBSRC grant (BB/P003745/1) respectively.

Acknowledgements: Marcus Pearce and Jens Hjortkjær for useful discussion. Gaelle Rouvier for her help with the data collection. Part of the data analysis and discussions were conducted at the Telluride Cognitive Neuromorphic Engineering Workshop.

Summary

Humans engagement in music rests on underlying elements such as the listeners' cultural background and general interest in music, all shaping the way music is processed in the brain and perceived. Crucially, these factors modulate how listeners anticipate musical events, a process inducing instantaneous neural responses as the music confronts these expectations. Measuring such neural correlates would represent a direct window into high-level brain processing of music. Here we recorded electroencephalographic and electrocorticographic brain responses as participants listened to Bach melodies. We assessed the relative contributions of the acoustic *versus* melodic components of the music to the neural signal. Acoustic features included envelope and its derivative. Melodic features included information on melodic progressions (pitch) and their tempo (onsets), which were extracted from a Markov model predicting the next note based on a corpus of Western music and the preceding proximal musical context. We related the music to brain activity with a linear temporal response function, and demonstrated that cortical responses to music encode melodic expectations. Specifically, individual-subject neural signals were better predicted by a combination of acoustic *and* melodic expectation features than by either alone. This effect was most pronounced at response latencies up to 350ms, and in both *planum temporale* and *Heschl's gyrus*. Finally, expectations of pitch and onset-time of musical notes exerted independent cortical effects, and such influences were modulated by the listeners' musical expertise. Overall, this study demonstrates how the interplay of experimental and theoretical approaches can yield novel insights into the cortical encoding of melodic expectations.

Experiencing music as a listener, performer, or a composer is an active process that engages perceptual and cognitive faculties, endowing the experience with memories, joy, and emotion. Through this active auditory engagement, humans analyze and comprehend complex musical scenes by invoking the cultural norms of music, segregating sound mixtures, and marshaling expectations and anticipation. However, this process rests on the “musical knowledge” that listeners acquire and encode through frequent exposure in their daily lives (Tillmann et al. 2000). Ultimately, this knowledge shapes their expectations, determining what constitutes a “familiar” musical style that they are likely to understand and appreciate (Morrison et al. 2008; Hannon et al. 2012; Pearce 2018). A similar process has been demonstrated in many domains of human learning (e.g., language; Saffran et al. 1997; Toro et al. 2005; Finn et al. 201s4), as well as in animals that must learn their species-specific songs and vocalizations (Woolley 2012). Theoretically, this is often conceptualized as learning the statistical regularities of the sensory environment (Romberg and Saffran 2010; Erickson and Thiessen 2015; Bretan et al. 2017; Skerritt-Davis and Elhilali 2018) so as to predict upcoming events, in a process referred to as "statistical learning". Supporting evidence is that failed predictions due to deviations in the sensory sequence produce measurable brain activations that have been associated with the detection and strength of these irregularities, i.e. prediction error (Vuust et al. 2012; Clark 2013; Moldwin et al. 2017; Omigie et al. 2019; Quiroga-Martinez et al. 2019), and with learning (Storkel and Rogers 2000; Attaheri et al. 2015; Qi et al. 2017).

Music sequences induce a wide range of violation strengths (Pearce and Wiggins 2012) because the sequential events are not all equally likely (and hence not equally predictable). However, widely reported ERP studies (Besson and Macar 1987; Paller et al. 1992; Miranda and Ullman 2007; Pearce, Ruiz, et al. 2010; Carrus et al. 2013) often limit the range of expectations’ violations strength that are tested to severe violations because of the need for repeated stimulus presentations. One issue with this approach is that notes eliciting strong violations could be considered by the listener as production mistakes (e.g., a pianist playing the wrong note), thus the corresponding cortical correlates may characterise only a limited aspect of the neural underpinnings of melodic perception. Instead, here we study the brain responses to continuous naturalistic musical stimuli, thus allowing us to explore the full range of expectations’ strengths. This study seeks to determine whether the neural response reflects this range of strengths by regressing the musical information with it.

Music is a complex, multi-layered signal that have structures allowing predictions of a variety of properties at multiple time-scales. Pearce et al. designed a framework based on variable-order Markov models that learns statistics describing the temporal sequences in monophonic music at these various time-scales (IDyOM; Pearce 2005; Hansen and Pearce 2014). This framework attempts to optimally predict the next note in a melodic sequence by combining predictions based on 1) long-term statistics learnt from a large corpus of western music and 2) short-term statistics from the previous notes of the current musical stream. In turn, this provides us with likelihood values for each note in a melody that have been shown to tightly mirror listeners’ expectations (Pearce 2005).

Here, we regressed these estimates with electrophysiological data to investigate the impact of expectations on auditory perception and cortical processing. We recorded scalp electroencephalography (EEG; 20 subjects; about 1 hour and 15 minutes of data) signals and invasive electrocorticography (ECoG; 3 patients; about 25 minutes of data each) signals as participants listened to monophonic piano music from Bach that was generated from MIDI scores

(see **Methods**). According to the predictive coding theory (Friston and Kiebel 2009; Clark 2013), cortical signals partly reflect the mismatch between a participant's prediction and the actual sensory input. If this is the case, less expected musical notes should produce relatively stronger cortical responses. We expected individuals with musical expertise to internally generate predictions on a next note that more closely relate the ones of a specialized statistical model than non-musicians. We tested this hypothesis on our EEG dataset, which was recorded from both non-musicians and expert pianists, and investigated specific details of the cortical response using our ECoG dataset.

Our stimuli were regular musical streams, rather than artificially-constructed repeated patterns such as those usually utilized for the induction and detection of mismatched negativity (MMN) responses (Garrido et al. 2009; Clark 2013; Fishman 2014; Lecaignard et al. 2015; Southwell and Chait 2018). Musical streams are imbued with melodic events that routinely violates listeners' expectation to some degree (Pearce and Wiggins 2012; Salimpoor et al. 2015). Here we sought to test the hypothesis that the listeners' expectation violations produce cortical responses that change with the degree of the violations and that are measurable with EEG during naturalistic music listening. By tackling these questions, our experiment aims to provide novel insights into the precise spatio-temporal dynamics of the integration of melodic expectations and sensory input during naturalistic musical listening. In turn, we aim to provide evidence of the neurophysiological validity of predictive statistical models of music structure. In particular, we aim to assess the relative contribution and dissociation of distinct melodic expectation features (such as pitch and note onset) to the cortical responses to music.

To establish the contribution of the various signal features (**Fig. 1*A***) to the neural responses, we used multivariate ridge regression (**Fig. 1*B***) to quantify how well the acoustic (**A**) factors (e.g., signal envelope and its derivative) and melodic expectation or surprise (**M**) factors (e.g., pitch and onset-timing) can predict the EEG and ECoG responses to music (Crosse et al. 2016). Since the prediction quality is considered to be an estimate of how strongly a stimulus property is encoded in the EEG data (Di Liberto et al. 2015, 2019; Brodbeck, Presacco, et al. 2018; Somers et al. 2018; Verschueren et al. 2019), and since cortical signals are assumed to be modulated by the various **A** and **M** factors above, we consequently expected the combination of both the acoustic and surprise features to predict the neural responses better than either set of features alone (**Fig. 1*C***). Validating these hypotheses would therefore provide physiological support for the melodic expectations generated according to the statistical learning model.

Results

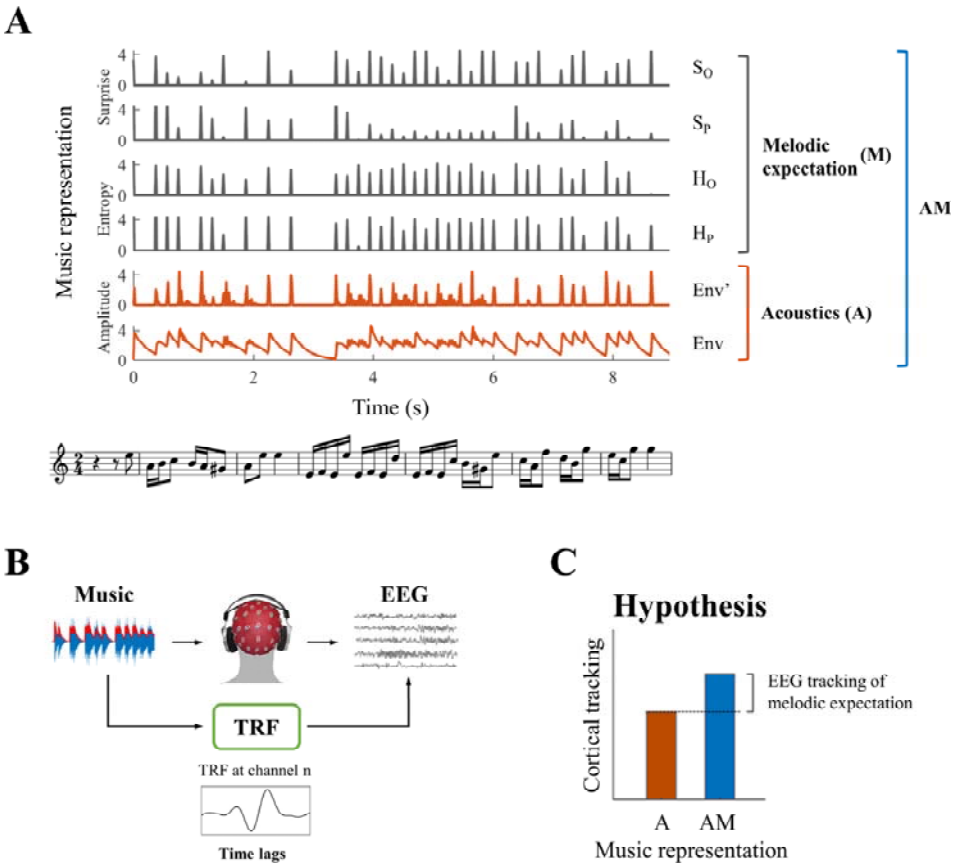

Figure 1. System identification framework for isolating neural correlates of melodic expectations. (*A*) Music score of a segment of auditory stimulus, with its corresponding features (from bottom to top): acoustic envelope (Env), half-way rectified first derivative of the envelope (Env'), and the four melodic expectation features: entropy of note-onset (H_o) and pitch (H_p), surprise of note-onset (S_o) and pitch (S_p). (*B*) Regularized linear regression models were fit to optimally describe the mapping from stimulus features (Env in the example) to each EEG and ECoG channel. This approach, called the Temporal Response Function (TRF), allows us to investigate the spatio-temporal dynamics of the linear model by studying the regression weights for different EEG and ECoG channels and time-latencies. (*C*) TRFs were used to predict EEG and ECoG signals on unseen data by using only acoustic features (A) and a combination of acoustic and melodic expectation features (AM). We hypothesised that cortical signals encode melodic expectations, therefore we expected larger EEG and ECoG predictions for the combined feature-set AM.

Neural data were recorded from twenty healthy EEG participants and three ECoG epilepsy patients as they listened to monophonic excerpts of music from Bach sonatas and partitas that were synthesized with piano sound. The melodic expectation of each note was estimated from the musical score of the stimulus with IDyOM, the model for predictive statistical modelling of musical structure. Specifically, given a musical piece at time t_0, the model estimates the likelihood of having a note with a particular pitch at time t_0 given short-term information for $t < t_0$ and long-term information from a large corpus of Western music. Based on these estimates we calculated four measures (referred to as melodic features M, see **Methods** for details) that capture distinct aspects of expectation and surprise at each new note within a melody: entropy of pitch (H_p), entropy of onset-time (H_o), surprise of pitch (S_p), and surprise of onset-time (S_o). The first two measures refer to the Shannon entropy at a particular position in a melody, before the musical note is observed. Intuitively, the entropy indicates the amount of uncertainty represented by the

distribution of pitch and onset-time for the next note, where the most uncertain scenario is when all possible notes have equal likelihood and the least uncertain scenario is when the next note is known. The latter two measures refer to the *inverse* probability of occurrence of a particular note pitch or onset-time, with smaller values for more predictable transitions (see **Methods**).

Melodic expectation encoding in low-rate cortical signals. In all of the analyses and results below, we focused on the EEG and ECoG responses in the low-rate bands between 1 and 8 Hz, filtering out the remainder of the bands (see **Methods**; note that inclusion of rates down to 0.1 Hz and up to 30 Hz did not alter any of the results that follow). Because of potential interactions between the responses to the succession of notes (which would complicate the interpretation of the event-related potentials (ERPs) time-locked to note onsets), we began by utilizing a linear modelling framework known as the temporal response function (TRF) (Ding and Simon 2012a; Crosse et al. 2016) as depicted in **Figure 1*B***. This approach 1) explicitly dissociates the effects of expectations from those due to changes in the acoustic envelope on the neural responses to music and 2) allows us to investigate neural responses to rapidly presented stimuli by accounting for the dependence among the sequences of input notes. Specifically, TRFs were derived by using ridge regression between suitably parameterized stimuli and their neural responses. These were then used to predict unseen EEG data (with leave-one-out cross-validation) based on either the acoustic properties alone (**A** predictions) or a combination of acoustics and melodic expectation features (**AM** predictions). The predictive models included time-lagged versions of the stimulus accounting for delays between the stimulus and corresponding neural response. The time-lag window was limited to [0, 350] ms as longer latencies had little impact on the predictive power of the TRF model (see **Methods**).

In **Figure 2** we illustrate the average (**Fig. 2*A***) and individual (**Fig. 2*B***) EEG prediction correlations for all subjects using either the **A** features alone (envelope and its derivative) or combined with the melodic expectation features, **AM.** The **A** correlations were significantly positive ($p < 0.05$, permutation test) for all subjects but one (S20), confirming that neural responses to monophonic music track the stimulus envelopes (henceforth 'envelope tracking'). Crucially, **AM** correlations were significantly larger than **A**, implying that melodic expectations explained EEG variance that was not captured by acoustic information alone. Specifically, the average EEG prediction correlation over all electrodes was significantly larger for **AM** than for **A** both at the group level ($r_{AM} > r_A$: permutation test, $p < 10^{-6}$, $d = 1.64$; **Fig. 2*A***) and at the individual-subject level (16 out of 20 subjects, permutation test, $p < 0.05$; **Fig. 2*B***), and this difference was even larger when computed from selected single electrodes (e.g., Cz channel: $d = 1.80$). This supports the hypothesis that melodic expectation is directly reflected in the EEG responses to music.

While these results indicate that **AM** is a better descriptor of the signal than **A**, it should also be noted that TRF_{AM} has higher dimensionality than TRF_A. Correlations are measured on unseen data, and thus should be immune to overfitting, nonetheless to verify that the better predictions for **AM** are not due simply to the higher degrees of freedom afforded by the addition of more **M** components, we assessed the performance of our TRF_{AM} model using less elaborate **M** functions that nevertheless maintained the TRF_{AM} with the same dimensionality and value distributions. Specifically, we built melodic expectation estimates by relying on progressively smaller amount of local context (or short-term memory) (**Fig. 2*C***), with **M** based on 1, 2, 4, 8, 16, and 32 musical bars, as opposed to the unbounded ('∞') estimates in our predictions of **Figure 2*A***. In each of these cases, fitted TRFs of the same dimensionality performed better as the memory increased

(**Fig. 2*C***; ANOVA: $F(4.6,101.1) = 4.52$, $p = 0.0003$), indicating that the contribution of melodic expectation estimates to the prediction accuracy was not due to increased TRF dimensionality per se.

Despite the significant positive effects of melodic expectation on prediction correlations, we found no differences between the corresponding EEG topographical distributions for **A**, **AM**, or their difference **AM-A** ($DISS_{A,AM} = 0.017$, $p = 0.33$; $DISS_{A,AM-A} = 0.214$, $p = 0.61$; $DISS_{AM,AM-A} = 0.197$, $p = 0.57$; **Fig. 2*D***). By contrast, melodic expectations induced new *long temporal latencies* in the linear regression weights of the TRF_{AM} model (**Fig. 2*E***), which were mostly centered around 200 ms compared to the 50 ms latency of the acoustic TRF_A Env component ($p < 0.05$, FDR correction).

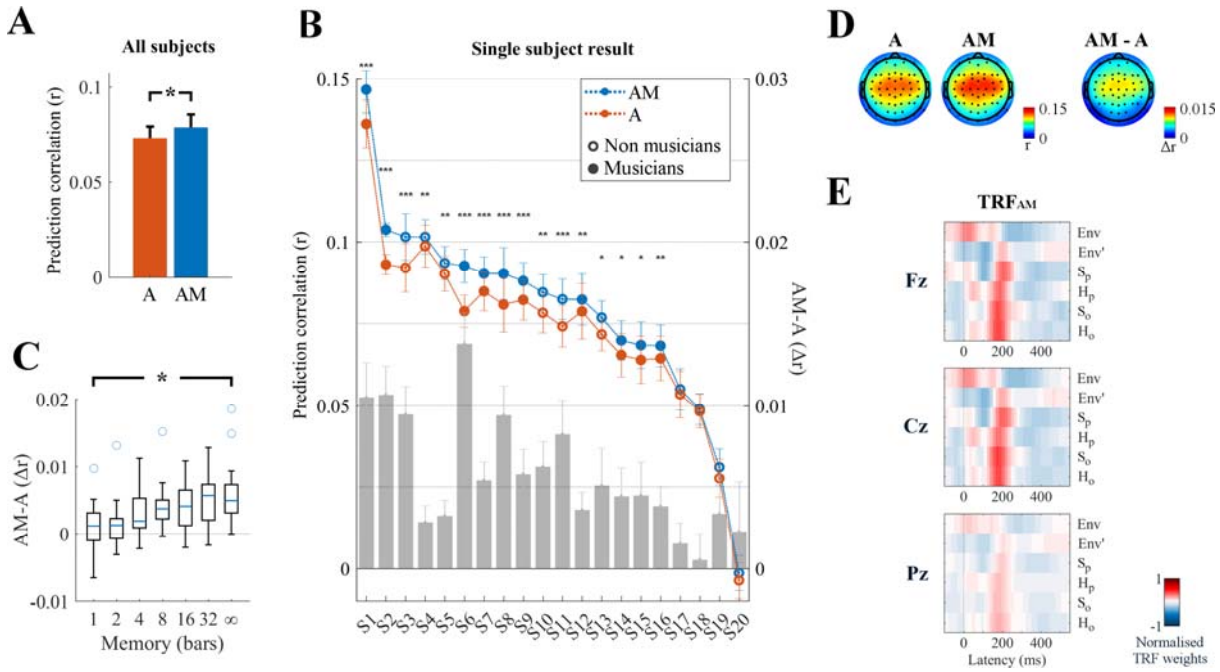

Figure 2. Low-rate (1 - 8 Hz) cortical signals reflect melodic expectations. Scalp EEG data were recorded while participants listened to monophonic music. Forward ridge regression models were fit to assess what features of the stimulus were encoded in the low-rate EEG signal. The link between music features and EEG was assessed by using these models to predict unseen EEG data. (*A*) Prediction correlations were greatest when the stimulus was described by a combination of acoustic information (**A**: envelope Env, and its half-way rectified first derivative Env') and melodic expectations (**M**: S_P, S_O, H_P, H_O). This effect of expectations was significant on the average prediction correlation across all 64 EEG electrodes ($*p < 10^{-6}$). The error bars indicate the SEM across participants. (*B*) The enhancement due to melodic expectations emerged at the individual subject level. The gray bars indicate the predictive enhancement due to melodic expectation. Error bars indicate the SEM over trials ($***p < 0.001$, $**p < 0.01$, $*p < 0.05$, permutation test). (*C*) Predictive enhancement due to melodic expectations (**AM-A**; *y*-axis) increased with the length of the local context (in bars; *x*-axis) used to estimate the expectation of each note (ANOVA: $*p = 0.0003$). (*D*) The effect of melodic expectations (r_{AM}-r_A) emerged bilaterally on the same scalp areas that showed also envelope tracking. (*E*) Ridge regression weights for TRF_{AM}. Red and blue colors indicate positive and negative TRF components respectively.

Melodic expectations modulate auditory responses in higher cortical areas. Since melodic expectations reflect regularities within a musical tone sequence at multiple time-scales that depend

on the extent of knowledge and exposure of the subject listening to them, we hypothesized that neural signals correlated with the melodic properties of the music would be generated at higher hierarchical cortical levels than those strictly due to the acoustics (Sammler et al. 2013; Bianco et al. 2016; Nourski et al. 2018). EEG lacks the spatial resolution needed to test this hypothesis, but the test was possible in spatially localized ECoG recordings from three patients who had electrodes over the early primary auditory areas in the *anterior transverse temporal gyrus*, also called *Heschl Gyrus* (HG; patients 1 and 3), the belt regions along *planum temporale* (PT) and the *superior temporal gyrus* (STG), as well as the *supra-marginal gyrus* (SMG) in the parietal lobe (see **Tables S1-S3** for details on the channel locations). The subjects listened to the same monophonic music described earlier for the EEG experiments.

We first identified 21/241, 25/200, and 33/285 electrodes in Patients 1, 2, and 3 respectively that exhibited reliable auditory responses (stronger responses to monophonic music than to silence: Cohen's $d > 0.5$) in the form of either low-rate (1 – 8 Hz) local field potentials (similar bands to those in EEG analyses above), or power in the high-γ (70 – 150 Hz) field potentials which are thought to reflect local neuronal activity (Miller et al. 2007; see **Methods** for details). A similar TRF analysis was conducted to identify responses that were sensitive to melodic expectations (**Fig. 3*A*** and **Fig. S1*A***). Envelope tracking was significant (permutation test on the ECoG prediction correlations over trials, $p < 0.05$) in STG, PT, and HG channels. The predictive enhancement due to melodic expectations was small but significant ($p < 0.01$, FDR-corrected) on several electrodes in PT in all patients, on one electrode in the *transverse temporal sulcus* (TTS), but not on the three HG electrodes in Patient 1, and was also significant on five bilateral HG electrodes in Patient 3. A Wilcoxon rank sum test indicated that the effect of expectations **AM-A** is larger in PT than HG ($p = 0.011$; all electrodes in the two cortical areas from Patients 1 and 3 were combined, while Patient 2 was excluded as there was no coverage in HG). Similar effects as in PT were also measured in right parietal cortical areas (SMG and the *postcentral gyrus*) for Patient 2 but not for Patient 3. In addition, the effect of expectations seen in STG, PT, and HG was right lateralized in Patient 3 ($p = 0.038$), while envelope tracking did not show an hemispheric bias ($p = 0.85$).

Figure 3*B* depicts the TRF_{AM} weights for selected electrodes in SMG, PT, TTS, and HG. The TRF_{AM} weights for ECoG exhibited low-rate temporal patterns very similar to those measured with EEG. Specifically, strong correlations were found with the TRF_{AM} measured with EEG at Cz ($r = 0.60$, 0.50, and 0.45 in left PT, TTS, and HG respectively – e2, e6, and e9 from Patient 1; $r = 0.61$ and 0.80 in right PT and SMG – e4 and e10 from Patient 2 respectively). Overall, the TRF analysis of the ECoG recordings demonstrates that low-rate cortical responses to music encode melodic expectations, suggesting that the similar EEG responses to melodic expectations may well originate from temporal regions between PT and HG, and or parietal regions such as SMG.

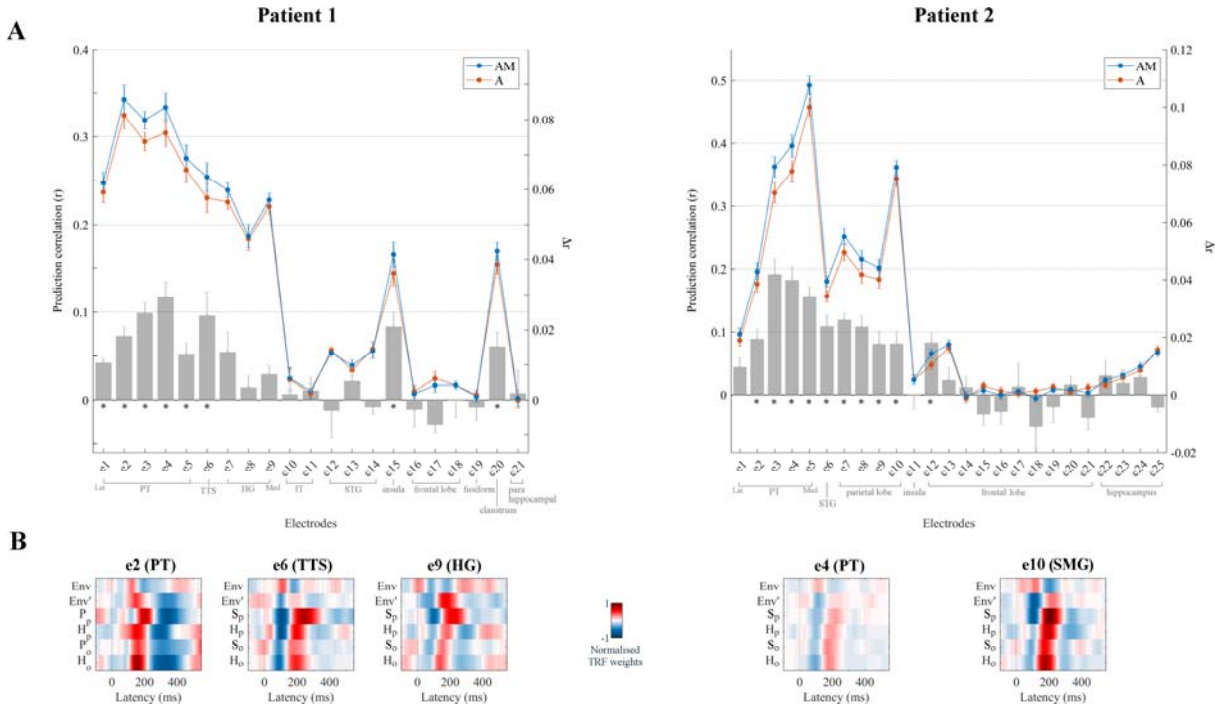

Figure 3. Low-rate (1 - 8 Hz) cortical signals in bilateral temporal cortex reflect melodic expectations. Electrocorticography (ECoG) data were recorded from three epilepsy patients undergoing brain surgery. Magnetic resonance imaging was used to localise the ECoG electrodes. Electrodes with stronger low-rate or high-γ (70 – 150 Hz) responses to monophonic music than to silence were selected (Cohen's $d > 0.5$). (*A*) ECoG prediction correlations for individual electrodes for **A** and **AM**. Electrodes within each group, as indicated by the gray square brackets, were sorted from lateral to medial cortical sites. The gray bars indicate the predictive enhancement due to melodic expectation (r_{AM}-r_A). Error bars indicate the SEM over trials (*$p < 0.01$, FDR-corrected permutation test). (*B*) TRF weights for selected electrodes. For Patient 1, PT and TTS electrodes (e1-e5 and e6 respectively) exhibited large effects of musical expectations, while HG electrodes (e7-e9) had strong envelope tracking (**A**) but showed smaller effects of expectations that did not reach statistical significance. Patient 2 showed also strong envelope tracking and a significant effect of melodic expectations in the right temporal and parietal lobes (for example, the PT electrode e4 and the SMG electrode e10 respectively).

ECoG recordings also allowed us to investigate more directly the link between local neuronal activity and melodic expectations, since these signals are available in the instantaneous power of the high-γ field potentials (Crone et al. 2001; Edwards et al. 2009; Ray et al. 2008; Steinschneider et al. 2008). Again, TRF analysis was used in **Figure 4** (and in **Fig. S1*B***) to disentangle the contributions of envelope tracking and melodic expectations. As before, left HG (in Patients 1 and 3), left TTS (Patient 1), bilateral STG (Patients 2 and 3), bilateral PT (Patients 1, 2, and 3), and right SMG (Patient 2 but not Patient 3) electrodes exhibited substantial envelope tracking. By contrast, the effect of expectations ($\Delta r = r_{AM} - r_A$) was largest in PT, TTS, and in HG electrodes close to the junction between PT and HG, with a predictive enhancement up to ~50% (e.g., Δr_{e6} = 0.09 in Patient 1, which corresponds to a prediction enhancement $r_{AM,e6}/r_{A,e6}$ of 149%), in contrast to an enhancement of only 6% in the HG electrode with strongest envelope tracking (e9). Similar patterns emerged for Patient 3, with a predictive enhancement up to ~20% in PT and an enhancement of only 5% in the HG electrode with strongest envelope tracking (e9; but with stronger effects in other HG electrodes with weaker envelope tracking). Also, the temporal latencies in the TRF weights (**Fig. 4*B***) were rather different from what was previously seen for low-

rate EEG and ECoG signals. In fact, the TRF_A weights corresponding to the acoustic features exhibited sharp, short-latency dynamics while those of the melodic expectation features (TRF_{AM}) pointed to more temporally extended and strong neural responses.

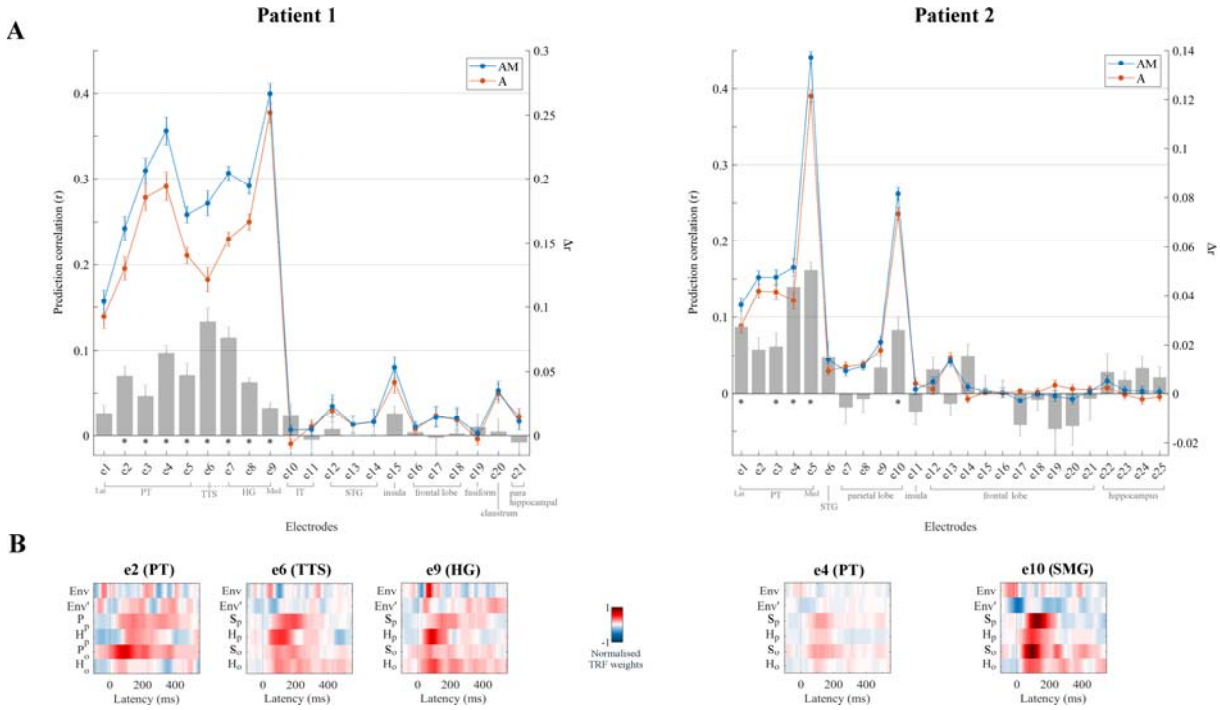

Figure 4. High-γ neural signals in bilateral temporal cortex reflect melodic expectations. Electrodes with stronger low-rate (1 – 8 Hz) or high-γ (70 – 150 Hz) responses to monophonic music than to silence were selected (Cohen's d > 0.5). (*A*) ECoG prediction correlations for individual electrodes for **A** and **AM**. Electrodes within each group, as indicated by the gray square brackets, were sorted from lateral to medial cortical sites. The gray bars indicate the predictive enhancement due to melodic expectation (r_{AM}-r_A). Error bars indicate the SEM over trials (*p < 0.01, FDR-corrected permutation test). (*B*) Normalised TRF weights for selected electrodes (same electrodes as for Fig. 3). For Patient 1, the HG electrode e9 showed the strongest envelope tracking and small effect of melodic expectations, while e6 in TTS exhibited the largest effect of expectations ($\Delta r_6 > \Delta r_9$, $p = 1.8e^{-4}$, $d = 2.38$). For Patient 2, both e4 (PT) and e10 (SMG) electrodes showed strong envelope tracking and a significant effect of melodic expectations.

Explicit encoding of melodic expectations in the evoked-responses.

So far, melodic effects were extracted in terms of the temporally extended analysis of the TRF, and indirectly validated through assessment of prediction accuracy. A more direct measure of these effects is possible by examining whether event-related potentials (ERPs) time-locked to note onsets are specifically modulated by melodic expectations, i.e., *beyond* what is expected from the acoustic features of the stimuli. For instance, here we specifically demonstrate that the cortical responses evoked by tones of identical envelope can produce significantly different responses that are modulated proportionately to the melodic values of the tones. To do so, we selected notes with equal acoustic envelopes corresponding to the median peak envelope amplitude (25% of all notes), but had large disparity in their surprise values S_p according to the IDyOM model, namely the top 20% and bottom 20% (**Fig. 5**, blue and red respectively). As illustrated in **Figure 5**, the two groups (red and blue curves) had identical average signal envelopes (**Fig. 5*A***), but displayed significantly disparate EEG responses (**Fig. 5*B***), with significantly larger responses to the notes

with the high pitch surprise (blue > red ; $p < 0.05$ at the N1 and P2 peaks, permutation test; $p = 0.001$ on the power of the average ERP across all channels for latencies between 0 and 200ms). A similar effect emerged for H_p and H_o (average power ERP within 0–200ms, high surprise > low surprise with $p = 0.0425$ and $p = 0.006$ for H_p and H_o respectively*; not shown*), while no effect was measured for S_o ($p = 0.8764$). Note that the ERPs showed large responses at pre-stimulus latencies (before zero latency). This is due to the temporal regularities that are intrinsic in music, which results in large average envelope before the note of interest (see **Fig. 5*A***). In fact, limiting the ERP calculation to musical events with preceding inter-note-interval longer than 200 ms eliminated such pre-stimulus responses from the ERPs (*not shown*). However, this selection procedure reduced the number of EEG epochs, and thus our choice to include short inter-note-interval in the analysis in **Figure 5**.

Similar analyses for both low-rate and high-γ ECoG data revealed that ERP responses in TTS electrodes to musical notes with *equal* envelopes were modulated in proportion to the S_p (**Fig. 5*C***) stats). This effect of melodic surprise was absent in the electrode with strongest envelope tracking e9 in Patient 1 (in the left HG; **Fig. 5*C***). These results are consistent with previous findings on melodic expectations (Omigie et al. 2013, 2019) and complements the TRF result by confirming that the effect of melodic expectations on the cortical responses can be disentangled from changes in the amplitude of the stimulus envelope. It should be emphasized, however, that compared to the TRF approach, this analysis may in many cases suffer from the potential of interactions between the responses to the sequence of notes, e.g., if the inter-note-interval is shorter than the duration of the neural response of interest. It also cannot isolate among the interactions and modulations due to the various melodic expectation features. Nevertheless, the validity of these results is confirmed by the parallel TRF findings in **Figures 2-4**, that the encoding of melodic expectations in the cortical responses is different from responses due to stimulus acoustics.

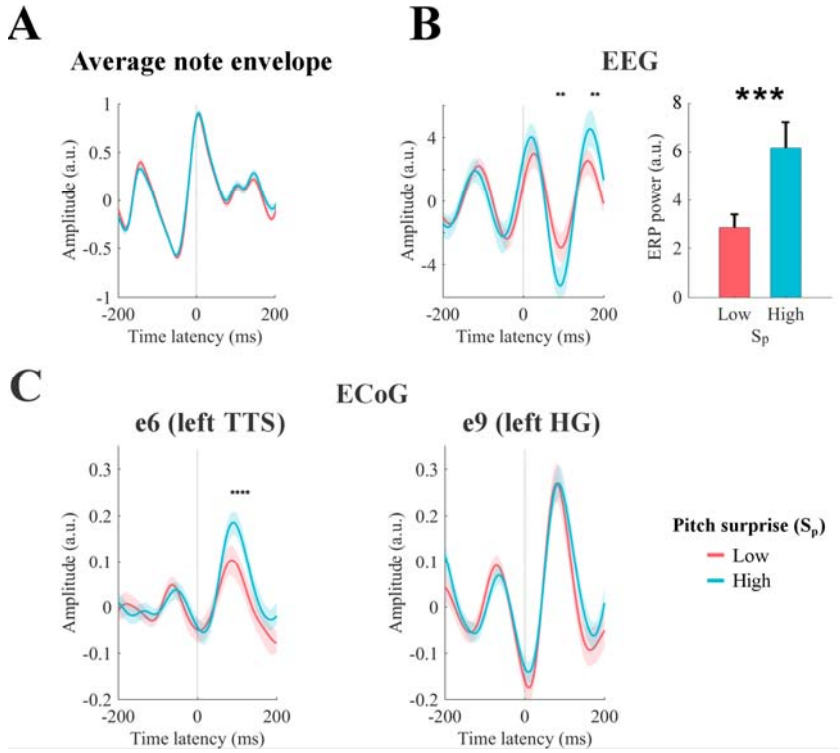

Figure 5. Event-related potentials (ERP) analysis. (*A*) Notes with equal peak envelope were selected (median envelope amplitude across all notes with a tolerance of ±5%). Together, the selected elements were

25% of all notes. Notes were grouped according to the corresponding pitch surprise values (S_p). The figure shows the average sound envelope for the 20% notes with lowest and highest surprise values. Shaded areas indicate the 95% confidence interval. (*B*) Low-rate EEG signals time-locked to note-onset were selected for high and low S_p values. ERPs for channel Cz are shown on the left. Shaded areas indicate the 95% confidence interval (across subjects). Stars indicate significant differences between ERPs for high and low surprise for a given note onset-EEG latency (permutation test, $p < 0.05$, FDR-corrected). The right panel shows the total ERP power for the latencies from 0 to 200 ms (* $p < 0.001$, permutation test). Error-bars indicate the SEM across subjects. (*C*) ERPs for high-γ ECoG data from left TTS and HG (Patient 1). Stars indicate significant differences between ERPs for high and low surprise for a given note onset-EEG latency (permutation test, $p < 0.05$, FDR-corrected). Shaded areas indicate the 95% confidence interval (across individual trials, i.e. responses to single notes).

Pitch and onset-time induce distinct musical expectations

So far, we have parameterized melodic expectations in terms of surprise and entropy features, each for pitch and note-onsets. Surprise and entropy were expected to interact as they convey complementary information. Entropy provides information on the uncertainty of the prediction of the next note before observing the event, thus it describes the overall probability distribution. Surprise depends on that same distribution but is specific to the observed event. For this reason, we expected the responses to entropy and surprise to be dissociable in their temporal dynamics. This hypothesis was tested in our EEG data by measuring the contrast in the TRF_{AM} weights for surprise *versus* entropy (**Fig. 6*A***, top; weights were averaged as follows: $(S_p+S_o)/2$ vs. $(H_p+H_o)/2$). The results showed that responses with latencies up to 350 ms were significantly dominated by both surprise and entropy in alternation ($p < 0.05$, permutation test, FDR-corrected).

A second analysis was conducted to test the relative contribution of pitch and onset-time expectations to the TRF_{AM} model. As previous studies suggested a dissociation between pitch and sound onset processing (Schönwiesner and Zatorre 2008; Coffey et al. 2017), we expected a similar dissociation in the processing of their expectations in early auditory cortical regions. We tested for such a dissociation in our EEG data by measuring the contrast in the TRF_{AM} weights for pitch *versus* onset time (**Fig. 6*A***, bottom; $(S_p+H_p)/2$ vs. $(S_o+H_o)/2$). Note-onset dominant responses emerged only up to 200ms, while pitch dominant responses persisted for much longer latencies up to 400 ms. The latency differences for pitch and note-onset TRFs suggests a certain level of dissociation between pitch and onset-time expectations. Specifically, we hypothesised that the magnitude (and SNR) of the EEG responses to music would increase with the overall note-onset surprise values, thus leading to larger EEG prediction correlations. To test this hypothesis, we sorted the ten musical pieces according to their average S_o (**Fig. 6*B***) and measured whether such an index correlated with the envelope tracking (a by-product of larger EEG responses). The results confirmed the presence of such a correlation by measuring the Spearmen coefficient between the average S_o of a piece and the neural signal prediction correlations for EEG ($r = 0.98$, $p < 0.001$ for non-musicians; $r = 0.96$, $p < 0.001$ for musicians; **Fig. 6*C***) and high-γ ECOG data ($r = 0.88$, $p = 0.002$ for e6 in the left TTS of Patient 1; $r = 0.88$, $p = 0.002$ for e9 in the left HG of Patient 1; **Fig. 6*D***). These effects were specific to onset-time surprise. In fact, Spearman correlations of comparable magnitude emerged with H_o, while no significant correlations were measured for S_p and H_p for these pieces. **Fig. 6*C* and *D*** also illustrates the prediction correlations for AM_p, showing that small (nearly zero) envelope tracking due to small average S_o does not hamper the encoding of pitch expectations on the same ECoG electrode (see **Fig. 6*D* left**), thus further highlighting the dissociation of processes underlying expectation of pitch and onset-time.

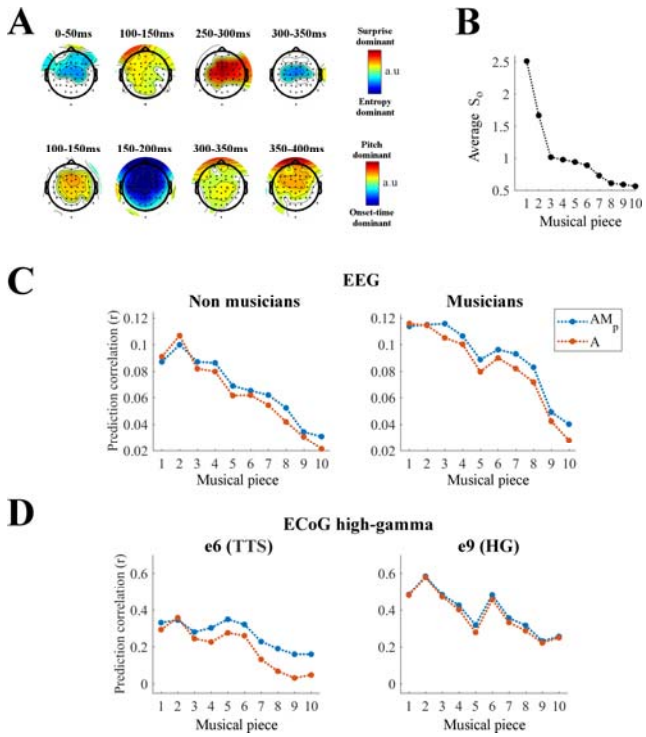

Figure 6. Cortical tracking of music changes with overall surprise of note onset-time within a musical piece. (*A*) Contrasts at each EEG channel of the TRF weights for surprise vs. entropy (top) and pitch vs. onset-time (bottom) in TRF_{AM}. Colors indicate significant differences ($p < 0.05$, permutation test, FDR-corrected) (*B*) Average surprise of note-onsets (S_o) for each musical piece. Lower average S_o indicates musical pieces with more predictable tempo. (*C*) Single-trial EEG prediction result (average across all channels) for musicians ($N_m = 10$) and non-musicians ($N_n = 10$). Trials were sorted as in panel B. (*D*) Single-trial ECoG prediction correlations for the surgery Patient 1 for two electrodes of interest.

Effect of musical expertise on the encoding of melodic expectations

We were also able to shed light on the effect of musical expertise on the encoding of melodic expectations. Specifically, by design, half of the EEG participants had no musical training, while the others were expert pianists that studied for at least ten years (**Fig. 2**). In **Fig. 7*A*** we show a comparison between the two EEG groups. A cluster statistics indicated that melodic expectation was larger for musicians than non-musicians for frontal EEG channels (**Fig. 7*B***; see Di Liberto et al. *in press*, for comparisons that are specific to music envelope tracking).

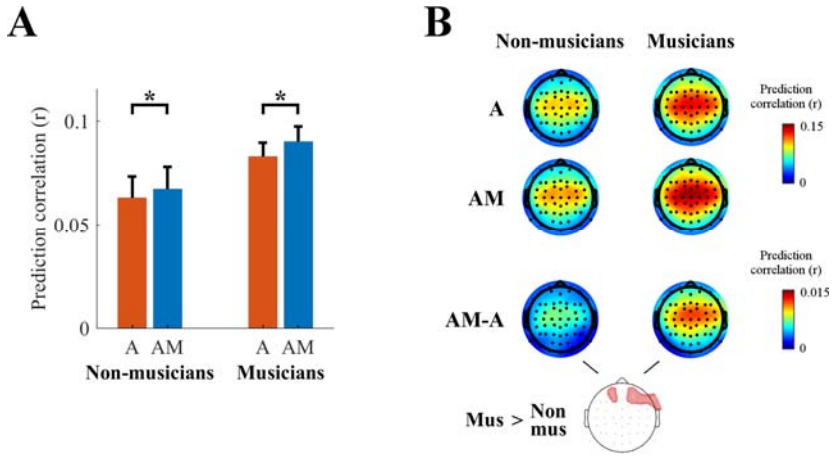

Figure 7. Effect of musical expertise on the low-rate encoding of melodic expectations. (*A*) EEG prediction correlations (average across all scalp channels) for musicians and non-musicians (*$p < 10^{-5}$). The error bars indicate the SEM across participants. (*B*) EEG prediction correlations for individual scalp electrodes (red color indicates stronger prediction correlations) for **A**, **AM**, and **AM-A**. A cluster statistics found a stronger effect of melodic expectations for musicians.

Discussion

Musical perception is strongly influenced by expectations (Bar et al. 2006; Huron 2006; Kok et al. 2012; Pearce 2018; Henin et al. 2019). Violation of these expectations elicits distinct neural signatures that may underlie the emotional and intellectual engagement with music (Zatorre and Salimpoor 2013). Here, we exploited such neural signatures of melodic expectations during listening to Bach piano pieces to demonstrate that cortical responses encode explicitly subtle changes in note predictability during naturalistic music listening. In doing so, we 1) demonstrated a novel methodology to assess the cortical encoding of melodic expectations that is effective at the individual subject level with non-invasive EEG signals recorded during passive music listening, 2) provided detailed insights into the spatial selectivity and temporal properties of that encoding in ECoG recordings, and 3) physiologically evaluated the statistical learning framework of melodic expectation, exemplified by models such as IDyOM (Pearce 2005).

Our findings have several implications for current theories and understanding of sensory perception in general, and musical cognition in particular. First, we show that the neural responses are consistent with the statistical learning theoretical frameworks in that response dynamics are proportional to the predictability of several signal attributes, including pitch and onset-timing. For musical sequences, additional attributes, such as timbre, loudness, and complex patterns of harmony are likely to be relevant, although we did not investigate them here. These properties may all contribute in tandem to music perception, or sometimes in a dissociated manner, as we demonstrated for the relative contribution of pitch and onset-timing (see also Coffey et al. 2017). Second, melodic expectations attributes are encoded in both low-rate and high-γ responses of higher cortical regions (e.g., bilateral PT in humans). The effect of expectations extended to the junction between the PT with HG (TTS) and to neighbouring electrodes in HG, an early auditory cortical areas that favor the accurate representation of the acoustic properties of the music. In addition, neural signals in some SMG electrodes showed similar encoding of melodic entropy and surprise, a result that well aligns with previous findings suggesting a role of the left SMG in short-term pitch memory (Vines et al. 2006; Schaal et al. 2015, Schaal et al. 2017). Third, we have demonstrated a methodology to objectively assess the cortical encoding of melodic expectations using non-invasive neural recordings and passive listening of ecologically valid stimuli. This opens new avenues for investigation of the neurophysiological bases of music perception and its link with emotions (Salimpoor et al. 2009, 2013).

This work builds upon previous ERP findings on simple note sequences (Omigie et al. 2013, 2019), but circumvents one of the main issues tied to the ERP approach, namely that it does not allow in general to isolate responses to continuous or rapidly presented stimuli. While such limitation may be overcome for stimuli with particular statistics, such as speech (Khalighinejad et al. 2017), the temporal regularities inherent in music hamper the ability to dissociate the late ERP components in response to a note from the early responses to subsequent ones. Instead, the TRF framework is more resistant to this issue as it assumes that the transformation of a stimulus into the corresponding brain responses can be captured by a linear time-invariant system. Although the human brain is neither linear nor time-invariant, these assumptions can be reasonable in certain

cases, and this approach was shown to be effective in previous studies with stimuli that were either discrete and rapidly presented, or continuous (Lalor et al. 2006; Ding and Simon 2012b; Di Liberto et al. 2015; Crosse et al. 2016; Fiedler et al. 2017; Brodbeck, Hong, et al. 2018b; Broderick et al. 2018; Vanthornhout et al. 2018; Wong et al. 2018). Using this approach, we have demonstrated how information on melodic expectations, as well as levels and fidelity of envelope tracking, can be extracted from both EEG and ECoG recordings. There is ample evidence that melodic expectations of various origins modulate responses to stimulus acoustics, possibly increasing the magnitude of those responses according to the degree of violations of expectations. This raises the possibility that melodic expectations are encoded also on cortical sites with low or no acoustic sensitivity, where the TRF methodology was not effective in capturing the cortical response patterns. Further studies with different methodologies are required to test this possibility.

Furthermore, since subjects' expectations modulate musical responses differently depending on their cultural experience and musical exposure, it is expected that musical expertise may significantly enhance these modulations and hence reveal stronger encoding of melodic expectations. The evidence we present in this study is in line with that hypothesis (**Fig. 7**). While this finding is preliminary, the results in **Figure 7** leave open a key question: do musicians in general encode melodic expectations better (more strongly or accurately), or is the estimate of the IDyOM model more in tune with that of musicians' predictions than non-musicians'? Furthermore, our results suggest the possibility of a right hemispheric bias in the processing of melodic expectations, and the separate analysis of low-rate and high-γ neural signals seems crucial to investigate such an effect (**Fig. 7*B*** and **Fig. S1**). Other possibilities need to be controlled for in future studies with larger sample sizes and more types of stimuli, e.g., whether the experiment may have been more engaging for musically trained participants, which would explain their stronger envelope tracking! It will also be important to assess the effect of musical expertise on the optimal amount of memory for estimating melodic expectations.

Finally, the effects of melodic expectations exhibited different temporal and spatial dynamics from those that were measured in a similar manner for phonotactic- and semantic-level predictions in the case of natural speech perception (Brodbeck, Hong, et al. 2018a; Broderick et al. 2018; Di Liberto et al. 2019). One noticeable difference is that the strongest TRF components for phonotactic- and semantic-TRFs (Broderick et al. 2018 and Di Liberto et al. 2019 respectively), both in centro-parietal scalp areas, were absent in the responses to melodic surprise and entropy. Such a discrepancy may reflect a fundamental difference between music and speech perception. One tentative explanation is in the different use of predictions in the two domains. In speech, expectations are important to successfully understand the meaning of a sentence (e.g., phonemic restoration, priming; Leonard et al. 2016; Norris et al. 2016), especially in noisy, multi-talker environments (McGettigan et al. 2012; Strauß et al. 2013). In music, expectations may have a stronger link to emotions and musical engagement (Dunsby 2014; Salimpoor et al. 2015). Another element that could explain the different expectation signals measured in speech and music is that statistical learning is not the only contributor to expectations (Morgan et al. 2019). In fact, although statistical learning (IDyOM) was shown to have a prominent role on melodic expectations based on behavioral data, an independent contribution of a rule-like music-theoretically motivated approach was found (Temperley Probabilistic Model of Melody Perception; Temperley 2008). It is possible that the melodic expectation signals presented here and the speech prediction responses shown in the past originate from distinct sources. These conjectures can be readily tested by carrying out analogous experiments with richer prediction models and comparing directly music with speech, with the present findings providing a key starting point to tackle these questions.

This study presented novel detailed insights on the impact of melodic expectations on the neural processes underlying music perception and informed us on the physiological validity of models of melodic expectations based on Markov chains. In the process, we introduced the first solution to investigate the neural underpinnings of music perception in ecologically-valid listening conditions. As a result, this work constitutes a platform where research in cognitive neuroscience and musicology meet and can inform each other.

Methods

EEG data acquisition and preprocessing

Twenty healthy subjects (10 male, aged between 23 and 42, M = 29) participated in the EEG experiment. Ten of them were highly trained musicians with a degree in music and at least ten years of experience, while the other participants had no musical background. Each subject reported no history of hearing impairment or neurological disorder, provided written informed consent, and was paid for their participation. The study was undertaken in accordance with the Declaration of Helsinki and was approved by the CERES committee of Paris Descartes University (CERES 2013-11). The experiment was carried out in a single session for each participant. EEG data were recorded from 64 electrode positions, digitized at 512 Hz using a BioSemi Active Two system. Audio stimuli were presented at a sampling rate of 44,100 Hz using Sennheiser HD650 headphones and Presentation software (*http://www.neurobs.com*). Testing was carried out at Ecole Normale Supérieure, in a dark room, and subjects were instructed to maintain visual fixation on a crosshair centered on the screen, and to minimize motor activities while music was presented.

Neural data were analysed offline using MATLAB software (The Mathworks Inc). EEG signals were digitally filtered between 1 and 8 Hz using a Butterworth zero-phase filter (low- and high-pass filters both with order 2 and implemented with the function *filtfilt*), and down-sampled to 64 Hz. Results were also reproduced with high-pass filters down to 0.1 Hz and low-pass filters up to 30 Hz. EEG channels with a variance exceeding three times that of the surrounding ones were replaced by an estimate calculated using spherical spline interpolation. All channels were then re-referenced to the average of the two mastoid channels with the goal of maximising the EEG responses to the auditory stimuli (Luck 2005).

Electrodes responsive to sound inputs were identified by comparing portion of EEG responses to music with signals recorded during the pre- and post-stimulus silence. 25 chunks of data, each with duration 200 ms, were selected for each of the two conditions and Cohen's *d* effect-size was calculated to quantify the effect of a monophonic music stimulus on the EEG data. Nineteen of the twenty subjects showed responses with $d > 0.5$.

ECoG data acquisition and preprocessing

We recorded cortical activity from three adult human patients (1 male) implanted with stereotactic EEG electrodes as part of their clinical evaluation for epilepsy surgery. The research protocol was approved by the Institutional Review Board of Columbia University and written informed consent of the patients was obtained before surgery. The first patient (P1) was a highly trained musician with about twenty years of experience; the second patient (P2) had no musical background; and the third patient (P3) studied clarinet for 8 years while in secondary school and had not played for 25 years. As a part of their clinical diagnosis of epileptic focus, P1 was implanted with a total of 241 electrodes in the left hemisphere, P2 with 200 electrodes in the right hemisphere, and P3 with 285 electrodes in both left and right hemispheres. Patients had self-reported normal hearing. Electrocorticography signals with sampling rate of 3000Hz were recorded with a multichannel amplifier connected to a digital signal processor (Tucker-Davis Technologies). All data were montaged again to common average reference (Crone et al. 2001).

Channel positions were mapped to brain anatomy using registration of the postimplantation computed tomography (CT) to the preimplantation MRI via the postoperative MRI (Groppe et al.

2017). The CT was first coregistered with the postimplantation structural MRI and, subsequently, with the preimplantation MRI. The coregistration was performed by means of the automated procedure FSL's FLIRT (Mehta and Klein 2010). Channels were assigned to anatomical areas according to the Destrieux atlas (Destrieux et al. 2010) and confirmed by expert inspection blinded to the results of this study.

Neural responses were transformed using Hilbert transform to extract the high-γ band (70-150Hz) for analysis (Edwards et al. 2009). This signal is known to correlate with neural spiking activity (Ray et al. 2008; Steinschneider et al. 2008) and was shown to reliably reflect auditory responses (Kubanek et al. 2013; Mesgarani et al. 2014). Secondly, low-rate responses were extracted from the raw unfiltered data by digitally filtering between 1 and 8 Hz using a Butterworth zero-phase filter (low- and high-pass filters both with order 2 and implemented with the function *filtfilt*). Both high-γ and low-rate signals were then down-sampled to 100 Hz.

Electrodes responsive to sound inputs were identified by comparing portion of ECoG responses to music with signals recorded during the pre-stimulus silence. 25 chunks of data, each with duration 200 ms, were selected for each of the two conditions and Cohen's *d* effect-size was calculated to quantify the effect of a monophonic music stimulus on the ECoG data. Electrodes with $d > 0.5$ (medium effect-size) were marked as responsive to the sound stimulus (21, 25, and 34 electrodes for patients 1, 2, and 3 respectively).

Stimuli and procedure

Monophonic MIDI versions of ten musical pieces from Bach's monodic instrumental corpus were partitioned into short snippets of approximately 150 seconds. The selected melodies were originally extracted from violin and flute scores and were synthesised by using piano sounds with MuseScore 2 software (MuseScore BVBA). This was done in order to reduce familiarity for the expert pianist participants while enhancing their neural response by using their preferred instrument timbre (Pantev et al. 2001). Each 150s piece, corresponding to an EEG/ECoG trial, was presented three times throughout the experiment, adding up to 30 trials that were presented in a random order. At the end of each trial, participants were asked to report on their familiarity to the piece (from 1: unknown; to 7: know the piece very well). This rating could take into account both their familiarity with the piece at its first occurrence in the experiment, as well as the build-up of familiarity across repetitions. Behavioural results confirmed that participants reported repeated pieces as more familiar (paired t-test on the average familiarity ratings for all participants across repetitions: $rep_2 > rep_1$, $p = 7.2\text{x}10^{-5}$; $rep_3 > rep_2$, $p = 0.015$, Bonferroni correction; Di Liberto et al. *in press*). No significant difference emerged between musicians and non-musicians on this account (two-sample *t*-test, $p = 0.4$, 0.5, 0.6 for repetitions 1, 2, and 3 respectively). EEG participants undertook the entire experiment (30 trials: ten stimuli repeated three times), ECoG patients were presented with 10 trials (ten stimuli, played with random order).

IDyOM

The Information Dynamics of Music model (IDyOM; Pearce 2005) is a framework based on variable-order hidden Markov models. Given a note sequence of a melody, the probability distribution over every possible note continuation is estimated for every *n*-gram context up to a given length *k* (model order). The distributions for the various orders were combined according to an entropy-based weighting function (IDyOM; Pearce 2005, Section 6.2). Here, we used an unbounded implementation of IDyOM that builds *n*-grams using contexts up to the size of each musical piece. In addition, predictions were the result of a combination of long- and short-term models (LTM and STM respectively), which yields better estimates than either LTM or STM alone. The LTM was the result of a pre-training on a large corpus of Western music that did not include the stimuli presented during the EEG experiment, thus simulating the statistical knowledge of a listener that was implicitly acquired after a life-time of exposure to music. The STM, on the other hand, is constructed online for each individual musical piece that was used in the EEG experiment.

Our choice of IDyOM was motivated by the empirical support that Markov model-based frameworks received as a model of human melodic expectation (Pearce and Wiggins 2006; Pearce, Müllensiefen, et al. 2010). Specifically, among other evidence, previous work has indicated that it predicts human ratings of uncertainty during music listening (Hansen and Pearce 2014; Moldwin et al. 2017).

Music features

In the present study, we have assessed the coupling between the EEG data and various properties of the musical stimuli. Of course, this required the extraction of such properties from the stimulus data in the first place. First, we defined a set of descriptors summarizing *low-level acoustic properties* of the music stimuli (**A**). Since the specific set of stimuli was monophonic, broadband envelope and fundamental frequency (f_0) of each note fully characterize the sound acoustics. However, only the envelope descriptor was used in the present study as the frequency information did not explain additional EEG variance. The broadband amplitude envelope was extracted from the acoustic waveform using the Hilbert transform (**Fig. 1*A***). In addition, **A** included the half-way rectified first-derivative of the envelope, which was shown to contribute to the stimulus-EEG mapping when using linear system identification methods (Daube et al. 2019).

In order to investigate the cortical processing of melodic expectations, we estimated *melodic surprise* and *entropy* for each individual note of a given musical piece by using IDyOM. Given a note e_i, a note sequence $e_{1..n}$ that immediately precedes that note, and an alphabet E describing the possible pitch or note-onset values for the note, *melodic surprise* $S(e_i|e_{1..i-1})$ refers to the inverse probability of occurrence of a particular note at a given position in the melody. In other words, this surprise indicates the degree to which a note appearing in a given context in a melody is unexpected, or information content (MacKay and Kay 2003; Pearce, Müllensiefen, et al. 2010):

$$S(e_i|e_{1..i-1}) = log_2 \frac{1}{p(e_i|e_{1..i-1})}.$$

The second feature that was extrapolated from IDyOM is the entropy in a given melodic context. This measure was defined as the Shannon entropy (Shannon 1948) computed by averaging the surprise over all possible continuations of the note sequence, as described by E:

$$H(e_{1..i-1}) = \sum_{e \in E} p(e|e_{1..i-1}) S(e|e_{1..i-1}).$$

Inverse probability and entropy are complementary in that the first indicates the level of expectedness of a note, while the second clarifies whether an unexpected note occurred in a context that was more or less uncertain, thus corresponding to a weaker or stronger note sequence violation respectively.

IDyOM simulates implicit melodic learning by estimating the probability distribution of each upcoming note. This model can operate on multiple viewpoints, meaning that it can capture the distributions of various properties of music. Here, we focused on two such properties that are considered the most relevant to describe a melody: the *pitch* and the *onset time* of a note. IDyOM generates predictions of upcoming musical events based on what is learned, allowing the estimation of surprise and entropy values for the properties of interest. This provided us with four features describing the prediction of an upcoming note: surprise of pitch ($\mathbf{S_p}$), entropy of pitch ($\mathbf{H_p}$), surprise of onset time ($\mathbf{S_o}$), and entropy of onset time ($\mathbf{H_o}$). Each of these features was encoded into time-series by using their values to modulate the amplitude of a note-onset vector i.e., vectors of zeros marking with value one all note onsets, with length matching that of the corresponding musical piece and with the same sampling frequency as the EEG (or ECoG) data. The matrix composed of the four resulting vectors is referred to as *melodic expectations* feature-set (**M**).

In order to assess and quantify the contribution of melodic expectations to the music-EEG mapping, the main analyses were conducted on **A** and the concatenation of **A** and **M** (**AM**). The rationale is that the inclusion of M will improve the fitting score if the EEG responses to music are modulated by melodic expectations i.e., if **M** describes dynamics of the EEG signal that are not redundant with A.

Control analysis

The concatenation of acoustic and melodic expectation features **AM** constitutes a richer representation of a musical piece than **A** or **M** alone, and we hypothesized that it would be a better descriptor of the neural responses to music. However, it is also true that **AM** has more dimensions than **A**, which could be a confounding factor when comparing their coupling with the neural signal. In order to factor out dimensionality from this comparison, we have built stimulus descriptors with the same dimensionality as **AM** that carry the same acoustic information but less meaningful melodic expectation values, the hypothesis being that such descriptors would be less coupled with the neural signal. Such vectors were obtained by degrading the STM model by imposing memory restrictions on the local musical piece, while leaving untouched the LTM model, which represents the participants' prior knowledge on Western music. The memory restrictions on the STM model were introduced by subdividing each musical piece in chunks of exponentially longer lengths (1, 2, 4, 8, 16, and 32 musical bars) and then calculating the melodic expectations in each chunk separately. Similar results were obtained by reducing the model order *k*. However, the model order restricts the memory in terms of number of notes, while the first approach works in the musical bar dimension, which we considered more relevant and comparable across musical pieces.

The same analysis was conducted by using a stimulus descriptor consisting of the concatenation of **A** with the **M** descriptor after randomly shuffling the surprise and entropy values in time (but by preserving the note onset times; $\mathbf{AM_{shu}}$), providing us with a feature-set with the same dimensionality and surprise- and entropy-values distributions of **AM** that contains **A** but not **M** information and that was outperformed by **AM** ($r_{AM} > r_A$: permutation test, $p < 10^{-6}$, $d = 1.31$).

Computational model and data analysis

A system identification technique was used to compute the channel-specific music-EEG mapping. This method, here referred to as the temporal response function (TRF; Lalor et al. 2009; Ding et al. 2014), uses a regularized linear regression (Crosse et al. 2016) to estimate a filter that optimally describes how the brain transforms a set of stimulus features into the corresponding neural response (forward model; **Fig. 1*B***). Leave-one-out cross-validation (across trials) was used to assess how well the TRF models could predict unseen data while controlling for overfitting. The quality of a prediction was quantified by calculating Pearson's correlation between the preprocessed recorded signals and the corresponding predictions at each scalp electrode.

The interaction between stimulus and recorded brain responses is not instantaneous, in fact a sound stimulus at time t_0 affects the brain signals for a certain time-window $[t_1, t_1+t_{win}]$, with $t_1 \geq 0$ and $t_{win} > 0$. The TRF takes this into account by including multiple time-lags between stimulus and neural signal, providing us with model weights that can be interpreted in both space (scalp topographies) and time (music-EEG latencies). First, a time-lag window of -150–750 ms was used to fit the TRF models. The temporal dynamics of the music responses were inferred from the TRF model weights, as shown in **Figures 2*E***, **3*B***, and **4*B***. We then performed the EEG prediction analysis by restricting the TRF model fit to the window [0,350] ms, thus reducing the dimensionality of the data and the risk for overfitting. This time-lag window was identified by means of a backward elimination procedure that quantified the relevance of the various stimulus-EEG latencies to the TRF mapping.

Backward elimination is a method to perform feature selection on multivariate data (Guyon and Elisseeff 2003) (only the first iteration of this approach was run for computational reasons). In our context, the relevance of a feature (where feature includes both stimulus properties and time-lags)

is quantified as the loss in EEG prediction correlation due to its exclusion from the TRF model. Specifically, TRF were fit for the time-lag window -150 and 750 ms after excluding a 50-ms window of time-lags [t_i,t_i+50] ms. Then, the loss was calculated as $r_{LOSS} = r_{[-150,750]} - r_{[-150,750] \setminus [ti,ti+50]}$. Ultimately, this allowed us also to isolate the temporal dynamics of the effect of melodic surprise AM-A. Note that this procedure is similar to a single-lag analysis (O'Sullivan et al. 2014; Das et al. 2016), with the difference that latencies capturing information that is redundant with other lags will not produce a large r_{LOSS}, which is a useful property when the goal is to minimise the time-latency window.

Statistical analysis

Statistical analyses were performed using two-tailed permutation tests for pair-wise comparisons. Correction for multiple comparisons was applied where necessary via the false discovery rate (FDR) approach. One-way ANOVA was used to assess when testing the significance of an effect over multiple (> 2) groups (e.g., memory size in **Fig. 2*C***). The values reported use the convention $F(df, df_{error})$. Greenhouse-Geisser corrections was applied when the assumption of sphericity was not met (as indicated by a significant Mauchly's test). Cohen's *d* was used as a measure of effect size.
Topographical dissimilarity scores were calculated according to Murray et al. (2008):

$$DISS = \sqrt{2 * (1 - r)},$$

where *r* is the correlation between two topographical distributions of interest. Significance was assessed by means of a one-sided *p*-values based on a randomization test with 100 permutations.

Author Contributions

The study was conceived by G.D.L., C.P., and S.S.; the experiments were designed by G.D.L., C.P., R.B., S.S. and A.d.C.; G.D.L. programmed the tasks; G.D.L. and C.P. collected the EEG data; P.P., A.D.M., J.L.H., and N.M. collected the ECoG data; G.D.L., R.B., C.P, and P.P. analyzed the data; G.D.L. and S.S. wrote the first draft of the manuscript; N.M., A.d.C., C.P., R.B., P.P., A.D.M., J.L.H. edited the manuscript.

References

Attaheri A, Kikuchi Y, Milne AE, Wilson B, Alter K, Petkov CI. 2015. EEG potentials associated with artificial grammar learning in the primate brain. Brain Lang. 148:74–80.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E. 2006. Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A. 103:449–454.

Besson M, Macar F. 1987. An event-related potential analysis of incongruity in music and other non-linguistic contexts. Psychophysiology. 24:14–25.

Bianco R, Novembre G, Keller PE, Kim S-G, Scharf F, Friederici AD, Villringer A, Sammler D. 2016. Neural networks for harmonic structure in music perception and action. Neuroimage. 142:454–464.

Bretan M, Oore S, Eck D, Heck L. 2017. Learning and Evaluating Musical Features with Deep Autoencoders.

Brodbeck C, Hong LE, Simon JZ. 2018b. Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. Curr Biol. 28:3976–3983.e5.

Brodbeck C, Hong LE, Simon JZ. 2018a. Transformation from auditory to linguistic representations across auditory cortex is rapid and attention dependent for continuous speech. bioRxiv. 326785.

Brodbeck C, Presacco A, Simon JZ. 2018. Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. Neuroimage. 172:162–174.

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC. 2018. Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Curr Biol.

Carrus E, Pearce MT, Bhattacharya J. 2013. Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. Cortex. 49:2186–2200.

Clark A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 36:181–204.

Coffey EBJ, Musacchia G, Zatorre RJ. 2017. Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. J Neurosci. 37:830–838.

Crone NE, Boatman D, Gordon B, Hao L. 2001. Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. Clin Neurophysiol. 112:565–582.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC. 2016. The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. Front Hum Neurosci. 10.

Das N, Biesmans W, Bertrand A, Francart T. 2016. The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. J Neural Eng. 13:056014.

Daube C, Ince RAA, Gross J. 2019. Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. Curr Biol. 29:1924–1937.e9.

Destrieux C, Fischl B, Dale A, Halgren E. 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage. 53:1–15.

Di Liberto GM, O'Sullivan JA, Lalor EC. 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr Biol. 25.

Di Liberto GM, Pelofi C, Shamma S, de Cheveigné A. n.d. Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening. Acoust Sci Technol.

Di Liberto GM, Wong D, Melnik GA, de Cheveigne A. 2019. Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. Neuroimage. 196:237–247.

Ding N, Chatterjee M, Simon JZ. 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage. 88:41–46.

Ding N, Simon JZ. 2012a. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol. 107:78–89.

Ding N, Simon JZ. 2012b. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A. 109:11854–11859.

Dunsby J. 2014. On Repeat: How Music Plays the Mind. By Elizabeth Hellmuth Margulis. Music Lett. 95:497–499.

Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT. 2009. Comparison of Time–Frequency Responses and the Event-Related Potential to Auditory Speech Stimuli in Human Cortex. J Neurophysiol. 102:377–386.

Erickson LC, Thiessen ED. 2015. Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. Dev Rev. 37:66–108.

Fiedler L, Wöstmann M, Graversen C, Brandmeyer A, Lunner T, Obleser J. 2017. Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. J Neural Eng. 14:036020.

Finn AS, Lee T, Kraus A, Hudson Kam CL. 2014. When it hurts (and helps) to try: the role of effort in language learning. PLoS One. 9:e101806–e101806.

Fishman YI. 2014. The mechanisms and meaning of the mismatch negativity. Brain Topogr. 27:500–526.

Friston K, Kiebel S. 2009. Predictive coding under the free-energy principle. Philos Trans R Soc L B Biol Sci. 364:1211–1221.

Garrido MI, Kilner JM, Stephan KE, Friston KJ. 2009. The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol. 120:453–463.

Groppe DM, Bickel S, Dykstra AR, Wang X, Mégevand P, Mercier MR, Lado FA, Mehta AD, Honey CJ. 2017. iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. J Neurosci Methods. 281:40–48.

Guyon I, Elisseeff A. 2003. An Introduction to Variable and Feature Selection André Elisseeff, Journal of Machine Learning Research.

Hannon EE, Soley G, Ullal S. 2012. Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. J Exp Psychol Hum Percept Perform. 38:543–548.

Hansen NC, Pearce MT. 2014. Predictive uncertainty in auditory sequence processing. Front Psychol. 5.

Henin S, Turk-Browne N, Friedman D, Liu A, Dugan P, Flinker A, Doyle W, Devinsky O, Melloni L. 2019. Statistical learning shapes neural sequence representations. bioRxiv. 583856.

Huron DB. 2006. Sweet anticipation□: music and the psychology of expectation. MIT Press.

Khalighinejad B, Cruzatto da Silva G, Mesgarani N. 2017. Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech. J Neurosci.

Kok P, Jehee JFM, de Lange FP. 2012. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron. 75:265–270.

Kubanek J, Brunner P, Gunduz A, Poeppel D, Schalk G. 2013. The Tracking of Speech Envelope in the Human Cortex. PLoS One. 8:e53398.

Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ. 2006. The VESPA: a method for the rapid estimation of a visual evoked potential. Neuroimage. 32:1549–1561.

Lalor EC, Power AJ, Reilly RB, Foxe JJ. 2009. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. J Neurophysiol. 102:349–359.

Lecaignard F, Bertrand O, Gimenez G, Mattout J, Caclin A. 2015. Implicit learning of predictable sound sequences modulates human brain responses at different levels of the auditory hierarchy. Front Hum Neurosci. 9:505.

Leonard MK, Baud MO, Sjerps MJ, Chang EF. 2016. Perceptual restoration of masked speech in human cortex. Nat Commun. 7:13619.

Luck SJ. 2005. An introduction to the event-related potential technique.

MacKay D, Kay D Mac. 2003. Information theory, inference and learning algorithms.

McGettigan C, Faulkner A, Altarelli I, Obleser J, Baverstock H, Scott SK. 2012. Speech comprehension aided by multiple modalities: behavioural and neural interactions. Neuropsychologia. 50:762–776.

Mehta AD, Klein G. 2010. Clinical utility of functional magnetic resonance imaging for brain mapping in epilepsy surgery. Epilepsy Res. 89:126–132.

Mesgarani N, Cheung C, Johnson K, Chang EF. 2014. Phonetic Feature Encoding in Human Superior Temporal Gyrus. Science (80-). 343:1006–1010.

Miller KJ, Leuthardt EC, Schalk G, Rao RPN, Anderson NR, Moran DW, Miller JW, Ojemann JG. 2007. Spectral Changes in Cortical Surface Potentials during Motor Movement. J Neurosci. 27:2424–2432.

Miranda RA, Ullman MT. 2007. Double dissociation between rules and memory in music: an event-related potential study. Neuroimage. 38:331–345.

Moldwin T, Schwartz O, Sussman ES. 2017. Statistical Learning of Melodic Patterns Influences the Brain's Response to Wrong Notes. J Cogn Neurosci. 29:2114–2122.

Morgan E, Fogel A, Nair A, Patel AD. 2019. Statistical learning and Gestalt-like principles predict melodic expectations. Cognition. 189:23–34.

Morrison SJ, Demorest SM, Stambaugh LA. 2008. Enculturation Effects in Music Cognition. J Res Music Educ. 56:118–129.

Murray MM, Brunet D, Michel CM. 2008. Topographic ERP Analyses: A Step-by-Step Tutorial Review. Brain Topogr. 20:249–264.

Norris D, McQueen JM, Cutler A. 2016. Prediction, Bayesian inference and feedback in speech recognition. Lang Cogn Neurosci. 31:4–18.

Nourski K V, Steinschneider M, Rhone AE, Kawasaki H, Howard MA, Banks MI. 2018. Processing of auditory novelty across the cortical hierarchy: An intracranial electrophysiology study. Neuroimage. 183:412–424.

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC. 2014. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb Cortex. bht355.

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. J Cogn Neurosci. 31:855–873.

Omigie D, Pearce MT, Williamson VJ, Stewart L. 2013. Electrophysiological correlates of melodic processing in congenital amusia. Neuropsychologia. 51:1749–1762.

Paller KA, McCarthy G, Wood CC. 1992. Event-related potentials elicited by deviant endings to melodies. Psychophysiology. 29:202–206.

Pantev C, Roberts LE, Schulz M, Engelien A, Ross B. 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport. 12:169–174.

Pearce MT. 2005. THE CONSTRUCTION AND EVALUATION OF STATISTICAL MODELS OF MELODIC STRUCTURE IN MUSIC PERCEPTION AND COMPOSITION.

Pearce MT. 2018. Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. Ann N Y Acad Sci. 1423:378–395.

Pearce MT, Müllensiefen D, Wiggins GA. 2010. The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. Perception. 39:1367–1391.

Pearce MT, Ruiz MH, Kapasi S, Wiggins GA, Bhattacharya J. 2010. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. Neuroimage. 50:302–313.

Pearce MT, Wiggins GA. 2006. Expectation in Melody: The Influence of Context and Learning. Music Percept. 23:377–405.

Pearce MT, Wiggins GA. 2012. Auditory Expectation: The Information Dynamics of Music Perception and Cognition. Top Cogn Sci. 4:625–652.

Qi Z, Beach SD, Finn AS, Minas J, Goetz C, Chan B, Gabrieli JDE. 2017. Native-language N400 and P600 predict dissociable language-learning abilities in adults. Neuropsychologia. 98:177–191.

Quiroga-Martinez DR, Hansen NC, Højlund A, Pearce MT, Brattico E, Vuust P. 2019. Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. Cortex. 120:181–200.

Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS. 2008. Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. J Neurosci. 28:11526–11536.

Romberg AR, Saffran JR. 2010. Statistical learning and language acquisition. Wiley Interdiscip Rev Cogn Sci. 1:906–914.

Saffran JR, Newport EL, Aslin RN, Tunick RA, Barrueco S. 1997. Incidental Language Learning: Listening (and Learning) Out of the Corner of Your Ear. Psychol Sci. 8:101–105.

Salimpoor VN, Benovoy M, Longo G, Cooperstock JR, Zatorre RJ. 2009. The Rewarding Aspects of Music Listening Are Related to Degree of Emotional Arousal. PLoS One. 4:e7487.

Salimpoor VN, van den Bosch I, Kovacevic N, McIntosh AR, Dagher A, Zatorre RJ. 2013. Interactions between the nucleus accumbens and auditory cortices predict music reward value. Science. 340:216–219.

Salimpoor VN, Zald DH, Zatorre RJ, Dagher A, Mcintosh AR. 2015. Predictions and the brain: how musical sounds become rewarding. Trends Cogn Sci. 19:86–91.

Sammler D, Koelsch S, Ball T, Brandt A, Grigutsch M, Huppertz H-J, Knösche TR, Wellmer J, Widman G, Elger CE, Friederici AD, Schulze-Bonhage A, Bianco R, Novembre G, Keller PE, Kim S-G, Scharf F, Friederici AD, Villringer A, Sammler D. 2013. Co-localizing linguistic and musical syntax with intracranial EEG. Neuroimage. 64:134–146.

Schaal NK, Pollok B, Banissy MJ. 2017. Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. Sci Rep. 7:42456.

Schaal NK, Williamson VJ, Kelly M, Muggleton NG, Pollok B, Krause V, Banissy MJ. 2015. A causal involvement of the left supramarginal gyrus during the retention of musical pitches. Cortex. 64:310–317.

Schönwiesner M, Zatorre RJ. 2008. Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. Exp Brain Res. 187:97–105.

Shannon CE. 1948. A Mathematical Theory of Communication. Bell Syst Tech J. 27:379–423.

Skerritt-Davis B, Elhilali M. 2018. Detecting change in stochastic sound sequences. PLOS Comput Biol. 14:e1006162.

Somers B, Verschueren E, Francart T. 2018. Neural tracking of the speech envelope in cochlear implant users. J Neural Eng. 16:16003.

Southwell R, Chait M. 2018. Enhanced deviant responses in patterned relative to random sound sequences. Cortex. 109:92–103.

Steinschneider M, Fishman YI, Arezzo JC. 2008. Spectrotemporal Analysis of Evoked and Induced Electroencephalographic Responses in Primary Auditory Cortex (A1) of the Awake Monkey. Cereb Cortex. 18:610–625.

Storkel HL, Rogers MA. 2000. The effect of probabilistic phonotactics on lexical acquistion [WWW Document]. Clin Linguist Phon. URL https://wordlearning.ku.edu/storkel-hl-rogers-ma-2000

Strauß A, Kotz SA, Obleser J. 2013. Narrowed expectancies under degraded speech: revisiting the N400. J Cogn Neurosci. 25:1383–1395.

Temperley D. 2008. A Probabilistic Model of Melody Perception. Cogn Sci A Multidiscip J. 32:418–444.

Tillmann B, Bharucha JJ, Bigand E. 2000. Implicit learning of tonality: A self-organizing approach. Psychol Rev. 107:885–913.

Toro JM, Sinnett S, Soto-Faraco S. 2005. Speech segmentation by statistical learning depends on attention. Cognition. 97:B25–B34.

Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T. 2018. Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. J Assoc Res Otolaryngol. 19:181–191.

Verschueren E, Somers B, Francart T. 2019. Neural envelope tracking as a measure of speech understanding in cochlear implant users. Hear Res. 373:23–31.

Vines BW, Schnider NM, Schlaug G. 2006. Testing for causality with transcranial direct current stimulation: pitch memory and the left supramarginal gyrus. Neuroreport. 17:1047–1050.

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. 50:1432–1443.

Wong DDE, Fuglsang SA, Hjortkjaer J, Ceolini E, Slaney M, De Cheveigne A. 2018. A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding. Front Neurosci. 12:531.

Woolley SMN. 2012. Early experience shapes vocal neural coding and perception in songbirds. Dev Psychobiol. 54:612–631.

Zatorre RJ, Salimpoor VN. 2013. From perception to pleasure: Music and its neural substrates. Proc Natl Acad Sci U S A. 110:10430–10437.

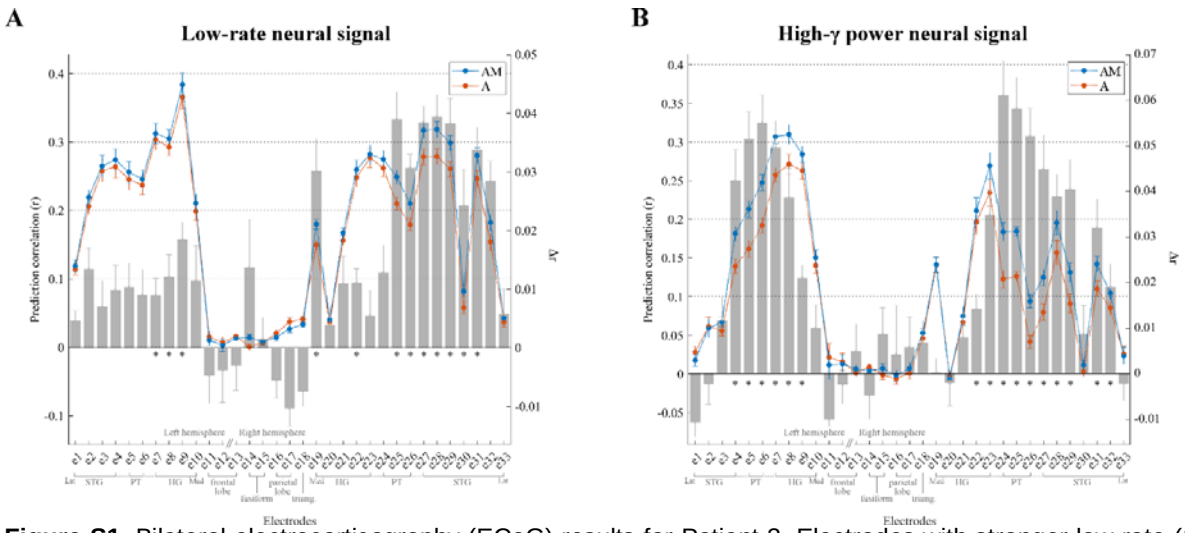

Figure S1. Bilateral electrocorticography (ECoG) results for Patient 3. Electrodes with stronger low-rate (1 - 8 Hz) or high-**γ** (70 – 150 Hz) responses to monophonic music than to silence were selected (Cohen's *d* > 0.5). (*A*) Low-rate ECoG prediction correlations for individual electrodes for **A** and **AM**. Electrodes within each group, as indicated by the gray square brackets, were sorted from lateral to medial cortical sites. The gray bars indicate the predictive enhancement due to melodic expectation (r_{AM}-r_A). Error bars indicate the SEM over trials (*p < 0.01, FDR-corrected permutation test). (*B*) High-**γ** ECoG prediction correlations for individual electrodes for **A** and **AM**.