Low-Level Audio Feature Correlates to Physiological Arousal: A Secondary Analysis of the DEAP Dataset

Daniel Ethridge
ATLAS Institute, University of Colorado Boulder
Boulder, Colorado, USA
Daniel.Ethridge@colorado.edu

Luis X. de Pablo
Department of Ecology and Evolutionary Biology, University of Colorado Boulder
Boulder, Colorado, USA
luis.depablo@colorado.edu

Grace Leslie
College of Music and ATLAS Institute, University of Colorado Boulder
Boulder, Colorado, USA
grace@colorado.edu

ABSTRACT

Music invites measurable changes in a listener's physiology. Heart rate variability (HRV), defined as the beat to beat variation across successive heartbeats, is a frequently used physiological measure of the activity of the autonomic nervous system. Galvanic skin response (GSR) is defined by changes in electrical potentials in the skin and has also been shown to serve as a biomarker of autonomic nervous system activity. Prior work has shown that listening to various genres of music will elicit measurable changes in heart rate, HRV, and GSR. However, what remain underexplored are any potential correlations between low level audio features of music and their associated physiological responses in listeners. We begin filling this gap in knowledge by performing a secondary analysis of the DEAP dataset. We extract heart rate, HRV features, and GSR features from participant data corresponding to when they listened to music. Then we extract low-level audio features from the music that participants listened to. Using mixed effects models and multiple regression analyses, we find correlations between two low-level audio features and extracted physiological data which suggest that vocal or vocal-like sounds and percussive or noisy sounds in music may have particularly strong effects on the body.

CCS CONCEPTS

• **Information systems** → **Music retrieval**; • **Human-centered computing**; • **Applied computing** → **Bioinformatics**;

KEYWORDS

Physiological Signals, Heart Rate Variability, HRV, Galvanic Skin Response, GSR, Mixed Effects Models, Exploratory Data Analysis

ACM Reference Format:
Daniel Ethridge, Luis X. de Pablo, and Grace Leslie. 2024. Low-Level Audio Feature Correlates to Physiological Arousal: A Secondary Analysis of the DEAP Dataset. In *Audio Mostly 2024 - Explorations in Sonic Cultures (AM '24), September 18–20, 2024, Milan, Italy*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3678299.3678327

AM '24, September 18–20, 2024, Milan, Italy

ACM ISBN 979-8-4007-0968-5/24/09
https://doi.org/10.1145/3678299.3678327

1 INTRODUCTION

For most people, listening to music may invite an emotional response. Music serves as a focus aid for many people, it is increasingly utilized in music therapy contexts, and it plays a large role in social development [33]. Additionally, music has been repeatedly shown to elicit responses from our physiology [16, 20, 43]. In recent years, researchers have explored the mechanisms that underlie the brain's response to music, attempting to explain how emotional responses to music manifest in the brain [9] to how music may be a regulator of pain [27]. However, an equally interesting question concerns how music manifests in our peripheral physiological signals. One piece of music might cause our heart to beat faster while another may cause an increase in the conductivity of our skin.

Analyzing these physiological states over time can tell us a surprising amount about our internal state and stress response. Two physiological responses in particular are heart rate variability (HRV) and galvanic skin response (GSR). HRV is the natural change in time intervals between heart beats[1], and a higher level is typically an indicator of health [41]. HRV has become a focus in psychophysiological research due to its link with the parasympathetic nervous system which is relevant to "several self-regulation mechanisms linked to cognitive, affective, social, and health phenomena" [24]. Colloquially, the parasympathetic nervous system is referred to as the "rest and digest" portion of the nervous system and pulls the body into more relaxed states. Galvanic Skin Response (GSR) is defined as "a change in the electrical properties of the skin", and it is considered an indicator of sympathetic activity [42]. Sympathetic activity is a portion of the general arousal pattern that is present at the onset of emergencies which helps ready us for stressing stimuli [42]. In a similar colloquial fashion, the sympathetic nervous system is associated with the "fight or flight" response and activates in response to various stressors. While acute sympathetic activation is helpful and natural, long-term sympathetic activation from chronic stress is linked to cardiovascular disease and other ailments. These insights and studies on the sympathetic nervous system have greatly influenced clinical practice [44]. The parasympathetic and sympathetic nervous systems are two components of the broader autonomic nervous system[2] which governs involuntary motor functions. As stated prior, the parasympathetic nervous system is constantly pulling the body towards a more relaxed state while

[1]The heart does not beat at a constant rate but rather it continuously beats faster and slower over time. We invite the reader to feel for their pulse and then take slow, deep breaths. Heart rate increases during inhalation and decreases on exhalation. This is a basic demonstration of HRV.
[2]The third and final component is the enteric nervous system whose discussion lies outside the scope of this work.

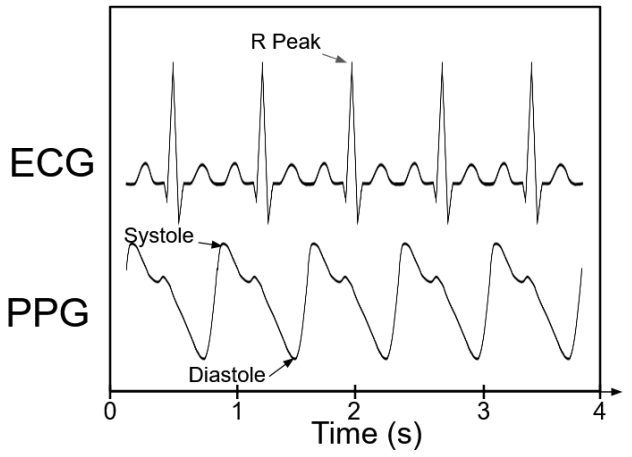

Figure 1: Electrocardiogram (ECG) is measured via electrodes placed on the body. Photoplethysmogram (PPG) uses optics to measure blood volume through the skin of typically a finger tip, earlobe, or wrist. PPG is the technology used by smartwatches to gather cardiac data.

the sympathetic nervous system is constantly pulling the body towards a more aroused state. The exact balance at any given time is determined by different levels of various neurotransmitters in the nervous system. Through measuring physiological responses such as heart rate, HRV, and galvanic skin response, we can measure relative effects and activity of the ANS, giving us insight into health and stress information about an individual. More detail about the ANS can be found in [14].

Prior work has introduced numerous extractable features for both HRV and GSR. HRV features can be extracted from both electrocardiogram and photoplethysmogram signals (see figure 1). Time- and frequency-domain features exist for HRV. The time domain features rely on time differences between consecutive R waves (see figure 2), while frequency domain features are calculated by performing a Fourier transform on the signal. GSR signals can be divided into two main components: skin conductance response (SCR) and skin conductance level (SCL). SCR, or the tonic component, represents a slowly changing baseline of the signal that modulates based on hydration, autonomic regulation, and skin moisture. SCL, or the phasic component, is a faster-changing component which is sensitive to emotionally arousing stimuli [42]. Common features to extract from a GSR signal include the first four statistical moments, minimum and maximum values, and the number of peaks over a defined period of time from the SCL component [35]. Below is a selection of HRV metrics found in [41] and [33]. Formulas and in-depth explanations for those utilized in our analysis are given in section 3.1.

- Time domain
  - SDNN: Standard Deviation of NN intervals
  - pNN50: Percentage of successive RR intervals that differ by more than 50 milliseconds
  - RMSSD: Root Mean Square of Successive Differences
- Frequency domain

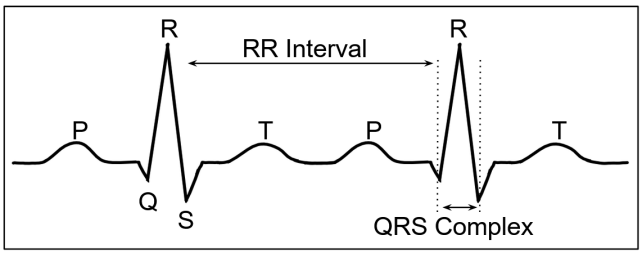

Figure 2: In an ECG signal, a QRS complex represents depolarization (or contraction) of the left and right ventricles in the heart [40]. The duration between successive R peaks is measured in milliseconds. HRV can be seen in different durations between multiple successive R peaks.

  - ULF power: Absolute power of the ultra-low-frequency band ($\leq$ 0.003 Hz) (ms^2)
  - VLF power: Absolute power of the very-low-frequency band (0.003 - 0.04 Hz) (ms^2)
  - LF power: Absolute power of the low-frequency band (0.04 - 0.15 Hz) (ms^2)
  - HF power: Absolute power of the high-frequency band (0.15 Hz - 0.4 Hz) (ms^2)

The choice to focus solely on physiological signals stemming from the heart and the skin is a matter of scope. Other physiological signals such as electroencephalography (EEG) for brainwaves, electromyography (EMG) for muscle activity, and respiration rate are studied to great effect in the context of music perception and cognition. Of these, EEG is the most common, and several datasets are available online to study music perception with EEG [7]. Due to the popularity of studying EEG in this context and the aforementioned connections of peripheral physiological signals to the autonomic nervous system, we chose to focus on HRV and GSR.

1.1 Music, Emotion, the Heart, and the Skin

Scholarly research has explored the effects of different music and sounds on both heart rate and HRV. White noise has been shown to increase the LF power and the ratio of LF power to HF power ($\frac{LF}{HF}$) [25]. In [43], researchers tested how speech noise, traffic noise, and mixed noise affect HRV compared to background noise. The researchers described speech noise as multiple people talking simultaneously, traffic noise as aircraft and road traffic sounds, mixed noise as a combination of speech and traffic noise. No description of background noise is given. They reported that speech noise decreased participants' LF power and $\frac{LF}{HF}$ when compared with traffic noise and mixed noise, but results were not significant. Musical structure was shown to play a role when a group of singers had both higher RMSSD values and synced HRV when singing a mantra designed to produce a 0.1 Hz breathing rate (ie one full breath every 6 seconds) compared to independent humming (no coordinated breathing) or the hymn *Fairest Lord Jesus* (semi-coordinated breathing) [47]. In a study comparing sedative music (Erik Satie's Gymnopedie No. 1, arranged by Claude Debussy) to excitative music (Igor Stravinsky's "Sacrificial Dance" from *The Rite of Spring*), participants had increased HF power when listening to

sedative music when compared to excitative music while LF power and $\frac{LF}{HF}$ increased with the number of repetitions [19]. Separately, [20] came to similar conclusions, finding lower heart rate and higher RMSSD values when participants listened to ambient music instead of metal music.

We note that the meaning of LF power and $\frac{LF}{HF}$ is debated. While LF power and $\frac{LF}{HF}$ traditionally refer to joint sympathetic and parasympathetic nerve activation and the sympatho-vagal balance[3] [25, 43], respectively, these ideas are challenged. Billman provides an argument against the relationship between sympatho-vagal balance and the LF-to-HF ratio [4], and Laborde et. al. state that the relationship between LF power and sympathetic nervous system activity is loose. Due to this uncertainty in LF power and $\frac{LF}{HF}$, we choose to focus on two HRV metrics recommended by [24]: RMSSD and SDNN. Mathematical definitions for these features are given in section 3.1.

In addition to HRV metrics, [20] also investigated GSR. They found increased rapid fluctuations in GSR according to greater amount of stimuli and an increase in skin conductivity correlated with rhythm changes. The authors do not report numerical values for these specific observations. In [21], engineering students were tasked with one of three tasks: yogic breathing, religious hymn listening, or flute music listening. GSR values were recorded, and, when compared with a control group that only received instruction to silently remain seated, the researchers reported statistically significant lower values of GSR for the experimental conditions than for the control group. In [15], GSR was found a useful feature in discriminating whether or not men and women were listening to music. The idea of GSR as a sympathetic index has been successfully utilized in both emotion recognition [35, 51] and in the recognition of perceived valence and arousal of affective sounds [6].

1.2 Our Contribution

Despite music's several influential roles in social relationships [38] and its ability to evoke numerous basic emotions in us, a mystery of why music possess such power remains. With insights about the influence of musical attributes like genre [20] and structure [47] on physiological state, we address potential answers to the broader mystery. But with these higher level investigations, whether or not specific qualities of music like frequency distributions, sound brightness, and rhythmic variation lead to physiological changes is largely unknown. Researchers in [43] perform a basic audio analysis alongside the physiological analysis. The researchers stated that the speech noise had different specific loudness, specific roughness, fluctuation strength, and tonality when compared with traffic noise and mixed noise, though the authors do not report formulae for these features. It is important to expand on the investigation of these low level features and the physiological correlates. There exists a controversy in the music perception community about why (or even if) music induces emotion, and [22] believes that the controversy is fueled by a lack of understanding about the underlying mechanisms contributing to musically induced emotion and physiological arousal. In this work, we aim to go beyond the higher level genre and structure tasks routinely seen in the literature, and we investigate low level audio features that could provide insight into pertinent building blocks of sound that correlate with physiological response.

[3]Sympatho-vagal balance simply refers to the balance betweent the sympathetic and parasympathetic nervous systems. The vagus nerve is the one of the main components in the parasympathetic nervous system.

2 THE DEAP DATASET

The DEAP (Database for Emotion Analysis using Physiological Signals) dataset [23] was originally developed in 2012 to explore "the possibility of classifying emotion dimensions induced by showing music videos to different users." The researchers recorded EEG and peripheral physiological signals (PPG, GSR, and respiration) from 32 participants. A 2-minute baseline recording was conducted, and then each participant watched a minute long excerpt from 40 different music videos. After each video, participants gave subjective ratings to each music video based on perceived valence, arousal, familiarity, liking, and dominance. The researchers presented a discussion on found correlations between the EEG and the emotions ratings, and they later used extracted features from the audio and video alongside the peripheral physiological data to train classification models for emotion prediction. Both the peripheral physiological signals and the audio features served as inputs to the DEAP models, whereas we utilize the peripheral physiological signals as outcome variables in our correlational analyses.

3 PHYSIOLOGICAL SIGNAL ANALYSIS

We chose to analyze the GSR data and the PPG data of the DEAP dataset. As stated in section 1.1, GSR and HRV are indicators of sympathetic and parasympathetic nervous system activity, respectively. We also analyzed the mean interbeat interval (IBI) given by

$$IBI_{mean} = 1000 * \frac{60}{HR_{mean}}, \tag{1}$$

where HR_{mean} is average heart rate over a period of time. 60 refers to the number of seconds in a minute, and multiplying by 1000 provides IBI_{mean} in milliseconds.

3.1 HRV

Each PPG signal was filtered using the $\mathrm{filter\_signal}$ function from the Python package Heartpy [46]. The filter was designed as a 4th order bandpass filter with lower and upper frequency bounds of 0.5Hz and 10Hz, respectively. The signals were then segmented into the appropriate portions corresponding to music listening timestamps. The HRV metrics that we extracted from the PPG data are the **R**oot **M**ean **S**quare of **S**uccessive **D**ifferences (RMSSD) and the **S**tandard **D**eviation of **NN** intervals (SDNN). Both of these features are explained in figure 3.

In addition to the two raw HRV feature vectors for RMSSD and SDNN, we created two additional feature vectors: ΔRMSSD and ΔSDNN. These are defined as

$$HRV_{\Delta} = HRV_s - HRV_b, \tag{2}$$

where HRV_s is an HRV value during stimulus presentation, and HRV_b is baseline HRV.

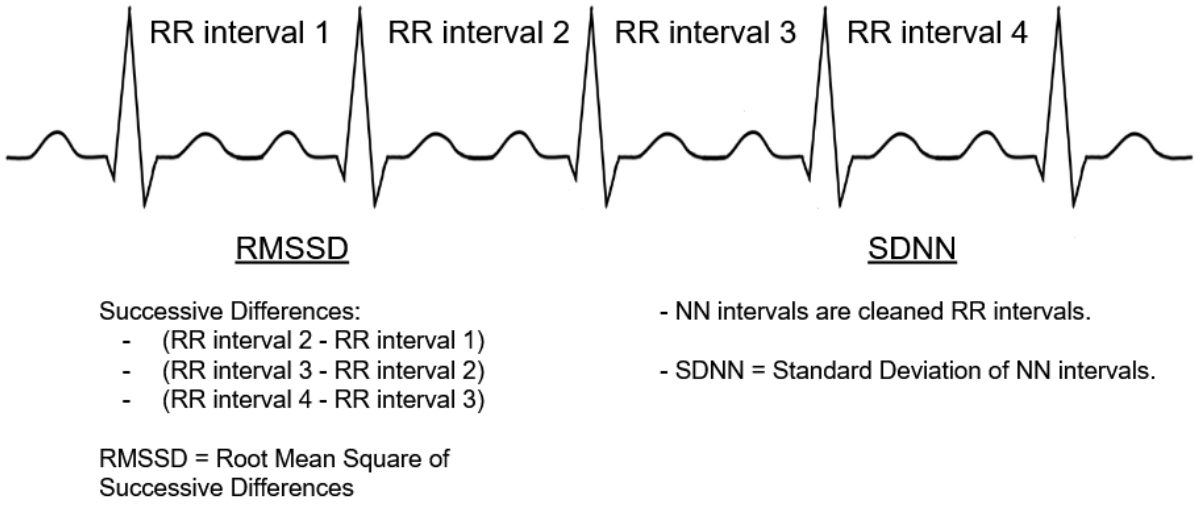

Figure 3: RMSSD and SDNN are two of the features extracted from the cardiac data in the DEAP dataset, and they both rely on RR intervals. An RR interval is simply the time in milliseconds from one R peak to the next. RMSSD is calculated by first calculating the time differences between successive RR intervals. We then perform a root mean square on the resulting set of successive differences. SDNN is simply the standard deviation of the NN intervals. A set of NN intervals is defined as a set of RR that has been cleaned and is free of irregular beats. Cleaning can be performed in a variety of ways. Expert visual analysis of the signal can be used to manually verify beat timings, digital filters can be utilized to reduce noise, and various outlier detection algorithms can be used to discard irregular beats [46].

3.2 GSR

The GSR signals were preprocessed using the eda_process function from the Python package Neurokit2 [30]. Afterwards, the phasic component was extracted using the eda_phasic function. We then segmented each GSR signal corresponding to music listening. The features that we extracted from GSR include the first four statistical moments, the first and second derivatives of the first two statistical moments, root mean square, and the first and second derivatives of root mean square.

4 AUDIO ANALYSIS

Methodologies in this section were largely informed by chapter 3 in *An Introduction to Audio Content Analysis* by Alexander Lerch [26]. Lerch is a leader in the field of music information retrieval, and his book provides clear instruction on validated methodologies.

4.1 Audio Preprocessing

The DEAP dataset provides the audio stimulus in the form of YouTube.com video links. We utilized the open source PyTube[4] package to download the videos from YouTube. About half of the video links did not work because of age-restricted errors, videos being taken down, etc. In these instances, we manually found what we believed to be the same or similar video at a different link, and used those to finish the audio downloading process.

Each audio file was read into Python using the Scipy [48] library. Each audio file was downmixed to mono if necessary, and the DC offset was removed. All audio was normalized such that sample amplitudes ranged between -1 and 1. Lastly, the 60 second long sections as denoted by the DEAP metadata were extracted and saved as new WAV files. Of the 40 audio stimuli, six of the audio segments were slightly less than 60 seconds.

[4]https://github.com/pytube/pytube

4.2 Feature Extraction and Preprocessing

We chose to utilize Librosa [31] for audio feature extraction since it is a validated and widely used Python package. This limited the features that we extracted from the audio. Where necessary, a block size of 2048 and a hop size of 512 were used in the Librosa feature extraction methods.

The 17 extracted features were:

- Spectral Features [26]
  - Spectral Centroid: The center of gravity, or mean, of the magnitude spectrum of a signal.
  - Spectral Spread: The concentration of the magnitude spectrum around the spectral centroid.
  - Spectral Contrast: The difference between the spectral peak and spectral valley in each frequency band.
  - Spectral Flatness: The ratio of the geometric mean and the arithmetic mean.
  - Spectral Rolloff: A measure of bandwidth of a block of n audio samples.
- Cepstral Features: Typically, anywhere from 4 to 20 Mel-Frequency Cepstral Coefficients (MFCCS) are extracted, and it has been shown that the first few coefficients contain the priciple information [26]. Based on this, we extracted the first 10 MFCCs and used the Librosa-default 128 mel bands.

- Zero Crossing Rate (ZCR): The number of times in a block of audio samples that a signal changes sign.
- Tempo *(Note: Tempo was determined manually. In checking the accuracy of Librosa tempo estimates, many were inaccurate.)*

We then calculated the mean and the standard deviation across audio blocks for each feature except tempo, resulting in 33 feature vectors per minute-long audio segment. Finally, we applied z-score normalization as described in [26] to each feature vector.

5 STATISTICAL METHODS

All statistical analyses were performed using R statistical software (v4.3.3) [37]. The final stages of data preprocessing and formatting for use in R were performed and visualizations generated using Tidyverse [49] packages, ggplot2 [17], and cowplot [50] unless otherwise stated.

Some of the physiological features were positively skewed. We applied a natural log transformation to remedy this. In order to determine which covariates to include in our model, we utilized two separate tools: random forest and a model selection exercise. We note that while random forest algorithms are typically associated with large amounts of data, they are also useful with smaller amounts of data [12]. For the model selection exercise, we utilized the dredge function from the MuMIN [2] package for R. Dredge takes a linear model that includes all candidate covariates as an input and outputs the AIC of models built using every possible combination of covariates. We created these models, and all later models, using the lme4 package [3] in R. We repeated these methods using each physiological feature as the outcome variable, and both methods consistently produced two audio features of interest:

- Standard deviation of zero crossing rate (ZCR) and
- Standard deviation of MFCC number 6.

To determine whether or not mixed effects models would be necessary for our analysis, we calculated the AIC using the AIC function in R for the linear fixed effects models and the corresponding linear mixed effects models utilizing these audio feature predictors. Consistently, the mixed effects models produced lower AICs which confirmed the need for mixed effects models.

Next, we performed model comparisons using the Flexplot [11] package in R. Flexplot calculates the AIC, BIC, and Baye's factor for both models in a comparison. Each model comparison used a full model (consisting of predictors) (equation 3) and a reduced model (consisting only of the data mean) (equation 4) created using lme4:

$$outcome_{ij} = b_{0j} + b_{1j}(predictor1_i) + b_{2j}(predictor2_i) + \epsilon_{ij} \quad (3)$$

$$outcome_{ij} = b_{0j} + \epsilon_{ij} \quad (4)$$

where i represents trial number, j represents participant number, and ϵ represents the residual.

We determined that the most likely outcome variables were the natural logarithm (ln) of the standard deviation (σ) of GSR and mean interbeat interval.

With this insight, we created two multiple regression models. Model 1 examined the relationship between $ln(\sigma(\text{GSR}))$ with σ(MFCC 6) and σ(ZCR). Model 2 examined the relationship between mean interbeat interval with σ(MFCC 6) and σ(ZCR). Derived from further model comparisons, both models utilize random intercepts while only model 1 utilizes random slopes. The random intercepts and slope account for variability between participants. The final models that present a global fit across all participants are shown in equations 5 and 6.

$$ln(\sigma(GSR)) = -4.18 + 0.12(\sigma(MFCC\ \ 6)) + 0.04(\sigma(ZCR)) + \epsilon \quad (5)$$

$$Mean\ \ IBI = 853.13 - 2.65(MFCC\ \ 6) - 3.734(ZCR) + \epsilon \quad (6)$$

6 RESULTS

To test model significance, we utilized the report [29] package in R to compute p-values using a Wald t-distribution approximation. In model 1, σ(ZCR) did not show a significant effect on $ln(\sigma(\text{GSR}))$ ($\beta = 0.04, p = 0.316$) while σ(MFCC 6) showed a significant positive effect ($\beta = 0.12, p = 0.021$). For model 2, both audio features show significant negative effects ($\sigma(\text{ZCR}) : \beta = -3.74, p < 0.001$; σ(MFCC 6) : $\beta = -2.66, p = 0.013$). We tested for co-linearity between σ(ZCR) and σ(MFCC 6) and found that the predictor variables were independent. These results are summarized in table 1 and visualized in figure 4. Residual plots (created with Flexplot [11]) are shown in figure 5.

7 DISCUSSION

The two audio features which showed significant correlations were the standard deviations of MFCC 6 and zero crossing rate. While there is discussion in the academic community that zero crossing rate is correlated with the noisiness of a signal [26] and that the MFCCs are correlated with timbre [32], instantaneous features such as these do not have agreed upon perceptual meaning [26]. There is no clear parallel between physiological response and known perceptual dimensions. Attempts to use popular features to classify music into perceptual categories has performed poorly [36]. While progress has been made in similar areas like emotion classification using neural networks [8], automatically classifying music into subjective bins such as emotion and perception is still an open research area. As of 2022, state-of-the-art accuracy is 69% [18]. That being said, there exist patterns throughout the music information retrieval literature where the features in question repeatedly present themselves. In other words, each feature has proven to be an effective candidate for specific audio classification tasks.

7.1 Zero Crossing Rate

Zero crossing rate (ZCR) has been a common feature used in audio classification tasks for decades due to its simple calculation. ZCR provides an indirect estimate of the fundamental frequency of a signal [1, 26, 39]. In [13], ZCR, the MFCCs, spectral centroid, and energy are utilized for musical instrument classification. ZCR only performed worse than the MFCCs, and it particularly excelled in drum recognition. The authors postulate about the good performance in noting that the drum is an unharmonic, noisy sound. In [39], ZCR, the MFCCs, and other features were used as an input into various support vector machine models for drone detection based on audio. ZCR, being a single-value feature, performed worse

Model	R^2	Outcome	Predictors	Estimate (β)	p-value
1	0.5	$ln(\sigma(\text{GSR}))$	σ(ZCR)	0.04	0.316
			σ(MFCC 6)	0.12	0.021
2	0.92	Mean IBI	σ(ZCR)	-3.74	< 0.001
			σ(MFCC 6)	-2.66	0.013

Table 1: Summary of results. ln = natural log, σ = standard deviation

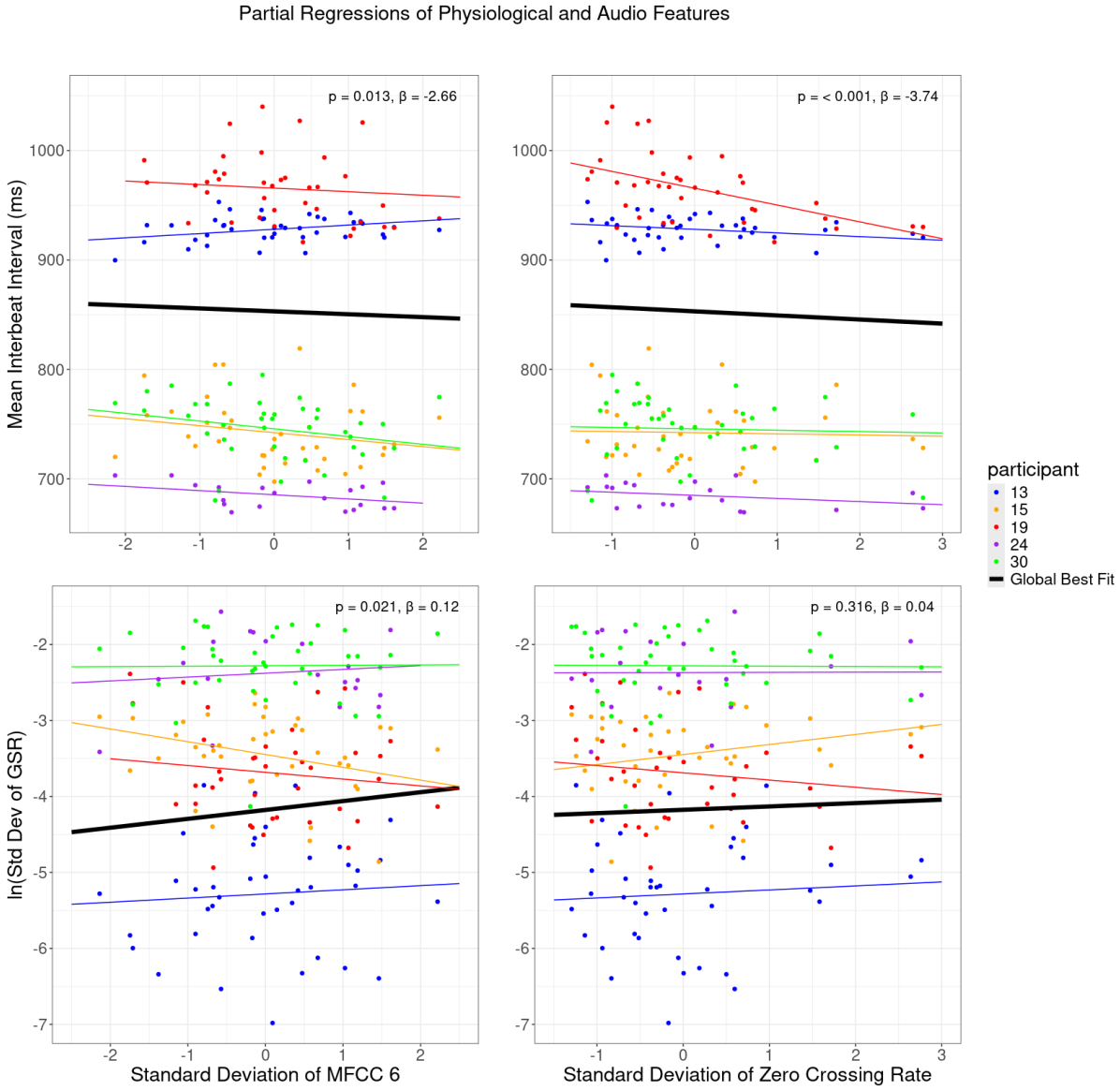

Figure 4: Four plots showing the partial regressions of the audio features plotted against the physiological features. The global best fit line (black) is the average of all individual participant best fit lines (colored). Five participants chosen at random are also included on the plots. The audio features are normalized with the horizontal axes representing z-scores. See section 4.2.

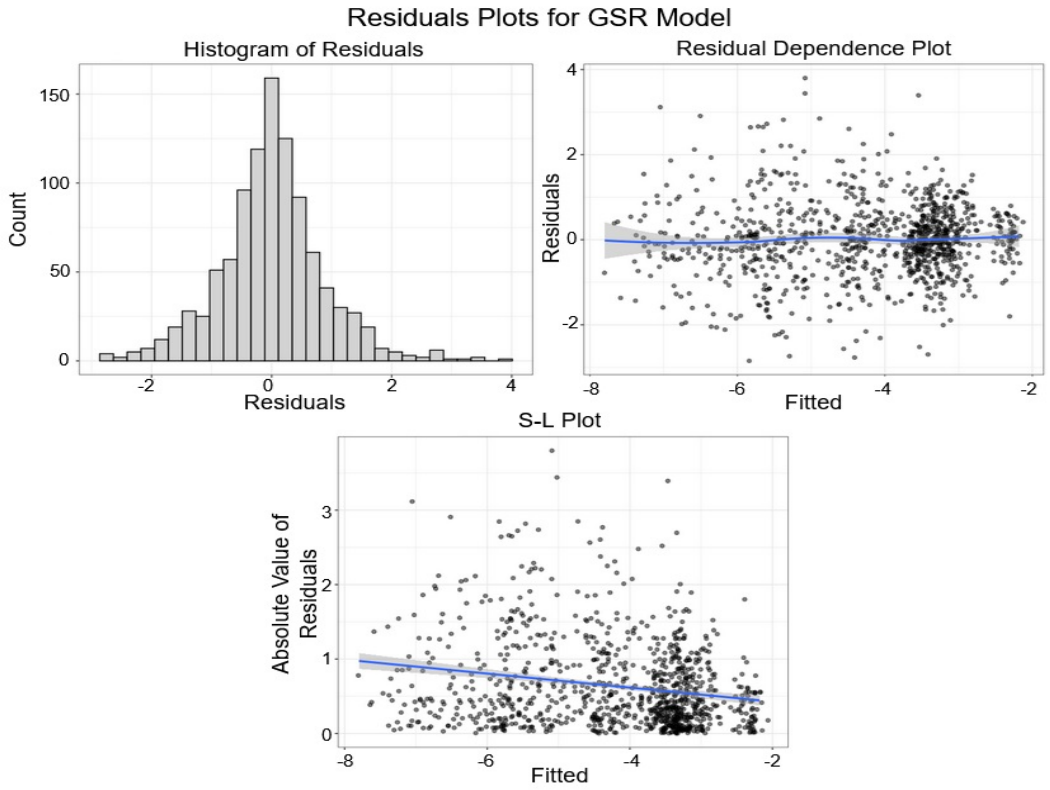

Figure 5: The utilized statistical methods make four assumptions about the data: normality, linearity, heteroskedasticity, and independence. These three plots address the first three assumptions for the GSR model (see equation 5) and visually show the degree to which the data satisfies these assumptions. The fourth, independence, was addressed in a separate col-linearity test where the assumption was deemed satisfied. A second set of residual plots for the mean IBI model (equation 6) is located in supplementary materials. Plots generated using Flexplot [11].

than the MFCCs and the other multivalue features. It did, however, improve classification performance when utilized alongside the MFCCS and other multi-value features.

Various authors note that a higher ZCR is indicative of a more noisy signal, while a lower ZCR implies a more periodic signal [13]. This could partially explain why ZCR was a useful feature in drum classification and drone sound classification. The feature has also been utilized in works aiming at discriminating speech from music [1, 34]. A potential explanation for the utility of ZCR in music-speech discrimination is that speech signals may be more inherently noisy than music signals, raising the ZCR [52].

7.2 The Mel-Frequency Cepstral Coefficients

While only the standard deviation of the 6th MFCC showed any significant correlation to physiological values in our work, it is worth discussing the MFCCs more generally. The coefficients are calculated by taking the discrete cosine transform of a log-mel spectrogram. In this way, they are a "spectrum of a spectrum." They have performed excellently in musical instrument recognition [13, 28] and music-speech discrimination [52] tasks, among others. Historically, they have been widely used in speech signal processing [13, 26, 52]. Such speech signal processing tasks include word recognition in continuous speech [10] and health monitoring via detecting speech changes in patients with Parkinson's disease [45]. Some researchers have stated a link between the MFCCs and timbre [32] which makes sense given that the features are a description of the spectral envelope of an audio signal [26].

7.3 Conclusions

Considering the links between fundamental frequency and noisiness with ZCR and the link between the MFCCs and timbre, we are left to reason why the standard deviation of these features correlated with the standard deviation of GSR and the mean interbeat interval. Let us first return to the discussion about GSR. In general, changes in GSR reflect activity in the sympathetic nervous system, and we know that the phasic component is particularly responsive to acute stimuli. More variation in ZCR across audio blocks is consistent with less periodicity in a signal [26] (and thus more randomness), and this variation across blocks could be interpreted physiologically as increased amounts of acute emotional stimuli. Perhaps this lack of periodicity could also be associated with more unpredictability. It is possible that experiencing less predictability, and thus more music

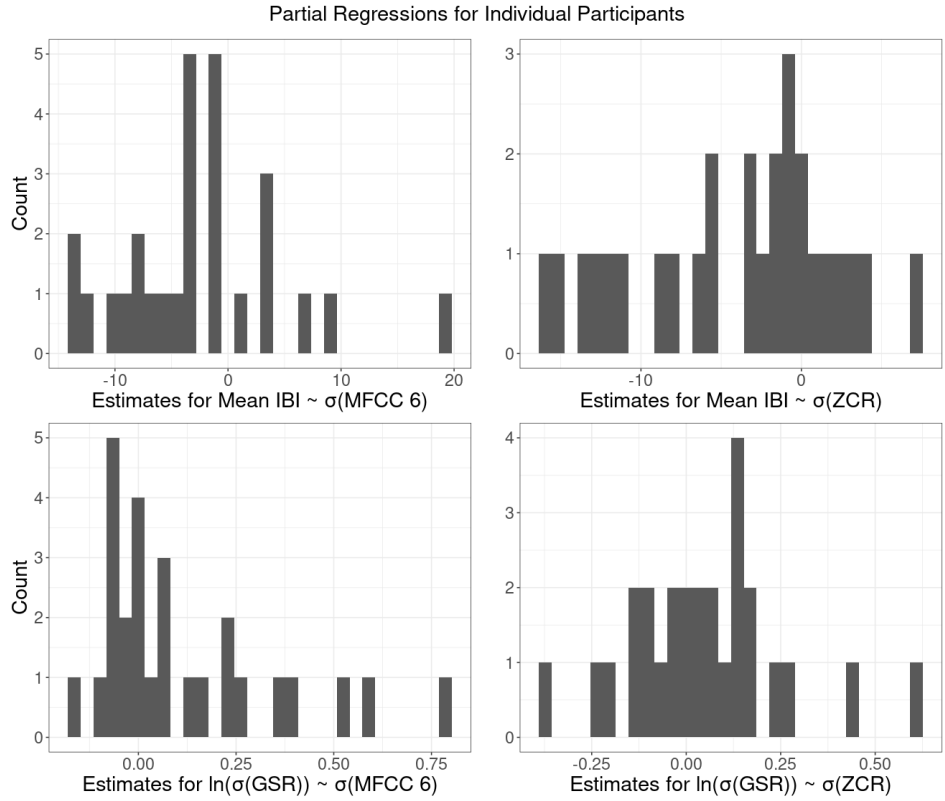

Figure 6: The global best fit lines in figure 4 represent the average slope estimates across all participants. These histograms show the various individual participant slope estimates. Note the variation and sparsity in the histograms due in part to the limited sample size.

expectancy violation, is associated with increased physiological arousal [22]. The observed negative correlation between mean IBI and the audio features can also be explained through this logic. Increased sympathetic response has been repeatedly shown to decrease mean IBI (i.e. increase heart rate).

We may postulate in a similar fashion about the correlation between the physiological features and the standard deviation of MFCC 6. The triangular filter for MFCC 6 (using 128 mel bands) corresponds to the frequency range of approximately 152Hz to 213Hz[5]. While we reiterate that the MFCCs have no real perceptual dimension correlate, we note that the much of the human voice spectral power lies near this range. We also note that whether or not music has lyrics has been shown to have differing effects on brain response to music [5]. If MFCC 6 can be attributed to a portion of the spectral envelope where a large portion of the human voice's frequencies reside, then variation in these voice-like frequencies could be linked to higher physiological arousal. More generally, it may be that increased frequency variation in the lower frequency bands may correlate with increased physiological arousal.

As a small post hoc experiment, we extracted the amplitude values of the log-mel spectrogram across the 128 mel bins. We noted negative correlations between the standard deviation of GSR and average decibel measurements between approximately 152Hz and 426Hz. While further study is needed to support this correlation, the potential link between physiological arousal and average power in the aforementioned frequency band is worth noting. Following this train of thought, applying a band stop filter on the audio from 152Hz to 426Hz removed any correlation between physiological arousal and the standard deviation of MFCC 6. Applying a low pass filter with a cutoff frequency of 750Hz to the audio preserved the correlations between the audio and physiological features.

The discussed correlations between physiological arousal and the variation in ZCR and MFCC 6 seem to point towards the ratio of periodicity to noisiness and the amount of the spectral power from about 150Hz to 420Hz as being correlated to an individual's physiological response to music. There is also the question of the human voice such that variation of innate human frequencies might correlate with physiological arousal

8 LIMITATIONS

Of the total 32 participants in the dataset, we only analyzed the data from 28 of them. Data from participants 31 and 32 was unavailable

[5]In audio engineering, too much of this frequency band is often associated with the "boominess" or "muddiness" of a mix, and it is often cut using an audio equalizer. However, boosting this frequency band when necessary can add a desired fullness to the mix.

to us, and we were unable to extract some of the physiological data from participants 23 and 29 due to potential noise. For some PPG segments, the heart rate output by HeartPy [46] was above 90bpm or below 50bpm. We considered these values noise. Some of the output HRV values also fell outside of accepted ranges and were thrown out. Inherently from our methodology of analyzing data collected for a different purpose, there may be confounding variables present, but we believe that attaining significant results given the limitations present speaks positively of our results. Analyzing data from a larger set of participants would be beneficial. We can visualize the benefit of more participants in figure 6. Variation across participants exists, and a larger sample size would help mitigate that effect. Lastly, there might be issues with data heteroskedasticity (see figure 5).

9 FUTURE WORK

In this study, we analyzed correlations between low-level audio features and peripheral physiological features. A future study could expand on this by including data from other open source datasets with audio(visual) stimuli and peripheral physiological data. Further study with a custom curated dataset is needed to determine any causal effects. There is also an opportunity to collect data designated for measuring the effects of music on peripheral physiological signals. There are numerous open source datasets that contain electroencephalogram data and music, and there are many that contain audiovisual stimuli and peripheral physiological signals[7]. The DEAP dataset was used largely because we were unable to locate other open source datasets that contained audio stimuli and peripheral physiological data with sufficient duration for calculating HRV metrics [24].

Another potential route to explore is physiological data collected from commercially available sensors such as smartwatches. These devices are becoming increasingly ubiquitous in society, and they could potentially allow for more ecologically valid data despite being lower quality than what would be expected from research grade equipment. Methods to consider for analyzing this type of data largely stem from the fields of data mining and machine learning. Such methods would provide ways to deal with and glean insight from large amounts of physiological data collected while listening to music.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation grant 2313518 "CAREER: Multimodal Brain and Body Music Interfaces to Promote Entrainment, Connection, and Creative Science Education."

REFERENCES

[1] Muhammad Alnadabi and Sherri Johnstone. 2008. Discrimination between speech and music using time series events. In *2008 9th International Conference on Signal Processing* (Beijing, China). IEEE, 565–570. https://doi.org/10.1109/ICOSP.2008.4697196

[2] Kamil Bartoń. 2023. *MuMIn: Multi-Model Inference.* https://CRAN.R-project.org/package=MuMIn R package version 1.47.5.

[3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[4] George E. Billman. 2013. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. 4, Article 25 (February 2013). https://doi.org/10.3389/fphys.2013.00026

[5] Elvira Brattico, Vinoo Alluri, Brigitte Bogert, Thomas Jacobsen, Nuutti Vartiainen, Sirke Nieminen, and Mari Tervaniemi. 2011. A Functional MRI Study of Happy and Sad Emotions in Music with and without Lyrics. 2, Article 308 (December 2011). https://doi.org/10.3389/fpsyg.2011.00308

[6] Aaron Frederick Bulagang, Ng Giap Weng, James Mountstephens, and Jason Teo. 2020. A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. 20, Article 100363 (June 2020). https://doi.org/10.1016/j.imu.2020.100363

[7] Vybhav Chaturvedi, Arman Beer Kaur, Vedansh Varshney, Anupam Garg, Gurpal Singh Chhabra, and Munish Kumar. 2022. Music mood and human emotion recognition based on physiological signals: a systematic review. 28 (February 2022), 21–44. https://doi.org/10.1007/s00530-021-00786-6

[8] Deepti Chaudhary, Niraj Pratap Singh, and Sachin Singh. 2021. Development of music emotion classification system using convolution neural network. 24 (September 2021), 571–580. https://doi.org/10.1007/s10772-020-09781-0

[9] Ian Daly, Asad Malik, Faustina Hwang, Etienne Roesch, James Weaver, Alexis Kirke, Duncan Williams, Eduardo Miranda, and Slawomir J. Nasuto. 2014. Neural correlates of emotional responses to music: An EEG study. 573 (June 2014), 52–57. https://doi.org/10.1016/j.neulet.2014.05.003

[10] S. Davis and P. Mermelstein. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. 28, 4 (August 1980), 357–366. https://doi.org/10.1109/TASSP.1980.1163420

[11] Dustin Fife. 2024. *flexplot: Graphically Based Data Analysis Using 'flexplot'.* R package version 0.20.3.

[12] Dustin A. Fife and Juliana D'Onofrio. 2023. Common, uncommon, and novel applications of random forest in psychological research. 55 (August 2023), 2447–2466. https://doi.org/10.3758/s13428-022-01901-9

[13] Seema Ghisingh and V. K. Mittal. 2016. Classifying musical instruments using speech signal processing methods. In *2016 IEEE Annual India Conference (INDICON)* (Bangalore, India). IEEE, 1–6. https://doi.org/10.1109/INDICON.2016.7839034

[14] Christopher H. Gibbons. 2019. Basics of autonomic nervous system function. In *Clinical Neurophysiology: Basis and Technical Aspects*, Kerry H. Levin and Patrick Chauvel (Eds.). Handbook of Clinical Neurology, Vol. 160. Elsevier, 407–418. https://doi.org/10.1016/B978-0-444-64032-1.00027-8

[15] Atefeh Goshvarpour, Ataollah Abbasi, and Ateke Goshvarpour. 2014. Impact of Music on College Students: Analysis of Galvanic Skin Responses. 35 (December 2014), 11–20. https://api.semanticscholar.org/CorpusID:73082965

[16] Alberto Greco, Gaetano Valenza, Luca Citi, and Enzo Pasquale Scilingo. 2017. Arousal and Valence Recognition of Affective Sounds Based on Electrodermal Activity. 17, 3 (February 2017), 716–725. https://doi.org/10.1109/JSEN.2016.2623677

[17] Wickham Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org

[18] Donghong Han, Yanru Kong, Jiayi Han, and Guoren Wang. 2022. A survey of music emotion recognition. 16, Article 166335 (January 2022), 11 pages. https://doi.org/10.1007/s11704-021-0569-4

[19] Makoto Iwanaga, Asami Kobayashi, and Chie Kawasaki. 2005. Heart rate variability with repetitive exposure to music. 70, 1 (September 2005), 61–66. https://doi.org/10.1016/j.biopsycho.2004.11.015

[20] Mariana C. Jacob Rodrigues, Octavian Postolache, and Francisco Cercas. 2023. The Influence of Stress Noise and Music Stimulation on the Autonomous Nervous System. 72 (June 2023), 1–19. https://doi.org/10.1109/TIM.2023.3279881

[21] Anurag Joshi and Ravi Kiran. 2020. Gauging the effectiveness of music and yoga for reducing stress among engineering students: An investigation based on Galvanic Skin Response. 65, 3 (2020), 671–678. https://doi.org/10.3233/WOR-203121

[22] Patrik N. Juslin and Daniel Västfjäll. 2008. Emotional responses to music: The need to consider underlying mechanisms. 31, 5 (October 2008), 559–575. https://doi.org/10.1017/S0140525X08005293

[23] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A Database for Emotion Analysis Using Physiological Signals. 3, 1 (January 2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15 Publisher: IEEE.

[24] Sylvain Laborde, Emma Mosley, and Julian F. Thayer. 2017. Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting. 8, Article 213 (February 2017). https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00213

[25] Guo-She Lee, Mei-Ling Chen, and Gin-You Wang. 2010. Evoked response of heart rate variability using short-duration white noise. 155, 1 (June 2010), 94–97. https://doi.org/10.1016/j.autneu.2009.12.008

[26] Alexander Lerch. 2023. Input Representation. In *An Introduction to Audio Content Analysis: Music Information Retrieval Tasks and Applications.* Wiley-IEEE Press, 17–89. https://doi.org/10.1002/9781119890980.ch3

[27] Xuejing Lu, William Forde Thompson, Libo Zhang, and Li Hu. 2019. Music Reduces Pain Unpleasantness: Evidence from an EEG Study. 12 (December 2019), 3331–3342. https://doi.org/10.2147/JPR.S212080

[28] Saranga Kingkor Mahanta, Abdullah Faiz Ur Rahman Khilji, and Partha Pakray. 2021. Deep Neural Network for Musical Instrument Recognition Using MFCCs. 25, 2 (May 2021), 351–360. https://doi.org/10.13053/cys-25-2-3946

[29] Dominique Makowski, Daniel Lüdecke, Indrajeet Patil, Rémi Thériault, Mattan S. Ben-Shachar, and Brenton M. Wiernik. 2023. Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption. (2023). https://easystats.github.io/report/

[30] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. 53, 4 (February 2021), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[31] Brian McFee, Matt McVicar, Daniel Faronbi, Iran Roman, Matan Gover, Stefan Balke, Scott Seyfarth, Ayoub Malek, Colin Raffel, Vincent Lostanlen, Benjamin van Niekirk, Dana Lee, Frank Cwitkowitz, Frank Zalkow, Oriol Nieto, Dan Ellis, Jack Mason, Kyungyun Lee, Bea Steers, Emily Halvachs, Carl Thomé, Fabian Robert-Stöter, Rachel Bittner, Ziyao Wei, Adam Weiss, Eric Battenberg, Keunwoo Choi, Ryuichi Yamamoto, C. J. Carr, Alex Metsai, Stefan Sullivan, Pius Friesch, Asmitha Krishnakumar, Shunsuke Hidaka, Steve Kowalik, Fabian Keller, Dan Mazur, Alexandre Chabot-Leclerc, Curtis Hawthorne, Chandrashekhar Ramaprasad, Myungchul Keum, Juanita Gomez, Will Monroe, Viktor Andreevitch Morozov, Kian Eliasi, nullmightybofo, Paul Biberstein, N. Dorukhan Sergin, Romain Hennequin, Rimvydas Naktinis, beantowel, Taewoon Kim, Jon Petter Åsen, Joon Lim, Alex Malins, Darío Hereñú, Stef van der Struijk, Lorenz Nickel, Jackie Wu, Zhen Wang, Tim Gates, Matt Vollrath, Andy Sarroff, Xiao-Ming, Alastair Porter, Seth Kranzler, Voodoohop, Mattia Di Gangi, Helmi Jinoz, Connor Guerrero, Abduttayyeb Mazhar, toddrme2178, Zvi Baratz, Anton Kostin, Xinlu Zhuang, Cash TingHin Lo, Pavel Campr, Eric Semeniuc, Monsij Biswal, Shayenne Moura, Paul Brossier, Hojin Lee, and Waldir Pimenta. 2024. *librosa*. https://doi.org/10.5281/zenodo.11192913

[32] Zhihang Meng. 2021. Research on timbre classification based on BP neural network and MFCC. 1856, Article 012006 (April 2021). https://doi.org/10.1088/1742-6596/1856/1/012006

[33] Helia Mojtabavi, Amene Saghazadeh, Vitor Engrácia Valenti, and Nima Rezaei. 2020. Can music influence cardiac autonomic system? A systematic review and narrative synthesis to evaluate its impact on heart rate variability. 39, Article 101162 (May 2020). https://doi.org/10.1016/j.ctcp.2020.101162

[34] C. Panagiotakis and G. Tziritas. 2005. A speech/music discriminator based on RMS and zero-crossings. 7, 1 (February 2005), 155–166. https://doi.org/10.1109/TMM.2004.840604

[35] Maria S. Perez-Rosero, Behnaz Rezaei, Murat Akcakaya, and Sarah Ostadabbas. 2017. Decoding emotional experiences through physiological signal processing. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (New Orleans, LA, USA). IEEE, 881–885. https://doi.org/10.1109/ICASSP.2017.7952282

[36] Tim Pohle, Elias Pampalk, and Gerhard Widmer. 2005. Evaluation of Frequently Used Audio Features for Classification of Music into Perceptual Categories. (2005).

[37] R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

[38] Peter J. Rentfrow. 2012. The Role of Music in Everyday Life: Current Directions in the Social Psychology of Music. 6, 5 (May 2012), 402–416. https://doi.org/10.1111/j.1751-9004.2012.00434.x

[39] Soha Salman, Junaid Mir, Muhammad Tallal Farooq, Aneeqa Noor Malik, and Rizki Haleemdeen. 2021. Machine Learning Inspired Efficient Audio Drone Detection using Acoustic Features. In *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)* (Islamabad, Pakistan). IEEE, 335–339. https://doi.org/10.1109/IBCAST51254.2021.9393232

[40] Michael Sampson and Anthony McGrath. 2015. Understanding the ECG. Part 1: Anatomy and physiology. 10, 11 (November 2015), 548–554. https://doi.org/10.12968/bjca.2015.10.11.548

[41] Fred Shaffer and J. P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. 5, Article 258 (September 2017). https://doi.org/10.3389/fpubh.2017.00258

[42] Mahima Sharma, Sudhanshu Kacker, and Mohit Sharma. 2016. A Brief Introduction and Review on Galvanic Skin Response. 2, 5 (December 2016), 254–257. https://doi.org/10.21276/ijmrp.2016.2.6.003

[43] Chang Sun Sim, Joo Hyun Sung, Sang Hyeon Cheon, Jang Myung Lee, Jae Won Lee, and Jiho Lee. 2015. The Effects of Different Noise Types on Heart Rate Variability in Men. 56, 1 (January 2015), 235–243. https://doi.org/10.3349/ymj.2015.56.1.235

[44] M Sinski, J Lewandowski, P Abramczyk, K Narkiewicz, and Z Gaciong. 2006. Why Study Sympathetic Nervous System. 57, 11 (November 2006), 79–92.

[45] Brian Tracey, Dmitri Volfson, James Glass, R'mani Haulcy, Melissa Kostrzebski, Jamie Adams, Tairmae Kangarloo, Amy Brodtmann, E. Ray Dorsey, and Adam Vogel. 2023. Towards interpretable speech biomarkers: exploring MFCCs. 13, Article 22787 (December 2023). https://doi.org/10.1038/s41598-023-49352-2

[46] Paul van Gent, Haneen Farah, Nicole Nes, and B. Arem. 2018. Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data, N. Van Nes and C. Voegelé (Eds.). NL HUMANIST publications.

[47] Björn Vickhoff, Helge Malmgren, Rickard Åström, Gunnar Nyberg, Seth-Reino Ekström, Mathias Engwall, Johan Snygg, Michael Nilsson, and Rebecka Jörnsten. 2013. Music structure determines heart rate variability of singers. 4, Article 34 (July 2013). https://doi.org/10.3389/fpsyg.2013.00334

[48] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[49] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. Welcome to the tidyverse. 4, 43, Article 1686 (2019), 6 pages. https://doi.org/10.21105/joss.01686

[50] Claus O. Wilke. 2024. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. https://CRAN.R-project.org/package=cowplot R package version 1.1.3.

[51] Qifei Zhang, Xiangwei Lai, and Guangyuan Liu. 2016. Emotion Recognition of GSR Based on an Improved Quantum Neural Network. In *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)* (Hangzhou, China, 2016-08-27). IEEE, 488–492. https://doi.org/10.1109/IHMSC.2016.66

[52] Huiyu Zhou, Abdul Sadka, and Richard M. Jiang. 2008. Feature extraction for speech and music discrimination. In *2008 International Workshop on Content-Based Multimedia Indexing* (London, UK). IEEE, 170–173. https://doi.org/10.1109/CBMI.2008.4564943