Toward a Musical Sentiment (MuSe) Dataset for Affective Distant Hearing

Christopher Akiki[a], Manuel Burghardt[a]

[a] Leipzig University, Augustusplatz 10, 04109 Leipzig, Germany

Abstract

In this short paper we present work in progress that tries to leverage crowdsourced music metadata and crowdsourced affective word norms to create a comprehensive dataset of music emotions, which can be used for sentiment analyses in the music domain. We combine a mixture of different data sources to create a new dataset of 90,408 songs with their associated embeddings in Russell's model of affect, with the dimensions *valence*, *dominance* and *arousal*. In addition, we provide a Spotify ID for the songs, which can be used to add more metadata to the dataset via the Spotify API.

Keywords

music information retrieval, music emotion recognition, music sentiment

1. Introduction

The study of sentiments and emotions and more concretely their impact on human cognitive processes, for instance decision making (Schwarz, 2000), has been part of the research agenda of psychology and cognitive studies for a long time. More recently, data-driven approaches in sentiment analysis that have been developed mainly in the fields of computational linguistics and social media analytics, have been adopted by the Digital Humanities, to investigate emotional characteristics of socio-cultural artefacts. Sentiment analysis so far has been primarily used in text-based domains, such as digital history and most notably computational literary studies (Kim & Klinger, 2019), but is also gaining importance in other domains, such as computational musicology, where researchers are interested in understanding how music can affect humans emotionally.

Music and emotion Music can both convey the emotion of the artists and modulate the listeners' mood (Yang and Chen 2011). The importance of the emotional dimension of music is further made salient by the role it plays in discovering or looking up songs. Accordingly, Lamere (2008) found mood tags to be the third most prevalent descriptors of tracks on Last.fm, whereas 120 out of 427 polled people said they would use mood and emotional descriptors to look up music (Schedl, Gómez, and Urbano 2014). Music emotion recognition (MER) has therefore been gaining track as an important research direction in music information systems (MIR), both in the industry and academia. As an example for the latter, the established MIREX (Music Information Retrieval Evaluation eXchange) hosts a yearly competition where participants are invited to classify music according to mood (Downie et al. 2010).

CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands

christopher.akiki@uni-leipzig.de (C. Akiki); burghardt@informatik.uni-leipzig.de (M. Burghardt)
0000-0002-1634-5068 (C. Akiki); 0000-0003-1354-9089 (M. Burghardt)

Related work As for related work on that topic, Delbouys et al. (2018), Yang & Chen (2011) as well as Turnbull (2010) provide a concise overview of the different approaches to emotion detection and classification in music. To that end, a multitude of datasets have been put together over the years, with different machine learning approaches to try and classify them based on numerous audio and non-audio features. This multitude of datasets is the result of a lack of consensus as to what categories ought to be used, infamous copyright issues that hinder the field, and the diversity of machine learning approaches that require different sorts of data. Those studies that do not rely on small datasets of public domain music, such as Lu, Liu & Zhang (2006) and Laurier, Grivolla & Herrera (2008), either use the Million Song Dataset (MSD) (Bertin-Mahieux et al. 2011), their own in-house private collection, which is not open to the public (MIREX audio data is only available to submission algorithms), or both, like a study by Deezer researchers in Delbouys et al. (2018). It is worth noting that a number of studies use lyrics, either fully or in part, as a proxy to analyze musical sentiment (Wang et al. 2011, Parisi et al., 2019). The MTG-Jamendo dataset (Bogdanov et al., 2019) is a prominent example for a large audio-based dataset. Among other kinds of information, the dataset also contains mood annotations, which are also used in the MediaEval[1] task on emotion and theme recognition in music. However, a major drawback of the MTG-Jamendo collection is that it exclusively consists of royalty free music.

Focus of this work While we see the value of a lot of these existing approaches, we try and adopt a method that does not rely on the inherent assumptions made by the respective methods. Using already compiled datasets, like the MSD, forces us to deal with whatever bias the original gathering method forces on the mood distribution of the ensuing dataset. Indeed, some have noted a bias toward positively charged music when collecting the social tags (Cano & Morisio, 2017). With this in mind, we think it is more important to focus on the collection of a new dataset, rather on the methods used to classify it. Such a dataset can pave the way for sentiment analyses in the Digital Humanities that goes well beyond text-based media. Combined with other metadata, such as genre, chart placement or gender of the artist(s), a variety of research questions come to mind that may be investigated in an empirical way and enable some kind of "distant hearing".

In this paper we present the MuSe (*Music Sentiment*) dataset, a collection of 90,408 songs, sentiment information and various metadata. We use the Allmusic mood taxonomy of music, the creation of which involves human experts, to seed a scraping of social tags-based Last.fm data which we then enrich with all available tags[2]. We then filter the additional tags based on a corpus of affect-related terms. We also provide the Spotify ID, which can be used to obtain all kinds of metadata via the Spotify API.

2. Modeling emotions in music

Conceptualizing the understanding of emotions in general and music in particular is not a new problem. Different approaches to the subject involve different representations. Some are based on a categorical approach that considers emotions as categories or classes which one experiences separately (Yang and Chen 2011), or as a cluster of discrete categories (Downie et al. 2010).

[1] http://www.multimediaeval.org/mediaeval2019/music/

[2] A similar approach was used by Delbouys et al. (2018), who used a Last.fm-based filtering approach on the MDS to create a corpus of about 18k songs with affective information.

Others use a dimensional approach, which can be achieved by asking subjects to rate music-related concepts on a continuous scale for several dimensions (Warriner, Kuperman & Brysbaert 2013). While a dimensional approach may include various factors of scale, such as *tension-energy*, *gaiety-gloom*, *solemnity-triviality*, *intensity-softness*, *pleasantness-unpleasantness*, etc. (Yang and Chen 2011), Scherer (2004) shows that these various factors can mostly be reduced to three fundamental dimensions of emotion:

1. *valence* (pleasantness and positivity)
2. *arousal* (energy and stimulation)
3. *dominance* (potency and control)

This approach can be traced back to Russell's seminal "circumplex model" (Russell 1980), a two-dimensional vectorial space with *valence* on the abscissa and *arousal* on the ordinate. This simple yet powerful model provides a way to embed emotions into a space where one dimension represents physiology (arousal), and the other represents affect (valence). This approach also allows for a straightforward comparison of different emotions within the space in terms of Euclidean distance. In this study, we also opt for Russell's circumplex model and extend it by the third dimension of *dominance*, as suggested by Scherer (2014).

3. Methodology: Creating the dataset

This section describes the different stages of creating a comprehensive dataset that contains songs along with values for the dimensions valence, arousal and dominance.

Seed stage: Collecting mood labels from Allmusic.com We start building our dataset by collecting mood labels that are available from Allmusic.com[3]. There is a total of 279 labels that were created manually by Allmusic.com editors. These will serve as the starting point of our analysis and provide an overarching set of mood descriptors which we will use to guide our collecting of further tags. If we had let other factors guide our collection efforts, we might have ended up with emotional biases where certain moods are over-represented. This approach will ensure that we collect a balanced dataset that encompasses a large variety of moods.

Expansion stage: Collecting user-generated tags from Last.fm Now that we have 279 terms which can be used to address different categories of musical emotions, we can use each of these terms to collect song objects from the Last.fm API. For every one of the 279 seed moods, we collect a maximum of 1,000 songs, which currently is the official limit of the Last.fm API.

For every single song, we then collect the top-100 tags[4] assigned to it by the users of Last.fm[5] as well as other metadata such as artist name, album name and number of listeners. Using this approach, we collected a total of 131,883 songs. We did not collect the theoretical amount of $279 * 1,000 = 279,000$ songs, because we could not find 1,000 songs on Last.fm for each of the 279 seed labels. The total number is also reduced, as some songs were collected for multiple tags (but obviously kept only once in the dataset). After removing duplicate songs, the corpus

[3] https://www.allmusic.com/moods; It is worth noting that the MIREX emotion categories were originally also derived by clustering the co-ocurrence of moods in songs from the Allmusic.com moods taxonomy.

[4] https://www.last.fm/api/show/track.getTopTags

[5] For some songs, the initial seed mood tags were not among the Last.fm top-100 tags. As we did not want to lose these seed moods, we included them for every song, regardless of their ranking on Last.fm.

contains a total of 96,499 songs. These songs altogether have more than 3 million tags, from which about 261k tags are unique. At this stage, songs in our collection have a mean of 33.83 Last.fm tags (std: 31.01; min: 1, max: 100). The most frequent tags are typically genre-related tags, such as the top-3 tags *rock* (29,810 times), *alternative* (24,763 times) or *indie* (23,006). However, there are also affective tags among the highly frequent tags, for instance *chill* (11,841 times), *sad* (8,350 times) and *melancholy* (8,344 times).

Filtering stage: Identifying mood tags with WordNet-Affect In order to detect which of the collected tags correspond to mood categories, we apply the pre-processing step described by Hu, Downie & Ehmann (2009) and employed by Delbouys et al. (2018), which is to compare the list of tags of a song to the lemmas of WordNet-Affect (Strapparava & Valitutti 2004). WordNet-Affect is an extension of WordNet, an English lexical database which itself provides sets of synonyms for English words. WordNet-Affect restricts those words to "affective concepts correlated with affective words" (Strapparava & Valitutti 2004). Hu, Downie & Ehmann use a categorical approach and therefore use the WordNet-Affect categories themselves to further cluster the mood terms into several groups. Our study takes on a continuous approach and aims to embed affective words in a continuous Euclidean space, similarly to Delbouys et al.

From the 261k unique tags in our corpus, only 873 tags (for some example tags and their frequencies see Table 1) could be matched to the WordNet-Affect list, which altogether contains a collection of 1,606 affective words[6].

Mapping stage: Embedding mood labels into Russell space Having a way to filter out tags based on whether or not they refer to mood leaves us with the task of estimating the valence, arousal and dominance of each one of these tags. Warriner, Kuperman & Brysbaert (2013) mention the hegemony of the ANEW norms dataset (Bradley et al. 1999) for assigning emotive values to words for many fields of study. The ANEW dataset is one of 1,034 words and conceived for small-scale experiments. Warriner et al. (2013) recognize the promise of crowdsourcing a newer and more complete word norms dataset than the ANEW collection. To that end, they recruited native English speakers through Amazon Mechanical Turk who provided 1,085,998 affective ratings of 13,915 lemmas across all three dimensions.

We make use of the Warriner et al. (2013) wordlist – which we will refer to as the V-A-D list – to map the mood labels of our song collection to the dimensions valence, arousal and dominance. Each of the mood tags from each song is looked up in the V-A-D list and is assigned a 3-dimensional coordinates triple for a word. If the word is not present, it will return a value of 0 for each dimension. From the overall 1,827 unique mood labels in our dataset, only 765 (41.9%) are matched with the lemmas in the V-A-D list. Taking a closer look at the mood tags collected from Last.fm and the V-A-D list, it becomes clear that a basic lemmatization procedure might further improve the match[7]. We will look into this optimization as we will further develop the dataset. For now we were content to observe that, although many of the Last.fm tags did not match with the V-A-D lemmas, these were typically tags with very low frequencies. In total, 88.1% of the mood tags used to characterize the songs in our dataset are actually matched with the V-A-D list, i.e. Last.fm tags that occur frequently are typically already in lemmatized form and thus have a higher chance of matching the V-A-D list.

[6]As we believe our initial seed moods collected from Allmusic.com are important affect words in any case, we add those that are not already included on top of the WordNet-Affect list, resulting in a total of 1,827 affect tags.

[7]Some examples: fevered (Last.fm)/fever (V-A-D), horrify/horrific, sickish/sick, peacefulness/peaceful

Table 1
Snippet from the top-10 and the bottom-10 mood tags and their frequencies.

rank	tag	frequency	rank	tag	frequency
1	love	14,363	...	...	...
2	mellow	13,825	859	worrying	1
3	chill	11,841	860	urge	1
4	sad	8,351	861	estranged	1
5	melancholy	8,344	862	lovesome	1
6	sexy	7,926	863	disgust	1
7	atmospheric	7,641	864	dismay	1
8	dark	7,560	865	pose	1
9	cool	6,888	866	aversion	1
10	epic	6,811	867	quiver	1
...	...	...	868	stew	1

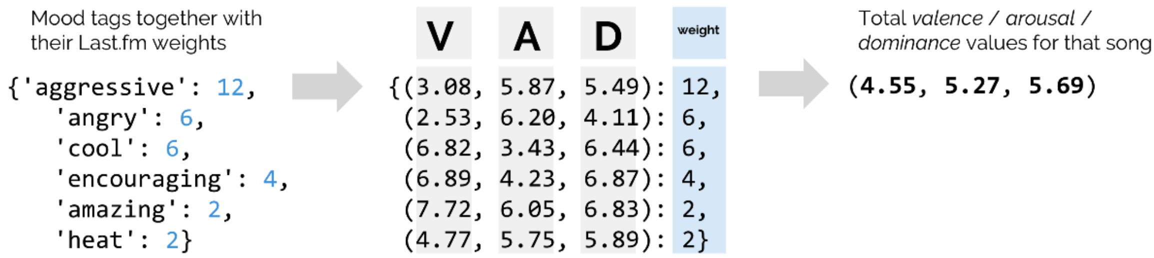

Figure 1: Example weighted average calculation for valence, arousal and dominance of the song "Till I Collapse", by Eminem.

From the overall 96,499 songs, 90,408 (93,7%) are matched with a least one tag that is also present in the V-A-D list, with a mean of 3.36 mood tags (std: 3.71; min: 1, max: 42). As songs typically will have multiple mood tags, we calculate the weighted average for every dimension of every word separately, whereas each word is weighted according to scores we retrieved for each tag via the Last.fm API (see Figure 1 for an example). Higher Last.fm weights indicate a higher relevance of a tag.

Metadata stage: Adding further metadata via the Spotify API Having collected all the affective metadata we now add the Spotify ID whenever we were able to retrieve one. In order to find the appropriate ID for each of the songs, we performed an API lookup for a track based on its title and artist. Some of the problems we encountered here include untranslated artist names for languages such as Japanese or Korean, as well as search queries that are too wordy. In the end, we were able to track down a Spotify ID for a total of 61,484 songs. This allows us to append any additional information to our dataset that is available via the Spotify API[8], for instance:

- *metadata*: release date, popularity, available markets, explicit lyrics (boolean), etc.

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

Table 2
Snippet of the final dataset (not all of the rows and columns are displayed).

Title	*Artist*	*Spotify ID*	...	*Emotion*	*Valence*	*Arousal*	*Dominance*
Moop Bears	Momus	000PUfi7X3otImjyjJXvFS	...	[naive]	4.22	3.86	5.00
Await The King's Justice	Ramin Djawadi	001VMKfkHZrlyj7JIQbQFL	...	[dramatic, joy, score]	6.63	6.16	6.36
Et Tu	Siddharta	002SF61pDJexw3oSRcUgnE	...	[monumental, easy]	6.61	4.53	6.43
Magnolia	The Hush Sound	004VU4cWTkRqVMrlv8KW3D	...	[introspective, optimism, love, cool]	5.51	3.44	4.82
The Angels	Melissa Etheridge	004ddQGTS8w7sDEKuwXZhi	...	[earthy]	6.50	3.29	6.56
Eternity	Robbie Williams	7zw9OxtVotLfxlfavSADXQ	...	[bittersweet, love, sad, romantic, melancholy,...]	4.99	4.18	4.78
Cupid De Locke	The Smashing Pumpkins	7zwwvrJAWGjfc9wFD3bVzZ	...	[dreamy, dreamy, love, melancholic, melancholy,...]	4.92	4.09	4.83
Breaking News - Reflections	Anuj Rastogi	7zxG55mLhPAx2EIMj6loEg	...	[philosophical]	6.21	3.50	5.39
I Love A Man In A Uniform	Gang of Four	7zy6jG8RIUI8qNYYVuLGbY	...	[cynical, cynical]	3.30	4.78	4.95
Mosquito Brain Surgery	Spastic Ink	7zzbwY4h8Q6kI1ZX3gB1B3	...	[complex]	4.21	4.60	5.25
...	...	...	...	...	...	...	...

- *low-level audio features*: bars, beats, sections, duration, etc.
- *mid-level audio features*: acousticness, danceability, tempo, energy, valence[9], etc.

The final song dataset (for a snippet see Table 2) contains basic metadata, such as artist, title and genre, as well as the mood tags, the three affective dimensions as well as the Spotify ID that allows us to further extend the dataset with additional information via the Spotify API. Figures 2 and 3 plot the overall distribution of different artists and genres in the dataset.

Figure 4 shows a Pearson correlation matrix for the three affective dimensions and various Spotify parameters. While the matrix reveals a strong correlation between Spotify features such as loudness and energy (0.78) or acousticness and energy (-0.74), there seems to be no positive or negative correlation between valence / arousal / dominance and other acoustic features. However, this might well be different, if we calculate correlations for different genres. Another interesting insight is that Spotify's valence value – which very generally describes a song's acoustic sentiment – is only very weakly correlated with our valence / arousal / dominance dimensions, indicating that sentiment based on user generated tags is different to sentiment that is exclusively derived from audio features. Another interesting observation in the correlation plot is the strong correlation of the valence and dominance values (0.87) we calculated. This suggests that the emotion data might be plotted also in a 2-dimensional way, without loosing to much information on the general distribution of songs. Figure 5 shows an example that plots all songs in our collection on the axes arousal and valence, indicating a slightly positive trend for most of the songs, which aligns with the notion of a bias toward

[9]Note: Energy and valence obviously point toward a similar direction as our calculated arousal and valence dimensions. We want to stress that Spotify uses audio features exlusively while we make use of user-generated tags, which might be considered a more holistic description of a song.

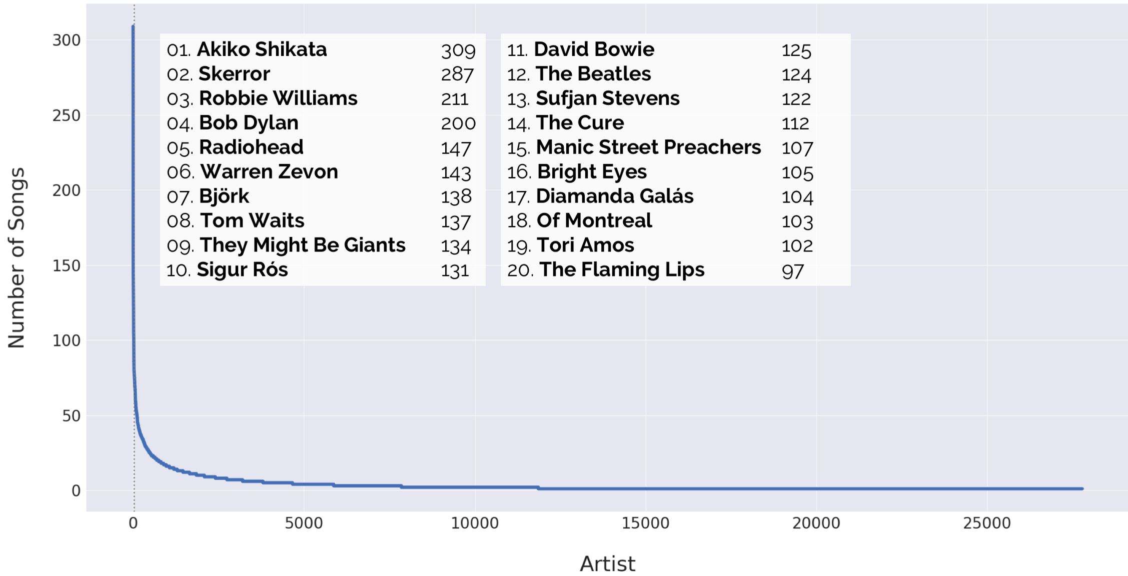

Figure 2: All of the 27,782 unique artists in the dataset are plotted by their individual number of songs. The dotted line highlights the top 20 artists, for which we provide the exact frequencies.

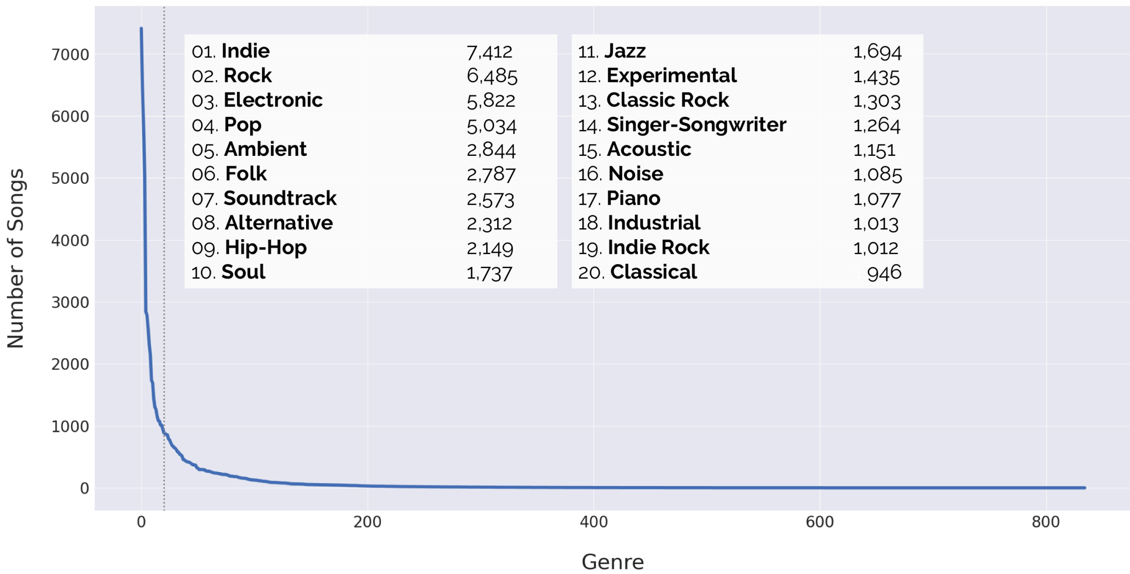

Figure 3: All of the 835 unique genres in the dataset are plotted by their individual number of songs. The dotted line highlights the top 20 genres, for which we provide the exact frequencies.

positively charged music by Cano & Morisio (2017).

4. Conclusion: Toward affective distant hearing with the MuSe dataset

Moretti's (2000) notion of distant reading – as an antipode to the more traditional close reading – by now has become a dictum in the Digital Humanities, summarizing any kind of

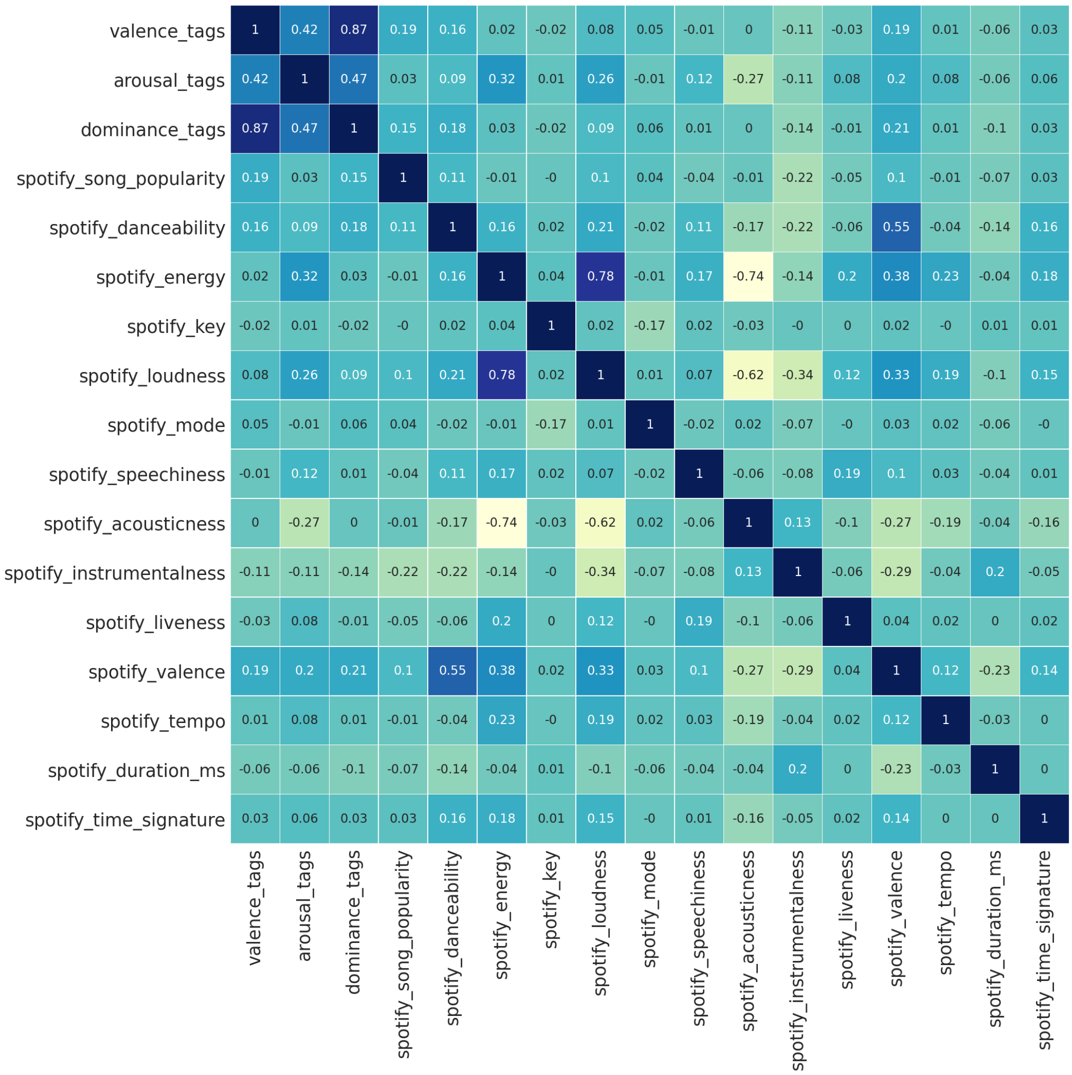

Figure 4: Correlation matrix for the V-A-D dimensions and various Spotify parameters.

computational, empirical approach to analyzing text. Tilton & Arnold (2019) even adopted the concept for the field of image and video analysis, proclaiming a distant viewing paradigm. Along these lines, it feels natural to coin *distant hearing* as a corresponding analogy for the field of music.

In this paper we have described the reasonings and practical steps for the creation of a music sentiment dataset that is not solely based on an analysis of lyrics or audio features, but rather takes into account actual human judgement of a song's emotional characteristics by mining user-generated mood tags from the social music platform Last.fm. With our current MuSe dataset we provide a resource that enables different kinds of research questions that may be subsumed as *affective distant hearing*. These research questions may extend existing studies, such as Elvers' (2018) sentiment analysis of European chart music or Argstatter's (2015) study of emotions in music across-cultures. Further empirical research questions that come to mind might investigate the relation of music emotion and genres, gender (either of the artist or the

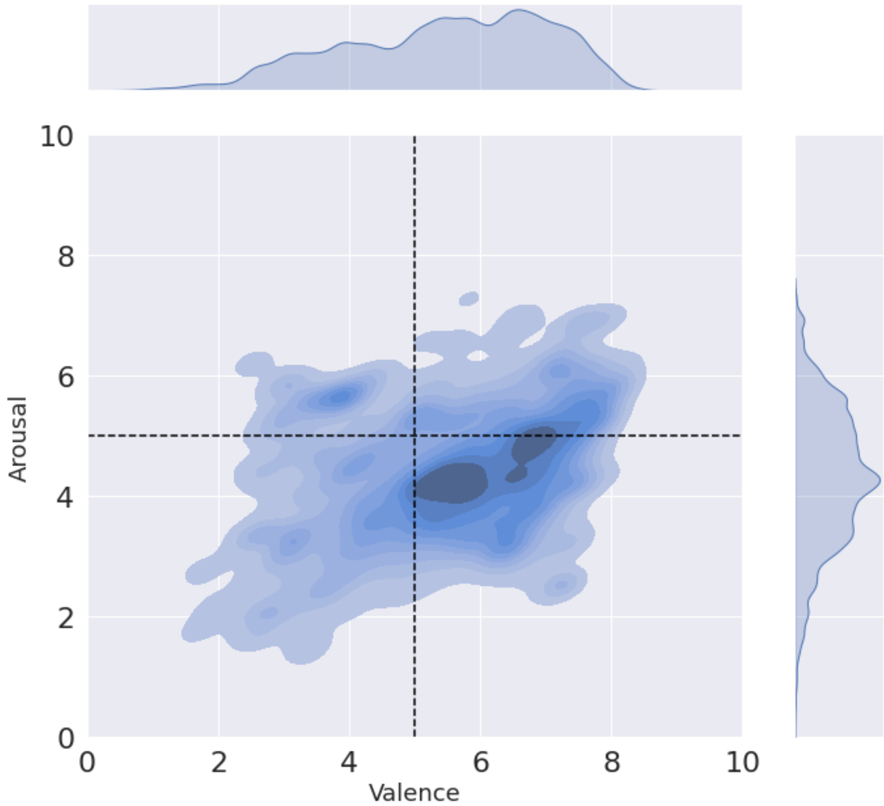

Figure 5: 90,408 songs embedded into 2D Russel space for the dimensions arousal and valence.

predominant audience), chart placement, artist collaborations or audio features available via the Spotify API, e.g. *danceability* or *acousticness*.

The current dataset is available upon request. Although it can already be used to empirically investigate research questions in music emotions, we consider this to be work in progress and plan to further enhance the dataset in the near future. More concretely, we think about extending the scope of songs we collect from Last.fm, by enhancing the number of seed tags by the 873 mood tags we identified in our current dataset. We also plan to enhance the match between Last.fm tags and the V-A-D list by experimenting with lemmatization procedures. In addition, we plan to integrate further metadata, for instance from Discogs[10] and MusicBrainz[11], into the dataset. The latter seems particularly interesting, as Last.fm readily provides a MusicBrainz ID for many songs. All in all, we hope the MuSe dataset will help to advance the field of computational musicology and thus provide an incentive for more quantitative studies on the role of emotions in music.

[10]https://www.discogs.com/developers
[11]https://musicbrainz.org/doc/Developer_Resources

References

[1] H. Argstatter. "Perception of basic emotions in music: Culture-specific or multicultural?" In: *Psychology of Music* 44 (June 2015), pp. 1–17.

[2] T. Arnold and L. Tilton. "Distant Viewing: Analysing Large Visual Corpora". In: *Digital Scholarship in the Humanities* (Mar. 2019), pp. 3–16. ISSN: 2055-7671.

[3] J. P. Bello and J. Pickens. "A Robust Mid-Level Representation for Harmonic Content in Music Signals". In: *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), London, UK*. 2005, pp. 304–311.

[4] T. Bertin-Mahieux et al. "The million song dataset". In: *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), Miami, Florida*. 2011, pp. 591–596.

[5] D. Bogdanov et al. "The MTG-Jamendo dataset for automatic music tagging". In: *Machine Learning for Music Discovery Workshop, International Conference on Machine Learning (ICML 2019)*. 2019, pp. 1–3.

[6] M. M. Bradley and P. J. Lang. *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Tech. rep. Center for Research in Psychophysiology, University of Florida, 1999.

[7] E. Çano and M. Morisio. "Music mood dataset creation based on last.fm tags". In: *International Conference on Artificial Intelligence and Applications, Vienna, Austria*. 2017, pp. 15–26.

[8] R. Delbouys et al. "Music Mood Detection Based on Audio and Lyrics with Deep Neural Net". In: *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR), Paris, France*. 2018, pp. 370–375.

[9] J. S. Downie et al. "The Music Information Retrieval Evaluation eXchange: Some Observations and Insights". In: *Advances in Music Information Retrieval*. Vol. 274. Studies in Computational Intelligence. Springer, 2010, pp. 93–115.

[10] P. Elvers. *Sentiment analysis of musical taste: a cross-European comparison*. 2018. URL: https://paulelvers.com/post/emotionsineuropeanmusic/.

[11] X. Hu, J. S. Downie, and A. F. Ehmann. "Lyric Text Mining in Music Mood Classification". In: *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan*. 2009, pp. 411–416.

[12] P. N. Juslin and J. A. Sloboda. *Music and emotion: Theory and research*. Oxford University Press, 2001.

[13] E. Kim and R. Klinger. "A Survey on Sentiment and Emotion Analysis for Computational Literary Studies". In: *CoRR* (2018).

[14] Y. E. Kim et al. "State of the Art Report: Music Emotion Recognition: A State of the Art Review". In: *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), Utrecht, Netherlands*. 2010, pp. 255–266.

[15] P. Lamere. "Social Tagging and Music Information Retrieval". In: *Journal of New Music Research* 37.2 (2008), pp. 101–114.

[16] C. Laurier, J. Grivolla, and P. Herrera. "Multimodal Music Mood Classification Using Audio and Lyrics". In: *Seventh International Conference on Machine Learning and Applications, (ICMLA), San Diego, California, USA*. IEEE Computer Society, 2008, pp. 688–693.

[17] L. Lu, D. Liu, and H. Zhang. "Automatic mood detection and tracking of music audio signals". In: *IEEE Trans. Speech Audio Process.* 14.1 (2006), pp. 5–18.

[18] R. Malheiro et al. "Music Emotion Recognition from Lyrics: A Comparative Study". In: *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD), Prague, Czech Republic.* 2013, pp. 47–50.

[19] F. Moretti. "Conjectures on World Literature". In: *New Left Review* 1 (Jan. 2000), pp. 54–68.

[20] L. Parisi et al. "Exploiting Synchronized Lyrics And Vocal Features For Music Emotion Detection". In: *arXiv preprint arXiv:1901.04831* (2019).

[21] J. Russell. "Core Affect and the Psychological Construction of Emotion". In: *Psychological Review* 110 (Feb. 2003), pp. 145–172.

[22] J. A. Russell. "A Circumplex model of affect". In: *Journal of Personality and Social Psychology* 39.6 (1980), pp. 1161–1178.

[23] M. Schedl, E. Gómez, and J. Urbano. "Music Information Retrieval: Recent Developments and Applications". In: *Found. Trends Inf. Retr.* 8.2-3 (2014), pp. 127–261.

[24] K. R. Scherer. "Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?" In: *Journal of New Music Research* 33.3 (2004), pp. 239–251.

[25] N. Schwarz. "Emotion, cognition, and decision making". In: *Cognition & Emotion* 14.4 (2000), pp. 433–440.

[26] C. Strapparava and A. Valitutti. "WordNet Affect: An Affective Extension of WordNet". In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC), Lisbon, Portugal.* European Language Resources Association, 2004, pp. 1083–1086.

[27] X. Wang et al. "Music Emotion Classification of Chinese Songs based on Lyrics Using TF*IDF and Rhyme". In: *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR), Miami, Florida, USA*. 2011, pp. 765–770.