Does Culture Loop? An Exploratory Investigation of Spotify Audio Features in Hit Songs Over Time

Maria Nissen Byg (201906706)

Mia Jacobsen (201906908)

Characters: 45513

Code available on GitHub at: https://github.com/Kalkiara/data-science-exam

Abstract

Investigating a theory in popular media of 20-year cultural loops, this study uses the Billboard Hot 100 tracks from 1960 to 2022 and corresponding Spotify audio features to assess this notion. Specifically, the analysis builds on a preliminary set of visual assessments of cycles within each audio feature over time using autocorrelation and STL decomposition. Additionally, we create a series of harmonic regression models to assert the evidence for the hypothesis of 20-year loops and forecast past the range of our available data. We find that each audio feature seems to consist of both a general trend (either positive or negative) and some larger, cyclical patterns within the trend. Specifically for the hypothesis, the time series model with a 20-year Fourier component was the best model according to AICc for two out of six audio features, whereas for the remaining features, a 10-year model best captured the data. The paper concludes with a discussion of the methodology of the paper, specifically concerning whether 63 years of data is enough to adequately capture the actual size of underlying cultural loops, and whether the models are adequately capturing the nature of the available data.

Key words: Spotify audio features, harmonic regression, cultural loops, time series

1. Introduction (MJ)

On May 4, 2023, singer-songwriter Ed Sheeran won a lawsuit put against him by the heirs of Ed Townsend – the co-writer of the Marvin Gaye 1973 hit 'Let's Get It On'. Specifically, the heirs of Ed Townsend accused Sheeran of copyright infringement, as the chord progression in Sheeran's 2014 hit 'Thinking Out Loud' was similar to one in 'Let's Get It On' (Sisario, 2023). Over the past decade, the music industry has faced several of these lawsuits, starting with the 2015 hit 'Blurred Lines' being accused of plagiarizing another Marvin Gaye hit from 1973 (Sisario, 2023). As the music industry tries to navigate what constitutes basic building blocks of music, and what constitutes original ideas in need of copyright protection, it also begs the question: Why do we see these songs from 20-40 years ago being comparable enough to popular music today, that the cases can go all the way to court? Constituting a possible explanation, several popular media outlets report a so-called 20-year trend cycle, pertaining to the seeming repetition of culture in decade-long loops – be that fashion, music, or cinematography (Ewens, 2022; Friedman, 2022). Several popular media outlets speculate that the cause of such 20-year cycles pertains to nostalgification of a generation's youth as they enter adulthood and report the existence of such trend cycles as matter of fact (Baksh, 2022; Ulaby, 2022; Wickman, 2012). However, to our knowledge, only few scientific studies have attempted to look into the notion, all of which are arguably either outdated or limited in scope.

Specifically, two recent studies looked at the idea of fashion trends as they pertain to cultural transmission in society (Acerbi et al., 2012; Yoganarasimhan, 2017). They focused on why trends occur, i.e., the rise and fall of a single fashion trends like popular names or dog breeds, and why some might become stable parts of culture. Their results build on modeling the volatile dynamics of fashion trends (Acerbi et al., 2012), and on the algorithmic identification of cycles in data (Yoganarasimhan, 2017). These studies, however, do not take a more in-depth perspective on what these cycles mean for culture.

However, in a seminal 1975 study, Petersen and Berger investigated this notion of culture going through cycles by looking at popular music production. They looked at the market concentration of various firms and the level of innovation and diversity in popular music from 1948 to 1973. The authors found evidence for cycles in music production, with longer periods of homogeneity followed by brief bursts of competition and innovation leading to more diversity, until a new period of gradually increasing homogeneity started again (Petersen & Berger, 1975). However, as argued by Ross (2006), looking at the market concentration of record companies has become an inadequate way of investigating music diversity and homogeneity. This is both due to the new nature of music consumption – streaming platforms – and the current state of the market in which five companies own the vast majority of recorded music (Ross, 2006). There is therefore an argument to be made on how to investigate these cultural cycles in our modern world.

1.2 Spotify Audio Features (MNB)

Since 1958, the Billboard magazine has kept track of the most popular songs every week on their 'Hot 100' list (Petersen & Berger, 1975). It thus provides a systematic and reliable way of investigating which songs dominate the popular music scene over time. This becomes especially interesting with the addition of streaming platforms such as Spotify. As part of Spotify's metadata on the content of its platform, a range of audio features are provided for each track. These features consist of both quantitative musical concepts such as time signature, key, and tempo, but also features that Spotify themselves estimate based on characteristics of the track such as danceability and energy. Our analysis will focus on the second kind, and specifically the following features: Acousticness, danceability, energy, instrumentalness, speechiness, and valence.

Common to these six features is that they range from 0-1, where 1 describes a high degree of the feature in question and 0 describes a total lack of the feature. The specifics of how Spotify estimates their audio features are not publicly available information, and so we rely on the information provided by their developer site (Spotify, n.d.). Here, acousticness is described as a confidence measure of whether the track is acoustic; danceability uses characteristics such as beat stability, tempo, and rhythm to estimate how suitable a track is for dancing; energy seeks to describe the perception of intensity and activity in a track using loudness, tempo, general entropy, and more; instrumentalness is another confidence measure of whether a track contains vocals, excluding "ooh" and "aah" sounds; speechiness describes the presence of spoken words, where scores above 0.66 are likely to be tracks made entirely of spoken words, scores between 0.33 and 0.66 are likely to be a mix of music and spoken words, and scores below 0.33 are likely to be music and non-spoken-word tracks; finally, valence is a measure of musical positiveness conveyed by the track, i.e., how happy the song sounds (Spotify, n.d.). These six audio features are all algorithmically determined using a method developed by Berenzweig and colleagues (2004), and afterwards propagated by Spotify when they acquired The Echo Nest Company, which commercialized the algorithm (Whitman & Jehan, 2011). It claims to be able to simulate how people listen to music by incorporating music perception principles, psychoacoustics, and adaptive learning (Bourreau et al., 2022). Since the scores have (to our knowledge) not been evaluated by human evaluators, we will discuss the implications of using algorithmically determined audio features later in the paper.

Using these features, one can investigate both the current and historical cultural zeitgeist as it pertains to music. Previous research using the Spotify audio features has primarily focused on the former, by attempting to predict which songs will become hits using a range of machine learning techniques (Nijkamp, 2018; Gulmatico et al., 2022; Middlebrook & Sheik, 2019; Al-Beitawi et al., 2020;

Adeagbo, 2020; Georgieva et al., 2018). However, previous research has also looked into the relation between Spotify audio features and specific musical phenomena such as protest songs (Jiang & Jin, 2022), the Turkish music market (Pinarbaşi, 2019), the relation between dance music and reasons for listening (Duman et al., 2022), and so-called sleep music (Scarratt et al., 2023). Additionally, Elena Georgieva and Blair Kaneshiro (2018) used the Spotify audio features to describe the evolution of pop music from 1988 to 2018. They found valence and energy to be decreasing over time, but danceability to be increasing, and that after a large degree of similarity in 2008, a greater diversity in the audio features of the songs investigated was observed in 2018 (Georgieva & Kaneshiro, 2018). This finding of increased diversity is supported by Bourreau and colleagues (2022), who used the Spotify audio features and ordinary least squares regression to investigate whether the digitization of music – i.e., the introduction of platforms such as iTunes and Spotify – increased the homogenization of popular music. They found that based on weekly music charts from 1990-2015 from 10 different countries, acoustic diversity increased when music streaming became the dominant way of distributing music (Bourreau et al., 2022). This finding is supported by Bello and Garcia (2021), who found increased diversity over time, with diversity measured as distinct units such as songs, artists, and labels. Inspired by these frameworks, our paper seeks to look at the association between cultural evolution and cyclic patterns in time, specifically investigating how the audio features of trending music evolve using time series analysis and harmonic regression.

1.3 Time Series Analysis (MJ)

In order to capture changes in time with an underlying looping structure, various methods for analyzing time series data can be applied. Specifically, time series analyses can be conducted by fitting linear models with a time dimension, meaning that the time series outcome variable is linearly associated with other time series variables used as predictors (Hyndman & Athanasopoulos, 2021). Accordingly, time series linear models allow for variables such as trend and seasonality to inform the analysis depending on the characteristics of the data, which is what sets this modeling apart from regular linear modeling approaches (Hyndman et al., 2023). As such, trends and seasonality are some of the main contributors when analyzing time series data compared to data without a temporal component. These trend and seasonality components can be analyzed using a plethora of methods, including Seasonal-Trend decomposition using Loess (STL), dynamic harmonic regression, and autocorrelation functions.

1.3.1 Seasonal-Trend Decomposition Using Loess (MNB)

Originally developed by Cleveland and colleagues (1990), the method of STL decomposition allows for the fragmentation of the different components of a time series into trend-cycle, seasonality, and remaining noise. By splitting the time series into different underlying categories of patterns, a more thorough understanding of the data can be obtained, which may ultimately be associated with greater model fit and forecast accuracy. STL is generally considered versatile and robust and works by applying what is known as the Loess method for estimating nonlinear relationships. Any type and length of seasonality can be handled, seasonal components can change over time, and the researcher is allowed to control the rate of change for the season and smoothness of the trend-cycle. Furthermore, the method is robust to outliers (Hyndman & Athanasopoulos, 2021).

1.3.2 Dynamic Harmonic Regression (MJ)

Dynamic harmonic regression, in turn, builds on the mathematical property that periodic functions can be made up of different combinations of sine and cosine functions (Foley, n.d.). These are known as Fourier terms. The Fourier terms take a parameter K, which controls the smoothness of the patterns by defining the number of sine and cosine pairs. As such, smaller values of K are associated with a smoother cyclic pattern. This dynamic modeling approach allows for any seasonal length and is generally considered flexible. It has however been criticized for not allowing a flexible pattern of seasonality over time (Hyndman & Athanasopoulos, 2021). Such assumed fixedness can be a limitation for long time series data that may change over time, but is useful for the investigation of a fixed cyclic pattern length as in the current analysis.

1.3.3 Autocorrelation (MJ)

Besides linear regression time series models, one can use Autocorrelation Functions (ACFs) to look at how later data points may correlate to earlier ones. By measuring the correlation coefficient between lagged values of a time series, trends and seasonality can become apparent (Hyndman & Athanasopoulos, 2021). Data with a trend will show large and positive autocorrelations for small lags, with a slow decrease in positive values as the lags increase. With seasonality, the autocorrelations will be larger for seasonal lags, i.e., at lag points which are the multiples of the seasonal period, than for other lags (Hyndman & Athanasopoulos, 2021).

Using a combination of these three methods – STL decomposition, harmonic regression, and autocorrelation, we wish to investigate the prevalence of 20-year loops in Spotify audio features in hit songs. Instead of using the audio features as predictor variables, as most previous research has done, we will use them as outcome variables. In doing so, we will explore the hypothesis that musical components loop in approximately 20-year cyclic intervals, and try to otherwise determine the length, nature, and existence of these loops.

2. Methods (MNB)

The full analysis pipeline, including scraping script and dataset, is available on GitHub. See appendix 1. A combination of R version 4.2.2 (Core Team, 2020) and Python 3.9 (2020) was used.

2.1. Data Acquisition (MNB)

Data acquisition was carried out in Python 3.9 (2020) by combining top-ranked songs over time and available audio features, algorithmically defined by Spotify (n.d.). Using the link structures for the Billboard Hot 100 webpage (Billboard, 2023), the content of the monthly charts was scraped for the first date of every month ranging from the years 1960 to the end of 2022. This was done using the HTML parser BeautifulSoup (Richardson, 2014), in consortium with the Requests package (Reitz, 2023) accessing the relevant URLs. The monthly top charts were then fed into the Spotify Web API by use of the lightweight Python library Spotipy (Lamere, 2023). In total, 21790 unique data entries were acquired.

2.2 Preprocessing (MJ)

For preprocessing of the data, audio feature scores were aggregated per month, so that each month had one score per audio feature, which was the mean of the top 100 songs from that month. This left us with 756 data points. As often done in time series analyses (Engel, 2022), the aggregation allowed us to better capture the general trends across time and reduce some of the variability within each point in time.

2.3 Preliminary Analysis (MJ)

In order to properly investigate the data prior to modeling, each audio feature time series was plotted. A simple line plot was made in order to get an overview of the monthly aggregated data across time. Additionally, the data was plotted using the ACF with a maximum lag of the length of the data, corresponding to 756 months. This choice was made in order to visually assess whether the audio features behave in a cyclical manner as hypothesized, or if other larger or smaller patterns should be accounted for in modeling.

To further assess the patterns of the data prior to modeling, we applied STL decomposition to the various audio features. As mentioned in the introduction, this method allows us to plot the underlying components of the data, i.e. trend, seasonality, and the remaining noise, independently of each other. Assessing whether everything looked as expected, we gradually changed the window defining how rapidly the sea-sonal com-po-nent changes. We settled on setting it to 35, which satisfied the decomposition into a smooth trend, seasonal fluctuations, and remaining noise.

2.4 Main Analysis (MNB)

A testing and training split was made, of which the test set consisted of the last 120 data points, corresponding to 10 years of Billboard Hot 100 data aggregated by month, ranging from 2013 to 2022. The choice of the size of the split was made in order to balance the available data for training the models, while also keeping a sizable set for comparison when forecasting the data. The choice of the size of the dataset will be further considered in the discussion section.

Modeling was carried out on the training dataset using the TSLM function from the fable package (O'Hara-Wild et al., 2023). The data were fit using the trend and periodic Fourier terms at varying lengths as predictors for each of the six audio features. The period was set to 120, 240, and 300, corresponding to 10, 20, and 25 years respectively. This choice was made in order to investigate our hypothesis of the 20-year cultural cycles, and possible alternative hypotheses that could explain the loops in the data better. The wavenumber K within the period was varied in the range of 20, 30, 40, 60, 80, 120, and 150 for the different periods, corresponding to a maximum of half of the period, but allowed to vary down to only a few waves. All of the models were fit with and without a predictor for seasonality.

In order to reduce any remaining trend or cyclic pattern when plotting the residuals, various maximal models were also attempted. However, due to warnings of rank-deficiency when fitting more than one Fourier term as predictors, we settled on the simpler models made up of only one Fourier term predictor, despite the cyclic behavior in the residuals that remained unaccounted for. The choice was deemed reasonable for the purpose and scope of the current analysis. It however constitutes a major limitation to the inference of the analysis, which will be discussed later in this paper.

2.4.1 Model Comparison (MJ)

Model comparison was carried out using the corrected Akaike information criterion (AICc), further aided by additional model comparison metrics such as the adjusted R^2 and the regular AIC. The model comparison metrics are reported in table 1 in the results section, additionally detailing for which audio features the 20-year period was deemed the best fit for the data.

2.4.2 Prediction Against Testing Data (MJ)

For each of the fitted models pertaining to the 20-year hypothesis, a prediction was made using a forecast horizon of 120 months, corresponding to the 10 years' worth of testing data. The predictions were plotted against the true data, and accuracy measures were calculated for each of the predictions for the six audio features. Specifically, the predicted values were compared to the true data using the mean absolute error (MAE) and root mean squared error (RMSE) metrics.

2.4.3 Forecast Using Full Dataset (MNB)

After having tested and assessed the validity of the predictions for the model fits pertaining to our hypothesis, an additional exploratory analysis was carried out using the full dataset. For this exploration, forecasts were made into the future with a forecast horizon of 120 months, resulting in a monthly forecast for the years 2023 to 2032. Whilst using the full monthly dataset ranging from 1960 to 2022, we also decided to switch to a bottom-up driven analysis. We therefore set aside the hypothesis and utilized the best model for each feature as found in the model comparison. Each of the forecasts was plotted and will be discussed further in the discussion section of the paper.

3. Results

3.1 Results for The Preliminary Investigation (MJ)

Below, we report the results of the preliminary investigation, which mainly focuses on plotting the time series data and assessing its behavior across time. Figure 1 displays the aggregated monthly mean for each of the audio features, while figure 2 shows the ACF plots. Finally, figure 3 showcases the STL decomposition for each of the audio features.

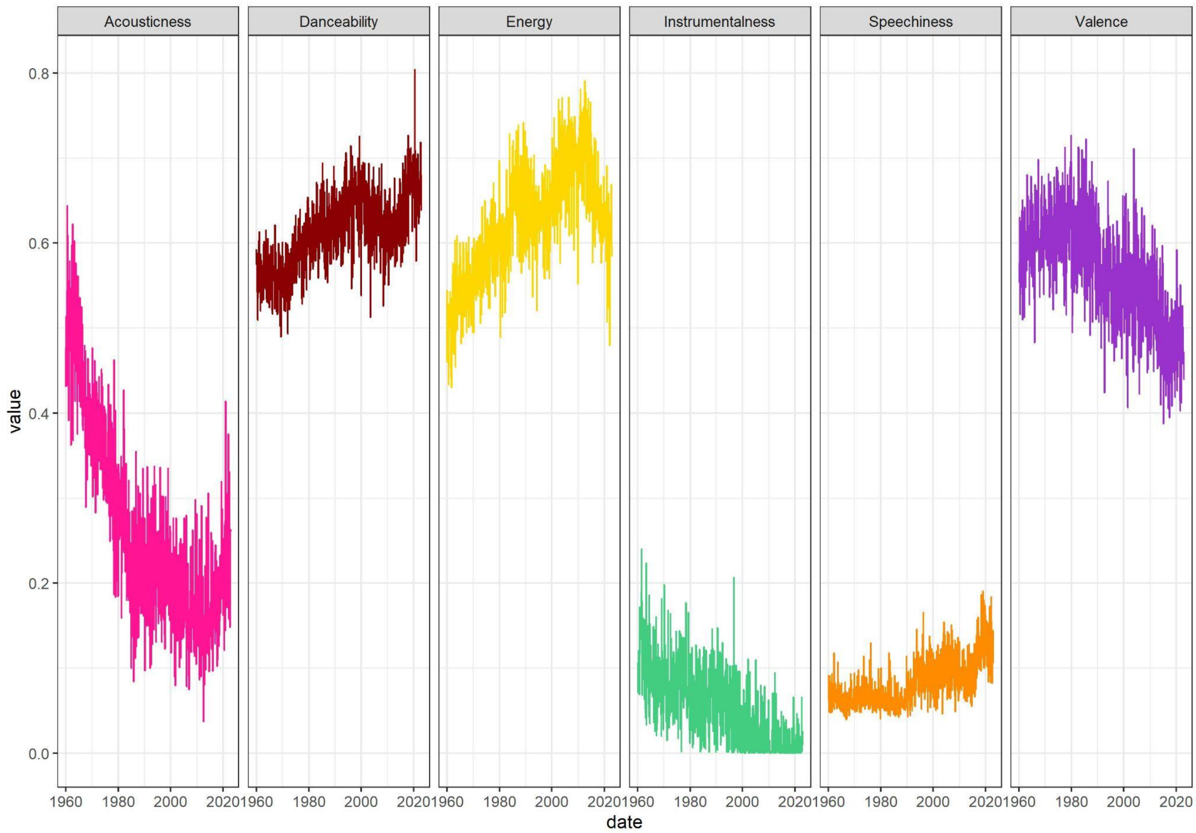

Figure 1: Monthly aggregated scores for each Spotify audio feature over time from January 1960 to December 2022.

From the lineplots of the aggregated monthly data, we are able to observe and compare the different audio features and their trends over time to one another. While some features, such as acousticness, vary greatly across time, speechiness is more stable, fluctuating less.

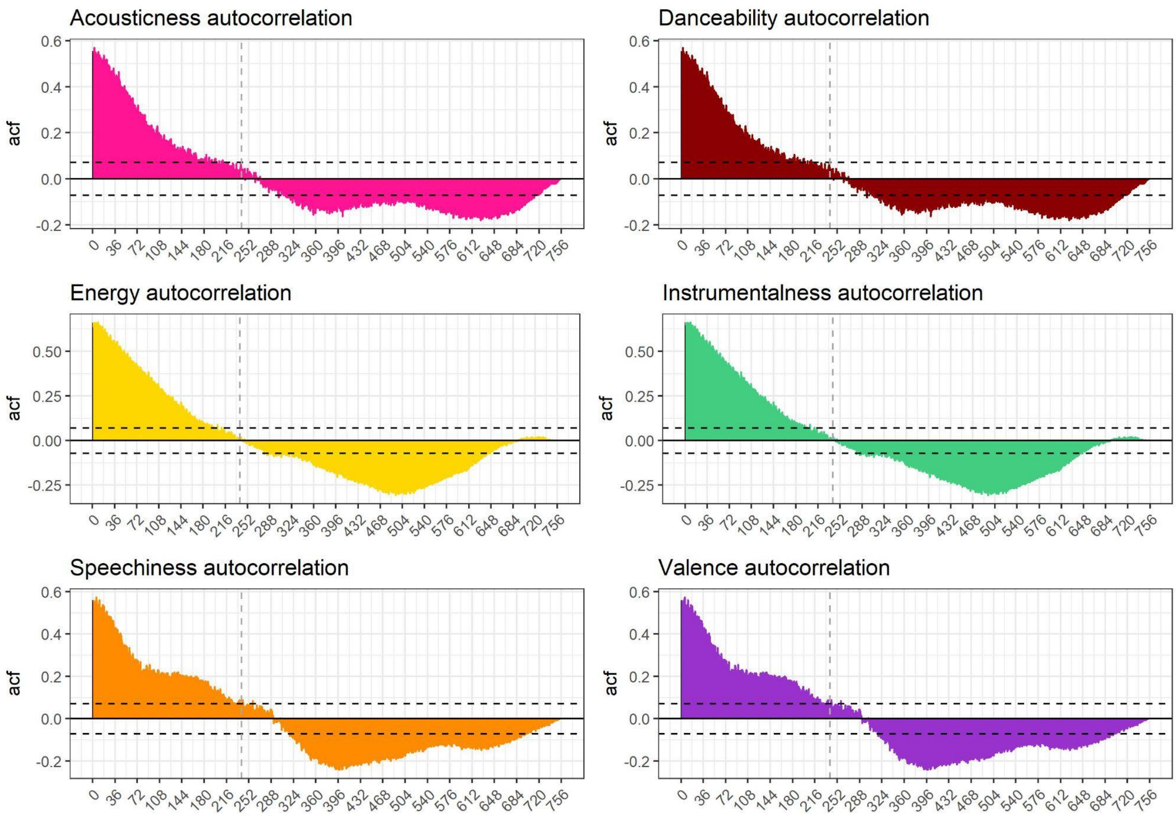

Figure 2: Autocorrelation function plots for each of the Spotify audio features with up to 756 months of lag.

The ACF plot, with each audio feature plotted in its own panel, showcases the trends in autocorrelation across time. For the first proportion of time for each of the features, the correlation starts quite high, after which it steadily declines. The vertical dashed line marks the first 20-year lag of the correlations corresponding to the hypothesis of the current investigation. At roughly this 20-year lag, the audio features start being negatively correlated with themselves. For the negatively correlated parts of the autocorrelation, the audio features energy and instrumentalness show a unimodal distribution, whereas the audio features acousticness, danceability, speechiness, and valence showcase bimodal distributions, suggesting a somewhat cyclic pattern. Note also that both of the unimodal negative correlations, energy, and instrumentalness, have started being positively correlated again during the most recent years. A similar trend seems to be observed for the rest of the audio features.

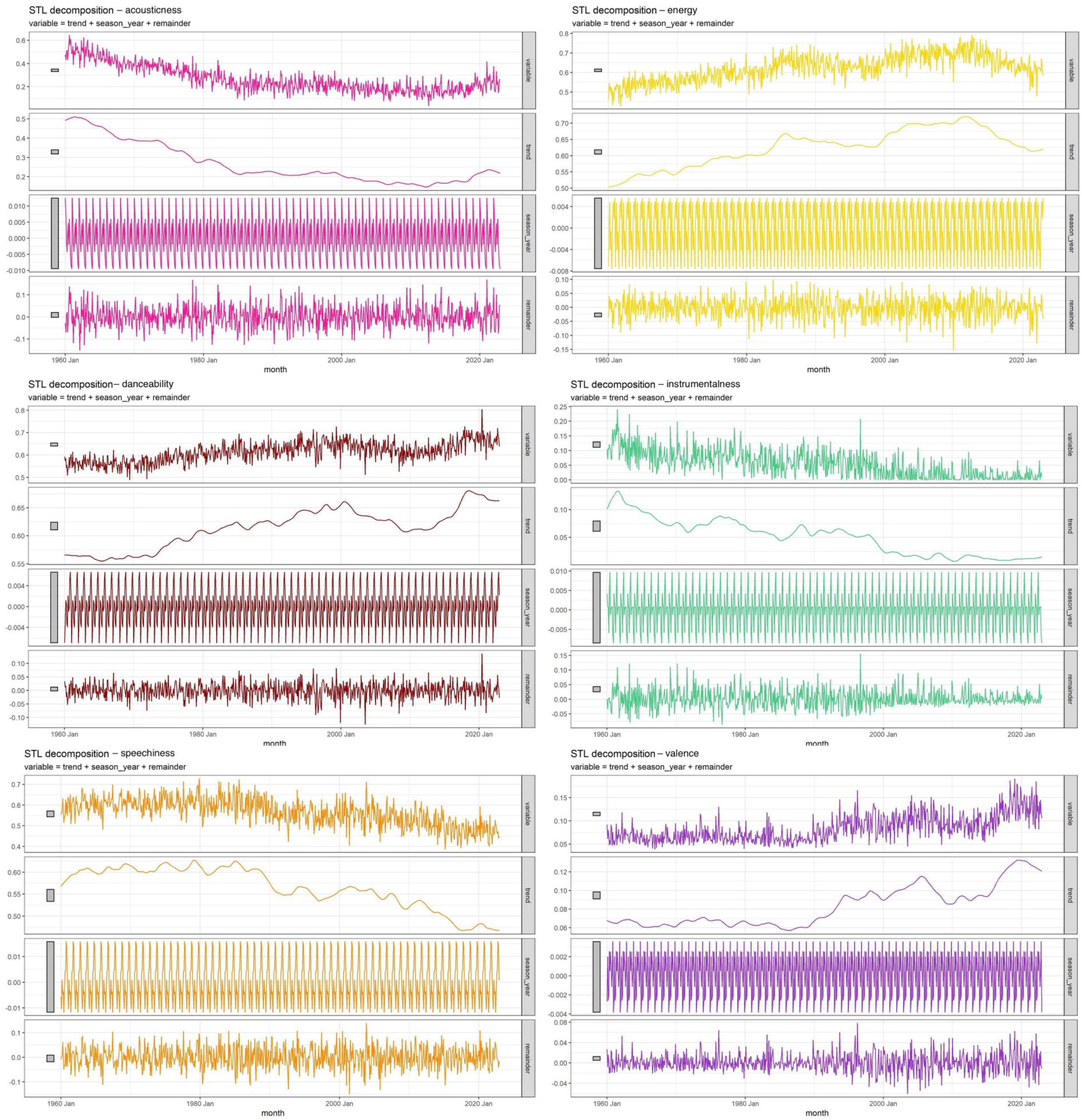

Figure 3: STL decomposition plots for each of the Spotify audio features. Window for seasonal components is set to 35.

For the STL decomposition plotting, we were able to extract a relatively smooth trend line when setting the window for the seasonal components to 35. As seen in the plots, some features exhibit a generally positive trend, whereas others exhibit a negative trend over time. For some of the audio features, specifically acousticness and danceability, the general trend seems to fluctuate steadily across time, suggesting a possible cyclic behavior. For all features, the seasonal component fluctuates rapidly, with quite short oscillations when compared to the general trend as marked using gray boxes.

3.2 Results for The Main Analysis (MNB)

After the preliminary analysis, we constructed a series of harmonic regression models for the main analysis. The models were used both to inform the 20-year hypothesis, and for subsequent prediction and forecasting. In order to assess which model was best, the models were compared using AICc and adjusted R^2.

3.2.1 Model Comparison Results (MJ)

We used the AICc for the main model comparison criteria, along with the adjusted R^2 value. The score for AICc and adj. R^2 for the top three models for each feature can be seen in table 1.

Model	**AICc**	**Adj. R^2**
acousticness ~ trend() + fourier(K = 20, period = 120)	-3484.4	0.72
acousticness ~ trend() + season() + fourier(K = 20, period = 120)	-3471.7	0.72
acousticness ~ trend() + fourier(K = 30, period = 240)	-3455.4	0.72
danceability ~ trend() + fourier(K = 30, period = 240)	-4269.8	0.40
danceability ~ trend() + season() + fourier(K = 30, period = 240)	-4251.0	0.39
danceability ~ trend() + fourier(K = 20, period = 120)	-4228.1	0.34
energy ~ trend() + fourier(K = 30, period = 240)	-3980.3	0.65
energy ~ trend() + fourier(K = 20, period = 120)	-3972.5	0.63
energy ~ trend() + season() + fourier(K = 30, period = 240)	-3960.8	0.65
instrumentalness ~ trend() + fourier(K = 20, period = 120)	-4185.1	0.38
instrumentalness ~ trend() + season() + fourier(K = 20, period = 120)	-4180.3	0.38
instrumentalness ~ trend() + fourier(K = 30, period = 240)	-4162.4	0.38

Model	**AICc**	**Adj. R^2**
speechiness ~ trend() + fourier(K = 20, period = 120)	-4936.7	0.28
speechiness ~ trend() + fourier(K = 30, period = 240)	-4928.0	0.30
speechiness ~ trend() + season() + fourier(K = 20, period = 120)	-4923.1	0.28
valence ~ trend() + fourier(K = 20, period = 120)	-3776.8	0.22
valence ~ trend() + season() + fourier(K = 20, period = 120)	-3764.4	0.22
valence ~ trend() + fourier(K = 30, period = 240))	-3758.8	0.23

Table 1 – AICc and adjusted R^2 for the top three models for each Spotify audio feature.

As seen in table 1, for two out of the six audio features, specifically danceability, and energy, the best scoring model in terms of AICc was one that used a Fourier predictor with a period corresponding to 20 years. For the remaining four of the six audio features, specifically acousticness, instrumentalness, speechiness, and valence, the best-scoring model had a Fourier predictor with a period corresponding to 10 years. For all of the best-performing models, the seasonal component was not included as a predictor.

3.2.2 Prediction Results (MNB)

To assess the evidence for our hypothesis of the 20-year cycles, we tested the model with a K of 30 and period of 240 for each audio feature on their ability to predict 120 months (10 years) into the future based on the training data. The model selection was based on both our motivated hypothesis, and the fact that the model with a K of 30 and period of 240 was in the top three for all audio features. We calculated the MAE and RMSE between the model predictions and the test data. The model predictions can be seen in figure 4, and the MAE and RMSE values for each audio feature can be seen in figure 5.

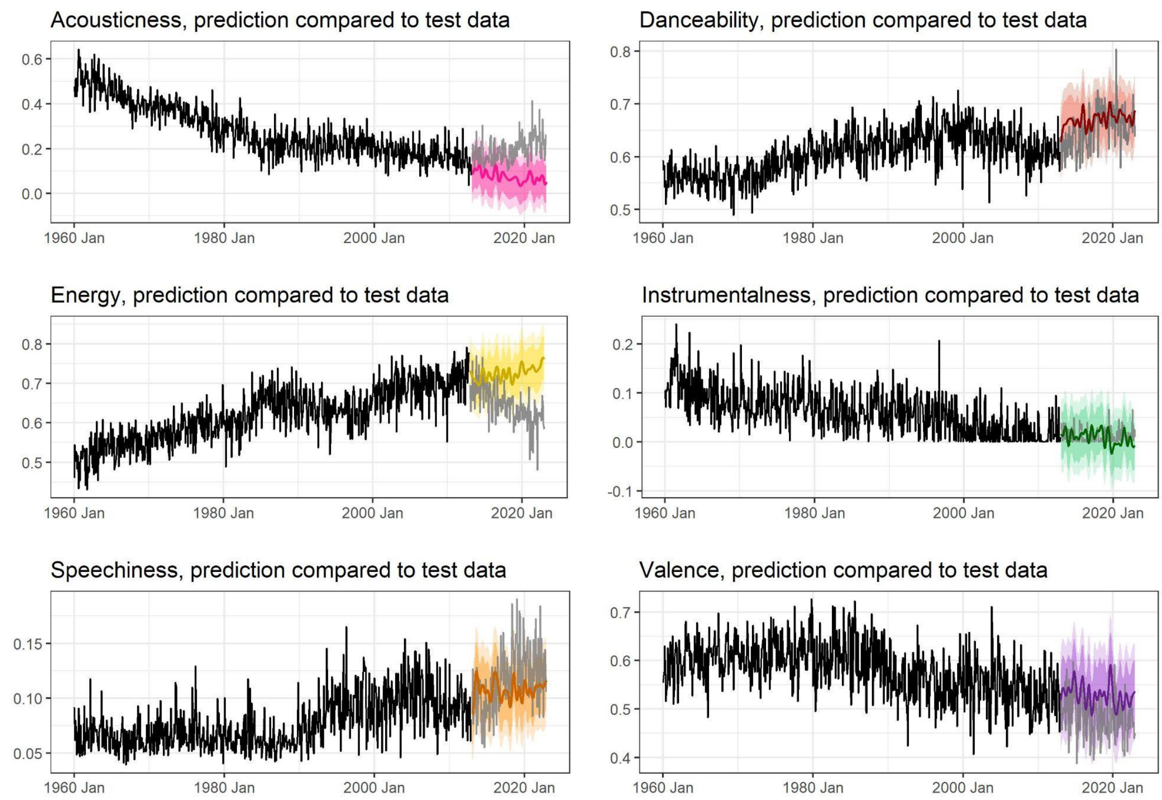

Figure 4: Monthly aggregated data for each Spotify audio feature until 2013 cutoff (black line), with 10-year prediction based on the best, motivated 20-year model (colored line), plotted against test data (gray line).

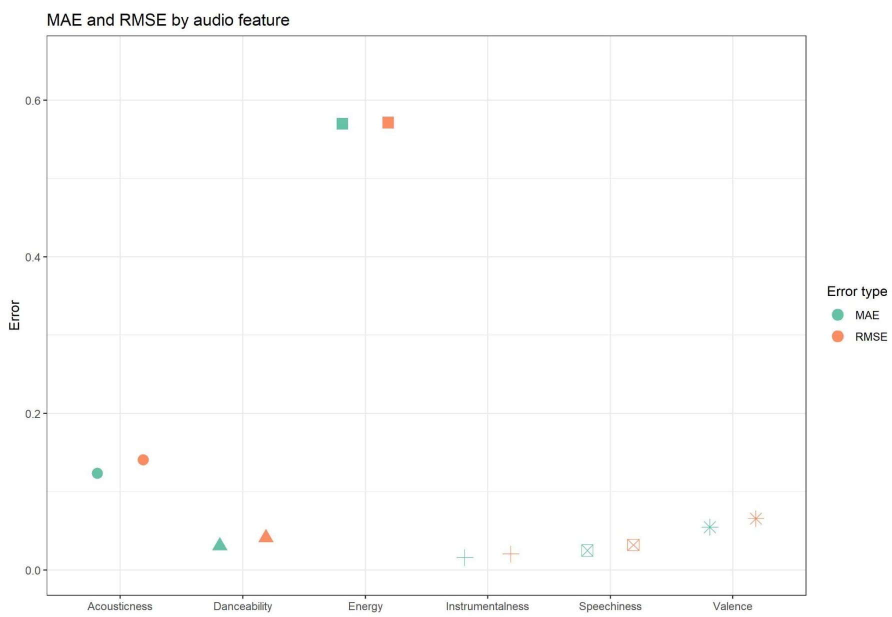

Figure 5: Mean absolute error (MAE) and root mean squared error (RMSE) for each of the Spotify audio features' model prediction.

As shown in figure 5, the accuracies of the forecasts generally lie in the realm of 0.02 to 0.14. For the 10-year prediction for energy, the model is not properly capturing the trend, resulting in error terms at 0.57 for both MAE and RMSE. This is further supported by the energy prediction plotted in figure 4.

3.3 Exploratory Forecasting (MJ)

Finally, our last analysis was an exploratory forecasting, to assess what each model estimates the future will look like. For this analysis, we used the best-performing model as determined by the AICc.

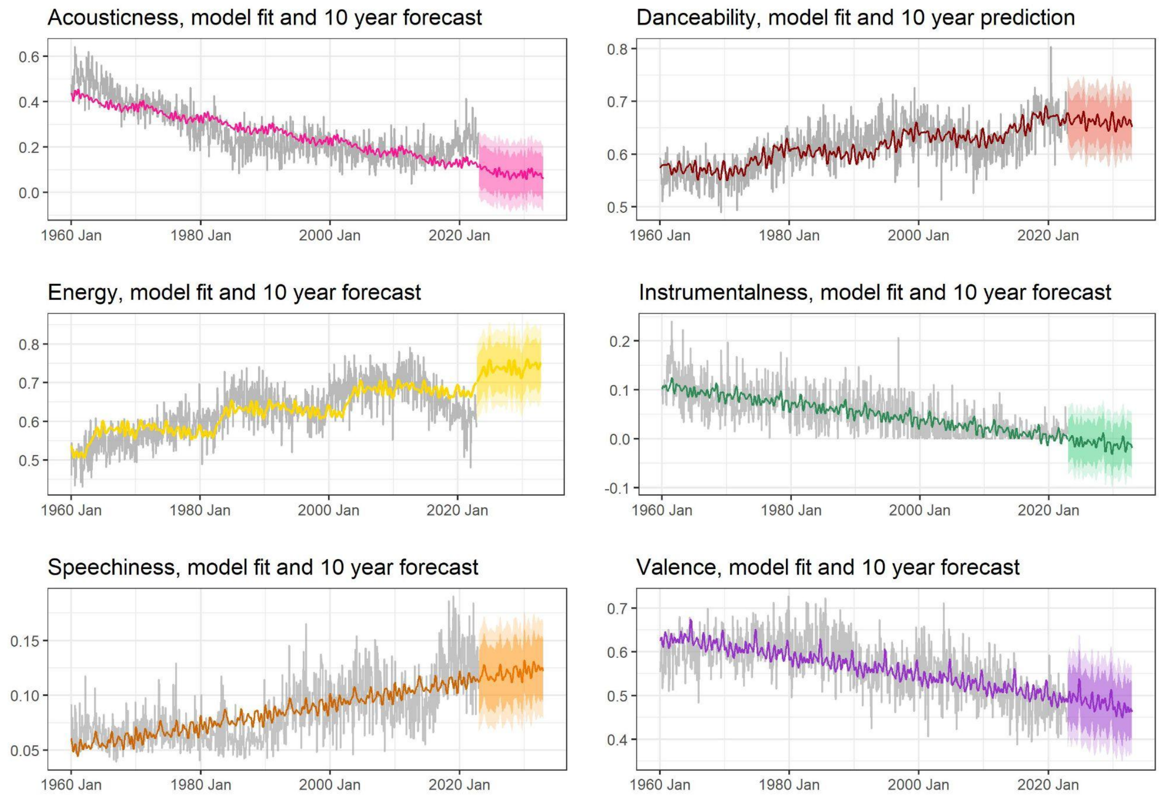

Figure 6: 10-year forecasting based on full dataset and the best model for each Spotify audio feature as determined by the AICc (colored line with prediction intervals), plotted against real data (grey line) and model fit (colored line).

Plotted in figure 6 is the 10-year forecast until 2032 using the best-performing models from the main analysis according to the AICc on the full dataset. The colored line represents the model fit, and the shaded area shows the uncertainty of the forecast. The light gray line shows the actual data until present day (end of 2022).

4. Discussion

4.1 Preliminary Analysis Findings (MJ)

From the data plotted in figure 1, each feature seems to have its own general trend containing a cyclical loop structure. Specifically, as danceability, energy, and speechiness generally increase in score over time, they also seem to have local peaks and troughs indicating a cyclical nature. The same goes for acousticness, valence, and instrumentalness, which show a similar pattern, however with a generally decreasing trend. This is similar to the findings of Georgivia and Kaneshiro (2018), although we find energy to be increasing over time up until 2010, where it starts to decrease again. Additionally, this interpretation of a general trend with a cyclical structure is supported by the STL decomposition. Here, the Loess smoothing makes the general trend with the cyclical pattern more evident. However, from the STL decomposition, the data mostly consists of these general trends and no seasonality, as the seasonality plots show miniscule, rapid peaks rather than the longer year- and decade-spanning cycles we hypothesized.

The autocorrelation plots also support these findings. As mentioned earlier, data with a trend will show a large, positive correlation at early lags with decreasing size as lags increase. This is the exact pattern of all the autocorrelation plots for all audio features. Additionally, for most features, when the 20-year mark is hit (as indicated by the vertical dashed line), the correlation switches sign and becomes negative. For energy and instrumentalness, it results in a new, negative peak, which towards the final lags seems to be going upwards again, possibly indicating the beginning of a new cycle. For acousticness, danceability, speechiness, and valence, two bumps can be seen, which indicates smaller cycles within the negative autocorrelation. However, there is also some indication that these features might be heading upwards as observed with energy and instrumentalness.

The preliminary analysis thus shows that the audio features generally seem to consist of cyclical loops within a larger trend. Additionally, the autocorrelation pattern suggests the existence of a larger cyclical pattern that extends beyond the range of data available for analysis. This implies that there may be additional cycles or trends in the data that are not fully captured or explored due to the limitations of the available data.

4.2 Main Analysis Findings (MNB)

As reported in table 1 in the results section, the best-performing models according to the AICc corresponded to our 20-year hypothesis for only two out of six audio features – specifically danceability and energy. The rest of the audio features, acousticness, instrumentalness, speechiness, and valence, were best captured with a period corresponding to only 10 years, i.e., half of what popular media tends to suggest. This indicates that the 20-year cycle hypothesis cannot account for the development of these audio features over time, at least not on its own. Drawing from the preliminary analysis findings as stated above, we however suspect a combination of differently spanning Fourier periods to better capture the data, as evident from both the STL and ACF plotting. This is further supported by the notion that our residuals were non-satisfactory as shown in appendix 2, in which some cyclic behavior seems to remain unaccounted for. As briefly mentioned in the methods section, adding several differently spanning Fourier terms to our models was attempted but consequently abandoned due to problems of rank-deficiency. This crucial limitation is further addressed in the section on limitations.

4.2.1 The Lacking Value of Short-Term Seasonality (MJ)

The fact that none of the best-performing models in terms of the AICc included a predictor variable dealing with seasonality outside of the Fourier predictor suggests that a meaningful seasonality term is not present in the data, at least not over the timespan available for the analysis at hand. This further ties into the discussed findings from the STL decomposition, in which we found only very rapid and short-term seasonal fluctuations. Apart from this notion, the implications of the seasonality predictor not adding value to the models seem to further indicate that a fixed cyclic pattern of seasonality is not properly able to capture the true behavior of the data, at least not on top of the seasonality already captured within the Fourier terms. Additionally, we emphasize that the interpretation of the seasonality should be assessed cautiously, as the duration of seasonality in harmonic regression is assumed fixed, as briefly mentioned in section 1.3 on time series analysis – a fixedness that does not necessarily correspond to the true behavior of the data spanning over the course of 63 years.

4.2.2 10-Year Prediction Accuracy (MNB)

For the attempted 10-year prediction compared to the test dataset, visual assessment leads to somewhat promising results for half of the analyzed audio features, specifically danceability, instrumentalness, and valence. As evident in figure 4, the colored lines seem to follow the true trend somewhat faithfully, however with much more narrow fluctuations, indicating that much of the fluctuation of the true data can be accounted for by noise. For the rest of the audio features, acousticness, energy, and speechiness, a more worrying divergence from the true data is observed in which the predicted values shown in colored lines separate from the true data in opposite directions. This may suggest that the true cycles span over longer periods of time than what is present in the training data, as the cutoff in the train and test split seems to fall more or less exactly at the point in which the trend changes direction.

When assessing the MAE and RMSE terms plotted in figure 5, the points from the visual assessment are further supported. While the accuracy of the predictions is generally low, in the range of 0.02 to

0.14, the sudden directional change of the energy time series results in a sizeable error term that make it clear that the fit cannot be used for accurate time series forecasting, at least not for the size of dataset currently available.

The relatively good fit for half of the audio features, despite limited training data, which likely does not span over the complete underlying trend cycle, holds promise for future research. We imagine that a larger dataset could entail an even better model fit and prediction accuracy when periods of varying lengths are successfully implemented in modeling.

4.3 Exploratory Future Forecast (MJ)

Building on the notion that the train and test split may severely limit the forecasting ability of the model fits, forecasting from the full dataset may improve the forecast accuracy into the future. Although the current attempt at forecasting into the future cannot be immediately assessed, we are still able to visually assess the fit and its plausibility.

As mentioned earlier, due to the exploratory nature of this part of the analysis, we decided to use the best model according to the AICc, regardless of the 20-year hypothesis. Visual assessment shows some mixed results. For speechiness, instrumentalness, and valence, the model fit primarily looks like a straight regression line with some minor, regular fluctuations. The models therefore do not seem able to capture the larger cycles we found in the preliminary analysis, and one should as such be cautious in trusting this forecast. This is especially relevant considering that instrumentalness is forecasted to go below 0, which is not possible as the features are bounded between 0 and 1.
A similar pattern emerges for acousticness, which however seems to show a smaller, cyclical pattern with 10-year bumps in its otherwise linear pattern. We however again caution against these predictions, primarily due to the slight upwards going line at around January 2020, which the model does not capture.

This inability to capture a divergence late in the data seems to recur in the energy model. Here, a clearer 20-year cyclic pattern occurs in the model fit. However, the model predictions take a sharp turn compared to the last ten years of the real data, making us hesitant that the model has adequately captured the nature of the data.

Lastly, the danceability model is the one model we visually assess to have a decent model fit and probable predictions. The 20-year cycles apparent in the model fit seem to correspond to cycles in the real data, and the predictions seem to follow naturally from where the true data stops.

Overall, this analysis has shown the complicated task of predicting culture and shows that not even 60 years of data can adequately capture the nature of how cultural evolution occurs as it pertains to music.

4.3 Limitations

4.3.1 Non-satisfactory Residuals (MNB)

A major limitation to the current investigation is the residuals of the fitted models, which we report in appendix 2. For all of the different audio features, a clear cyclic pattern persists in the residuals. This suggests that the predictors do not capture the entirety of the deterministic component of the data (Frost, 2017). In order to account for the cyclic pattern of the residuals, fitting the models with more than one Fourier term was attempted, as we identified the missing information to be related to loops at varying lengths. A variety of different maximal models was attempted, which did indeed produce more satisfactory looking plotting of the residuals. However, we were not able to successfully avoid problems of rank-deficiency when fitting the data, suggesting that our predictors were either perfectly correlated or that the dataset did not include enough observations compared to the model parameters (Zach, 2021).

Central to the discussion of the non-random residual plotting is then the consideration to add other simpler predictors that may capture the underlying pattern. For instance, one could imagine accounting for other trends or changes within society or the music industry coded as dummy variables. The exploration of non-time-based predictor variables was outside the scope of the current analysis but holds promise for future investigations of musical culture.

4.3.2 Using Spotify Audio Features (MJ)

As mentioned in the introduction, the Spotify audio features are algorithmically determined based on the characteristics of a given track. As mentioned in the introduction, the specifics of how these features are constructed is not publicly available information, and to our knowledge has not been evaluated by human evaluators. Some research has looked into the relation between the audio features and the music people report as ‘music to move to’ and found that these songs perceived as dance music generally have a higher danceability score as well (Duman et al., 2022). Although this partly verifies the robustness of the features, there might still be reason to be skeptical of the degree to which the audio features are musically meaningful (Heggli et al., 2021). Especially since other aspects of music listening behavior cannot be accounted for only using track characteristics – i.e., factors such as demography and time of day, which are known to influence listening behavior and perception of music (Heggli et al., 2021).

4.3.3 Size of Dataset (MNB)

The observation that the training set did not span what we assess to be the full cultural cycle of the audio features speaks to the choice of the training and testing split as mentioned in the methods section. Utilizing a cutoff for the most recent 10 years corresponds to roughly 16 percent of the full time series data. This is in contrast to the often-recommended 20 percent of a given dataset, ranging all the way to 30 or even 50 for older sources and textbooks (Tokuç, 2021). However, as the training set already spanned over too short a range of data to capture the changing directionality, the choice of a less sizable test set is to some extent merited. This justification can be drawn according to the discussed interpretation of the 10-year prediction accuracy. In fact, until more data becomes available, we estimate that the balancing of choosing a cutoff for the testing and training split could in principle have favored even more data for the training part than what was the case in our investigation.
Furthermore, the lack of data previous to the year 1960 can be argued to be both a limiting factor and an asset to the analysis. The Billboard Hot 100 began ranking songs in the end of the 1950s and thus constituted a natural beginning for the analysis at hand. Even though a longer-spanning dataset would have been preferable, we however also wish to stress the consequence that comes with using music that is in some sense outdated. For instance, the way in which we listen to music has changed over the decades, in part due to the rise of streaming. Thus, we also imagine that the analysis would be benefitted from adding predictors taking account of the shifting ways of music production and consumption.

4.4 Future Research (MJ)

Based on the results and limitations of the investigation at hand, we would recommend future research to better account for the varying cyclical patterns of the data in modeling. This could be done by identifying a way in which additional Fourier terms can be added while keeping the robustness of the analysis. Alternatively, the cyclic pattern may be accounted for by other complementary modeling approaches. We also strongly suggest adding other relevant predictors to be compared in the model comparison, for instance pertaining to the underlying mechanisms that may result in fluctuations other than the hypothesized repetition at fixed lengths of time.
Our finding that cycles may be spanning over the course of several decades, ranging almost the entirety of the dataset, further merits the conduction of future research. This however entails the availability of more data, either by identifying a way of expanding the data horizon into the past or by waiting a few more decades to investigate the suggestions found in this paper.
Other potential areas of research relating to the investigation at hand would be to look at the relation between the various audio features and how they correlate or impact each other. This was outside the scope and research question of our investigation but would be a relevant addition to the understanding of how audio features develop over time.

5. Conclusion (MJ & MNB)

This paper has investigated the lay theory that culture consists of 20-year cycles. Motivated by the increasing number of plagiarism lawsuits in the music industry, as well as the general lack of systematic research into this perception of cultural loops, we analyzed the Spotify audio features from the Billboard Hot 100 every month from 1960 to 2022. The analysis found some evidence for this cyclical nature across the six audio features chosen for the analysis at hand. For two out of the six features, the time series model with a 20-year Fourier component was the best model according to the AICc, whereas for the remaining features a 10-year model fit the data better. However, our analysis was severely limited, as both the data seemingly did not span enough years to adequately capture the longevity of these cultural loops, and because the residuals of our models suggest further predictors needed to be added. A case can therefore be made for future research to build upon the exploratory analysis framework proposed in the paper at hand, in order to both verify the robustness of the algorithmically generated Spotify audio features, as well as the cultural evolution over time as it pertains to popular music. Especially considering the fickle nature of culture, trends, and human behavior.

6. References

Acerbi, A., Ghirlanda, S., Enquist, M. (2012). The Logic of Fashion Cycles. *PLoS ONE,* 7(3): e32541. https://doi.org/10.1371/journal.pone.0032541

Adeagbo, A. (2020). Predicting Afrobeats Hit Songs Using Spotify Data. *arXiv preprint arXiv:2007.03137*.

Al-Beitawi, Z., Salehan, M., & Zhang, S. (2020). What makes a song trend? Cluster analysis of musical attributes for Spotify top trending songs. *Journal of Marketing Development and Competitiveness*, *14*(3), 79-91.

Baksh, J. (2022, April 12). The nostalgification of pop culture: Our inevitable return to the 2000s. The Michigan Daily. http://www.michigandaily.com/michigan-in-color/trend-cycles-in-response-to-politics-the-70s-2000s-2020s-and-their-recurring-pop-culture/

Bello, P., Garcia, D. (2021). Cultural Divergence in popular music: the increasing diversity of music consumption on Spotify across countries. *Humanit Soc Sci Commun* 8, 182. https://doi.org/10.1057/s41599-021-00855-1

Berenzweig, A., Logan, B., Ellis, D.P.W. & Whitman, B. (2004) A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*, 28(2), 63– 76.

Billboard. (2023). Hot 100, 1960-2022. Billboard.com. Retrieved April 19, 2023, from https://www.billboard.com/charts/hot-100/{year}-{month}-01/

Bourreau, M., Moreau, F., & Wikström, P. (2022). Does digitization lead to the homogenization of cultural content?. *Economic Inquiry*, *60*(1), 427-453.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. J. Off. Stat, 6(1), 3-73.

Duman, D., Neto, P., Mavrolampados, A., Toiviainen, P., & Luck, G. (2022). Music we move to: Spotify audio features and reasons for listening. *Plos one*, *17*(9), e0275228.

Engel, A. (2022, March 8). Weekly aggregation of time series data. Medium. https://towardsdatascience.com/weekly-aggregation-of-time-series-data-f9bdcc495a58

Ewens, H. (2022, December 14). Trends Used to Come Back Round Every 20 Years. Not Anymore. Vice. https://www.vice.com/en/article/bvmkm8/how-the-20-year-trend-cycle-collapsed

Foley, M. (n.d.). Chapter 6 Dynamic Harmonic Regression | Time Series Analysis. Retrieved May 17, 2023, from https://bookdown.org/mpfoley1973/time-series/dynamic-harmonic-regression.html.

Friedman, A. (2022). Will personal style put an end to the 20-year trend cycle?. A Magazine. https://theamag.com/6494/culture/will-personal-style-put-an-end-to-the-20-year-trend-cycle%ef%bf%bc/

Frost, J. (2017, April 5). Check Your Residual Plots to Ensure Trustworthy Regression Results! Statistics By Jim. http://statisticsbyjim.com/regression/check-residual-plots-regression-analysis/

Georgieva, E., & Kaneshiro, B. (2018). Using Spotify Audio Features to Study the Evolution of Pop Music. *WiMIR, 1st Annual Workshop*. Paris, France.

Georgivia, E., Suta, M., and Burton, N. (2018). Hitpredict: Predicting Hit Songs Using Spotify. *Data Stanford Computer Science 229: Machine Learning*.

Gulmatico, J. S., Susa, J. A. B., Malbog, M. A. F., Acoba, A., Nipas, M. D., and Mindoro, J. N. (2022) SpotiPred: A Machine Learning Approach Prediction of Spotify Music Popularity by Audio Features. *Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, Raipur, India, 2022, pp. 1-5, doi: 10.1109/ICPC2T53885.2022.9776765.

Heggli, O. A., Stupacher, J., & Vuust, P. (2021). Diurnal fluctuations in musical preference. *Royal Society open science*, *8*(11), 210885.

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed May 2023. Chapters 7 and 10.5.

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2023). _forecast: Forecasting functions for time series and linear models_. R package version 8.21, <URL: https://pkg.robjhyndman.com/forecast/>.

Jiang, Y., Jin, X. (2022). Using *k*-Means Clustering to Classify Protest Songs Based on Conceptual and Descriptive Audio Features. In: Rauterberg, M. (eds) Culture and Computing. HCII 2022. Lecture Notes in Computer Science, vol 13324. Springer, Cham. https://doi.org/10.1007/978-3-031-05434-1_19

Lamere, P. (2023). Spotipy (Version: 2.23.0) [Software]. Available from https://github.com/spotipy-dev/spotipy

Middlebrook, K, & Sheik, K. (2019). Song Hit Prediction: Predicting Billboard Hits Using Spotify Data. *arXiv preprint arXiv:1908.08609.*

Nijkamp, R. (2018). *Prediction of product success: explaining song popularity by audio features from Spotify data* (Bachelor's thesis, University of Twente).

O'Hara-Wild, M., Hyndman, R., Wang, E., implementation), G. C. (NNETAR, Bergmeir, C., Hensel, T.-G., & Hyndman, T. (2023). fable: Forecasting Models for Tidy Time Series (0.3.3). https://cran.r-project.org/web/packages/fable/index.html

Peterson, R. A., & Berger, D. G. (1975). Cycles in symbol production: The case of popular music. *American sociological review*, 158-173.

Python Software Foundation. (2020). Python (Version 3.9) [Software]. Available from https://www.python.org/downloads/release/python-390/

Reitz, K. (2023). Requests (Version 2.28.2) [Software]. Available from https://github.com/requests/requests

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richardson, L. (2014). Bs4. BeautifulSoup (Version 4.10.0) [Software]. Documentation at https://www.crummy.com/software/BeautifulSoup/bs4/doc/

Ross, P. G. (2005). Cycles in symbol production research: Foundations, applications, and future directions. *Popular Music and Society*, *28*(4), 473-487.

Scarratt, R. J., Heggli, O. A., Vuust, P., & Jespersen, K. V. (2023). The audio features of sleep music: Universal and subgroup characteristics. *PloS one*, *18*(1), e0278813.

Sisario, B. (2023, May 4). Ed Sheeran Won His Copyright Trial. Here's What To Know. *The New York Times.*

Spotify. (n.d.). *Get Track's Audio Features*. Retrieved May 16, 2023, from: https://developer.spotify.com/documentation/web-api/reference/get-audio-features

Tokuç, A. A. (2021, January 14). Splitting a Dataset into Train and Test Sets | Baeldung on Computer Science. https://www.baeldung.com/cs/train-test-datasets-ratio

Ulaby, N. (2022, March 1). From Tumblrcore to 2014core, the nostalgia loop is getting smaller and faster. NPR. https://www.npr.org/2022/03/01/1081115609/from-tumblrcore-to-2014core-the-nostalgia-loop-is-getting-smaller-and-faster

Yoganarasimhan, H. (2017). Identifying the Presence and Cause of Fashion Cycles in Data. *Journal of Marketing Research*, *54*(1), 5-26.

Whitman, B. & Jehan, T. (2011) Determining the similarity of music using cultural and acoustic information. *US Patent 8,073,854*. Available from: https://patents.google.com/patent/US8073854B2/en

Wickman, F. (2012, April 17). The 20-Year Nostalgia Cycle—or Is It 40 Years? 15? Slate. https://slate.com/culture/2012/04/the-golden-forty-year-rule-and-other-nostalgia-cycles-could-trends-possibly-return-every-40-years-20-years-and-12-15-years.ht

Zach. (2021, October 1). How to Fix: Prediction from a rank-deficient fit may be misleading. Statology. https://www.statology.org/prediction-from-rank-deficient-fit-may-be-misleading/

7. Appendices

7.1. Appendix 1

Code available on GitHub at: https://github.com/Kalkiara/data-science-exam

7.2 Appendix 2

The plotted residuals of each of the best models. Each one indicates some cyclic behavior that is not being accounted for by the model. However, the models were deemed appropriate for the analysis due to their theoretically motivated construction, and because of warning of rank-deficiency.