**Changes in American Popular Music (1950 - 2015): An Analysis Based on Billboard’s Hot 100**

Yuehan (Hailey) Liu

March 16, 2022

The current analysis aims to provide insights on the changes in musical features and song lyrics in American popular music from 1950 to 2015, based on data extracted from Billboard’s Hot 100 and the Spotify API. Two research questions are addressed in the following analysis: (1) how has American popular music changed in terms of musical features, and (2) how has American popular music changed in terms of song lyrics.

The study is divided into two major sections according to the two research questions. The first section is centered on the shifts in 9 musical features. Stacked bar plots grouped by Gender were produced to explicitly show the difference between year bins for each musical feature. Danceability and Energy seem to have similar fluctuations over the years, with the lowest scores in the 50s and highest in the 00s. The measurements of Loudness seem to peak from the 60s to the 90s, and experienced a sharp decline entering the millennium. Speechiness sustained a gradual increase from the 50s to the 90s, and demonstrated drastic increase in the 00s, followed by a sudden drop in the 10s. Acousticness is observed to diminish starting from the 60s, reaching its lowest point in the 10s. Instrumentalness witnessed a great leap from the 50s to the 60s, but dwindled quickly since its culmination in the 80s. Liveness achieved significant enhancement as it entered the 60s and maintained relatively steady. Valence and Tempo followed similar distribution pattern, which went through subtle variation after the 50s. When viewed vertically within each year bin, a gender disparity can also be observed. Male artists and bands take up a greater portion of the overall data, and therefore manifest more dramatic transformation over time. In addition to the bar plot demonstrations, violin box plots were also produced to provide complimentary information about the distribution and density of each numerical variable. However, the values of some variables like Speechiness and Instrumentalness are too centered and the outliers are too scattered for the plots to present useful results.

To synthesize the changes in different musical features, a scatterplot matrix and a correlation matrix of all variables were generated. Particularly, in the correlation matrix, several interesting relationships can be found. For instance, Energy has a strong positive correlation with Loudness, whereas a negative correlation with Acousticness. These results are relatively intuitive, considering the definition and scope of each musical feature. Furthermore, a Principal component analysis (PCA) was conducted on the musical features to reduce dimensionality of feature vectors and show the clustering situation. Generally speaking, the ellipses of musical features of each year bin overlap a lot, indicating no dramatic distinctions in size or dimension. However, it is still evident that the 60s seems to have the most scattered distribution, while the value points in the 10s are the most concentrated. This pattern indicates a potential homogenizing tendency in musical qualities throughout the decades.

In addition, in order to capture the predominant change in musical features in a more simplified way, the Cluster variable resulted from spectral clustering of song features was used to reflect chronological difference. Percentage distribution of the two clusters “String Lover” and “Poetic” of each year bin was shown in the filled bar plot. It can be interpreted that the String Lover cluster dominated the 50s through the 80s with decreasing proportion. Starting from the 90s, the Poetic cluster began to flourish, and basically became the prevailing majority in the 00s and 10s. This pattern corresponds to the gradual increase of Speechiness and the decline of Instrumentalness observed in previous bar plots.

Several reasons can be accounted for the shifting trends in musical features. First, these differences can be attributed to the iteration of music style and genre across time. The rise and fall of rock music, the emergence and popularization of hip-pop and pop music and other genre shifts have all contributed to the resulting phenomenon. Secondly, form of artist (solo singer or band) has also been through some changes. Rock bands appeared to be the general trend in the 1970s and 80s, whereas more and more singers decided to go solo after the 90s. Thirdly, the advancement of sound recording technology and the development in the entertainment industry might have also given rise to the discrepancy in musical features, but this hypothesis still requires more data and further research in order to be substantiated.

Switching gears to the second section of the study, the focus of analysis pivots to just the first column “lyrics” in the dataset. To perform a content analysis of the shifts in message and theme in song lyrics, several NLP approaches were adopted in this research. Some text preprocessing procedures were carried out to exclude punctuations, stop words, and so on. Looking at the sentence length distribution and POS percentage table, generally speaking, short sentences remain the most frequent in song lyrics throughout decades, and ADJ, ADV, and ADP continue to be the most dominant part-of-speech. But variations can still be found across different year bins. For example, corresponding to previous observations, as the Speechiness estimate increases, the limit of the y-axis in the sentence distribution histogram grows acutely as well (from 50 to 500), rendering all sentence frequencies in later decades (00s, 10s) markedly higher than previous ones.

The most interesting results came from the top 50 lemmas table and the topic modeling analysis. By comparing the top 50 lemmas tables from different year bins, a shift in choice of lexicon and style of expression in composing song lyrics can be observed. To be specific, there is a tendency of moving from a lyrical and reserved style to one that is physical and unrestrained, revealing a sense of sexual explicitness. Words like “bitch”, “fuck”, “ass” and “money” in the 2010s have replaced “sweet”, “kiss”, “lonely” and “tear” which prevailed in the 1950s and 60s.

Despite the aforementioned transition, when examined from a macro level regarding the overall themes and topics of the songs, such dramatic difference has disappeared. A LDA topic modeling was conducted on the whole document of lyrics to generate two most salient topics and their most weighted tokens. As it turns out, song lyrics in the past sixty years have been disproportionately concentrated on “love”, “baby”, “yeah”. Lyrics tokens that are most different between the two topics also did not show clear-cut separation of centralized themes. It seems like songs have been consistently expressing the feeling of “Yeah, I love baby girl”, but in notably different ways.

In conclusion, musical features of American popular songs have witnessed shifts in nearly all dimensions. But the most major changes are centered on the two measurements of Instrumentalness and Speechiness. The vocabulary and style of expression of song lyrics have transformed drastically over the decades, however, the contents and themes have not undergone much changes. The messages of American popular songs have homogeneously been about romantic relationships (or fantasies?) and affection for women. This study serves as a starting point for investigating the changes in American popular music industry and the cultural patterns and connotations these changes denote. Interesting questions can be raised concerning the dynamics of ideology behind the changes in musical features and lyrical content, as well as the cultural and social motives that might have stimulated the shifts in popular music.