HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 3 (2023)

Relating Musical Genre and Lyrical Repetition

Sovannjet Lim[1] and Soo Park[1#]

[1]Advanced Technologies Academy
[#]Advisor

ABSTRACT

In the literature on music and repetition, it has been found that both repetition and genre influence liking for music. Naturally, this leads to the question of whether repetition and genre influence each other. The purpose of this study was to test the hypothesis that there is an association between songs' musical genre and their amount of lyrical repetition. Using the percent compressibility of lyrics as a measure of repetitiveness, the average repetitiveness of ten songs for each of ten genres was compared across genres. Based on these measures of average repetitiveness, it was possible to separate the genres into categories that were statistically significantly different: R&B, pop, EDM, and hip-hop were most repetitive; Latin and country were moderately repetitive; and metal, rock, folk, and blues were least repetitive. This confirmed the hypothesis that there is an association between musical genre and lyrical repetition and establishes repetitiveness as a new criterion by which musical genres can be classified.

Introduction

In music, *repetition* refers to the recurrence of phrases within a song or piece. Repetition is omnipresent in music. According to ethnomusicologist Bruno Nettl (1983), repetition is a "musical universal" common among cultures across the world. In fact, repetition is a feature that is essential to the way people perceive and listen to music, due to a variety of psychological phenomena. For example, Robert Zajonc (1968) showed that repeated exposure to stimuli such as words and symbols improves attitudes toward them—known as the *mere exposure effect*. In addition, in many experiments, repeating words or numbers made it more difficult to perform cognitive tasks involving them, indicating that repetition of words weakens the interpretation of their meaning, which Jakobovits (1962) calls *semantic satiation*. This goes hand-in-hand with the *speech-to-song illusion*, in which repeating a spoken phrase over and over, with the pitch and order of the syllables preserved, causes it to be increasingly perceived as a musical phrase (Deutsch et al., 2008, 2011). In short, semantic satiation and the speech-to-song illusion show that repetition of a word or phrase shifts attention away from its meaning and toward aspects of its sound. These psychological effects, along with the mere exposure effect, mean that repetition is powerful and essential to how music works.

Literature Review

Repetition influences the perception of music. For instance, when two lines of music are playing simultaneously and one line contains the immediate, exact repetition of a short segment, attention tends to shift to the other line, which varies more; in general, novelty commands more attention and emotional response than repetition (Taher et al., 2016). As another example, Elizabeth Hellmuth Margulis (2012), a Professor of Music and Director of the Musical Cognition Lab at Princeton University who has extensively researched the psychology of music, found that repetition of musical excerpts improves people's ability to detect repetition for longer repeated units and weakens it for shorter units, indicating that repetition might cause attention to move to "progressively higher temporal levels." Margulis and Simchy-Gross (2016) also found that randomly generated musical excerpts that were looped were rated higher in musicality than excerpts that were unlooped. Both of Margulis's findings agree with the aforementioned speech-to-song illusion.

Repetition also influences liking for music. Repetition increases familiarity (Getz, 1966; Hargreaves, 1984), and in a study by Getz (1966), repetition-based familiarity influenced preferences for musical pieces more than any single musical factor. Similarly, Williams (2018) found moderate, uniformly significant correlations between familiarity and preference. According to the "inverted-U" model of liking, as repetition increases, liking initially increases and then eventually decreases (Hargreaves, 1984; Margulis & Simchy-Gross, 2016). David Hargreaves (1984), an Emeritus Professor of Education and a Music Psychology researcher at the University of Roehampton, partially confirmed this in two experiments that found that liking increased then decreased for "easy listening music" with continuous repetition and for popular and classical music when continuous repetition and repetition at intervals were combined, although, for avant-garde jazz, liking increased with continuous repetition and stayed the same when both types of repetition were combined.

In these experiments by Hargreaves (1984), the fact that people responded differently to different genres indicates that genre also influences liking for music; in fact, the different genres, which varied in complexity, stayed separate in terms of levels of both liking and familiarity. Indeed, genre (as well as musical training) significantly influences preference (Williams, 2018). Moreover, Getz (1966) found that "fast tempo, variety of volume, melodic repeats, flowing rhythm, jumpy melody, variety of melodies, and mode" were common reasons for liking and that "too loud volume, jumpy melody, dissonance, and minor mode" were common reasons for disliking (both lists are ordered by how often subjects mentioned the reasons, from most often to least often). Since different genres can be characterized by musical features similar to these reasons, this again suggests that genre affects liking. In fact, an audio-based analysis of music of various genres showed that different genres of music elicit different emotions (Eerola, 2011).

Beyond musical features, genres have characteristic lyrical features as well. For example, Nadav Appel (2014) identifies many examples in which pop-rock music uses nonsense words, gibberish, baby talk, and similar childlike speech patterns expressively, such as to shift attention to the physical and bodily aspects of singing. Appel (2014) even notes that various forms of repetition can make nonsense words somewhat make sense. Genres can also have characteristic lyrical features in that artist of certain genres, such as folk and rock, write lyrics after instrumentals, matching words to the mood or feel, while artists of other genres compose instrumentals after lyrics, translating stresses and meter—the "shape of the lyric"—to rhythm (Negus & Astor, 2015). Lyrical differences manifest themselves in the content of songs as well. In a linguistic analysis of 200 popular contemporary songs sampled from *Billboard*, Hart and Day (2020) found that rock music contains sexual content moderately correlated with lyrics expressing anxiety, while rap and R&B music have higher frequencies of sexual content, strongly correlated with lyrics expressing anger. Evidently, lyrical features can characterize genres, raising the question of whether lyrical *repetition* can also do so.

So far, it has been established that repetition influences liking (and perception) and that genre influences liking. Naturally, this leads to the question of whether repetition and genre influence each other. However, while there are many analyses of repetition's *effects* on listening experiences (perception) and responses (liking), there is little research on the repetition itself—its presence in actual songs and how it varies by genre. Abeßer et al. (2012) were relatively successful in classifying genres based on musical features, like rhythm and tonality, of their repetitive basslines. Meanwhile, Thompson (2021) classified genres using linguistic features instead of audio features; in doing so, considering lyrical features like rhymes, readability, and profanity improved the accuracy of computer predictions of genre. However, neither of these studies used *lyrical repetition* to classify genres. Therefore, this study will fill this gap by investigating how lyrical repetition varies among different genres of music—that is, how genre influences repetition—directly connecting repetition and genre in the process. It will seek to answer the following research question: *How do different genres of music differ in how frequently they use lyrical repetition?*

To test the hypothesis that there is an association between songs' musical genre and their amount of repetition—for example, that pop, country, and rock songs are more repetitive than folk, hip-hop, and blues songs—it is necessary to operationally define "repetition." This starts with identifying the type of repetition to be investigated. This study focused specifically on *lyrical repetition*, in which certain word phrases occur multiple times in a song.

Computerized audio analysis techniques have been developed for identifying chords, which can be used to investigate *instrumental repetition*, in which phrases in the background music recur (Fujishima, 1999). Computerized methods for transcribing musical lines have also been created, which can be used to investigate *melodic repetition*, in which phrases of notes in the melody recur (Klapuri, 1998). However, it is difficult to obtain direct access to these computer programs in order to be able to input audio files for analysis. It is even more challenging to replicate such programs or to manually identify chords and transcribe melodies by ear, which would also be less reliable due to human error. Therefore, to keep the method viable, this study only examined lyrical repetition.

Methods

The purpose of this study was to test the hypothesis that there is an association between songs' musical genre and their amount of repetition. This hypothesis was tested by following a quantitative procedure to analyze the amount of repetition in songs. On top of measuring repetitiveness quantitatively and objectively, this study adds to the existing knowledge about the characteristics of various musical genres, establishing the motivation for this research. This new knowledge may be of interest to musical artists and listeners of music, both of whom engage significantly with song lyrics and their repetition.

Materials

This study looked at pop, rock, hip-hop, Latin, electronic dance music (EDM), R&B, country, folk, metal, and blues. These are the musical genres listed on Chosic (https://www.chosic.com/list-of-music-genres/). Chosic also lists classical, jazz, easy listening, new age, and world/traditional as genres, but they were not investigated since most songs in these genres lacked lyrics. Chosic is a public website that offers a variety of free online tools for discovering new music. The public nature of the site makes the site appropriate for studying popular music (of all genres), which is largely a public phenomenon. More importantly, the site is more accessible than other sources that list songs by genre, such as *Billboard* (https://www.billboard.com/charts/), whose genre-based charts require a paid subscription to access.

For each genre listed on its site, Chosic includes a 100-song playlist created on Spotify. From each of these playlists, a simple random sample of 10 songs was drawn, creating a dataset of 100 songs (10 songs per genre × 10 genres). A sample size of 10 songs for each genre was chosen in order to provide a sizeable representation for each genre while keeping the total number of songs (100) manageable in data collection. In addition, a sample size of 10 maintains independence in the random sampling for each genre ($n = 10 \leq 10\%$ of $p = 100$ songs per playlist).

Lyrics for each song were obtained from Genius (https://genius.com/), a public website that is a popular source of reference for song lyrics. The site's published lyrics for many songs are reviewed, enhancing its reputability. Moreover, Genius has song lyrics available for a large number and a wide variety of songs, and consistently using one source throughout the study in order to obtain lyrics served as a measure of standardization.

Instruments

In order to quantitatively analyze and compare the repetitiveness of songs, a method of operationally quantifying the amount of repetition in the lyrics of a song needed to be developed. Multiple options were considered for this method.

One such option was the *Hirsch-Popescu-point*, which "is defined as the point in which the ranking of a word in [a] distribution matches its frequency"—for example, the point where the tenth most frequently occurring word occurs ten times (Nunes et al., 2015). The Hirsch-Popescu-point reliably indicates the repetitiveness of a word in a text and is commonly used in lexical analysis (Nunes et al., 2015). However, while this measure of repetition analyzes the repetition of individual words, this study was interested in the repetition of entire phrases. Therefore, the Hirsch-

Popescu-point would not have been completely appropriate for answering the research question, and it was not chosen for this study.

Another option was using data compression. Data compression attempts to reduce the size of a file by representing data more compactly (McGeoch, 1993). One way of doing so is replacing repeated sequences of units—such as bits, characters, or letters—with keys that are shorter than those sequences. Then, a dictionary can be created, associating each key with *one* instance of its corresponding repeated sequence. This shortens the data, as the original data contained *multiple* instances of the repeated sequences. Such a process is undertaken by the *Lempel-Ziv algorithm* (Ziv & Lempel, 1977; Zeeh, 2013). Since greater compression (i.e., the compressed data is much shorter than the original data) indicates a greater presence of repeated phrases in the original data, the Lempel-Ziv algorithm can be used to quantify the repetitiveness of the lyrics of a song. The greater the compression (i.e., the more the lyrics could be compressed), the more repetitive the song is.

In a study of whether lyrics in pop songs have become more repetitive over time, Morris (2017) used the Lempel-Ziv algorithm to measure the amount of repetition in pop songs. The study described in this paper used the algorithm in a similar fashion; while Morris's (2017) study measured repetition in pop songs and compared this measure over time, this paper's study measured repetition in songs of a variety of genres and compared this measure across genres. Since both methodologies required the measurement of repetition for similar comparative purposes, the Lempel-Ziv algorithm was reasoned to be appropriate for measuring repetition in this study. The exact implementation of the Lempel-Ziv algorithm used in this study was sourced from Rosetta Code (*LZW*, 2023) and is shown in Appendix A: Computer Code.

Procedure

For each song, the lyrics were obtained from Genius and copied and pasted into a .txt file. The text content of this file was programmatically processed by removing elements that were reasoned to not be essential to the actual words of the lyrics. Such elements included bracketed notes (e.g., "[Verse 1: Artist]"), line breaks, and "subjective" punctuation that might be inconsistent across different sources of lyrics (i.e., all punctuation except apostrophes and hyphens, which are part of English conventions, and parentheses, which consistently indicate quieter, secondary lyrical lines). Then, the processed lyrics were compressed using the Lempel-Ziv algorithm. The original length of the processed lyrics was compared with the length of the compressed lyrics to calculate the percent compression ((original length − compressed length) ÷ original length), and therefore the measure of repetitiveness, for the song. This process was repeated for each song in the sample. The computer code for the implementation of this process is shown in Figure A2.

Next, each genre was examined individually, averaging the measures of repetitiveness across the songs in the genre. This was repeated for each genre, giving a measure of average repetitiveness for each genre. Finally, these average measures were compared among genres to identify how the amount of lyrical repetition varies among them. This comparison was performed by using statistical analysis to find the statistically significant differences among the genres. Ultimately, this method allowed the research question to be answered through quantitative comparison.

Results

Because this study used 100 songs, the complete data for all the songs is too long to reasonably present in this section. Instead, the complete data are listed in Appendix B: Data for All Songs. Meanwhile, Table 1 presents a condensed version of the data, showing the genres studied in order of mean percent compression, from most to least compression (most to least repetitive). Throughout this section, terms such as *mean percent compression* and *average amount of repetition* will be used interchangeably.

Table 1. Mean and standard deviation of percent compression for each genre.

Genre	Mean % Compression	S.D. % Compression
R&B	59.24%	4.71%
Pop	58.94%	3.13%
EDM	58.63%	7.12%
Hip-Hop	58.10%	3.44%
Latin	56.61%	1.54%
Country	54.90%	2.95%
Metal	53.60%	8.02%
Rock	53.13%	7.55%
Folk	52.95%	5.32%
Blues	49.64%	2.67%

R&B was the most repetitive genre (59.24% compression), followed closely behind by pop (58.94%) and EDM (58.63%). Blues was the least repetitive genre by a considerable margin—the mean percent compression for folk, the second least repetitive genre, was over 3% greater than the mean percent compression for blues. The means did not vary largely among genres, staying within a fairly tight range between 49.5% and 59.5%. Therefore, even if there exists an association between genre and lyrical repetition, the association may not be very strong.

Genres differed notably in the variability of the percent compression among different songs. This is evident when looking at the standard deviation of the percent compression for each genre. In particular, metal, rock, and EDM had the highest variability in repetitiveness, as measured by percent compression ($s_x > 7\%$). Meanwhile, Latin had the lowest variability ($s_x = 1.54\%$), followed by blues ($s_x = 2.67\%$).

Since the means did not vary largely among genres, significance tests were performed in order to determine whether the differences in the means among genres were statistically significant (i.e., resulting from an actual difference in the genres' level of repetition and not simply from random chance alone). In each of these significance tests, two genres were compared at a time, with the null hypothesis (i.e., initial assumption) that their actual mean percent compressions—those of the entire population (a 100-song playlist) for each genre—are equal. Finding statistical significance in a test would mean that the alternative hypothesis is true: the two genres actually differ from each other in their average amount of repetition. In other words, the difference was not simply due to random chance, indicating that there is, in fact, an association between genre and lyrical repetition—a confirmation of the hypothesis described in the discussion of the purpose of this study.

For the significance tests, a 95% significance level was used ($\alpha = .05$). Since the data involved means, a *t*-test was used. In order to perform a two-sample *t*-test, certain conditions must be met: 1) the samples must have been selected randomly and independently of each other; 2) for each sample, the sample size (10) must be less than or equal to 10% of the population size (100) so that random sampling without replacement does not interfere with the independence of each random selection from the others; and 3) the population distribution must be a Normal distribution, or the sample size must be greater than or equal to 30 (Starnes et al., 2015). The first two conditions are met. However, the third condition is not met, since the population distribution is unknown and the sample size is 10. In addition, there are outliers in the samples for R&B, pop, rock, folk, and blues (see Appendix B), indicating that the population distribution is not Normal. Since the last condition for using a *t*-test was not fulfilled, caution must be exercised in proceeding with the significance tests and using the results.

Table 2 shows the *p*-values resulting from the significance tests for every pair of different genres. For each pair of genres, the *p*-value represents the probability of obtaining, by chance alone, the difference in repetition that

was seen in the data, assuming that the true difference is 0. This probability was calculated as the area to the right of the test statistic $t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma}$ under the *t*-distribution with a mean of 0 and a standard deviation of $\sigma = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, where $\bar{x}_1$ is the mean percent compression for the first genre, $\bar{x}_2$ is the mean percent compression for the second genre, σ_1 is the standard deviation for the first genre, σ_2 is the standard deviation for the second genre, and $n_1 = n_2 = 10$ (the sample size for each sample), following the procedure described by Starnes et al. (2015). The degrees of freedom for the *t*-distribution was automatically calculated based on these values by a computer, as part of the *t*-test.

A *p*-value less than the significance level of $\alpha = .05$ indicates a low probability of obtaining the observed difference by chance alone, meaning that there is an actual statistically significant difference in the amount of repetition between the two genres. These *p*-values are highlighted green, with darker shades of green representing greater statistical significance. On the other hand, a *p*-value greater than $\alpha = .05$ indicates a moderate or high probability of obtaining the observed difference by chance alone, meaning that the observed difference in the amount of repetition between the two genres could have reasonably occurred simply by chance, without any actual difference (not statistically significant). These *p*-values are highlighted red, with darker shades of red representing less statistical significance.

Table 2. Significance tests for each pair of genres.

	R&B	Pop	EDM	Hip-Hop	Latin	Country	Metal	Rock	Folk	Blues
R&B		0.4342	0.4117	0.2717	0.0607	0.0129	0.0376	0.0231	0.0060	0.0000
Pop	0.4342		0.4506	0.2869	0.0273	0.0041	0.0372	0.0221	0.0040	0.0000
EDM	0.4117	0.4506		0.4175	0.2009	0.0759	0.0781	0.0557	0.0300	0.0015
Hip-Hop	0.2717	0.2869	0.4175		0.1176	0.0194	0.0646	0.0408	0.0106	0.0000
Latin	0.0607	0.0273	0.2009	0.1176		0.0633	0.1363	0.0924	0.0311	0.0000
Country	0.0129	0.0041	0.0759	0.0194	0.0633		0.3208	0.2524	0.1646	0.0003
Metal	0.0376	0.0372	0.0781	0.0646	0.1363	0.3208		0.4470	0.4164	0.0833
Rock	0.0231	0.0221	0.0557	0.0408	0.0924	0.2524	0.4470		0.4754	0.0974
Folk	0.0060	0.0040	0.0300	0.0106	0.0311	0.1646	0.4164	0.4754		0.0510
Blues	0.0000	0.0000	0.0015	0.0000	0.0000	0.0003	0.0833	0.0974	0.0510	

All *p*-values were evaluated at the $\alpha = .05$ level for a one-tailed test. No genre was compared to itself.

Looking at examples can help to understand these results. For example, the significance test between R&B and pop yielded a *p*-value of 0.4342, meaning that there is a high (43.42%) probability of obtaining the difference in repetition between R&B and pop (59.24% – 58.94% = 0.30%) by chance alone. Indeed, a 0.30% difference is quite small. Meanwhile, the significance test between hip-hop and country yielded a *p*-value of 0.0194, meaning that there is only a 1.94% probability of obtaining the difference between hip-hop and country (58.10% – 54.90% = 3.2%) by chance alone. This is statistically significant, suggesting that there actually *is* a difference in repetition between hip-hop and country beyond the samples. Accordingly, a 3.2% difference is relatively large.

The results of the significance tests allow the musical genres to be grouped into more distinct categories, where each category is distinctly different from the others in terms of average amount of repetition. On the other hand, the genres themselves are not always distinctly different from each other; for example, R&B and pop did not differ significantly in their average amount of repetition. Grouping sections of statistically insignificant *p*-values (red) together and making the divisions between categories reflect the divisions between insignificant (red) and significant (green) *p*-values, a reasonable categorization of the genres is as follows: *R&B, pop, EDM, and hip-hop are the most*

repetitive; Latin and country have moderate levels of repetition; and metal, rock, folk, and blues are the least repetitive. This categorization places genres whose measures of repetitiveness differ significantly from each other in the same category while placing genres whose measures of repetitiveness do not differ significantly from each other in separate categories.

In conclusion, looking at the mean percent compression for various genres, performing significance tests, and being able to use them to organize the genres into categories by repetitiveness shows that there is an association between musical genre and lyrical repetition, confirming this study's hypothesis. However, these results must be used with caution, as the population distributions for the mean percent compression of each genre might not be Normal, especially for R&B, pop, rock, folk, and blues.

Discussion

Seminal literature in the study of music established that repetition influences liking and that genre influences liking (Eerola, 2011; Hargreaves, 1984; Getz, 1966; Margulis & Simchy-Gross, 2016; Williams, 2018). This study was able to fill a gap in the research by connecting these two ideas, confirming the original hypothesis that there is an association between repetition and genre.

Looking at all the genres, the mean percent compression ranged from 49.5% to 59.5%. This roughly means that, for most songs in most genres, over half of the lyrics represent repeated material (with the exception of blues [49.64% compression], which nonetheless comes very close to having over 50% compression). Seeing as repetition influences liking—most notably, repetition increased ratings of musicality in randomly generated musical excerpts (Margulis and Simchy-Gross, 2016)—the fact that this study found such high amounts of repetition in actual songs of all genres suggests, as a general phenomenon, that musical artists use repetition to increase the appeal of their songs.

The methodology and the outcomes of this study add to those of other studies that looked at common features within genres. For example, Hart and Day (2020) found that rock contains sexual content moderately correlated with lyrics expressing anxiety, while rap and R&B has higher frequencies of sexual content, strongly correlated with lyrics expressing anger. Meanwhile, Abeßer et al. (2012) were able to classify genres based on the rhythm, tonality, and other musical features of their repetitive basslines. Additionally, Thompson (2012) was able to classify genres based on linguistic features in song lyrics, such as rhymes, readability, and profanity. This study was able to develop a new criterion by which musical genres can be classified: repetitiveness.

Interestingly, folk and rock—genres in which artists compose lyrics after instrumentals, according to Negus and Astor (2015)—were placed in the same category (least repetitive) in this study. Additionally, rap and R&B, both of which Hart and Day (2020) found to have frequent sexual content strongly correlated with lyrics expressing anger, were also placed in the same category (most repetitive). This suggests that lyrical similarities among genres extend to a broad range of lyrical features beyond repetition and that there may be an association between lyrical repetition and these other lyrical features.

Conclusion

This study involved a number of limitations. First, in generalizing the results of the study from the samples to the playlists to the entire populations for each genre, Chosic's genre playlists were assumed to be representative of the entire population of songs belonging to each genre, though this is not necessarily the case. The difficulty in obtaining representative samples of songs for each genre arises partially from the fact that the boundaries between various genres are not clearly and unambiguously defined, which means that the classification of certain songs into various genres can vary among sources.

Another limitation is that the results of the study must be used with caution since the conditions for the *t*-tests were not fully met. For example, the sample data for R&B, pop, rock, folk, and blues showed outliers. This would

not be a problem if the sample size were larger; a sample size of 30 or more would allow the *t*-distribution to be used as an approximation of the sampling distribution, regardless of the shape of the population distribution, when performing the statistical inference with the *t*-tests (Starnes et al., 2015). However, this study sampled only 10 songs per genre since the playlists only had 100 songs. Repeating this study with a greater sample size would yield stronger evidence of the association between musical genre and lyrical repetition, suggesting a direction for future research.

This study also implies another direction for future research. In developing the methodology for this study, different types of repetition—lyrical, instrumental, and melodic—were identified and defined. Ultimately, however, this study only focused on lyrical repetition due to time constraints, as well as a lack of existing tools for transcribing instrumentals and melodies and measuring their repetitiveness. In order to form a more comprehensive picture of the relationship between musical genre and repetition, it would be valuable for future studies to investigate this relationship in terms of instrumental repetition and melodic repetition. Instrumental repetition might be examined by recording the chord progressions of songs and measuring the repetitiveness of these progressions. Melodic repetition might be examined by transcribing the melodies of songs—including the pitch and rhythm of individual notes—and measuring the repetitiveness of these melodies. In both cases, the measure of repetitiveness might be obtained using a method similar to the one used in this study, which utilized the Lempel-Ziv algorithm, to measure lyrical repetition.

The connection between genre and liking represents another direction for future research. It has been found that repetition increases familiarity and preference and that genre influences liking (Getz, 1966; Hargreaves, 1984; Williams, 2018). However, Hargreaves (1984) only studied a few relatively uncommon genres: "easy-listening" music, avant-garde jazz, popular music, and classical music. In addition, Getz (1966) only studied the effect of musical characteristics—such as tempo, volume, and melodic features—and not actual genres. It would be valuable for future research to investigate liking for actual genres that are relatively common, such as the genres examined in this study, in order to further clarify the relationship among repetition, liking, and genre (although the subjectivity of preference for different genres may present a challenge in conducting such research).

Ultimately, the outcomes of this study have some implications. For instance, the results suggested that musical artists use repetition to increase the appeal of their songs (since the repetition levels were quite high across all genres). This reinforced existing knowledge within the music industry, which makes both musical artists and listeners of music more aware of the influence of repetition on their experience of music, as well as how this influence varies based on the genre that artists are producing or that listeners are listening to.

References

Abeßer, J., Lukashevich, H., & Bräuer, P. (2012). Classification of Music Genres Based on Repetitive Basslines. *Journal of New Music Research, 41*(3), 239–257. https://doi.org/10.1080/09298215.2011.641571

Appel, N. (2014). "Ga, ga, ooh-la-la": the childlike use of language in pop-rock music. *Popular Music*, *33*(1), 91–108. http://www.jstor.org/stable/24736973

Deutsch, D., Henthorn, T., & Lapidis, R. (2008). The speech-to-song illusion. *Journal of the Acoustical Society of America, 124*(4), 2471. https://doi.org/10.1121/1.4808987

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America, 129*(4), 2245–2252. https://doi.org/10.1121/1.3562174

Eerola, T. (2011). Are the Emotions Expressed in Music Genre-specific? An Audio-based Evaluation of Datasets Spanning Classical, Film, Pop and Mixed Genres. Journal of New Music Research, 40(4), 349–366. https://doi.org/10.1080/09298215.2011.602195

Fujishima, T. (1999). Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music. *Proceedings of the 1999 International Computer Music Conference* (pp. 464–467). San Francisco: International Computer Music Association.

Getz, R. P. (1966). The Effects of Repetition on Listening Response. *Journal of Research in Music Education*, *14*(3), 178–192. https://doi.org/10.2307/3344051

Hargreaves, D. J. (1984). The Effects of Repetition on Liking for Music. *Journal of Research in Music Education*, *32*(1), 35–47. https://doi.org/10.2307/3345279

Hart, C. B., & Day, G. (2020). A Linguistic Analysis of Sexual Content and Emotive Language in Contemporary Music Genres. *Sexuality & Culture, 24*(3), 516–531. https://doi.org/10.1007/s12119-019-09645-z

Huffman, D. (1952). A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE, 40*(9), 1098–1101. https://doi.org/10.1109/JRPROC.1952.273898

Jakobovits, L. A. (1962). *Effects of Repeated Stimulation on Cognitive Aspects of Behavior: Some Experiments on the Phenomenon of Semantic Satiation* [Doctoral dissertation, McGill University]. https://escholarship.mcgill.ca/concern/theses/c821gp587

Klapuri, A. (1998). *Automatic Transcription of Music*. M.Sc. Thesis, Tampere University of Technology, Finland.

LZW compression. (2023, January 25). Rosetta Code. https://rosettacode.org/wiki/LZW_compression

Margulis, E. H. (2012). Musical Repetition Detection Across Multiple Exposures. *Music Perception: An Interdisciplinary Journal*, *29*(4), 377–385. https://doi.org/10.1525/mp.2012.29.4.377

Margulis, E. H., & Simchy-Gross, R. (2016). Repetition Enhances the Musicality of Randomly Generated Tone Sequences. *Music Perception: An Interdisciplinary Journal*, *33*(4), 509–514. https://www.jstor.org/stable/26417434

McGeoch, C. C. (1993). Data Compression. *The American Mathematical Monthly*, *100*(5), 493–497. https://doi.org/10.2307/2324310

Morris, Colin. (2017, May). *Are Pop Lyrics Getting More Repetitive?* The Pudding. https://pudding.cool/2017/05/song-repetition/

Negus, K., & Astor, P. (2015). Songwriters and song lyrics: architecture, ambiguity and repetition. *Popular Music*, *34*(2), 226–244. http://www.jstor.org/stable/24736910

Nettl, B. (1983). *The study of ethnomusicology: Thirty-one issues and concepts*. Urbana-Champaign, IL: University of Illinois Press.

Nunes, J. C., Ordanini, A., & Valsesia, F. (2015). The power of repetition: repetitive lyrics in a song increase processing fluency and drive market success. *Journal of Consumer Psychology, 25*(2), 187–199. https://doi.org/10.1016/j.jcps.2014.12.004

Starnes, D. S., Tabor, J., Yates, D.S., & Moore, D.S. (2015). *The practice of statistics* (5th ed.). Ann Heath.

Taher, C., Rusch, R., & McAdams, S. (2016). Effects of Repetition on Attention in Two-Part Counterpoint. *Music Perception: An Interdisciplinary Journal*, *33*(3), 306–318. https://www.jstor.org/stable/26417409

Thompson, C. (2021). Lyric-Based Classification of Music Genres Using Hand-Crafted Features. *Reinvention: An International Journal of Undergraduate Research, 14*(2), 1–19. https://doi.org/10.31273/reinvention.v14i2.705

Williams, M. L. (2018). The Influence of Genre and Musical Training on Ratings of Listener Enjoyment. *Bulletin of the Council for Research in Music Education*, *217*, 67–83. https://doi.org/10.5406/bulcouresmusedu.217.0067

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zeeh, C. (2013, January 16). *Famous Algorithms: The Lempel Ziv Algorithm* [PDF]. Retrieved from https://ece.uwaterloo.ca/~ece611/LempelZiv.pdf

Ziv, J. & Lempel, A. (1977). A Universal Algorithm for Sequential Data Compression. *IEEE Transactions on Information Theory, 23*(3), 337–343. https://doi.org/10.1109/TIT.1977.1055714