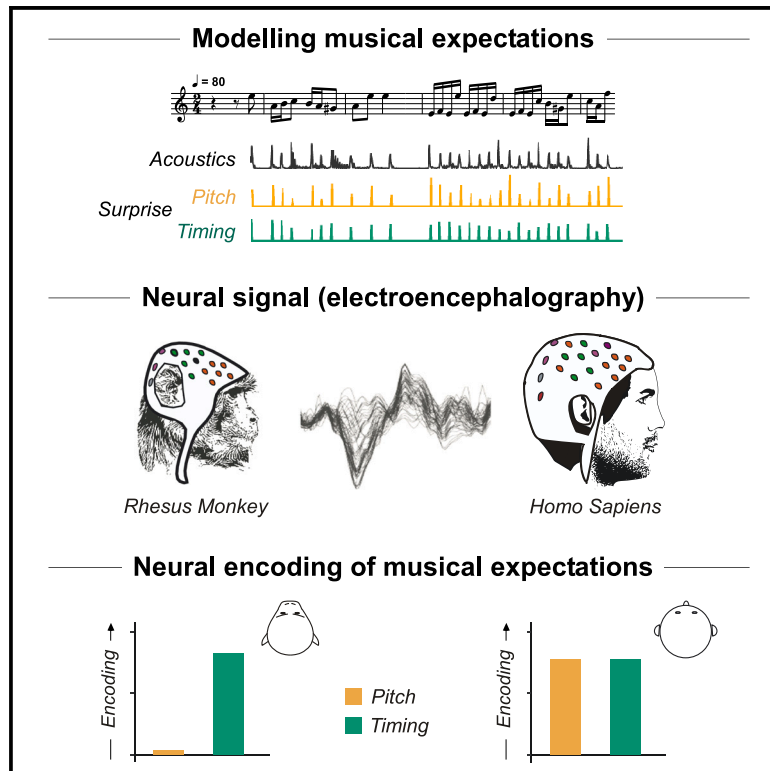# Neural encoding of musical expectations in a non-human primate

Graphical abstract

Authors

Roberta Bianco, Nathaniel J. Zuk,
Félix Bigand, ..., Andrea Ravignani,
Alexandra Battaglia-Mayer,
Giacomo Novembre

Correspondence

roberta.bianco@iit.it (R.B.),
giacomo.novembre@iit.it (G.N.)

In brief

Humans naturally anticipate rhythmic and pitch patterns in music. By modeling musical expectations and neural signals, Bianco et al. compare this ability across humans and rhesus monkeys. They reveal a conserved capacity to anticipate rhythmic but not pitch patterns. This finding sheds light on the phylogeny of music perception.

Highlights

- Musically naive monkeys generate expectations while passively listening to music

- Neural encoding of expectations depends on musical structure beyond acoustics

- Monkeys' expectations are driven by timing rather than pitch structures

- Humans' expectations rely on both timing and pitch

Current Biology

Report

Neural encoding of musical expectations in a non-human primate

Roberta Bianco,[1,7,10,*] Nathaniel J. Zuk,[2] Félix Bigand,[1] Eros Quarta,[3] Stefano Grasso,[3] Flavia Arnese,[1] Andrea Ravignani,[4,5,6] Alexandra Battaglia-Mayer,[3,9] and Giacomo Novembre[1,8,9,*]

[1]Neuroscience of Perception & Action Lab, Italian Institute of Technology, Viale Regina Elena 291, 00161 Rome, Italy
[2]Department of Psychology, Nottingham Trent University, 50 Shakespeare Street, Nottingham NG1 4FQ, UK
[3]Department of Physiology and Pharmacology, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[4]Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands
[5]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music, Universitetsbyen 3, 8000 Aarhus, Denmark
[6]Department of Human Neurosciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[7]X (formerly Twitter): @RobertaBianco0
[8]X (formerly Twitter): @NovembreGiacomo
[9]Senior author
[10]Lead contact
*Correspondence: roberta.bianco@iit.it (R.B.), giacomo.novembre@iit.it (G.N.)
https://doi.org/10.1016/j.cub.2023.12.019

SUMMARY

The appreciation of music is a universal trait of humankind.[1–3] Evidence supporting this notion includes the ubiquity of music across cultures[4–7] and the natural predisposition toward music that humans display early in development.[8–10] Are we musical animals because of species-specific predispositions? This question cannot be answered by relying on cross-cultural or developmental studies alone, as these cannot rule out enculturation.[11] Instead, it calls for cross-species experiments testing whether homologous neural mechanisms underlying music perception are present in non-human primates. We present music to two rhesus monkeys, reared without musical exposure, while recording electroencephalography (EEG) and pupillometry. Monkeys exhibit higher engagement and neural encoding of expectations based on the previously seeded musical context when passively listening to real music as opposed to shuffled controls. We then compare human and monkey neural responses to the same stimuli and find a species-dependent contribution of two fundamental musical features—pitch and timing[12]—in generating expectations: while timing- and pitch-based expectations[13] are similarly weighted in humans, monkeys rely on timing rather than pitch. Together, these results shed light on the phylogeny of music perception. They highlight monkeys' capacity for processing temporal structures beyond plain acoustic processing, and they identify a species-dependent contribution of time- and pitch-related features to the neural encoding of musical expectations.

RESULTS

Music is not just any sound wave. It encompasses sequential dependencies between tones and durations that determine the probability of each sound as a function of past auditory context.[13–15] By tracing these intricate patterns, human listeners recognize melodies, synchronize with rhythms, and generalize prior experience to novel stimuli. Is the ability to track melodic and rhythmic progressions phylogenetically conserved? With few exceptions, research on non-human primates has either focused on isolated rather than holistic music features or on naturalistic music stimuli that are difficult to control experimentally.[16–25] Here, we bridged this gap by adopting the best of both approaches: we used a method that combines naturalistic music with multi-feature stimulus and neural signal modeling. We tested whether, similarly to humans,[26,27] monkeys encode the regular relationships between musical notes to predict when the next event is likely to occur and what it will be.[14] Although

monkeys can extract probabilistic spectral-relations within a small set of events,[28] this ability remains unexplored in the richer context of naturalistic music.

Prediction-related neural signals have been widely measured in humans using non-invasive electroencephalography (EEG). Traditional methods selectively focus on EEG activity evoked by musically unexpected events that violate the rhythmic or melodic patterns of the preceding context, in particular event-related potentials (ERPs).[29–31] Here, we combined this traditional approach with a more recent analytical method[32,33] that no longer examines single events but rather models the continuous EEG response to music by predicting neural activity from both acoustic features[34] and probabilistic rhythmic-melodic descriptors of music.[26,27] We examined the neural data of two monkeys (*Macaca mulatta*) passively listening to real melodies composed by J.S. Bach. In control conditions, monkeys listened to shuffled melodies matched in pitch content and average note timing but disrupted in terms of musical structure (STAR Methods).
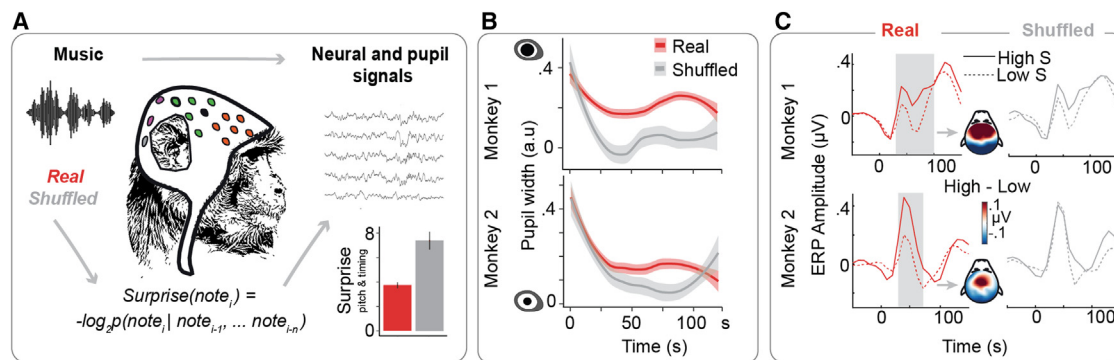
Figure 1. Assessing whether musically naive monkeys generate musical expectations with pupillometry and EEG

(A) EEG and pupillometry were recorded from two rhesus monkeys passively listening to real or shuffled music. We used an unsupervised statistical learning model to calculate values of surprise (S) and entropy (E, not illustrated in the figure) associated with pitch and timing-onset of each note (S_p, S_t, E_p, and E_t). The bar plot shows the mean overall surprise ($S_p + S_t$) associated with real and shuffled music. Error bars represent 1 SEM.

(B) Sustained pupil dilation during exposure to real and shuffled melodies over 2 min (duration of the shortest melody; shadowed areas represent 95% CI).

(C) ERPs evoked by high versus low S notes for real and shuffled music (electrode FCz), aligned to note onset (0 ms). Cluster-based permutation test yielded a significant cluster capturing higher EEG amplitude in response to unexpected notes, specifically associated with real music (cluster spanning from +30 to +90 ms in monkey 1 and from +30 to +60 ms in monkey 2). Cluster topographies showing the difference in amplitude between the two conditions are pointed out by the arrows, and time windows are shaded in gray.

See also Figure S1 and Table S1.

Monkeys had been reared without musical exposure. We simulated the expectations of these "naive listeners" using a variable-order Markov model (information dynamic of music [IDyOM])[13] that can successfully predict the neural encoding of musical expectations in humans.[26,35–37] This model leverages past (long- and short-term) music exposure to compute the Shannon's surprise (S) and entropy (E) of each note, specifically with respect to pitch and onset timing (S_p, S_t, E_p, E_t) (Figure 1A; STAR Methods). Confirming that our control condition reduced the predictability of the original stimuli, the model indicated that shuffled music had a higher overall surprise ($S_p + S_t$) than real music (W = 40, p = 0.005, confidence interval [CI] [1.92, 5.21], median shuffled = 7.53 ± 1.46, real = 3.67 ± 0.77) (Figures 1A and S1A).

We also recorded pupil dilation as a covert indicator of the monkeys' internal state and engagement.[38,39] The pupil diameter decreased throughout listening, less so in response to real than to shuffled music (Figure 1B) (music type, χ^2 (1) = 1.78, p = 0.181; time bins, χ^2 (1) = 159.90, p < 0.001; music type × time bins, χ^2 (1) = 12.64, p < 0.001; the latter interaction was also observed in a control analysis including gaze as a covariate: music type × time bins, χ^2 (1) = 24.77, p < 0.001). This indicates that the monkeys' internal state[38] changed across real and shuffled music and suggests that their nervous system was relatively more engaged by the real music. This interpretation is supported by evidence showing that sustained pupil dilation is induced by regular auditory patterns in monkeys[39] and by attended,[40] familiar,[41] or pleasant[42,43] musical stimuli in humans.

Did real and shuffled music also lead to distinguishable neural activities? We examined ERP modulations in response to a subsection of events (i.e., musical notes) associated with the highest and lowest 20% quantiles of surprise values (high S versus low S; Figure S1B). ERP amplitude was enhanced in response to unexpected (high S) compared with expected (low S) notes when monkeys were listening to real but not shuffled music (Figure 1C),

suggesting a sensitivity to the regularities present in the real music. Specifically, we observed a significant cluster capturing higher EEG amplitude in response to unexpected notes (monkey 1: p < 0.001, electrodes: AF3-4, F1-z-2, FC5-3-1-z-2-4, Cz; monkey 2: p < 0.001, electrodes: F1-z-2, FC3-1-z-2, Cz). The time window of this modulation is reminiscent of ERP components associated with prediction error, previously characterized in monkeys[44] and humans.[45] This effect of surprise observed while listening to real music indicates optimized detection of unexpected events within structured contexts, in line with human[46,47] and rodent studies.[48] Conversely, surprise was not tracked while listening to shuffled music, possibly because these stimuli were characterized by higher entropy that might have flattened the effect of surprise (Figure S1C). Our finding highlights that the structure of musical stimuli allowed the monkeys to generate expectations.

To distinguish EEG modulations driven by musical expectations from those driven by processing acoustic features, we used the multivariate temporal response function (mTRF) approach (Figure 2A). Compared with traditional ERP analysis, mTRF allows us to model how multiple features of a continuous stimulus predict the EEG signal[26,27,33] (Figure 2A; STAR Methods). We first quantified prediction accuracy of a baseline acoustic model (A model) comprising only low-level acoustic features (i.e., note onsets, spectral flux[27,34]). We then tested whether better prediction accuracy could be achieved using a music model (AM model) comprising both low-level acoustic features (A model) as well as four musical features quantifying pitch- and timing-based expectations (S_p, S_t, E_p, E_t,[26] or only S_p, S_t; Figures 2 and S2). A control model (AMc) included the same features of AM but shuffled musical features (S, E). We observed a gain in AM relative to A, and selectively during real as opposed to shuffled music (Figure 2B). Despite having the same dimensionality as AM, no gain was found in AMc. This was confirmed by a linear mixed model yielding main effects of model (AM versus
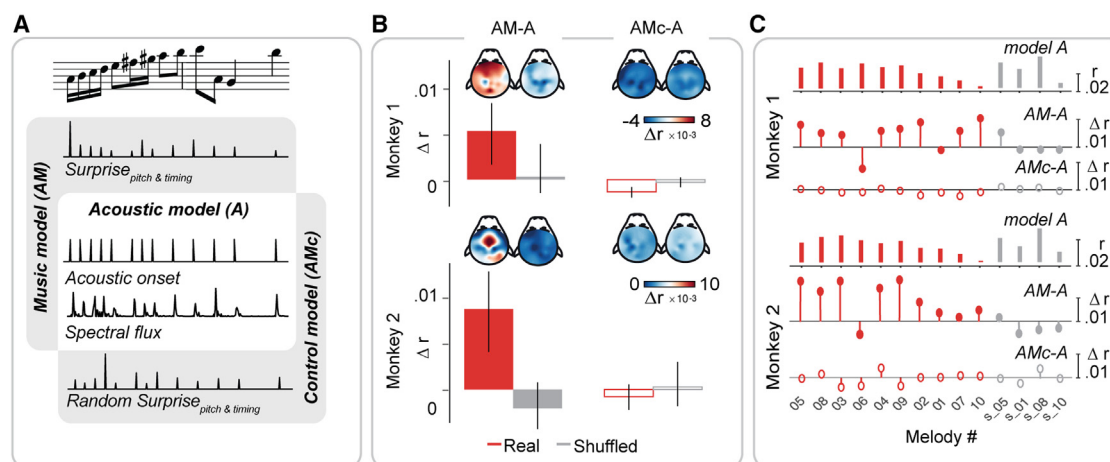
Current Biology
Report

Figure 2. Neural encoding of expectations depends on musical structure beyond acoustics

(A) Schematic of mTRF models regressing different representations of the continuous stimulus on the EEG data: the acoustic model (A, binary note onset and spectral flux as predictors); the music model (AM, A plus music predictors S_p, S_t, E_p, E_t of each note; E vectors are not illustrated in the figure); and the control model (AMc, A plus randomized music predictors with preserved note onset times).

(B) Delta values representing the gain in EEG prediction accuracy (and associated topographies) obtained by subtracting the baseline A model from the AM and AMc models for real and shuffled music, respectively. Error bars represent 1 SEM.

(C) Monkey- and melody-specific EEG prediction accuracy (ranked according to mean surprise; red, real melodies; gray, shuffled melodies). The upper panel shows the prediction accuracy scores yielded by the A model. The middle and bottom panels show the prediction accuracy gain of the AM and AMc models relative to the A model, respectively.

See also Figure S2.

AMc: $\chi^2(1) = 39.78$, $p < 0.001$) and condition (real versus shuffled: $\chi^2(1) = 14.127$, $p < 0.001$), as well as a significant interaction between model and condition ($\chi^2(1) = 14.192$, $p < 0.001$). Follow-up contrasts indicated that the prediction accuracy gain of AM (relative to A) in real music was greater than in shuffled music ($b = 0.008$, $SE = 0.002$, $p = 0.005$). It was also greater than the control AMc (AM versus AMc in real music, $b = 0.008$, $SE = 0.001$, $p < 0.001$; AM versus AMc in shuffled music, $b = 0.007$, $SE = 0.002$, $p = 0.014$). There were no other significant contrasts (all $p > 0.96$). Melodies varied in their mean predictability (Figure S1A), yet these effects were not driven by any specific melody (Figure 2C). This analysis demonstrates neural encoding of subtle variations of surprise while listening to real but not shuffled melodies. Thus, the structure of the musical sequences played a role in the formation of expectations in monkeys over and above mere acoustic tracking.

Both pitch- and timing-related features drive musical expectations in humans.[12,13,26] Which of these features drives musical expectations in monkeys? We recomputed EEG prediction accuracy estimating the selective contribution of pitch and timing. We then related monkeys' neural data (real music only) to that of humans exposed to the same music (data from Di Liberto et al.[26]) and found a species-dependent contribution of pitch and timing to the formation of musical expectations (Figure 3A). For each species, we tested the gain in prediction accuracy (relative to the baseline A model) of reduced versions of AM, including either only pitch- (AMp) or timing-related (AMt) features. In monkeys, the model including only timing (AMt) outperformed the baseline A model ($V = 52$, $p = 0.005$). In contrast, the model including only pitch (AMp) did not yield a significant gain ($V = 38$, $p = 0.161$). A main effect of the model yielded by a linear mixed model (AM versus AMp versus AMt: $\chi^2(2) = 30.8$, $p < 0.001$)

indicated that AMt performed similarly to the full AM model (AM versus AMt: $b = 5.73\mathrm{e}{-5}$, $SE = 0.001$, $p = 0.99$), while AMp prediction accuracy gain dropped compared both to AM (AM versus AMp: $b = 6.47\mathrm{e}{-3}$, $SE = 0.001$, $p < 0.001$) and AMt (AMt versus AMp: $b = 6.41\mathrm{e}{-3}$, $SE = 0.001$, $p < 0.001$). Instead, in humans, both AMt and AMp models outperformed the baseline A model (AMp, $V = 181$, $p = 0.001$; AMt, $V = 169$, $p = 0.007$). A main effect of the model (AM versus AMp versus AMt: $\chi^2(2) = 16.88$, $p < 0.001$) indicated a drop in prediction accuracy gain for both reduced models (AM versus AMp, $b = 1.77\mathrm{e}{-3}$, $SE = 0.5\mathrm{e}{-3}$, $p = 0.002$; AM versus AMt, $b = 1.90\mathrm{e}{-3}$, $SE = 0.5\mathrm{e}{-3}$, $p < 0.001$), which performed similarly to one another (AMt versus AMp: $b = 0.13\mathrm{e}{-3}$, $SE = 0.5\mathrm{e}{-3}$, $p = 0.96$). This analysis demonstrates that, while pitch-related features seem not to be relevant for generating monkeys' expectations, timing-based expectations, exhibited by both humans and monkeys, are phylogenetically conserved (see also supporting ERP evidence; Figure S3A).

Species-specific mnemonic constraints might be captured by the length of the musical context used to optimally predict upcoming events. Given that both species exhibited timing-based expectations, we tested how such expectations change with progressively smaller short-term memory. We achieved this by progressively shortening the length of the Markov chains (i.e., the number of preceding notes or n-gram) used to simulate timing-based expectations in both species. The optimal EEG prediction (the maximum AMt gain among models relying on contexts of different lengths) was associated with an n-gram of 4 in monkeys and of 16 in humans (Figure 3B). This suggests that monkeys might have relied on a shorter musical context compared with humans, presumably due to short-term memory limitations.[49,50] However, with such a small number of data points, the reliability of this effect cannot be confirmed.
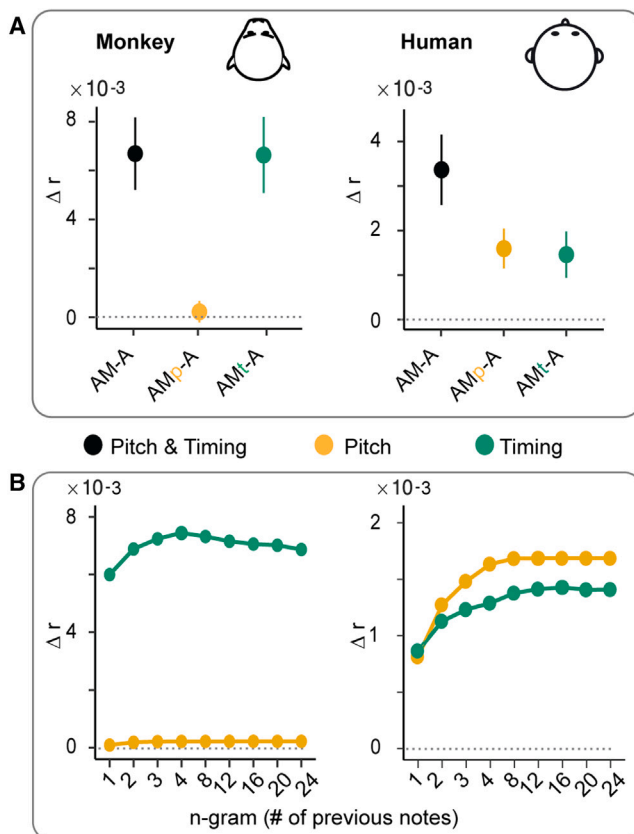
Figure 3. Contribution of pitch, timing, and mnemonic constraints to the generation of musical expectations (monkeys and humans)

(A) Gain in EEG prediction accuracy obtained by computing the delta between the baseline A model and reduced versions of AM, including either only pitch-(AM*p*) or only timing-related (AM*t*) features. Dots represent mean gain across melodies (real music only) and subjects. Cross-species differences in the absolute size of gain might be attributed to differences in EEG signal-to-noise ratio (note that monkeys listened to the melodies across more repetitions [n = 22] compared with humans [n = 3] to compensate for the different group numerosity [monkeys = 2, humans = 20]). Error bars represent 1 SEM.

(B) Mean gain as a function of the number of preceding consecutive notes (n-gram) used to estimate note surprise (for each n-gram, surprise is estimated by combining its sub n-grams up to the given length). Qualitatively, the largest gain in EEG prediction accuracy (AM*t* versus A) was associated with n-gram = 4 in monkeys and n-gram = 16 in humans (the average inter-inset interval across melodies was 0.315 s, hence we estimate that n-grams equal to 4 and 16 correspond approximately to a temporal interval of ∼0.9 and ∼4.7 s, respectively).

See also Figure S3.

DISCUSSION

This study tested whether non-human primates generate auditory expectations while listening to music. Monkeys were passive listeners and were reared without musical exposure. Despite this, monkeys exhibited larger pupil dilation, as well as neural encoding of musical expectations beyond mere acoustic tracking, when listening to real as opposed to shuffled music. Thus, a species that diverged from our ancestors 20–30 mya shows evidence for a conserved auditory mechanism underlying music perception, i.e., a primordial sensitivity to temporal

regularities that may have served as an intermediate step toward the evolution of human musicality. This work further sets the stage for future cross-species and developmental comparative studies to non-invasively track neural encoding of naturalistic musical stimuli.

Are we musical animals because of species-specific predispositions?[51,52] Our results show that musically naive monkeys strongly rely on timing- rather than pitch-based expectations; this speaks for an evolutionarily conserved sensitivity to timing structure (presumably the relationship between successive intervals; STAR Methods), probably evolved for general-purpose auditory scene analysis.[53,54] Such sensitivity is built upon the capacity to compare the duration of successive intervals, enabling an animal to recognize a rhythmic pattern among others and across different temporal scales. This capacity has been observed in other mammals, such as rodents,[55] and could even support communicative behavior in some species (i.e., sperm whales and lemurs[56,57]). As such, it should not be considered exclusive to music; it likely extends to various temporal structures within the auditory environment.[58,59] Phylogenetically conserved timing-based expectations might explain why the capacity to process temporal—and notably musical—structures is observed early during human ontogeny and universally across cultures.[2,3]

The neural responses observed here suggest that monkeys generated timing-based expectations while listening to real music that contained a variety of temporal intervals. This highlights temporal structure processing capacities that go beyond mere acoustic processing or detection of isochronous rhythms.[18,19,60] Despite inevitable species-specific differences in stimulus relevance, both species showed integration of past temporal intervals to generate temporal expectations, showing that in the presence of regularities, the past can help to predict the future.[61] Future studies should test whether, compared with humans, monkeys rely on a lower number of past adjacent intervals. This hypothesis stems from evidence indicating reduced short-term auditory memory in monkeys[49,50] and weaker connections between auditory and motor cortico-thalamic circuits linked with temporal and sequential processing.[53,62]

Unlike their sensitivity to timing, monkeys' sensitivity to pitch regularities seems limited. Although monkeys can learn simple statistical relationships in the spectral domain (i.e., transition probabilities within a small set of events),[28] this capacity might not extend to larger arrays of events such as musical sequences. Monkeys' limited sensitivity to pitch regularities might be linked to an attentional bias favoring timing over pitch, an underrepresentation of complex harmonic pitches in the auditory cortex,[20] or an underdeveloped fronto-temporal dorsal pathway for sequential binding.[62–65] Are pitch-based probabilistic expectations a human innovation within the primate lineage? To address this question, future studies should measure experience-dependent effects[66] after extended music exposure and extend our approach to other non-human primates or human newborns. Negative evidence from enculturated non-human primates or positive evidence from human newborns would strongly argue that pitch-based expectations are uniquely predisposed in humans among primates. Outcomes in this direction will dovetail with the hypothesis that pitch processing convergently evolved in response to similar demands in vocal learning species,

Current Biology
Report

CellPress
OPEN ACCESS

regardless of their evolutionary lineages.[67–69] This would explain the scarce use of pitch cues in poor vocal learning species, such as rhesus monkeys.

Our work paves the way to integrated experiment-modeling and non-invasive pipelines, which can facilitate future cross-species and ontogenetic studies. This will help in mapping the biological building blocks of human musicality, its evolution, and its spread across species.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
- METHOD DETAILS
  - Stimuli
  - Information Dynamics of Music model
  - Experimental procedures
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Pupil analysis
  - Monkey EEG pre-processing
  - ERP analysis
  - TRF analysis
  - Human EEG data pre-processing and analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.12.019.

REFERENCES

1. Zatorre, R.J., and Salimpoor, V.N. (2013). From perception to pleasure: music and its neural substrates. Proc. Natl. Acad. Sci. USA *110*, 10430–10437.

2. Singh, M., and Mehr, S.A. (2023). Universality, domain-specificity and development of psychological responses to music. Nat Rev Psychol. *2*, 333–346.

3. Trehub, S.E. (2003). The developmental origins of musicality. Nat. Neurosci. *6*, 669–673.

4. Savage, P.E., Brown, S., Sakai, E., and Currie, T.E. (2015). Statistical universals reveal the structures and functions of human music. Proc. Natl. Acad. Sci. USA *112*, 8987–8992.

5. Ravignani, A., Delgado, T., and Kirby, S. (2017). Musical evolution in the lab exhibits rhythmic universals. Nat. Hum. Behav. *1*, 1–7.

6. Jacoby, N., Undurraga, E.A., McPherson, M.J., Valdés, J., Ossandón, T., and McDermott, J.H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. Curr. Biol. *29*, 3229–3243.e12.

7. Mehr, S.A., Singh, M., Knox, D., Ketter, D.M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A.A., Hopkins, E.J., et al. (2019). Universality and diversity in human song. Science *366*, 1–17.

8. Edalati, M., Wallois, F., Safaie, J., Ghostine, G., Kongolo, G., Trainor, L.J., and Moghimi, S. (2023). Rhythm in the premature neonate brain: very early processing of auditory beat and meter. J. Neurosci. *43*, 2794–2802.

9. Perani, D., Saccuman, M.C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., and Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. Proc. Natl. Acad. Sci. USA *107*, 4758–4763.

10. Winkler, I., Háden, G.P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. Proc. Natl. Acad. Sci. USA *106*, 2468–2471.

11. Hauser, M.D., and McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. Nat. Neurosci. *6*, 663–668.

12. Krumhansl, C.L. (2000). Rhythm and pitch in music cognition. Psychol. Bull. *126*, 159–179.

13. Pearce, M.T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. Ann. N. Y. Acad. Sci. *1423*, 378–395.

14. Huron, D. (2006). Sweet Anticipation: Music and the Psychology of Expectation. MIT Press.

15. Tillmann, B., and Bigand, E. (2004). The relative importance of local and global structures in music perception. J. Aesthet. Art Crit. *62*, 211–222.

16. Brosch, M., Selezneva, E., Bucks, C., and Scheich, H. (2004). Macaque monkeys discriminate pitch relationships. Cognition *91*, 259–272.

17. Fishman, Y.I., Volkov, I.O., Noh, M.D., Garell, P.C., Bakken, H., Arezzo, J.C., Howard, M.A., and Steinschneider, M. (2001). Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J. Neurophysiol. *86*, 2761–2788.

18. Honing, H., Merchant, H., Háden, G.P., Prado, L., and Bartolo, R. (2012). Rhesus monkeys (Macaca mulatta) detect rhythmic groups in music, but not the beat. PLoS One *7*, e51369.

19. Honing, H., Bouwer, F.L., Prado, L., and Merchant, H. (2018). Rhesus monkeys (Macaca mulatta) Sense isochrony in rhythm, but not the

beat: additional support for the gradual audiomotor evolution hypothesis. Front. Neurosci. *12*, 475.

20. Norman-Haignere, S.V., Kanwisher, N., McDermott, J.H., and Conway, B.R. (2019). Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. Nat. Neurosci. *22*, 1057–1060.

21. Selezneva, E., Deike, S., Knyazeva, S., Scheich, H., Brechmann, A., and Brosch, M. (2013). Rhythm sensitivity in macaque monkeys. Front. Syst. Neurosci. *7*, 49.

22. Wright, A.A., Rivera, J.J., Hulse, S.H., Shyan, M., and Neiworth, J.J. (2000). Music perception and octave generalization in rhesus monkeys. J. Exp. Psychol. Gen. *129*, 291–307.

23. Hattori, Y., and Tomonaga, M. (2020). Rhythmic swaying induced by sound in chimpanzees (Pan troglodytes). Proc. Natl. Acad. Sci. USA *117*, 936–942.

24. McDermott, J., and Hauser, M.D. (2007). Nonhuman primates prefer slow tempos but dislike music overall. Cognition *104*, 654–668.

25. Criscuolo, A., Schwartze, M., Prado, L., Ayala, Y., Merchant, H., and Kotz, S.A. (2023). Macaque monkeys and humans sample temporal regularities in the acoustic environment. Prog. Neurobiol. *229*, 102502.

26. Di Liberto, G.M., Pelofi, C., Bianco, R., Patel, P., Mehta, A.D., Herrero, J.L., de Cheveigné, A., Shamma, S., and Mesgarani, N. (2020). Cortical encoding of melodic expectations in human temporal cortex. eLife *9*, e51784.

27. Kern, P., Heilbron, M., de Lange, F.P.D., and Spaak, E. (2022). Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. eLife *11*, e80935.

28. Kikuchi, Y., Sedley, W., Griffiths, T.D., and Petkov, C.I. (2018). Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. Curr. Opin. Behav. Sci. *21*, 145–153.

29. Vuust, P., Ostergaard, L., Pallesen, K.J., Bailey, C., and Roepstorff, A. (2009). Predictive coding of music – brain responses to rhythmic incongruity. Cortex *45*, 80–92.

30. Koelsch, S., Gunter, T., Friederici, A.D., and Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. J. Cogn. Neurosci. *12*, 520–541.

31. Besson, M., and Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. J. Exp. Psychol. Hum. Percept. Perform. *21*, 1278–1296.

32. Lalor, E.C., Power, A.J., Reilly, R.B., and Foxe, J.J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. J. Neurophysiol. *102*, 349–359.

33. Crosse, M.J., Di Liberto, G.M., Bednar, A., and Lalor, E.C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. *10*, 604.

34. Weineck, K., Wen, O.X., and Henry, M.J. (2022). Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. eLife *11*, e75515.

35. Pearce, M.T., Ruiz, M.H., Kapasi, S., Wiggins, G.A., and Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. NeuroImage *50*, 302–313.

36. Omigie, D., Pearce, M., Lehongre, K., Hasboun, D., Navarro, V., Adam, C., and 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.

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

38. Suzuki, T.W., Kunimatsu, J., and Tanaka, M. (2016). Correlation between pupil size and subjective passage of time in non-human primates. J. Neurosci. *36*, 11331–11337.

39. Barczak, A., O'Connell, M.N., McGinnis, T., Ross, D., Mowery, T., Falchier, A., and Lakatos, P. (2018). Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. Proc. Natl. Acad. Sci. USA *115*, E7605–E7614.

40. Kang, O., and Wheatley, T. (2015). Pupil dilation patterns reflect the contents of consciousness. Conscious. Cogn. *35*, 128–135.

41. Weiss, M.W., Trehub, S.E., Schellenberg, E.G., and Habashi, P. (2016). Pupils dilate for vocal or familiar music. J. Exp. Psychol. Hum. Percept. Perform. *42*, 1061–1065.

42. Bowling, D.L., Graf Ancochea, P., Hove, M.J., and Fitch, W.T. (2019). Pupillometry of groove: evidence for noradrenergic arousal in the link between music and movement. Front. Neurosci. *12*, 1039.

43. Bianco, R., Gold, B.P., Johnson, A.P., and Penhune, V.B. (2019). Music predictability and liking enhance pupil dilation and promote motor learning in non-musicians. Sci. Rep. *9*, 17060.

44. Novembre, G., Lacal, I., Benusiglio, D., Quarta, E., Schito, A., Grasso, S., Caratelli, L., Caminiti, R., Mayer, A.B., and Iannetti, G.D. (2023). A cortical mechanism linking saliency detection and motor reactivity in rhesus monkeys. J. Neurosci.

45. Schwartze, M., Farrugia, N., and Kotz, S.A. (2013). Dissociation of formal and temporal predictability in early auditory evoked potentials. Neuropsychologia *51*, 320–325.

46. Quiroga-Martinez, D.R., C Hansen, N., Højlund, A., Pearce, M., Brattico, E., and Vuust, P. (2020). Musical prediction error responses similarly reduced by predictive uncertainty in musicians and non-musicians. Eur. J. Neurosci. *51*, 2250–2269.

47. Southwell, R., and Chait, M. (2018). Enhanced deviant responses in patterned relative to random sound sequences. Cortex *109*, 92–103.

48. Yaron, A., Hershenhoren, I., and Nelken, I. (2012). Sensitivity to complex statistical regularities in rat auditory cortex. Neuron *76*, 603–615.

49. Fritz, J., Mishkin, M., and Saunders, R.C. (2005). In search of an auditory engram. Proc. Natl. Acad. Sci. USA *102*, 9359–9364.

50. Scott, B.H., Mishkin, M., and Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. Proc. Natl. Acad. Sci. USA *109*, 12237–12241.

51. Fitch, W.T. (2006). The biology and evolution of music: a comparative perspective. Cognition *100*, 173–215.

52. Huron, D. (2003). Is music an evolutionary adaptation? In The Cognitive Neuroscience of Music, I. Peretz, and R.J. Zatorre, eds. Oxford University Press.

53. Merchant, H., and Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. Front. Neurosci. *7*, 1–8.

54. Trainor, L.J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140089.

55. Celma-Miralles, A., and Toro, J.M. (2020). Non-human animals detect the rhythmic structure of a familiar tune. Psychon. Bull. Rev. *27*, 694–699.

56. De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., Giacoma, C., Ravignani, A., and Gamba, M. (2021). Categorical rhythms in a singing primate. Curr. Biol. *31*, R1379–R1380.

57. Hersh, T.A., Gero, S., Rendell, L., Cantor, M., Weilgart, L., Amano, M., Dawson, S.M., Slooten, E., Johnson, C.M., Kerr, I., et al. (2022). Evidence from sperm whale clans of symbolic marking in non-human cultures. Proc. Natl. Acad. Sci. USA *119*, e2201692119.

58. Lenc, T., Merchant, H., Keller, P.E., Honing, H., Varlet, M., and Nozaradan, S. (2021). Mapping between sound, brain and behaviour: four-level framework for understanding rhythm processing in humans and non-human primates. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200325.

59. Rajendran, V.G., Teki, S., and Schnupp, J.W.H. (2018). Temporal processing in audition: insights from music. Neuroscience *389*, 4–18.

🔗 **CellPress**
OPEN ACCESS

60. Celma-Miralles, A., and Toro, J.M. (2020). Discrimination of temporal regularity in rats (Rattus norvegicus) and humans (Homo sapiens). J. Comp. Psychol. *134*, 3–10.

61. Winkler, I., Denham, S.L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn. Sci. *13*, 532–540.

62. Rauschecker, J.P., and Scott, S.K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. *12*, 718–724.

63. Balezeau, F., Wilson, B., Gallardo, G., Dick, F., Hopkins, W., Anwander, A., Friederici, A.D., Griffiths, T.D., and Petkov, C.I. (2020). Primate auditory prototype in the evolution of the arcuate fasciculus. Nat. Neurosci. *23*, 611–614.

64. Belin, P., and Zatorre, R.J. (2000). "What", "where" and "how" in auditory cortex. Nat. Neurosci. *3*, 965–966.

65. Rilling, J.K., Glasser, M.F., Preuss, T.M., Ma, X., Zhao, T., Hu, X., and Behrens, T.E.J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. Nat. Neurosci. *11*, 426–428.

66. Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat. Neurosci. *6*, 1216–1223.

67. Ravignani, A. (2021). Isochrony, vocal learning, and the acquisition of rhythm and melody. Behav. Brain Sci. *44*, e88.

68. Wagner, B., and Hoeschele, M. (2022). The links between pitch, timbre, musicality, and social bonding from cross-species research. Comp. Cogn. Behav. Rev. *17*, 13–32.

69. Patel, A.D. (2017). Why doesn't a songbird (the European Starling) use pitch to recognize tone sequences? The informational independence hypothesis. Comp. Cogn. Behav. Rev. *12*, 19–32.

70. Harrison, P.M.C., Bianco, R., Chait, M., and Pearce, M.T. (2020). PPM-Decay: a computational model of auditory prediction with memory decay. PLoS Comput. Biol. *16*, e1008304.

71. Cheung, V.K.M., Harrison, P.M.C., Meyer, L., Pearce, M.T., Haynes, J., and Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. Curr. Biol. *29*, 1–9.

72. Gold, B.P., Pearce, M.T., Mas-Herrero, E., Dagher, A., and Zatorre, R.J. (2019). Predictability and uncertainty in the pleasure of music: a reward for learning? J. Neurosci. *39*, 9397–9409.

73. Koelsch, S., Vuust, P., and Friston, K. (2019). Predictive processes and the peculiar case of music. Trends Cogn. Sci. *23*, 63–77.

74. Barto, A., Mirolli, M., and Baldassarre, G. (2013). Novelty or surprise? Front. Psychol. *4*, 907.

75. Pearce, M.T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. PhD thesis, p. 267.

76. Bianco, R., Ptasczynski, L.E., and Omigie, D. (2020). Pupil responses to pitch deviants reflect predictability of melodic sequences. Brain Cogn. *138*, 103621.

77. Baayen, R.H., Davidson, D.J., and Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. *59*, 390–412.

78. Lenth, R. (2018). Emmeans: estimated marginal means, aka least-squares means. ge version 1.2.1. https://CRAN.R-project.org/package=emmeans. R Package.

79. Selezneva, E., Brosch, M., Rathi, S., Vighneshvel, T., and Wetzel, N. (2021). Comparison of pupil dilation responses to unexpected sounds in monkeys and humans. Front. Psychol. *12*, 754604.

80. Zhao, S., Chait, M., Dick, F., Dayan, P., Furukawa, S., and Liao, H.-I. (2019). Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. Nat. Commun. *10*, 4030.

81. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869.

82. Kothe, C.A.E., and Jung, T.-P. (2016). Artifact removal techniques with signal reconstruction. US patent application WO2015047462A2. filed June 3, 2014, and published April 2, 2015.

83. Plechawska-Wojcik, M., Kaczorowska, M., and Zapala, D. (2019). The Artifact Subspace Reconstruction (ASR) for EEG signal correction. A comparative study. In Information Systems Architecture and Technology Proceedings of the 39th International Conference on Information Systems Architecture and Technology – ISAT, J. Świątek, L. Borzemski, and Z. Wilimowska, eds. (Springer International Publishing), pp. 125–135.

84. Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. NeuroImage *198*, 181–197.

85. Woodman, G. (2012). Homologues of human ERP components in nonhuman primates. The Oxford Handbook of Event-Related Potential Components. (Oxford University Press).

86. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190.

87. Nelson, P.C., and Carney, L.H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J. Acoust. Soc. Am. *116*, 2173–2186.

Current Biology
Report

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Monkey data	Dataverse	https://doi.org/10.48557/U5SHX6
Human EEG data	Di Liberto et al.[26]	https://doi.org/10.5061/dryad.g1jwstqmh
Melodies	N/A	http://www.jsbach.net
Software and algorithms		
Analysis code	Github	https://github.com/robilobi/monkey_eeg_music
Information Dynamics Of Music (IDyOM) model	Github	https://github.com/mtpearce/idyom/wiki
Multivariate Temporal response function (mTRF)	Crosse et al.[33]	https://sourceforge.net/projects/aespa/

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Roberta Bianco (roberta.bianco@iit.it).

Materials availability
All melodic sequences and metadata are publicly available at https://doi.org/10.48557/U5SHX6.

Data and code availability

- Neurophysiological data collected from monkeys have been deposited to IIT Dataverse. The accession numbers are listed in the key resources table.
- This paper analyzes existing, publicly available human EEG data. The accession numbers are listed in the key resources table.
- All original code has been deposited to Github and is publicly available. The link is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants
Two male rhesus monkeys (*Macaca mulatta*) reared without music exposure were tested: Monkey M (11 years old, 10 Kg) and Monkey T (11 years old, 9.8 Kg). The animals were previously involved in other studies[44] and trained to sit on a primate chair, with their head fixed, wearing an ad hoc tailored EEG cap (http://www.easycap.de). The animals had no restrictions on food and water, both before and after the experimental sessions. Animal care and housing, and experimental procedures complied with European (EU Directive 63-2010) and Italian (DL. 26/2014) laws on the use of non-human primates in scientific research.

METHOD DETAILS

Stimuli
Stimuli consisted of 14 monophonic piano melodies: 10 melodies (Real music) written by Johann Sebastian Bach (previously used in Di Liberto et al.,[26] see Table S1 for details), and 4 control melodies (Shuffled music) that entailed disruption of pitch order and timing regularities from four of the original musical melodies. The length of the melodies was variable (average duration = 158.07 s ± 24.06) and the tempo ranged between 47 and 140 bpm (average tempo = 106.5 bpm ± 34.7). The original melodies (Real music) can be found at http://www.jsbach.net/ (see other details in Table S1). The four shuffled melodies were built from four of the real melodies. We generated only 4 shuffled melodies to constrain the experiment within a reasonable duration. These shuffled melodies were specifically built from a selection of original melodies, i.e. those associated with the relatively highest (melodies 05 and 08) and lowest (melodies 01 and 10) temporal-onset mean surprise (see Figure S1 and below for a definition of "surprise"). This decision was driven by evidence showing that music carrying relatively higher timing surprise evokes stronger brain responses in humans,[26] so our selection aimed at balancing these effects across real and shuffled music. Shuffled melodies were matched with the real ones in pitch content, average note duration and inter-onset intervals (IOIs), but their predictable structure was disrupted along two musical

CellPress
OPEN ACCESS

features. Pitch regularities were disrupted by shuffling the temporal order of the original notes. Rhythmic regularities were disrupted by generating a new pool of IOIs Gaussian-distributed around the original mean IOI ± the distance between the mean and the minimum original IOI. The random IOIs were then quantized in MuseScore software (https://musescore.org, version 3.3.4.24412) to the 16th note to remain organized in integer ratios. In MuseScore, the MIDI velocity of all notes (which roughly corresponds to the note's loudness in MIDI notation) was set constant (value = 100), and piano sound wav files were synthesized at a sampling rate of 44,100 Hz. All melodies were flanked with a beep (800 Hz pure tone, linearly ramped with 5 ms fade-in and fade-out) and 5 s silence according to the following structure: beep-silence-music-silence-beep. Finally, the.wav files were converted to mono, and amplitude-normalized by dividing by the standard deviation using Matlab (R2019, The MathWorks, Natick, MA, USA).

Information Dynamics of Music model

Stimuli were modelled in terms of note-by-note unexpectedness (as represented by surprise) and uncertainty (as represented by entropy) using the Information Dynamics of Music model (IDyOM, https://www.marcus-pearce.com/idyom/).[13] IDyOM is an unsupervised variable-order Markov model of statistical learning and probabilistic predictions that learns the stimulus statistics by storing previous events and sub-sequences of events (n-grams) from given input sequences. The algorithm reads a musical sequence (one event, or note, at a time) and it generates a probability distribution for each new event by combining predictions from n-gram models of different orders. For example, a Markov model of order n–1, known as an n-gram model, calculates the probability of a note based on the context of the previous n–1 notes. Thus, the probability distribution over every possible note continuation is estimated for every n-gram context up to a given length (model order-bound). The model outputs *Surprise* (S, or Information Content) and *Entropy* (E) as a function of time. Entropy at time 't' is based on the probability distribution of potential notes considering observations before the event at 't' is observed. It is computed as the sum of all (negative) log-probabilities across all possible notes previously experienced by the model before time 't'. Surprise at 't' is the specific negative log-probability of the note actually observed at time 't' in continuation of the previous sequence. It uses the same probability distribution but incorporates the event at 't'. Note that the term 'surprise', used here as in previous work relying on IDyOM,[26,43,70–72] simply refers to the negative log probability of the event[73] and should be distinguished from 'Bayesian surprise[74]'. In sum, E is often used to reflect the "uncertainty" about the event at 't' before it occurs, whilst S reflects its unexpectedness once the event has occurred. IDyOM can incorporate both a short-term model and a long-term model using a geometric weighted mean.[75] The short-term model is trained on the current sequence to simulate the statistics learned from the specific stimulus at hand. The long-term model is incrementally trained on stimulus material seen by the model before submitting a specific sequence to simulate the listener's prior musical exposure. In this way, IDyOM can generate both schematic and dynamic predictions[14] based on both global regularities (from the training material) and local regularities (from the portion of the sequence heard up to that point). In line with previous literature,[26,43,72] we used an unbounded n-gram configuration that uses contexts up to the size of each musical piece. We also run additional analyses using IDyOM configurations with reduced model order-bound (1 to 4, 8, 12, 16, 20, 24) to reproduce differences in short-term memory. To simulate the statistical knowledge that the (naïve) monkeys would acquire through mere exposure to the stimuli, predictions were derived from a combination of short and long-term models, with the latter being trained only on the stimuli used in the experiment i.e. via resampling (10-fold cross-validation) (in IDyOM terminology: no pretraining, "both+" model configuration). To simulate the long-term statistical knowledge of human subjects, which instead is acquired through lifetime exposure to Western tonal music, predictions were derived from a combination of short- and long-term models, with the latter being additionally trained on a large corpus of Western tonal music (comprising 152 Canadian folk songs, 566 German folk songs from the Essen folk song collection, and 185 chorale melodies harmonised by Bach; as in previous applications of IDyOM).[26,72,76] Control analyses using the same IDyOM model for both species lead to a similar pattern of results (Figure S3B). Amongst the several facets of music that IDyOM can account for, here, we focused on two fundamental dimensions of music that best describe piano monophonic melodies: the pitch and timing of each note. Accordingly, time series representing pitch and inter-onset interval ratios (unlinked 'cpitch and ioi-ratio' IDyOM viewpoints) were independently used by IDyOM to estimate note-by-note surprise (S) and entropy (E) for pitch and timing onset (Sp, Ep, and St, Et, respectively), before combining these as the joint (sum) probability of the note (S, E).

Experimental procedures

The experimental procedure was conducted within a darkened and acoustically insulated room. The two monkeys were tested simultaneously whilst sitting in custom-made primate chairs (Crist Instrument, USA). The monkey's head, which was restrained through a head post, faced a screen (40-inch monitor, 100 Hz, 800-600 resolution, 32-bit color depth; monitor-eye distance: 140 cm). Monkey M (M1) had the monitor slightly on its right side (approximately 30 degrees from the midline, i.e. at 1 o'clock) while Monkey T (M2) slightly on its left side (i.e. at 11 o'clock). The stimuli were presented free field from two audio speakers (Creative Inspire T10) placed on either side of the monitor and delivered at ∼75 dB SPL. The monkeys watched a silent movie consisting of a sequence of different images (displaying natural and urban landscapes), slowly fading in and out (image duration 7 s, fade in / out 2 s), presented in random order.

EEG was recorded using the 26-channel Biosemi Active-2 system and pupil diameter was recorded with an infrared eye-tracking camera (Arrington Research, ViewPoint EyeTracker software). The procedure started with EEG preparation (taking approximately 30 min). This included securing the EEG cap on the monkeys' (shaved) heads and placing all recording electrodes by keeping electrode offset value < 20 (cf. Biosemi ActiveView software). This was followed by calibration of the eye tracker, which took approximately 10 min, and then by 45-min of EEG recording, hereafter referred to as "Session". During each Session, the animals passively

listened to the 14 stimuli, each presented once and in random order (inter-stimulus interval jittered between 500 and 2000 ms), whilst watching the silent movies. Monkeys underwent 26 sessions.

The stimulus presentation and the synchronization with EEG and eye-tracker acquisition devices were controlled by Presentation software (Version 23.0, Neuro- behavioural System, Berkeley, CA).

Pupil recording

We recorded pupil dilation as a covert indicator of the monkeys' attentional state and engagement.[39] The infrared eye-tracking camera, positioned just below the monitor, continuously tracked gaze position and estimated pupil diameter at a sampling rate of 225.79 Hz. The camera focused monocularly on each animal and recorded M1's right eye and M2's left eye. A nine-point calibration procedure was conducted before the recording started following the 9-points built-in procedure of the ViewPoint EyeTracker software. Estimated gaze points were obtained by presenting monkeys with sudden visual stimuli on a black background, leveraging their natural behavior of saccading and fixating on such stimuli. The calibration was considered successful when the following conditions were met: (1) a relatively rectilinear and well-separated configuration of the geometry grid of the estimated points, which had to display the same aspect ratios as those presented on the screen, and (2) an overlap between the estimated and the actual gaze points when the stimuli were presented again.

EEG recording

We recorded neural activity with electroencephalography (EEG) at a 1024-Hz sampling rate. The electrodes were mounted on two custom-made caps (http://www.easycap.de), tailored to fit each animal's head. The BioSemi system replaces the ground and reference electrodes with two electrodes named CMS (Common Mode Sense, active electrode) and DRL (Driven Right Leg, passive electrode). According to the system's guidelines, CMS should (ideally) be placed in the center of the measuring electrodes, while DRL should be placed relatively away from them. While placing CMS, we also had to consider the position of the headpost (approximately over Cz in M1, and CPz in M2). Therefore, CMS was placed on CPz (in monkey M1) and Cz (in monkey M2). DRL was always placed on the frontal-left side of the animal's head. Two other electrodes ('Fz', 'FCz' in M1; 'Fz', 'F2' in M2) were not recorded due to the presence of the recording chamber (not used for the current study) positioned over the right frontal hemisphere in both animals.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data were pre-processed and analyzed in Matlab R2019. Statistical analyses were run in R (version 4.1.3, 2022-03-10) and included non-parametric tests or linear mixed-effects models (lme4 package[77]). All models comprised Random Effects being Monkey (M1, M2) and Melody IDs (1-14). The Fixed effects were specific to each model (see below). Statistical significance was evaluated by likelihood-ratio tests conducted using the 'anova' function in the stats package. Follow-up contrasts were conducted using the 'emmeans' package and the Tukey method to account for the increased risk of type I error resulting from multiple comparisons.[78] The adjusted p-values were calculated to determine significant differences between conditions. A significance level of $\alpha = 0.05$ was used.

Pupil analysis

Time series representing pupil width were segmented into 120 s long epochs (corresponding to the duration of the shortest melody), each locked to stimulus onset. Pupil width was analyzed throughout each epoch to evaluate the overall sustained response to each melody. Compromised data points during which the eye was either fully or partially closed were set to NaN. Although the monkeys' head was fixed to minimize gaze dispersion, they were not trained to fixate on any specific point on the screen. Therefore, to reduce noise due to large eye movements,[42,79] we computed the median gaze point, presumably reflecting the center of the screen, and set to NaN the samples with x-y coordinates located outside a circular region (diameter 0.5 a.u., which corresponds to 35 × 20 degrees of visual angle) centered on the median gaze. Gaps of missing data points shorter than 1 s, due to blinks or averted gaze, were estimated using shape-preserving piecewise cubic interpolation[80] (in a control analysis, we confirmed that interpolation is not necessary to replicate the presented results). The resulting mean percentage of interpolated data was 9.36 % ± 5.20 for M1; 22.46 % ± 8.19 for M2. Time courses were binned by calculating the median every 500 ms time bins (a bin was set to NaN if it contained less than 50% of data). Finally, if NaN intervals constituted 25 s or more out of the 120 s epoch duration, then this epoch was rejected entirely. Following these procedures, 20% of M1's epochs were excluded, while all M2's epochs were retained. Next, the data were z-scored (within each animal and session), and time series (one per melody) representative of pupil width were entered into a linear mixed effect model with Fixed Effects (and their interaction) being Music (Real, Shuffled) and Time-bins (1-240), with covariate being Session (1-26), and Random Effects being Monkey (1, 2) and Melody IDs (1-14). For control purposes, we run another model controlling for gaze at each data point. This model additionally included gaze y and gaze x (median per Time-bin) as covariates.

Monkey EEG pre-processing

Electrophysiological signals such as EEG are notoriously noisy when recorded from an awake animal. This is mostly because, unlike humans, monkeys do not comply with task instructions, and may generate spontaneous movements that leak artefacts in the EEG recordings. For this reason, we used a combination of open-access denoising algorithms and developed a fully data-driven pipeline for the pre-processing of monkey EEG data. EEG data from 4 out of 26 sessions were corrupted. Thus, the EEG time series associated with 22 sessions were analyzed by applying the same pipeline to the two monkeys' datasets separately. The analysis was conducted primarily using Fieldtrip[81] combined with EEGLAB toolboxes (http://sccn.ucsd.edu/). The continuous EEG data were

CellPress
OPEN ACCESS

bandpass filtered between 1 and 30 Hz (Butterworth filter, zero-phase, order 3), down sampled to 100 Hz, and segmented into epochs starting with the first onset and ending with the last offset of the musical piece. Before re-referencing the data using common average reference (CAR), we provisionally discarded faulty or noisy electrodes in order not to leak noise to all electrodes through CAR. Specifically, mean, STD and peak-to-peak were calculated over time within a trial for each electrode; if any of these variables for one electrode was 2.75 STD away from the mean of the other electrodes, that electrode was flagged noisy or faulty. This process was iterated without the outlier electrode(s) until a distribution without outliers was found. AF3 and AF4 were not excluded because they were necessary for the automatic detection of eye-related artefacts (see below). The data were further de-noised in EEGLAB using a validated algorithm for automatic artefact-correction: Artefact Subspace Reconstruction (ASR, threshold value 5),[82] previously applied to human and monkey EEG data.[44,83] Eye-movement artefacts were corrected using another automatic artefact-correction algorithm implemented in EEGLAB: ICLabel.[84] For this, following independent component analysis (ICA, using EEGlab's "runica" function), independent components that ICLabel categorised as related to eye movements (with > 90% likelihood) were rejected. At this stage, the electrodes initially excluded from the analysis (being either faulty or noisy) as well as those corresponding to the scalp position where the headpost and the recording chamber were placed ('Cz', 'CPz', 'Fz', 'FCz' for M1 or 'Cz', 'CPz', 'Fz', 'F2' for M2) were interpolated by replacing their voltage with the average voltage of the neighboring electrodes (25 mm distance). If, following the above pre-processing, noisy electrodes were still automatically identified, the interpolation step was repeated (the number of such iterations varied between 1 and 2 across sessions).

ERP analysis

Event-related potential (ERP) analyses were conducted by segmenting the EEG data into 200 ms epochs, starting 50 ms before the onset of each note, and ending 150 ms following the onset. This time window was chosen based on the observation that monkeys' ERPs (namely the P30, the N70 and the P130) exhibit earlier latencies than their human homologous (P50, N100 and P200) possibly due to anatomical differences in the length of the recruited fiber tracts.[44,85] To quantify the ERP modulation as a function of note surprise, we selected the notes associated with the highest and lowest 20% S values for each melody. For each session, the epochs of interest were then averaged across melodies and baseline corrected using the -50 to 0 ms pre-stimulus interval. The effect of surprise was tested separately for Real and Shuffled music. Statistics were corrected for multi-comparisons using cluster-based permutation tests (as implemented in Fieldtrip; 1000 permutations[86]). A cluster had to be composed of at least two consecutive time points with a p-value <.05 on at least 3 neighboring EEG electrodes.

TRF analysis

We used Temporal Response Functions (TRF) to model the EEG responses to the continuous acoustic and musical features of the presented stimuli (see below for details). This analysis was carried out using the mTRF Matlab toolbox.[33] Each stimulus feature was normalized across time for each melody such that the root mean square of each feature was equal to 1. The EEG data time series corresponding to the same melody were averaged across sessions to increase the signal-to-noise ratio. We ran a forward model that predicts the ongoing EEG response from stimulus features with a time lag window of -50 and +150 ms to capture the fluctuations in the EEG evoked by changes in the stimulus feature. In practice, the model weights capture ERP-like responses,[32,33] so the specific time window was chosen to equal the relevant time window based on the ERP analysis. We used ridge regression to avoid overfitting (range of lambda values: 10^{-4} - 10^{8}). The TRF was fit using leave-one-melody-out cross-validation across all melodies (using the function "mTRFcrossval"). The EEG time course of the left-out melody was predicted using the function "mTRFpredict". Next, the TRF prediction accuracy was quantified by computing the Pearson correlation between the predicted and actual EEG data for each electrode (*r* correlation values). To compare the prediction accuracies (associated with each melody) across different models representing the stimulus, we averaged the *r* values for the ten electrodes displaying the highest TRF correlations for the baseline acoustic model A (see below). These electrodes also corresponded to a canonical auditory fronto-central topography (Figure 2B).

In line with the approach used in previous human work,[26,27,34] we used the TRF procedure to test three different models (see Figure 2A): (i) a baseline Acoustic model accounting for the low-level acoustic features of the music stimuli (model A; predictors: acoustic onset, spectral flux); (ii) a Music model, that includes the features of the A model as well as features with impulses at the note onsets but whose amplitudes are set to the pitch and time-onset surprise and entropy values from IDyOM (model AM; predictors: acoustic onset and spectral flux, plus Surprise *pitch*, Surprise *timing* and Entropy *pitch*, Entropy *timing* – S_p, S_t and E_p, E_t); and (iii) a Control model, where the note onset times of the Surprise and Entropy features were maintained but the Surprise and Entropy values were shuffled independently from one another (model AMc; predictors: acoustic onset, spectral flux, and shuffled S_p, S_t and E_p, E_t). Whilst we adhere to Di Liberto et al.[26] for the Musical features (S_p, S_t, E_p, E_t), we updated the acoustic model, originally including envelope and its derivative, based on recent research.[27,34] Specifically, a reanalysis of the data by Di Liberto et al.[26] demonstrated that incorporating an acoustic onset vector alone explained a non-negligible amount of additional variance in the data (see also Kern et al.[27]). Moreover, spectral flux has been demonstrated to outperform envelope and its derivative in neural music tracking.[34] Upon inclusion of the onset vector, we observed that adding spectral flux or adding envelope and its derivative produced comparable results.

For each melody, the prediction accuracy of model A was subtracted from the prediction accuracies of models AM and AMc to quantify the unique contribution of the combined four musical features in the EEG prediction. We expected model AM to have a higher prediction accuracy than model A, suggesting that musical surprise and entropy account for dynamics of the EEG signal that are not accounted for by the acoustic features. However, because simply adding features to a model can generally improve performance, we also compared the gain in prediction accuracy of model AM to that of AMc, which has the same dimensionality as AM but

meaningless musical information. If surprise and entropy are encoded in the brain response, the gain of AM is expected to be higher than AMc. We quantified this effect statistically by using a linear mixed effect model with two Fixed effects: Model ID (AM-A, AMc-A) and Music (Real, Shuffled) and two Random effects: Monkeys (1, 2) and Melody IDs (1-14).

In addition, we quantified the separate contribution of pitch and timing features in the TRF model performance to assess how well the individual features were tracked by the EEG. We created reduced models from the full AM model by selectively including the Surprise and Entropy features representing either only pitch or only timing, hence yielding AMp (only S_p and E_p) and AMt (only S_t and E_t) models, respectively. The difference in prediction accuracies between the full and the reduced model provided an estimate of the specific contribution of the included variable. Also, for comparability to our earlier analyses, the TRF prediction accuracies of model A were subtracted from those of models AM, AMp and AMt (Figure 3A). The obtained delta values were entered into a linear mixed effect model with two Fixed effects: Model ID (AM-A, AMp-A, AMt-A) and Music (Real, Shuffled), and two random effects: Monkey (1, 2) and Melody IDs (1-14). In a control analysis, we repeated the above procedure while relying on a richer model A, including predictors simulating midbrain auditory responses.[87] This control analysis corroborated the results reported in the main text (compare Figure 3 with Figure S3C).

Finally, to assess the contribution of local context (or short-term memory) to the TRF analysis, we created AMp and AMt models with surprise and entropy values derived using a progressively smaller n-gram context (model order k, see information dynamics of musical model above). We thus assessed the prediction accuracy gain (relative to model A) obtained from AMp and AMt models relying on IDyOM estimates based on 1, 2, 3, 4, 8, 12, 16, 20, and 24 model order.

Human EEG data pre-processing and analysis

We analyzed an open EEG dataset from a previously published study (to which the first author contributed[26]). This consisted of EEG data recorded from 20 human participants listening to the same set of musical pieces used in the current study, except for the shuffled melodies created for the current study. In the human study, each melody was presented three times in pseudo-randomized order. EEG was acquired using a 64-electrode BioSemi Active Two system (the same one used for recording the monkey EEG data) and digitized at a sampling rate of 512 Hz. For more details about the acquisition settings refer to the original paper.

The human EEG data were pre-processed and analyzed following the same pipeline applied to the monkey EEG data (see monkey EEG pre-processing above). Before computing participant specific TRFs, the time series were averaged over the three repetitions of the melodies. Next, we ran a forward model for each subject using ridge regression (range of lambda values: 10^{-4} - 10^{8}), with 0-350 ms time lag, in line with the optimal time window found in Di Liberto et al.[26] on the same dataset.