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*[1,2,3], & Pablo Ripollés*[1,2,3]

Music and Audio Research Laboratory, NYU (1); Center for Language, Music, and Emotion, NYU (2); Department of Psychology, NYU (3)

* Denotes shared last authorship

Contact: ea84@nyu.edu

1. Introduction

Surprisal is a cognitive process which involves the relative expectations listeners have of ongoing auditory events, as well as their *predictability*.

- The relationship between **surprisal** and **pleasure** has been described by an inverted U-shape curve (the Wundt effect).[1]
- Thus, perceptual measures may have useful implications for improving **music recommender systems** above and beyond auditory features.

Given the role that **prediction** plays in **musical enjoyment** and **preference,** we explore **whether the brain tracks surprisal** using a decoding method which relates the neural signal during music listening to the output of a cognitive model of surprisal.

Dynamic **R**egularity **EX**traction model (**D-REX**)

- Uses Bayesian sequential prediction to model surprisal.[2]
- Surprisal may be defined as the mismatch between an observation x_{t+1} and its predictive probability given previous observations:

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

2. Methods

Data Collection & Procedure

- 20 adult participants listened to 30 one-minute musical excerpts.
- After each excerpt, participants provided ratings across five measures using a 4-point scale (1 lowest, 4 highest): pleasure, valence, recognition, familiarity, and surprisal.
- Brain activity recorded with 157-channel MEG system.
  - MEG measures magnetic fields generated by electrical activity of neurons in the brain (higher spatial resolution than EEG).

Stimuli

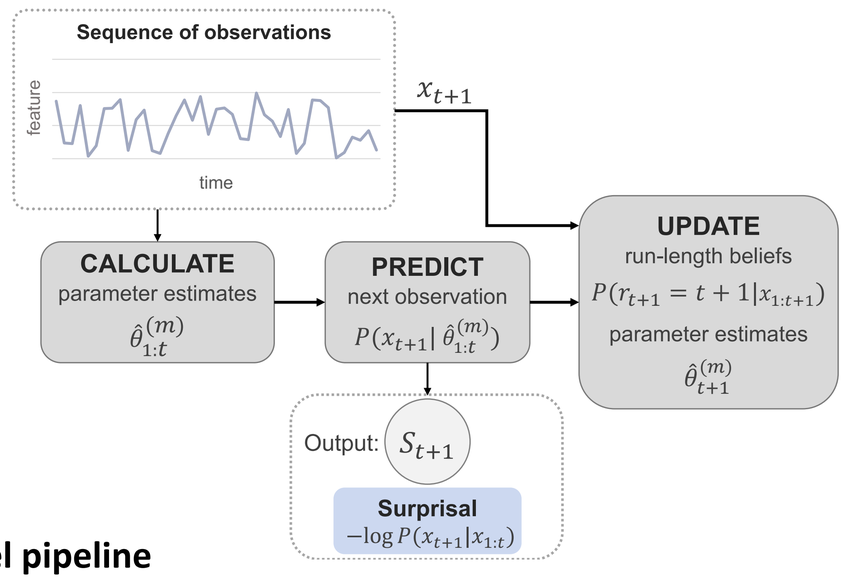
- Stimuli consisted of musical excerpts used in a previous experiment validating D-REX behaviorally.[3]
- 15 classical music excerpts, 15 elevator music excerpts.
- D-REX surprisal extracted from each musical piece.

3. Modeling musical surprise

Acoustic features

- D-REX takes as input a set of continuous acoustic features.
- We used a cortical model of sound processing to generate a multi-resolution spectrotemporal representation of each waveform.
- 15 features, including *spectral energy, bark loudness, spectral brightness, pitch value, pitch salience, spectral centroid,* and others.

Cognitive model: D-REX



Model pipeline

- Surprisal is calculated separately for each feature vector.
- A summary measure of surprisal, called **joint surprisal**, is obtained by calculating the product of predictive probability across all features.

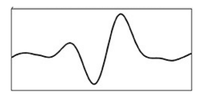
4. Decoding method

Temporal Response Function

- Assumes the neural response is the convolution of a stimulus feature with a **kernel**.

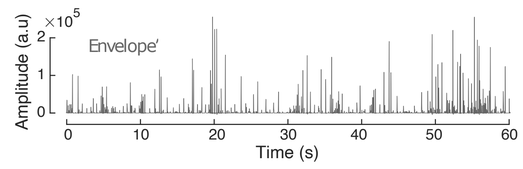
$$r(t,n) = (s * w_n)(t) + \varepsilon(t,n)$$

- For time-point $r(t,n)$, stimulus feature $s(t)$, and kernel w_n.
- Kernel weights $w(\tau, n)$ are estimated for time lags, τ, used to capture an auditory Event-Related Potential (ERP; e.g. -100-600 ms) by minimizing the difference between the predicted vs. actual neural response.
  - Continuous weights $w(n)$ for each τ are a **modeled-ERP.**[4]



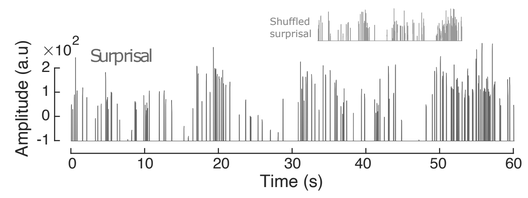
Inputs

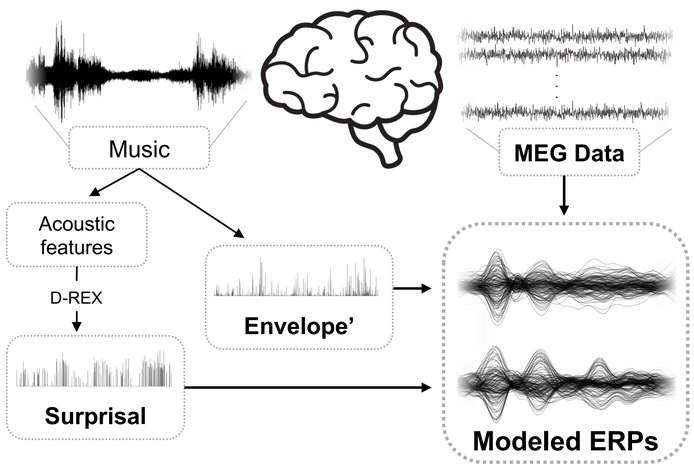
- Two analyses: using **Envelope'** (first derivative of Hilbert transform) vs. **Surprisal** (joint surprisal indexed at Envelope' onsets). Inputs shown on right.



Outputs

- *r*-value for each channel and each musical piece, which are then averaged across pieces to obtain an average *r*-value per channel.
- Modeled-ERPs for each input (Envelope' or Surprisal)





How can we interpret these outputs?

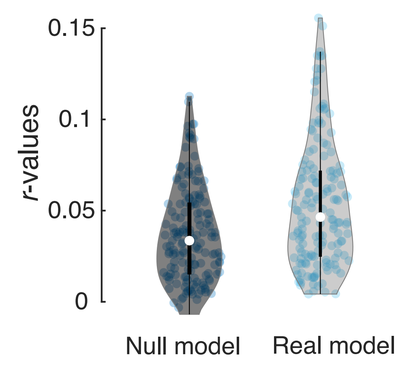
1. Distributions of *r*-values indicate **how well** each input is encoded in the brain.
2. Modeled-ERPs depict the **dynamics** of the predicted **neural response** for each input.

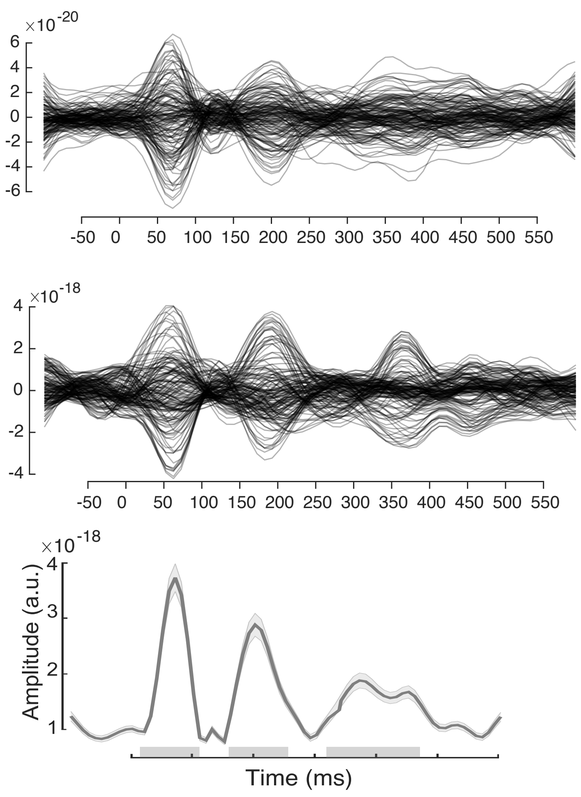
5. Results

Cortical encoding of surprisal

Create a baseline distribution by shuffling each song's surprisal values and running the same decoding analysis (100 permutations, one shown in Analysis figure), then compare true *r*-values from Surprisal.

➡ ***r*-values obtained from the real model were higher than those from the null model** (via permutation test, shown on right; $t(156) = 16.25$; $p < .001$, $d_s = 1.29$).





Higher-level processing

Examine the shape of the modeled-ERPs associated with **Envelope'** (top left) and **Surprisal** (middle left), respectively.

1. Calculate difference signal: z-scored difference between Surprisal and Envelope' (bottom left).
2. Conduct t-tests on the distribution of values across channels at each timepoint of the difference signal.

➡ Both modeled-ERPs exhibit typical auditory responses at 50-100ms and 150-200ms. Surprisal ERP exhibits a third peak around 350ms, which resembles a P3 response reflecting **syntactic processing**.

Modeled-ERP of Surprisal is significantly higher (FDR correction, $p < .05$), showing an enhanced neural response for all three peaks.

6. Conclusions & future directions

- Musical surprisal, a high-level, cognitive measure, enhances decoding of the neural signal **above and beyond what may be explained by acoustic features alone.**
- The brain tracks statistical regularities as modeled by D-REX, a Bayesian computational model of surprisal.
- Open-source database of neural data for twenty participants listening to 30 musical excerpts as well as their acoustic features and surprisal.[5]

Music recommender systems

- D-REX may be used to characterize individuals' music corpora and predict liking.
- The cognitive neuroscience methods here highlight the importance of incorporating cognitively relevant outcomes to content- and metadata-based recommender systems.

7. References & Acknowledgements

[1] 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.

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

[3] 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.

[4] 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.

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

https://www.ripolleslab.com/

Thank you to the Grammy Museum Research Award.