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

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-language comments across 233 genres. Each comment was preprocessed and analyzed using VADER for sentiment scoring and LIWC for psycholinguistic feature extraction.

The results revealed some clear differences between genres in terms of sentiment analysis. Music videos from genres characterized by spirituality or smoothness generally evoke stronger positive reactions, while hip-hop and aggressive genres rank lowest despite high engagement metrics. Furthermore, emoji usage was found to correlate strongly with positive sentiment, although excessive use diminished this effect.

Contents

Introduction

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 to examine how music genres differ in their online discourse.

2 Objectives and Approach

The overarching objective of this thesis is to analyze sentiment and linguistic patterns in YouTube comments across musical genres. The study pursues following guiding research questions:

1. Can correlations be identified between sentiment distributions and the number of likes received by comments?
2. How do sentiment distributions, as measured by VADER, vary across different musical genres?
3. What linguistic and psycholinguistic features, captured through LIWC, characterize genre-specific communication styles?
4. To what extent do conventional sentiment analysis tools succeed or fail in capturing authentic audience expressions within particular music communities?

To address these questions, a multi-stage research approach was implemented. First, a representative dataset of music videos was created by combining the JKU dataset of YouTube music links with the Music4All-Onion dataset for genre classification. A conflict-resolution algorithm ensured balanced representation across 233 genres, yielding 11,500 unique videos. From these videos, 100 recent comments per video were collected using the YouTube Data API, resulting in 356,973 English-language comments across 233 genres.

The collected comments were then preprocessed to remove noise and standardized for analysis. Sentiment scores were assigned using VADER, while psycholinguistic features were extracted using LIWC. Statistical comparisons and visualization techniques were subsequently applied to uncover genre-specific differences in sentiment expression and linguistic behavior. This pipeline ensures that the findings are grounded in both computational text analysis and established psycholinguistic methods.

3 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 [10].

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 [8]. Other work has extended sentiment analysis beyond simple positive–negative classifications to capture more fine-grained emotional categories, showing that comments can provide insights into deeper psychological and social dynamics.

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 [3].

VADER-based Sentiment Analysis

A substantial body of work uses VADER for analyzing short, informal, and emoji-rich social media text because it handles intensifiers, negation, punctuation and emoticons directly in its rule set [6]. Beyond the original Twitter-style validation, VADER has been widely applied to YouTube comments to quantify audience polarity and track shifts in aggregate sentiment. In these settings, the compound score serves as a practical and interpretable score for overall valence, enabling large-scale comparisons across topics and channels.

LIWC-based Linguistic Profiling

Complementary 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 [7, 9]. In the context of YouTube, LIWC has been employed to uncover community norms—such as social bonding, swearing/hostility markers, or religiosity—that are not captured by sentiment alone, thereby offering a richer view of how users express themselves.

Why Combine VADER and LIWC?

VADER and LIWC answer complementary questions. VADER provides a robust, social-media-tuned estimate of comment valence (positive/negative/neutral) with a single, comparable compound score suitable for large-scale analyses [6]. LIWC, in turn, 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 [7]. Combining both methods makes it possible to (i) map sentiment hierarchies across genres, (ii) explain those differences through concrete linguistic markers, and (iii) relate community endorsement (likes) to either valence or specific stylistic/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 demonstrated 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 based on explicit listening intents and seek playlists aligned with their goals, as shown by the ExIM study [4]. 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 toward the artist, highlighting the social dimension of 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 users' preferences align with either 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

Together, 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.

4 Dataset

4.1 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** [**music4all**], available at https://zenodo.org/records/6609677, connects these songs to their respective 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.

4.2 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 .1.1.

4.3 Comment Collection & Metadata

Using the YouTube Data API v3, exactly 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.

4.4 Preprocessing Pipeline

Language Filtering

Given that both VADER sentiment analysis and LIWC require English text for accurate processing, only English-language 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-language comments.

Text Standardization

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

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.

4.5 Feature Extraction

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 [6]. 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 [5].

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.

4.6 Final Dataset Characteristics

The final dataset comprises 356,973 English-language 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 discourse.

Genre representation varies significantly, with popular genres like pop and rock contributing thousands of comments while niche genres approach the 500-comment minimum threshold. This distribution reflects natural music consumption patterns while maintaining analytical feasibility across all included genres.

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.

5 Analysis

5.1 Engagement–Sentiment Correlation

The first step of the analysis examines the relationship between expressed sentiment and audience engagement, measured through the number of likes a comment receives. For each genre, Pearson correlation coefficients were calculated between comment like counts and the four VADER sentiment dimensions (positive, negative, neutral, and compound).

Positive Sentiment (VADER_pos)

In genres such as *Thrash Metal* and *New Jack Swing*, even slight positivity increases the likelihood of likes ($r \approx 0.04$–0.05). In contrast, genres like *Lounge* and *Pop Rock* show negative correlations ($r \approx -0.09$ to -0.10), where positivity is not rewarded.

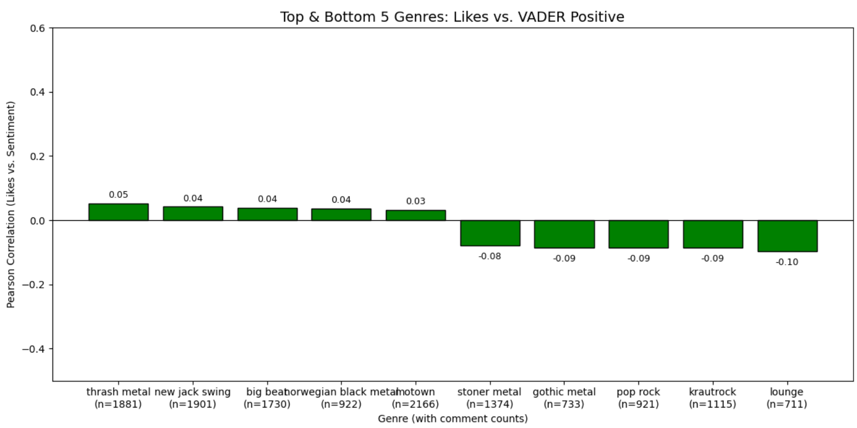

Figure 1: Top and bottom five genres by correlation between likes and VADER positive sentiment.

Negative Sentiment (VADER_neg)

Genres like *Piano Rock* and *Vocal Jazz* show weak positive correlations ($r = 0.05$–0.08), meaning mildly negative comments may still be endorsed. Conversely, genres such as *Latin Rock* or *Melodic Black Metal* show small negative correlations ($r \approx -0.09$), where negativity is less socially validated.

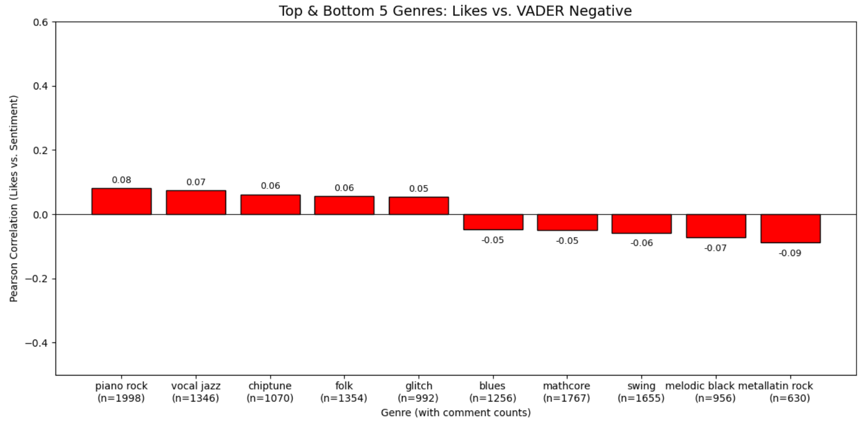

Figure 2: Top and bottom five genres by correlation between likes and VADER negative sentiment.

Neutral Sentiment (VADER_neu)

Here, *Lounge*, *Stoner Metal*, and *Krautrock* show the highest positive correlations ($r = 0.08$–0.09), suggesting neutral-toned comments attract likes. In contrast, *Norwegian Black Metal* and *Dark Ambient* show negative correlations ($r \approx -0.06$).

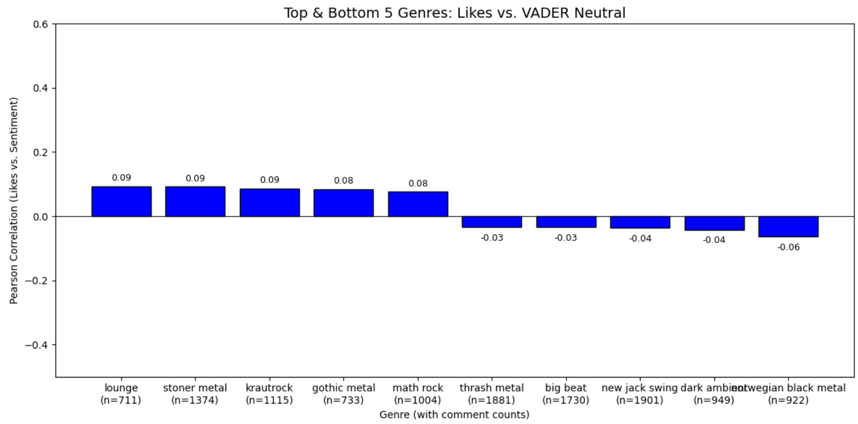

Figure 3: Top and bottom five genres by correlation between likes and VADER neutral sentiment.

Compound Sentiment (VADER_compound)

The compound score provides the overall view. Genres such as *Korean Pop*, *Latin Rock*, and *Baroque Pop* show positive correlations ($r \approx 0.09$–0.10), while *Folk*, *Chiptune*, and *Noise* display negative ones ($r \approx -0.07$ to -0.10).

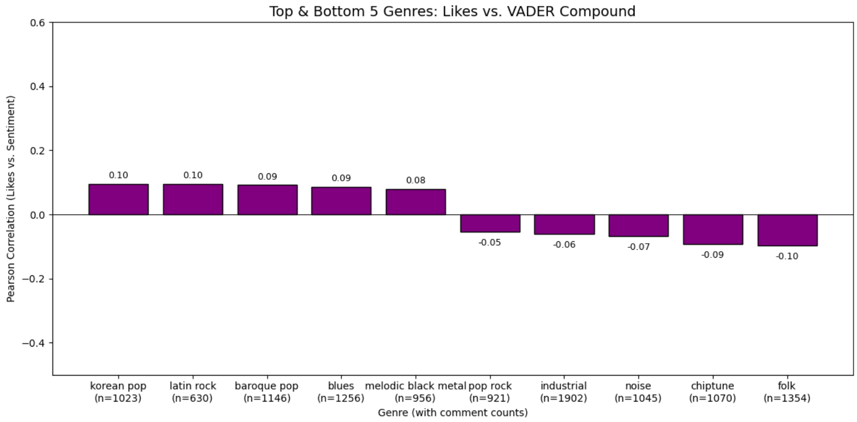

Figure 4: Top and bottom five genres by correlation between likes and VADER compound sentiment.

General Summary. Across all four sentiment dimensions, correlations between likes and sentiment remain weak in absolute size, but they consistently reveal genre-specific endorsement patterns. Positive sentiment is modestly rewarded in genres such as Korean Pop, Thrash Metal, and Baroque Pop, while neutral or even negative comments are relatively more valued in genres like Folk, Chiptune, and Noise. The neutral dimension highlights that in some communities (e.g., Lounge, Krautrock) balanced or non-emotional comments attract the most engagement. Taken together, these findings suggest that audience validation on YouTube is not uniformly driven by positivity but shaped by shared writing convention within each genre, where different communities reward distinct expressive styles.

5.2 Overall Analytical Strategy

The analytical strategy employs multiple complementary approaches to understand sentiment patterns across music genres. Primary analysis focuses on average compound sentiment scores per genre to identify consistent emotional response patterns. Outlier analysis examines potential data anomalies, including individual user behavior that might skew genre-level results. Cross-comparison analysis integrates VADER sentiment scores with LIWC linguistic features to understand the psychological and stylistic factors underlying observed sentiment patterns.

This multi-faceted approach enables both broad genre comparisons and detailed investigation of specific anomalies, such as the unexpected sentiment patterns observed in hip-hop genres, providing comprehensive insights into how different musical communities engage in online discourse.

6 Results

6.1 Sentiment Distributions Across Genres

The analysis of average compound scores across 233 music genres reveals substantial variation in audience emotional responses, with scores ranging from 0.475 to 0.062 - a difference of over 0.4 points on the average comment

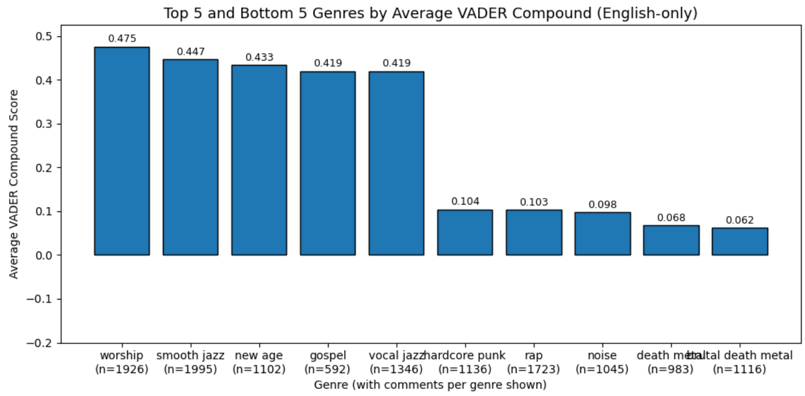

Figure 5: Top 5 Most Positive and Bottom 5 Most Negative Genres by Average Compound Score

Spiritual and smooth genres dominate the positive sentiment rankings, with worship music achieving the highest average compound score (0.475), followed by smooth jazz (0.447), new age (0.433), gospel (0.419), and vocal jazz (0.419). This pattern suggests that genres emphasizing spiritual expression, relaxation, and sophisticated musical forms consistently generate appreciative and uplifting audience responses.

Conversely, heavier and more aggressive musical styles occupy the lowest sentiment positions. Death metal (0.062) and brutal death metal (0.068) rank at the bottom, followed by noise (0.098), rap (0.103), and hardcore punk (0.104). Notably, despite ranking as the most negative genres, all maintain positive compound scores, indicating that even critically discussed genres generate more positive than negative sentiment overall.

6.2 Hip-Hop Anomaly Investigation

Hip-hop genres present an unexpected pattern that warrants detailed examination. Despite representing one of the most popular musical categories globally, all hip-hop subgenres rank in the bottom 18 genres for compound sentiment scores.

Table 1: Sentiment Analysis Results for 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 a average compound score of 0.173 (rank 215/233), while experimental hip-hop ranks lowest at 0.124 (rank 227/233), notably achieving the 7th highest negative sentiment score across all genres.

Data Anomaly Analysis

Investigation of potential data skewing revealed one user contributing 66 out of 5,418 total hip-hop comments. Analysis of this outlier's impact showed:

- Average compound score (all comments): 0.273
- Average compound score (Hip-Hop genres only): 0.150
- Average compound score (outlier user only): -0.019 (n=66)
- Average compound score (Hip-Hop excluding outlier): 0.152 (n=5,352)

The minimal difference between scores with and without the outlier (0.150 vs. 0.152) demonstrates that individual user behavior cannot explain hip-hop's low sentiment rankings.

Hip-Hop vs. All Genres Comparison

Comparative analysis across 136 features reveals hip-hop's distinctive characteristics:

Higher Engagement Metrics:

- View Count: 16.5M vs. 10.5M across all genres (+57%)
- Like Count: 188,898 vs. 82,196 across all genres (+130%)
- Comment Count: 11,693 vs. 3,907 across all genres (+199%)

Distinct Communication Patterns:

- Lower overall tone (66.25 vs. 74.45, -11%)
- Reduced positive emotional tone (5.37 vs. 7.61, -29%)
- Longer comments (20.7 vs. 17.6 words, +17%)
- Higher emoji usage (6.20 vs. 4.09, +52%)
- Higher authentic language use (60.5 vs. 59.1, +2%)

These findings reveal that hip-hop generates significantly higher engagement while exhibiting lower traditional positivity indicators, creating a paradox between popularity and measured sentiment.

6.3 Views and Sentiment Correlation

To examine whether video popularity is associated with comment sentiment, a Pearson correlation was calculated between the logarithm of video view counts and the mean VADER compound score per video. The analysis revealed a very weak negative correlation ($r = -0.05, p < 0.001$), suggesting that more popular videos tend to attract slightly less positive comments on average. Despite the large sample size, the small negative correlation is negligible, indicating that video popularity does not meaningfully shape comment sentiment.

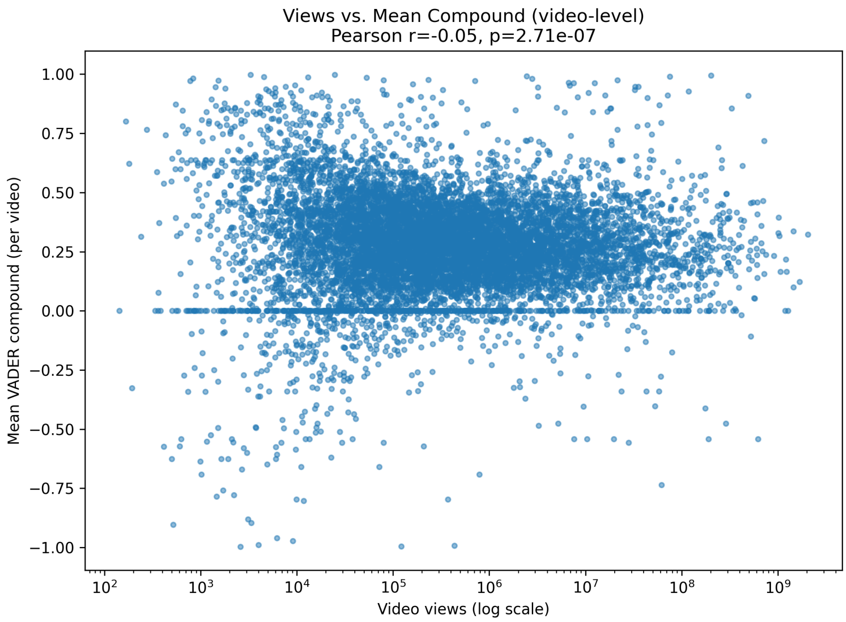

Figure 6: Scatterplot of video views (log scale) and mean VADER compound scores per video. Each point represents a video. A weak but statistically significant negative correlation ($r = -0.05, p < 0.001$) was observed.

6.4 Emoji Usage and Sentiment Correlation

Analysis of emoji usage across the entire dataset reveals a strong positive correlation with sentiment scores:

- Average compound score with emojis: **0.417**
- Average compound score without emojis: **0.255**
- Difference: **+64% higher sentiment for emoji-containing comments**

This finding contradicts the hypothesis that hip-hop's higher emoji usage might explain their lower sentiment scores. Instead, it makes hip-hop's poor sentiment performance more anomalous, as their 52% higher emoji usage should theoretically improve their compound scores.

Emoji Count Analysis

Threshold analysis identified the optimal emoji range for maximizing positive sentiment:

Table 2: Top Emoji Count Thresholds by Sentiment Difference

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 containing 1-5 emojis achieve the highest sentiment scores (0.7896), while excessive emoji usage (>5) correlates with lower scores (0.5166). Hip-hop's average emoji count of 6.20 places many comments beyond this optimal range, potentially contributing to their lower compound scores despite emoji usage being generally positive.

6.5 Linguistic and Psycholinguistic Dimensions (LIWC Analysis)

Beyond sentiment distributions, additional analyses were conducted to examine linguistic and psychological dimensions of user comments. Using LIWC features, genres were compared not only in terms of sentiment but also in their broader stylistic and cognitive patterns.

Positive vs. Negative Balance By comparing LIWC `emo_pos` and `emo_neg` scores across genres, a positive–negative balance index was computed. Genres such as worship and jazz appeared at the top of this scale, while hip-hop again ranked among the lowest. This reinforced the earlier finding that hip-hop communities express sentiment in ways that conventional lexicon-based tools classify as less positive.

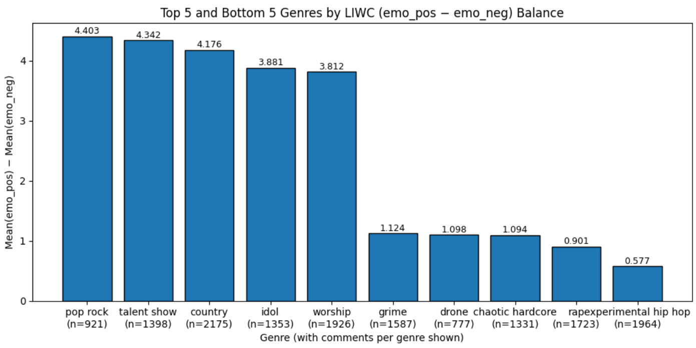

Figure 7: Top and bottom scoring genres based on the LIWC `emo_pos` minus `emo_neg` balance.

Genre-specific LIWC Patterns. Individual LIWC features provided further insight into genre-level differences. For example, religious and smooth genres exhibited high scores for `Religion` and `Tone`, while aggressive styles such as death metal and hardcore punk ranked highest in categories like `Swear` and `Anger`. Features related to anxiety and health also showed distinct distributions across genres. These genre-specific linguistic markers suggest that community norms and cultural contexts strongly shape language use.

6.6 Cultural Linguistic Markers: Swear, Conflict, and Social

To complement the sentiment-based results, additional LIWC features were analyzed to capture linguistic markers that reflect deeper cultural and community-specific patterns. Three particularly informative features are presented here: `Swear`, `Conflict`, and `Social`.

Swear. A striking contrast was observed between aggressive and smooth genres. As shown in Figure 8, genres such as brutal death metal, hardcore, and stoner metal exhibited the highest frequencies of swear words, while spiritual or smooth genres like worship, gospel, and smooth jazz showed the lowest. This split reinforces cultural and linguistic contrasts between music communities, where aggressive genres embrace explicit language while spiritual genres avoid it.

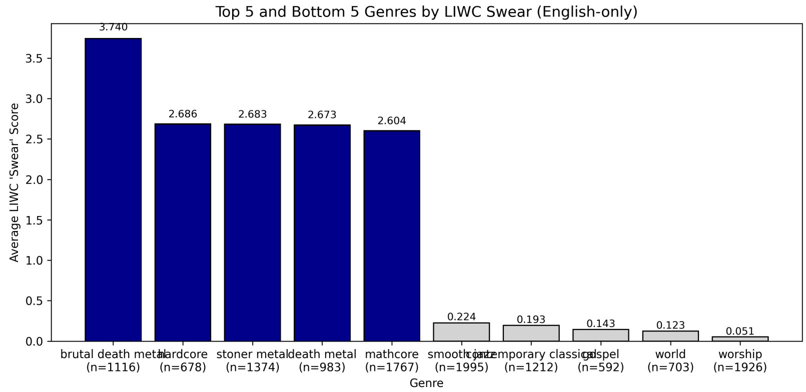

Figure 8: Top 5 and bottom 5 genres by LIWC `Swear` score. Aggressive metal genres dominate swearing, while worship and gospel show minimal usage.

Conflict. A similar pattern emerged in the `Conflict` category (Figure 9). Thrash metal and brutal death metal scored highest, reflecting a language style centered on hostility and confrontation. In contrast, genres such as baroque, trance, and dance ranked lowest, indicating their communities employ less conflict-related vocabulary. Together with the `Swear` feature, this demonstrates that aggressive genres cluster around hostility markers, while others remain linguistically neutral.

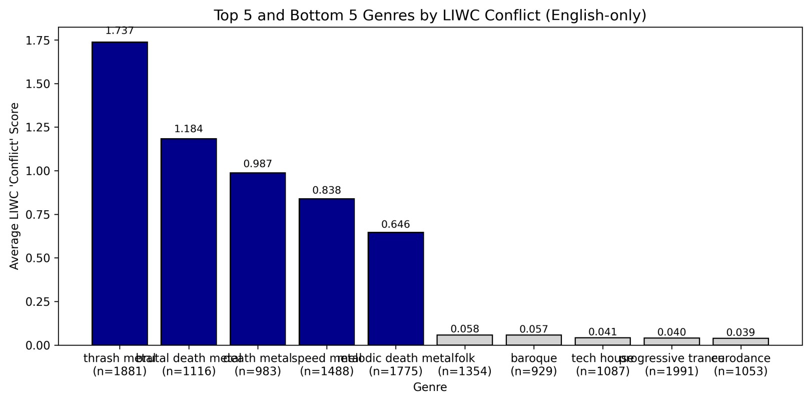

Figure 9: Top 5 and bottom 5 genres by LIWC `Conflict` score. Thrash and death metal dominate conflict-related vocabulary, while baroque and trance show minimal presence.

Social. The `Social` feature provided an important counterbalance (Figure 10). Worship, country pop, and K-pop ranked highest, consistent with their roles in fostering community, spirituality, and group identity. In contrast, extreme metal subgenres such as technical death metal and black metal ranked lowest, pointing to more individualistic communication patterns. This result highlights not only negativity in certain communities but also the richness of socially oriented language in others, suggesting distinct cultural functions of music genres.

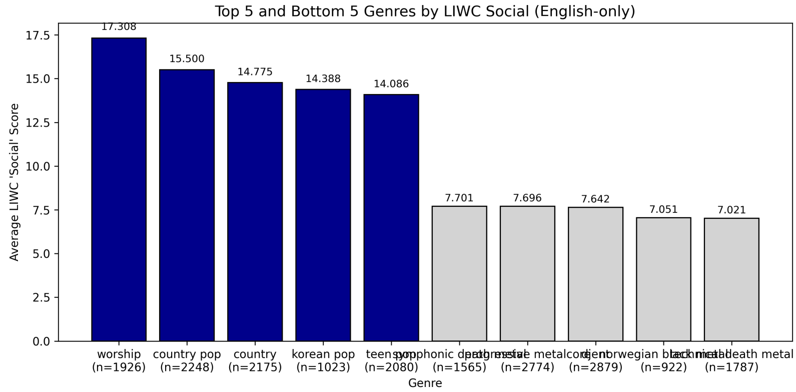

Figure 10: Top 5 and bottom 5 genres by LIWC `Social` score. Worship and country pop emphasize social language, while extreme metal genres show lower social word usage.

Variance and Clustering. A variance analysis identified the features most responsible for differences across genres, including `Tone`, `Clout`, `Authenticity`, `Emoji`, and cognitive process indicators. Normalized heatmaps further revealed genre clusters: spiritually oriented genres grouped together, while hip-hop and aggressive genres formed distinct clusters. These findings point toward shared communicative identities within genre communities.

6.7 Correlation between LIWC Features and VADER Sentiment Scores

To better understand the relationship between linguistic features and sentiment, correlations were computed between all LIWC dimensions and the four VADER sentiment scores (`positive`, `negative`, `neutral`, and `compound`). The results are summarized in Figure 11, which presents a heatmap of LIWC–VADER correlations.

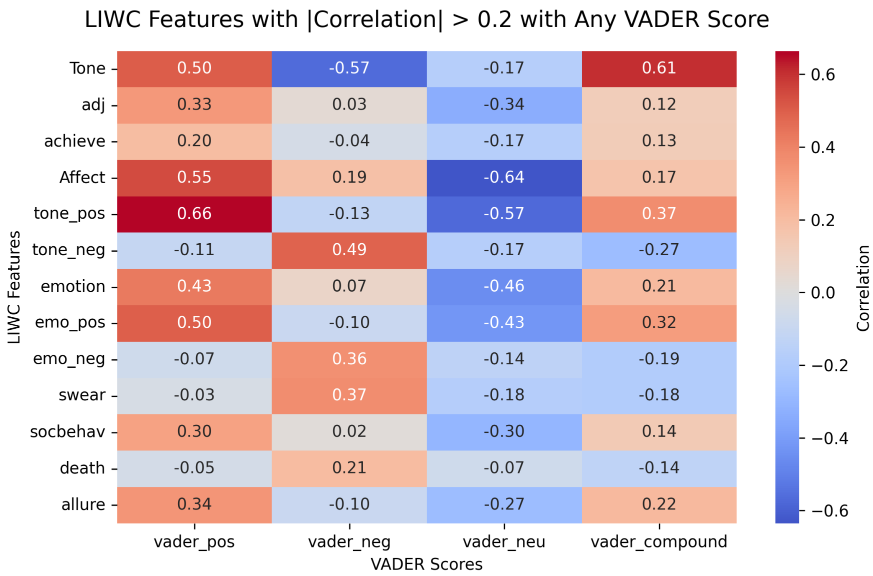

Figure 11: Correlation of LIWC features with VADER sentiment scores across all comments. Red indicates positive correlations, blue indicates negative correlations.

6.8 Expected and Unexpected Correlation Patterns

The heatmap in Figure 11 highlights both expected and unexpected relationships between LIWC features and VADER sentiment scores.

Expected results:

- LIWC *Affect*, *Emotion*, and *Positive Tone* correlate strongly with VADER positive sentiment ($r = 0.43$ to 0.66).
- LIWC *Negative Tone* and *Negative Emotion* correlate with VADER negative sentiment ($r = 0.36$ to 0.49).
- The *Swear* category is positively associated with VADER negative sentiment ($r = 0.37$), as expected for hostile or aggressive language.
- Neutral VADER scores show inverse correlations with affective LIWC categories, reflecting the distinction between neutrality (VADER) and emotional expression (LIWC).

Unexpected results:

- LIWC *Allure* shows a positive correlation with VADER compound ($r = 0.22$), suggesting that attractiveness-related language is systematically linked with positivity.
- The *Social Behavior* category correlates only weakly with sentiment ($r \approx 0.30$ with VADER positive), indicating that social language is not strongly tied to positive sentiment.
- Some linguistic dimensions (e.g., *Adjectives*, *Achieve*) exhibit modest but consistent correlations with VADER sentiment, their non-emotional focus within LIWC.

6.9 Key Findings Summary

The analysis reveals several critical patterns in music genre sentiment:

1. **Genre Hierarchy**: Spiritual and smooth genres consistently generate higher positive sentiment than aggressive or complex musical genres.
2. **Hip-Hop Paradox**: Even after checks for sampling bias, outliers, and emoji effects, hip-hop genres remain among the lowest in sentiment scores, suggesting systematic differences in discourse conventions and how users communicate within this genre.
3. **Emoji Optimization**: Sentiment scores are highest when comments contain between 1 and 5 emojis, while using more than 5 reduces the positive effect. This pattern may influence genres with higher emoji usage.
4. **Cultural Communication Differences**: Hip-hop communities exhibit distinct linguistic patterns—longer comments, stronger authenticity markers, and fewer conventional positive words in their language.

These findings indicate that traditional sentiment analysis may fall short in capturing the complexity of genre-specific communication.

Chapter 1

Conclusion and Outlook

Space for your summary, central conclusions, and an outlook on potential future work.

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] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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).

.1 Appendix

.1.1 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.