- **POPSTAR- a biopsychological geometric model for human musical experience**

- **Abstract**: We have developed a geometric model for human musical experience that combines an underlying biological signal of Darwinian fitness with a model of psychological impacts. The triangular architecture of this model seems to describe very well the existing genres of music. In the future, we want to statistically validate this model using statistical analysis and machine learning applied to sound files.  The POPSTAR software maps the multivariate acoustic fitness signals of sound files in live time and displays the signal as a movie containing an animated Chernoff face diagram and a data trajectory within a ternary diagram. POPSTAR is useful for many types of comparisons (e.g. different artists, different genres, experts vs. novices, live vs studio performances etc)

- **Description of the research team**: Drs. GA Babbitt (GSOLS) and EP Fokoue (SMS). We will accept as many student volunteers as possible who are interested in the project to work on python code, visualization, client-side application, and data collection. Students from biology, bioinformatics, math and perhaps performing arts programs will be recruited for this project.

- **Project Narrative**:

What makes a sound musical? What is the origin and purpose of music?  Is music a purely human endeavor, or part of the larger fabric of nature? These questions and similar others have been around for thousands of years, perhaps as long as music itself. The evolutionary origin of musical and artistic behavior in humans has long been a subject of archeology and many popular books. One point of view, borrowed from the field of behavioral ecology has been far less applied. Behavioral ecologists generally ask the basic question, "To what advantage does a behavior give to an individual (or group) when compared to the absence of the behavior?" We would hypothesize that musical behavior in humans might have similar evolutionary function to other forms of social communication in animals. We conjecture that music may act as a form of a Darwinian fitness signal akin to other kinds of animal communication and display; one where individuals signal through performances that allow others to honestly assess biological fitness (i.e. energy reserves, cognitive ability, and motor skills …aka 'good genes' models) when making choices regarding potential mates and/or group inclusion. While modern music obviously does not always serve this function in present times, it may still contain ancestral elements of fitness signaling from the deep evolutionary past. In our preliminary work, we show that autocorrelative features defining music from other natural sounds exhibit feature shifts indicating that musicality in the human voice may have predated the evolution of speech and language.  This work is preprinted at the link below, and will probably be under journal review at the time of this proposal (Babbitt et al., 2025).

https://www.biorxiv.org/content/10.1101/2025.03.13.643054v2

What are the main components of fitness signaling in music? …and what is the cognitive impact on the mind?

We put forth a geometric theoretical model for the relationship between three major axes of fitness signaling and three major cognitive effects of music (Figure 1).  In animal communication, behavioral signals of fitness often evolve towards multivariate structure to promote honest signaling. For example, female birds assessing male song will often attune to several acoustic components at once, such as loudness, bout frequency, and control of species-specific elements in the song, as the combination of these elements produces a far less cheatable signal of male fitness than any single element alone. In musical sound generation as well as human physical movements in response to music (i.e. dance), we theorize that elements of fitness are combined along three major axes; energy level, motor control, and surprise (Figure 1 – upright gray triangle).  Behavioral elements evolving under all of these axes would provide clear selective advantages in individual and group conflicts (i.e. in the context of individual fights or physical battles between social groups). Different combinations of emphasis on these axes of fitness create adjacent cognitive effects related to the stimulation of emotional, physical, and intellectual impacts on the mind (Figure 1 – inverse red triangle).  Interestingly, the general feelings invoked by various music genres would seem to be easily explained by this basic geometry via either singular or binary emphasis on the different points of the gray fitness triangle (Figure 2).
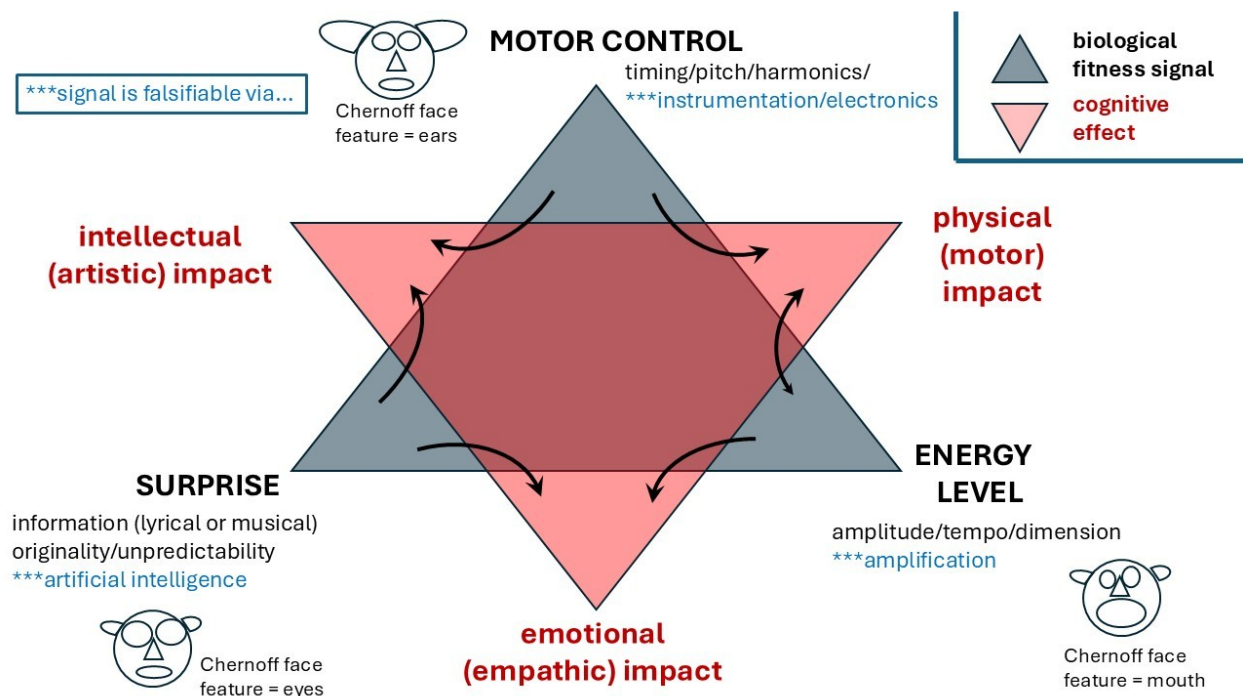
**Figure 1. A geometric theory for the proximate stimulus and ultimate effects of musical stimulation on human minds.** Proximate stimulation is a biological signal of fitness, much like signaling in animal communication. Three components of biological fitness are energy level, motor control, and surprise or novelty. For example, a female bird might choose among prospective males based upon how loud or fast the 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). These three outward aspects of biological fitness signaling by an individual musician or group of musicians (upright gray triangle) combine to affect the minds of performers and listeners by stimulating physical, emotional, and intellectual sensations (inverted red triangle). We also include a plan for general Chernov face representation of each potential multivariate input feature for the fitness signal when measured in various sound files. (Chernoff, 1973).
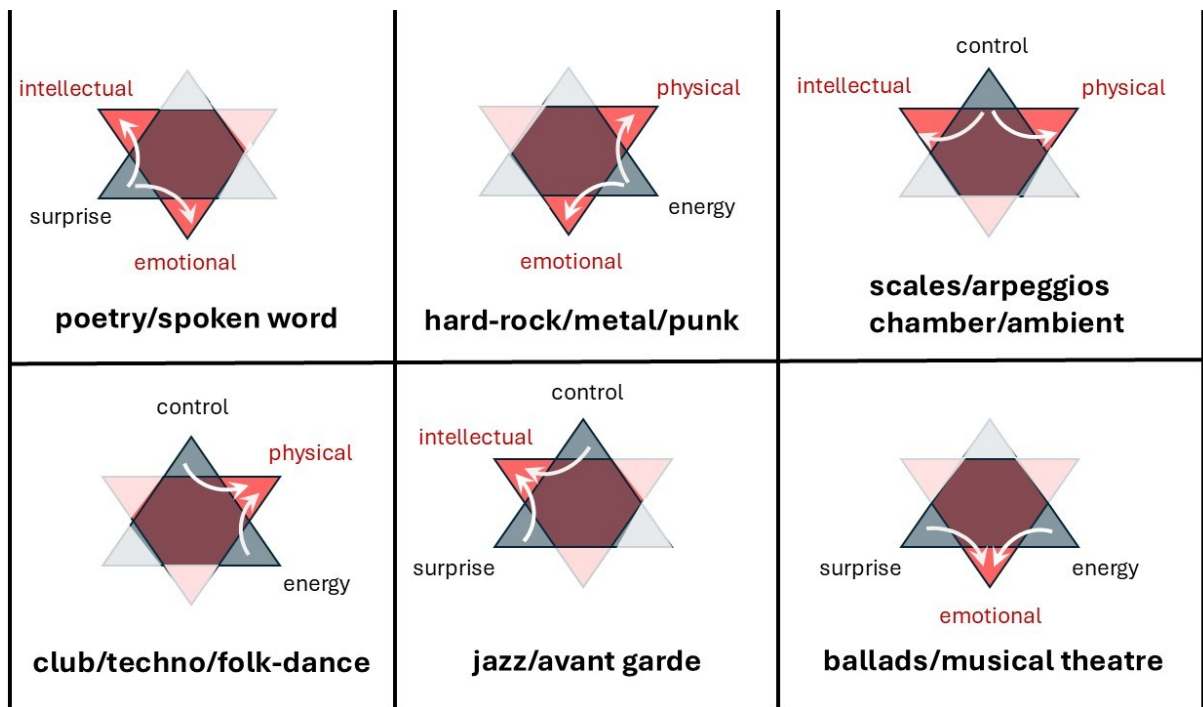


Figure 2. The effects of increased emphasis on one or two proximate stimuli (gray) can explain the cognitive reactions (red) to various artistic genres through their adjacency in the geometric pattern. For example, emphasis on just energy level in music with de-emphasis on novelty/surprise and fine motor 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 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, classical and popular music would seem to be well balanced between all three stimuli (energy, control and surprise).

**Implementation and Output** – POPSTAR runs as a python pipeline with a PyQt5 graphical user interface (GUI) that is launched from the terminal as 'python popstar.py'. A single sound file or a folder with multiple sound files can be entered on the GUI. Several user options such as beat tracking, lyrical analysis, and averaging vs maximum value plotting are available. For more details and download go to https://github.com/gbabbitt/POPSTAR-signaling-in-music
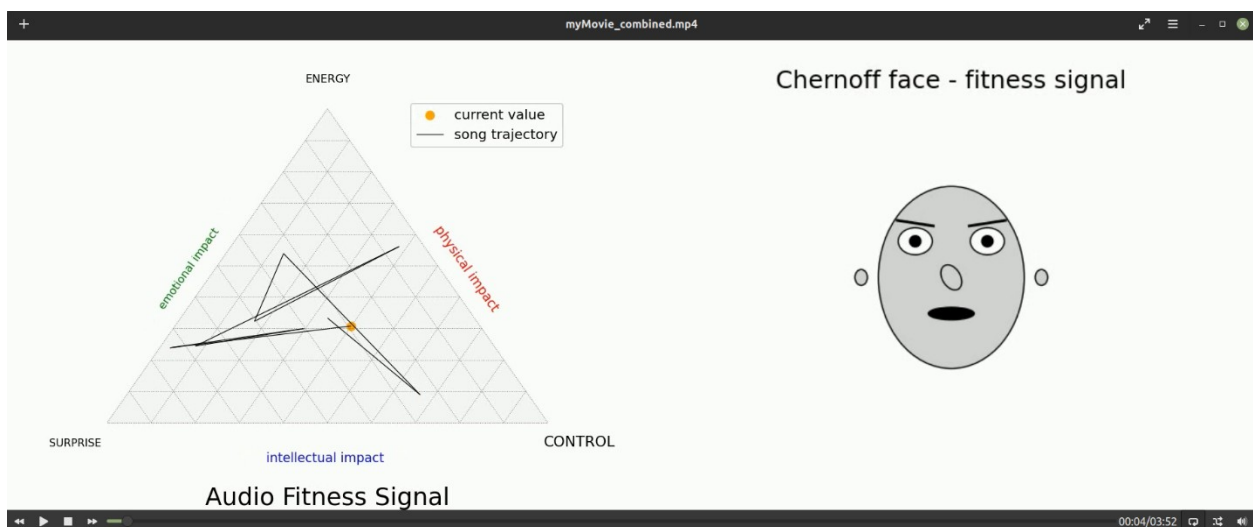


**Chernoff face details**
Eyes = surprise/complexity of the acoustic signal
   Eye height = note variability index (derived from spectral cross-correlation)
   Eye width = Lempel-Ziv complexity (derived from substring counts)
   Pupil area = Complexity Index (derived from refined multi-scale entropy)
Eyebrows = energy level of the acoustic signal
   Slant level = tempo (beats per minute)
Mouth = energy level of the acoustic signal
   Mouth height = amplitude volume
   Mouth width = sound size or ambient dimension (derived from $1^{st}$ order autocorrelation)
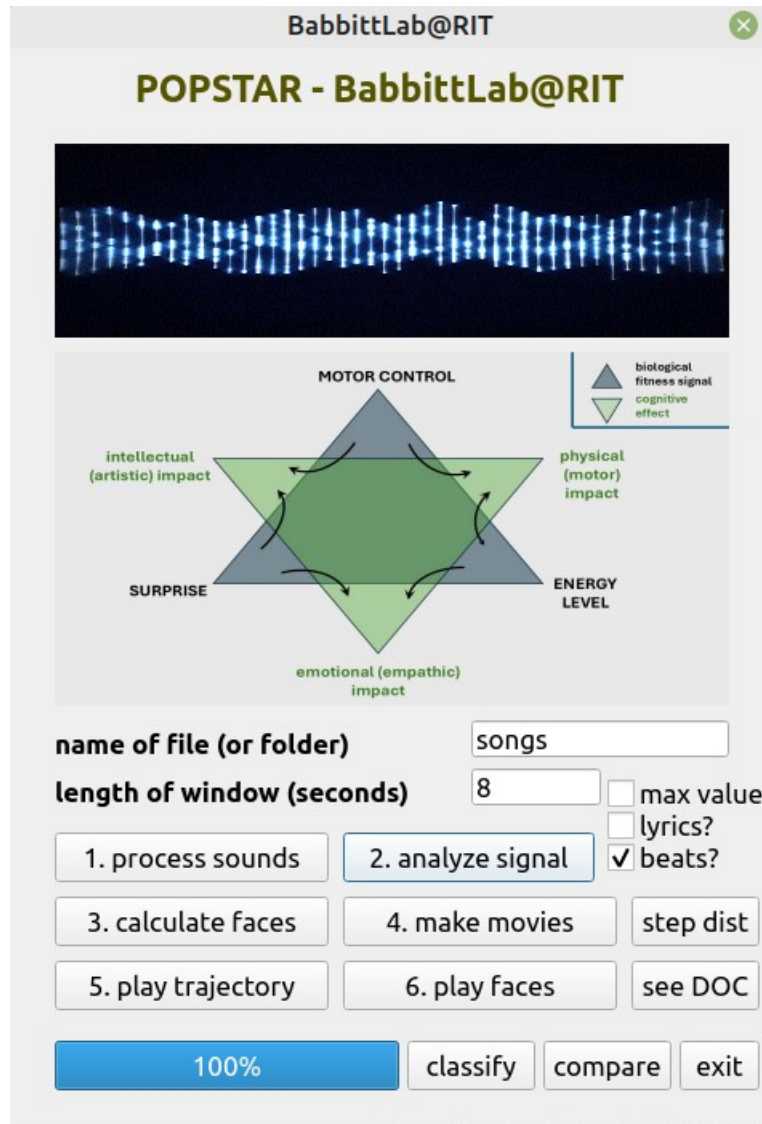Ears = motor control of the acoustic signal
   Ear height = fundamental frequency (f0) control or control of pitch
   Ear width = evenness or control of beat intervals (derived from sums of square differences)
Nose = energy and control of the acoustic signal
   Slant = level of harmonic energy or number of harmonic frequencies (fN)

# The POPSTAR main user interface



**Usage – Choose a file (.mp3 or .wav) OR a folder containing multiple files and run the buttons in order 1-4.   If you ran a single file the playback buttons (5-6) are available.  Results are located in a new folder 'popstar_results'. If you processed an entire folder, playback is not available, but the movie files are available in both the system folders created for each file and all together in the 'popstar_results' folder. The 'step dist' button can be run after step 4 (making movies) and produces histograms of the step distributions in the dynamic ternary plot in both a time dependent (1st order) and time indepenedent (0th order) fashion. The 'classify' and 'compare' open new interfaces to run the analyses described below.**

best model waterfall.mp3|1st order =lognorm|0th order =lognorm
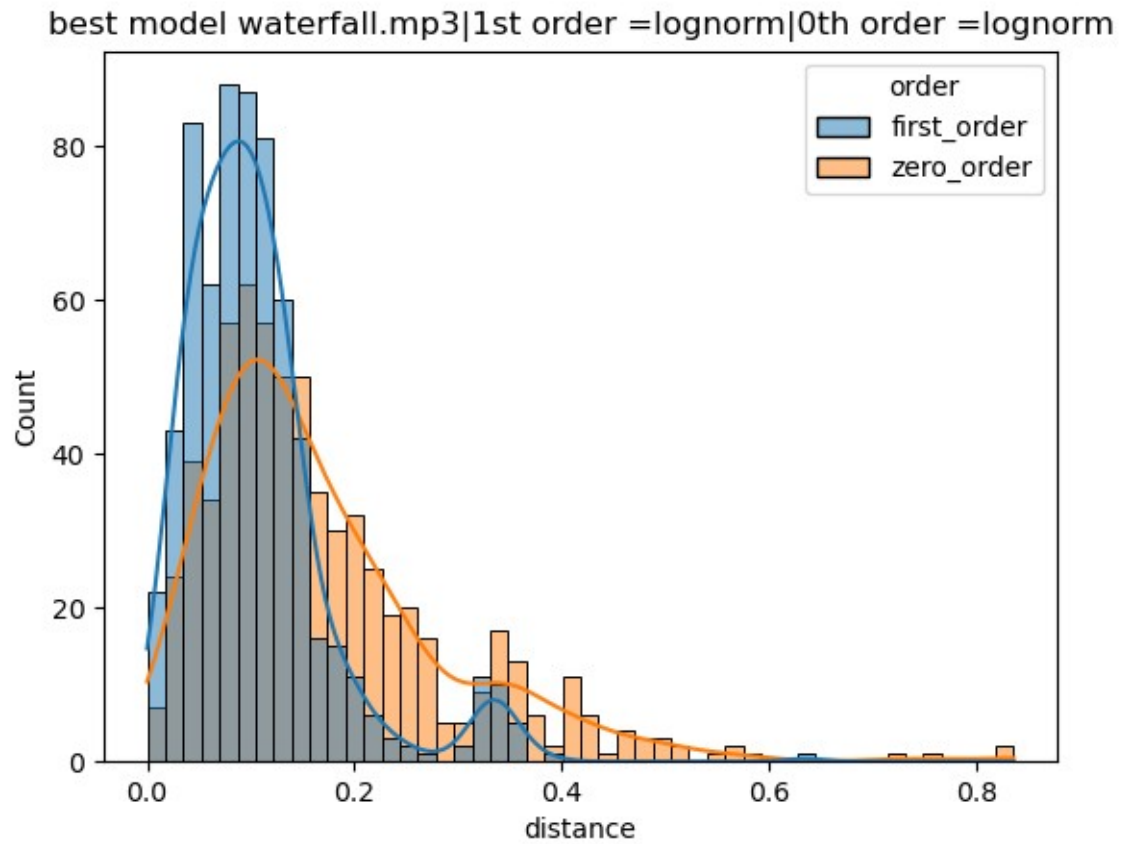
The 'step dist' histogram is accompanied by multi-model inference (AIC and BIC lists) in the file mm_inference_myFile.txt. The lowest negative number indicates best model that describes the data. The best model is also shown on the top of the histogram itself.

**The 'comparison' interface only works upon multiple folders containing multiple files relating to categories of interest (e.g. human voices compared to animal vocalizations or jazz music compared to classical music). Overall composite histograms are compared using KL divergence of the 1st order step distributions and average values of the three axes on the ternary plots (i.e. energy, control, surprise).**
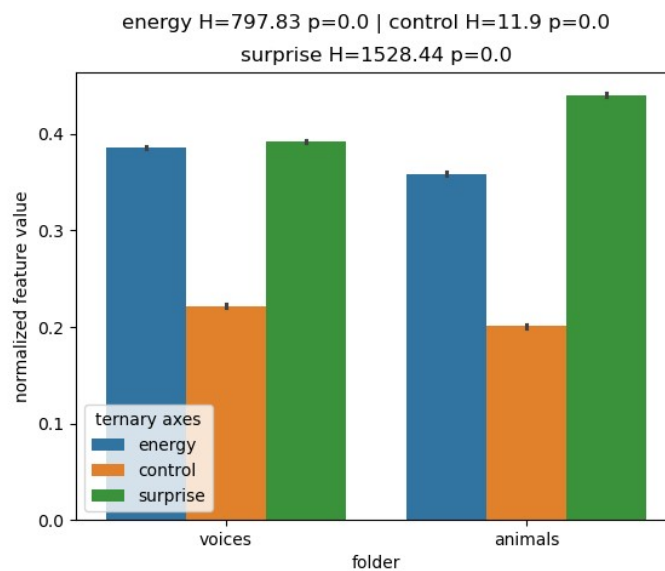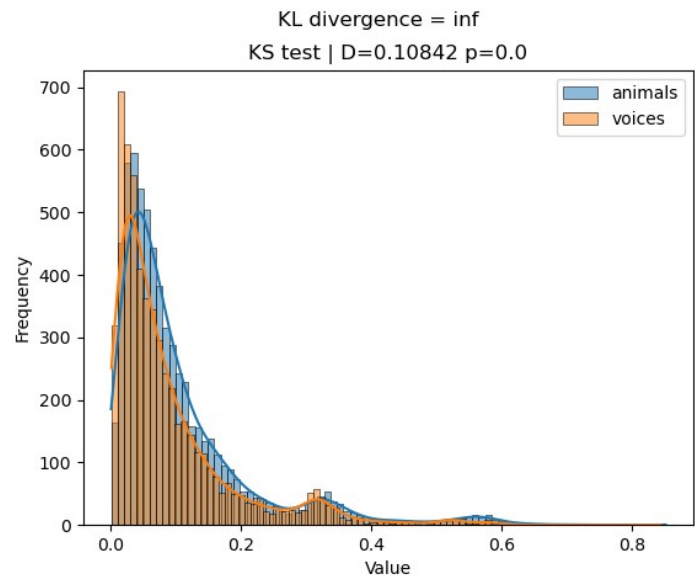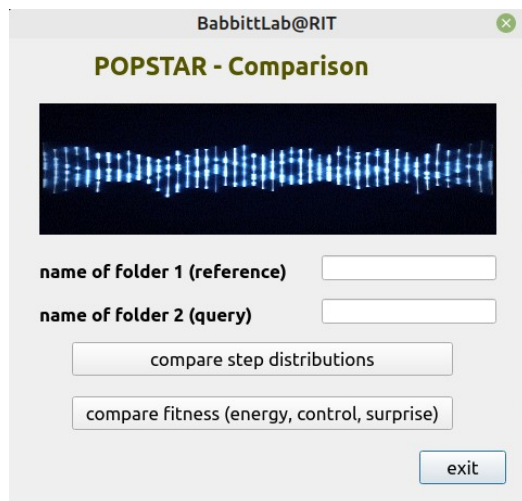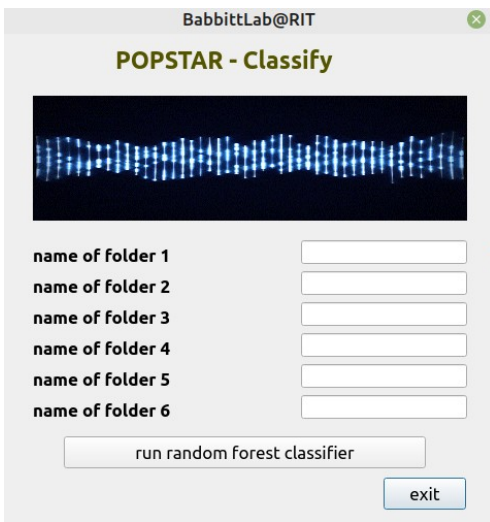
**The 'classify' interface invokes a random forest classifier bootstrapped 100 times to discriminate between 2-6 folders (i.e. categories) of sound files. An overall performance as well as feature importances are reported in a .txt file output and as a barplot. This can be used to investigate what multivariate features of the acoustics are most useful for distinguishing between categories that are compiled and analyzed.**



Feature Importance from Random Forest Model (500 trees, 100 bootstraps)
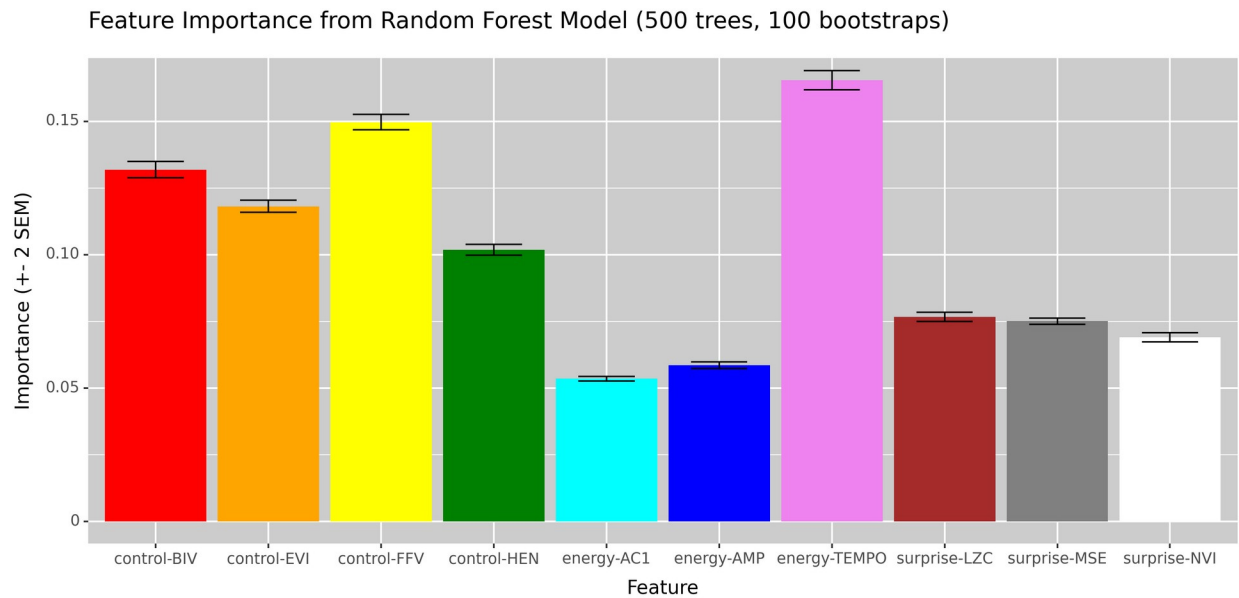
# References

Babbitt, G.A., Wang, L., Fokoue, E.P., 2025. Musicality in protein interaction dynamics informs the multi-scale evolution of prosocial behavior. https://doi.org/10.1101/2025.03.13.643054

Chernoff, H., 1973. The Use of Faces to Represent Points in k-Dimensional Space Graphically. Journal of the American Statistical Association 68, 361–368. https://doi.org/10.1080/01621459.1973.10482434

Jarne, C., 2019. A method for estimation of fundamental frequency for tonal sounds inspired on bird song studies. MethodsX 6, 124–131. https://doi.org/10.1016/j.mex.2018.12.011

Sawant, S., Arvind, C., Joshi, V., Robin, V.V., 2022. Spectrogram cross-correlation can be used to measure the complexity of bird vocalizations. Methods in Ecology and Evolution 13, 459–472. https://doi.org/10.1111/2041-210X.13765