An ancient evolutionary calculus for attention signaling retained in modern music

Gregory A. Babbitt[1] and Ernest P. Fokoue[2]

[1] *T.H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester NY, USA*
[2] *School of Mathematical Sciences, Rochester Institute of Technology, Rochester NY, USA*

May 2026

Abstract

What is in a song? Although the evolution of music has been subject to philosophical and scientific study, the exact mechanism(s) of fitness signaled through the quality of music has not. As music and dance functionally aim to hold an observer's attention, we investigate the features of sound that stimulate attentive behavior in an audience. We propose a calculus of audiovisual attention to determine an individual's f(x) = Control=position, f'(x) = Energy=change in position and f''(x)=Surprise=change in direction (CES). We develop theory and software to, (A) extract audio CES, (B) dynamically map audio trajectories within the ternary space of CES, and (C) measure performer stability (i.e. quality) and song individuality. We validate that both human music and human/animal vocalizations all maintain high stability in the space of CES, with professional musicians being the most stable. Stability of CES also reflects differences in expert vs. novice singing ability in both humans and canaries and imposes constraints during live audience interaction as well. We conclude that quality signaling via this calculus of attention common to behavioral display reflects the physics of observing audiovisual motion, probably originating from selective pressure of intra/interspecific conflicts formed after the Ediacaran-Cambrian boundary, and retained in music today.

0.1 key words

evolution, music, attention, calculus, bird song, primate vocalization, machine learning

0.2 lay summary

Most animals (including humans) pay attention to specific audiovisual stimuli within their environment. Here we explore common features of these stimuli and propose a single mathematical framework (with software) for detecting and measuring the quality of performance regarding attention signaling in the context of sound. We show that musical sounds are exceptionally attractive to our attention and that attention signaling in music probably has ancient evolutionary roots that connect to when early life forms started to evolve swift directed movement and the sensory capacities to detect it.

1 Introduction

What is 'song'? What is in a 'song'? Why have we applied the word 'song' not only to our own human vocal-social behavior, but also to the vocal behavior of some other animal groups as well? Why are vocalizations in animal groups such as suboscine passerines (i.e. songbirds), some cetaceans (i.e. humpback whales), some primates such as Hylobates (i.e. gibbons and siamangs), and some anurans (i.e. chorus frogs) viewed as more musical to our ears than most other animal groups? Is this difference purely an artifact of human perception or is there a real acoustic difference that we can detect with a technological approach? If singing is the oldest form of human musical behavior, then what can a mathematical comparison of human singing to what we call 'song' in the vocalizations of other animals inform us about what music is and perhaps what it is not?

The origin, purpose, and nature of music have long been viewed through very different intellectual perspectives. Natural philosophers have repeatedly observed that music has potential roots in both cosmic and Earthly forms of nature [41][91], and like mathematics[92], music also seems to possess elements of both an internally-represented language in our minds as well as an external physical reality that we can directly and objectively measure. More recently scientists across diverse fields have also explored the origin and function of music. Recent consensus across scientific disciplines might conclude that music contains features very close to and yet predating human language [30][29][63][71], is culturally diverse and yet rooted to some degree in genetics [10][11][38][59],

has structural features (e.g. isochrony, improvisational dueting, harmonic overtone series, intentional composition) that occur in some animal vocalizations outside humans [7][17][22][23][18][6][34][35][68][72][79] and can even can occur spontaneously within artificial intelligence [42]. Music fosters human social connection in the absence of shared language [69][75][80][93], and evokes stimulation and development across large parts of the human brain [3][9][31][37][74][84][90]. Anthropologists and biomusicologists have linked the evolution of music to Darwinian fitness [2][36][60] and the rise of bipedalism [80], the formation of prosocial behavior in human societies [64][75] and perhaps molecular systems of interaction as well [5][4]. Musical behavior has also been linked to the evolution of self-awareness and the ability for cognitive abstraction via theory of the mind [23][80]. One perspective that has been less often applied to the subject of music is that of behavioral ecology. Here, we would begin with basic questions about what fitness advantage is ultimately conferred to individuals engaging in musical behavior [36], and more specifically, what are the exact proximate mechanisms that signal individual quality (i.e fitness) in the performance display of musical behavior. While scientists have proposed many plausible ultimate evolutionary biological and cultural drivers of musical behavior, to our knowledge, the proximate mechanisms by which individual quality is honestly signaled through musical performance has not yet been thoroughly investigated.

Whether or not a song is originated by a human voice, at its most basic level, a song (and dance) most often functions to hold the attention of one or more observers [46]. Therefore, we might define the proximate function of song to simply hold the attention of others. In this respect, it shares a common function with all other forms of behavioral display. Attention or focus in the brain is controlled largely by the dorsal and ventral attention networks, an evolutionary ancient but recently expanded set of connected brain regions [81][89][88][61] with strong anti-correlated activity with the mammalian default mode network [45][27], a set of connected brain regions activated during resting states and responsible for internal mental representations of self-awareness [53]. Interestingly, while music has features that stimulate attention [48] and emotion [94], it also has been shown to stimulate the default mode network as well [83], perhaps suggesting some interesting but as yet unexplored causal links between the evolution of human music and the several largely unexplained brain size expansion events in hominid evolution [28] [77].

When we consider musical performance within the context of other types of animal behavioral display, we can see that there are some potential common features or rules that govern the evolution and dynamics of an individual's performance display. In all animal displays, audiences expect a form of authenticity or honesty in a performer. In animal behavior, this requires that a display conveys a more-or-less honest signal of Darwinian fitness. For example, in socially monogamous bird song, the male's singing is an honest indicator of not only his physical quality but also the quality of resources that his defended territory may hold. Males on better territories can sing louder and more frequently than those on lesser territories [43]. In human singing, particularly opera and other live musical performances, we seem to also respond directly to the 'live' presence of the performer [8] perhaps because we appreciate the physicality of the performance by responding well when performers give enhanced effort. And we reject forms of deception like lip-syncing in live performance [65], even though it is quite accepted, even necessary in other forms of performance such as TV or music video. In the context of animal behavior, the cheating or falsification of specific single indicators of fitness often lead to the evolution of behavioral displays with many multi-variate or multi-modal components fitness, as a combination of feature components are harder to cheat as a whole than any single feature [15] [54]. This unwritten contract of authenticity between the performer and their audience also leads to a increased variance or individuality in behavioral display across a population as well. The observation of increased variance in genetic, behavioral and/or morphological traits is one way evolutionary biologists have traditionally used to detect sexual selection acting on a species [66] [82] [78]. Similarly, in music, increased individuality and variance among professional performers can equate to a 'signature sound' that can be very essential for long-term mainstream succuss [25].

While compelling in its simplicity, a performance-based view of the function of music does not explain everything. For example, many forms of 'ambient' music have evolved where there is no apparent performer and the music seems to serve to create a space where the listener can enter and occupy. In this sense, perhaps ambient forms of music gives the listener the illusion of being in the center of a performance space as opposed to a seat in the audience. There is some evidence that this kind of music might also be less likely to stimulate the attention network of the brain and more likely to play upon self-awareness via the default mode network [40] [70] [62]. Music also has many cultural functions unrelated to individual performance that may have contributed to its evolution as well. These are mother-infant bonding (i.e. lullaby) [52], organization and bonding during social enterprise (i.e. work songs) [73] and spatial-behavioral organization during military campaigns (i.e.. marching) [57]. These many culturally-defined aspects of music, perhaps layered over more ancient behavioral functions make it a challenging and complex subject of study.

Here we propose that some basic investigation of acoustical features of music that arouse attentive behavior in the listener could prove especially insightful regarding the common behavioral mechanism(s) that probably underlie the diverse cultural functions of music. The need for attentive behavior probably first arose during the transition from the Ediacaran to Cambrian (E-C) period about 541 million years ago when animal life first evolved swift directed movements as well as the more complex brains and sensory organs required to

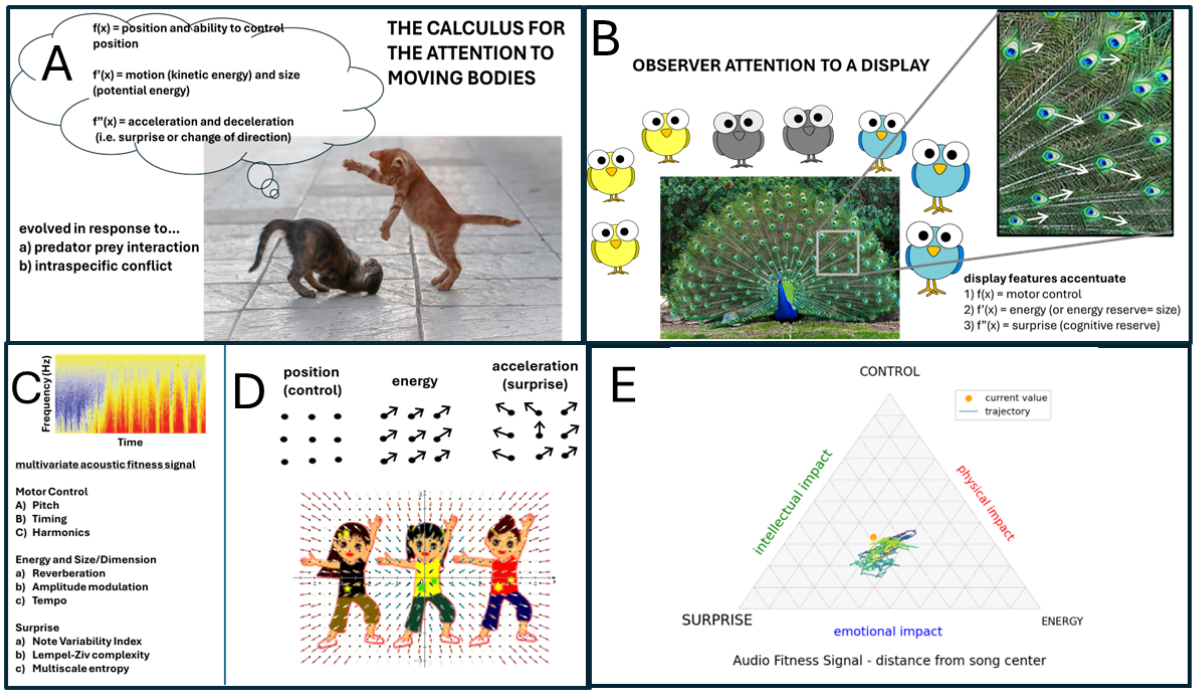

Figure 1: **The theory of the calculus of attention examined through musical stimulation of human minds.** (A) The calculus of attention focuses upon three components of biological fitness; motor control, energy level, and cognitive capacity demonstrated via surprise or novelty. (B) It is captured by many proximate stimuli present in animal displays. For example, a female bird might choose among prospective males based upon how loud or fast they sing (indicator of energy), how well they conform to a species-specific call (indicator of motor control), and in some species such as mockingbirds, nightingales, and lyrebirds, how well they can learn to improvise (indicator of cognitive reserve). (C) Human music and dance also capture attention using these same three elements of display. (D) These outward displays of attention signaling by an individual musician or group of musicians are plotted dynamically over time by the POPSTAR software and represented in a moving image ternary plot where the vertices are normalized multivariate measures of control, energy and surprise (likely processed via the dorsal attention network in the brain) and the sides combine to affect the internal focused physical, emotional, and intellectual sensations (likely processed via the default mode network of the brain). Example .mp4 videos can be found at our website https://gbabbitt.github.io/POPSTAR-signaling-in-music/

detect these movements. Associated with this E-C boundary are the rise of many new body forms, extensive burrowing behavior, and the evolution of protective shells covering soft body parts [16][55]. Attentiveness to swiftly moving bodies can be described mathematically by a simple calculus where an observer must accurately gauge the position of another body in time and space (Figure 1A). This requires accurate reading of the control of position of the other organism (f(x)), its energy or velocity relative to the observer (f'(x)), and its change of direction through acceleration and deceleration relative to the observer (f"(x)). An incomplete ability of the observer to perfectly account for changes in direction would create an element of surprise and unpredictability on the part of the observer. We can also similarly describe the quality of performance displays of an observed individual directly through their exhibition of motor control, kinetic or potential energy, and its cognitive reserve (i.e. ability to surprise). In the context of perhaps all behavioral displays of individual fitness across many animal species, including humans, we propose that this calculus of attention is always present to some degree, and that audiences of behavior displays of individual quality will be attentively engaged by any components of the display that also convey honest signaling regarding motor control, energy expenditure, and surprise (i.e. complexity) (Figure 1B). We can directly relate this general idea to multi-modal tail display in peacocks (Pavo cristatus), where males perform a tail train rattling movements via feather stridulation at 25.6 Hz providing a visual tail display where stationary eyespots are presented in front of a background of feather undulations to hold the attention of females and affect mate choice. [15]. In human music and dance, we can also readily observe elements of control, energy, and surprise (CES) and measure them in sound itself using a variety of audio processing methods (Figure 1C). We can mathematically define the CES of external sounds in a ternary space and map them dynamically over time Figure 1D). We hypothesize that the more physically, energetically, and cognitively fit an individual musician or animal vocalist is, the better they can (A) control and stabilize their sound in the ternary space of CES and (B) move more intentionally within it (i.e. poor performance will create more randomness in the CES trajectory). We introduce an open-source, user-friendly software application named the POPSTAR PROJECT (https://gbabbitt.github.io/POPSTAR-signaling-in-music/) to dynamically map and analyze sound files (.wav or .mp3) in the space of CES. We introduce a stability indicator function for CES that measures performance with respect to the calculus of attention. We then compare (A) musical and non-musical sounds, (B) various animal vocalizations, (C) musical genres with different emphasis on solo performance, and (D) differences in musical performance ability and environment (e.g. expert vs. novice, live vs. studio). We find compelling evidence that this ancient attention signaling is representative of an individual's physical, energetic, and cognitive ability and is still retained in most modern musical performance today.

2 Methods

The POPSTAR software begins with the audio file in .mp4 or .wav format and segments it into user defined windows spaced on either (A) every beat using python's librosa library beat tracker algorithm [20] (if user checks the box labeled 'beats" on the GUI), or (B) every $\frac{1}{4}$ second if the user leaves 'beats' unchecked. We extract nine acoustic features of attention signaling from each audio segment, 3 each for control, energy, and surprise (i.e. CES). Features are extracted from a sliding window of user defined length that moves through the entire audio file. The default that is recommended by the software is a sliding window of 8s. We define performance quality directly in terms of acoustic measurement of physical and cognitive indicators of individual quality (i.e. fitness) such as motor control, energy, and complexity. Note that this is quite different than mathematical descriptions of Darwinian fitness as heritable phenotypic changes in populations (e.g. the Price equation [26]). While it is an obvious conjecture that direct indicators of physical and cognitive quality might translate to Darwinian fitness within populations, we do not measure or model this here.

2.1 Acoustic features of control (C)

2.1.1 Control of pitch or frequency

Pitch control (PC) is defined as

$$\mathbf{PC} = \log\left(1/\left(\sum_{i=1}^{N}\left|\frac{(O-E)^2}{E}\right|/N\right)\right) \tag{2.1.1}$$

Where N = number of notes discovered with a f0 or fundamental frequency and O = f0 in Hz determined by librosa pYIN algorithm [14][49] and E = expected frequency of the nearest value on the western music scale at A440 Hz or else is determined by nearest cluster mean frequency determined via expectation-maximization (EM) clustering. If the user chooses music' or 'single' as sound-type on the GUI, then the Western music scale is used. If the user choses 'other' as sound-type on the GUI then the clustering approach is used.

2.1.2 Control of harmonics

Harmonics control (HC) is determined as the harmonic energy (HE) summed over N notes and L levels of harmonics above f0 detected by the librosa harmonic spectrum analyzer [67]. The harmonic energy (HE) of a given note is measured by the number of harmonic levels f(L) detected above the fundemental frequency of the a given note f0. Harmonic control is the harmonic energy averaged over the number of notes detected.

$$\mathbf{HE} = \log\left(\sum_{i=1}^{L} f(L)\right) \tag{2.1.2}$$

$$\mathbf{HC} = \log\left(\frac{\sum_{i=1}^{N} HE}{N}\right) \tag{2.1.3}$$

Timing control (TC is determined as

$$\mathbf{TC} = \log\left(1/\left(\sum_{i=1}^{N}\left|\frac{(O-E)^2}{E}\right|/B\right)\right) \tag{2.1.4}$$

Where O is observed beat interval and E is mean beat interval returned by librosa beat tracking software [20] and B is total number of beats

2.2 Acoustic features of energy (E)

2.2.1 Tempo

Tempo (TP) is defined as

$$\mathbf{TP} = B/t \tag{2.2.1}$$

where B = total beats in audio segment and t = time in minutes

2.2.2 Reverberation

Reverberation (RV) is determined as

$$\mathbf{RV} = r_{L=1} \tag{2.2.2}$$

Where r is the Pearson correlation coefficient at lag (L) = 1 or the nearest peak to the center taken from the autocorrelation function (ACF) of the time domain audio signal. The ACF with time lag (k) is computed across the time domain (x) as

$$\text{ACF} = p(k) = \frac{cov(x_t, x_{t-k})}{cov(x_t, x_t)} \tag{2.2.3}$$

2.2.3 Amplitude

The amplitude volume (AM) is determined as

$$\mathbf{AM} = \log\left(\sum_{i=1}^{T} |S|\right) \tag{2.2.4}$$

Where S is the audio signal in time domain and T is the time length of the audio segment

2.3 Acoustic features of surprise (S)

2.3.1 Multi-scale entropy

The multi-scale entropy (MSE) is defined as the complexity index returned from the refined multi-scale entropy (rMSE) function for univariate data from python's EntropyHub library [24] using an Entropy object with embedding dimension =4 and radius threshold = 1.25 to return a Complexity Index over 5 temporal scales using 3rd order Butterworth Filter with normalized corner frequency of 0.6 at each temporal scale.

2.3.2 Lempel-Ziv complexity

The Lempel-Ziv complexity (LZC) is defined as

$$\mathbf{LZC} = \log(s) \tag{2.3.1}$$

Where s = number of distinct substrings in the binary representation of the audio file

2.3.3 Note variability index

Note variability is defined according to the method of [76] for analyzing bird song complexity. The spectral cross-correlation is converted to a Note Variability Index (NVI) ranging between 0 and 1 where

$$\mathbf{NVI} = \frac{\sum_{i=1}^{N_C}\sum_{j=1}^{N_C}(1 - C_{\Delta(T)})}{N_C(N_C - 1)} \tag{2.3.2}$$

forming an N_C by N_C similarity matrix where N_C is the number of notes and $C_{\Delta(T)}$ is the association between two notes. Thus, a lack of cross-correlation across all pairs of notes or data columns in a spectrogram represents high variability in the sound file while a tone repeated constantly in the time domain has high cross-correlation but no complexity. It is important to understand that variability in the signal increases to the point of unpredictability, thus the variability eventually increases from a complex pattern to memoryless white noise (where the next value in a sequence cannot be predicted at all).

2.4 Min max (self) and Z score normalized average values of 9 features of CES

If the user chooses 'single' file under sound-type on the GUI, the CES feature matrix constructed listed above is min-max normalized by column where

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{2.4.1}$$

If the user chooses 'music' under sound-type, the CES feature matrix is z-score normalizes using the average feature parameter values (μ, σ) for human music where

$$x' = \frac{x - \sigma}{\mu} \tag{2.4.2}$$

If the user chooses 'other' under sound-type, the CES feature matrix is z-score normalizes using the average feature values (μ, σ) for human speech.

2.5 Dynamic ternary plot representation of CES over time

Control, energy and surprise features for each audio segment in time (t) are averages of the normalized features from that audio segment

$$C_t = \text{mean(PC,HC,TC)} \tag{2.5.1}$$

$$E_t = \text{mean(TP,RV,AM)} \tag{2.5.2}$$

$$S_t = \text{mean(MSE,LZC,NVI)} \tag{2.5.3}$$

and thus

$$(C_t, E_t, S_t), t = 1 + w, 2 + w, ...T \tag{2.5.4}$$

where t is user defined as t = $\frac{1}{4}$ second OR beats tracked in the time of the music and w = analysis widow length determined by user (default is 8 seconds) and values of (CES)t are normalized to sum to 1

The single points in CES ternary space are thus

$$(C_t, E_t, S_t), t = 1, 2, ...T \xrightarrow{yeilds} P_t, t = 1, 2, ...T \tag{2.5.5}$$

A dynamic ternary plot is generated as well as a dynamic Chernoff face [13] and saved in movie file (.mp4) format (Figure S1). In the Chernoff face representation, Eye height = NVI, Eye width = LZC, Pupil area = MSE, Eyebrow slant level = TP, Mouth height = AM, Mouth width = RV, Ear height = PC, Ear width = TC, and Nose slant = HC. Examples of .mp4 (movies) containing dynamic Chernoff faces and dynamic ternary plots with sound overlay for a variety of bird song, human/animal vocalization, and human music can be found in examples dynamicPlots.zip. Thus, the POPSTAR software produces a dynamic Chernoff face using all 9 features tracked over time where the three features of control are represented by the ear dimensions (and nose angle) , the three features of energy are represented by the mouth dimensions (and eyebrow slant), and the 3 features of surprise are represented by the eye dimensions. It has been observed that human minds struggle to conceptualize statistical relationships that extend much beyond 3-dimensional space and have trouble tracking more than a handful of objects in real time [12]. Herman Chernoff invented facial representation of complex data because faces are one of the only multivariate structures that evolution has equipped us to rapidly and easily interpret [13]. The normalized averages of the control, energy and surprise are also plotted dynamically as a song trajectory within a ternary plot. See the POPSTAR software landing website and Figure S1 for more details.

2.6 Stability indicator (i.e. performance fitness indicator) for the tracking of CES over time

We develop a simple permutation-based indicator of the stability of the performer's trajectory within the space of CES (Figure 2). It can be observed that both animal vocalizations broadly characterized across anuran, avian, and primate taxa have histograms where most steps in the ternary space of CES are quite small. This distribution tails off with a power lognormal decay often with some multimodal behavior observed in the tail (Figure 2A). Human music is quite similar with even more stability in CES step length (Figure 2B). We hypothesize that high levels of performance quality and resulting musicality of sound (i.e. fitness of performance) should result in more stable and intentional movement within the ternary space of CES. To measure the stability of any given audio signal we create permutations of the observed CES trajectory by randomly shuffling time steps and comparing the resulting observed and randomized distributions of step length in the ternary space of CES (Figure 2C-D). Note this does not alter the values of points in the ternary space, only the order in which they occur, thus

$$(P_1, P_2, ...P_T) \xrightarrow{yeilds} (P_{\sigma 1}, P_{\sigma 2}, ...P_{\sigma T}) \tag{2.6.1}$$

Each observed time step in the ternary space

$$\delta_t = P_{t+1} - P_t \tag{2.6.2}$$

is compared to each corresponding shuffled (i.e. random) time step

$$\delta_{\sigma(t)} = P_{\sigma(t+1)} - P_{\sigma(t)} \tag{2.6.3}$$

and we count instances in which the observed step is less than or greater than its corresponding shuffled step

$$\kappa(A) = \sum_{t=1}^{T} \mathbb{1}(\delta_t < \delta_{\sigma(t)}) \tag{2.6.4}$$

$$\kappa(B) = \sum_{t=1}^{T} \mathbb{1}(\delta_t \geq \delta_{\sigma(t)}) \tag{2.6.5}$$

Thus. an indicator of stability ranging from 0 to 100 percent in the CES trajectory becomes

$$\eta(A) = \left[\left(\frac{\kappa(A)}{\kappa(A) + \kappa(B)}\right) - 0.5\right] * 200 \tag{2.6.6}$$

Nonparametric hypothesis testing on average Ct, Et, St and stability = $\eta(A)$ can be achieved using Kruskal-Wallis H test where

$$H = (n-1) * \frac{\sum_{i=1}^{G} n_t(\bar{r_i} - \bar{r})^2}{\sum_{i=1}^{G} n_t \sum_{j=1}^{n_i} n_t(\bar{r_{ij}} - \bar{r})^2} \tag{2.6.7}$$

Where N is total observations across all groups, g is number of groups, n_i is number of observations in group i, r_{ij} is rank of observation j from group i, $\bar{r}_i$ is average rank of all observations in group I and $\bar{r}$ is average of all r_{ij} r_{ij} Multi-model inference can also be performed on the observed step distributions using Bayesian information criterion

$$BIC = -2ln(L) + k * ln(n) \tag{2.6.8}$$

where L is likelihood of model given the data, k is the number of parameters in the model and n is the number of data points

2.7 Comparative functional data analysis of song trajectory in the dimensions of control, energy, and surprise

For comparative analysis of different groups of songs (i.e. human artists, musical genres, characteristic species songs, etc), we fit b-spline functions to C_t, E_t, S_t trajectories of each song using an optimal tuned bandwidth determined via the highest group classification accuracy achieved when employing the K nearest-neighbors classifier function from python scikit-fda module(). The time-series functions were aligned (registered) using Fisher-Rao elastic registration applied to all songs in the groups being compared.

Elastic registration aligns functional data by separating phase and amplitude variability using the Fisher-Rao Riemannian metric. It works by transforming functions into a Square-Root Velocity Function (SRVF) representation,

$$q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|} \tag{2.7.1}$$

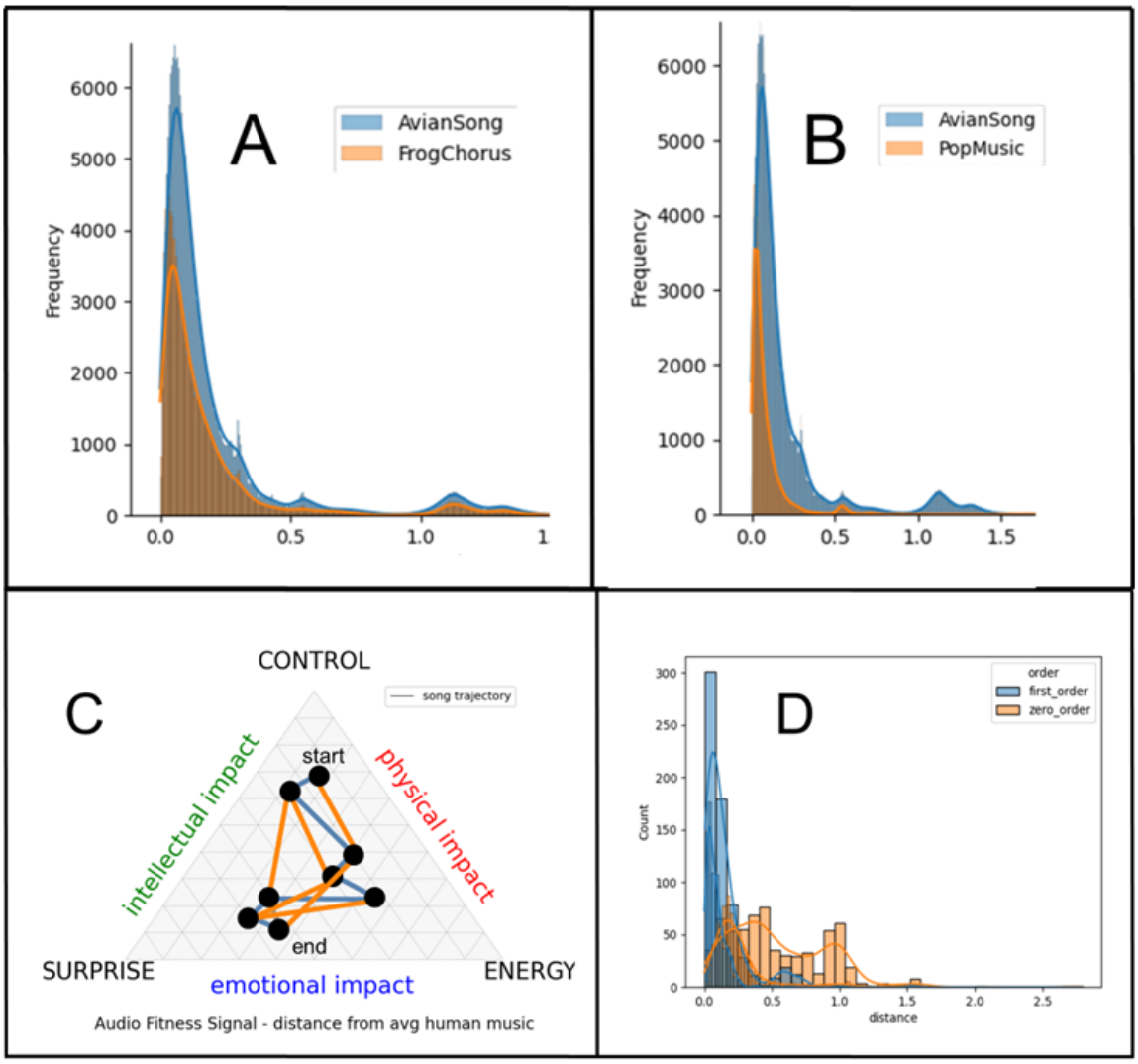

Figure 2: **The shape of the distribution of step distances in the ternary space of control, energy, and surprise (CES) in sound** demonstrates a strikingly common pattern across (A) anuran chorus, avian song that is even more stable in professional (B) human musicians. (C-D) The stability of song trajectory in the space of CES is determined by comparing the step distribution of the song trajectory in the space of CES (blue) to random walks of the song trajectory created by the shuffling of the order of steps (yellow). An expertly created musical sound typically shows a multi-modal lognormal, gamma or power lognormal distribution of steps where step distances are usually very short and controlled until the musician intends to move somewhere else in the space of CES. When the time dependence of the signal is destroyed by shuffling (C) , the distribution of steps (D) becomes more uniform with a larger mean step distance. In the case of brown noise (not shown), no difference in observed and shuffled step distributions is evident. The percentage of the stability in CES can be determined through repeated permutations of random draws from the real- and shuffled-time step distributions and tallying the number of times the time step in the observed trajectory is less than the that of the shuffled trajectory. The significance of this difference can be determined via the Kruskal-Wallis test.

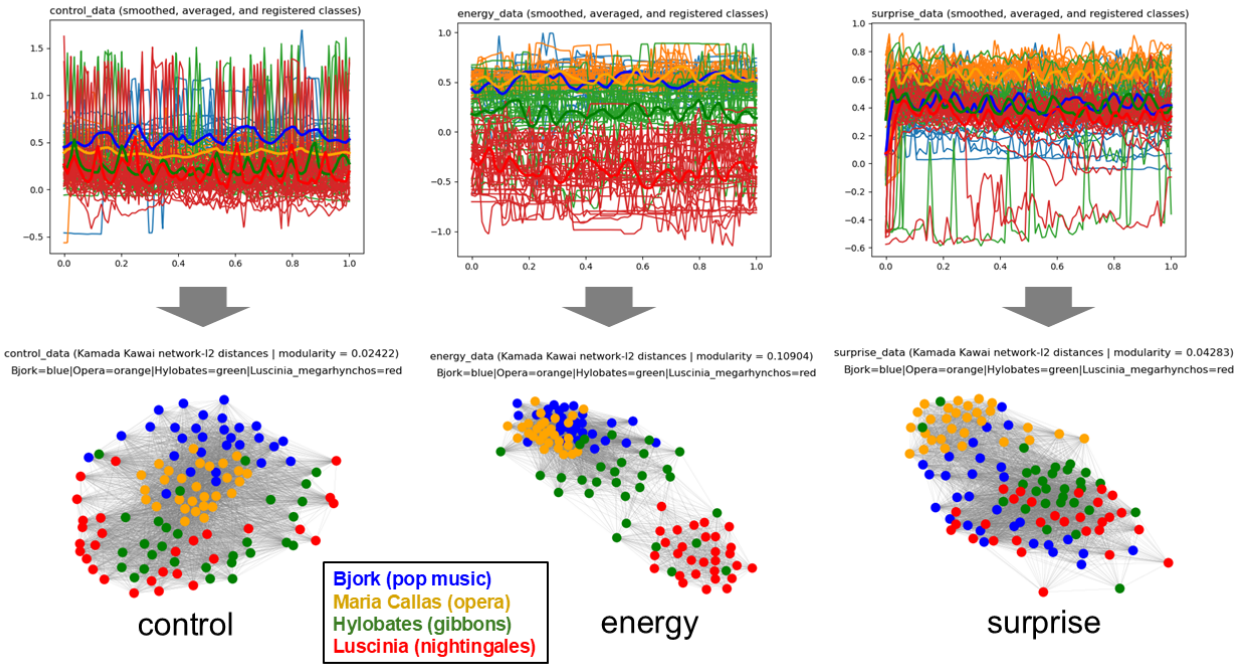

Figure 3: **Functional Data Analysis (FDA) and adjacency networks for comparatively analyzing the similarity/individuality of single song time series trajectories in the space of control, energy, and surprise (CES).** First b-spline functions are generated and aligned using optimal bandwidths determined via machine learning classification of the grouping classes after 70/30 training-test split (method = nearest centroid classifier from python scikit-fda). L2 or Euclidian distances are calculated between all pairs of spline functions in the FDA to create an adjacency matrix. The network is mapped using the Kamada-Kawai method (python – network module) with edges placed between all L2 distances below the median L2 value.

transforming the complex Riemannian metric into a simpler L^2 norm. The optimal warping function γ is found by minimizing

$$d_{FR}(f_1, f_2 \circ \gamma) = ||q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}|| \tag{2.7.2}$$

Group machine learning classifications were then conducted using four methods in scikit-fda (KNN, nearest centroid classifier, maximum depth classifier, and a functional quadratic discriminant analysis classifier). Overall classification performance was determined using these four methods as a multi-agent stacked model (Table 1). The individuality or dissimilarity between any two songs was defined as the L^2 distance between the elastic registered b-spline functions for each song in each of the three dimensions C_t, E_t, S_t. An adjacency matrix was constructed for all pairwise comparisons between songs and represented as a Kamada-Kawai force directed graph network() with edges constructed between allL^2 distance below the median of all L^2 distances. Thus the 50 percent closest nodes in the network graph have an attractive force and the highest 50 percent have a repulsive force. Modularity of the network was computed using python network. See Figure 3 for a schematic example of this. The algorithm defines a cost function, often called "stress," which represents the total energy in the network of springs: $E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{ij}(d_{ij} - l_{ij})^2$ Where: E: The total energy of the layout (to be minimized). n: Number of nodes. x_i, y_i: The coordinates of vertex i. d_{ij}: The actual Euclidean (geometric) distance between node i and node j in the layout. l_{ij}: The target (desired) length for the spring between nodes i and j, defined by their shortest path in the graph, usually $l_{ij} = L \cdot d_{ij}^{\text{shortest_path}}$. k_{ij}: The spring constant (or stiffness) between nodes i and j, typically $k_{ij} = K/(d_{ij}^{\text{shortest_path}})^2$.

2.8 Random forest classification

The 9 normalized acoustic features are used to train subsequent random forest models that were used to assess comparative classification of sound types derived from our protein sonification when compared within the context of human speech, music, and other natural and computer-generated sounds. All sound files used are listed in folder list.txt and included in Supplemental Data. Note that some music files are omitted due to copyright, but the titles of the music tracks analyzed can be seen in the first column of featureData allFolders.dat. Our random forests consisted of 500 decision trees using 80/20 percent training/testing data split. It is implemented

with python scikit-learn RandomForestClassifier function which implements the ID3 with CART algorithm to create each decision tree. Splits on the trees are called using the Gini impurity or Gini index (GI).

$$GI = \sum_{i=1}^{N} f_i(1 - f_i) \tag{2.8.1}$$

Where f_i is the frequency of label i at a node and N is the number of labels. The importance for each feature (I) in the random forest is given by

$$I = \frac{\sum_{\text{j=all trees}}^{N} \text{norm}I_j}{T} \tag{2.8.2}$$

where the normalized feature importance is for a give tree is

$$\text{norm}I_{ij} = \frac{I_i}{\sum_{\text{j=all features}}^{N} I_j} \tag{2.8.3}$$

where the feature importance for a given tree is

$$I_i = \frac{\sum_{\text{j=node j splits on feature i}}^{N} ni_j}{\sum_{\text{k=all nodes}}^{N} ni_k} \tag{2.8.4}$$

where n_j the importance of each node j is

$$n_j = w_j G_j - w_{j(left)} G_{j(left)} - w_{j(right)} G_{j(right)} \tag{2.8.5}$$

where w is the number of samples reaching node j, G is the impurity value of node j and left(j) and right(j) indicate the child nodes of node j

2.9 Data retrieval and statistical comparisons

In our research, we use the POPSTAR PROJECT code, popstar.py available at

$$https://gbabbitt.github.io/POPSTAR-signaling-in-music/$$

to generate dynamic ternary maps and make statistical comparisons of CES and attention signaling (i.e. CES stability) across musical and non-musical sounds, musical genres, and several artist's discography, and selected examples of human, primate, bird, and anuran(frog) vocalizations. All animal vocalizations were extracted from the iNaturalist dataset() audio repository iNat sounds available at

$$https://github.com/visipedia/inat_sounds/$$

All iNaturalist audio files were manually curated for consistency of song type and cleanness of audio, removing any with background sounds of humans, other animals and/or excessive wind noise. Movie output for specific examples of the dynamic ternary plots and dynamic Chernoff faces can be observed at the website above. We make specific comparisons of attention signaling in expert vs. novice vocalists in both humans and Yellow Canary (owners often post the first songs of their pet canaries on YouTube after training them with videos of 'exceptional' adult birds singing). We compare attention signaling in piano concertos, where the ability of soloists is key, to ambient piano works of the Impressionist era, where the music emphasizes the environment of the listener rather than the skill of the soloist. We also compare live vs. studio performances of the popular artist Bjork Guomundsdottir, to see if direct feedback of the audience affects the resulting attention signaling of the artist.

In cases where audio was extracted from YouTube we avoided videos with excessive background noise and performed batch audio extraction and batch loudness normalization using our script **(yt-dlp batch.py)** available in the POPSTAR Project Github repository. All ternary plot movie output (with sound) and statistical results are available at

$$https://doi.org/10.5281/zenodo.17218769$$

.

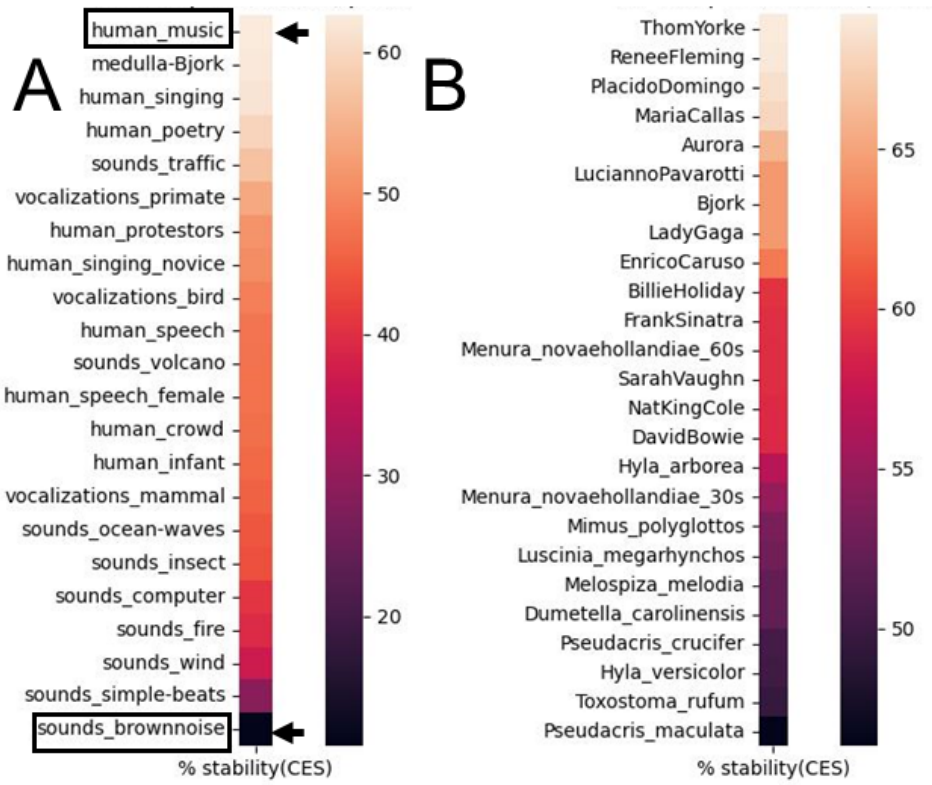

Figure 4: **Relative stability of different sound trajectories in the space of control, energy, and surprise (CES)** for (A) a selection of natural sounds ranging from noise, non-biological sounds in nature, and various animal (and human) vocalizations including music. (B) We also compare a selection of famous human musicians of various genres (Pop, jazz and opera) to chorus frogs (genus Hyla and Pseudacris) and several birds with highly notable songs (i.e. Song Sparrow (Melospiza melodia), Lyrebird (Menura novaehollandiae), Nightingale (Luscinia megarhynchos), and several species of Mimidae (i.e. mimics) including Northern Mockingbird (Mimus polyglottos), Brown Thrasher (Toxostoma rufum) and Catbird (Dumetella carolinensis). Note that (A) noise has 0 percent stability in the space of CES while music has the most (typically ¿ 50 percent) which is roughly 10 percent higher than spoken words.

3 Results

3.1 Validation of musicality of sound via stability of control, energy, and surprise

In comparisons of stability of audio trajectory in the ternary space of CES (i.e. signaling of individual performance quality) across a large variety of musical and non-musical sounds (Figure 4), we find that music has the highest stability in CES (¿60-70 percent stable) while samples of brown noise have near 0 percent stability (Figure 4A). Therefore, our indicator function for stability in CES is a valid measure of the human perception of musicality in sound. We find that famous human musicians in the genres of jazz, opera, and popular music have significantly different and higher CES trajectory stability than other animal vocalizations (H = 416.945, p ¡0.001) with some birds with very complex song such as the Lyrebird of Australia (Menura novaehollandiae) and the European Nightingale (Luscinia megarhynchos) having CES stability that is comparable or even surpassing some human singers (Figure 4B). We also find that the career discography of a single pop musician (Bjork Guomundsdottir) also exhibits significant variation in CES stability (H = 85.492, p¡ 0.001) (Figure S2) indicating that our metric of musicality is not only valid regarding our perceptions of song, but has potentially interesting variation within individual performances as well.

3.2 Comparison of control, energy, and surprise in human vs. animal song

A broad taxonomic comparison of individuality in song across humans, birds and chorus frogs across each dimension of CES (Figure 5) reveals that while professional human singers have less variation in control and

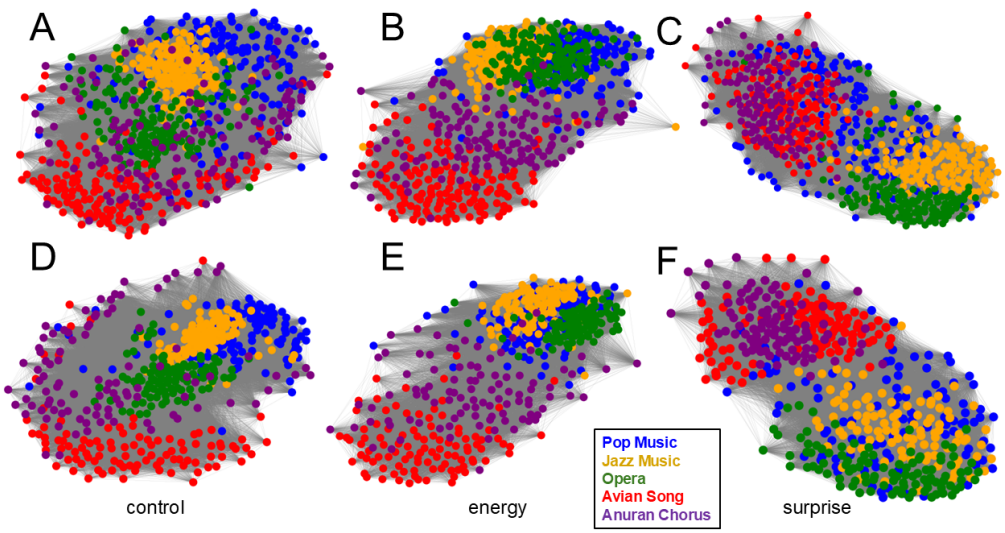

Figure 5: **Adjacency networks of individual songs of frogs, birds, and humans mapped within the three dimensions of control, energy and surprise (CES)**. In A-C the human singers are female and in D-F they are male. In the animals, all singers are assumed to be male. Functional data analysis is applied to generate a b-spline interpolation of the time series of an individual song in the specified dimension (either C, E or S). In each network, each node represents an individual song and each edge represents an L2 distance between the functions for two songs. The network representations utilize Kamada-Kawai topology with edges formed using all L2 distances below the median distance observed (i.e. closest 50 percent of the nodes attract and the farthest 50 percent of the nodes repulse).

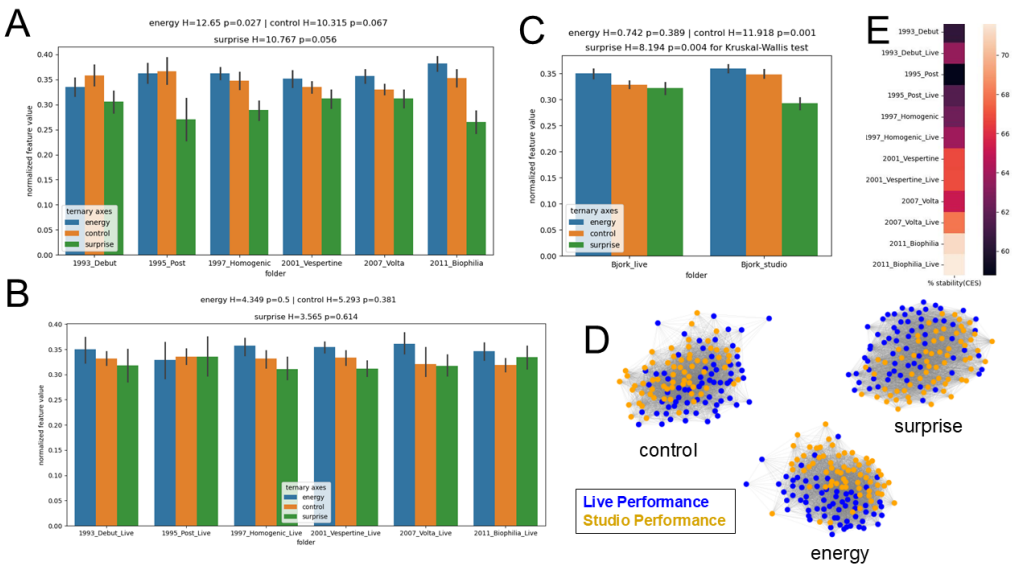

Figure 6: **Comparison of stability in musical control, energy and surprise (CES) in live vs. studio performances while controlling for individual song track.** Here we examine the Icelandic pop star Bjork's six studio LPs compared to their corresponding live tour LPs. We find that (A) the studio LP's exhibit significant or borderline significant differences in CES while (B) the corresponding live LP's do not. This suggests that live performances feedback may constrain Bjork's performance to a more consistent level of relative CES. (C) Overall, the live performances show significantly less control and more surprise than the corresponding studio song tracks, showing a trend similar to that distinguishing novice from expert performance. This suggests that fitness signaling via CES is less manipulatable on stage than in studio. (D) Random forest classification indicates with moderate accuracy that differences between Bjork's live and studio performance involves all nine features of CES to large extent. (E) While each studio project and its accompanying tour vary significantly in CES stability, the live performances seem to be slightly more stable in CES, again suggesting some creative constraints or added fitness requirements when performing live.

energy than these two animal groups, the human genres are quite discernable from each other with opera being closer to animal vocalization in terms of energy and control than other musical genres. We also see that opera is closer to anuran than avian song (Figure 5A-B, 5D-E) perhaps due to more similar vocal architecture between mammalian and anuran groups, both having a larynx as opposed to the avian syrinx. In anurans the larynx is used within a closed cycle air space moving air repeatedly over the vocal chords between a storage sac and the lungs. In mammals, air is inhaled and exhaled over the vocal chords. In birds, this air cycling is similar to mammals, but the avian syrinx sits deeper in the lungs with vocal chords on each side that can often be controlled independently allowing for more complexity of sound. In the dimension of surprise (Figure 5C, 5F), we still see differentiation between human singers of different genres and our two main animal groups, but the variability between human singers is more pronounced indicating more human variation in song complexity than with control and energy. This is perhaps not surprising given the language components of human song as well as the falsifiability of energy and control signaling in human music via the use of instrumentation and other technology and professional training. We do not observe much difference in these trend based upon whether the human singers are female (Figure 5A-C) or male (Figure 5D-F), however we do note that female human singers have somewhat more individual variation in song control and energy, while male human singers have more variation in song complexity or surprise. These trends are all still apparent when we include a limited number of other ape songs (i.e. Hylobates – gibbons) in the analysis (Figure S3) and we note that primate song is also generally closer to humans as expected. Within chorus frog and avian species (Figure S4 and S5) we observe less individuality than across major taxonomic groups, but differences are still apparent especially with regards to control and surprise in anurans and energy and surprise in birds. Within human musical genres (Figure S6) opera and jazz tend to be more distinct emphasizing surprise (jazz) and energy+surprise (opera) while popular music emphasizes energy+control, with popular music having wider variation in CES distances between individual songs. This is perhaps not surprising given that popular music borrows from many other genres and thus may appear less distinct in our network representations. Within each musical genre (Figures S7-S9), while significant differences in the levels of CES are evident, we find little difference in the network distances (i.e. individuality of each song).

3.3 The role of audience feedback on control, energy, and surprise

We analyze the career of the pop musician Bjork who is very interesting from the context of our research because of the large range of 'controlled' experimentation that she has recorded. Across a nearly 50-year career and 6 World tours, she is well known to be an excellent live vocalist performer and as well a highly experimental composer touching upon many unique genres (pop, punk, trip-hop, jazz, folk, electronica, ambient, and avant-garde). Her vocal exploration of the boundary between her own human speech and her more conventional music is also important for our study, particularly her fifth studio album Medulla, which attempts to produce fully modern music using only the human voice alone (i.e. no instrumentation) (Figure S10). After each of her world tours she released a live album corresponding with the same tracks as her studio work prior to touring, thus providing a nicely controlled experiment where we can examine the effect of the feedback of live audiences on the different components of attention signaling in her performances (Figure 6). While we find no significant differences between live and studio performance regarding the stability of CES trajectories, we do see different levels of CES evident across 6 studio LPs recorded over 18 years (Figure 6A) that are absent in the corresponding live performance LPs (Figure 6B). This indicates more consistent combination of control, energy and surprise when facing a live audience. Combined comparison of live vs studio LP's (Figure 6C) finds significantly increased control (H=11.918, p=0.001) and decreased surprise (H= 8.194, p=0.004) again indicative that audience feedback greatly effects attention signaling in a musical performer. We also note that the differences between live vs. studio performance is not reliant on any single feature of CES (Figure 6D) and is quite apparent in our network analyses of individuality of songs (Figure 6E). Lastly, we compared Bjork's attention signaling across normal speech (i.e. TV interviews), her Medulla LP, and her Greatest Hits LP (Figure S10) and find significant differences in all levels of CES (control: H=39.034, p¡0.001 — energy: H=15.014, p=0.002 — surprise: H=26.961, p¡0.001) and stability of CES (H=16.033,p=0.003) with clear evidence from the Medulla LP indicating that the attention signaling present in the audio is not dependent upon electronic instrumentation, but can be the product of human vocalization alone.

3.4 Comparison of control, energy, and surprise in expert vs. novice singers

In small comparisons of expert and novice human singers (Figure 7A-C) reveal that expert musicians exhibit borderline significantly more CES stability (H=4.339, p=0.037), more control (H= 3.429, p=0.064) and less surprise (H=5.357, p=0.021) than novice musicians (Figure 7A-B). A machine learning random forest classifier identifies expert vs. novice human singers with 99 percent accuracy using mostly the features of Lempel-Ziv (substring) complexity and note variability and amplitude modulation to a lesser extent (Figure 7C). In a similar analysis comparing adult Yellow Canary to immature birds singing their first songs (Figure 7D-F,

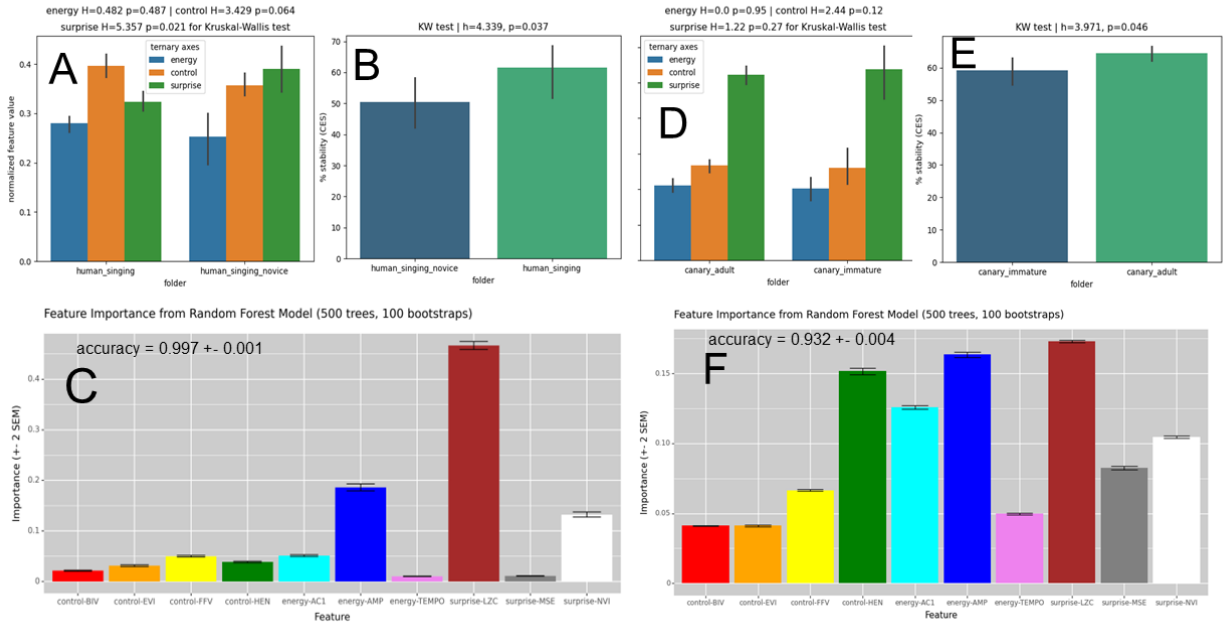

Figure 7: **Comparison of stability in musical control, energy and surprise (CES) in expert vs. novice singers (A-C) and adult vs. immature birds (D-F).** (A) Expert singers exhibit significantly more control and less surprise or complexity than novice singers. (B) Expert singers also have significantly more stability in song trajectory in the space of CES than novice singers. (C) Random forest classification indicates with high accuracy that differences between experts and novices are largely driven by Lempel-Ziv (substring) complexity, amplitude modulation, and note variability in this study.

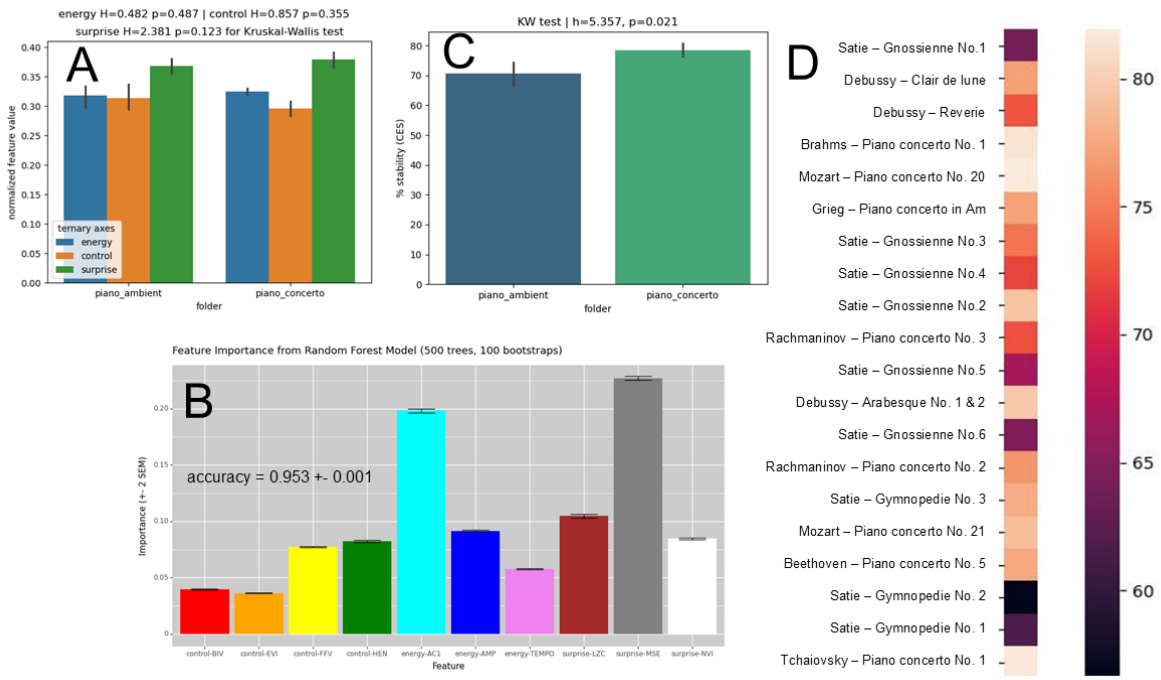

Figure 8: **Comparison of stability in musical control, energy and surprise (CES) in piano concertos vs. ambient piano pieces** (i.e. furniture music of Erik Satie and Claude Debussy). (A) While there are no significant differences in the levels of CES between the two genres, (B) the piano concertos have significantly more stability in song trajectory in the space of CES than ambient music. (C) Random forest classification indicates with high accuracy that differences between the two genres are largely driven by multiscale entropy and reverberation.

Figure S11) we find no differences in overall levels of control, energy or surprise, but we do find that adults have significantly more musicality indicated by the stability in CES (H=3.971, p=0.046)(Figure 7E). Machine learning classification can distinguish adult from immature song with 93 percent accuracy using a wide variety of features of CES (Figure 7F).

3.5 Comparison of control, energy, and surprise in piano concerto vs. ambient works for piano

As not all music places equal emphasis on solo performance (i.e. individual performance quality), we compared a small sample of the most difficult piano concertos for soloists to perform (Rachmaninov Nos. 2-3, Beethoven No. 5, Mozart Nos. 20-21, Brahms No. 1, Tchaikovsky No. 1, Grieg Am concerto) with a small sample of ambient piano pieces by the composers Satie (Gymnopedie 1-3, Gnossienne 1-6), and Debussy (Arabesque No. 1-2, Clair de lune, Reverie) (Figure 8). Here we find that while the levels of control energy and surprise are similar (Figure 8A), the stability in CES is significantly higher in the concertos designed to demonstrate soloist virtuosity than in the ambient pieces which emphasize music as a space which the listener occupies (H= 5.357, p=0.021) (Figure 8B). Random forest classification can discriminate concertos from ambient works on piano with 95 percent accuracy indicating that multiscale entropy and reverberation are the most important features for learning the difference here (Figure 8C).

4 Discussion

We have proposed a novel multivariate measure of vocal performance (i.e. individual quality) in sound defined directly at the phenotypic level through motor control, energy, and surprise/cognitive complexity. We propose that these three sets of acoustic features (i.e. control, energy and surprise or CES) combine to form a mathematical calculus of attention that individuals can manipulate to stimulate attentive behavior in an observer (or larger audience). We further propose that the quality of attention signaling can be indicated by the level of non-randomness in the trajectories of sounds in the ternary space of CES. We demonstrate that attention signaling defined in this way is common across many forms of human and animal vocalization and also suggest that it is associated with visual displays as well. We investigate auditory attention signaling in human and animal vocalizations compared to other non-biological sounds and demonstrate that stability in CES corresponds to our human perception of musicality in both sounds we culturally define as music, as well as animal vocalizations we perceive as more musical than others (i.e. songbirds, gibbons, and chorus frogs). Overall, we find that professional human musicians have less variance than other animal 'song' in their song signatures (i.e. distances derived from functional data analysis of time series trajectories) of control and energy, but not in their signatures of complexity or surprise. This is probably due to the self-selected talent and subsequent training that human singers go through. Unlike chorus frogs, gibbons and songbirds, not all humans have to learn to sing to succeed in life. Our distance network analysis of song signatures also shows that the genres of opera and jazz are more distinct and well defined than popular music, with opera being closer to that of other animals (especially gibbons and chorus frogs, rather than songbirds). This result likely reflects the functional similarity of vocal architecture (i.e. frogs and mammals both have a single larynx while bird use a more complex structured syrinx). Popular music also uses more technological means of achieving near perfect control and energy through amplification, fretboards and keyboards, autotuning, and electronic drums and synthesizers. Jazz relies on less of this and opera even less so. This is also consistent with the observation that we find less individuality of song signature within popular musicians than within other genres (jazz, opera) or within the boundaries of broader taxonomic groups. We also demonstrate that expert vocalists, in both humans and at least one species of bird (Yellow Canary), have higher quality levels of attention signaling than novice individuals as evidenced by less randomness of song trajectory in the space of CES. This finding is supported by a recent study showing that daily vocal exercise is as important for birds as it is for human singers [1] and therefore is likely to be an honest signal of fitness through performance as well. We also find attention signaling is stronger in musical performances that emphasize the ability of soloists over the perceptions or moods of the listener (i.e. piano concertos vs. ambient piano music). Lastly, we demonstrate that in the career of one famous popular performer (Bjork), live audience feedback seems to constrain the diversity of her exploration of the space of CES when compared to her work in the studio. This suggests there is something consistent to attention signaling that requires audience presence and helps to explain previous observations in neuroscience that show that people are often more stimulated by live vs. recorded music [85]. In conclusion, we find striking mathematical similarity among animal 'song' vocalizations and human musical sounds regarding the stability of CES, along with some well-defined differences across evolutionary distant taxa that evolved singing behavior independently (i.e. anurans, birds, primates) We believe this validates our mathematical definition of fitness in animal vocal performance and human music suggesting further investigations at the intersection of aesthetics and evolutionary biology could prove valuable. In future studies, we plan to investigate the utility of our soft-

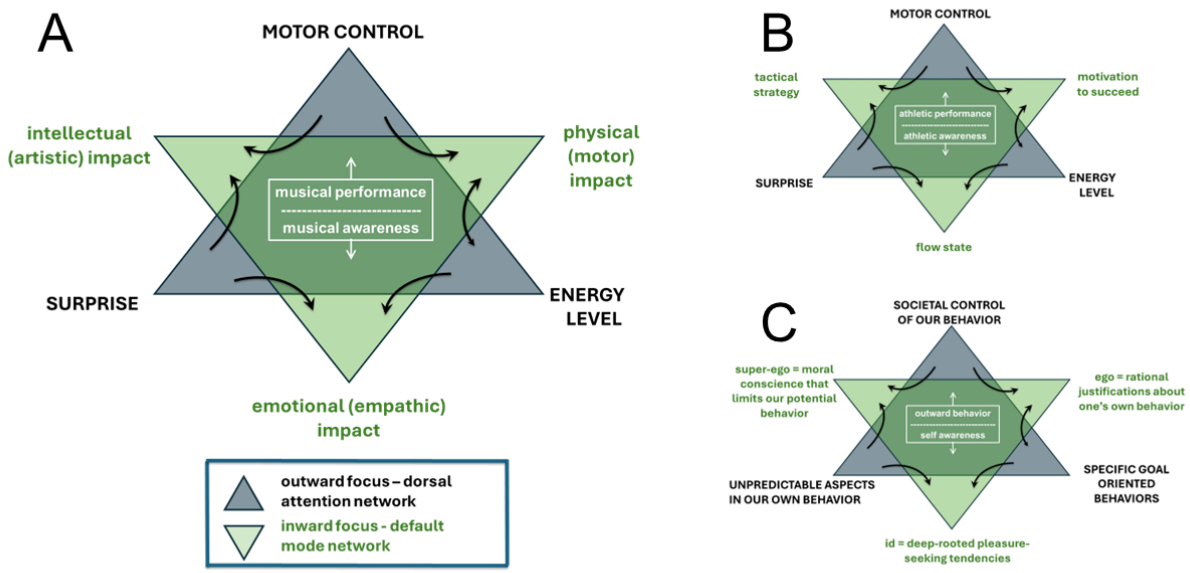

Figure 9: **Theoretical relationships between outward and inward attention or focus with respect to (A) music and dance, (B) sports contests, and (C) human behavior more generally.** We theorize that context-dependent features of control, energy, and surprise (gray triangle) act generally to stimulate the dorsal attention network of the brain, while inward attention, likely mediated by the default mode network of the brain, stimulates an inverse geometry of self-awareness (green triangle) specific to that activity. (A) In music and dance, inward self-awareness is balanced between physical, emotional and intellectual (artistic) impacts on the mind. (B) In the context of a sporting event, inward self-awareness is balanced between physical motivation to succeed, emotional flow state and intellectual (tactical) strategy in the mind. (C) In a broader context of all human behavior, features of societal control, individual goal-oriented behavior, and unpredictability in our own and others behavior (i.e. a generalized form of control, energy, and surprise) create an inward geometry of self-awareness that replicates Freud's concepts of id (i.e. pleasure seeking), ego (i.e. rationalization of one's own behavior) and super-ego (i.e. moral conscience or judgement of one's own behavior in the context of societal rules).

ware and methodology for predicting historic and current trends in the popularity of music (e.g. peak Billboard magazine chart positions, YouTube views, or Spotify downloads) and in training amateur musicians and/or designing new kinds of instrumental or electronic sounds that stimulate the attention of audiences in interesting ways.

Musical sound stimulates the neurophysiology of the human brain more than any other neurological stimulus [74] and may yet hold the answer to some of the yet unanswered questions around the major evolutionary expansion events in hominid brain size[19][28]. We intuitively understand that music affects us greatly in both an external physical sense via motion and attention as well as an internal cognitive sense via emotion and intellect. The geometry of the calculus of attention provides us with a simple and yet novel way to envision this complex cognitive duality. If we view the ternary plot representing the external acoustic features of musical control, energy and surprise as the foundation for an complementary triangular internal representation of how music might affect our internal awareness, we can build a six-sided star (Figure 9A) that balances (A) the physical impact of music on our motor cortex between the vertices for motor, control and energy, the (B) emotional impact of music between the vertices of energy and surprise (i.e. complexity), and the (C) intellectual impact of music between the vertices of control and surprise. If we place unbalanced emphasis on the three vertices of control, energy, and surprise, either singularly or in pairs, we can seemingly and yet surprisingly describe the general feelings invoked by hearing different genres of music, simple musical scales, and spoken words (Figure 10). In this internal-external cognitive geometry, the upward triangle represents the outward focused activity of the ancestral dorsal and ventral attention networks [88] while the downward triangle represents the inward focus of the more recently evolved default mode network [53]. In the context of music, this model provides a simple demarcation of the difference between outward musical performance and inward musical awareness (Figure 9A). What is also very interesting is that if we consider song to be one of many other possible types

of behavioral display, we can apply this geometry much more broadly. For example, like musical performance, any sporting contest observed by an audience might be also be considered an external display of motor control of the body, energy on the field of play, and elements of surprise that relies upon an internal awareness of a physical motivation to succeed, an emotional flow state of body-mind, and an intellectual manipulation of tactics within the rules of the game. (Figure 9B). Most interestingly, this geometry can be broadened even more to include general human behavioral psychology regarding external aspects of societal control of behavior (i.e. control), specific goal-oriented motivations of individuals (i.e. energy), and unpredictability in peoples behavior (i.e. surprise), to reproduce the classic internal Freudian architecture of the id, ego, and super-ego (Figure 9C). While modern neuroscience has largely rejected these Freudian terminologies because of their lack of localization in the brain [33] [56] and studies citing Freudian concepts are in decline [95], interestingly, many clinical psychologists still rely upon them in therapeutic interventions [39] where classic psychoanalysis has evolved into a more modern psychodynamic therapy focused upon patient outcome [58]. Our geometric model hypothesizes the existence many potential common features of human and animal behavioral display that manifest from an ancient common calculus of attention in the animal brain. This might underpin our internal self-awareness of not only music, but many other combinations of physical, emotional, and intellectual awareness in response to many other human social behaviors. Perhaps future work within the fields of behavioral ecology and neurophysiology might be able to further confirm or reject the ternary architecture of our mathematical model. Experimental science in general, and neuroscience more specifically, has often struggled with its inability to understand complex emergent phenomena born of simple system dynamics [51]. However, the mathematics of luminaries like Robert May, Ed Lorenz, Benoit Mandelbrot and Mitchell Feigenbaum have demonstrated that complex emergent behavior originating from simple feedback mechanisms are universal in nature, even if not well understood[50][44][21][47]. While our conclusions here might seem tenuous to laboratory scientists trained in reductionist experimental methods based in molecular biology, we would argue that some attempts to understand complex emergent phenomena such as those involved in art and aesthetics can benefit from from some basic theoretical grounding in mathematics when it is combined with data driven computational methods like machine learning. We have attempted to demonstrate this in our work presented here.

In conclusion, our mathematical modeling and comparative analyses of the behavior of human vs. animal 'song' are highly suggestive as to how our attentive behaviors and perhaps even some aspects of our self-awareness were probably molded by ancient evolutionary history. The Ediacaran-Cambrian transition 541 million years ago created selective pressures underlying a singular calculus of attention, thereby limiting the dimensionality of our minds in ways that computer algorithms behind machine learning and artificial intelligence are not. Yet, when attention has been implemented in the transformer neural network architecture of modern generative artificial intelligence (GAI), it has vastly improved its capability to mimic and interact with our own neural architecture [86]. The concept of a calculus of attention has broad implications beyond popular music, especially in an emerging internet economy where people's attention is increasingly monetized for exploitation by GAI [32][87]. We suggest that further investigating the underlying mathematical patterns and fundamental biology of attentive behavior can potentially help guard our society against GAI applications aimed at overstimulating the cognitive mechanisms that control our attention. Humans not only direct their attention towards popular music, but also popular athletes, public speakers, teachers, preachers, politicians, and newscasters, all of whom might also be exploiting behavioral signals of control, energy, and surprise designed by evolutionary biology to hold our attention. This can potentially affect us in ways that can prove either productive or unproductive to our lives depending upon the context with which we engage in internet outlets for modern media. As the attention and default mode networks of our brains are so strongly anti-correlated in their activity[53], we can perhaps too easily lose our sense of self when we are heavily immersed in stimuli that constantly demand our immediate attention.

References

[1] Iris Adam et al. "Daily vocal exercise is necessary for peak performance singing in a songbird". In: *Nature Communications* 14.1 (Dec. 12, 2023). Publisher: Nature Publishing Group, p. 7787. ISSN: 2041-1723. DOI: 10.1038/s41467-023-43592-6. URL: https://www.nature.com/articles/s41467-023-43592-6 (visited on 09/20/2025).

[2] Christoph Adami. "Adaptive walks on the fitness landscape of music". In: *Proceedings of the National Academy of Sciences* 109.30 (July 24, 2012). Publisher: Proceedings of the National Academy of Sciences, pp. 11898–11899. DOI: 10.1073/pnas.1209301109. URL: https://www.pnas.org/doi/10.1073/pnas.1209301109 (visited on 09/25/2025).

[3] Hussain-Abdulah Arjmand et al. "Emotional Responses to Music: Shifts in Frontal Brain Asymmetry Mark Periods of Musical Change". In: *Frontiers in Psychology* 8 (Dec. 4, 2017). Publisher: Frontiers. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2017.02044. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.02044/full (visited on 01/21/2025).

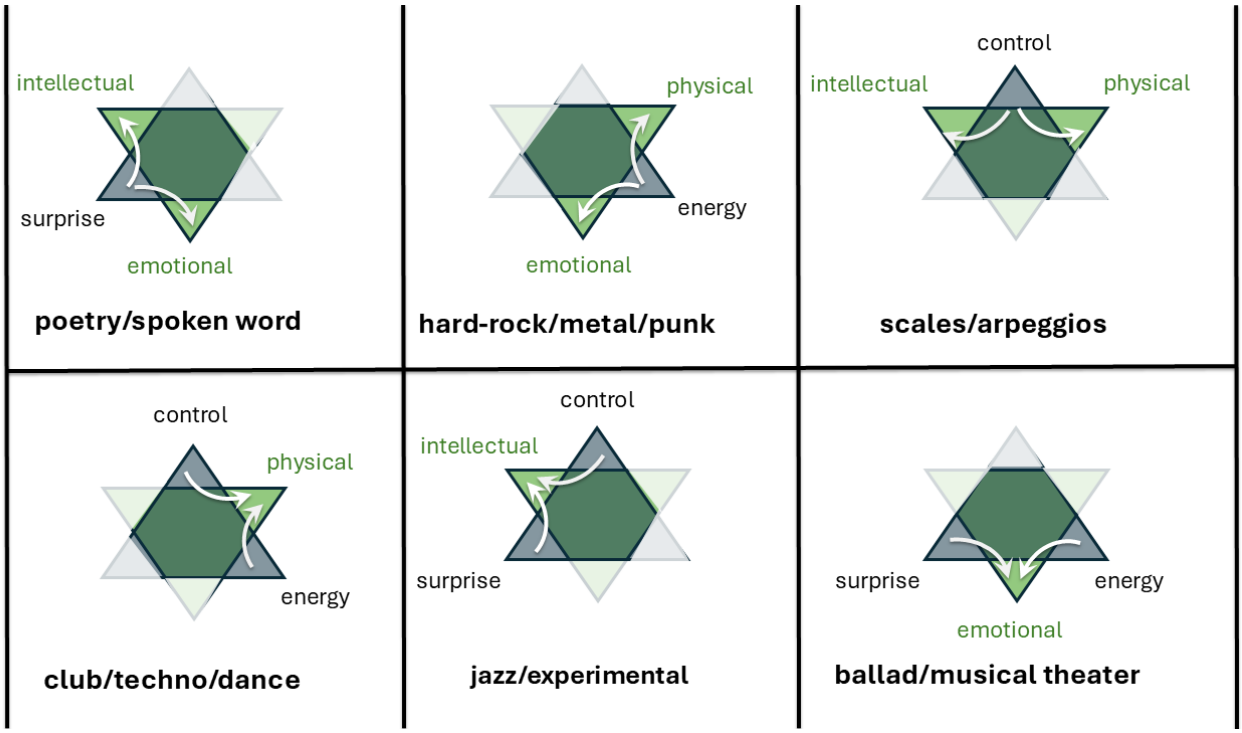

Figure 10: **The effects of unbalanced geometry in the calculus of attention (gray) can explain the internal cognitive effects (green) of various musical genres.** For example, emphasis on just energy level in music with de-emphasis on novelty/surprise and fine motor control can result in the strong feelings of physical plus emotional impact that we often tend to associate with heavy metal/punk rock. Similarly, a singular emphasis on motor control produces a consequent lack of emotion when we listen to scales or arpeggios, and a singular focus on lyrical novelty, without emphasis energy or motor control explains the cognitive differences we perceive when comparing poetry/spoken word versus music itself. When adjacent pairs of fitness signals are emphasized in combination, it has the effect of focusing on the cognitive impact in a single area. For example, the primarily physical impact we hear in dance music is created by a combination of balanced emphasis on energy level and motor control, with little novelty or surprise. Similarly, the intellectual stimulation of modern jazz is created by the balanced combination of improvisation (surprise) with motor control. And the enhanced emotional impact of song ballads and musical theatre are caused by the balance of energy level and unexpected features in the lyrical or musical elements. Note: not pictured here, popular music would seem to require musicians to maintain balance between all three features of the model (energy, control and surprise).

[4] Gregory A. Babbitt, Lin Wang, and Ernest P. Fokoue. *Musicality in protein interaction dynamics informs the multi-scale evolution of prosocial behavior*. Pages: 2025.03.13.643054 Section: New Results. Mar. 27, 2025. DOI: 10.1101/2025.03.13.643054. URL: https://www.biorxiv.org/content/10.1101/2025.03.13.643054v2 (visited on 04/01/2025).

[5] Gregory A. Babbitt et al. "ATOMDANCE: Kernel-based denoising and choreographic analysis for protein dynamic comparison". In: *Biophysical Journal* 123.17 (Sept. 3, 2024), pp. 2705–2715. ISSN: 1542-0086. DOI: 10.1016/j.bpj.2024.03.024.

[6] Fiona Backhouse et al. "Higher-order sequences of vocal mimicry performed by male Albert's lyrebirds are socially transmitted and enhance acoustic contrast". In: *Proceedings of the Royal Society B: Biological Sciences* 289.1970 (Mar. 9, 2022), p. 20212498. ISSN: 0962-8452. DOI: 10.1098/rspb.2021.2498. URL: https://doi.org/10.1098/rspb.2021.2498 (visited on 04/13/2026).

[7] Luis F. Baptista and Robin A. Keister. "Why birdsong is sometimes like music". In: *Perspectives in Biology and Medicine* 48.3 (2005), pp. 426–443. ISSN: 0031-5982. DOI: 10.1353/pbm.2005.0066.

[8] Antonia S. Becker et al. "Musician presence and its effects on physiological and psychological well-being in live versus livestreamed concerts". In: *Scientific Reports* 16.1 (Feb. 9, 2026). Publisher: Nature Publishing Group, p. 7889. ISSN: 2045-2322. DOI: 10.1038/s41598-026-38194-3. URL: https://www.nature.com/articles/s41598-026-38194-3 (visited on 04/16/2026).

[9] D. Bedoya et al. "Even violins can cry: specifically vocal emotional behaviours also drive the perception of emotions in non-vocal music". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1840 (Nov. 2021). Publisher: Royal Society, p. 20200396. DOI: 10.1098/rstb.2020.0396. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0396 (visited on 01/21/2025).

[10] Giacomo Bignardi et al. "Twin modelling reveals partly distinct genetic pathways to music enjoyment". In: *Nature Communications* 16.1 (Mar. 25, 2025). Publisher: Nature Publishing Group, p. 2904. ISSN: 2041-1723. DOI: 10.1038/s41467-025-58123-8. URL: https://www.nature.com/articles/s41467-025-58123-8 (visited on 09/20/2025).

[11] Steven Brown et al. "Correlations in the population structure of music, genes and language". In: *Proceedings of the Royal Society B: Biological Sciences* 281.1774 (Jan. 7, 2014). Publisher: Royal Society, p. 20132072. DOI: 10.1098/rspb.2013.2072. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2013.2072 (visited on 01/19/2025).

[12] Cristina R. Ceja and Steven L. Franconeri. "Difficulty limits of visual mental imagery". In: *Cognition* 236 (July 1, 2023), p. 105436. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2023.105436. URL: https://www.sciencedirect.com/science/article/pii/S0010027723000707 (visited on 09/25/2025).

[13] Herman Chernoff. "The Use of Faces to Represent Points in k-Dimensional Space Graphically". In: *Journal of the American Statistical Association* 68.342 (June 1, 1973). Publisher: ASA Website _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1973.10482434, pp. 361–368. ISSN: 0162-1459. DOI: 10.1080/01621459.1973.10482434. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1973.10482434 (visited on 04/01/2025).

[14] Alain de Cheveigné and Hideki Kawahara. "YIN, a fundamental frequency estimator for speech and music". In: *The Journal of the Acoustical Society of America* 111.4 (Apr. 2002), pp. 1917–1930. ISSN: 0001-4966. DOI: 10.1121/1.1458024.

[15] Roslyn Dakin et al. "Biomechanics of the Peacock's Display: How Feather Structure and Resonance Influence Multimodal Signaling". In: *PLoS ONE* 11.4 (Apr. 27, 2016), e0152759. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0152759. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4847759/ (visited on 04/16/2026).

[16] Simon A. F. Darroch et al. "Causes and consequences of end-Ediacaran extinction: An update". In: *Cambridge Prisms: Extinction* 1 (Jan. 2023), e15. ISSN: 2755-0958. DOI: 10.1017/ext.2023.12. URL: https://www.cambridge.org/core/journals/cambridge-prisms-extinction/article/causes-and-consequences-of-endediacaran-extinction-an-update/5D8669E8599A9065C0B2F35A0BF2EA8A (visited on 09/25/2025).

[17] Chiara De Gregorio et al. "Isochrony as ancestral condition to call and song in a primate". In: *Annals of the New York Academy of Sciences* 1537.1 (July 2024), pp. 41–50. ISSN: 1749-6632. DOI: 10.1111/nyas.15151.

[18] Emily L. Doolittle et al. "Overtone-based pitch selection in hermit thrush song: unexpected convergence with scale construction in human music". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.46 (Nov. 18, 2014), pp. 16616–16621. ISSN: 1091-6490. DOI: 10.1073/pnas.1406023111.

[19] Andrew Du et al. "Pattern and process in hominin brain size evolution are scale-dependent". In: *Proceedings of the Royal Society B: Biological Sciences* 285.1873 (Feb. 21, 2018), p. 20172738. ISSN: 0962-8452. DOI: 10.1098/rspb.2017.2738. URL: https://doi.org/10.1098/rspb.2017.2738 (visited on 04/15/2026).

[20] Daniel P. W. Ellis. "Beat Tracking by Dynamic Programming". In: *Journal of New Music Research* 36.1 (Mar. 1, 2007). Publisher: Routledge _eprint: https://doi.org/10.1080/09298210701653344, pp. 51–60. ISSN: 0929-8215. DOI: 10.1080/09298210701653344. URL: https://doi.org/10.1080/09298210701653344 (visited on 09/07/2025).

[21] Mitchell J. Feigenbaum. "Universal behavior in nonlinear systems". In: *Physica D Nonlinear Phenomena* 7 (May 1, 1983). Publisher: Elsevier ADS Bibcode: 1983PhyD....7...16F, pp. 16–39. ISSN: 0167-2789. DOI: 10.1016/0167-2789(83)90112-4. URL: https://ui.adsabs.harvard.edu/abs/1983PhyD....7...16F (visited on 04/16/2026).

[22] W. Tecumseh Fitch. "Four principles of bio-musicology". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370.1664 (Mar. 19, 2015), p. 20140091. ISSN: 1471-2970. DOI: 10.1098/rstb.2014.0091.

[23] W. Tecumseh Fitch. "The biology and evolution of music: a comparative perspective". In: *Cognition* 100.1 (May 2006), pp. 173–215. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2005.11.009.

[24] Matthew W. Flood and Bernd Grimm. "EntropyHub: An open-source toolkit for entropic time series analysis". In: *PloS One* 16.11 (2021), e0259448. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0259448.

[25] Giovanni Formilan, Gino Cattani, and Simone Ferriani. "Trajectories of Consecration: Signature Style and the Pace of Category Spanning". In: *Organizing Creativity in the Innovation Journey*. Ed. by Elke Schuessler, Patrick Cohendet, and Silviya Svejenova. Vol. 75. Emerald Publishing Limited, Sept. 17, 2021, p. 0. ISBN: 978-1-83982-875-1. DOI: 10.1108/S0733-558X20210000075005. URL: https://doi.org/10.1108/S0733-558X20210000075005 (visited on 04/16/2026).

[26] Andy Gardner. "The Price equation". In: *Current Biology* 18.5 (Mar. 11, 2008). Publisher: Elsevier, R198–R202. ISSN: 0960-9822. DOI: 10.1016/j.cub.2008.01.005. URL: https://www.cell.com/current-biology/abstract/S0960-9822(08)00008-0 (visited on 09/25/2025).

[27] Clément M. Garin et al. "An evolutionary gap in primate default mode network organization". In: *Cell Reports* 39.2 (Apr. 12, 2022), p. 110669. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2022.110669. URL: https://www.sciencedirect.com/science/article/pii/S2211124722004211 (visited on 04/18/2026).

[28] Mauricio González-Forero and Aida Gómez-Robles. "Why did the human brain size evolve? A way forward". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 380.1929 (June 26, 2025), p. 20240114. ISSN: 0962-8436. DOI: 10.1098/rstb.2024.0114. URL: https://doi.org/10.1098/rstb.2024.0114 (visited on 04/15/2026).

[29] Morgan L. Gustison and Thore J. Bergman. "Vocal complexity influences female responses to gelada male calls". In: *Scientific Reports* 6.1 (Jan. 21, 2016). Publisher: Nature Publishing Group, p. 19680. ISSN: 2045-2322. DOI: 10.1038/srep19680. URL: https://www.nature.com/articles/srep19680 (visited on 10/07/2025).

[30] Morgan L. Gustison, Aliza le Roux, and Thore J. Bergman. "Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1597 (July 5, 2012), pp. 1847–1859. ISSN: 0962-8436. DOI: 10.1098/rstb.2011.0218. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3367700/ (visited on 10/07/2025).

[31] Eleanor E. Harding et al. "Musical neurodynamics". In: *Nature Reviews Neuroscience* 26.5 (May 2025). Publisher: Nature Publishing Group, pp. 293–307. ISSN: 1471-0048. DOI: 10.1038/s41583-025-00915-4. URL: https://www.nature.com/articles/s41583-025-00915-4 (visited on 09/20/2025).

[32] Maxi Heitmayer. "The Second Wave of Attention Economics. Attention as a Universal Symbolic Currency on Social Media and beyond". In: *Interacting with Computers* 37.1 (Jan. 1, 2025), pp. 18–29. ISSN: 1873-7951. DOI: 10.1093/iwc/iwae035. URL: https://doi.org/10.1093/iwc/iwae035 (visited on 09/25/2025).

[33] Marion Hendrickx, Quentin Tousart, and Emmanuel Drouin. "If Freud had known..." In: *Frontiers in Psychology* 16 (May 20, 2025), p. 1556279. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2025.1556279. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12129887/ (visited on 04/16/2026).

[34] Henkjan Honing. "On the biological basis of musicality". In: *Annals of the New York Academy of Sciences* (Mar. 15, 2018). ISSN: 1749-6632. DOI: 10.1111/nyas.13638.

[35] Henkjan Honing. "Without it no music: beat induction as a fundamental musical trait". In: *Annals of the New York Academy of Sciences* 1252.1 (2012). _eprint: https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.2011.06402.x, pp. 85–91. ISSN: 1749-6632. DOI: 10.1111/j.1749-6632.2011.06402.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2011.06402.x (visited on 04/16/2026).

[36] David Huron. "Is Music an Evolutionary Adaptation?" In: *Annals of the New York Academy of Sciences* 930.1 (2001). _eprint: https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.2001.tb05724.x, pp. 43–61. ISSN: 1749-6632. DOI: 10.1111/j.1749-6632.2001.tb05724.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2001.tb05724.x (visited on 09/25/2025).

[37] Patricia Izbicki, Andrew Zaman, and Elizabeth L. Stegemöller. "Music Form but Not Music Experience Modulates Motor Cortical Activity in Response to Novel Music". In: *Frontiers in Human Neuroscience* 14 (Apr. 16, 2020), p. 127. ISSN: 1662-5161. DOI: 10.3389/fnhum.2020.00127. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7179827/ (visited on 01/21/2025).

[38] Nori Jacoby et al. "Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries". In: *Nature Human Behaviour* 8.5 (May 2024). Publisher: Nature Publishing Group, pp. 846–877. ISSN: 2397-3374. DOI: 10.1038/s41562-023-01800-9. URL: https://www.nature.com/articles/s41562-023-01800-9 (visited on 09/20/2025).

[39] Sofia Jesus. "Freedom and Anxiety: A Psychoanalytic Exploration of the Super-ego in the Postmodern Era". In: *Trends in Psychology = Temas EM Psicologia* (June 13, 2023), pp. 1–20. ISSN: 2358-1883. DOI: 10.1007/s43076-023-00303-w. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10263376/ (visited on 04/16/2026).

[40] Benjamin P. Kay et al. "Moderating effects of music on resting state networks". In: *Brain Research* 1447 (Apr. 4, 2012), pp. 53–64. ISSN: 0006-8993. DOI: 10.1016/j.brainres.2012.01.064. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3307876/ (visited on 04/16/2026).

[41] Johannes Kepler et al. *The harmony of the world*. Memoirs of the American Philosophical Society. OCLC: 36826094. Philadelphia, Pa.: American Philosophical Society, 1997. 549 pp. ISBN: 978-0-87169-209-2. URL: http://www.books.google.co.uk/books?vid=ISBN0871692090 (visited on 08/30/2025).

[42] Gwangsu Kim, Dong-Kyum Kim, and Hawoong Jeong. "Spontaneous emergence of rudimentary music detectors in deep neural networks". In: *Nature Communications* 15.1 (Jan. 2, 2024). Publisher: Nature Publishing Group, p. 148. ISSN: 2041-1723. DOI: 10.1038/s41467-023-44516-0. URL: https://www.nature.com/articles/s41467-023-44516-0 (visited on 09/20/2025).

[43] Katalin Krenhardt et al. "Birdsong as an Indicator of Habitat Structure and Quality". In: *Ecology and Evolution* 15.6 (June 4, 2025), e71510. ISSN: 2045-7758. DOI: 10.1002/ece3.71510. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12137194/ (visited on 04/16/2026).

[44] Edward N. Lorenz. *The Essence Of Chaos*. Google-Books-ID: j5Ub6sMCoOsC. University of Washington Press, 1995. 244 pp. ISBN: 978-0-295-97514-6.

[45] Hanbing Lu et al. "Rat brains also have a default mode network". In: *Proceedings of the National Academy of Sciences* 109.10 (Mar. 6, 2012). Publisher: Proceedings of the National Academy of Sciences, pp. 3979–3984. DOI: 10.1073/pnas.1200506109. URL: https://www.pnas.org/doi/10.1073/pnas.1200506109 (visited on 04/18/2026).

[46] Robert M. MacCallum et al. "Evolution of music by public choice". In: *Proceedings of the National Academy of Sciences* 109.30 (July 24, 2012). Publisher: Proceedings of the National Academy of Sciences, pp. 12081–12086. DOI: 10.1073/pnas.1203182109. URL: https://www.pnas.org/doi/10.1073/pnas.1203182109 (visited on 09/25/2025).

[47] Benoit B. Mandelbrot. *The fractal geometry of nature*. In collab. with Internet Archive. San Francisco : W.H. Freeman, 1983. 506 pp. ISBN: 978-0-7167-1186-5. URL: http://archive.org/details/fractalgeometryo00beno (visited on 04/16/2026).

[48] Cassia Low Manting et al. "How musicality enhances top-down and bottom-up selective attention: Insights from precise separation of simultaneous neural responses". In: *Science Advances* 11.38 (Sept. 19, 2025), eadz0510. ISSN: 2375-2548. DOI: 10.1126/sciadv.adz0510.

[49] Matthias Mauch and Simon Dixon. "PYIN: A fundamental frequency estimator using probabilistic threshold distributions". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). ISSN: 2379-190X. May 2014, pp. 659–663. DOI: 10.1109/ICASSP.2014.6853678. URL: https://ieeexplore.ieee.org/document/6853678 (visited on 09/07/2025).

[50] Robert M. May. "Simple mathematical models with very complicated dynamics". In: *Nature* 261.5560 (June 1976). Publisher: Nature Publishing Group, pp. 459–467. ISSN: 1476-4687. DOI: 10.1038/261459a0. URL: https://www.nature.com/articles/261459a0 (visited on 04/16/2026).

[51] Fulvio Mazzocchi. "Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory". In: *EMBO Reports* 9.1 (Jan. 2008), pp. 10–14. ISSN: 1469-221X. DOI: 10.1038/sj.embor.7401147. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC2246621/ (visited on 04/16/2026).

[52] Samuel A. Mehr and Max M. Krasnow. "Parent-offspring conflict and the evolution of infant-directed song". In: *Evolution and Human Behavior* 38.5 (Sept. 1, 2017), pp. 674–684. ISSN: 1090-5138. DOI: 10.1016/j.evolhumbehav.2016.12.005. URL: https://www.sciencedirect.com/science/article/pii/S1090513816301969 (visited on 04/16/2026).

[53] Vinod Menon. "20 years of the default mode network: A review and synthesis". In: *Neuron* 111.16 (Aug. 16, 2023), pp. 2469–2487. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2023.04.023.

[54] Clémentine Mitoyen, Cliodhna Quigley, and Leonida Fusani. "Evolution and function of multimodal courtship displays". In: *Ethology* 125.8 (Aug. 2019), pp. 503–515. ISSN: 0179-1613. DOI: 10.1111/eth.12882. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC6618153/ (visited on 04/16/2026).

[55] Lyle L. Nelson et al. "Pushing the boundary: A calibrated Ediacaran-Cambrian stratigraphic record from the Nama Group in northwestern Republic of South Africa". In: *Earth and Planetary Science Letters* 580 (Feb. 15, 2022), p. 117396. ISSN: 0012-821X. DOI: 10.1016/j.epsl.2022.117396. URL: https://www.sciencedirect.com/science/article/pii/S0012821X22000322 (visited on 09/25/2025).

[56] Georg Northoff. "Psychoanalysis and the Brain – Why Did Freud Abandon Neuroscience?" In: *Frontiers in Psychology* 3 (Apr. 2, 2012), p. 71. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2012.00071. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3317371/ (visited on 04/16/2026).

[57] Eamonn O'Keeffe. "Military music and society during the French wars, 1793–1815*". In: *Historical Research* 97.275 (Feb. 1, 2024), pp. 108–128. ISSN: 0950-3471. DOI: 10.1093/hisres/htad027. URL: https://doi.org/10.1093/hisres/htad027 (visited on 04/16/2026).

[58] Caitlin Opland and Tyler J. Torrico. "Psychodynamic Therapy". In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2026. URL: http://www.ncbi.nlm.nih.gov/books/NBK606117/ (visited on 04/16/2026).

[59] Sam Passmore et al. "Global musical diversity is largely independent of linguistic and genetic histories". In: *Nature Communications* 15.1 (May 10, 2024). Publisher: Nature Publishing Group, p. 3964. ISSN: 2041-1723. DOI: 10.1038/s41467-024-48113-7. URL: https://www.nature.com/articles/s41467-024-48113-7 (visited on 09/20/2025).

[60] Aniruddh D. Patel. "The Evolutionary Biology of Musical Rhythm: Was Darwin Wrong?" In: *PLoS Biology* 12.3 (Mar. 25, 2014), e1001821. ISSN: 1544-9173. DOI: 10.1371/journal.pbio.1001821. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3965380/ (visited on 09/25/2025).

[61] Gaurav H. Patel et al. "Functional evolution of new and expanded attention networks in humans". In: *Proceedings of the National Academy of Sciences* 112.30 (July 28, 2015). Publisher: Proceedings of the National Academy of Sciences, pp. 9454–9459. DOI: 10.1073/pnas.1420395112. URL: https://www.pnas.org/doi/10.1073/pnas.1420395112 (visited on 04/18/2026).

[62] Mangor Pedersen et al. *Music Affects State Anxiety and Brain Connectivity*. Pages: 2023.05.18.541357 Section: New Results. May 18, 2023. DOI: 10.1101/2023.05.18.541357. URL: https://www.biorxiv.org/content/10.1101/2023.05.18.541357v1 (visited on 04/16/2026).

[63] Isabelle Peretz et al. "Neural overlap in processing music and speech". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1664 (Mar. 19, 2015). Publisher: Royal Society, p. 20140090. DOI: 10.1098/rstb.2014.0090. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2014.0090 (visited on 01/19/2025).

[64] Stefan Pfattheicher, Yngwie Asbjørn Nielsen, and Isabel Thielmann. "Prosocial behavior and altruism: A review of concepts and definitions". In: *Current Opinion in Psychology* 44 (Apr. 1, 2022), pp. 124–129. ISSN: 2352-250X. DOI: 10.1016/j.copsyc.2021.08.021. URL: https://www.sciencedirect.com/science/article/pii/S2352250X21001548 (visited on 01/21/2025).

[65] David Pogue. *Is Lip-Synching a Form of Cheating?* Scientific American. July 1, 2017. URL: https://www.scientificamerican.com/article/is-lip-synching-a-form-of-cheating/ (visited on 04/16/2026).

[66] Jacek Radwan, Leif Engqvist, and Klaus Reinhold. "A Paradox of Genetic Variance in Epigamic Traits: Beyond "Good Genes" View of Sexual Selection". In: *Evolutionary Biology* 43 (2016), pp. 267–275. ISSN: 0071-3260. DOI: 10.1007/s11692-015-9359-y. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4860406/ (visited on 04/16/2026).

[67] Zafar Rafii. "The Constant-Q Harmonic Coefficients: A timbre feature designed for music signals [Lecture Notes]". In: *IEEE Signal Processing Magazine* 39.3 (May 2022), pp. 90–96. ISSN: 1558-0792. DOI: 10.1109/MSP.2021.3138870. URL: https://ieeexplore.ieee.org/document/9770278 (visited on 09/07/2025).

[68] Teresa Raimondi et al. "Isochrony and rhythmic interaction in ape duetting". In: *Proceedings. Biological Sciences* 290.1990 (Jan. 11, 2023), p. 20222244. ISSN: 1471-2954. DOI: 10.1098/rspb.2022.2244.

[69] Andrea Ravignani, Daniel L. Bowling, and W. Tecumseh Fitch. "Chorusing, synchrony, and the evolutionary functions of rhythm". In: *Frontiers in Psychology* 5 (2014), p. 1118. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.01118.

[70] Mark Reybrouck, Peter Vuust, and Elvira Brattico. "Brain Connectivity Networks and the Aesthetic Experience of Music". In: *Brain Sciences* 8.6 (June 12, 2018), p. 107. ISSN: 2076-3425. DOI: 10.3390/brainsci8060107. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC6025331/ (visited on 04/16/2026).

[71] Bruce Richman. "Rhythm and melody in gelada vocal exchanges". In: *Primates* 28.2 (Apr. 1, 1987), pp. 199–223. ISSN: 1610-7365. DOI: 10.1007/BF02382570. URL: https://doi.org/10.1007/BF02382570 (visited on 10/07/2025).

[72] David Rothenberg et al. "Investigation of musicality in birdsong". In: *Hearing research* 308 (Feb. 2014), pp. 71–83. ISSN: 0378-5955. DOI: 10.1016/j.heares.2013.08.016. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3947120/ (visited on 01/21/2025).

[73] Gordon Marc le Roux. ""Whistle While You Work": A Historical Account of Some Associations Among Music, Work, and Health". In: *American Journal of Public Health* 95.7 (July 2005), pp. 1106–1109. ISSN: 0090-0036. DOI: 10.2105/AJPH.2004.042564. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC1449326/ (visited on 04/16/2026).

[74] Oliver Sacks. *Musicophilia: Tales of Music and the Brain, Revised and Expanded Edition*. Revised & enlarged edition. New York: Vintage, Sept. 23, 2008. 425 pp. ISBN: 978-1-4000-3353-9.

[75] Patrick E. Savage et al. "Music as a coevolved system for social bonding". In: *Behavioral and Brain Sciences* 44 (Jan. 2021), e59. ISSN: 0140-525X, 1469-1825. DOI: 10.1017/S0140525X20000333. URL: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/music-as-a-coevolved-system-for-social-bonding/F1ACB3586FD3DD5965E56021F506BC4F (visited on 10/07/2025).

[76] Suyash Sawant et al. "Spectrogram cross-correlation can be used to measure the complexity of bird vocalizations". In: *Methods in Ecology and Evolution* 13.2 (2022). _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf 210X.13765, pp. 459–472. ISSN: 2041-210X. DOI: 10.1111/2041-210X.13765. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13765 (visited on 09/07/2025).

[77] Susanne Shultz, Emma Nelson, and Robin I. M. Dunbar. "Hominin cognitive evolution: identifying patterns and processes in the fossil and archaeological record". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1599 (Aug. 5, 2012), pp. 2130–2140. ISSN: 0962-8436. DOI: 10.1098/rstb.2012.0115. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3385680/ (visited on 10/08/2025).

[78] Leigh W. Simmons et al. "Evolutionary Response to Sexual Selection in Male Genital Morphology". In: *Current Biology* 19.17 (Sept. 15, 2009), pp. 1442–1446. ISSN: 0960-9822. DOI: 10.1016/j.cub.2009.06.056. URL: https://www.sciencedirect.com/science/article/pii/S0960982209013323 (visited on 04/16/2026).

[79] Charles T. Snowdon and David Teie. "Affective responses in tamarins elicited by species-specific music". In: *Biology Letters* 6.1 (Sept. 2, 2009). Publisher: Royal Society, pp. 30–32. DOI: 10.1098/rsbl.2009.0593. URL: https://royalsocietypublishing.org/doi/10.1098/rsbl.2009.0593 (visited on 01/19/2025).

[80] Michael Spitzer. *The Musical Human: A History of Life on Earth*. New York: Bloomsbury Publishing USA, 2021. 1 p. ISBN: 978-1-63557-624-5 978-1-63557-625-2.

[81] Nicholas J. Strausfeld and Frank Hirth. "Deep homology of arthropod central complex and vertebrate basal ganglia". In: *Science* 340.6129 (Apr. 12, 2013). Place: New York, N.Y., pp. 157–161. ISSN: 1095-9203. DOI: 10.1126/science.1231828.

[82] D. Stuart-Fox and A. Moussalli. "Sex-specific ecomorphological variation and the evolution of sexual dimorphism in dwarf chameleons (Bradypodion spp.)" In: *Journal of Evolutionary Biology* 20.3 (2007). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1420-9101.2007.01295.x, pp. 1073–1081. ISSN: 1420-9101. DOI: 10.1111/j.1420-9101.2007.01295.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1420-9101.2007.01295.x (visited on 04/16/2026).

[83] Liila Taruffi et al. "Effects of Sad and Happy Music on Mind-Wandering and the Default Mode Network". In: *Scientific Reports* 7.1 (Oct. 31, 2017). Publisher: Nature Publishing Group, p. 14396. ISSN: 2045-2322. DOI: 10.1038/s41598-017-14849-0. URL: https://www.nature.com/articles/s41598-017-14849-0 (visited on 04/16/2026).

[84] Michael Trimble and Dale Hesdorffer. "Music and the brain: the neuroscience of music and musical appreciation". In: *BJPsych International* 14.2 (May 1, 2017), pp. 28–31. ISSN: 2056-4740. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5618809/ (visited on 01/21/2025).

[85] Wiebke Trost et al. "Live music stimulates the affective brain and emotionally entrains listeners in real time". In: *Proceedings of the National Academy of Sciences of the United States of America* 121.10 (Mar. 5, 2024), e2316306121. ISSN: 1091-6490. DOI: 10.1073/pnas.2316306121.

[86] Ashish Vaswani et al. *Attention Is All You Need*. Aug. 2, 2023. DOI: 10.48550/arXiv.1706.03762. arXiv: 1706.03762[cs]. URL: http://arxiv.org/abs/1706.03762 (visited on 09/23/2025).

[87] Paul Hendriks Vettehen and Gabi Schaap. "An attention economic perspective on the future of the information age". In: *Futures* 153 (Oct. 1, 2023), p. 103243. ISSN: 0016-3287. DOI: 10.1016/j.futures.2023.103243. URL: https://www.sciencedirect.com/science/article/pii/S0016328723001477 (visited on 09/25/2025).

[88] Simone Vossel, Joy J. Geng, and Gereon R. Fink. "Dorsal and Ventral Attention Systems". In: *The Neuroscientist* 20.2 (Apr. 2014), pp. 150–159. ISSN: 1073-8584. DOI: 10.1177/1073858413494269. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4107817/ (visited on 04/18/2026).

[89] Simone Vossel et al. "Deconstructing the Architecture of Dorsal and Ventral Attention Systems with Dynamic Causal Modeling". In: *Journal of Neuroscience* 32.31 (Aug. 1, 2012). Publisher: Society for Neuroscience Section: Articles, pp. 10637–10648. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0414-12.2012. URL: https://www.jneurosci.org/content/32/31/10637 (visited on 04/18/2026).

[90] Peter Vuust et al. "Music in the brain". In: *Nature Reviews Neuroscience* 23.5 (May 2022). Publisher: Nature Publishing Group, pp. 287–305. ISSN: 1471-0048. DOI: 10.1038/s41583-022-00578-5. URL: https://www.nature.com/articles/s41583-022-00578-5 (visited on 09/20/2025).

[91] Holly Watkins. *Musical Vitalities: Ventures in a Biotic Aesthetics of Music*. New Material Histories of Music. Chicago, IL: University of Chicago Press, Nov. 2018. 240 pp. ISBN: 978-0-226-59470-5. URL: https://press.uchicago.edu/ucp/books/book/chicago/M/bo31043611.html (visited on 08/30/2025).

[92] Eugene P. Wigner. "The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959". In: *Communications on Pure and Applied Mathematics* 13.1 (1960). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160130102, pp. 1–14. ISSN: 1097-0312. DOI: 10.1002/cpa.3160130102. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160130102 (visited on 01/21/2025).

[93] Joe Wolfe. "Speech and Music, Acoustics and Coding, and What Music Might Be 'For'". In: (2002).

[94] Zengyao Yang et al. "Music tempo modulates emotional states as revealed through EEG insights". In: *Scientific Reports* 15.1 (Mar. 10, 2025). Publisher: Nature Publishing Group, p. 8276. ISSN: 2045-2322. DOI: 10.1038/s41598-025-92679-1. URL: https://www.nature.com/articles/s41598-025-92679-1 (visited on 04/16/2026).

[95] Andy Wai Kan Yeung. "Is the Influence of Freud Declining in Psychology and Psychiatry? A Bibliometric Analysis". In: *Frontiers in Psychology* 12 (Feb. 18, 2021), p. 631516. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2021.631516. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7930904/ (visited on 04/16/2026).

Table 1: **Overall agreement for the FDA machine learning classification tasks with the various categories comparisons in our research.** Classification was performed with a multi-agent model including a maximum depth, nearest centroid, K nearest neighbor and functional quadratic discriminant analysis. Model accuracy was converted to Cohen's kappa based upon the number of grouping factors it was deployed upon. The live vs studio comparison involved only the popular musician Bjork. The musical genres compared were popular music, jazz and opera. The anuran (chorus frog) comparisons included 4 species from the genera Hyla and Pseudacris. The avian comparisons included Nightingale, Superb Lyrebird, Northern Mockingbird, Brown Thrasher, Catbird, and either Song Sparrow or House Sparrow. The comparison of all animal groups included the anuran, avian, opera, jazz, and popular musicians combined within each group.

category	dimension	groups	Cohen's kappa	agreement
live VS studio	control	2	0.35	fair
live VS studio	energy	2	0.735	substantial
live VS studio	surprise	2	0.525	moderate
jazz-female	control	2	0.71	moderate
jazz-female	energy	2	0.516	moderate
jazz-female	surprise	3	0.813	perfect
jazz-male	control	2	0.363	fair
jazz-male	energy	3	0.628	substantial
jazz-male	surprise	3	0.797	substantial
opera-female	control	5	0.391	fair
opera-female	energy	5	0.686	substantial
opera-female	surprise	5	0.619	substantial
opera-male	control	5	0.613	substantial
opera-male	energy	5	0.784	substantial
opera-male	surprise	5	0.707	substantial
pop-female	control	4	0.338	fair
pop-female	energy	4	0.486	moderate
pop-female	surprise	4	0.728	substantial
pop-male	control	3	0.416	moderate
pop-male	energy	3	0.384	fair
pop-male	surprise	3	0.453	moderate
genre-female	control	3	0.633	substantial
genre-female	energy	3	0.673	substantial
genre-female	surprise	3	0.73	substantial
genre-male	control	3	0.636	substantial
genre-male	energy	4	0.477	moderate
genre-male	surprise	3	0.666	substantial
anuran	control	4	0.697	substantial
anuran	energy	4	0.38	fair
anuran	surprise	4	0.49	moderate
avian(song sp)	control	4	0.727	substantial
avian(song sp)	energy	4	0.451	moderate
avian(song sp)	surprise	4	0.595	moderate
avian(house sp)	control	6	0.727	substantial
avian(house sp)	energy	6	0.436	moderate
avian(house sp)	surprise	5	0.574	moderate
all (female)	control	5	0.606	substantial
all (female)	energy	5	0.719	substantial
all (female)	surprise	5	0.681	substantial
all (male)	control	5	0.675	substantial
all (male)	energy	5	0.748	substantial
all (male)	surprise	5	0.747	substantial