EXPLORING ACOUSTIC SIMILARITY FOR NOVEL MUSIC RECOMMENDATION

Derek Cheng
Cornell University
dsc252@cornell.edu

Thorsten Joachims
Cornell University
tj@cs.cornell.edu

Douglas Turnbull
Ithaca College
dturnbull@ithaca.edu

ABSTRACT

Most commercial music services rely on collaborative filtering to recommend artists and songs. While this method is effective for popular artists with large fanbases, it can present difficulties for recommending novel, lesser known artists due to a relative lack of user preference data. In this paper, we therefore seek to understand how content-based approaches can be used to more effectively recommend songs from these lesser known artists. Specifically, we conduct a user study to answer three questions. Firstly, do most users agree which songs are most acoustically similar? Secondly, is acoustic similarity a good proxy for how an individual might construct a playlist or recommend music to a friend? Thirdly, if so, can we find acoustic features that are related to human judgments of acoustic similarity? To answer these questions, our study asked 117 test subjects to compare two unknown candidate songs relative to a third known reference song. Our findings show that 1) judgments about acoustic similarity are fairly consistent, 2) acoustic similarity is highly correlated with playlist selection and recommendation, but not necessarily personal preference, and 3) we identify a subset of acoustic features from the Spotify Web API that is particularly predictive of human similarity judgments.

1. INTRODUCTION

Suppose you want to recommend a novel artist B to a friend who you know likes some artist A. Which of B's songs should you recommend that they listen to first? As we will argue in the following, the answer to this question involves concepts of acoustic similarity, playlist context, and personal preference as they all relate to the task of music recommendation [18].

The motivation for asking this question comes from an application we are developing called Localify [1] that recommends artists from a listener's hometown based on the listener's favorite artists [7]. While most of these favorite artists will tend to be popular, well-known artists, the vast majority of local artists are likely to be relatively obscure *long-tail* artists [1, 3, 11]. There is often limited user preference data (listening histories, like/dislike ratings) associated with these long-tail artists and even less associated with their individual songs.

Automatic playlist generation is one of the core components of our locally-focused music recommender system. As we know from recommender systems research, providing the user with a contextual explanation of the recommendation is important because it provides information that is both useful for decision making and for developing trust in the system [13, 21]. We also know from music psychology research that listeners tend to prefer familiar music [6, 15]. To this end, our playlist algorithm creates a list of songs by alternating between a song by a favorite artist and a song by a local artist. This allows the algorithm to balance exploiting familiarity from known artists with exploring novel artists from the local music scene.

Most commercial recommender systems make use of listening histories from a large number of listeners to make accurate recommendations [2]. This technique is referred to as collaborative filtering (CF). CF systems suffer the cold-start problem [19]: little or no historical user preference data exists for new or obscure artists. As a result, a CF-based recommender system cannot recommend these artists with sufficient confidence. An alternative to CF systems are content-based (CB) recommender systems that make use of the audio signal for recommendation.

Although many of the local artists we want to recommend are relatively obscure, we have informally found that our CF system is able to accurately recommend local artists in most cases. However, it tends to do a poor job of recommending relevant songs for many of our obscure local artists. To this end, we have been exploring the suitability of CB recommendation based on acoustic similarity between a given reference song and a set of candidate songs in order to improve our playlist algorithm.

1.1 Research Questions

To better understand the role of acoustic similarity in recommending long-tail artists, we address the following three research questions in this paper:

RQ1 Are human judgments about acoustic similarity consistent across users?

RQ2 Is acoustic similarity a good proxy for how an individual might construct a playlist, recommend music to a friend, or prefer one song over another?

[1] https://www.localify.org

RQ3 If so, what are some of the measurable acoustic properties that correlate with how humans judge acoustic similarity?

The first question acknowledges that music is subjective and people will naturally have differing opinions [5]. However, if there is little or no consistency in how listeners perceive songs and compare them to one another, then modeling music on the basis of acoustic similarity will be unreliable.

While the answer to the second question may seem obvious, many listeners have eclectic tastes and expect some amount of variety, so it is possible that having music that is too acoustically similar could result in homogeneous playlists and boring recommendations. Additionally, listeners bring non-musical context to bear when listening to music. This includes the meaning of the lyrics, the artist's social persona, visual media (album covers, music videos), trust in the source of the recommendation, the listener's emotional state, etc. In addition, an individual's perception of music is influenced by a variety of socio-demographic factors including age, race, ethnicity, gender, social class, family, political beliefs, and values [12]. To this end, it is entirely possible that the acoustic content is only a small part of the equation when it comes to making a successful recommendation [17].

We address the first two questions by asking individuals to choose among two unknown candidate songs (B1 and B2) from an unknown artist B in the context of a third familiar reference song (A1) by a known artist A. Specifically, we ask which one (B1 or B2) is more acoustically similar to A1, which one they would recommend to a friend who likes A1, which one they would include in a playlist after A1, and which one they personally prefer. Throughout this paper, we will refer to a specific set of three songs $\{A1, B1, B2\}$ as a *song tuple*, and we asked a large number of test subjects to each evaluate a subset of 12 total song tuples (see Table 1) based on their genre preferences.

For the third question, there has been a good deal of research within the music information retrieval community that focuses on using digital signal processing and machine learning to estimate acoustic similarity (e.g., MIREX Audio Similarity Task [2], [10,14,22]). We will not advance the state-of-the-art in this paper but rather simply explore how acoustic properties such as estimated tempo, danceability, and valence correlate with a human listener's judgments related to acoustic similarity. For this, we both analyze open-ended responses from listeners and examine correlations between 11 song-level audio features that we obtain from the public Spotify API for each of the songs in our study.

1.2 Related work

Our work is related to other studies that explore how people interact with music recommender systems. One recent study by Lee et al. [13] found that there are many factors, including aesthetic qualities, familiarity, trust in the recommender, and contextualization, that affect whether a person will adopt a music recommendation. This is consistent with our findings that while acoustic similarity is important for music recommendation, other non-content-based information also plays a big role (see Table 2.)

Zhang et al. [23] stress the importance of serendipity and warn about the dangers of self-reinforcing "filter bubbles" when music recommender systems focus too much on optimizing accuracy. Our recommender system embraces these ideas by attempting to introduce novel artists through locally-focused music recommendation. In addition, the design of our survey was influenced by their suggestion that a recommender system should be akin to having a trusted friend recommend music.

Our application also reflects the findings of Jun et al. [8] who suggest that blending songs in a specific order can improve the quality of the playlist. Specifically, we would like to play a familiar song followed by an acoustically similar song by a local artist so that they flow together. The idea is that a user will be more likely accept the local music recommendation if it sounds similar to something they already enjoy.

Finally, we refer the reader to both Lee et al. [13] and Laplante [12] for recent and comprehensive literature reviews on studies related to human-centered music recommendation.

2. METHODOLOGY OF STUDY

In this section, we describe both how we selected the song tuples and how we designed the user study.

2.1 Song Tuple Selection

We began by collecting 12 song tuples of 3 songs each, for a total of 36 songs. We first selected 4 genres (pop, rock, hip hop, and R&B) with 3 tuples assigned to each genre. Each tuple consisted of a song by a popular artist from its genre, as well as two songs by a relatively lesser-known artist from the same genre. We refer to the popular artist and song respectively as the A artist and A1 song, and the lesser-known artist and songs respectively as the B artist and B1 & B2 songs. For example, in one of the song tuples from the pop genre, the A artist is Billie Eilish and the A1 song is *bad guy*, while the B artist is Gabbie Hanna and the B1 & B2 songs are *Honestly* and *Butterflies*.

The A artists were determined by finding popular artists associated with each of our four genres using the Spotify API [3]. The A1 songs were chosen from the most popular tracks for each artist to maximize the likelihood of them being recognized by our study's participants. The B artists were chosen from a large corpus of artists from our own application such that each B artist was listed as being related to artist A according to the Spotify Web API [4]. We also ensured that the B artists and their B1/B2 songs had limited popularity so that they would likely not be familiar to our study's participants to better simulate long-tail music recommendation.

[2] https://www.music-ir.org/mirex

[3] https://developer.spotify.com/documentation/web-api/reference/artists/get-artist/

[4] https://developer.spotify.com/documentation/web-api/reference/artists/get-related-artists/

While we generally selected the most popular songs for both the A and the B artists, we skipped over songs in the popularity ranking under the following conditions. For A1 song selection, we skipped over a song if:

1. The song is a cover of a song by another artist
2. The song is actually by another artist where the current artist is only featured
3. The song was very recently released and thus high in popularity but still relatively low in number of streams

For B1/B2 song selection, criteria 1 and 2 were applied as well. In addition, we also skipped over a potential B1/B2 song if it had an excessively high number of streams due to being a "one hit wonder" so as to reduce the likelihood that test subjects would recognize it.

2.2 Survey Design

Our three section study was conducted through an online Qualtrics survey. In the first section, each participant was asked to provide demographic and psychographic data (e.g., age, gender, time spent listening to music daily, preferred streaming services). Additionally, we asked participants for the two genres they were most familiar with from our four genres of pop, rock, hip hop, and R&B.

The second section collected quantitative data regarding the participants' preferences for playlist selection, recommendation to a friend, acoustic similarity, and personal preference. In this section, participants were presented with six song tuples — three song tuples associated with each of their two selected genres — and asked to answer questions regarding each song tuple. The song tuples were shown by genre, with the ordering of the song tuples within each genre as well as the ordering of the genres themselves being randomized. Additionally, within each song tuple, the ordering of the B1 and B2 songs was also randomized for every participant.

For each song tuple, the participants were first asked to listen to clips of the first 30 seconds of the A1 song and the two B1/B2 songs. The length of 30 seconds was chosen firstly to minimize fatigue and restrict the survey to a reasonable length, and secondly to give the user a sufficiently long sample to form a solid impression of the music [20]. Once they had listened to the clips, they were prompted to answer the following questions regarding the two B1/B2 songs:

1. If you were creating a playlist with Song A1 and either B1 or B2, which one would you pick?
2. If you had a friend who likes Song A1 by Artist A and you wanted to introduce them to Artist B, which song would you recommend to them first?
3. Which song is most acoustically similar to Song A1?
4. Which song do you prefer?

In addition to choosing either B1 or B2, participants were also allowed to answer "About the same" if they could not decide. Finally, we also asked participants if they recognized any of the songs or artists presented in the song tuple, for which they could answer "Yes", "No", or "Maybe".

In the third section, participants were asked the following open-ended questions:

1. When deciding to pick specific songs for a playlist, what do you consider to be most important?
2. When deciding songs to recommend to a friend, what do you consider to be most important?
3. When comparing songs in terms of acoustic similarity, what do you consider to be most important?

These three questions were intended to collect qualitative data regarding how participants made their decisions in the second section of the survey.

3. SURVEY DATA

Participants were recruited on a voluntary basis with no compensation via email lists and social media from the authors' academic and social circles, based primarily within the United States. We received responses from 113 participants, with 103 of these considered valid. From these participants, the youngest represented age group was 17 or younger, while the oldest was 61 - 70, with the median age group being 21 - 25. 58.25% of participants identified as male, while 41.75% identified as female. Participants indicated usage of seven different streaming services, with Spotify being by far the most popular. Some participants also indicated usage of older music playback technologies, such as iTunes libraries, CDs, and vinyl records.

3.1 Song Tuple Responses

From the 103 valid responses, we counted a total of 612 song tuple evaluations, with 317 of these considered valid. We considered a song tuple evaluation valid based on the following two criteria. Firstly, the participant must recognize the A1 song and must not recognize the B1 and B2 songs. This is because our set of questions regarding a song tuple were intended to be asked under the assumption that the A1 song was known, and the B1 and B2 songs were both unknown. Secondly, the participant must spend at least 60 seconds evaluating the tuple. We defined 60 seconds per tuple as the threshold at which a participant is considered to have faithfully answered our questions. Based on these criteria, we discarded responses which did not contain any valid tuple evaluations, and we discarded any invalid tuple evaluations from the remaining responses. Each song tuple was evaluated by at least 14 participants, with a mean of 26.4 participants.

3.2 Qualitative Feedback

One author coded the responses to the three open-ended questions from the third section of the survey. Initial categories (e.g., texture content, rhythmic content, context, preference, playlist mix) and subcategories (e.g., energy, tempo, variety) were formed after making a first pass over 95 non-empty responses out of a total of 103 survey responses. Each response was then coded according to these categories and subcategories. The results are found in Table 2.

The main takeaway from our coding exercise is to note that, as expected, acoustic similarity (Q1) is almost entirely related to audio content, while playlist song selection (Q2) and song recommendation (Q3) involve both audio content

Table 1. The 12 song tuples across four genres, consisting of 3 songs per tuple, used in our survey. For each song tuple, the first song listed is the well-known reference song (A1), followed by the more acoustically similar B1/B2 song, and then the less acoustically similar B1/B2 song. The first line for each of our four concepts (acoustic similarity, playlist selection, recommendation, and personal preference) represents the number of participants respectively that selected the first place song / indicated that they were the same / selected second place song based on acoustic similarity. The second line represents the p-value for a binomial hypothesis test in which the null hypothesis assumes that B1/B2 songs are equally likely to be selected by a participant. Song tuples are sorted by these p-values for acoustic similarity. Bold font indicates statistically significant differences at the $\alpha < 0.05$ level. Italics indicate that participants generally preferred the less acoustically similar song.

Genre	Artist	Song	Acoustic Similarity	Playlist Selection	Recommend	Preference
Rock	The Beatles Aviator Stash Aviator Stash	Here Comes The Sun Hype Tyler the Beat	45 / 12 / 5 **0.000**	42 / 15 / 5 **0.000**	42 / 15 / 5 **0.000**	31 / 18 / 13 **0.006**
Hip Hop	Nicki Minaj Mulatto Mulatto	Anaconda B*tch From Da Souf Longway	19 / 0 /1 **0.000**	14 / 3 / 3 **0.010**	14 / 5 / 1 **0.001**	*5 / 6 / 9* *0.244*
Rock	Paramore Tonight Alive Tonight Alive	Still into You Lonely Girl The Other Side	14 / 2 / 1 **0.001**	13 / 3 / 1 **0.002**	12 / 1 / 4 0.056	8 / 7 / 2 0.088
Pop	Post Malone Lil Xan Lil Xan	rockstar Lies Color Blind	22 / 8 / 5 **0.001**	17 / 8 / 10 0.126	15 / 11 / 9 0.156	*10 / 10 / 15* *0.195*
Pop	Billie Eilish Gabbie Hanna Gabbie Hanna	bad guy Honestly Butterflies	29 / 5 / 10 **0.002**	29 / 6 / 9 **0.001**	28 / 5 / 11 **0.006**	20 / 13 / 11 0.079
Hip Hop	Cardi B Kash Doll Kash Doll	I Like It Doin Too Much No Lames	13 / 2 / 3 **0.017**	12 / 2 / 4 0.056	8 / 4 / 6 0.367	*6 / 5 / 7* *0.419*
Rock	Imagine Dragons The Score The Score	Believer Unstoppable Legend	13 / 5 / 3 **0.017**	8 / 10 / 3 0.161	10 / 8 / 3 0.070	*3 / 9 / 9* *0.107*
Hip Hop	Drake Kahiem Rivera Kahiem Rivera	One Dance Smokin' Weed with the Devil Good Winter	14 / 2 / 4 **0.023**	13 / 3 / 4 **0.036**	11 / 6 / 3 **0.044**	7 / 9 / 4 0.322
Pop	The Weeknd Myer Clarity Myer Clarity	Starboy Love Me When I'm High All the Way Down	21 / 5 / 10 **0.041**	13 / 8 / 15 0.279	18 / 6 / 12 0.161	*13 / 6 / 17* *0.223*
R&B	Beyoncé Keri Hilson Keri Hilson	Halo Energy - Main Final Got Your Back	9 / 2 / 3 0.107	12 / 0 / 2 **0.011**	11 / 0 / 3 **0.044**	7 / 4 / 3 0.234
R&B	Camila Cabello Ally Brooke Ally Brooke	Havana Lips Don't Lie No Good	9 / 2 / 4 0.175	9 / 2 / 4 0.175	10 / 0 / 5 0.183	10 / 1 / 4 0.122
R&B	Frank Ocean Syd Syd	Thinkin Bout You Getting Late Over	9 / 1 / 5 0.244	10 / 2 / 3 0.070	9 / 3 / 3 0.107	10 / 2 / 3 0.070

Table 2. Coded responses from 95 participants to questions about acoustic similarity, playlist selection, and music recommendation. The numbers in parenthesis reflect the number responses for each label.

Q1: When comparing songs in terms of **acoustic similarity**, what do you consider to be most important?	
Textural content (71)	Instrumentation (37), Vibe/tone (15), Vocal tone (13), Production/mix (6)
Rhythmic content (65)	Tempo (37), Beat/rhythm (24), Bass line (3), Repetitiveness (1)
Dynamic content (13)	Energy (8), Dynamic range(2), Volume (2), Brightness/intensity (1)
Musicological concepts (8)	Genre/style (4), Mood/expression (4)
Harmonic content (7)	Key (5), Chords (2)
Context (3)	Lyrics (3)
Q2: When deciding to pick specific songs for a playlist, what do you consider to be most important?	
Textural content (49)	Vibe/tone (25), Acoustic similarity (18), Instrumentation (5), Vocal tone (3)
Musicological concepts (32)	Mood/expression (19) , Genre/style (12), Time period (1)
Preference (17)	Personal preference (9), Personal mood (6), Catchy (1)
Playlist mix (17)	Flow between songs (10), Variety (7)
Context (13)	Lyrics (7), Theme (6)
Rhythmic content (13)	Tempo (9), Beat (4)
Dynamic content (7)	Energy (7)
Q3: When deciding songs to recommend to a friend, what do you consider to be most important?	
Preference (59)	Friend will like (32), Personal preference (16), Interesting to me (5), Catchy (5), Originality (1)
Textural content (30)	Acoustic similarity (20), Vibe/tone (5), Instrumentation (3), Vocal tone (2)
Musicological concepts (11)	Genre/style (11)
Context (6)	Lyrics (3), Meaning (2), Artist background (1)
Rhythmic content (3)	Beat (3)

as well as information that is not directly related to audio content such as personal preference, artist background, and lyrical meaning. That is, while audio similarity is important when creating playlists and recommending music, it is not the only factor that listeners use to make decisions.

4. DISCUSSION OF RESULTS

In this section, we address the three questions that we initially posed using the data that we collected from our user study.

4.1 RQ1: Consistency of Acoustic Similarity Judgments

The first question we explore is the extent to which judgments about acoustic similarity are consistent across many listeners. To do this, we examine how often one of the B1/B2 songs is designated as being more acoustically similar to the A1 song for each of the 12 song tuples. The third column of Table 1 reports the raw counts of how often the *winning* B1/B2 song is designated as being more, equally, or less acoustically similar by our study's participants. We also conducted a binomial hypothesis test where the null hypothesis assumes that both B1/B2 songs are equally similar to the A1 song.

In 9 of the 12 tuples, the similarity judgment was significantly ($\alpha < 0.05$) pointing in one direction [5] , suggesting that there was a winner between the B1/B2 songs. The other three tuples showed majorities at or above 64% of the vote. These three tuples were the three R&B tuples that received the fewest number of evaluations, reducing the statistical power of the tests. Overall, the results suggest that listeners are somewhat consistent in their judgment of acoustic similarity even when comparing songs by similar artists.

[5] When applying a Bonferroni correction for multiple hypothesis tests, we observe that 5 of the 12 tuples have p-values less than $\alpha < 0.004$.

Table 3. Correlation coefficients for 317 song tuple trials when comparing pairs of acoustic similarity, playlist selection, song recommendation, and personal preference.

	Playlist	**Recommend**	**Pref.**
Aco. Similarity	0.716	0.595	0.387
Playlist		0.596	0.386
Recommend			0.116

4.2 RQ2: Relationship between Playlist Selection and Recommendation

To answer our second research question, we look at the correlation between how participants voted for B1 or B2 relative to A with respect to assessing acoustic similarity, song selection for playlist creation, music recommendation, and personal preference. In Table 3, we report the correlation coefficients between pairs of these four concepts. We also conducted a two-tailed hypothesis test for each of these correlation coefficients and found all six to be highly statistically significant, with five p-values less than 0.001 and the p-value for recommendation and preference equal to 0.038.

We see that acoustic similarity is most highly correlated with playlist selection, suggesting that listeners consider acoustic similarity to be important when constructing playlists. This high correlation is also supported by the qualitative feedback (see Table 2) in which *acoustic similarity* is explicitly mentioned as being an important consideration by 18 participants. The participants also mention a large number of acoustic concepts related to texture, rhythm, and dynamics.

We also observe a high level of correlation between acoustic similarity and recommendation. This is consistent with the coded qualitative feedback (see Table 2) in which 20 participants explicitly mention the role of acoustic similarity in music recommendation. However, the slightly lower level of correlation might be attributed to the fact that preference seems to play a greater role in recommendation than acoustic similarity.

Finally, we note a lower level of correlation with personal preference. This is unsurprising since in the previous two cases, we asked participants to make judgments between B1 and B2 relative to the A1 song. It is reasonable that one of the B1/B2 songs would fit more naturally, but that the participant would personally prefer the other B1/B2 song. This, in fact, occurred in 5 of our 12 tuples (see the right most column of Table 1) suggesting that making a recommendation relative to a reference song is different from simply picking preferred songs in isolation.

4.3 RQ3: Acoustic Similarity with Acoustic Features

Here, we explore how a set of song-level content-based features are related to human judgments about acoustic similarity. The specific set of features that we use are obtained using the *Audio Features for a Track* endpoint from the Spotify Web API [6]. These eleven mid-level song features include danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, and tempo.

Our approach involved conducting a one-tailed paired t-test over the 12 song tuples for each of the 11 audio features. For each tuple, we calculate the magnitude (absolute value) of the difference of an audio feature between the the winning B1/B2 song and the reference track A1, and the magnitude of the difference between the losing B1/B2 song and A1. Here, our alternative hypothesis is that the deviation between the winning B1/B2 song and A1 will be smaller than losing B1/B2 and A1 for a given feature. This would imply that the audio feature encodes information that is used by listeners to assess acoustic similarity.

None of the t-tests for any of the 11 audio features revealed any statistically significant differences at the 0.05 confidence level, suggesting that there is no obvious single audio feature that can be used directly to predict acoustic similarity. However, five features shown in Table 4 have a p-value less than $\alpha < 0.15$, suggesting that these features may be related to acoustic similarity. We should note that our sample size with $n = 12$ tuples is small and, as a result, the power of our hypothesis test was limited. A future study with a larger number of song tuples (i.e., more statistical power) may result in more statistically significant results.

It is interesting to note that two of these features, valence and energy, roughly correspond to the first two dimensions of Russell's classic circumspect model of affect [16] that is frequently used to model mood in emotion [9]. Tempo is a rhythmic audio feature and was explicitly mentioned by a large number of our study's participants as being important for assessing acoustic similarity (see Table 2) as well as for song selection when creating playlists. The other two features, speechiness and liveness, are less typical features that were engineered by researchers first at The Echo Nest [7] and now at Spotify [8] to describe the texture of a song. Taken together, this set of five features reflects song texture, rhythmic properties, and perceived mood, suggesting that acoustic similarity is likely to be multi-faceted.

Table 4. Spotify audio features for p-values less than $\alpha < 0.15$.

Acoustic Feature	p-value
valence	0.07
speechiness	0.11
tempo	0.12
liveness	0.13
energy	0.13

5. CONCLUSION

In this paper, we have explored the relationships between the concepts of acoustic similarity, song selection for playlist creation, music recommendation, and personal preference. Through our user study, we have found that:

1. There is a degree of consistency among human judgments of acoustic similarity.
2. Acoustic similarity is highly correlated with playlist selection and recommendation, but not personal preference.
3. While certain acoustic features obtained seem to be related to acoustic similarity, additional evidence (i.e., more tuples) is necessary to support this with statistical certainty.

These findings are especially significant for the task of recommending songs by obscure, long-tail artists. They provide empirical support for the usage of content-based recommender systems when lack of user preference data precludes the effective functioning of collaborative filtering-based recommender systems. This can apply to both more general recommendation tasks, as well as specific ones like next-song selection for playlist generation.

Building upon the findings described in this paper, a potential avenue for further investigation would be another user study, in a similar vein to our study, but covering a much wider range of music than the 36 songs included in our study, and perhaps with fewer human evaluations per song tuple. This would allow us to determine with statistical certainty whether specific acoustic features encode information about acoustic similarity, therefore providing insight into what acoustic features should be taken into account when building a content-based recommender system [4]. Ultimately, the findings from this paper combined with additional research will aid in the development of a more effective novel-artist recommender system for Localify and other similar music recommendation services.

[6] https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

[7] https://github.com/echonest/pyechonest

[8] https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/

6. ACKNOWLEDGMENTS

All survey data collected for this paper can be found at the author's website. [9] We have also created a shared Spotify playlist [10] that contains the 36 songs that were used in this study to help the reader better interpret our survey results. This project is supported by NSF grant IIS-1615706/1615679 and IIS-1901168/1901330.

7. REFERENCES

[1] Chris Anderson. *The long tail: Why the future of business is selling less of more*. Hachette Books, 2006.

[2] Luke Barrington, Reid Oda, and Gert RG Lanckriet. Smarter than genius? human evaluation of music recommender systems. In *ISMIR*, volume 9, pages 357–362. Citeseer, 2009.

[3] Oscar Celma. Music recommendation. In *Music recommendation and discovery*, pages 43–85. Springer, 2010.

[4] Joseph Cleveland, Derek Cheng, Michael Zhou, Thorsten Joachims, and Douglas Turnbull. Content-based music similarity with siamese networks. In *Machine Learning for Media Discovery (ML4DM) Workshop at the International Conference on Machine Leanring (ICML)*, 2020.

[5] Arthur Flexer and Taric Lallai. Can we increase inter- and intra-rater agreement in modeling general music similarity? 2019.

[6] Patrick G Hunter and E Glenn Schellenberg. Interactive effects of personality and frequency of exposure on liking for music. *Personality and Individual Differences*, 50(2):175–179, 2011.

[7] Vianca Hurtado, Thorsten Joachims, and Douglas Turnbull. Changing how we value local music using personalized recommendation. In *ACM Tapia*, 2020.

[8] Sanghoon Jun, Daehoon Kim, Mina Jeon, Seungmin Rho, and Eenjun Hwang. Social mix: automatic music recommendation and mixing scheme based on social network analysis. *The Journal of Supercomputing*, 71(6):1933–1954, 2015.

[9] Youngmoo E Kim, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull. Music emotion recognition: A state of the art review. In *ISMIR*, volume 86, pages 937–952, 2010.

[10] Peter Knees and Markus Schedl. *Music similarity and retrieval: an introduction to audio-and web-based strategies*, volume 36. Springer, 2016.

[9] https://dougturnbull.org/index.php/publications/

[10] https://open.spotify.com/playlist/4zuHH222bWJFMBaMv3TiyY?si=LmI1wkWGRDqf5xZn7ssCEA

[11] Dominik Kowald, Markus Schedl, and Elisabeth Lex. The unfairness of popularity bias in music recommendation: A reproducibility study. In *European Conference on Information Retrieval*, pages 35–42. Springer, 2020.

[12] Audrey Laplante. Improving music recommender systems: What can we learn from research on music tastes? In *ISMIR*, pages 451–456, 2014.

[13] Jin Ha Lee, Liz Pritchard, and Chris Hubbles. Can we listen to it together?: Factors influencing reception of music recommendations and post-recommendation behavior. 2019.

[14] Brian McFee, Luke Barrington, and Gert Lanckriet. Learning content similarity for music recommendation. *IEEE transactions on audio, speech, and language processing*, 20(8):2207–2218, 2012.

[15] Adrian C North and David J Hargreaves. Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition*, 14(1-2):77, 1995.

[16] J. A. Russell. A circumspect model of affect. *Journal of Psychology and Social Psychology*, 39(6):1161, 1980.

[17] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.

[18] Markus Schedl, Peter Knees, Brian McFee, Dmitry Bogdanov, and Marius Kaminskas. Music recommender systems. In *Recommender systems handbook*, pages 453–492. Springer, 2015.

[19] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7(2):95–116, 2018.

[20] Yading Song, Simon Dixon, Marcus T Pearce, and Andrea R Halpern. Perceived and induced emotion responses to popular music: Categorical and dimensional models. *Music Perception: An Interdisciplinary Journal*, 33(4):472–492, 2016.

[21] Nava Tintarev and Judith Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender systems handbook*, pages 479–510. Springer, 2011.

[22] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. In *Advances in neural information processing systems*, pages 2643–2651, 2013.

[23] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. Auralist: Introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 13–22, 2012.