Telling Stories with Soundtracks: An Empirical Analysis of Music in Film

Jon Gillick
School of Information
University of California, Berkeley
jongillick@berkeley.edu

David Bamman
School of Information
University of California, Berkeley
dbamman@berkeley.edu

Abstract

Soundtracks play an important role in carrying the story of a film. In this work, we collect a corpus of movies and television shows matched with subtitles and soundtracks and analyze the relationship between story, song, and audience reception. We look at the content of a film through the lens of its latent topics and at the content of a song through descriptors of its musical attributes. In two experiments, we find first that individual topics are strongly associated with musical attributes, and second, that musical attributes of soundtracks are predictive of film ratings, even after controlling for topic and genre.

1 Introduction

The medium of film is often taken to be a canonical example of narrative multimodality: it combines the written narrative of dialogue with the visual narrative of its imagery. While words bear the burden alone for creating compelling characters, scenes, and plot in textual narratives like literary novels, in film, this responsibility is shared by each of the contributors, including the screenwriter, director, music supervisor, special effects engineers, and many others. Working together to support the overall story tends to make for more successful component parts; the Academy Award winner for Best Picture, for example, often also collects many other awards and nominations—including acting, cinemotography, and sound design.

While film has recently been studied as a target in natural language processing and computer vision for such tasks as characterizing gender representation in dialogue (Ramakrishna et al., 2015; Agarwal et al., 2015), inferring character types from plot summaries (Bamman et al., 2013), measuring the memorability of phrasing (Danescu-Niculescu-Mizil et al., 2012), question answering (Guha et al., 2015; Kočiskỳ et al., 2017), natural language understanding (Frermann et al., 2017), summarization (Gorinski and Lapata, 2015) and image captioning (Zhu et al., 2015; Rohrbach et al., 2015, 2017; Tapaswi et al., 2015), the modalities examined are almost exclusively limited to text and image. In this work, we present a new perspective on multimodal storytelling by focusing on a so-far neglected aspect of narrative: the role of music.

We focus specifically on the ways in which soundtracks contribute to films,[1] presenting a first look from a computational modeling perspective into soundtracks as storytelling devices. By developing models that connect films with musical parameters of soundtracks, we can gain insight into musical choices both past and future. While a great film score is in part determined by how well it fits with the context of the story (Byrne, 2012), we are also interested in uncovering musical aspects that, in general, work better *in support of a film*.

To move toward understanding both what makes a film fit with a particular kind of song and what musical aspects can broadly be effective in the service of telling a story, we make the following contributions:

1. We present a dataset of 41,143 films paired with their soundtracks. Metadata for the films is drawn from IMDB, linked to subtitle information from OpenSubtitles2016 data (Lison and Tiedemann, 2016), and soundtrack data is linked to structured audio information from Spotify.
2. We present empirical results demonstrating the relationship between audio qualities of the soundtrack and viewers' responses to

[1] We use the word *film* to refer to both movies and television shows interchangeably.

Proceedings of the First Workshop on Storytelling, pages 33–42
New Orleans, Louisiana, June 5, 2018.

the films they appear in; soundtracks with more instrumental or acoustic songs generate higher ratings; "danceable" songs lower the average ratings for films they appear in.

3. We present empirical results demonstrating the relationship between the topics that make up a film script and the audio qualities of the soundtrack. Films with settings in high school or college, for example, tend to have electric instrumentation and singing; soundtracks with faster tempos appear both in films about zombies and vampires and in films in which the word *dude* appears frequently.

2 The Narrative Role of Music in Film

The first films appeared around 1890, before the development of technology that enabled synchronization of picture with sound (Buhler et al., 2010). While silent films featured no talking or music in the film itself, they were often accompanied by music during live performances in theatres. Rather than playing set scores, these live accompaniments were largely improvised; practical catalogues for such performances describe the musical elements appropriate for emotions and narrative situations in the film (Becce, 1919; Erdmann et al., 1927). For example, Lang and West (1920) note that a string accompaniment with tremolo (trembling) effect is appropriate for "suspense and impending disaster"; an organ tone with heavy pedal is appropriate for "church scenes" and for generally connoting "impressive dignity"; flutes are fitting for conveying "happiness," "springtime" or "sunshine."

With the rise of talkies in the late 1920's (Slowik, 2012), music could be incorporated directly into the production of the film, and was often composed specifically for it; Gorbman (1987) describes that in the classical model of film production, scored music is "not meant to be heard consciously," primarily acts as a signifier of emotion, and provides referential and narrative cues, such as establishing the setting or character. The use of Wagnerian leitmotif—the repeated association of a musical phrase with a film element, such as a character—is common in original scores, especially in those for epic films (Prendergast, 1992).

Works from the "Golden Age" of film music (the period between 1935–1950, shortly after the rise of synchronized sound) set the standard for cinematic scoring practices and have been extensively analyzed in the film music literature (Slowik, 2012). Following this period, with the rise of rock and roll, popular music began to make its way into film soundtracks in addition to the scores written specifically for the movie. As Rodman (2006) points out, this turn coincided with directors seeing the potential for songs to contribute to the narrative meaning of the film:

> In *The Blackboard Jungle*, Bill Haley's rock and roll anthem, 'Rock Around the Clock,' was used in the opening credits, not only to capture the attention of the teenage audience, but also to signify the rebellious energy of teenagers in the 1950s.... *The Graduate* relied upon the music and poetry of Simon and Garfunkel to portray the alienation of American youth of the 1960s. *Easy Rider* took a more aggressive countercultural stance by using the rock music of Hoyt Axton, Steppenwolf, The Byrds, and Jimi Hendrix to portray the youth rebellion in American society, complete with communes, long hair and drugs (Rodman, 2006, 123)

In recent years, the boundaries between popular music and film music in the traditional sense have become increasingly blurred, pushed forward especially by more affordable music production technology including synthesizers and prerecorded samples that allow a broad range of composers to use sounds previously reserved for those with access to a full orchestra (Pinch et al., 2009). Though electronic music pioneers like Wendy Carlos have been composing for film since the late 1960's (Pinch et al., 2009), pop and electronic musicians have only gradually been recognized as film composers in their own right, with Daft Punk's original score for *Tron: Legacy* in 2010 marking a breakthrough into the mainstream (Anderson, 2012).

3 Data

In order to begin exploring the relationship between films and their soundtracks, we gather data from several different sources. First, we draw on the OpenSubtitles2016 data of parallel texts for film subtitles (Lison and Tiedemann, 2016); this dataset includes scripts for a wide variety of movies and episodes of television shows (106,609 total in English) and contains publicly available subtitle data. Each film in the OpenSubtitles2016 data is paired with its unique IMDB identifier; using this information, we extract IMDB metadata for the film, including title, year of release, average user rating (a real number from 0-10), and genre (a set of 28 categories, ranging from drama and comedy to war and film-noir).

Most importantly, we also identify soundtrack information on IMDB using this identifier; soundtracks are listed on IMDB in the same form as they appear in the movie/television credits (generally also in the order of appearance of the song). A typical example is the following:

Golden Slumbers
Written by John Lennon and Paul McCartney
Performed by Jennifer Hudson
Jennifer Hudson appears courtesy of Epic Records
Produced by Harvey Mason Jr.

This structured format is very consistent across films (owing to the codification of the appearance of this information in a film's closing credits, which is thereby preserved in the user transcription on IMDB[2]). For each song in a soundtrack for a film, we extract the title, performers and writers through regular expressions (which are precise given the structured format).

We then identify target candidate matches for a source soundtrack song by querying the public Spotify API for all target songs in the Spotify catalogue with the same title as the source song in the IMDB soundtrack. The names of performers are not standardized across datasets (e.g., IMDB may list an artist as *The Velvet Underground*, while Spotify may list the same performance as *The Velvet Underground and Nico*). To account for this, we identify exact matches between songs as those that share the same title and where the longest common substring between the source and target performers spans at least 75% the length of either entity; if no exact match is found, we identify the best secondary match as the target song with the highest Spotify popularity among target candidates with the same title as the source. In the example above, if this particular performance of *Golden Slumbers* by Jennifer Hudson (from the movie *Sing*) were not in Spotify's catalogue, it would match the performance by The Beatles on *Abbey Road*.

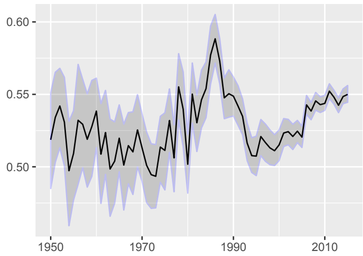
Spotify provides a number of extracted audio features for each song; from a set of 13 we chose 5 that we hypothesized would be predictive of viewer preferences and whose descriptors are also interpretable enough to enable discussion. Those that we include in our analysis are the following, with descriptions drawn from Spotify's Track API:[3]

- **Mode**. "Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0."
- **Tempo**. "The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration." In the raw data, these range from 36 to 240; we divide by the maximum value of 240 to give a range between 0.15 and 1.
- **Danceability**. "Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable."
- **Instrumentalness**. "Instrumentalness predicts whether a track contains no vocals. 'Ooh' and 'aah' sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly 'vocal.' The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0."
- **Acousticness**. "A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic."

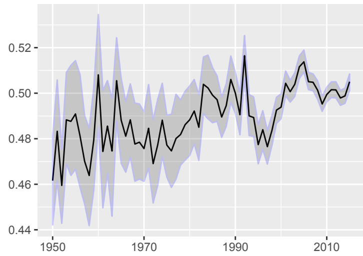
The dataset totals 189,340 songs (96,526 unique) from 41,143 movies/television shows,

[2] https://help.imdb.com/article/contribution/titles/soundtracks/GKD97LHE9TQ49CZ7

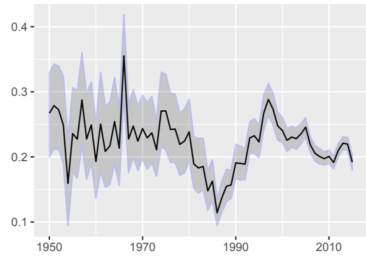
[3] https://developer.spotify.com/web-api/get-audio-features/



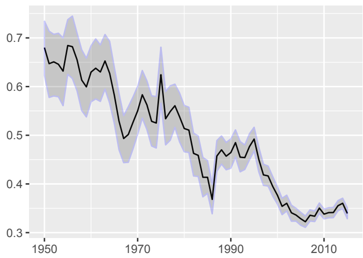
(a) Average danceability



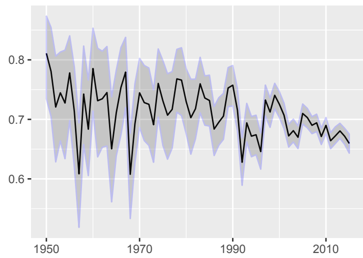
(b) Average tempo



(c) Average instrumentalness



(d) Average acousticness



(e) Average mode (major = 1)

Figure 1: Change in audio features over time, 1950–2015. Films and TV shows in the late 1980s peaked for danceable soundtracks, with electric instrumentation and singing. Each plot displays the average value for that feature, with 95% bootstrap confidence intervals.

along with a paired script for each movie/show. Figures 1 and 2 provide summary statistics of this dataset (using only the metadata and audio features) and begin to demonstrate the potential of this data. Figure 1 illustrates the change in the average value of each feature between 1950–2015. Soundtracks featuring acoustic songs naturally decline over this time period with the rise of electric instruments; as time progresses, soundtracks feature quicker tempos and include more songs in minor keys. The 1980s in particular are peaks for danceable soundtracks, with electric instrumentation and voice, while the 1990s appear to react against this dominance by featuring songs with comparatively lower danceability, higher acoustic instrumentation, and less singing.

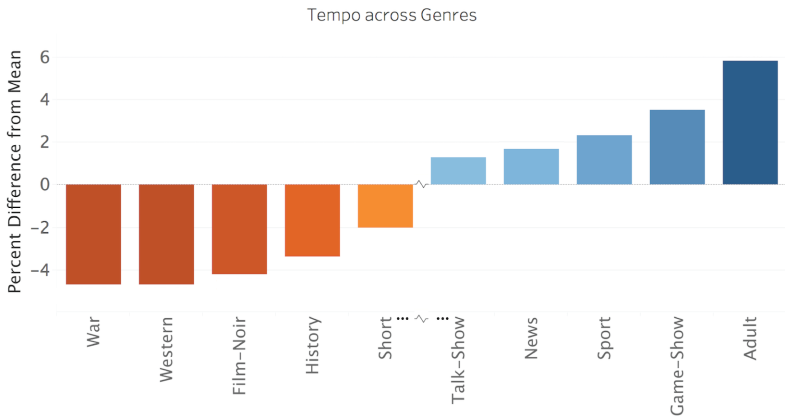
Figure 2, in contrast, displays variations in selected audio features across genres. Movies and television shows tagged by IMDB users with genre labels of "war" and "western" tend to have songs that are slow and in major keys, whereas game shows and adult films more often have faster songs in minor keys. We can also see from figure 2 that different audio characteristics can have different amounts of variation across genres; mode varies more with genre than tempo does.

4 Analysis

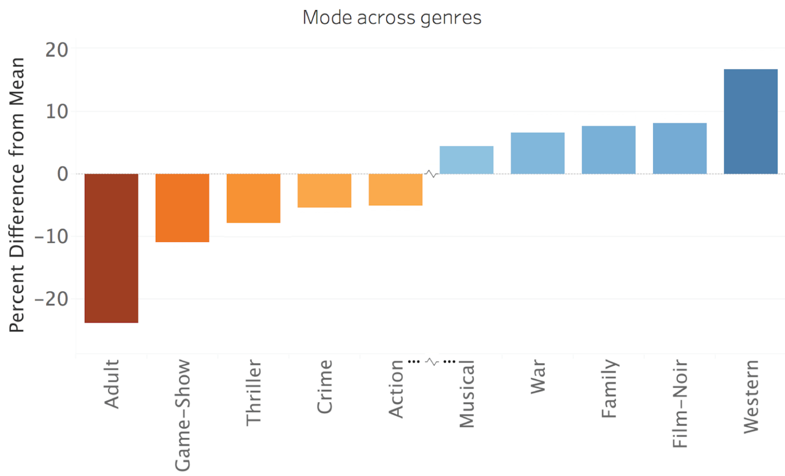
We present two analyses here shedding some light on the role that music plays in the narrative form of films, demonstrating the relationship between fine-grained topics in a film's script and specific audio features described in §3 above; and measure the impact of audio features in the soundtrack on the *reception* to the storytelling in the form of user reviews, attempting to control for the topical and generic influence of the script using topically-coarsened exact matching in causal inference (Roberts et al., 2016).

4.1 Topic analysis of audio features

While we can expect to see trends in the music employed in films over time and across genres, these surface descriptors do not tell us about the actual *contents* of the film. What kind of stories have soundtracks in minor keys? Is there a relationship between the content of a movie or television show and the tempo of its soundtrack? We investigate this by regressing the content of the script against each audio feature; rather than representing the script by the individual words it contains, we seek instead to uncover the relationship between broad thematic topics implicit in the script and those fea-



(a) Top and bottom 5 film genres with the fastest and slowest soundtracks. Blue indicates faster tempos and red indicates slower tempos.



(b) Top and bottom 5 film genres whose soundtracks are most major and most minor. Blue indicates major and red indicates minor.

Figure 2: Top and bottom 5 film genres in terms of average tempo and mode. Heights of the bars represent percentage difference from the mean across the entire dataset. Actual values are at the tops of the bars.

tures.

To do so, we identify topics in all 41,143 scripts using LDA (Blei et al., 2003), modeling each script as a distribution over 50 inferred topics, removing stopwords and names.

We model the impact of individual topics on audio features by representing each document as its 50-dimensional distribution of topic proportions; in order to place both very frequent and infrequent topics on the same scale, we normalize each topic to a standard normal across the dataset. For each of the four real-valued audio features $y = \{danceability, acousticness, instrumentalness, \text{and } tempo\}$, we regress the relationship between the document-topic representations in turn using OLS; for the binary-valued *mode* variable (major vs. minor key), we model the relationship between the topic representations using binary logistic regression.

Table 1 illustrates the five strongest positive and negative topic predictors for each audio feature.

While the latent topics do not have defined categories, we can extract salient aspects of the stories based on the words in the most prevalent topics. The words describing the topics most and least associated with an audio feature can give us some insight into how songs are used in soundtracks.

- **Mode**. Major versus minor is typically one of the most stark musical contrasts, with the major key being characteristically associated with joy and excitement, and the minor key being associated with melancholy (Hevner, 1935). We see songs in major being used in

0.141	captain ship sir
0.074	town horse sheriff
0.055	sir colonel planet
0.050	boy huh big
0.050	mr. sir mrs.
-0.045	sir brother heart
-0.052	leave understand father
-0.054	gibbs mcgee boss
-0.066	agent security phone
-0.083	baby yo y'all

(a) mode (major/minor)

0.044	mr. sir mrs.
0.028	boy huh big
0.020	christmas la aa
0.019	sir dear majesty
0.019	leave understand father
-0.017	um work fine
-0.018	kill dead blood
-0.020	fuck shit fucking
-0.021	school class college
-0.024	dude cool whoa

(b) acousticness

0.017	baby yo y'all
0.010	woman married sex
0.008	sir brother heart
0.007	dude cool whoa
0.006	spanish el la
-0.006	father lord church
-0.007	remember feel dead
-0.008	sir dear majesty
-0.011	captain ship sir
-0.015	sir colonel planet

(c) danceability

0.003	dude cool whoa
0.003	sir colonel planet
0.002	music show sing
0.002	um work fine
0.002	kill dead blood
-0.002	baby yo y'all
-0.002	remember feel dead
-0.002	sir dear majesty
-0.004	mr. sir mrs.
-0.005	captain ship sir

(d) tempo

0.025	sir colonel planet
0.020	mr. sir mrs.
0.017	years world work
0.017	captain ship sir
0.016	agent security phone
-0.011	school class college
-0.012	baby yo y'all
-0.013	woman married sex
-0.014	sir brother heart
-0.015	music show sing

(e) instrumentalness

Table 1: Script topics predictive of audio features. For each feature, the 5 topics most predictive of high feature values (1-5) and the 5 topics most predictive of low feature values (46-50) are shown. Topics are displayed as the top three most frequent words within them. Coefficients for mode (as a categorical variable) are for binary logistic regression; those for all other features are for linear regression.

films with polite greetings, those with sheriffs and horses, and with ship captains. Minor songs appear more often in stories involving separation from parents, FBI agents, or viruses. The strongest topic associated with major key is *"captain ship"*; productions mostly strongly associated with this topic include episodes from the TV show *Star Trek: The Next Generation*.

- **Acousticness**. The acousticness of a soundtrack captures the degree of electric instrumentation; as figure 4(d) shows, acousticness shows the greatest decline over time (corresponding to the rise of electric instruments); we see this also reflected topically here, with the topic most strongly associated with acoustic soundtracks being *"mr. sir mrs."*; this topic tends to appear in period pieces and older films, such as *Arsenic and Old Lace* (1944).
- **Danceability**. The danceability of a song is the degree to which is it suitable for dancing. The topics most strongly associated with consistently danceable soundtracks including *"baby yo"*—dominant in movies like *Malibu's Most Wanted* (2003), *Menace II Society* (1993) *Hustle & Flow* (2005)—and *"women married"*—dominant in episodes of *Friends* and *Sex and the City*.
- **Tempo**. Musical tempo is a measure of pace; the strongest topic associated with fast pace is the *"dude cool"* topic, include episodes from the TV show *Workaholics* and *The Simpsons*. Perhaps unsurprisingly, the mannered *"mr. sir mrs."* topic is associated with a slow tempo.
- **Instrumentalness**. Instrumentalness measures the degree to which a song is entirely instrumental (i.e., devoid of vocals like singing), such as classical music. The *"mr. sir mrs."* topic again rates highly along this dimension (presumably corresponding with the use of classical music in these films); also highly ranking is the *"sir colonel"* topic, which is primarily a subgenre of science fiction, including episodes from the TV show *Stargate* and the movie *Star Wars: Episode III: Revenge of the Sith* (2005).

4.2 Impact on ratings

While the analysis above examines the internal relationship between a film's soundtrack and its narrative, we can also explore the relationship between the soundtrack and a film's reception: how do audiences respond to movies with fast soundtracks, to acoustic soundtracks, or to soundtracks that are predominantly classical (i.e., with high instrumentalness)? We measure response in this work by the average user rating for the film on IMDB (a real value from 0-10).

One confound with this kind of analysis is the complication with the content of the script; as §4.1 demonstrates, some topics are clearly associated with audio features like "acousticness," so if we identify a relationship between acousticness and a film's rating, that might simply be a relationship between the underlying topic (e.g., *"dude," "cool," "whoa"*) and the rating, rather than acousticness in itself.

In order to account for this, we employ methods from causal inference in observational studies, drawing on the methods of coarsened exact matching using topic distributions developed by Roberts et al. (2016). Conventional methods for exact matching aim to identify the latent causal experiment lurking within observational data by eliminating all sources of variation in covariates except for the variable of interest, identifying a subset of the original data in which covariates are balanced between treatment conditions; in our case, if 100 films have high tempo, 100 have low tempo and the 200 films are identical in every other dimension, then if tempo has a significant correlation with a film's rating, we can interpret that significance causally (since there is no other source of variation to explain the relationship).

True causal inference is dependent on accurate model specification (e.g., its assumptions fail if an important explanatory covariate is omitted). In our case, we are seeking to model the relationship between audio features of the soundtrack and IMDB reviewers' average rating for a film, and include features representing the content of a film through a.) its topic distribution and b.) explicit genre categories from IMDB (a binary value for each of the 28 genres). We know that this model is mis-specified—surely other factors of a film's content impacting its rating may also be correlated with audio features—but in using the machinery of causal inference, we seek not to make causal claims but rather to provide a stricter criterion against which to assess the significance of our results.

Here, let us define the "treatment variable" to be the variable (such as "acousticness") whose relationship with rating we are seeking to establish. The original value for this variable is real; we binarize it into two treatment conditions (0 and 1) by thresholding at 0.5 (all values above this limit are set to 1; otherwise 0). To test the relationship between audio features and user ratings in this procedure, we place each data point in a stratum defined by the values of its other covariates; we coarsen the values of each covariate into a binary value: for all numeric audio features, we binarize at a threshold of 0.5; for topic distributions, we coarsen by selecting the argmax topic as the single binary topic value. For each stratum with at least 5 data points in each treatment condition, we sample data points to reflect the overall distribution of the treatment variable in the data; any data points in strata for which there are fewer than 5 points from each condition are excluded from analysis. This, across all strata, defines our matched data for analysis. We carry out this process once for each treatment variable $\{mode$, $danceability$, $acousticness$, $instrumentalness$, and $tempo\}$.

Coefficient	Audio feature
0.121*	Acousticness
0.117*	Instrumentalness
0.031	Tempo
0.024*	Mode
−0.103*	Danceability

Table 2: Impact of audio features on IMDB average user rating. Features marked * are significant at $p \leq 0.001$.

Table 2 presents the results of this analysis: all audio features of the soundtrack except tempo have a significant (if small) impact on the average user rating for the film they appear with. A highly danceable soundtrack would lower the score of a film from 9.0 to a 8.897; adding an acoustic soundtrack would raise it to 9.121; and adding an instrumental soundtrack with no vocals would raise it to 9.117. Our experiments suggest that certain musical aspects might be generally more effective than others in the context of a film score, and that these attributes significantly shape a viewer's reactions to the overall film.

5 Previous Work

Because the ability to forecast box office success is of practical interest for studios who want to decide which scripts to "Greenlight" (Eliashberg et al., 2007; Joshi et al., 2010) or where to invest marketing dollars (Mestyán et al., 2013), a number of previous studies look at predicting movie success by measuring box office revenue or viewer ratings. Eliashberg et al. (2007) used linear regression on metadata and textual features drawn from "spoilers", detailed summaries written by moviegoers, to predict return on investment in movie-making. Joshi et al. (2010) used review text from critics to predict revenues, finding n-gram and dependency relation features to informative complement to metadata-based features. Jain (2013) used Twitter sentiment to predict box office revenue, further classifying movies as either a Hit, a Flop, or Average. Oghina et al. (2012) used features computed from Twitter and YouTube comments to predict IMDB ratings.

Though a number of previous works have attempted to predict film performance from text and metadata, little attention has been paid to the role of the soundtrack in a movie's success. Xu and Goonawardene (2014) did consider soundtracks, finding the volume of internet searches for a movie's soundtrack preceding a release to be predictive of box office revenue. This work, however, only considers the popularity of the soundtrack as a surface feature; it does not directly measure whether the *musical characteristics* of the songs in the soundtracks are themselves predictive.

6 Conclusion

In this work, we introduce a new dataset of films, subtitles, and soundtracks along with two empirical analyses, the first demonstrating the connections between the contents of a story, (as measured by the topics in its script) and the musical features that make up its soundtrack, and the second identifying musical aspects that are associated with better user ratings on IMDB. Soundtracks using acoustic instruments, as measured by Spotify's "acousticness" descriptor, and those with instruments but no vocals, as measured by the "instrumentalness" descriptor, are each linked with more than a 0.11 increase in ratings on a 10-star scale, even when controlling for other musical dimensions, topic, and genre through Coarsened Exact Matching. Soundtracks that are more "danceable" point in the opposite direction, indicating a decrease of 0.1 stars.

We hope that one of the primary beneficiaries of the line of work introduced here will be *music supervisors*, whose job involves choosing existing music to license or hiring composers to create original scores. Understanding the connections between the different modalities that contribute to a story can be useful for understanding the history of film scoring and music licensing as well as for making decisions during the production process. Though traditionally the music supervisor plays a well-defined role on a film, in contemporary practice many people contribute to music supervision throughout the production process for all kinds of media, from movies and television to advertising, social media, and games.

There are several directions of future research that are worth further pursuit. First, while we have shown that strong relationships exist between films and their soundtracks as a whole and that a soundtrack is predictive of user ratings, this relationship only obtains over the entirety of the script and the entirety of the soundtrack; a more fine-grained model would anchor occurrences of individual songs at specific moments in the temporal narrative of the script. While our data does not directly indicate *when* a song occurs, latent variable modeling over the scripts and soundtracks in our collection may provide a reasonable path forward. Second, while our work here has focused on descriptive analysis of this new data, a potentially powerful application is *soundtrack generation*: creating a new soundtrack for a film given the input of a script. This application has the potential to be useful for music supervisors, by suggesting candidate songs that fit the narrative of a given script in production.

Music is a vital storytelling component of multimodal narratives such as film, television and theatre, and we hope to drive further work in this area. Data and code to support this work can be found at `https://github.com/jrgillick/music_supervisor`.

7 Acknowledgments

Many thanks to the anonymous reviewers for their helpful feedback. The research reported in this article was supported by a UC Berkeley Fellowship for Graduate Study to J.G.

References

Apoorv Agarwal, Jiehan Zheng, Shruti Kamath, Sriramkumar Balasubramanian, and Shirin Ann Dey. 2015. Key female characters in film have more to talk about besides men: Automating the bechdel test. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Denver, Colorado, pages 830–840. http://www.aclweb.org/anthology/N15-1084.

Stacey Anderson. 2012. Electronic dance music goes hollywood. *The New York Times* .

David Bamman, Brendan O'Connor, and Noah A. Smith. 2013. Learning latent personas of film characters. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Sofia, Bulgaria, pages 352–361.

Giuseppe Becce, editor. 1919. *Kinothek. Neue Filmmusik*. Schlesinger'sche Buchhandlung.

David M. Blei, Andrew Ng, and Michael Jordan. 2003. Latent dirichlet allocation. *JMLR* 3:993–1022.

James Buhler, David Neumeyer, and Rob Deemer. 2010. *Hearing the movies: music and sound in film history*. Oxford University Press New York.

David Byrne. 2012. *How music works*. Three Rivers Press.

Cristian Danescu-Niculescu-Mizil, Justin Cheng, Jon Kleinberg, and Lillian Lee. 2012. You had me at hello: How phrasing affects memorability. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '12, pages 892–901. http://dl.acm.org/citation.cfm?id=2390524.2390647.

Jehoshua Eliashberg, Sam K Hui, and Z John Zhang. 2007. From story line to box office: A new approach for green-lighting movie scripts. *Management Science* 53(6):881–893.

Hans Erdmann, Giuseppe Becce, and Ludwig Brav. 1927. *Allgemeines Handbuch der Film-Musik*. Schlesinger.

Lea Frermann, Shay B Cohen, and Mirella Lapata. 2017. Whodunnit? crime drama as a case for natural language understanding. *arXiv preprint arXiv:1710.11601* .

Claudia Gorbman. 1987. *Unheard melodies: Narrative film music*. Indiana University Press.

Philip John Gorinski and Mirella Lapata. 2015. Movie script summarization as graph-based scene extraction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pages 1066–1076.

Tanaya Guha, Che-Wei Huang, Naveen Kumar, Yan Zhu, and Shrikanth S Narayanan. 2015. Gender representation in cinematic content: A multimodal approach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, pages 31–34.

Kate Hevner. 1935. The affective character of the major and minor modes in music. *The American Journal of Psychology* 47(1):103–118. http://www.jstor.org/stable/1416710.

Vasu Jain. 2013. Prediction of movie success using sentiment analysis of tweets. *The International Journal of Soft Computing and Software Engineering* 3(3):308–313.

Mahesh Joshi, Dipanjan Das, Kevin Gimpel, and Noah A Smith. 2010. Movie reviews and revenues: An experiment in text regression. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 293–296.

Tomáš Kočiskỳ, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, and Edward Grefenstette. 2017. The narrativeqa reading comprehension challenge. *arXiv preprint arXiv:1712.07040* .

Edith Lang and George West. 1920. *Musical accompaniment of moving pictures; a practical manual for pianists and organists and an exposition of the principles underlying the musical interpretation of moving pictures, by*. The Boston Music Co., Boston.

Pierre Lison and Jrg Tiedemann. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA), Paris, France.

Márton Mestyán, Taha Yasseri, and János Kertész. 2013. Early prediction of movie box office success based on wikipedia activity big data. *PloS one* 8(8):e71226.

Andrei Oghina, Mathias Breuss, Manos Tsagkias, and Maarten de Rijke. 2012. Predicting imdb movie ratings using social media. In *European Conference on Information Retrieval*. Springer, pages 503–507.

Trevor J Pinch, Frank Trocco, and TJ Pinch. 2009. *Analog days: The invention and impact of the Moog synthesizer*. Harvard University Press.

Roy M Prendergast. 1992. *Film music: a neglected art: a critical study of music in films*. WW Norton & Company.

Anil Ramakrishna, Nikolaos Malandrakis, Elizabeth Staruk, and Shrikanth Narayanan. 2015. A quantitative analysis of gender differences in movies using psycholinguistic normatives. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 1996–2001. https://aclweb.org/anthology/D/D15/D15-1234.

Margaret E. Roberts, Brandon Stewart, and R. Nielsen. 2016. Matching methods for high-dimensional data with applications to text. Working paper.

Ronald Rodman. 2006. The popular song as leitmotif in 1990s film. *Changing Tunes: The Use of Pre-existing Music in Film* pages 119–136.

Anna Rohrbach, Marcus Rohrbach, Niket Tandon, and Bernt Schiele. 2015. A dataset for movie description. *CVPR* .

Anna Rohrbach, Atousa Torabi, Marcus Rohrbach, Niket Tandon, Christopher Pal, Hugo Larochelle, Aaron Courville, and Bernt Schiele. 2017. Movie description. *International Journal of Computer Vision* 123(1):94–120. https://doi.org/10.1007/s11263-016-0987-1.

Michael James Slowik. 2012. Hollywood film music in the early sound era, 1926-1934 .

Makarand Tapaswi, Martin Bauml, and Rainer Stiefelhagen. 2015. Book2movie: Aligning video scenes with book chapters. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Haifeng Xu and Nadee Goonawardene. 2014. Does movie soundtrack matter? the role of soundtrack in predicting movie revenue. In *PACIS*. page 350.

Yukun Zhu, Ryan Kiros, Richard S. Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. *ICCV* .