THE CHANGING SOUND OF MUSIC: AN EXPLORATORY CORPUS STUDY OF VOCAL TRENDS OVER TIME

Elena Georgieva[1], **Pablo Ripollés**[1,2,3], **Brian McFee**[1,2,4]
[1] Music and Audio Research Laboratory, New York University,
[2] Center for Language, Music, and Emotion, New York University,
[3] Department of Psychology, New York University,
[4] Center for Data Science, New York University
{elena, pripolles, brian.mcfee}@nyu.edu

ABSTRACT

Recent advancements in audio processing provide a new opportunity to study musical trends using quantitative methods. While past work has investigated trends in music over time, there has been no large-scale study on the evolution of vocal lines. In this work, we conduct an exploratory study of 145,912 vocal tracks of popular songs spanning 55 years, from 1955 to 2010. We use source separation to extract the vocal stem and fundamental frequency (f_0) estimation to analyze pitch tracks. Additionally, we extract pitch characteristics including mean pitch, total variation, and pitch class entropy of each song. We conduct statistical analysis of vocal pitch across years and genres, and report significant trends in our metrics over time, as well as significant differences in trends between genres. Our study demonstrates the utility of this method for studying vocals, contributes to the understanding of vocal trends, and showcases the potential of quantitative approaches in musicology.

1. INTRODUCTION

Current technologies for audio processing provide new opportunities to study musical trends using quantitative methods. While researchers have analyzed music for generations, studying the evolution of music at a large scale has only been possible recently, due to the availability of large datasets [1–3]. Additionally, recent improvements in source separation technology have allowed researchers to study individual instruments [4, 5]. However, the vocal lines of songs have been understudied, even though they are often the most salient part of a song [6, 7], and many popular songs are built around the vocal line.

In this study, we examine trends in the vocal lines of 145,912 songs over 55 years (from 1955 to 2010). We use modern source separation methods to isolate vocal lines of songs (30–60 second-long excerpts) from their respective accompaniments. Altogether, our dataset makes up over 59 days of continuous listening. This work is exploratory: we examine what trends and patterns can be observed from such a large corpus of vocal data. We have made our list of track IDs publicly available, along with our implementations. [1]

2. RELATED WORK

The transformation of music over time has received a lot of focus in recent years. This is partially thanks to the release of open-source resources such as The Million Song Dataset (MSD) [1]. The MSD is a free collection of audio features and metadata for one million contemporary music tracks. Datasets such as MSD allow researchers to quantitatively analyze patterns in music at a large scale.

Serrà *et al.* used musical 'codewords' based on MSD clips to identify changes in pitch, timbre, and loudness over time [2]. They found that newer songs have less variety in pitch transitions, more homogenized timbres, and increased loudness. Parmer *et al.* did similar work using the MSD to study musical complexity from 1960–2010. They found that pitch complexity has been generally stable over that time period, while loudness and rhythm complexity has decreased and timbral complexity has increased [8]. Parmer also studied the complexity of popular songs from the Billboard chart, [2] and found that the complexity of popular songs is concentrated around the mean complexity level of all songs. This supports the inverted U-shaped model for music complexity and likeability: that listeners prefer intermediate levels of complexity [9, 10].

Another team of researchers used the MSD songs along with quantitative modeling to study musical influence: the impact that a particular artist has on the music by other musicians [3]. They identified clusters of songs that were indicative of a genre, and studied how those clusters evolved over time. A different study used a corpus of 17,000 songs from Billboard to study the "Evolution of Popular Music" between 1960 and 2010 in the United States [11]. They used timbral and harmonic features derived from Billboard songs, and identified three musical stylistic revolutions in

[1] https://github.com/elenatheodora/ismir2024-changing-sound-of-music
[2] https://www.billboard.com/charts/hot-100

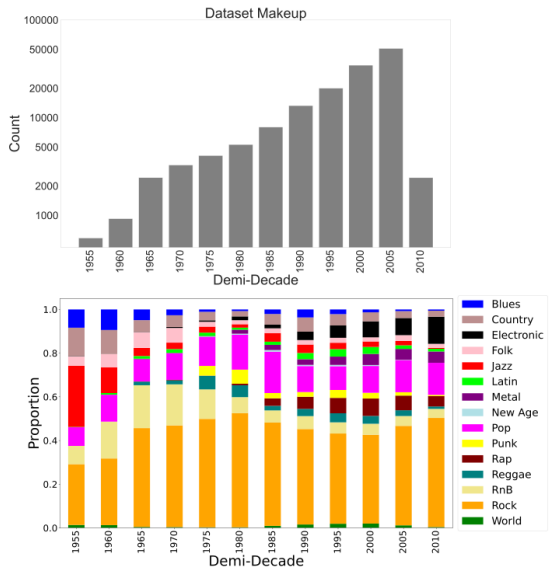

Figure 1: Top: Chronological distribution of the dataset organized in 5-year demi-decades. Bottom: Relative distribution of genres in the dataset by demi-decade.

Musical Genre & Number of Tracks					
Rock	64,203	Country	6,166	Punk	3,444
Pop	19,560	Metal	5,841	Latin	3,352
Rap	8,755	Reggae	4,689	Blues*	2,359
Electronic	8,754	Jazz	3,953	World*	2,044
RnB	8,647	Folk	3,630	New Age*	515

Figure 2: Number of tracks per musical genre. Blues, World, and New Age music (labeled '*'), were excluded from the by-genre analyses due to lower track count.

1964, 1983, and 1991. Other researchers have studied a more niche topic in detail over decades, including the evolution of a single band's performances [12], changes in dynamics/compression in mainstream music [13], or spectral characteristics of recordings over time [14].

In an early vocal corpus study, in 1959, Alan Lomax's Cantometrics project analyzed over 4,000 traditional vocal music songs from 400 cultures [15]. Researchers listened to songs and labeled them with 37 "style-factors," for example group cohesion in singing, and tense or relaxed vocal quality. The Cantometrics project suggested a correlation between song style and social norms of cultures.

In a more recent study, researchers developed a set of features to capture pitch and melodic embellishments of world vocal performances [16]. Using these features, they trained a classifier to distinguish vocal from non-vocal segments and learn a dictionary of singing style elements. Results showed that clusters were distinguished by characteristic uses of singing techniques such as vibrato and melisma. A different study categorized a collection of 360 Dutch folk songs, and found that the aspects of melody that are important for establishing similarity are contour, rhythm, and motifs [17]. Despite these previous works on vocal datasets, there has been no large-scale study on the evolution of the vocal lines of popular music over the years.

3. DATASET

We used a subset of the MSD [1] that has genre labels (the Tagtraum MSD annotations [18]). 278,619 tracks had genre labels available. Next, a group of songs was dropped due to a low presence of vocals in the excerpt, indicated by a low ratio of RMS (root mean square) energy of the separated vocal stem to RMS energy of the full audio file (see 4.1). Songs that did not have the release year available were also dropped. In a final filtering step, we chose to conduct analyses only starting in the year 1955, as data was sparse before 1955. The final dataset had 145,912 songs.

Figure 1 shows a chronological distribution of songs in demi-decade bins (i.e., 1990-1994). We observe a strong bias towards more recent songs. A relative distribution of genres across years shows fewer genres in earlier years, with a greater variety in more recent years. Figure 2 lists the number of tracks in each musical genre in the dataset. Blues, World, and New Age music (labeled with a '*'), were excluded from the by-genre analyses due to having a lower number of tracks.

Our dataset inherits biases from the MSD. The tracks in the MSD were selected based partly on their association with 'familiar' artists, as determined by The Echo Nest, followed by inclusion of tracks from similar artists. [3] The creators of the MSD also included artists that fit the 200 most frequently-occurring Echo Nest descriptive terms, as well as songs that were extreme in acoustic attributes. In general, songs in the dataset are generally widely listened-to, and the majority come from North America or Europe. There are much more data in recent years (1990s onward) than in earlier years. There is very little non-western and classical music in the dataset. The Latin music genre does contain non-western music, primarily performed in Spanish or Portuguese. Our findings apply to this dataset, not necessarily to music as a whole, and our work will have biases if applied to other datasets. Importantly, these dataset biases do not affect our methods.

4. METHOD

4.1 Source Separation

First, we used source separation to separate the vocal line of each song from the mix. For this, we use Hybrid Transformer Demucs (HT Demucs), a hybrid temporal/spectral bi-U-Net [5]. After computing the ratio of the vocal stem's RMS energy to the overall mix's RMS energy, we excluded any songs with a ratio below 0.08 (Figure 4). This ratio was set using a preliminary sub-sample of the data. These excerpts are either purely instrumental songs (non-vocal), or the clip happens to capture a part of the audio file with very few or no vocals (i.e., a guitar solo).

4.2 Pitch Characteristics

To study pitch characteristics, we did fundamental frequency (f_0) estimation on the estimated vocal stems using PYIN [19] as implemented in Librosa v0.8.1 [20].

[3] http://millionsongdataset.com/faq/

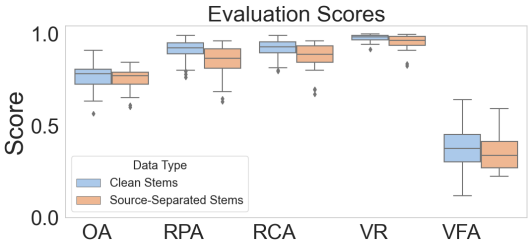

Figure 3: f_0 extraction evaluation scores for 29 clean and source-separated vocal stems from MedleyDB, when compared with the MedleyDB f_0 annotations.

We set the lower frequency limit at 70Hz and the upper limit at 900Hz, aligning with the human vocal range described in other works, while also extending one musical whole step in each extreme [21]. We chose PYIN over CREPE, another f_0-estimator, as it allows us to set a lower and higher pitch bound for f_0-estimation [22]. Other than changing the sampling rate to 44.1 kHz, we used the Librosa defaults: frame length 2048, hop length 512, number of thresholds for peak estimation 100, switch probability 0.01, and no-trough probability 0.01. We collect an f_0 estimate approximately every 12 milliseconds.

We evaluated the pitch tracking accuracy of the PYIN algorithm on source-separated audio by running PYIN on 29 monophonic vocal stems from MedleyDB. We used mir_eval to compute the standard evaluation metrics used in MIREX: Voicing Recall (VR), Voicing False Alarm (VFA), Raw Pitch Accuracy (RPA), Raw Chroma Accuracy (RCA) and Overall Accuracy (OA) [23]. First, we compared several different PYIN settings: the number of thresholds, switch probability, and no-trough probability parameters, and found the Librosa defaults performed among the best. Next, we compared the accuracy of PYIN on clean stems and source-separated stems, each respectively compared to the annotations included in MedleyDB, and observed only a small decrease in accuracy. The median evaluation metrics of our method on the 29 clean vocal stems were: for clean stems, OA 0.781, RPA 0.924, RCA 0.928, VR 0.984, VFA 0.375, and for source separated stems, OA 0.771, RPA 0.865, RCA 0.888, VR 0.964, VFA 0.340 (see Figure 3). The source separation process only slightly reduces the accuracy of our f_0-estimation.

Some tracks in the dataset have vocal harmonies. PYIN tends to track the pitch of the most prominent voice. We ran a query on last.fm, [4] and found that tags for vocal harmonies are present in less than 1% of songs in the dataset. We assume that the presence of vocal harmonies is uncorrelated with the variables we study: time and genre. Through informal listening, we found that Demucs and PYIN were comparably effective for older and newer audio recordings from the time period we study, 1955-2010.

PYIN also provides a voicing detection estimate, which we used to identify contiguous regions of pitched sound in the vocal stem. We converted f_0 values in hertz to cents

[4] https://www.last.fm/

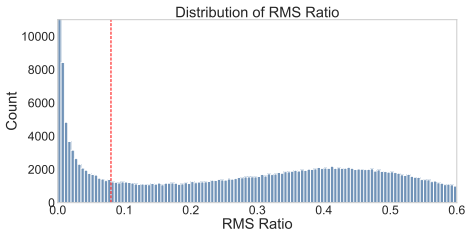

Figure 4: Distribution of the ratios of the vocal stem RMS energy to the full mix RMS energy in the data. A threshold of 0.08 was used to discard non-vocal clips.

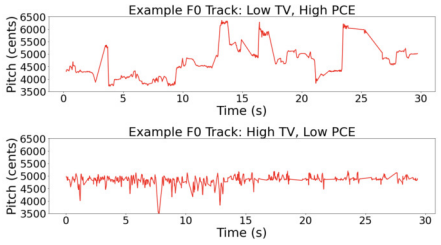

Figure 5: Example f_0 tracks. The top track, selected from the Metal genre, has a low TV and high PCE. The bottom track, from the Rap genre, has a high TV and low PCE.

using Eqn (1), where 16.35Hz is the frequency of C0:

$$f_0[¢] = 1200 \cdot \log_2 \frac{f_0[\text{Hz}]}{16.35}. \quad (1)$$

For example, "middle C" on a piano is 4800 cents, and $C\#/Db$ is 4900 cents. Using this information we extracted pitch features, dropping unvoiced frames. We calculated mean pitch (in cents) of each song, defined as the mean of each f_0 array.

We also calculated total variation (TV) [16]. TV summarizes the rate of pitch change and is defined in Eqn (2):

$$\text{TV}(x) = \frac{1}{N} \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (2)$$

for a given f_0 contour $x = (x_1, \ldots, x_N)$. TV is calculated independently for each voiced region within a song and then aggregated to a single total. Our TV calculations do not change the time interval between f_0 values.

4.3 Pitch Class Entropy

We calculated pitch class entropy (PCE) to measure the degree of unpredictability for the set of vocal pitches. Entropy was calculated over the probability of occurrence of each pitch class (independent of octave) in the vocal line [24]. Higher values of PCE indicate a greater spread in the pitch distribution, while lower values indicate a smaller and more predictable set of pitches. There is a theoretical maximum PCE of $\log_2(12) \approx 3.59$, achieved by a uniform distribution of the 12 pitch classes.

Figure 5 illustrates two example f_0 tracks with somewhat extreme TV and PCE values.

4.4 Statistical Analyses

We used R (4.2.2) and RStudio (2022.12.0+353) to implement linear regression with the *lm* function. Post-hoc tests

were implemented using the *emmeans* package with Tukey correction for multiple comparisons.

5. EXPERIMENT AND RESULTS

For each of our variables of interest (mean pitch, TV, PCE), we followed the same procedure. We first ran a linear regression to examine the relationship between the variable of interest (e.g., TV) and the year of track release (e.g., TV$\tilde{\text{y}}$ear). We then calculated a linear regression between the variable of interest and musical genre (e.g., TV$\tilde{\text{g}}$enre). Finally, we calculated independent linear regressions between the variable of interest and year of track release for the twelve most frequently-occurring genres. We calculated independent regressions because each of the genres becomes prevalent in the dataset during different years (i.e., Rap music starting in 1984).

When looking at musical genres, we chose to study the twelve genres with the most song entries in the dataset. We began analyzing each genre at the first year of a five-year period where at least ten songs were released in that genre annually. The twelve genres with corresponding start years were as follows: Country (1956), Electronic (1979), Folk (1963), Jazz (1955), Latin (1986), Metal (1980), Pop (1961), Punk (1977) Rap (1984), Reggae (1972), RnB (1957), and Rock (1956).

5.1 Mean Pitch

We found a significant positive relationship between mean pitch for a track and the year it was released (β = 0.957, t = 7.23, p < .001). For every one-year increase in the release year, the mean pitch of the track increased by a little less than one cent, on average (see Figure 6).

Next, we assessed mean pitch and musical genres. We found a significant main effect of genre (F(1, 140982) = 1378.6 , p < 0.001). All genres were significantly distinct (all p values <0.001) except for: electronic and pop music (t=1.891, p=0.765), jazz and Latin (t = 0.276, p=1.000), jazz and rock (t=-2.854, p = 0.158), and Latin and rock (t = -3.005, p = 0.107). Data for the mean pitch per song in each of these genres is illustrated in Figure 7.

We found a significant main effect of year for nine of the twelve musical genres, though the direction of the trends varied (see Figure 8). Specifically, country music (β = 3.991, t = 7.970, p <0.001), folk music (β = 1.374, t = 2.395, p <0.001), jazz (β = 2.901, t = 4.482, p =0.017), metal (β = 7.269, t = 4.612, p <0.001), punk (β = 3.301, t = 3.815, p <0.001), reggae (β = 2.053, t = 3.301, p <0.001) and rock (β = 1.097, t = 5.529, p <0.001) showed a significant positive relationship between year and mean pitch. Conversely, rap (β=-6.653, t=-7.757, p<0.001) and RnB (β=-3.800, t=-11.75, p<0.001) showed a significant negative relationship between year and mean pitch. No significant effect was found for electronic, Latin, or pop music.

5.2 Total Variation

The results for the TV and year regression between TV and year showed a significant negative relationship (β = -

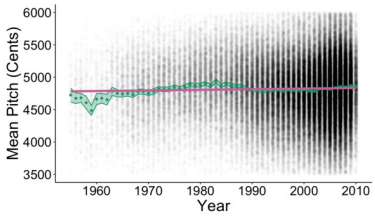

Figure 6: Mean pitch in cents as a function of year globally. Each dot represents a song. The red line represents the predicted slope with 95% confidence intervals. The green diamond and ribbon represent the mean per year and the standard error. This relationship was significant, with mean pitch increasing by approximately one cent per year.

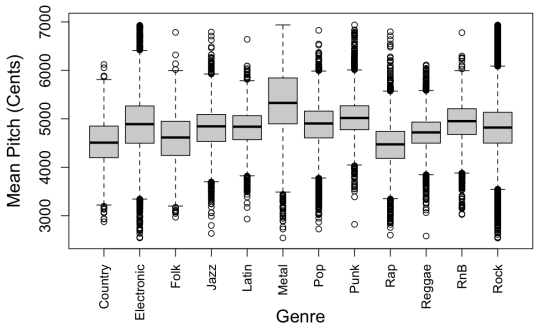

Figure 7: Mean pitch per song in each of the twelve genres across the dataset. Means are shown with boxes representing the interquartile range, error bars indicating the 95% confidence interval, and outliers as circles. There were significant differences between all genres except electronic and pop, jazz and Latin, jazz and rock, and Latin and rock

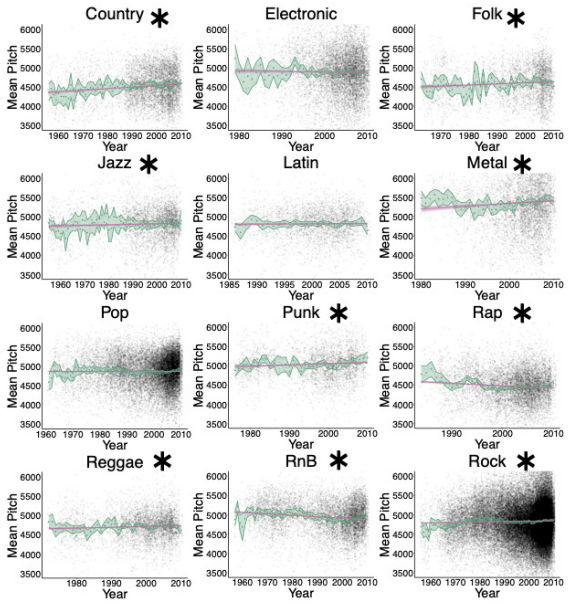

Figure 8: Relationship between mean pitch (in cents) and year for each genre. "*" denotes a significant effect of year. The red line represents the predicted slope with 95% confidence intervals. The green diamond and ribbon represent the mean per year and the standard error.

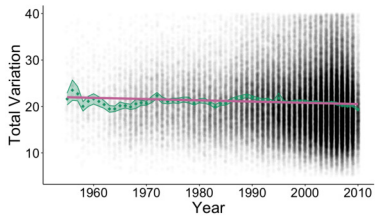

Figure 9: Total Variation as a function of year. Each dot represents a song. The red line represents the predicted slope with 95% confidence intervals. The green diamond and ribbon represent the mean TV per year and the standard error. There was a significant negative correlation between TV and year.

0.027, t = -10.47, p < .001; see Figure 9). When assessing the relationship between TV and musical genre, we found a significant main effect of genre (F(11, 140,980) = 1247.9, p < 0.001). Post-hoc tests showed that all genres were significantly different from one another (all p values <0.05) except for country and Latin (t=0.661, p = 1.000), country and punk (t = -3.214, p = 0.059), electronic and punk (t = 0.877, p = 0.999), electronic and reggae (t=0.269, p=1.000), folk and pop (t = -2.407, p = 0.4013), folk and rock (t = 0.185, p = 1.000), Latin and pop (t = 3.006, p = -.107), and punk and reggae (t = -0.569, p = 1.000; see Figure 10). Importantly, TV was significantly higher for rap music than for all other genres.

We found a significant main effect of year on TV for eleven of the twelve musical genres, though the direction of the trends varied (see Figure 11). Specifically, metal music (β = 0.066, t = 3.338, p <0.001), reggae music (β = 0.058, t = 6.512, p <0.001), and RnB (β = 0.013, t = 3.44, p <0.001) showed a significant positive relationship between year and TV. Conversely, electronic music (β=-0.126, t=-4.803, p<0.001), folk (β=-0.065, t=-7.852, p<0.001), jazz (β=-0.079, t=-6.418, p<0.001), Latin music (β=-0.041, t=-2.349, p=0.019), pop (β=-0.033, t=-8.698, p<0.001), punk (β=-0.045, t=-3.259, p=0.001), rap (β=-0.158, t=-11.06, p<0.001) and rock (β=-0.062, t=-13.79, p<0.001) showed a significant negative relationship between year and TV. No significant effect was found for country music.

5.3 Pitch Class Entropy

A linear regression showed a statistically significant negative relationship between PCE and year (β = -0.004, t = -50.02, p-value < 0.001; see Figure 12). There was a ceiling effect for PCE, with some of the tracks hitting close to the theoretical maximum of 3.59.

We ran a linear model with genre as the only main effect and found a significant main effect of genre on PCE (F(11, 140,982) = 759.64, p < 0.001). Post-hoc tests showed that all genres were significantly different than one another (all p-values <0.05) except for folk and Latin (t=-0.936, p=0.999), folk and pop (t=2.484, p=0.350), folk and reggae (t=1.255, p=0.984), jazz and RnB (t=3.123, p=0.077; approaching significance), Latin and rap (t=-2.952, p=0.123), Latin and reggae (t=2.218, p=0.536), and pop and reggae (t=-1.054, p=0.996; see Figure 13).

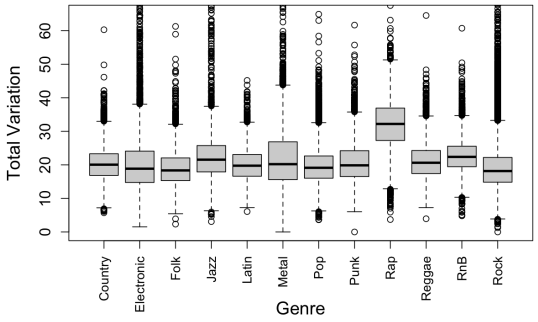

Figure 10: Total variation in each of the twelve genres across the whole dataset. Means are shown with interquartile range, 95% confidence interval error bars, and outliers. There were significant differences in TV between all genres except between country and Latin, country and punk, electronic and punk, electronic and reggae, folk and pop, folk and rock, Latin and pop, and punk and reggae.

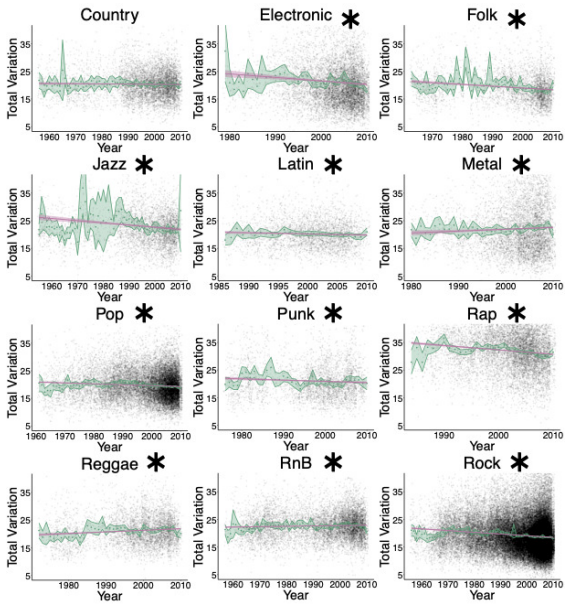

Figure 11: Relationship between TV and year for each genre. "*" denotes a significant effect of year. The red line represents the predicted slope with 95% confidence intervals. The green diamond and ribbon represent the mean per year and the standard error.

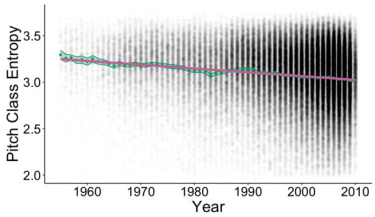

Figure 12: Pitch Class Entropy as a function of year. Each dot represents a song. The red line represents the predicted slope with 95% confidence intervals. The green diamond and ribbon represent the mean PCE per year and the standard error. There was a significant negative correlation between PCE and year

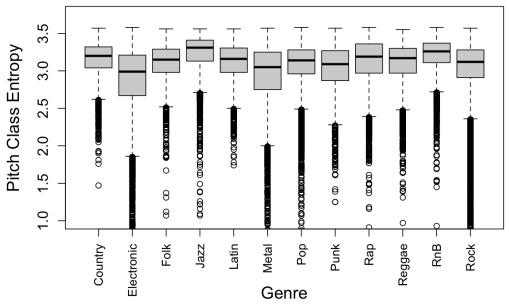

Figure 13: Pitch class entropy in each of the genres. Means are shown with interquartile ranges, 95% confidence interval error bars, and outliers. There were significant differences in PCE between all genres except between folk and Latin, folk and pop, folk and reggae, jazz and RnB, Latin and rap, Latin and reggae, and pop and reggae.

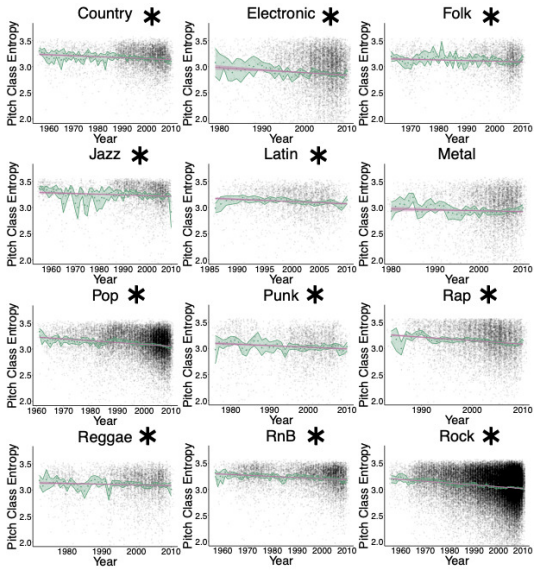

Figure 14: Pitch class entropy in each of the genres. "*" denotes a significant main effect of year. The red line represents the predicted slope with 95% confidence intervals. The green diamond and ribbon represent the mean per year and the standard error.

Finally, we found a significant negative relationship between PCE and year for eleven musical genres (country: β=-0.002, t=-7.847, p<0.001; electronic: β=-0.005, t=-5.126, p<0.001; folk: β=-0.001, t=-3.650, p<0.001; jazz: β=-0.001, t=-4.382, p<0.001; Latin: β=-0.004, t=-5.047, p<0.001; pop: β=-0.003, t=-18.61, p<0.001; punk: β=-0.003, t=-4.938, p<0.001; rap: β=-0.007, t=-11.61, p<0.001; reggae β=-0.001, t=-2.849, p=0.004; RnB: β=-0.002, t=-12.64, p<0.001; rock: β=-0.003, t=-28.75, p<0.001; see Figure 14). The effect of year for metal music (β=-0.002, t=-1.717, p=0.086) approached significance.

6. DISCUSSION

In this study, we analyzed vocal pitch characteristics across years and genres. We found musical genres are often significantly different from one another in mean pitch, total variation, and pitch class entropy. The data generally exhibited a significant negative relationship between year and total variation and year and pitch class entropy, respectively. This was the case both overall and for 8 and 11 musical genres, respectively (see Figure 10 and Figure 13).

If TV and PCE are taken to be measures of musical complexity, these findings could mean vocals, in this dataset at least, are getting less complex over time. This is somewhat in line with previous studies using the MSD. Serrà *et al.* found that newer songs have less variety in pitch transitions and more homogenized timbres, and Parmer *et al.* found that pitch complexity has been generally stable, but loudness and rhythm complexity have decreased [2] [8]. Our findings also parallel those of recent publications looking generally at Western popular music. In a recent study, authors found over five decades, lyrics have become simpler in their vocabulary richness, readability, complexity, and repetitiveness [25]. In another study analyzing popular melodies from 1950 to 2023, Hamilton and Pearce identified melodic revolutions that correspond to decreases in melodic complexity [26].

In our study, we observed that the rap genre had a higher TV than the other genres (see Figure 10), showing that rap songs feature more pitch variation than other musical genres, on average. This could be because rap vocals tend to have less sustained pitch than other genres. Previous work showed that pitch variance in rap music is a complex and significant feature of the genre [27, 28]. Rap music, only coming into prevalence in this dataset in 1984, may have influenced the genres that exhibit a significant positive relationship between year and total variation, counter to the all-genre-pooled negative trend: metal, reggae, and RnB.

We found mean pitch increased over time (see Figure 6). Gender and vocal range are key factors when considering pitch, and genre-specific gender prevalence may exist. However, we did not find a sufficiently reliable gender or vocal range classifier to support further analysis.

Interestingly, mean pitch was the highest for the metal genre, which has a low presence of female vocalists compared to other genres [29]. Therefore, the higher mean pitch of the metal genre cannot be fully explained by a higher prevalence of high-voiced singers. The average mean pitch of metal vocals sits quite high in a typical tenor range [21]. We hypothesize this is because screaming in metal music tends to have a higher f_0 than singing, but more investigation into metal vocals is needed [30].

7. CONCLUSION

In this exploratory research, we examined trends in the vocal lines of 143,152 songs spanning 55 years. Our work has identified relationships between vocal pitch and popular musical genres over time, providing valuable insights into the changing sound of music. We have demonstrated the utility of the methods presented here for studying vocals, and believe they have the potential to be applied to the study of other musical instruments as well as general musical phenomena including historical and cultural trends, changes in musical forms and structures, and stylistic differences across genres and periods.

8. REFERENCES

[1] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, Miami, Florida, USA*, A. Klapuri and C. Leider, Eds., 2011, pp. 591–596.

[2] J. Serrà, A. Corral, M. Boguñá, M. Haro, and J. L. Arcos, "Measuring the evolution of contemporary western popular music," *Scientific reports*, vol. 2, 05 2012.

[3] U. Shalit, D. Weinshall, and G. Chechik, "Modeling musical influence with topic models," in *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, ser. JMLR Workshop and Conference Proceedings, vol. 28. JMLR.org, 2013, pp. 244–252.

[4] D. Stoller, S. Ewert, and S. Dixon, "Wave-u-net: A multi-scale neural network for end-to-end audio source separation," in *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France*, E. Gómez, X. Hu, E. Humphrey, and E. Benetos, Eds., 2018, pp. 334–340.

[5] S. Rouard, F. Massa, and A. Défossez, "Hybrid transformers for music source separation," *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2023, Rhodes Island, Greece, June 4-10, 2023*, pp. 1–5, 2023. [Online]. Available: https://doi.org/10.1109/ICASSP49357.2023.10096956

[6] A. M. Demetriou, A. Jansson, A. Kumar, and R. M. Bittner, "Vocals in music matter: the relevance of vocals in the minds of listeners," in *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France, September 23-27, 2018*, E. Gómez, X. Hu, E. Humphrey, and E. Benetos, Eds., 2018, pp. 514–520.

[7] M. Bürgel, L. Picinali, and K. Siedenburg, "Listening in the mix: Lead vocals robustly attract auditory attention in popular music," *Frontiers in Psychology*, vol. 12, 12 2021.

[8] T. Parmer and Y. Ahn, "Evolution of the informational complexity of contemporary western music," in *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019, Delft, The Netherlands, November 4-8, 2019*, A. Flexer, G. Peeters, J. Urbano, and A. Volk, Eds., 2019, pp. 175–182.

[9] B. P. Gold, M. T. Pearce, E. Mas-Herrero, A. Dagher, and R. J. Zatorre, "Predictability and uncertainty in the pleasure of music: A reward for learning?" *Journal of Neuroscience*, vol. 39, no. 47, pp. 9397–9409, 2019.

[10] V. K. Cheung, P. M. Harrison, L. Meyer, M. T. Pearce, J.-D. Haynes, and S. Koelsch, "Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity," *Current Biology*, vol. 29, no. 23, pp. 4084–4092.e4, 2019.

[11] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi, "The evolution of popular music: Usa 1960–2010," *Royal Society Open Science*, vol. 2, no. 5, p. 150081, 2015. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsos.150081

[12] F. Thalmann, E. Nakamura, and K. Yoshii, "Tracking the Evolution of a Band's Live Performances over Decades," in *Proceedings of the 23rd International Society for Music Information Retrieval Conference*. Bengaluru, India: ISMIR, Dec. 2022, pp. 850–857. [Online]. Available: https://doi.org/10.5281/zenodo.7342596

[13] E. Deruty and F. Pachet, "The MIR perspective on the evolution of dynamics in mainstream music," in *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015, Málaga, Spain, October 26-30, 2015*, M. Müller and F. Wiering, Eds., 2015, pp. 722–727.

[14] P. D. Pestana, Z. Ma, J. D. Reiss, A. Barbosa, and D. A. A. Black, "Spectral characteristics of popular commercial recordings 1950-2010," *AES NY*, 2013.

[15] E. Oehrle, "Reviews - cantometrics: an approach to the anthropology of music by alan lomax. berkeley, ca: University extension media center, 1976. handbook and cassette available." *British Journal of Music Education*, vol. 9, no. 1, p. 83–86, 1992.

[16] M. Panteli, R. Bittner, J. P. Bello, and S. Dixon, "Towards the characterization of singing styles in world music," *EEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[17] A. Volk and P. van Kranenburg, "Melodic similarity among folk songs: An annotation study on similarity-based categorization in music," *Musicae Scientiae*, vol. 16, no. 3, pp. 317–339, 2012.

[18] H. Schreiber, "Improving genre annotations for the million song dataset," in *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015, Málaga, Spain, October 26-30, 2015*, M. Müller and F. Wiering, Eds., 2015, pp. 241–247.

[19] M. Mauch and S. Dixon, "Pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 659–663.

[20] B. McFee, A. Metsai, M. McVicar, S. Balke, C. Thomé, C. Raffel, F. Zalkow, A. Malek, Dana, K. Lee, O. Nieto, D. Ellis, J. Mason, E. Battenberg, S. Seyfarth, R. Yamamoto, viktorandreevichmorozov, K. Choi, J. Moore, R. Bittner, S. Hidaka, Z. Wei, nullmightybofo, D. Herénú, F.-R. Stöter, P. Friesch, A. Weiss, M. Vollrath, T. Kim, and Thassilo, "librosa/librosa: 0.8.1rc2," https://doi.org/10.5281/zenodo.4792298, May 2021.

[21] T. Stefan Kostka, T. Dorothy Payne, and B. Almén, *Tonal Harmony*. McGraw-Hill Education, 2017. [Online]. Available: https://books.google.com/books?id=Cs2UAQAACAAJ

[22] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, "Crepe: A convolutional representation for pitch estimation," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*, pp. 161–165, 2018. [Online]. Available: https://doi.org/10.1109/ICASSP.2018.8461329

[23] C. Raffel, B. McFee, E. J. Humphrey, J. Salamon, O. Nieto, D. Liang, and D. P. W. Ellis, "Mir_eval: A transparent implementation of common mir metrics," in *International Society for Music Information Retrieval Conference*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:17163281

[24] J. L. Snyder, "Entropy as a Measure of Musical Style: The Influence of A Priori Assumptions," *Music Theory Spectrum*, vol. 12, no. 1, pp. 121–160, 03 1990. [Online]. Available: https://doi.org/10.2307/746148

[25] E. Parada-Cabaleiro, M. Mayerl, S. Brandl, M. Skowron, M. Schedl, E. Lex, and E. Zangerle, "Song lyrics have become simpler and more repetitive over the last five decades," *Scientific Reports*, vol. 14, 03 2024.

[26] M. Hamilton and M. Pearce, "Trajectories and revolutions in popular melody based on u.s. charts from 1950 to 2023," *Scientific Reports*, vol. 14, 07 2024.

[27] M. Ohriner, "Analysing the pitch content of the rapping voice," *Journal of New Music Research*, vol. 48, pp. 413 – 433, 2019.

[28] R. Komaniecki, "Vocal pitch in rap flow," *Intégral*, vol. 34, pp. 25–46, 2020.

[29] A. Epps-Darling, H. Cramer, and R. T. Bouyer, "Artist gender representation in music streaming," in *International Society for Music Information Retrieval Conference*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:232317721

[30] S. Mesiä and P. Ribaldini, "Heavy metal vocals : A terminology compendium," in *Modern Heavy Metal: Markets, Practices and Culture, Helsinki: Aalto University*, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:204842147