JOHANNES KEPLER
UNIVERSITY LINZ

Author
DI **Firstname Lastname**,
BSc
12345678

Submission
Institute of
Networks and Security

Thesis Supervisor
Prof. Dr.
Firstname Lastname

Assistant Thesis Supervisor
Dr. **Firstname Lastname**,
MSc

September 2025

Sentiment and Linguistic Patterns in YouTube Music Video Comments Across Genres

Master's Thesis

to confer the academic degree of

Master of Science

in the Master's Program

Computer Science

JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Straße 69
4040 Linz, Austria
jku.at

Contents

Abstract iii

1 Introduction 1
- 1.1 Motivation 1
- 1.2 Objectives and Research Questions 1

2 Related Work 3

3 Dataset 6
- 3.1 Data Sources 6
- 3.2 Representative Song Selection 6
- 3.3 Comment Collection & Metadata 6
- 3.4 Preprocessing Pipeline 7
- 3.5 Feature Extraction 7
- 3.6 Final Dataset Characteristics 8

4 Methods 9
- 4.1 Sentiment Scoring (VADER) 9
- 4.2 Psycholinguistic Features (LIWC) 10
- 4.3 Engagement Metrics 10
- 4.4 Polarization Metrics 10
- 4.5 Emoji-Usage Analysis Design 11

5 Results 12
- 5.1 Engagement–Sentiment Correlation 12
- 5.2 Sentiment Distributions Across Genres 15
- 5.3 Polarity and Polarization 15
- 5.4 Hip-Hop Anomaly 17
- 5.5 Views and Sentiment 17
- 5.6 Emoji Usage and Sentiment 18
- 5.7 LIWC Dimensions and Cultural Markers 19
- 5.8 LIWC–VADER Correlations 20

6 Conclusion and Outlook 22
- 6.1 Key Findings 22
- 6.2 Limitations 23
- 6.3 Outlook 23

Bibliography 25

Appendix A Representative Song Selection Algorithm 27

Abstract

Music is closely linked to emotional experience and the emergence of social communities, and in the digital age, platforms such as YouTube have become central spaces for music consumption and discussion. YouTube comments in particular provide a large-scale record of spontaneous audience reactions, offering valuable insights into how listeners across different musical genres express themselves online.

This thesis investigates sentiment and linguistic patterns in YouTube comments on music videos, with a focus on differences between genres. A representative dataset was constructed by combining the JKU dataset of YouTube music links with the Music4All-Onion dataset for genre classification. Using a conflict-resolution algorithm, 50 representative videos were selected for each of 260 genres, yielding over 11,000 videos. From these, 100 recent comments per video were collected via the YouTube Data API, resulting in a dataset of 356,973 english comments across 233 genres. Each comment was preprocessed and analyzed using VADER for sentiment scoring and LIWC for psycholinguistic feature extraction.

The results highlight clear genre-specific differences in sentiment. Spiritual and smooth genres consistently received the most positive comments, while hip-hop and aggressive genres showed the lowest sentiment scores despite high engagement. Emoji usage was strongly associated with positive sentiment, though excessive use reduced this effect.

Chapter 1

Introduction

1.1 Motivation

Music is one of the most powerful and universal forms of human expression. It shapes cultural identity, regulates emotions, and forms social connection across communities. In the digital era, music consumption has shifted to online platforms, with YouTube emerging as the largest global centre for music streaming and community discussions. With more than 1.2 billion monthly users, YouTube not only gives people access to music but also give them the chance to present some feedback in form of comments, likes, and shares.

These comments represent a unique opportunity to study audience engagement and genre-specific communication styles. Unlike professional reviews, they capture spontaneous, authentic reactions from listeners across diverse genres and backgrounds. Understanding how these reactions differ across genres provides insight into how music communities construct meaning, express emotions, and build social connections.

However, analyzing such data is challenging. Social media comments are highly informal, often including slang, emojis, or irony, which complicates interpretation. Traditional sentiment analysis methods—such as lexicon-based models—may fail to capture the nuances of genre-specific language, particularly in culturally distinct communities such as hip-hop. This thesis addresses these challenges by combining computational sentiment analysis with psycholinguistic feature extraction, aiming to reveal how music genres differ in their online communication.

1.2 Objectives and Research Questions

The main objective of this thesis is to investigate how sentiment and linguistic patterns vary across YouTube music video comments, and how these patterns relate to engagement and genre. Based on this objective, the following research questions are addressed:

1. **Sentiment and Engagement:** How is the sentiment of YouTube music video comments related to audience engagement, such as the number of likes the comments receive?
2. **Sentiment Differences Across Genres:** How do average sentiment scores differ across musical genres, and which genres show the most positive or negative patterns?
3. **Polarization:** How polarized are comments within individual messages and across entire genres, and what does this reveal about community communication styles?

4. **Psycholinguistic Features:** Which psycholinguistic features (e.g., swearing, social words, conflict terms) distinguish genres, and how do these features correlate with sentiment?
5. **Emoji Usage:** How is emoji usage associated with the sentiment of comments, and is there an optimal range of emoji use linked to more positive expressions?
6. **Popularity and Sentiment:** Do more popular music videos (measured by view counts) tend to attract more positive or negative comments?

These questions guide the dataset construction, analysis design, and interpretation of results presented in the following chapters.

Chapter 2

Related Work

Social Media Comments Analysis

The analysis of social media comments has become an increasingly important research field. Most existing studies have focused on platforms such as Twitter and Meta-platforms, including Facebook and Instagram. With more than 1.2 billion monthly users, YouTube is currently the most visited social media platform in the world and the second most visited website overall after Google Search [13].

Research on YouTube comment analysis has primarily focused on sentiment analysis, spam detection, and user engagement patterns. Recent studies have proposed advanced methods for filtering spam content from comments [11].

Another growing research field is toxicity and hate-speech detection. Hartvigsen et al. introduced TOXIGEN, a large-scale dataset of machine-generated and human-written statements designed to capture both explicit and implicit toxic language. Their results showed that models fine-tuned on TOXIGEN outperform baselines in detecting subtle forms of hate speech [5].

VADER-based Sentiment Analysis

Many researchers use VADER for analyzing short, informal, and emoji-rich social media text because it handles intensifiers, negation, punctuation and emoticons directly in its rule set [8]. Beyond the original Twitter-style validation, VADER has been widely applied to YouTube comments to quantify audience polarity. For instance, Chalkias *et al.* (2023) examined 167,987 comments from educational YouTube channels using both VADER and TextBlob, finding that neutral sentiment dominates and that VADER tends to report more neutral statements compared to TextBlob [4]. Another recent study by Zhang (2025) compared VADER and TextBlob on approximately 18,000 YouTube video comments and correlated sentiment with likes and views, showing that while both tools provide useful signals, TextBlob produced more stable correlations in this context [14]. In these settings, the compound score serves as a practical and interpretable score for overall valence, enabling large-scale comparisons between topics and channels.

LIWC-based Linguistic Profiling

In addition to valence detection, LIWC provides psychologically grounded categories spanning affect, social processes, cognitive mechanisms, authenticity, and topical concerns. The approach has been extensively validated and used to profile communication styles and psychological correlates in natural language across online platforms [10, 12].

In the context of YouTube, LIWC has been used to uncover community norms, such as social bonding, swearing/hostility markers, or religiosity, that are not captured by sentiment alone, thus offering a richer view of how users express themselves. For example, Chae *et al.* (2024) analyzed COVID-19 video transcripts and comments from medical YouTubers using LIWC categories including analytical thinking and emotion (anxiety, anger, sadness), finding that these linguistic/emotional dimensions are associated with viewer engagement and emotional alignment between channel creators and audiences [3].

Why Combine VADER and LIWC?

VADER and LIWC answer complementary questions. VADER provides a robust social media-tuned estimate of the valence of comments (positive / negative / neutral) with a single comparable compound score suitable for large-scale analyses [8]. LIWC, decomposes language into psychologically meaningful dimensions (e.g., *Affect*, *Swear*, *Social*, *Cognition*), allowing one to interpret *how* communities communicate, not just *how positive or negative* they are [10].Combining both methods makes it possible to map sentiment hierarchies across genres, explain those differences through concrete linguistic markers, and relate community endorsement (likes) to either overall valence or specific stylistic and psycholinguistic features.

Music-related Social Media Research

Music-related videos on YouTube generate distinct engagement behaviors compared to other types of content. Research in music psychology has shown that music listening is driven by specific psychological functions such as mood regulation, identity expression, and social connection. These motivations strongly shape how listeners engage with and respond to music content. Listeners actively select music according to explicit listening intentions and seek playlists aligned with their goals, as shown by the ExIM study [6]. Similarly, Arif et al. (2024) conducted a content analysis of BTS music video comments and found strong evidence of parasocial interactions, with fans frequently expressing authenticity, affection, and social bonding with the artist, highlighting the social dimension of the music comment culture [1].

Extending these insights to YouTube more broadly, comments on music videos often contain explicit emotional expressions and personal accounts of why users listen to specific songs. This suggests that the psychological drivers of music listening identified in controlled studies can also be observed directly in user-generated content.

Bauer and Schedl (2019) further contributed to understanding genre-based behavior by introducing the concept of global and country-specific *mainstreaminess*. Using large-scale Last.fm data, they defined measures to capture how closely individual user preferences align with global or local popularity trends. Their results revealed strong cross-country variation, with some regions showing alignment with international mainstream genres, while others emphasized localized and niche listening patterns. This demonstrates that genre preferences are not only individual but also shaped by cultural and community-level contexts [2].

Summary

These studies show that music-related comments on YouTube provide more than casual reactions. They reflect emotional transmission, parasocial interaction, and culturally shaped genre preferences. Building on these insights and leveraging the complementary strengths of VADER and LIWC, this thesis analyzes YouTube music video comments to examine how sentiment and linguistic patterns vary across musical genres and how different communities express themselves within these online spaces.

Chapter 3

Dataset

3.1 Data Sources

This study combines three primary data sources to create a comprehensive dataset for analyzing sentiment patterns across music genres. The **JKU Dataset** (*id_youtube_url.csv*) provides unique identifiers paired with YouTube URLs linking to music videos. The Music4All-Onion dataset [9] connects these songs to genres based on audio features, enabling systematic genre classification. Finally, the **YouTube Data API v3** serves as the primary collection mechanism for user-generated comments and engagement metadata.

These sources were selected for their complementary strengths: the JKU Dataset ensures broad musical coverage, the Music4All-Onion Dataset provides objective, audio-feature-based genre classifications avoiding subjective labeling bias, and YouTube represents the world's largest music streaming platform with authentic user interactions.

3.2 Representative Song Selection

The primary objective was to achieve fair representation across all music genres while avoiding bias toward larger, more popular genres that typically dominate music datasets. Larger genres often exhibit more cross-genre overlap and achieve higher relevance scores, which can marginalize underrepresented musical genres.

The selection algorithm was therefore designed around three key principles: (i) ensuring exactly k unique videos per genre (preventing assignment of the same video to multiple genres), (ii) implementing fair conflict resolution when videos were contested by multiple genres, and (iii) maximizing representation of smaller genres through a fallback scoring mechanism that considers alternative options when conflicts arise.

In practice, the algorithm processed 260 initial genres, assigned the top-scoring videos to each genre, detected and resolved conflicts using fallback scores (where genres with weaker alternatives retained contested videos), and refilled losing genres with their next-best unassigned candidates. This iterative process continued until convergence was reached.

The outcome yielded 50 representative videos per genre across all qualifying genres, ensuring balanced representation while preserving video uniqueness in the final dataset. A detailed step-by-step description of the algorithm is provided in Appendix A.

3.3 Comment Collection & Metadata

Using the YouTube Data API v3, up to 100 comments were collected per representative video, targeting the most recent user interactions. This systematic approach ensured uniform sampling across all genres regardless of their individual popularity or engagement levels.

The collection process captured metadata at both the comment and video levels. Comment-level information included unique identifiers, authorship details, timestamps, engagement indicators such as likes, and reply structures. Video-level metadata comprised descriptive attributes (title, description, publication date), content characteristics (duration, definition, captions), engagement statistics (views, likes, comments, favorites), and technical properties (licensing status, privacy settings, and platform-specific flags).

This two-level metadata structure allows for the joint analysis of individual comment characteristics and the broader video context, providing a comprehensive basis for examining how genre-specific factors influence audience engagement. In total, the dataset contains 34 distinct features for each comment.

3.4 Preprocessing Pipeline

3.4.1 Language Filtering

Given that both VADER sentiment analysis and LIWC require English text for accurate processing, only english comments were retained for analysis. This decision, while necessary, introduces a cultural bias by excluding non-English speaking music communities and potentially underrepresenting genres popular in non-English speaking regions. The language filtering reduced the dataset from 620,736 initial comments to 356,973 english comments.

3.4.2 Text Standardization

A systematic text preprocessing pipeline standardized comment text through three stages: whitespace normalization (removing excess spaces, tabs, and line breaks) and case normalization (converting to lowercase for consistent analysis), and. This standardization ensures consistent input for the sentiment and linguistic analysis tools.

3.4.3 Genre Filtering

To ensure statistical reliability, genres with fewer than 500 English comments were excluded from analysis. This threshold balances the need for meaningful sample sizes with overall dataset coverage, reducing the genre count from 260 to 233 while keeping the analysis meaningful. The 27 excluded genres represent primarily niche or non-English-dominant musical styles.

3.5 Feature Extraction

3.5.1 Sentiment Analysis

VADER (Valence Aware Dictionary for sentiment Reasoning) was selected for sentiment analysis due to its specific optimization for social media text and its ability to handle informal language elements such as slang and emojis commonly found in online communication [8]. VADER generates four sentiment scores per comment: positive, negative, and neutral proportions, plus a compound score normalized between -1 (most negative) and +1 (most positive) that serves as the primary sentiment indicator [7].

3.5.2 Linguistic Features

LIWC (Linguistic Inquiry and Word Count) was employed to extract psycholinguistic features across multiple dimensions: psychological processes (cognitive, emotional, social), personal concerns (work, money, religion), linguistic dimensions (word count, pronouns, articles), and grammatical categories (verbs, adjectives, prepositions). These features enable analysis of communication styles and psychological patterns beyond basic sentiment.

3.6 Final Dataset Characteristics

The final dataset comprises 356,973 english comments distributed across 233 music genres, with each genre containing a minimum of 500 comments to ensure statistical validity. Despite efforts to minimize temporal dispersion by scraping only the most recent comments, the dataset still covers multiple years of YouTube activity, providing valuable insight into temporal dynamics in music communication.

Genre representation varies significantly, with popular genres like pop and rock contributing thousands of comments while niche genres approach the 500-comment minimum threshold.

The preprocessing pipeline successfully standardized textual content while preserving essential metadata, creating a robust foundation for comparative sentiment and linguistic analysis across diverse musical communities.

Chapter 4

Methods

This chapter explains the analysis design and statistical methods, describing what was computed and why. The actual results and figures are presented later in the Results chapter.

4.1 Sentiment Scoring (VADER)

We quantify comment-level sentiment using VADER (Valence Aware Dictionary and sEntiment Reasoner) [8]. VADER is a lexicon- and rule-based sentiment analysis tool specifically designed for short, informal, and emoji-rich social media text. Its foundation is a manually created lexicon of words, phrases, slang expressions, and emojis, each of which was assigned a valence score on a scale from -4 (most negative) to $+4$ (most positive) based on ratings from human annotators. This lexicon was validated to capture sentiment in social media communication.

For each comment, VADER returns four sentiment indicators:

- `positive` $\in [0, 1]$: proportion of text that conveys positive valence,
- `negative` $\in [0, 1]$: proportion of text that conveys negative valence,
- `neutral` $\in [0, 1]$: proportion of text that conveys neutral valence,
- `compound` $\in [-1, 1]$: a single summary score representing the overall sentiment score.

By definition, the three proportional scores sum to one:

$$\texttt{positive} + \texttt{negative} + \texttt{neutral} = 1.$$

The `compound` score is calculated by summing the valence scores of all words and emojis in the text, then adjusting them according to syntactic and semantic heuristics (such as negation, degree modifiers, punctuation, capitalization, and contrastive conjunctions). This raw sum is normalized using the following formula [8]:

$$\texttt{compound} = \frac{x}{\sqrt{x^2 + \alpha}},$$

where x is the unnormalized sum of valence scores and $\alpha = 15$ is a normalization constant. This normalization ensures the final `compound` score remains bounded in $[-1, 1]$, with values closer to -1 indicating strong negativity, 0 indicating neutrality, and $+1$ indicating strong positivity.

Importantly, VADER's lexicon also explicitly covers emojis and emoticons. These symbols are assigned valence values according to their typical affective meaning. Heuristic rules further take into account emoji placement, repetition, and punctuation, such that multiple positive emojis or exclamation marks amplify overall positivity, while repeated negative symbols strengthen negativity.

4.2 Psycholinguistic Features (LIWC)

We extract psycholinguistic features using LIWC. We focus on families that are interpretable for community style and affect, including (but not limited to): *Affect*, *Positive Emotion*, *Negative Emotion*, *Tone*, *Social*, *Swear*, *Anger*, *Anxiety*, *Clout*, and *Authentic*.

Normalization. LIWC outputs are proportions (%), which we treat as rates in $[0, 1]$ after dividing by 100 for analysis. For visualization we may display percentages; for statistical modeling we use the $[0, 1]$ rates.

4.3 Engagement Metrics

Engagement–sentiment analysis. We measure how comment sentiment relates to engagement (likes) in each genre using *Spearman rank correlations* (ρ) between like counts and the four VADER dimensions (positive, negative, neutral, compound). Spearman's ρ is less sensitive to outliers and skewed data than Pearson's correlation. For genres with at least $n \geq 500$ comments, we report ρ together with 95% confidence intervals.

4.4 Polarization Metrics

We measure polarization within comments and across comments to capture mixed sentiment at different levels.

4.4.1 Within-Comment Polarization

Let pos and neg denote a comment's VADER positive and negative proportions (with $\text{pos} + \text{neg} + \text{neu} = 1$). We define:

$$\text{polarity}_{\text{comment}} \ = \ 2 \times \min(\text{pos}, \text{neg}) \ \in \ [0, 1]. \tag{4.1}$$

This score approaches 1 when a comment simultaneously contains strong positive and negative components (high ambivalence) and is 0 when sentiment is one-sided.

4.4.2 Between-Comment Polarization

For between-comment polarization, we use proportions *aggregated across all comments in a genre*. Here, $\overline{\text{pos}}$ denotes the average VADER positive proportion and $\overline{\text{neg}}$ the average VADER negative proportion across all comments in that genre. We define:

$$\text{polarity}_{\text{genre}} \ = \ 2 \times \min(\overline{\text{pos}}, \overline{\text{neg}}) \ \in \ [0, 1]. \tag{4.2}$$

High values indicate a genre-level communication style that contains substantial amounts of both positive and negative sentiment across comments.

4.5 Emoji-Usage Analysis Design

We analyze how emoji frequency relates to sentiment. To optimize interpretability and highlight the range with the highest compound scores, we predefined three mutually exclusive groups at the comment level:

1. **No-emoji**: comments with 0 emojis.
2. **Moderate-emoji**: comments with 1–5 emojis.
3. **High-emoji**: comments with > 5 emojis.

Preprocessing Assumptions. All analyses are restricted to english comments, as required by VADER/LIWC, and are performed on the most recent $k = 100$ comments per selected video as per the dataset construction. We discuss implications of these choices in Limitations.

Chapter 5

Results

This chapter reports the empirical findings. All procedures and metrics are defined in Section 4. We first examine how engagement relates to sentiment, then present genre-wise sentiment distributions and polarization, analyze the hip-hop anomaly, study views versus sentiment, investigate emoji usage, and finally connect LIWC patterns with VADER sentiment.

5.1 Engagement–Sentiment Correlation

We analyze the relationship between expressed sentiment and audience engagement, measured by comment `like_count`. For each genre, we compute Spearman rank correlations between likes and the four VADER sentiment dimensions (`positive`, `negative`, `neutral`, and `compound`), with 95% confidence intervals.

Positive Sentiment (VADER_pos)

Genres such as *Nu Jazz*, *Disco*, and *Latin Rock* show positive correlations ($\rho \approx 0.12$), suggesting that positivity is modestly rewarded. By contrast, *Grime* ($\rho = -0.06$), *Teen Pop* ($\rho = -0.07$), and *Video Game Music* ($\rho = -0.10$) exhibit negative associations, where positive comments appear to be less endorsed.

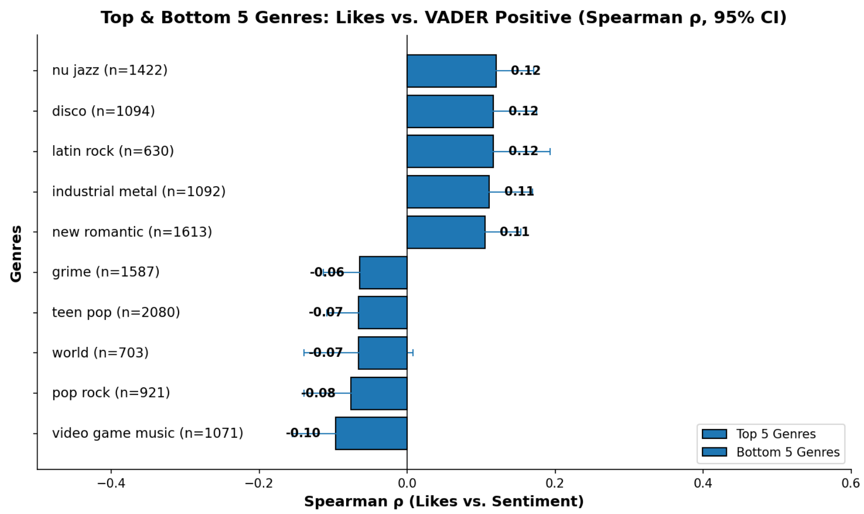

Figure 5.1: Top and bottom five genres by correlation between likes and VADER positive sentiment.

Negative Sentiment (VADER_neg)

Interestingly, *Singer-Songwriter* ($\rho = 0.12$), *Folk* ($\rho = 0.11$), and *Contemporary Classical* ($\rho = 0.10$) show positive correlations, indicating that more negative comments can still gain approval in reflective or critical communities. Conversely, *Country Pop*, *Swing*, and *Latin Rock* all show negative correlations around $\rho \approx -0.08$ to -0.10, where negativity is less validated.

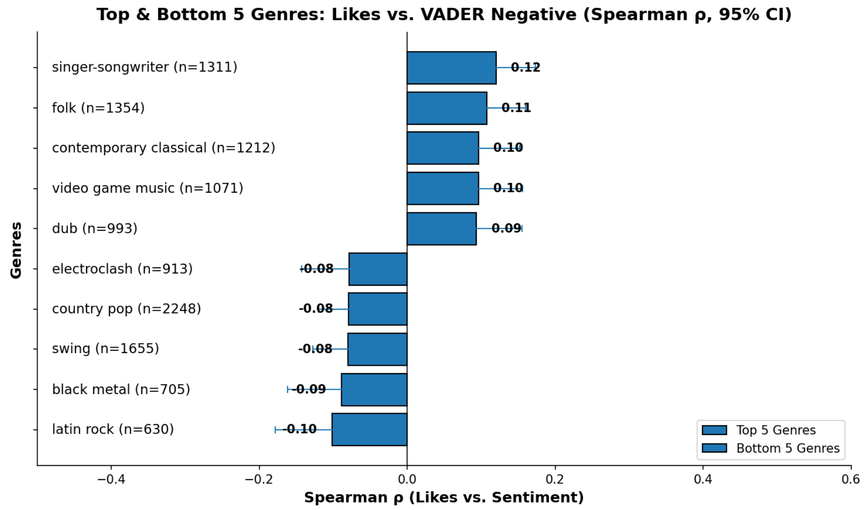

Figure 5.2: Top and bottom five genres by correlation between likes and VADER negative sentiment.

Neutral Sentiment (VADER_neu)

World ($\rho = 0.08$), *Breakbeat* ($\rho = 0.08$), and *Garage Punk* ($\rho = 0.07$) show weak positive correlations, suggesting that neutral-toned comments can still attract likes. By contrast, *Baroque* ($\rho = -0.08$), *Nu Jazz* ($\rho = -0.09$), and *Girl Group* ($\rho = -0.09$) exhibit small negative correlations, where neutrality appears less rewarded.

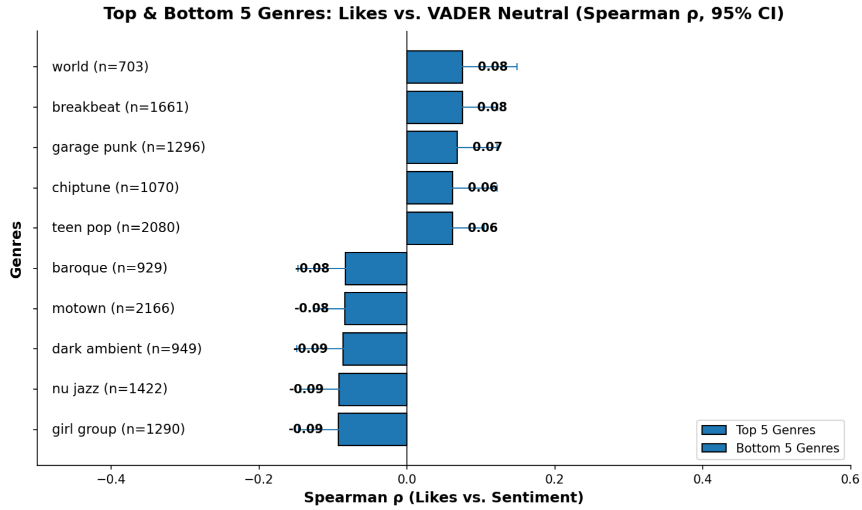

Figure 5.3: Top and bottom five genres by correlation between likes and VADER neutral sentiment.

Compound Sentiment (VADER_compound)

The overall sentiment score shows the strongest patterns. Genres such as *Nu Jazz*, *Disco*, *Jangle Pop*, and *Madchester* all show positive correlations ($\rho \approx 0.17$), while *Latin Rock* follows closely with $\rho = 0.16$. At the other end, *Grime* ($\rho = -0.03$), *Teen Pop* ($\rho = -0.02$), and *Video Game Music* ($\rho = -0.09$) show weak negative correlations.

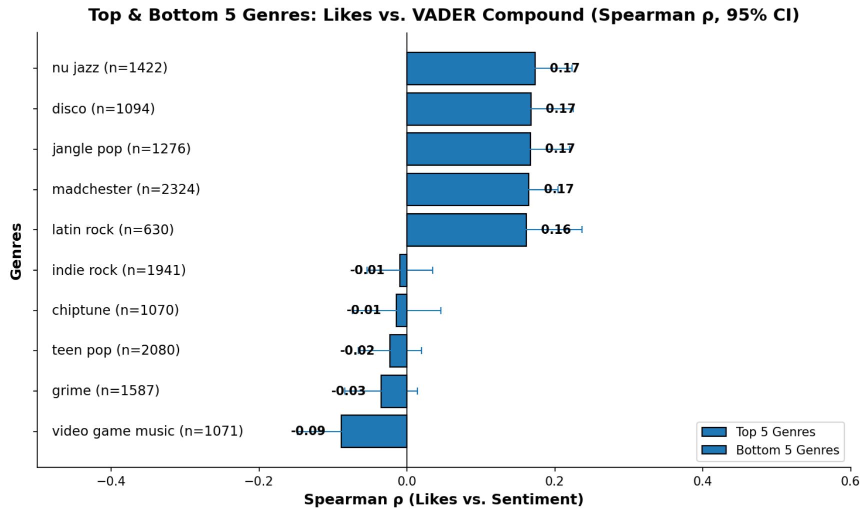

Figure 5.4: Top and bottom five genres by correlation between likes and VADER compound sentiment.

Summary Across all dimensions, correlations remain weak in absolute terms but reveal genre-specific endorsement norms. Positivity is rewarded in some communities (e.g., Nu Jazz, Disco), while others display preferences for neutrality or even negativity (e.g., Singer-Songwriter, Folk). These findings suggest that audience engagement is not uniformly driven by positive expression, but rather reflects culturally shaped communication patterns within each genre. As most comments received few or no likes, these results should be interpreted with caution, since the sparse engagement distribution may attenuate correlations.

5.2 Sentiment Distributions Across Genres

Average VADER `compound` scores across 233 genres vary substantially (from 0.475 to 0.062; a spread of ≈ 0.41 points on the average comment).

Spiritual and smooth genres dominate the positive end: *Worship* (0.475) is highest, followed by *Smooth Jazz* (0.447), *New Age* (0.433), *Gospel* (0.419), and *Vocal Jazz* (0.419). Conversely, heavier and more aggressive styles occupy the lower end: *Death Metal* (0.062) and *Brutal Death Metal* (0.068), followed by *Noise* (0.098), *Rap* (0.103), and *Hardcore Punk* (0.104). Notably, even the lowest averages remain above zero, implying comments are, on balance, slightly positive.

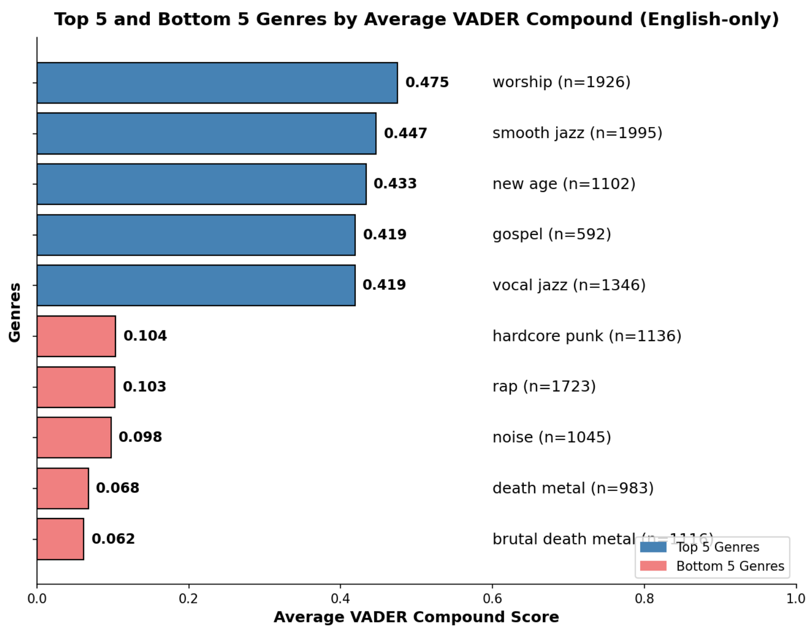

Figure 5.5: Top 5 most positive and bottom 5 lowest-scoring genres by average compound score.

5.3 Polarity and Polarization

We report *within-comment* polarization using Eq. (4.1) and *between-comment* polarization using Eq. (4.2).

Within-Comment Polarization

Extreme or intense styles (e.g., *Death Metal*, *Brutal Death Metal*, *Technical Death Metal*) show the highest within-comment ambivalence, indicating comments often mix strong positive and negative elements. *Spoken Word* and *Doom Metal* also appear among the most polarized, suggesting highly personal narratives or darker atmospheres elicit mixed emotions. By contrast, genres such as *World*, *Cool Jazz*, and *Disco* show the lowest ambivalence, with comments tending to be one-sided.

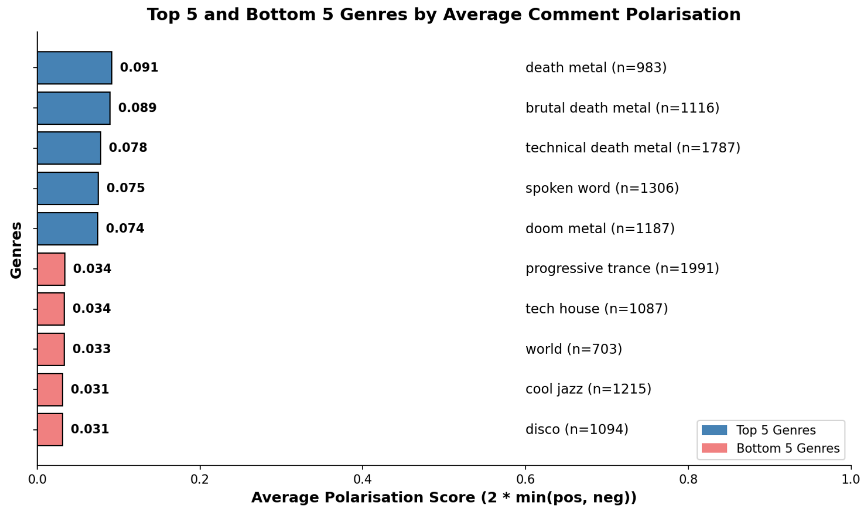

Figure 5.6: Top 5 and bottom 5 genres ranked by within-comment polarization ($2 \times \min(\text{pos}, \text{neg})$).

Between-Comment Polarization

Aggressive styles (*Brutal Death Metal*, *Death Metal*, *Deathcore*) are among the most polarized across comments, indicating community-level divisiveness (strong enthusiasm coexisting with strong criticism). *Noise* and *Emo Rap* also rank highly. By contrast, jazz-related genres as well as World and Progressive Trance show low polarization, indicating that comments there are more uniform in tone.

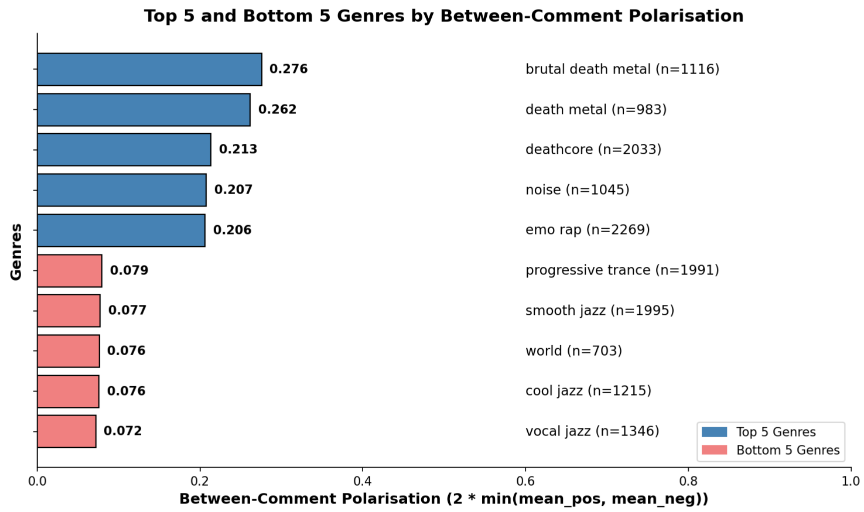

Figure 5.7: Top 5 and bottom 5 genres ranked by between-comment polarization ($2 \times \min(\overline{\text{pos}}, \overline{\text{neg}})$).

5.4 Hip-Hop Anomaly

Hip-hop genres present an unexpected pattern: despite global popularity, their compound sentiment ranks near the bottom across genres.

Table 5.1: Sentiment analysis for selected hip-hop genres.

Genre	**Positive Score (Rank)**	**Negative Score (Rank)**	**Neutral Score (Rank)**	**Compound Score (Rank)**
Hip Hop	0.169 (223/233)	0.080 (39/233)	0.750 (18/233)	**0.173 (215/233)**
Alternative Hip Hop	0.173 (217/233)	0.085 (23/233)	0.742 (36/233)	**0.154 (222/233)**
Experimental Hip Hop	0.156 (231/233)	0.101 (7/233)	0.743 (33/233)	**0.124 (227/233)**

Traditional hip-hop performs best among the subgenres with an average compound of 0.173 (rank 215/233), while experimental hip-hop ranks lowest at 0.124 (rank 227/233), also showing the 7th highest negative score overall.

Data Anomaly Check. One user contributed 66 of 5,418 hip-hop comments. The outlier's mean compound was -0.019; excluding the user shifts the hip-hop mean from 0.150 to 0.152 ($n = 5{,}352$), a negligible change, indicating the low scores are not driven by a single account.

Comparison with All Genres. Across 136 features, hip-hop stands out: *higher engagement* (views 16.5M vs. 10.5M, likes 188,898 vs. 82,196, comments 11,693 vs. 3,907) but *lower conventional positivity* (lower Tone and Positive Emotion), *longer comments* (20.7 vs. 17.6 words), *higher emoji usage* (6.20 vs. 4.09), and slightly *higher Authenticity*. This combination suggests communication conventions that differ from those captured by lexicon-based sentiment.

5.5 Views and Sentiment

We correlate per-video mean compound with $\log_{10}(1+\texttt{view\_count})$. A very weak negative correlation is observed ($r = -0.05, p < 0.001$), indicating that more popular videos tend to have slightly lower average positivity; however, the effect is practically negligible.

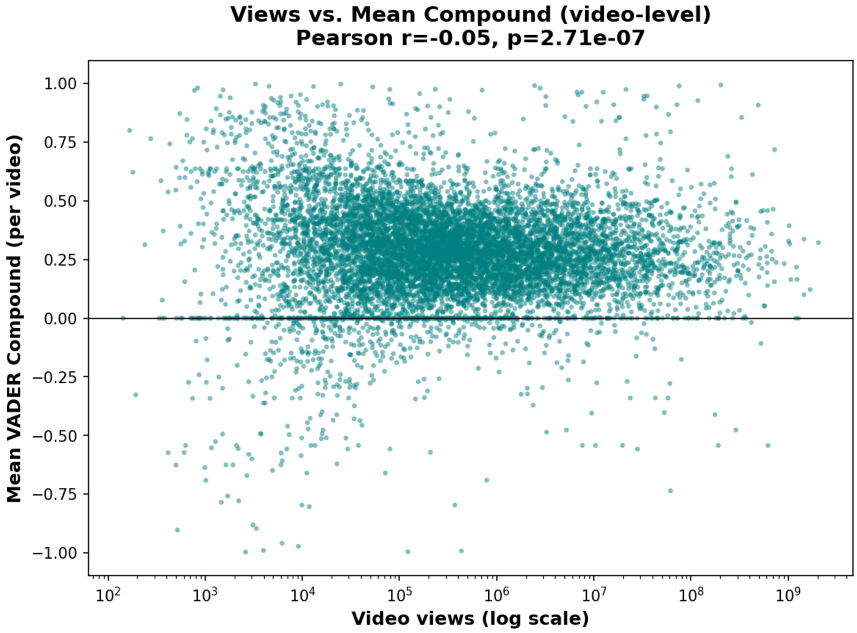

Figure 5.8: Video views (log scale) vs. mean VADER compound per video. Each point is a video; a weak negative association is visible ($r = -0.05$).

5.6 Emoji Usage and Sentiment

Across all comments, emoji presence is associated with higher sentiment: average compound with emojis is **0.417** versus **0.255** without (+64% relative increase). Thus, higher emoji usage in hip-hop does not explain its lower compound; if anything, emojis are generally linked to more positive tone.

Emoji Count Thresholds

We examine thresholds $x \in \{2, 3, 4, 5\}$ contrasting $\leq x$ vs. $> x$ emojis:

Table 5.2: Emoji threshold analysis: mean compound by group.

Threshold	Avg 1-to-x	Avg >x	Difference	n 1-to-x	n >x
5	0.7896	0.5166	**0.2730**	48	332
3	0.7796	0.5336	0.2460	27	353
2	0.7897	0.5446	0.2451	10	370
4	0.7560	0.5310	0.2250	34	346

Comments with **1–5** emojis achieve the highest sentiment (0.7896), whereas > 5 emojis corresponds to lower scores (0.5166). Hip-hop's mean of 6.20 places many comments beyond this optimal range.

Moderate emoji usage aligns with the highest median compound and smaller spread. No-emoji comments show lower medians and wider spread; > 5 emojis are still positively biased but less consistently so.

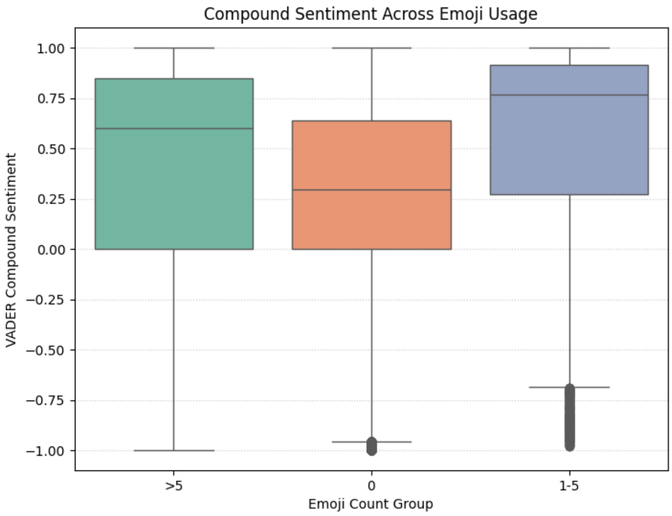

Figure 5.9: Distribution of compound scores across three emoji groups: no emojis, 1–5 emojis, and > 5 emojis.

5.7 LIWC Dimensions and Cultural Markers

We compare genres on LIWC features to contextualize stylistic and cultural differences beyond sentiment.

Positive vs. Negative Balance. The balance index (`emo_pos`−`emo_neg`) ranks spiritual and jazz genres highest, with hip-hop among the lowest, reinforcing lexicon-based lower positivity.

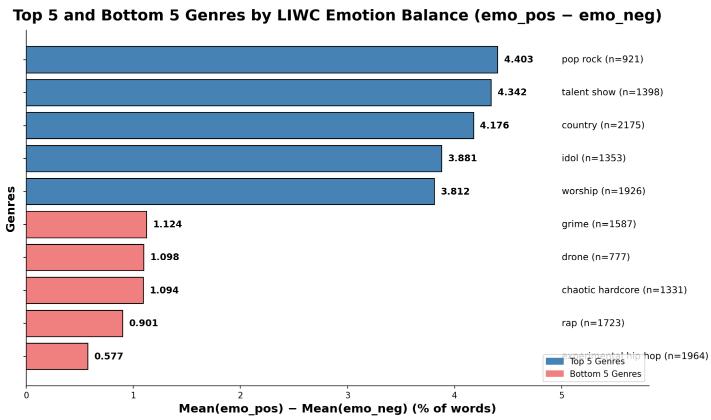

Figure 5.10: Top and bottom genres by LIWC `emo_pos` minus `emo_neg` balance.

Swear, Conflict, and Social. The `Swear` category captures the frequency of swear and offensive words, while `Conflict` refers to words denoting disagreement. Both are highest in aggressive metal genres and lowest in worship and gospel (Figures 5.11 and 5.12). In contrast, the `Social` category measures words related to social relations (e.g., friend, talk, share). This language is most frequent in *Worship*, *Country Pop*, and *K-pop*, and least common in extreme metal subgenres (Figure 5.13).

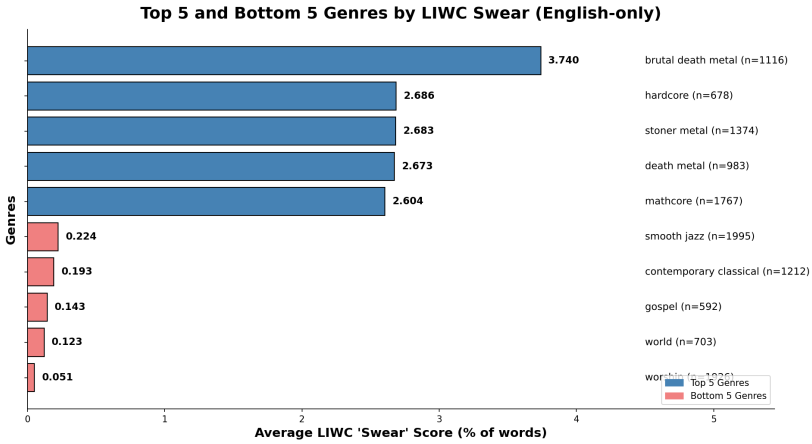

Figure 5.11: Top and bottom genres by LIWC `Swear`.

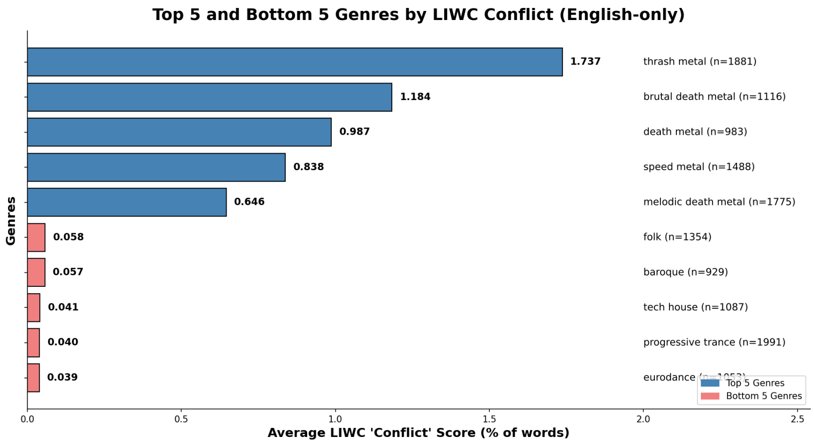

Figure 5.12: Top and bottom genres by LIWC `Conflict`.

5.8 LIWC–VADER Correlations

We correlate all LIWC dimensions with VADER sentiment (`positive`, `negative`, `neutral`, `compound`).

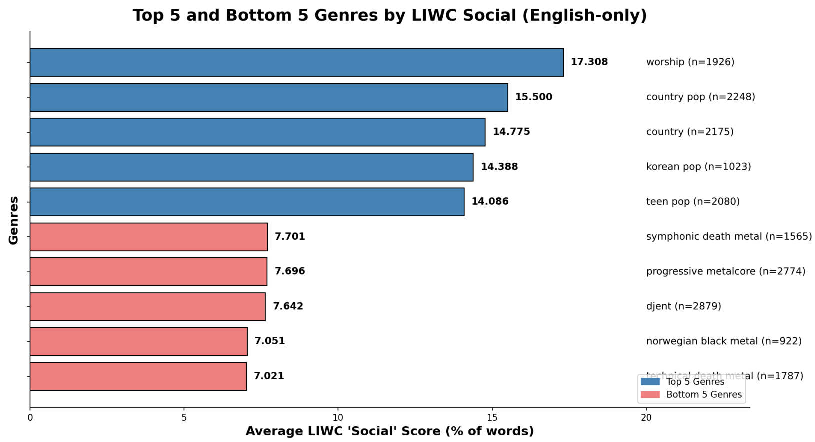

Figure 5.13: Top and bottom genres by LIWC `Social`.

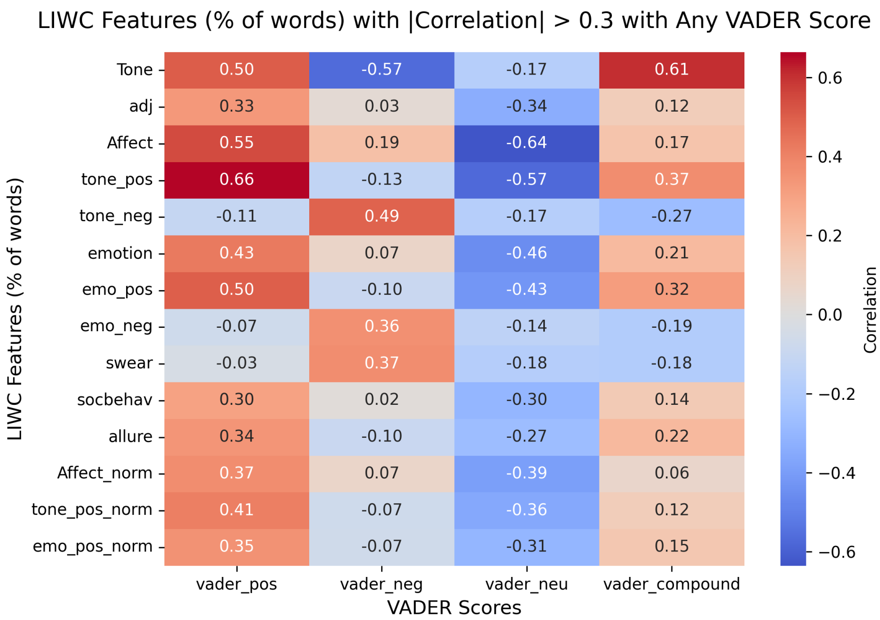

Figure 5.14: Correlation of LIWC features with VADER sentiment across all comments. Red = positive correlation; blue = negative.

Expected patterns. *Affect/Emotion/Positive Tone* correlate strongly with VADER positive ($r \approx 0.43$–0.66); *Negative Tone/Negative Emotion* correlate with VADER negative ($r \approx 0.36$–0.49). `Swear` is positively associated with VADER negative ($r \approx 0.37$). Neutral scores inversely track affective LIWC categories.

Unexpected patterns. The Emotion category (words directly expressing feelings such as happy, love, sad, or angry) shows a much stronger correlation with positive VADER scores than with negative ones. This is unexpected because one might assume emotional words would align equally with both valences, yet the results suggest that positive emotional language dominates in driving sentiment scores. A possible reason is that most comments are positively skewed, which could explain the stronger positive tendency.

Chapter 6

Conclusion and Outlook

6.1 Key Findings

This thesis set out to investigate sentiment and linguistic patterns in YouTube music video comments across genres, guided by six research questions. The main findings can be summarized as follows:

1. **Sentiment and engagement (RQ1).** Correlations between comment sentiment and likes were generally weak to modest across genres (mostly $|\rho| < 0.2$). Some genres (e.g., Nu Jazz, Disco, Latin Rock) showed that more positive or overall positive-toned comments received slightly more likes. In contrast, other genres rewarded neutral or even negative comments (e.g., Singer-Songwriter, Folk, Contemporary Classical), suggesting that critique or irony can also attract approval. This indicates that audience engagement on YouTube is not universally linked to sentiment but varies by cultural or stylistic context. These findings should be interpreted with caution, as many comments received no likes at all, which reduces the stability of correlation-based insights. Furthermore, engagement was measured only via likes, capturing just one limited aspect of audience interaction.

   Expected: at least a small positive correlation.

   Interesting: the direction of correlations varies by genre, with some communities valuing positive reinforcement and others rewarding critical or alternative engagement styles.

2. **Genre sentiment differences (RQ2).** Clear cross-genre differences emerged. Spiritual and smooth genres (e.g., worship, gospel, smooth jazz) showed the most positive average sentiment. Aggressive (e.g., death metal, noise, hardcore punk) scored lowest, yet still above zero, indicating that even the "most negative" genres lean slightly positive overall.

   Expected: spiritual music positive, aggressive music lower.

   Interesting: even the lowest genres remain net-positive.

3. **Polarization (RQ3).** Rap and hip-hop comments often contained both positive and negative elements, producing higher polarization scores. By contrast, worship comments were less polarized, dominated by unambiguously positive expressions. Between-comment polarization varied systematically across genres, with aggressive and rap-related styles showing the strongest coexistence of enthusiasm and critique. Within-comment polarization, while less pronounced, confirmed that individual comments also sometimes mix positive and negative sentiment, further reflecting ambivalence in community communication styles.

   Expected: some polarization in rap due to mixed reception.

Interesting: systematic differences in polarization levels between genre groups, with both within- and between-comment measures pointing to cultural contrasts.

4. **Psycholinguistic features (RQ4).** LIWC analysis revealed genre-specific linguistic markers. Aggressive genres had the highest frequencies of swearing and conflict terms, while worship and gospel showed high social and religious word use. These patterns align with distinct community norms and communication styles.

   Expected: more swearing in aggressive genres and more religious vocabulary in worship.

   Interesting: how consistently these markers separated genre clusters.

5. **Emoji usage (RQ5).** Comments with 1–5 emojis were on average much more positive than comments with none or with excessive emojis (>5). Moderate emoji use seems to optimize positive expression, while excessive use diminishes this effect.

   Expected: emojis linked to positivity.

   Interesting: the clear threshold effect with diminishing returns beyond five emojis.

6. **Popularity and sentiment (RQ6).** More popular videos tend to attract slightly less positive comments. Here too, engagement was only operationalized through view counts, which do not capture other relevant aspects such as replies or shares.

   Expected: more views → more positivity.

   Interesting: the opposite trend, possibly reflecting that viral or mainstream content invites more critical or mixed reactions.

Taken together, the results confirm that genre context matters greatly for interpreting online sentiment, and that lexicon-based methods should be applied with caution when analyzing diverse cultural communities.

6.2 Limitations

While the analyses provide novel insights, several limitations must be acknowledged:

- **Language filtering:** Only english comments were retained, excluding large non-English communities and possibly biasing results for globally popular genres.
- **Tool limitations:** VADER and LIWC cannot fully capture irony, sarcasm, or the positive use of swearing and slang. This limitation could be participially relevant for hip-hop, which may help explain the observed anomaly.
- **Temporal scope:** Restricting to the 100 most recent comments per video reflects current mood but not historical sentiment trends.
- **Genre coverage:** Genres with fewer than 500 comments were excluded, omitting smaller or emerging communities.
- **Engagement measures:** Likes and views are shaped by YouTube's algorithms and trends, and may not directly represent audience sentiment.

6.3 Outlook

Future research can build on this work in several ways:

- **Multilingual extension:** Including non-English comments would allow analysis of global communities and cross-cultural comparisons.
- **Advanced sentiment models:** Transformer-based or fine-tuned large language models could better capture slang, sarcasm, and cultural nuance than lexicon-based tools.
- **Temporal dynamics:** Tracking comment sentiment over time could reveal how community mood shifts around events like album releases or controversies.
- **Broader engagement measures:** Going beyond likes/views to include replies, shares, or watch time would provide a more complete picture of audience interaction.

Overall, this thesis provides the first large-scale, genre-aware analysis of sentiment and linguistic markers in YouTube music video comments. It demonstrates both expected and surprising patterns, and lays the groundwork for more culturally and methodologically robust future research.

Bibliography

[1] Rauf Arif, Yoonhyeung Cho, et al. 2024. Displays of Parasocial Interaction in K-Pop: A Content Analysis of YouTube Comments on BTS's Music Videos. *Nordicom Review*, 45, 1, 67–86. DOI: 10.2478/nor-2024-0005. https://www.researchgate.net/publication/382348670_Displays_of_Parasocial_Interaction_in_K-Pop_A_Content_Analysis_of_YouTube_Comments_on_BTS%27s_Music_Videos.

[2] Christine Bauer and Markus Schedl. 2019. Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems. In *Proceedings of the 17th IFIP International Conference on Human-Computer Interaction (INTERACT 2019)*. Springer, pp. 589–609. DOI: 10.1007/978-3-030-29387-1_35. https://doi.org/10.1007/978-3-030-29387-1_35.

[3] Seung Woo Chae, Noriko Hara, Harshit Rakesh Shiroiya, Janice Chen, and Ellen Ogihara. 2024. Being Vulnerable with Viewers: Exploring how medical YouTubers communicated about COVID-19 with the public. *PLOS ONE*, 19, 12, e0313857. DOI: 10.1371/journal.pone.0313857. https://doi.org/10.1371/journal.pone.0313857.

[4] Ioannis Chalkias, Katerina Tzafilkou, Panagiotis Karapiperis, and Christos Tjortjis. 2023. Exploring Sentiment Analysis Methods and Topic Clustering. *Electronics*, 12, 18, 3949. DOI: 10.3390/electronics12183949. https://www.mdpi.com/2079-9292/12/18/3949.

[5] Thomas Hartvigsen, Saadia Gabriel, Hamid Palangi, Maarten Sap, Dipankar Ray, and Ece Kamar. 2022. TOXIGEN: A Large-Scale Machine-Generated Dataset for Adversarial and Implicit Hate Speech Detection. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*. https://github.com/microsoft/ToxiGen.

[6] Hausberger, Strauß, and Schedl. 2025. ExIM: Exploring Intent of Music Listening for Retrieving User-Generated Playlists. Preprint. (2025).

[7] Ricardo Hernández-Pérez, Pablo Lara-Martínez, Bibiana Obregón-Quintana, Larry S. Liebovitch, and Lev Guzmán-Vargas. 2024. Correlations and Fractality in Sentence-Level Sentiment Analysis Based on VADER for Literary Texts. *Information*, 15, 11, (November 2024), 698. DOI: 10.3390/info15110698. https://www.mdpi.com/2078-2489/15/11/698.

[8] Clayton J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8, 1, (June 2014), 216–225. DOI: 10.1609/icwsm.v8i1.14550. https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

[9] Marta Moscati, Emilia Parada-Cabaleiro, Yashar Deldjoo, Eva Zangerle, and Markus Schedl. Music4All-Onion. Zenodo. https://zenodo.org/records/6609677.

[10] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. The Development and Psychometric Properties of LIWC2015. Technical report. University of Texas at Austin.

[11] Rahul Singha. 2024. Spam Detection on YouTube Comments Using Advanced Machine Learning Models: A Comparative Study. *arXiv preprint arXiv:2401.12345*. Accessed: 2024-07-31.

[12] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29, 1, 24–54.

[13] 2024. YouTube: number of monthly active users worldwide as of January 2024. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/. Accessed: 2024-07-31. (2024).

[14] Haowen Zhang. 2025. Sentiment Analysis on YouTube Data: A Comparison of TextBlob and VADER. *Frontiers in Humanities and Social Sciences*, 5, 1, 71–78. DOI: 10.26855/fhss.2025.01.010. https://www.hillpublisher.com/ArticleDetails/4437.

Appendix A

Representative Song Selection Algorithm

The algorithm operates in an iterative fashion to ensure fair distribution of representative videos across genres.

Input:

- For each genre: ranked list of candidate videos with relevance scores.
- Parameter k: number of representative videos per genre.

Output:

- For each genre: exactly k unique videos.

Procedure:

1. **Initial Assignment:** Assign the top k videos to each genre with at least k candidates.
2. **Conflict Detection:** Identify videos assigned to multiple genres.
3. **Conflict Resolution:**
   - Compute fallback scores for each genre (the score of the next unused candidate).
   - Retain the video for the genre with the lowest fallback score (i.e., the worst alternative).
   - Break ties using: (i) higher score on contested video, (ii) alphabetical order of genre name.
4. **Loser Refill:** Genres losing a conflict replace the video with their next best unused candidate.
5. **Iteration:** Repeat conflict detection and resolution until no conflicts remain or candidates are exhausted.
6. **Finalization:** Remove genres with fewer than k assigned videos and output the final assignment.