# RETRIEVING MUSICAL INFORMATION FROM NEURAL DATA: HOW COGNITIVE FEATURES ENRICH ACOUSTIC ONES

**Ellie Bean Abrams**[1,2,3] **Eva Muñoz Vidal**[1,2,3] **Claire Pelofi***[1,2] **Pablo Ripollés***[1,2,3]

[1] Music and Audio Research Laboratory, New York University
[2] Center for Language, Music, and Emotion, New York University
[3] Department of Psychology, New York University
* denotes shared last authorship
{ea84,elm8254,cp2830,pr82}@nyu.edu

## ABSTRACT

Various features – from low-level acoustics, to higher-level statistical regularities, to memory associations – contribute to the experience of musical enjoyment and pleasure. Recent work suggests that *musical surprisal*, that is, the unexpectedness of a musical event given its context, may directly predict listeners' experiences of pleasure and enjoyment during music listening. Understanding how surprisal shapes listeners' preferences for certain musical pieces has implications for music recommender systems, which are typically content- (both acoustic or semantic) or metadata-based. Here we test a recently developed computational algorithm, called the Dynamic-Regularity Extraction (D-REX) model, that uses Bayesian inference to predict the surprisal that humans experience while listening to music. We demonstrate that the brain tracks musical surprisal as modeled by D-REX by conducting a decoding analysis on the neural signal (collected through magnetoencephalography) of participants listening to music. Thus, we demonstrate the validity of a computational model of musical surprisal, which may remarkably inform the next generation of recommender systems. In addition, we present an open-source neural dataset which will be available for future research to foster approaches combining MIR with cognitive neuroscience, an approach we believe will be a key strategy in characterizing people's reactions to music.

## 1. INTRODUCTION

Musical surprisal, or, the relative expectations listeners have of ongoing musical events, is essential in understanding humans' engagement and experience with music [1]. Decades of theoretical and experimental work have defined the study of expectation and surprisal as cognitive processes, often in the context of language processing [2, 3]. Importantly, this line of inquiry has crucial implications for the study of music preference and pleasure [4, 5]. Recent studies have shown that the relationship between information-theoretic measures such as surprisal, entropy, or complexity, and enjoyment and pleasure may be described by an inverted U-shape curve (also referred to as the Wundt effect) where stimulus enjoyment is enhanced with an increase in complexity of the song. But as surprisal increases to higher levels, its effect becomes unpleasant [4, 6]. The perceptual measures (e.g. complexity, familiarity/novelty, surprisal) that have been shown to modulate musical preferences are referred to as *collative* variables [7–9]. Supposedly, when other high-level variables are controlled, collative variables explain a large portion of variance in listeners' musical preference [8, 10] following the aforementioned parabolic function. The "sweet spot" of this inverse U-shaped curve may shift left or right, with respect to surprisal measures, depending on factors such as personality, openness-to-experience, or genre preferences [9, 11]. While most music recommender systems rely on acoustic (extracted from a user's music library) and semantic features (derived from subjective behavioral ratings) to predict listeners' preferences, the use of cognitive measures such as music surprisal can remarkably improve their performance, especially because these are known to accurately predict musical pleasure. The results presented here suggest that cognitive neuroscience methods can be leveraged to elucidate new ways of extracting information from music and better predict user preferences.

Early theoretical work on bottom-up and top-down processes modulating expectations by Leonard Meyer and Eugene Narmour [12–14] brought about efforts to model musical prediction, evolving more concretely into explorations of musical tension [15, 16], entropy [17], and the neural bases of surprisal [18, 19]. Crucially, computational models may be tested against both subjective behavioral and objective neurophysiological measures in order to provide a deeper understanding of whether – and how precisely – information-theoretic measures of music inform the cognitive processes underlying music perception and enjoyment. The Information Dynamics of Music (IDyOM) model [20] generates variable-order Markov

probability distributions for each note in a melodic sequence by extracting statistics from both a music corpora and a short-term musical context, thereby incorporating long-term (top-down) and short-term (bottom-up) musical regularities. The IDyOM model has been shown to predict behavioral and physiological markers of listeners' expectations [21–25] and has recently been related to brain activity, showing that melodic expectation is also directly encoded in the neural signal [26–31]. While IDyOM is a well-validated, efficient, and widely used computational model of musical surprisal, it operates only on symbolic (MIDI) data and requires a training set of stimuli (the long-term component of the model) to generate predictions.

To circumvent these shortcomings and explore the behavioral and neural response to continuous audio signals, we turn to a computational model of surprisal recently developed by Skerritt-Davis and Elhilali at Johns Hopkins, the Dynamic Regularity Extraction (D-REX) model [32, 33]. D-REX uses a Bayesian framework to generate predictions and was originally designed to evaluate prediction errors over time for stochastic sound sequences. This model is relevant to the MIR community, as it can be run on any continuous audio input, which broadens its usability for the analysis of large, diverse collections of music. Our previous behavioral results have validated D-REX as predictive of subjective ratings of surprisal, showing that surprisal as calculated by D-REX predicted subjective behavioral surprisal ratings for 80 music excerpts [34]. Given the important role that prediction plays in musical pleasure, enjoyment, and engagement [4, 6, 35], D-REX is used here to explore whether *the brain* tracks musical surprisal. Concretely, we go beyond subjective behavioral responses and test whether musical surprisal as calculated by D-REX is directly represented in an objective neurophysiological signal. Specifically, we present an experimental work in which we recorded brain activity using magnetoencephalography (MEG) while twenty participants listened to musical excerpts. We then relate the neural signal to the D-REX model output using a decoding algorithm to determine whether surprisal is represented at the brain level.

## 2. DATA COLLECTION AND PROCEDURE

Twenty participants with self-reported normal hearing completed the experiment (11 female, 24.8 $\pm$ 2.9 years of age). Participants were presented with 30 one-minute-long musical excerpts (described in Section 3.1) while their brain activity was recorded using MEG. Participants began each trial by clicking a button to start playing each musical excerpt. At the end of each excerpt, participants moved to the next stage where they provided ratings across five measures using a 4-point scale (1 lowest to 4 highest): pleasure, valence, recognition, familiarity, and surprisal.

MEG measures the magnetic fields generated by the electrical activity of neurons in the brain. Unlike electrical currents captured by EEG devices, magnetic activity can pass through the cortex and skull without distortion, resulting in higher spatial resolution [36]. Continuous MEG data was collected using a 157-channel axial gradiometer system at NYU, at a sampling rate of 1000Hz with an online low-pass filter of 200Hz. Prior to conducting the main analysis, MEG data underwent preprocessing, starting with the noise-reduction of the signal using the continuously adjusted least squares method (CALM) with the MEG160 software [37]. The data was then exported into MNE-Python [38] and bad channels (e.g., channels which saturated during the recording) were removed through visual inspection and interpolated using a weighted sum of signals from neighboring channels. An independent component analysis (ICA) was fitted on the data using FastICA in MNE-Python to isolate independent sources of noise contaminating the channels. Components corresponding to system noise, heartbeat, and eye-blinks were removed from the raw recording after inspection of the topography and time-course of magnetic activity for each component. Finally, epochs were extracted from one second before stimulus onset to stimulus offset, resulting in 61-second-long epochs. For an additional data quality check, we inspected each participant's auditory response to a set of randomized 1000Hz tones and 250Hz tones and observed a higher amplitude response to the higher tone, thus confirming satisfactory data quality.

## 3. STIMULI

The musical stimuli used here have been previously used to validate D-REX as a predictor of behavioral subjective ratings of surprisal using different genres of music (classical and elevator music) [34]. Classical stimuli were taken from a list of musical excerpts rated by 65 participants for pleasantness in a previous study [39]. The other music stimuli were selected from a range of sources, including songs from Muzak Orchestra's Stimulus Progression albums, as well as more contemporary elevator music compositions (see Supplementary Table S1 [1] for a list of stimuli). The most interesting minute (highest accumulated surprisal) of each piece was selected using a shifting window of 60s across D-REX's surprisal output, so that any results showing lower correlations with brain activity would not simply be due to choosing a particularly unsurprising minute of the piece (see [34]). All excerpts were normalized to 70dB using Praat and python's AudioSegment package, and the sound faded 3s in 3s out.

## 4. MODELING MUSICAL SURPRISE

### 4.1 Acoustic Features

D-REX takes as input a set of acoustic features, which can be extracted using any existing MIR techniques. In our case, we extracted features using the NSL Auditory-Cortical Matlab Toolbox developed by the Neural Systems Laboratory at the University of Maryland. The toolbox is an implementation of a cortical model of sound processing, and outputs an estimate of sound as it is represented

[1] Supplementary materials may be downloaded at `https://osf.io/dbm49/`.

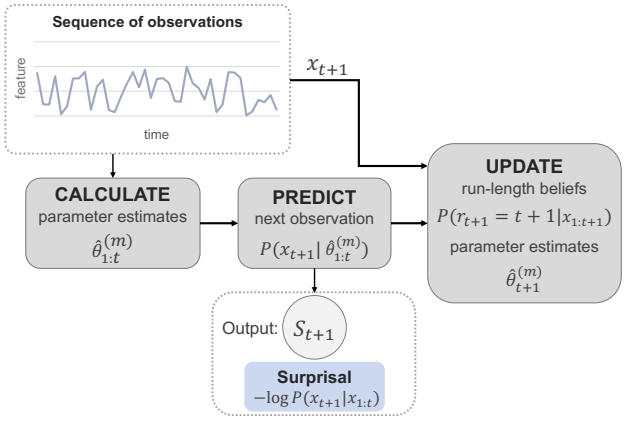


**Figure 1**. Simplified schematic of D-REX model. The model collects parameter estimates from run-lengths 0 until memory constraint $m$ and generates a prediction for the next observation. At $x_{t+1}$, the model updates its run-length beliefs and parameter estimates. The output of the model used in the current project is surprisal $S_{t+1}$.

at various stages of the auditory pathway. We followed the same procedure as detailed in Huang and Elhilali [40], where each acoustic waveform was processed through log-spaced asymmetrical cochlear filters (from 255 Hz to 10.3 KHz). A cortical, or multi-resolution spectrotemporal representation was then generated for a 4-D output (scale-rate-time-frequency). *Rate* and *scale* refer to the two bandwidths (temporal and spectral, respectively) which characterize the filter making up the typical spectrotemporal receptive field of an auditory neuron [41]. *Time* and *frequency* refer to the dimensions of a sound's spectrogram. Fifteen features were extracted from the resulting auditory cortical representation, including average spectral energy, average rate energy, average scale energy, bandwidth, average bark loudness, pitch value, pitch salience (harmonicity), spectral brightness, spectral flatness, spectral irregularity, maximum rate (maximum of temporal variations along each frequency channel), maximum scale (maximum of the spread of spectral energy along the logarithmic frequency axis), centroid rate (centroid of temporal variations along each frequency channel), centroid scale (centroid of the spread of spectral energy along the logarithmic frequency axis), and centroid rate using absolute value of rate (see Supplementary Table S2 and [40] for a complete list of how features were extracted) [34, 40].

### 4.2 Surprisal Output: Dynamic Regularity Extraction (D-REX) Model

D-REX uses Bayesian sequential prediction with perceptual constraints, such as memory ($m$) and observation noise ($n$) to model the brain processes which govern expectation-realization over time in response to sound sequences (as depicted in Figure 1) [32, 33]. Memory $m$ represents the working memory constraints of the listener, determining the maximum previous time points included in context hypotheses for the next time point. Observation noise $n$ consists of adding independent Gaussian noise with zero-mean and a constant variance $n^2$ to the input. The model takes as input a continuous feature over time $x_t$ extracted directly from the waveform as described in Section 4.1. From each feature, the model predicts the distribution for the next time point $x_{t+1}$ given previous ones $x_{1:t}$. Concretely, previous context is estimated after collecting local statistics $\hat{\theta}$, which correspond to the sample mean and sample variance of the input. Predictions for future time-points are based on these statistical representations of previous events:

$$P(x_{t+1}, |x_{1:t}) = P(x_{t+1}|\hat{\theta}_t) \tag{1}$$

The model assumes the parameters $\theta$ can change at any time. A $run$ represents the number of time points between change points of $\theta$. Thus, the model generates multiple hypotheses by gathering statistics across runs and integrates over all possible run lengths ($r_t$) to predict the observation at the next time point:

$$P(x_{t+1}, |x_{1:t}) = \sum_{r_t} P(x_{t+1}|r_t, x_{t-r_t+1:t}) P(r_t|x_{1:t}) \tag{2}$$

The output of interest is surprisal, $S_{t+1}$, which refers to how well the new observation was predicted by the model. Specifically, it is the mismatch between the observation $x_{t+1}$ and its predictive probability in bits:

$$S_{t+1} = -\log P(x_{t+1}|x_{1:t}) \tag{3}$$

As conveyed in Eqn (3), surprisal is inversely related to the event's probability: an event with low probability has high surprisal, an event with high probability has low surprisal, and an event with a probability of 1 has zero surprisal.

The D-REX model in this case takes a matrix of features (the 15 extracted features described in Section 4.1) and calculates surprisal as above *separately* across each feature vector. Then, D-REX combines all the outputs by getting the product of predictive probability (described in Eqn (3)) *across* features for a *summary measure* of joint surprisal. In our analysis, we use Surprisal to refer to *joint* surprisal, that is, the summary measure of surprisal across all 15 feature inputs (as in [34, 40]).

## 5. ANALYSIS

### 5.1 Decoding Method: Temporal Response Function

To decode acoustic (e.g., envelope) and cognitive (e.g., surprisal) features from the neural data, we used a Temporal Response Function (TRF) approach [42]. This decoding algorithm (`https://github.com/mickcrosse/mTRF-Toolbox`) describes the linear mapping of stimuli features onto a set of channels of neural activity. In this context, the input consists of the time-series collected at each MEG electrode $n$ sampled at times $t = 1, ...T$. The assumption is made that the neural response at some time-point $r(t, n)$ can be described as the convolution of a specific stimulus feature $s(t)$ with a channel-specific kernel $w_n$, as described in Eqn (4):

$$r(t,n) = (s * w_n)(t) + \varepsilon(t,n) \quad (4)$$

where $\epsilon(t,n)$ is the residual noise at each channel that is not explained by the model. The kernel $w_n$, thus, is essentially a filter that describes the linear transformation of the stimulus feature into the neural response. Its weights are estimated for a specified range of time lags, $\tau$, relative to the instantaneous occurrence of the stimulus feature $s(t)$, in order to capture the typical input-output activity of interest. In the context of auditory processing, the range of time lags over which to estimate $w(\tau,n)$ should be those used to capture the neural components of an auditory Event-Related Potential (ERP; e.g. -100-600 ms). Therefore, the weight at $\tau$ 100, for instance, describes how one unit of change in amplitude of a given input feature affects the neural response 100 ms later [43].

The weights $w(\tau,n)$ are estimated by minimizing the unexplained residual response $\epsilon(t,n)$, in this case by minimizing the difference (in terms of mean-squared error) between the actual neural response, $r(t,n)$, and the predicted $\hat{r}(t,n)$ response:

$$min\ \varepsilon(t,n) = \sum_t [r(t,n) - \hat{r}(t,n)]^2 \quad (5)$$

Concretely, this is achieved using ridge regression. This decoding method takes into account the high autocorrelation property of continuous stimuli such as music, thus avoiding the temporal smearing that would result from less sophisticated decoding approach, such as cross-correlation. We can interpret the continuous weights $w(n)$ for each $\tau$ as a *modeled-ERP*: a marker of the feature encoding into the neural data.

To summarize, the modeled-ERPs we will discuss describe the linear transformation of the stimulus feature into the neural response, and it can provide insights into the *dynamics* of the neural response to this specific feature [42]. As such, a modeled-ERP's shape may reveal cognitive processes usually observed through ERP analyses, such as the P1 and P2 components [43]. This method presents a useful alternative to averaging across segments of the neural response to derive ERP profiles from continuous signals, which is not precise nor informative enough. Speech envelope [44–46], phonemes [47, 48], semantics [49], and more recently, musical syntax [46] have been successfully decoded from brain data using TRF modeling [46–51].

### 5.2 Decoding Model Input

The decoding analysis was conducted twice using two different inputs: a purely acoustic signal (the first derivative of the amplitude envelope of the audio) and a cognitive signal (the surprisal D-REX output). The amplitude envelope of music carries modulations occurring in the 2-10 Hz range that capture the temporal patterns of critical rhythmic information [52]. Consistent results demonstrate a reliable cortical tracking of the music envelope [46, 53]. Therefore, we first conducted the decoding analysis using the envelope information, as a baseline for cortical tracking of low-level acoustical features [51, 53]. The broadband amplitude envelope was extracted using the Hilbert transform for each individual excerpt. The obtained signal was low-pass filtered at 30 Hz and down-sampled at 100 Hz. Then, the first derivative Envelope' was extracted (see Figure 2A), as it is well-known to enhance stimulus-neural response mapping when using linear system identification methods [54].

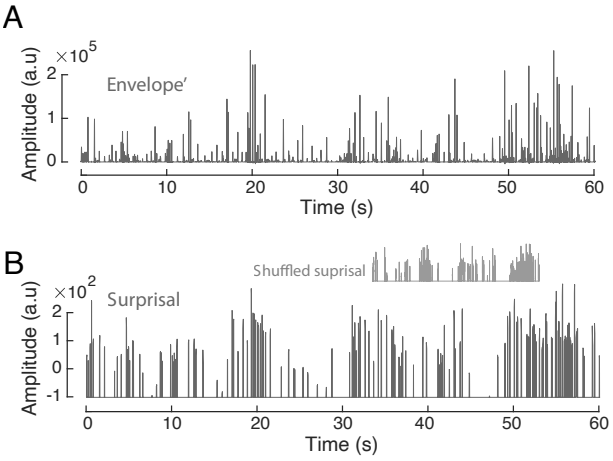


**Figure 2**. Example of decoding model input. **(A)** The first derivative of the envelope for one excerpt. **(B)** Surprisal as calculated using D-REX and indexed at each onset. Shuffled surprisal, shown in grey, is used for the null model.

The decoding analysis was then conducted using the Surprisal signal as input. Because the D-REX output is a smooth, continuous signal that does not constitute a good input for a TRF-based decoding method [42, 54], we used onset information to extract surprisal values associated with note onsets. From the Envelope', we extracted onset indexes by selecting indexes of amplitudes above a threshold determined individually for each excerpt. We then used the resulting onset vector to index Surprisal values (see Figure 2B), which resulted in a discrete, onset-based Surprisal signal.

### 5.3 Decoding Model Output

The modeled-ERP is estimated on a portion of the stimulus and neural data, and tested on an unseen part of the data through a leave-one-out cross-validation method. To evaluate the performance of the model in predicting neural data from stimulus features, a Pearson's correlation is then computed between the actual neural data (averaged across participants, which is typical for this type of experiment [51, 55]) and the data predicted by the convolution $s * w_n$. Thus, $r$-values are obtained for each channel and each musical piece and are then averaged across pieces to obtain one average $r$-value per channel. The distribution of these $r$-values is indicative of how well the particular feature used to obtain the modeled-ERP is encoded in the brain. For a visual representation of the flow of data processing, refer to Figure 3.

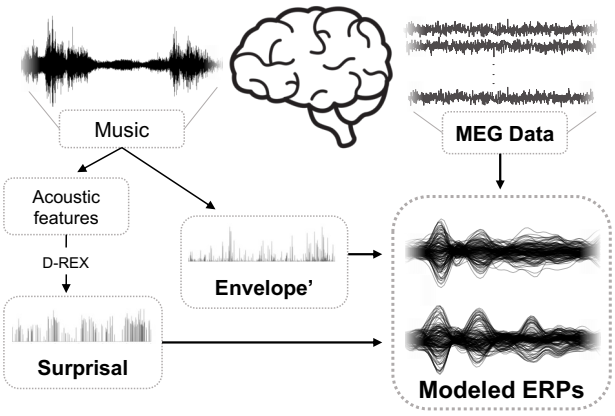


**Figure 3**. Diagram depicting the flow of the music to neural mapping described above.

## 6. RESULTS

The TRF decoding method capture the neural response to a feature (e.g. Envelope' or Surprisal) and informs us as to how well it is represented in the neural data. Here, this is evaluated by estimating a modeled-ERP that describes the mapping, at each channel, between a set of input features and the neural data. A correlation between predicted data obtained from the modeled-ERP and actual data is then computed as a marker of how well a feature is represented in the brain activity. Here we conducted two analyses to characterize the encoding of surprisal as modeled by D-REX in the neural data: we first tested the hypothesis that the encoding of surprisal was significantly different from its baseline, by computing a null model of Surprisal encoding and comparing the obtained $r$-value distribution between the null and the real model. Second, we sought to gain insight from the specific shape of the modeled-ERP by comparing it to the one obtained using the Envelope'.

### 6.1 Cortical Encoding of Surprisal

To properly baseline the encoding of features, the distribution of Pearson's correlation *r*-values obtained by decoding surprisal was tested against a distribution of $r$-values obtained by shuffling each song's surprisal and conducting the same decoding analysis on this shuffled surprisal. A significant enhancement of $r$-values obtained from the real model, as compared to the null one, would confirm that surprisal as predicted by D-REX is significantly encoded in the neural data. In practice, we examined the statistical difference between the Pearson's *r*-values for each MEG channel resulting from a real Surprisal model and its baseline. The real model was constructed using the Surprisal signal (plotted for one song in Figure 2B) as input to the TRF. The baseline model was obtained by shuffling the Surprisal values attributed to each onset over 100 permutations (one permutation is illustrated in Figure 2B).

For each electrode, we obtained an $r$-value by conducting a TRF analysis on the neural data averaged across participants, using either the real Suprisal input or a shuffled version of it (averaged over 100 permutations). We thus obtained one $r$-value for each channel and each musical piece, and the values obtained were consistent with previous reported musical stimuli encoding [46, 51]. We averaged these values across musical pieces and obtained two distributions of $r$-values (e.g. real and null model) for each channel, plotted in Figure 4D. They reveal that $r$-values obtained from the real model were higher, which was confirmed by a paired t-test where $r$-values for each MEG channel ($t(156) = 16.25; p < .001, d_s = 1.29$).

### 6.2 Higher-Level Processing

The second analysis consisted of examining the particular shape of modeled-ERPs obtained from the Envelope' and the Surprisal model, to gain insight into the differences in neural responses elicited by the two features. After averaging participants' neural data, the Envelope' and Surprisal modeled-ERPs were estimated for each individual MEG channel and each musical excerpt, and then averaged across musical excerpts. One modeled-ERP per MEG channel was thus obtained (see Figure 4A and 4B). As expected, both modeled-ERPs exhibit typical auditory responses at [50-100] ms and [150-200] ms [43]. However, the modeled-ERPs obtained from the Surprisal input exhibit an amplitude peak at around 350 ms. To statistically evaluate differences between the responses to the Envelope' and the Surprisal inputs, we first computed the absolute value of the difference Surprisal-Envelope' (z-scored). This difference signal is plotted in Figure 4C. The significant peaks were assessed over the entire signal ([-100-600] ms) using a series of t-tests that were conducted on each individual time sample, using the distribution of values over channels at each time point. The shaded grey areas on the *x*-axis indicate significance corrected for multiple comparisons (FDR correction, $p < .05$). This analysis revealed significant differences for the two main auditory components (at [50-100] ms and [150-200] ms). This may reflect the fact that the Surprisal signal was obtained by modulating values at note onsets, while the Envelope' model used a more continuous signal, yielding a more attenuated early modeled response. The third significant peak that occurs at around 350 ms indicates an enhanced neural response in the modeled-ERP obtained from the Surprisal. This response resembles a P3 response, a well-characterized, consistent, and reliable neurophysiological marker of syntactic processing in both the music [56, 57] and language [58] domains.

## 7. CONCLUSIONS

The central goal of the current project was to determine whether musical surprisal, a high-level, cognitive measure, enhances decoding performance of neural signal above and beyond what may be explained by acoustic features alone, thus showing that the human brain is tracking statistical regularities as modeled by D-REX. By validating D-REX, a computational model which simulates musical surprisal not from symbolic stimuli (e.g., MIDI) but directly from any audio file, we provide a valuable tool that can integrate perceptual and cognitive musical components relevant to

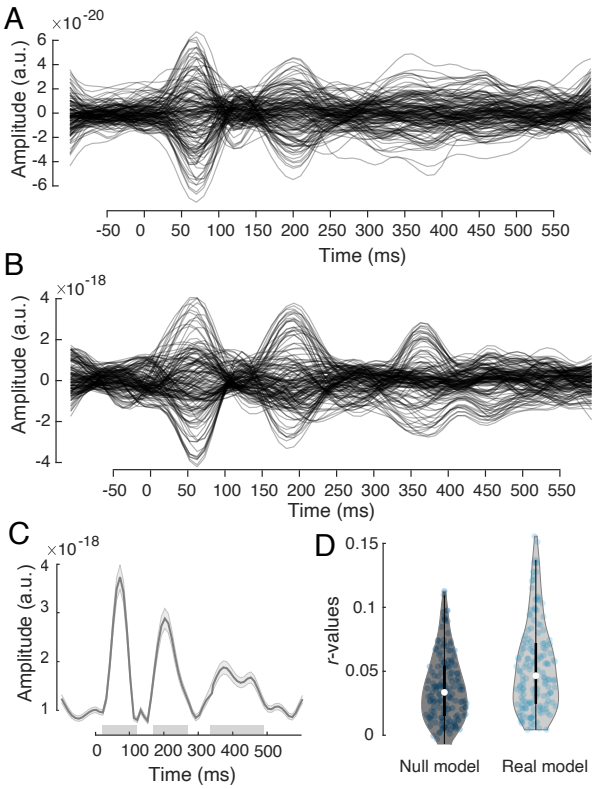


**Figure 4**. Results of decoding analysis. **(A)** The modeled-ERPs for each channel from the decoding model using Envelope' (the acoustic feature) as input. The weights for each channel are averaged across songs. **(B)** The modeled-ERPs for each channel from the decoding model using Surprisal (the cognitive feature) as input. **(C)** The absolute values of the Surprisal - Envelope' (z-scored) signals averaged across channels over time. Shaded areas around the curve indicate SEM over channels. Grey bars on the *x*-axis represent time windows during which a significant difference between the two averaged modeled-ERPs was found ($p < .005$, FDR-corrected). **(D)** Violin plots of *r*-value distributions for the real and baseline (i.e. shuffled) Surprisal models. Each dot corresponds to the *r*-value for one MEG channel.

recommender systems. Additionally, we provide an open-source database of neural data (twenty participants listening to 30 one minute long musical excerpts along with the audio files for these excerpts) with an excellent temporal resolution (at a sampling frequency of 1000 Hz) that can be used to further investigate the neural correlates of music processing. [2] In this paper, we used a decoding method to validate a Bayesian algorithm, D-REX, which models the continuous surprisal experienced by listeners during music listening. By describing a music-to-neural mapping between music stimuli and MEG data, we demonstrate that D-REX is an effective model to explain brain signals above and beyond mere acoustical features, capturing neural responses indicative of higher-level processing.

Our analysis also highlights a specific time window of auditory processing in the musical domain which is crucial to higher-level mechanisms involved in music listening. In the modeled-ERP derived from the Surprisal input (see Figure 4B), there is a significant peak around 350 ms, which undoubtedly evokes the P3 ERP [59, 60]. This ERP is implicated in syntax processing and context updating operations in both language [58] and music processing [56,57]. This further suggests that D-REX is capturing higher-level expectations generated by gathering statistics over time from acoustic information.

In the context of music listening, a recorded neural signal may be conceptualized as "a mid-level representation of the original music piece that has been heavily distorted by two consecutive black-box filters—the brain and the [MEG] equipment" [61]. However, while the brain does extract and represent the acoustic features contained in the musical excerpt, it also combines them to generate purely cognitive components, such as surprisal. In this experiment, we extract musical features and interface them with a cognitive model which abstracts beyond acoustic features into the psychological domain. Importantly, this cognitive model represents a measure which is one of the many factors predicting individual music preference and pleasure [4, 6]. Thus, we show that combining music information retrieval with cognitive neuroscience is an optimal way to measure people's responses to music.

## 8. FUTURE DIRECTIONS

### 8.1 Further Validation of D-REX Model

The current analysis was executed using *joint surprisal*, which is a summary measure across all 15 extracted acoustic features. Though our results show that D-REX joint surprisal is effectively encoded in the neural data recorded during music listening, it is possible that surprisal across specific features, such as pitch value or harmonicity, may be more relevant to cognitive aspects of music listening. Also, these representations may differ across participants. Thus, it is worth exploring across which dimensions surprisal is better represented. In addition, the present study used Western musical genres for listeners to validate D-REX, but future work may expand this exploration to non-Western genres.

### 8.2 Music Recommender Systems

This project presents a new line of inquiry for music recommender systems, which have typically been content-based (using extracted auditory features or subjective semantic ratings) or user-item/metadata-based [62]. Given the aforementioned inverse U-relationship between music surprisal and musical preferences, D-REX may be used to characterize individuals' musical corpora and, compiled with other variables, efficiently predict liking for new music. More broadly, we propose that harnessing cognitive neuroscience methods is potentially fruitful in improving and generating new methods for recommender systems by including cognitively relevant outcomes such as surprisal [4, 6].

[2] https://osf.io/dbm49/

## 9. REFERENCES


[1] D. B. Huron, *Sweet Anticipation: Music and the psychology of expectation*. MIT Press, 2007.

[2] R. Levy, "Expectation-based syntactic comprehension," *Cognition*, vol. 106, no. 3, pp. 1126–1177, Mar. 2008.

[3] R. M. Willems, S. L. Frank, A. D. Nijhof, P. Hagoort, and A. van den Bosch, "Prediction during natural language comprehension," *Cereb. Cortex*, vol. 26, no. 6, pp. 2506–2516, Jun. 2016.

[4] B. P. Gold, M. T. Pearce, E. Mas-Herrero, A. Dagher, and R. J. Zatorre, "Predictability and uncertainty in the pleasure of music: A reward for learning?" *J. Neurosci.*, vol. 39, no. 47, pp. 9397–9409, Nov. 2019.

[5] V. N. Salimpoor, D. H. Zald, R. J. Zatorre, A. Dagher, and A. R. McIntosh, "Predictions and the brain: how musical sounds become rewarding," *Trends Cogn. Sci.*, vol. 19, no. 2, pp. 86–91, Feb. 2015.

[6] V. K. M. Cheung, P. M. C. Harrison, L. Meyer, M. T. Pearce, J.-D. Haynes, and S. Koelsch, "Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity," *Curr. Biol.*, vol. 29, no. 23, pp. 4084–4092.e4, Dec. 2019.

[7] D. E. Berlyne, "Aesthetics and psychobiology," *Journal of Aesthetics and Art Criticism*, vol. 31, no. 4, 1973.

[8] A. Chmiel and E. Schubert, "Unusualness as a predictor of music preference," *Music Sci.*, vol. 23, no. 4, pp. 426–441, Dec. 2019.

[9] Y. Güçlütürk, R. H. A. H. Jacobs, and R. van Lier, "Liking versus complexity: Decomposing the inverted u-curve," *Front. Hum. Neurosci.*, vol. 10, p. 112, Mar. 2016.

[10] A. Chmiel and E. Schubert, "Back to the inverted-u for music preference: A review of the literature," *Psychol. Music*, vol. 45, no. 6, pp. 886–909, Nov. 2017.

[11] P. G. Hunter and E. G. Schellenberg, "Interactive effects of personality and frequency of exposure on liking for music," *Pers. Individ. Dif.*, vol. 50, no. 2, pp. 175–179, Jan. 2011.

[12] L. B. Meyer, *Emotion and meaning in music*. University of Chicago Press, 1961.

[13] E. Narmour, "The "genetic code" of melody: Cognitive structures generated by the implication-realization model," *Contemporary Music Review*, vol. 4, no. 1, pp. 45–63, 1989.

[14] ——, "The top-down and bottom-up systems of musical implication: Building on meyer's theory of emotional syntax," *Music Percept.*, vol. 9, no. 1, pp. 1–26, Oct. 1991.

[15] E. H. Margulis, "A model of melodic expectation," *Music Percept.*, vol. 22, no. 4, pp. 663–714, Apr. 2005.

[16] M. M. Farbood, "A parametric, temporal model of musical tension," *Music Percept.*, vol. 29, no. 4, pp. 387–428, Apr. 2012.

[17] D. Temperley, *Music and probability*. MIT Press, 2007.

[18] S. Koelsch, P. Vuust, and K. Friston, "Predictive processes and the peculiar case of music," *Trends Cogn. Sci.*, vol. 23, no. 1, pp. 63–77, Jan. 2019.

[19] L. Gebauer, M. L. Kringelbach, and P. Vuust, "Ever-changing cycles of musical pleasure: The role of dopamine and anticipation," *Psychomusicology*, vol. 22, no. 2, pp. 152–167, Dec. 2012.

[20] M. T. Pearce and G. A. Wiggins, "Auditory expectation: the information dynamics of music perception and cognition," *Top. Cogn. Sci.*, vol. 4, no. 4, pp. 625–652, Oct. 2012.

[21] N. Politimou, P. Douglass-Kirk, M. Pearce, L. Stewart, and F. Franco, "Melodic expectations in 5-and 6-year-old children," *Journal of Experimental Child Psychology*, vol. 203, p. 105020, 2021.

[22] D. Omigie, M. T. Pearce, and L. Stewart, "Tracking of pitch probabilities in congenital amusia," *Neuropsychologia*, vol. 50, no. 7, pp. 1483–1493, 2012.

[23] K. A. Corrigall, B. Tillmann, and E. G. Schellenberg, "Measuring children's harmonic knowledge with implicit and explicit tests," *Music Perception: An Interdisciplinary Journal*, vol. 39, no. 4, pp. 361–370, 2022.

[24] R. Bianco, B. Gold, A. Johnson, and V. Penhune, "Music predictability and liking enhance pupil dilation and promote motor learning in non-musicians," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[25] R. Bianco, L. E. Ptaszynski, and D. Omigie, "Pupil responses to pitch deviants reflect predictability of melodic sequences," *Brain and Cognition*, vol. 138, p. 103621, 2020.

[26] G. M. Di Liberto, C. Pelofi, R. Bianco, P. Patel, A. D. Mehta, J. L. Herrero, A. de Cheveigné, S. Shamma, and N. Mesgarani, "Cortical encoding of melodic expectations in human temporal cortex," *Elife*, vol. 9, Mar. 2020.

[27] G. Marion, G. M. Di Liberto, and S. A. Shamma, "The music of silence: Part i: Responses to musical imagery encode melodic expectations and acoustics," *Journal of Neuroscience*, vol. 41, no. 35, pp. 7435–7448, 2021.

[28] D. Omigie, M. Pearce, K. Lehongre, D. Hasboun, V. Navarro, C. Adam, and S. Samson, "Intracranial recordings and computational modeling of music reveal the time course of prediction error signaling in frontal and temporal cortices," *Journal of Cognitive Neuroscience*, vol. 31, no. 6, pp. 855–873, 2019.

[29] D. R. Quiroga-Martinez, N. C. Hansen, A. Højlund, M. Pearce, E. Brattico, and P. Vuust, "Decomposing neural responses to melodic surprise in musicians and non-musicians: evidence for a hierarchy of predictions in the auditory system," *NeuroImage*, vol. 215, p. 116816, 2020.

[30] D. R. Quiroga-Martinez, N. C. Hansen, A. Højlund, M. Pearce, E. Brattico, and P. Vuust, "Musical prediction error responses similarly reduced by predictive uncertainty in musicians and non-musicians," *European Journal of Neuroscience*, vol. 51, no. 11, pp. 2250–2269, 2020.

[31] D. Omigie, M. T. Pearce, V. J. Williamson, and L. Stewart, "Electrophysiological correlates of melodic processing in congenital amusia," *Neuropsychologia*, vol. 51, no. 9, pp. 1749–1762, 2013.

[32] B. Skerritt-Davis and M. Elhilali, "Detecting change in stochastic sound sequences," *PLoS Comput. Biol.*, vol. 14, no. 5, p. e1006162, May 2018.

[33] ——, "A model for statistical regularity extraction from dynamic sounds," *Acta Acust. United Acust.*, vol. 105, no. 1, pp. 1–4, Jan. 2019.

[34] E. B. Abrams, P. Ripolles, and D. Poeppel, "The rewards of muzak: elevator music as a tool for the quantitative characterization of emotion and preference," Oct 2021. [Online]. Available: psyarxiv.com/xqs8b

[35] R. Bianco, B. P. Gold, A. P. Johnson, and V. B. Penhune, "Music predictability and liking enhance pupil dilation and promote motor learning in non-musicians," *Nature News*, Nov 2019. [Online]. Available: https://www.nature.com/articles/s41598-019-53510-w

[36] P. Hansen, M. Kringelbach, and R. Salmelin, *MEG: An Introduction to Methods*. Oxford University Press, Jun. 2010. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780195307238.001.0001

[37] Y. Adachi, M. Shimogawara, M. Higuchi, Y. Haruta, and M. Ochiai, "Reduction of non-periodic environmental magnetic noise in meg measurement by continuously adjusted least squares method," *IEEE Transactions on Applied Superconductivity*, vol. 11, no. 1, pp. 669–672, 2001.

[38] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *Neuroimage*, vol. 86, pp. 446–460, Feb. 2014.

[39] N. Martínez-Molina, E. Mas-Herrero, A. Rodríguez-Fornells, R. J. Zatorre, and J. Marco-Pallarés, "Neural correlates of specific musical anhedonia," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 46, pp. E7337–E7345, Nov. 2016.

[40] N. Huang and M. Elhilali, "Auditory salience using natural soundscapes," *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 2163–2176, Mar. 2017.

[41] E. Hemery and J.-J. Aucouturier, "One hundred ways to process time, frequency, rate and scale in the central auditory system: a pattern-recognition meta-analysis," *Frontiers in Computational Neuroscience*, vol. 9, 2015. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2015.00080

[42] M. J. Crosse, G. M. Di Liberto, A. Bednar, and E. C. Lalor, "The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli," *Frontiers in human neuroscience*, vol. 10, p. 604, 2016.

[43] E. C. Lalor, A. J. Power, R. B. Reilly, and J. J. Foxe, "Resolving precise temporal processing properties of the auditory system using continuous stimuli," *Journal of neurophysiology*, vol. 102, no. 1, pp. 349–359, 2009.

[44] S. Akram, A. Presacco, J. Z. Simon, S. A. Shamma, and B. Babadi, "Robust decoding of selective auditory attention from meg in a competing-speaker environment via state-space modeling," *NeuroImage*, vol. 124, pp. 906–917, 2016.

[45] D. D. Wong, S. Fuglsang, J. Hjortkjær, E. Ceolini, M. Slaney, and A. de Cheveigné, "A comparison of temporal response function estimation methods for auditory attention decoding," *Biorxiv*, pp. 1–22, 2018.

[46] G. M. Di Liberto, C. Pelofi, S. Shamma, and A. de Cheveigné, "Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening," *Acoustical Science and Technology*, vol. 41, no. 1, pp. 361–364, 2020.

[47] G. M. Di Liberto, J. A. O'Sullivan, and E. C. Lalor, "Low-frequency cortical entrainment to speech reflects phoneme-level processing," *Current Biology*, vol. 25, no. 19, pp. 2457–2465, 2015.

[48] G. M. Di Liberto, V. Peter, M. Kalashnikova, U. Goswami, D. Burnham, and E. C. Lalor, "Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia," *NeuroImage*, vol. 175, pp. 70–79, 2018.

[49] M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech," *Current Biology*, vol. 28, no. 5, pp. 803–809, 2018.

[50] N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *Journal of neurophysiology*, vol. 107, no. 1, pp. 78–89, 2012.

[51] G. M. Di Liberto, C. Pelofi, R. Bianco, P. Patel, A. D. Mehta, J. L. Herrero, A. de Cheveigné, S. Shamma, and N. Mesgarani, "Cortical encoding of melodic expectations in human temporal cortex," *eLife*, vol. 9, 2020.

[52] N. Ding, A. D. Patel, L. Chen, H. Butler, C. Luo, and D. Poeppel, "Temporal modulations in speech and music," *Neuroscience & Biobehavioral Reviews*, vol. 81, pp. 181–187, 2017.

[53] K. B. Doelling and D. Poeppel, "Cortical entrainment to music and its modulation by expertise," *Proceedings of the National Academy of Sciences*, vol. 112, no. 45, pp. E6233–E6242, 2015.

[54] C. Daube, R. A. Ince, and J. Gross, "Simple acoustic features can explain phoneme-based predictions of cortical responses to speech," *Current Biology*, vol. 29, no. 12, pp. 1924–1937, 2019.

[55] D. H. Baker, G. Vilidaite, F. A. Lygo, A. K. Smith, T. R. Flack, A. D. Gouws, and T. J. Andrews, "Power contours: Optimising sample size and precision in experimental psychology and human neuroscience." *Psychological methods*, vol. 26, no. 3, p. 295, 2021.

[56] P. Regnault, E. Bigand, and M. Besson, "Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: evidence from auditory event-related brain potentials," *J. Cogn. Neurosci.*, vol. 13, no. 2, pp. 241–255, Feb. 2001.

[57] B. Poulin-Charronnat, E. Bigand, and S. Koelsch, "Processing of musical syntax tonic versus subdominant: an event-related potential study," *J. Cogn. Neurosci.*, vol. 18, no. 9, pp. 1545–1554, Sep. 2006.

[58] D. Friedman, R. Simson, W. Ritter, and I. Rapin, "The late positive component (p300) and information processing in sentences," *Electroencephalogr. Clin. Neurophysiol.*, vol. 38, no. 3, pp. 255–262, Mar. 1975.

[59] E. Donchin and M. G. H. Coles, "Is the p300 component a manifestation of context updating?" *Behavioral and Brain Sciences*, vol. 11, no. 3, p. 357–374, 1988.

[60] J. Polich, "Updating p300: An integrative theory of p3a and p3b," *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245707001897

[61] S. Stober, T. Prätzlich, and M. Müller, "Brain beats: Brain beats: Tempo extraction from eeg data," 01 2016, pp. 276–282.

[62] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, and M. Elahi, "Current challenges and visions in music recommender systems research," *Int. J. Multimed. Inf. Retr.*, vol. 7, no. 2, pp. 95–116, Jun. 2018.