

# EVOLUTION OF THE INFORMATIONAL COMPLEXITY OF CONTEMPORARY WESTERN MUSIC

Thomas Parmer, Yong-Yeol Ahn

School of Informatics, Computing, and Engineering  
Indiana University, Bloomington, IN, USA

tjparmer@indiana.edu, yyahn@iu.edu

## ABSTRACT

We measure the complexity of songs in the Million Song Dataset (MSD) in terms of pitch, timbre, loudness, and rhythm to investigate their evolution from 1960 to 2010. By comparing the Billboard Hot 100 with random samples, we find that the complexity of popular songs tends to be more narrowly distributed around the mean, supporting the idea of an inverted U-shaped relationship between complexity and hedonistic value. We then examine the temporal evolution of complexity, reporting consistent changes across decades, such as a decrease in average loudness complexity since the 1960s, and an increase in timbre complexity overall but not for popular songs. We also show, in contrast to claims that popular songs sound more alike over time, that they are not more similar than they were 50 years ago in terms of pitch or rhythm, although similarity in timbre shows distinctive patterns across eras and similarity in loudness has been increasing. Finally, we show that musical genres can be differentiated by their distinctive complexity profiles.

## 1. INTRODUCTION

Our everyday life is surrounded by cultural products; we wake up to a song, read a book on the subway, watch a movie with friends, or even travel far to admire a piece of art. Despite such pervasiveness, we cannot fully explain why we like a particular song over others or what makes something a great piece of art. Although the perceived quality of a piece is affected by numerous contextual factors, including one's cultural, social, and emotional background, theories suggest that preference, or 'hedonistic value', may also be affected by innate properties of the products, such as *novelty* and *complexity* [1, 18, 28]. In particular, a popular theory suggests there is a *Goldilocks principle* — that just the right amount of novelty or complexity elicits the largest amount of pleasure, whereas pieces with too little or too much complexity are less interesting and enjoyable [3]. On the other hand, cultural

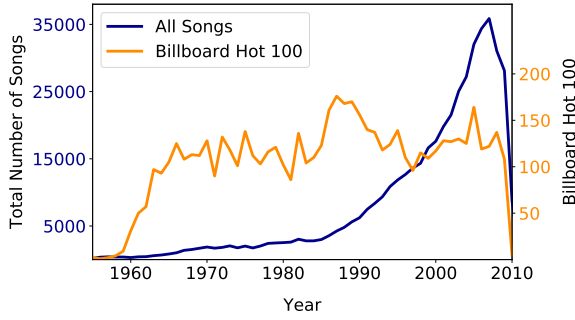
products are also *fashionable* — what is popular now may be completely out of fashion next month. Such seemingly contrasting observations prompt us to ask the following questions: *as fads come and go, is there still a consistent preference towards the optimal amount of complexity in cultural products? How has the complexity of contemporary cultural products changed over time?*

This question may apply to any type of cultural product, but we focus here on the complexity of contemporary Western songs. Although various studies have already reported evidence of the 'inverted U-shaped' relationship between perceived complexity and the pleasantness of music in terms of individual-level preference [1, 11, 32], evidence of this preference at the population-level is unclear [7, 22, 30], and many past studies have been limited by the size or extent of the data, in terms of genres or temporal range.

Recently, datasets such as the Million Song Dataset (MSD) began to allow researchers to systematically analyze patterns in music at a massive scale [6, 8, 26]. For example, Serra *et al.* used musical 'codewords' based on song segments in the MSD to identify changes in pitch, timbre, and loudness over time, finding that newer songs restrict pitch transitions, homogenize timbre, and increase loudness (without increasing the variability in loudness) [26]. Mauch *et al.* used a corpus of 17,000 songs from the Billboard Hot 100 to analyze how popular music has evolved between 1960 and 2010 in the United States; using timbral and harmonic features derived from songs on the Hot 100, they identified three stylistic revolutions that occurred in 1964, 1983, and 1991 [15].

In this paper, we analyze the large-scale evolution of complexity in contemporary music in terms of pitch, loudness, timbre, and rhythm, during the period from 1960 to 2010 using the Million Song Dataset (MSD) [4]. We find that complexity does seem to constrain popularity, as evidenced by the most popular songs (those on the Billboard Hot 100) clustering around average values and exhibiting smaller variance compared to a random sample. However, complexity values do fluctuate over time, as long-term trends are seen in loudness, timbre, and rhythm complexity and in the similarity between songs on the Billboard Hot 100. Finally, we compare the complexity of different genres and find that genres have characteristic complexity profiles, leading us to hypothesize that complexity may be a factor in an individual's musical selection.





**Figure 1.** The number of songs per year. ‘All Songs’ refers to the filtered MSD dataset, while ‘Billboard Hot 100’ refers to those songs whose title and artist we matched with songs on the Hot 100 as identified in [15].

## 2. METHODS

### 2.1 Data

The MSD is a dataset of one million songs created by Columbia University’s LabROSA in collaboration with The Echo Nest [4]. Each song in the dataset is divided into small temporal segments (based on note onsets) with detailed data derived from the song’s audio signal, and includes metadata such as title, artist, year, duration, and genre terms.

Prior to analysis, we filtered the MSD to remove duplicates, songs with missing genre or duration metadata, and songs likely to be commentary pieces (whose title included the tokens ‘interview’, ‘commentary’, ‘introduction’, ‘discuss’, ‘conference’, or ‘intro’), resulting in a dataset of 905,896 songs. Some songs also did not have all data types — pitch, loudness, timbre, rhythm, or year — that we examine here and were thus left out of the corresponding calculations. Due to a limited amount of data from the early years, we restricted our analysis to the period from 1960 to 2010. The genre of each song was determined by the term (Echo Nest tag) with the strongest weight, although we note that terms are assigned at the artist level so all songs by the same artist are grouped into the same genre.

To discover the most popular musical pieces in our dataset, we found 6,661 songs which charted on the Billboard Hot 100 as identified in a previous study [15]. The number of songs per year in our final dataset is shown in Figure 1.

### 2.2 Codewords

To estimate complexity, we defined “codewords” for each song across four dimensions (pitch, loudness, timbre, and rhythm), similar to a previous study [26]. Each codeword is based on a segment of the song. Pitch and timbre codewords are vectors containing the pitches (based on the binary presence of each of 12 pitches in the chromatic scale) and timbres (based on analysis of the audio signal, with 11 components thresholded into three bins) present in the segment. Loudness codewords are equal to the binned maximum decibel value of the segment. Similarly, rhythm

codewords are defined as the number of average sixteenth notes between segments, where the average sixteenth note is based on the time signature. We then defined a measure of complexity for each feature per song based on the conditional entropy of each type of codeword.<sup>1</sup>

### 2.3 Measuring Complexity

Although many studies have examined the relationship between the complexity of a piece and the derived pleasure from it [1, 19–21, 31, 32], there is no universally adopted way to measure the complexity of a song. Existing definitions of complexity include hierarchical complexity, dynamic complexity, information-theoretic complexity, and cognitive complexity [10, 23, 27, 34]. Information-theoretic measures are attractive because they capture the surprise inherent in a pattern, such as the notes played in a musical piece. Theories propose that music can be understood as the kinetics of expectation and surprise, and that composers seek to elicit emotions by fulfilling or denying these expectations [1, 17, 35]. In particular, Implication-realization (IR) theory posits that open intervals evoke expectations in a listener and the surprises of these expectations may be related to complexity [18, 32, 35].

Information-theoretic measures include Shannon entropy, joint entropy, conditional entropy, compression or algorithmic complexity [10, 16, 27, 29], and more complicated techniques such as pairwise predictability between time series, Hidden Markov Models, Normalized Compression Distance, and predictive information rate [1, 8, 9]. Previous studies have used information-theoretic quantities to estimate perceived complexity, identify piece similarity, derive psycho-acoustic features, and classify genres [5, 10, 13, 25, 27, 30].

We use conditional entropy as our measure of complexity, dependent on the immediately preceding symbol, as it is known that events during even short preceding intervals are enough to evoke strong expectations in the listener [1, 12, 35]. Other information-theoretic measures are either more complicated (e.g. predictive information measures), can only be approximated in practice (e.g. Kolmogorov or algorithmic complexity), or do not take past information into account (e.g. Shannon entropy).

Each song was assigned a complexity value for pitch, loudness, timbre, and rhythm, which is equal to the conditional entropy of the feature codewords:

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X=x) = - \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x) \quad (1)$$

where  $X$  and  $Y$  are possible codewords,  $p(x)$  is the probability of observing codeword  $x$  and  $p(y|x)$  is the probability of observing codeword  $y$  given the previous codeword  $x$ .

<sup>1</sup> Code is available at <https://github.com/tjparmer/music-complexity>.

### 3. RESULTS

#### 3.1 Complexity and Popularity

The complexity distribution of songs is approximately bell-shaped, although timbre is skewed towards zero complexity — unlike the other features, timbre becomes easily predictable after only one previous codeword. The distributions of the Hot 100 songs are similar, although the Hot 100 tends to exhibit statistically lower complexity in pitch and timbre and higher loudness complexity, compared to 95% confidence intervals of 1,000 bootstrap random samples of the same size (see Fig. 2). Furthermore, we found that the variances of the Hot 100 complexity values are smaller than for other songs (based on 95% confidence intervals of 1,000 bootstrap samples from the Hot 100 compared to 1,000 bootstrap samples from the overall distribution); thus, the popular songs tend to be located in a narrower range near the mean across pitch, loudness, and rhythm complexity. This result supports the theory for an inverted U-shaped curve where global popularity is maximized by medium complexity.

#### 3.2 Complexity Across Time

To examine the evolution of song complexity, we calculate the mean complexity values for each year (for all songs and the Hot 100 songs separately) in Fig. 3, which shows several long-term trends. Later years mark the appearance of songs with low loudness and rhythm complexity and songs with high timbre complexity, but they were not reflected strongly in the Hot 100 songs. The low loudness complexity may be due to the trend often called the “loudness war” [26], which describes the tendency to produce the entire song to be as loud as possible. Another possible reason may be the emergence of low complexity genres in recent years. For instance, terms associated with low loudness complexity outliers include ‘grindcore’, ‘hip hop’, and ‘black metal’, all of which are relatively newer genres in the dataset. Low rhythm and high timbre complexity may be due to pop or electronic music that contain modern production techniques with many different synthesized textures and strong dance beats. Terms associated with low rhythm outliers include ‘tech house’, ‘techno’, and ‘hard trance’, while terms associated with high timbre complexity outliers include ‘tech house’, ‘techno’, and ‘deep house’.

Previous research has indicated that the evolution of Western popular music experienced significant changes during three musical ‘revolutions’ in 1964, 1983, and 1991 [15]. The first was associated with rock and soul music, the second with disco, new wave, and hard rock, and the third with the emerging popularity of rap music over rock music. These three revolutions split our period of analysis into three ‘epochs’: 1964-1983, 1983-1991, and 1991-2010. With this reference frame, we examine our measures of complexity.

If we consider the entire dataset, each aspect of complexity shows a different pattern. The pitch complexity has been more or less stable across the whole period; the loud-

ness complexity has been decreasing overall, although the period from 1983 to 1991 shows a slight increase; the timbre complexity has been steadily increasing and reached a plateau after the 1990s; finally, the rhythm complexity was decreasing through the period from 1964 to 1983, and then stabilized.

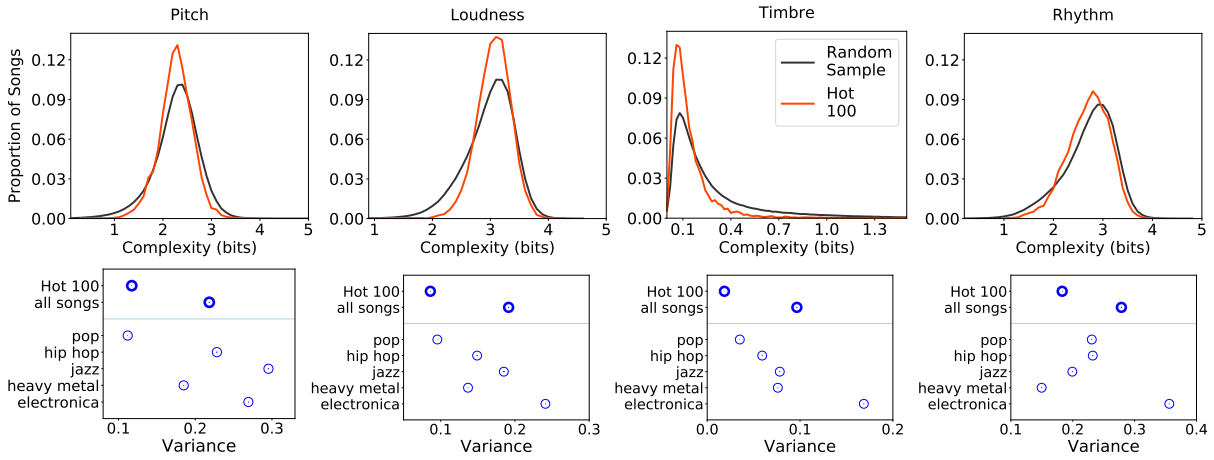
Meanwhile, we find that the temporal evolution of the Hot 100 songs does not follow the overall pattern. The largest difference can be observed in the timbre complexity. While the timbre complexity of the entire dataset has been steadily increasing, it has been almost completely flat for the most popular songs, diverging from the overall trend. This may indicate that the emergence of new genres with high timbre complexity primarily happened for more niche musical tastes. Pitch and loudness complexity, by contrast, have been higher for popular songs in recent years, while rhythm complexity was lower until the 2000s.

#### 3.3 Popular Song Similarity

The analysis of complexity over time suggests that modern day popular songs (at least from the 2000s) are more likely to have higher pitch, loudness, and rhythm complexity (and lower timbre complexity) than their less popular contemporaries. However, while this suggests that popular songs are not simpler than the average song, it does not necessarily indicate whether they sound more or less similar to their popular contemporaries (that is, other Hot 100 songs that are released in the same year). A recent report suggests that popular songs are sounding more and more similar to other songs on the charts [33]. In contrast, other research finds that songs that perform well on the charts do not sound too similar to their contemporaries but often have an optimal level of differentiation [2]. To analyze whether popular songs become more similar to their contemporaries over time, we measure the Kullback-Leibler (KL) divergence [14] from each song in the Hot 100 to other popular songs that were released in the same year. KL divergence captures the unexpectedness of a song’s codewords given the codewords present in other popular songs and thus indicates the spread of codeword usage per year.

In Fig. 4, we show the KL divergence per year for each feature, with each epoch marked. Larger KL divergence in the figure suggests that the songs in that year are more different from their contemporaries as compared to other years. Our measurement shows that the 1964-1983 and 1991-2010 period are similar to each other while 1983-1991 shows a reversing trend. Across features, the KL divergence was either decreasing (songs are more similar to each other) or stable during 1964-1983 and 1991-2010, while 1983-1991 marked either a positive trend reversal or a slow-down of the decreasing trend. These changing trends suggest that musical revolutions are not born equal; some may have spurred diversity among popular songs while some may have homogenized the field.

Despite these fluctuations, the divergence trend is roughly stable over time for pitch and rhythm, while timbre rebounds after similarity decreases; thus, our findings are consistent with research that songs that perform well



**Figure 2.** Feature complexities and variances. Complexity distributions are shown (in bins of 0.1 bits, except for timbre which is in bins of 0.02 bits). The variance plots include 95% confidence intervals in black (although confidence intervals are smaller than the symbol and not visible), based on 1000 bootstrap samples of 1000 songs from the respective genre.

on the charts do not sound too similar to their contemporaries but rather maintain a degree of uniqueness which is statistically consistent over time.

### 3.4 Complexity Across Genres

Let us turn our attention to musical genres and their complexity. As some genres may be characterized by complex harmonic structures or simple, repeated patterns, we expect to see differences across different genres in terms of complexity. For example, jazz is often considered to have complex patterns whereas dance music may be assumed to use simpler rhythmic patterns. Our measurement concurs with such speculation, but finds that different subsets of genres may be relatively complex across one or two features but not others. For instance, electronic and dance styles tend to have high pitch complexity values, whereas jazz and blues have high loudness complexity values. The highest timbre complexity values belong to electronic genres, although metal also scores highly, but electronic genres have reduced rhythmic complexity which is instead maximal in jazz, progressive and vocal genres.

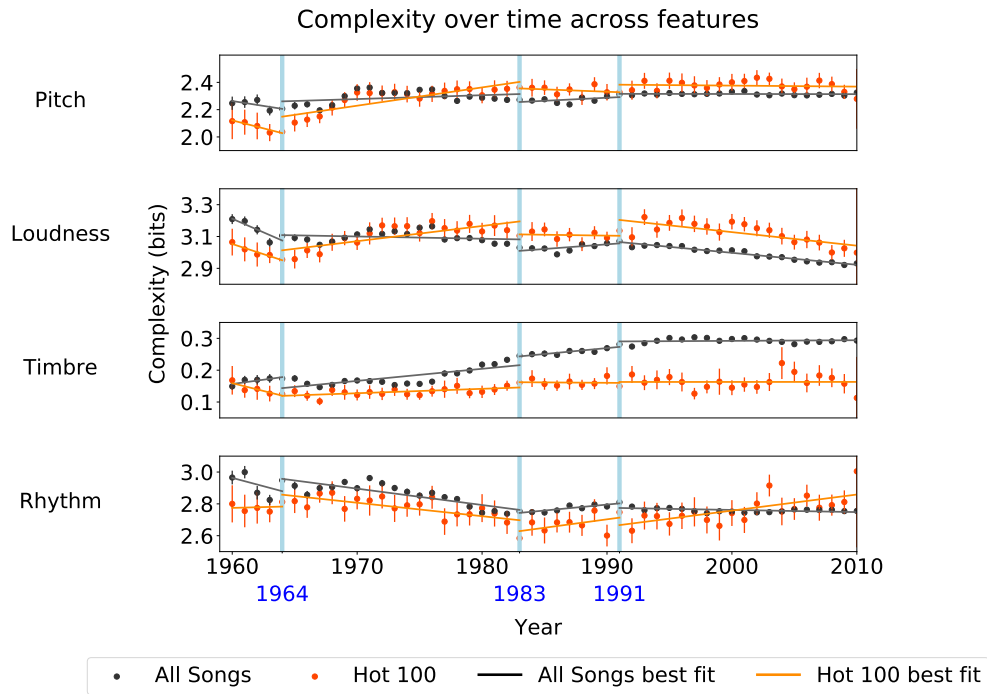
We found that a variety of common genres were significantly different from a random sample drawn from the overall distribution in terms of each feature complexity (based on a two-sample Kolmogorov-Smirnoff nonparametric test as well as 95% confidence intervals of the means of each feature), with the exception that pop was not rhythmically distinct. Thus each genre seems to have distinctive complexity features that describe its songs: jazz is relatively complex (except in terms of timbre), hip hop has higher than average pitch and loudness complexity, heavy metal has high rhythm complexity but low pitch and loudness complexity, and electronica has high timbre complexity but low rhythm complexity (Fig. 6).

This pattern may be indicative of some trade-offs that listeners make. If they prefer timbre at the expense of rhythmic complexity, they may prefer electronic genres. If

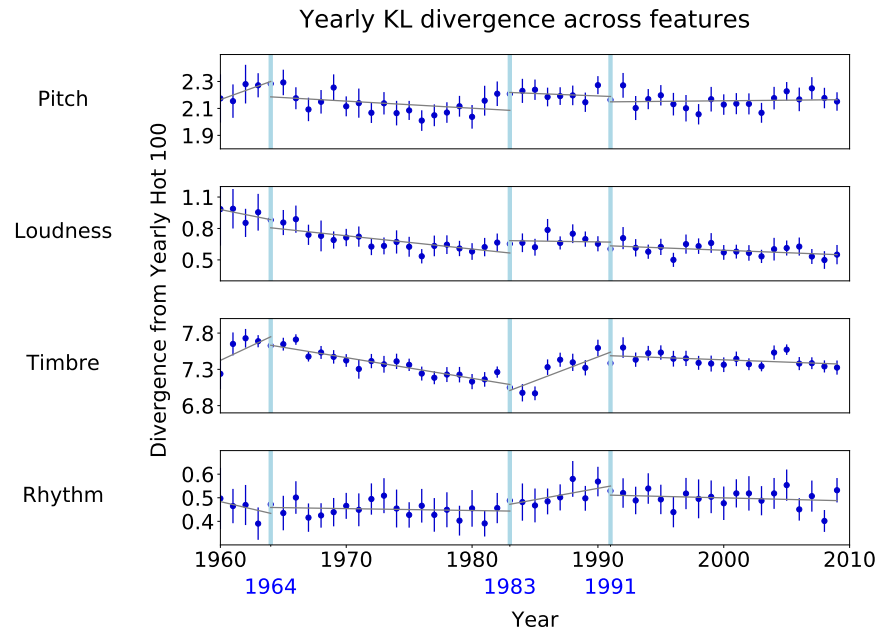
they prefer pitch and loudness complexity, they may prefer hip hop or jazz. If they care about timbre and rhythm over pitch and loudness, they may prefer heavy metal. There is a positive correlation between pitch and loudness complexity (Pearson’s  $r=0.77$ ) across all songs, suggesting that genres tend to have high pitch and loudness complexity (e.g. hip hop, jazz) or low pitch and loudness complexity (e.g. heavy metal). There is also a negative correlation seen between timbre and rhythm complexity (Pearson’s  $r=-0.55$ ), suggesting that rhythmic complexity decreases with higher timbre complexity (although this is not true for metal genres).

Interestingly, the Hot 100 is similar to the pop genre in feature means and variances (although statistically different). Both pop music (whose songs are given no genre-specific term with higher weight than ‘pop’) and the Hot 100 (whose songs are primarily classified as genres other than ‘pop’) have near average values of pitch, loudness, and rhythm complexity, and lower than average values of timbre complexity, while also having smaller variance than the other selected genres (refer to Figures 2 and 6). This may suggest that listeners expect the same from listening to the Hot 100 as they do when listening to music labeled as ‘pop’: mildly surprising songs that do not vary too much in complexity and which are sonically predictable.

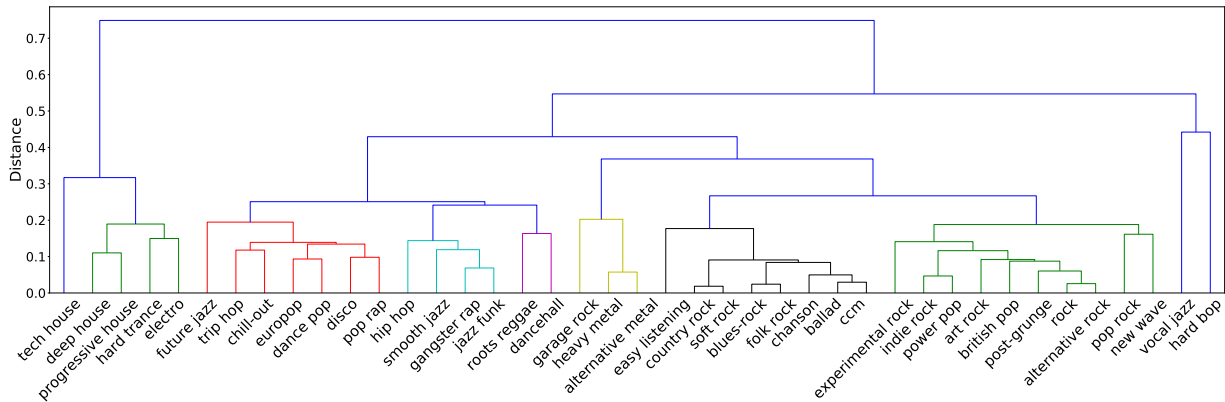
One may also expect similar genres to share similar complexity scores. We used agglomerative clustering on genres represented with over 5000 songs in the dataset (a total of 41 genres), using Euclidean distance between the genre mean complexity scores of each feature, and then used the silhouette coefficient [24, 28] to find nine optimal communities. The result matches intuitive expectations, such that rock genres are grouped together as are electronic genres; interestingly jazz is grouped with hip hop and rap, due to these genres having similar complexity scores. These results suggest that a genre may be defined, to some degree, by its pitch, loudness, timbre, and rhythmic complexity.



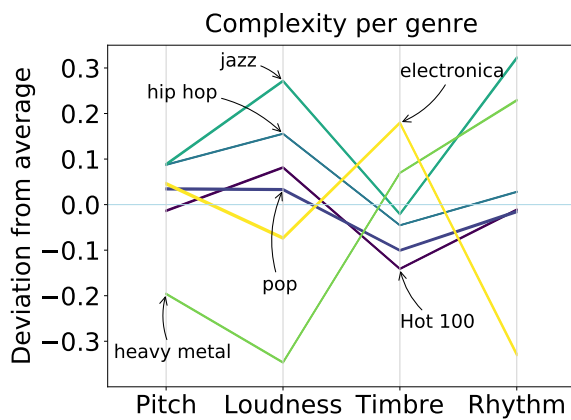
**Figure 3.** Average complexity over time across features. Linear trend lines (obtained using OLS linear regression) are shown for each epoch, along with 95% confidence intervals of the mean. Light blue lines indicate the musical revolutions found in [15].



**Figure 4.** Yearly KL divergence of Hot 100 songs across features. Complexity variance, as measured by KL divergence between a song's codewords and codewords from other songs released in that year, shows changing trends at the time of certain musical 'revolutions' in 1964, 1983, and 1991 (indicated by light blue lines). Linear trend lines (obtained using OLS linear regression) are shown for each epoch, along with 95% confidence intervals of the mean.



**Figure 5.** Dendrogram of top genres. Clustering is based on the Euclidean distance between complexity features. Colors indicate different communities.



**Figure 6.** Complexity of select genres across features. For each feature, the deviation of the genre complexity mean from the mean of 100 bootstrapped samples of the same size as the genre is shown. Note the similarity between the Hot 100 and the pop genre.

#### 4. DISCUSSION

Understanding and characterizing the complexity of music is an important area of study with both cultural and economic significance. Although music may seem complicated, songs quickly become predictable as you take previous knowledge into account. This suggests that conditional entropy may be a useful way to characterize musical complexity, although our approach here assumes that the uniform distribution of codewords is the state of maximum uncertainty and expectations are made based on only one previous symbol, which cannot distinguish counts of repeating patterns or phrasing [1]. Our approach is thus intrinsic to the song itself and ignores any *a priori* contextual information.

Using this measure, we find that pitch complexity has been generally stable over the period from 1960 to 2010, while loudness and rhythm complexity have decreased and timbre complexity has increased. Complexity norms seem to constrain the most popular songs, as those on the Hot

100 are distributed around the overall feature means with small variance in complexity, the exception being timbre which is lower than average. Indeed, the Hot 100 is similar to songs labeled as ‘pop’, in that pop also has average pitch, loudness, and rhythm complexity and low variance.

This result provides evidence of a global, inverted U-shaped relationship between popularity and complexity, where popular songs are, on average, the most pleasant to the population. Listeners may expect popular songs to be mildly complex, but not to deviate far from expected timbre or complexity norms. Complexity of the Hot 100 has in fact been consistent over fifty years in pitch and timbre, while increasing recently in rhythm and decreasing in loudness. Similarly, popular songs continue to maintain a consistent level of differentiation from their contemporaries in terms of pitch, timbre, and rhythm.

Certain genres do differ significantly across complexity features, suggesting that they have specific complexity profiles that help define them. We hypothesize that certain genres may ‘make up’ for lack of complexity in one area by increased complexity in another. Perhaps fans of electronic genres prioritize complexity in timbre but rhythmically simple dance beats, or metal fans prioritize rhythmic complexity and high volume at the expense of loudness complexity.

More research needs to be done to fully elucidate the relationship between complexity and musical appreciation. For example, future research can relate musical complexity to the listening habits of people on a large scale to determine a more fine-grained measure of song popularity. The consistency of popular songs over time suggests that, collectively, people tend towards songs that are a certain optimal level of complexity rather than being too simple or complicated. However, it remains an open question to what degree complexity plays a role in people’s cognitive appreciation of music.

#### 5. REFERENCES

- [1] Samer Abdallah and Mark Plumbley. Information dynamics: patterns of expectation and surprise in the per-



- ception of music. *Connection Science*, 21(2-3):89–117, 2009.
- [2] Noah Askin and Michael Mauskopf. What makes popular culture popular? Product features and optimal differentiation in music. *American Sociological Review*, 82(5):910–944, 2017.
- [3] Daniel E Berlyne. Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8(5):279–286, 1970.
- [4] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pages 591–596, 2011.
- [5] Rudi Cilibrasi, Paul Vitányi, and Ronald de Wolf. Algorithmic clustering of music based on string compression. *Computer Music Journal*, 28(4):49–67, 2004.
- [6] Sander Dieleman, Philémon Brakel, and Benjamin Schrauwen. Audio-based music classification with a pretrained convolutional network. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pages 669–674, 2011.
- [7] Tuomas Eerola and Adrian C North. Expectancy-based model of melodic complexity. In *Proceedings of the Sixth International Conference on Music Perception and Cognition*, 2000.
- [8] Peter Foster, Simon Dixon, and Anssi Klapuri. Identifying cover songs using information-theoretic measures of similarity. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(6):993–1005, 2015.
- [9] Peter Foster, Anssi Klapuri, and Simon Dixon. A method for identifying repetition structure in musical audio based on time series prediction. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 1299–1303. IEEE, 2012.
- [10] Barbra Gregory. *Entropy and complexity in music: some examples*. PhD thesis, University of North Carolina at Chapel Hill, 2005.
- [11] Ronald G Heyduk. Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception & Psychophysics*, 17(1):84–90, 1975.
- [12] David Brian Huron. *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2006.
- [13] Heather D Jennings, Plamen Ch Ivanov, Allan de M Martins, PC da Silva, and GM Viswanathan. Variance fluctuations in nonstationary time series: a comparative study of music genres. *Physica A: Statistical Mechanics and its Applications*, 336(3-4):585–594, 2004.
- [14] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.
- [15] Matthias Mauch, Robert M MacCallum, Mark Levy, and Armand M Leroi. The evolution of popular music: USA 1960–2010. *Royal Society open science*, 2(5):150081, 2015.
- [16] David Meredith. Music analysis and kolmogorov complexity. In *Proceedings of the 19th Colloquio d’Informatica Musicale (XIX CIM)*, 2012.
- [17] Leonard B Meyer. *Music, the Arts and Ideas: Patterns and Predictions in Twentieth-century Culture*. University of Chicago Press, 1967.
- [18] Eugene Narmour. *The analysis and cognition of basic melodic structures: The implication-realization model*. University of Chicago Press, 1990.
- [19] Adrian C North and David J Hargreaves. Experimental aesthetics and everyday music listening. In *The social psychology of music*, pages 84–103. Oxford University Press, Oxford, 1997.
- [20] Adrian C North and David J Hargreaves. Liking, arousal potential, and the emotions expressed by music. *Scandinavian journal of psychology*, 38(1):45–53, 1997.
- [21] Mark G Orr and Stellan Ohlsson. Relationship between complexity and liking as a function of expertise. *Music Perception: An Interdisciplinary Journal*, 22(4):583–611, 2005.
- [22] Robert Mitchell Parry. Musical complexity and top 40 chart performance. Technical report, Georgia Institute of Technology, 2004.
- [23] Jeffrey Pressing. Cognitive complexity and the structure of musical patterns. In *Proceedings of the 4th Conference of the Australasian Cognitive Science Society*, 1999.
- [24] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [25] Eric D Scheirer, Richard B Watson, and Barry L Vercoe. On the perceived complexity of short musical segments. In *Proceedings of the 2000 International Conference on Music Perception and Cognition*, 2000.
- [26] Joan Serrà, Álvaro Corral, Marián Boguñá, Martín Haro, and Josep Ll Arcos. Measuring the evolution of contemporary western popular music. *Scientific reports*, 2:521, 2012.
- [27] Ilya Shmulevich and D-J Povel. Measures of temporal pattern complexity. *Journal of New Music Research*, 29(1):61–69, 2000.

- [28] Higor YD Sigaki, Matjaž Perc, and Haroldo V Ribeiro. History of art paintings through the lens of entropy and complexity. *Proceedings of the National Academy of Sciences*, 115(37):E8585–E8594, 2018.
- [29] S.J. Simon. *A multi-dimensional entropy model of jazz improvisation for music information retrieval*. PhD thesis, University of North Texas, 2005.
- [30] Dean Keith Simonton. Computer content analysis of melodic structure: Classical composers and their compositions. *Psychology of Music*, 22(1):31–43, 1994.
- [31] Loren Steck and Pavel Machotka. Preference for musical complexity: Effects of context. *Journal of Experimental Psychology: Human Perception and Performance*, 1(2):170, 1975.
- [32] Sebastian Streich et al. *Music complexity: a multifaceted description of audio content*. Universitat Pompeu Fabra Barcelona, Spain, 2006.
- [33] Andrew Thompson and Matt Daniels. The musical diversity of pop songs, 2018. Available from: <https://pudding.cool/2018/05/similarity/>.
- [34] Eric Thul and Godfried T Toussaint. Rhythm complexity measures: A comparison of mathematical models of human perception and performance. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 663–668, 2008.
- [35] Pablo H Rodriguez Zivic, Favio Shifres, and Guillermo A Cecchi. Perceptual basis of evolving western musical styles. *Proceedings of the National Academy of Sciences*, 110(24):10034–10038, 2013.



# SUPPLEMENTARY MATERIALS

**Thomas Parmer, Yong-Yeol Ahn**

School of Informatics, Computing, and Engineering  
Indiana University, Bloomington, IN, USA  
tjparmer@indiana.edu, yyahn@iu.edu

## 1. FILTERING THE MSD

The MSD contained songs with bad metadata and segment data, duplicates, and tracks that were talking pieces rather than musical pieces (such as interviews with musicians). As such, the MSD was filtered to remove any song that met any of the following criteria:

- the song had no genre terms
- the song had no duration
- the song had duration less than three standard deviations below expectation for its genre (where standard deviations were calculated from a random 39100 song sample, directory ‘A’)
- the song title contained the words ‘interview’, ‘commentary’, ‘introduction’, ‘discussion’, ‘conference’, ‘intro’

Additionally, several tracks in the MSD were duplicates of the same song. Many of these are recognized in the official duplicate list of the MSD [1] while others were found by the authors by comparing song and artist names. Songs were considered duplicates if they shared the same song id or if their song name and artist name matched, subject to the following modifications:

- song and artist names were made lowercase
- whitespaces were removed
- the following characters were removed: " ' , . ? ! ( )
- any ‘&’ was replaced with ‘and’
- any instance of the word ‘featuring’ was replaced with ‘ft’

When considering duplicates, one track was arbitrarily chosen to keep in the filtered set and the rest were discarded. The resulting filtered dataset contained over 90% of the original songs. However not all songs had pitch, loudness, or rhythm data and these songs were kept out of

the respective analyses. Additionally, nearly half of the songs were missing year metadata and so were ignored during overtime analyses. A list of song ids comprising our dataset is available at <https://github.com/tjparmer/music-complexity>.

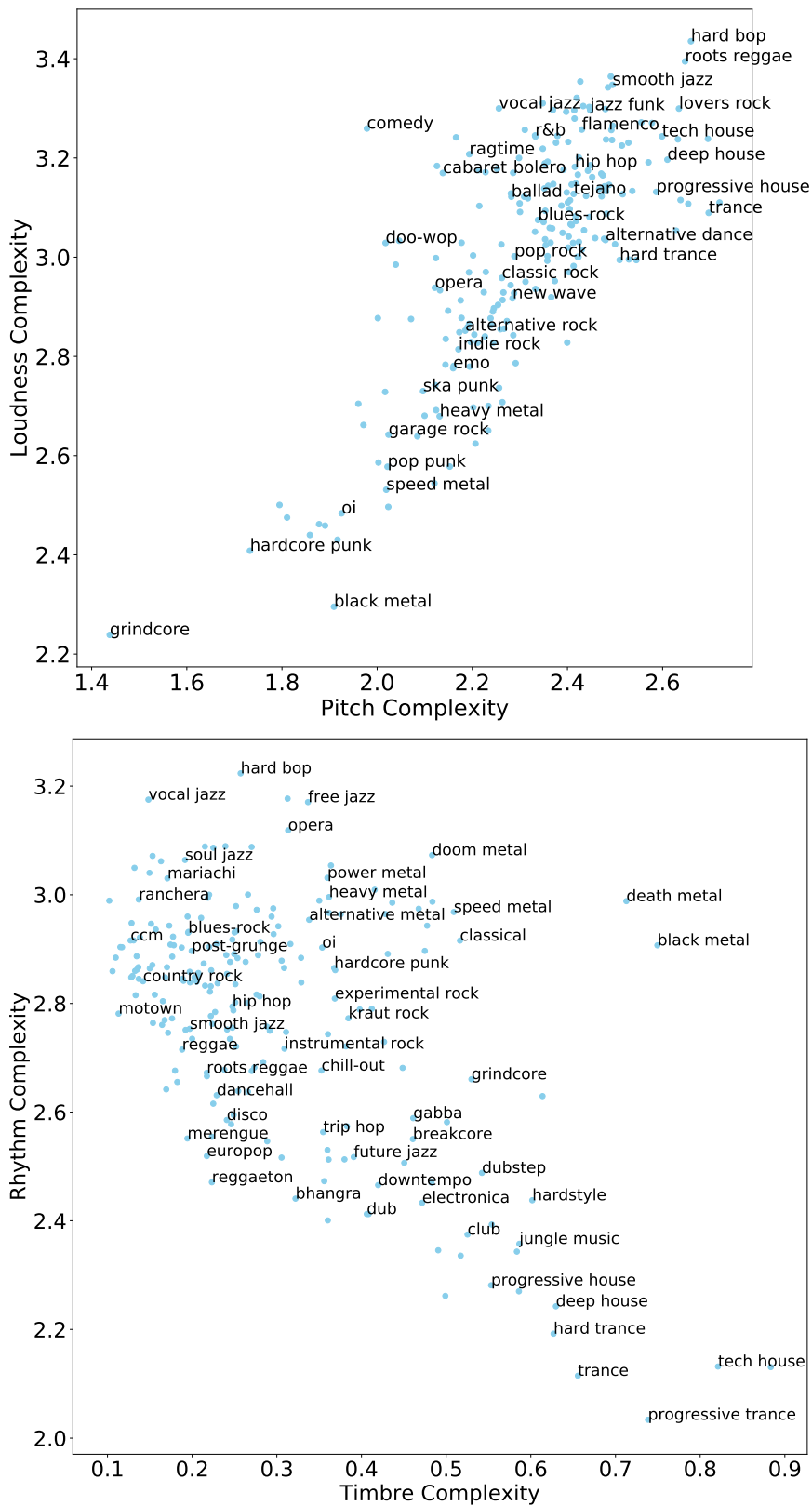
## 2. COMPLEXITY CORRELATIONS

In comparing between every pair of features, two interesting correlations were noted: the positive correlation between pitch and loudness and the negative correlation between timbre and rhythm. Correlation by genre was calculated using the complexity average of all songs in each genre. Graphs are shown with genres labeled (refer to Figure 1).

## 3. REFERENCES

- [1] Thierry Bertin-Mahieux. The 921,810 song dataset - duplicates, 2011.





**Figure 1.** Feature correlations across genres, where each data point represents one genre. A positive correlation is seen between pitch and loudness complexity (Pearson's  $r=0.77$ ) and a negative correlation is seen between timbre and rhythm complexity (Pearson's  $r=-0.55$ ). Only genres with more than 1000 songs are shown.