

## Research Reports on Thematic Focus

# “Evacuate the Dancefloor”: Exploring and Classifying Spotify Music Listening Before and During the COVID-19 Pandemic in DACH Countries

„Evacuate the Dancefloor“: Exploration und Klassifizierung von Spotify-Hörverhalten vor und während der COVID-19-Pandemie in den DACH-Ländern

Kework K. Kalustian\*<sup>1</sup>, Nicolas Ruth<sup>2</sup>

[1] Department of Music, Max Planck Institute for Empirical Aesthetics, Frankfurt/Main, Germany. [2] Department of Psychology, Goldsmiths, University of London, London, United Kingdom.

## Abstract

Many people used musical media via music streaming service providers to cope with the limitations of the COVID-19 pandemic. Accounting for such behavior from the perspective of uses-and-gratifications theory and situated cognition yields reliable explanations regarding people's active and goal-oriented use of musical media. We accessed Spotify's daily top 200 charts and their audio features from the DACH countries for the period during the first lockdown in 2020 and a comparable non-pandemic period situation in 2019 to support those theoretical explanations quantitatively with open data. After exploratory data analyses, applying a k-means clustering algorithm across the DACH countries allowed us to reduce the dimensionality of selected audio features. Following these clustering results, we discuss how these clusters are explainable using the arousal-valence-circumplex model and possibly be understood as (gratification) potentials that listeners can interact with to modulate their moods and thus emotionally cope with the stress of the pandemic. Then, we modeled a cross-validated binary SVM classifier to classify the two periods based on the extracted clusters and the remaining manifest variables (e.g., chart position) as input variables. The final test scenario of the classification task yielded high overall accuracy in classifying the periods as distinguishable classes. We conclude that these demonstrated approaches are generally suitable to classify the two periods based on the extracted mood clusters and the other input variables, and furthermore to interpret, by considering the model-related caveats, everyday music listening via those proxy variables as an emotion-focused coping strategy during the COVID-19 pandemic in DACH countries.

*Keywords:* API, COVID-19, interpretable machine learning, k-means clustering, popular music, SVM classifier, streaming behavior

## Zusammenfassung

Während der COVID-19-Pandemie bewältigten viele Menschen diese einschränkende Situation, indem sie Musikmedien über verschiedene Musikstreaming-Dienstleister nutzten. Eine Betrachtung dieses Verhaltens aus der Perspektive des Uses-and-Gratification-Ansatzes und der Situated Cognition liefert verlässliche Erklärungen für die aktive und zielgerichtete Nutzung von Musikmedien. Um solche Erklärungen quantitativ mit frei verfügbaren Daten zu untermauern, haben wir die täglichen Top-200-Charts von Spotify in den DACH-Ländern und die dazugehörigen Audio-Features gemäß den von Spotify bereitgestellten Audio-Feature-Informationen für die Zeiträume während des ersten Lockdowns im Jahr 2020 und dem gleichen Zeitraum ohne Pandemie in 2019 abgerufen. Nachdem wir explorative Datenanalysen durchgeführt haben, reduzierten wir die Dimensionalität ausgewählter Audio-Features durch Anwendung eines k-Means-Clustering-Algorithmus über die DACH-Länder hinweg. Basierend auf den Clustering-Ergebnissen diskutieren wir, wie diese Cluster nicht nur konzeptuell mit dem Arousal-Valence-Circumplex-Modell erklärbar sind, sondern auch womöglich als (Gratifikations-)Potenziale verstanden werden können, mit denen die Hörer\*innen interagieren können, um ihre Stimmungen zu modulieren und somit emotional mit dem Stress der Pandemie umzugehen. Darüber hinaus modellierten wir einen kreuzvalidierten binären SVM-Klassifikator, um die Zeiträume basierend auf den extrahierten Clustern und den verbleibenden manifesten Variablen (z.B. Chartpositionen) als Input-Variablen zu klassifizieren. Das endgültige Testscenario der Klassifizierungsaufgabe ergab eine hohe Gesamtgenauigkeit bei der Klassifizierung der Zeiträume als unterscheidbare Klassen. Abschließend kommen wir somit zu dem Schluss, dass diese aufgezeigten Ansätze grundsätzlich geeignet sind, die beiden Zeiträume auf Basis der extrahierten Mood-Cluster und der anderen Input-Variablen zu bestimmen und außerdem das alltägliche Musikhören über diese stellvertretenden Variablen als emotionsfokussierte Bewältigungsstrategie während der COVID-19-Pandemie in den DACH-Ländern unter modellbedingten Einschränkungen zu interpretieren.

*Schlüsselwörter:* API, COVID-19, interpretierbares maschinelles Lernen, k-Means Clustering, populäre Musik, SVM-Klassifikator, Streaming-Hörverhalten

Jahrbuch Musikpsychologie, 2021, Vol. 30: Musikpsychologie – Empirische Forschungen - Ästhetische Experimente, Artikel e95,  
<https://doi.org/10.5964/jbdgm.95>

Eingereicht: 2021-02-18. Akzeptiert: 2021-06-28. Publiziert (VoR): 2021-09-24.

Begutachtet von: Jochen Steffens; Holger Schwetter.

\*Korrespondenzanschrift: Department of Music, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322, Frankfurt am Main, Germany.  
E-Mail: [kework.kalustian@ae.mpg.de](mailto:kework.kalustian@ae.mpg.de)



Dieser Open-Access-Artikel steht unter den Bedingungen einer Creative Commons Namensnennung 4.0 International Lizenz, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.de>). Diese erlaubt für beliebige Zwecke (auch kommerzielle) den Artikel zu verbreiten, in jedwedem Medium zu vervielfältigen, Abwandlungen und Bearbeitungen anzufertigen, unter der Voraussetzung, dass der Originalartikel angemessen zitiert wird.

The global COVID-19 pandemic and its related lockdowns, curfews, and increased preventative measures dramatically changed everyday life. Many European countries implemented the same or similar measures to stop the virus' spread. For instance, the three major German-speaking countries (i.e., Austria, Germany, and Switzerland) enacted lockdowns, curfews, and travel restrictions between March and June 2020. These and other similar measures around the world changed everyday life, social and leisure activities—such as listening to and making music (Lades et al., 2020). In this vein, Spotify—as one of the biggest music streaming service providers (Statista, 2020)—has released mood-specific and topical playlists addressing this critical situation so that listeners can foster their mental well-being (King, 2020). Building on these general observations, the pandemic's influence on many facets of everyday life offers a rare opportunity to investigate music-listening behavior in times of a ubiquitous crisis and uncertainty. Accordingly, different research areas have scrutinized the pandemic's influence on everyday life in general and on music listening and making in particular: Research groups, such as the multi-national research network Musicovid<sup>1</sup>, or conferences, such as the UK Society for Education, Music and Psychology Research conference, featured panels dedicated to research on the topic. Furthermore, special editions of scientific journals like *Frontiers in Psychology* or the current *Yearbook* were planned to promote music and COVID-19 research. Despite the growing body of research on this topic, many questions remain unanswered.

Prior to answering any open questions, however, existing research on how the pandemic impacted music and how music helped affected people during the pandemic should first be reviewed starting with literature that describes how the pandemic influenced everyday life in general.

## The Pandemic's Impact on Everyday Life

Internet traffic increased by up to 15–20% in the first weeks of the pandemic, since many people started to work remotely or restrictions forced them to stay at home (Feldmann et al., 2020). Leisure activities as well as work processes such as meetings moved online, and Seetharaman (2020) predicts that, from an economic perspective, the pandemic will change individual jobs and work as well as many companies' business models. In addition to their work, people's mental health and well-being were expected to be affected most during the crisis (Bäuerle et al., 2020). In their study, Bäuerle and colleagues (2020) found that Germans experienced more anxiety, depression, and distress during the pandemic than usual. Nevertheless, Germans seem to have

1) See <https://www.aesthetics.mpg.de/forschung/abteilung-musik/musicovid-an-international-research-network.html>

coped better than expected with the situation (Entringer & Kröger, 2020). Still, a survey indicates that people spent a lot of time (4.45 hours per day on average) thinking about the pandemic and its effect (Petzold et al., 2020).

While many people were able to adapt their work and private lives to the situation, the pandemic had a strong impact on individuals working in the cultural industry. With no cultural events, many of them lost their jobs and income. The live music sector was especially affected since most concerts scheduled in 2020 were either cancelled or postponed. Some musicians found ways to re-engage in musical training (Antonini Philippe et al., 2020) or organized online concerts like Global Citizen's and Lady Gaga's TV concert, while fans also found ways to socially engage in music events online (Palamar & Acosta, 2021). Some scholars and journalists even believe that the pandemic was in some ways a catalyst for a change in the cultural industries and their relationship with the Internet (Lee et al., 2020).

## Effects of Music During the Pandemic

The pandemic influenced how music was released, performed, and used to communicate. Lehman (2020) shows how song lyrics were used to communicate certain rules (like how to social distance and wash your hands), for example. Not only did the pandemic affect music, but music influenced people during the pandemic, too, and not only by telling them how to properly apply certain hygiene concepts.

Music therapeutical interventions were used to successfully reduce stress and improve well-being in medical staff (Giordano et al., 2020). Another study showed how music should be used in a therapeutic way to improve the well-being of adults and children (Mastnak, 2020). It was argued that music helps to cope with a situation like the pandemic and could be applied in personal settings. Additionally, music offered a distraction and an opportunity to bond socially and engage with others and community events. Virtual music concerts were described as a sort of collective therapeutic hotline that helped people cope with social distancing and the lack of cultural events (Vandenberg et al., 2020). This indicates that individuals tend to turn to music for specific reasons and may adapt their music listening behaviors to an extraordinary situation.

## COVID-19 Pandemic and Music Listening

Research on the use of music streaming platforms shows how people use these providers to fulfill certain needs or for certain reasons and that service developers try to attend to these needs by incorporating social and creative functions into their software (e.g., Hampton-Sosa, 2017; Henning & Ruth, 2020). These findings may lead one to believe that music consumption increased during the pandemic. However, a recent study finds that consumption indeed decreased (Sim et al., 2020). The researchers interpret that the number of streaming users declined since travel and commuting was largely on pause during the lockdown. People may have also listened to CDs, vinyl records, radio, or other formats as they worked from home.

Comprehensive research explores how music is used in certain contexts like parties, social situations, commuting, sports (Greb et al., 2018; Rentfrow & Gosling, 2003; Skanland, 2011; Sloboda, 2010; van Goethem & Sloboda, 2011), or during daily routines (Greasley & Lamont, 2011; Lonsdale & North, 2011; North et al., 2004; Sloboda, 2010). Many of these situations changed during the pandemic, which arguably impacted music-listening behavior.

It is not unusual that external factors influence the overall music listening of certain populations. An analysis of the US billboard charts from 1955 to 2003, for example, explains why people listen to more meaningful music with social connotations during politically and socially threatening time-periods (Pettijohn & Sacco, 2009).

The selection and use of media, especially music, for satisfying certain needs and achieving goals is well documented and can be explained by different theoretical approaches.

## Theoretical Backgrounds

At least three intertwined approaches can describe music listeners' active and/or goal-orientated selection behavior toward musical media.

The first approach specifically addresses music listeners' selective exposure routines by assuming that the uses-and-gratifications approach can explain decision-making processes (Katz et al., 1973). This theory posits that individuals have personal and social needs, which are characterized by their personality and their knowledge of how to satisfy these needs through music or media (e.g., Delsing et al., 2008). Thus, people most frequently choose music they know or prefer to gain gratification from listening to it. Gratification from listening to music can be summarized into five categories, namely: surveillance, diversion, personal relationships, personal identity, and mood management (Lonsdale & North, 2011). Additionally, in certain situations people tend to listen to music that may help them cope with their need to feel nostalgic or secure, especially in times of crises (Pettijohn & Sacco, 2009; also Yeung, 2020). In this vein, it is unsurprising that streaming service providers, e.g., Spotify, take advantage of these needs and use pop music to compile playlists that intend to fulfill certain needs, such as enhancing well-being, helping to relax, and coping with stress, while users create quarantine playlists focused on song lyrics that seemed funny in the pandemic situation (like the eponymous song of this article). From a uses-and-gratifications perspective, the former playlists seem sensible in the idea of helping people cope, divert, and manage their mood.

Zillmann's (1988) mood management theory is a related conceptual model. It helps to explain how people choose media and music based on their current and targeted mood (e.g., Zillmann, 2000). For example, if a listener feels sad and wants to be happy, they will choose that kind of music that appears happy for them to reach their target mood (or vice versa according to the underlying dispositions in terms of compensation or cathartic principles, see Schramm, 2005). According to this theory, media or music can be used to enhance, decrease or to maintain a certain mood (Zillmann, 1988). In a listening context, people tend to use music more frequently to acquire or keep a positive mood than to reach or stay in a negative mood (Schramm, 2005). Listening to music to achieve a positive mood was most frequently associated with coping with negative affective states and low emotional health (Randall & Rickard, 2017).

The third concept that accounts for listeners' active role in choosing musical media to make sense of their surrounding environment is situated cognition (see e.g., Newen et al., 2018; Schiavio et al., 2017). One overarching branch within the music-related debates of situated cognition encompasses the enactive aspects of agents (e.g., listeners and/or musicians) when they interact with their musical environment to foster their well-being, either tacitly or explicitly. Although studies are often interested in how people can make sense of their environment (e.g., by humming, listening, dancing or simply *musicking*, see Small, 1998, or van der Schyff et al., 2018), enactivist approaches also value the felt or perceived emotionality as an essential part of musical cognition (see e.g., Krueger, 2009). Accordingly, this perspective can be used to interpret any music-related

coping strategy, such as the uses-and-gratification and/or mood-management approach, as a goal-oriented way why people listen to music and, foremost, how they bring forth their music-related gratifications while interacting with their surrounding musical environment.

## Research Question

If music is used for coping, mood regulation, and to fulfill certain needs during the pandemic, this behavior should be reflected in the data on streamed music during the pandemic in comparison to data from the same period in 2019. Therefore, we aim to answer the following research question:

RQ: To what extent can we estimate and classify the listening behavior during the pandemic and a comparable reference period based on Spotify's provided audio features for each track by taking the mood-related features particularly into account?

To answer the research question systematically, we examine the following hypotheses:

H<sub>1</sub>: The dimensionality of Spotify's mood-related audio features can be reduced to fewer clusters so that potential differences ( $r \geq .10$ ) can be observed in the stream counts of these clusters between the pre-pandemic and the pandemic period for each and across all DACH countries.

H<sub>2</sub>: The mood-related clusters and the remaining audio features can successfully be implemented in a classification task that aims to classify both periods in an interpretable way so that a high overall accuracy can be achieved ( $ACC \geq 90\%$ ).

## Method

To answer the research question by examining the hypotheses H<sub>1</sub> and H<sub>2</sub>, we automated data collection and then used exploratory data analysis, null hypothesis significance testing and (un)supervised machine learning techniques, and methods from the scope of interpretable machine learning to analyze the data.

## Data Retrieval

As Spotify is the leading music streaming service provider in many countries, including the German-speaking countries of Europe (e.g., [Statista, 2020](#)), it appears reasonable to examine the music streamed from it. Additionally, Spotify offers an application programming interface (API) that enables developers to retrieve metadata for every song on the platform. The DACH countries (i.e., Germany, Austria, Switzerland) were chosen because their charts feature comparable songs since both English- and German-speaking artists are featured. Moreover, leading music publishing companies, such as Sony Music, Universal Music, and Spotify itself consider these three countries as one target audience ([Sony Music, 2021](#); [Spotify, 2021b](#); [Universal Music, 2021](#)). By using web scraping techniques, we retrieved Spotify's daily top 200 charts from DACH countries as well as the respective song IDs. Audio features were collected by using Spotify's API and the R package "Spotifyr" ([Thompson et al., 2021](#)) on the previously retrieved song IDs. The period between March 11 (WHO declares COVID-19 a pandemic) and June 14, 2020, (travel restrictions lifted in the DACH countries) lasts 96 days with 200 daily chart positions for each country ( $N_{\text{Country (each)}} = 19,200$ ;  $N_{\text{Total, Pandemic}} = 57,600$ ). The same time frame was used to retrieve data from 2019 ( $N_{\text{Germany}} = 19,199$ ,  $N_{\text{Austria}} = 19,199$ ,  $N_{\text{Switzerland}} = 19,200$ ;  $N_{\text{Total, Pre-pandemic}} = 57,598$ ). Prior to compiling the final dataset ( $N = 115,198$ ), we controlled for any missing values or

observations. On May 17, 2019, in Germany and Austria one missing observation was found respectively for the chart positions 103 in Germany and 135 in Austria. Since those two missing observations in total could not be reconstructed, we removed them and cleaned the entire dataset by converting the cell information into the correct format (e.g., factors as factors, not character strings).

## Audio Features

Spotify provides information on several audio features for every song on its platform, which developers can access via an API. According to the description by [Spotify \(2021a\)](#), these features are estimated and calculated for each track. Some variables are represented as integers, such as loudness (in dB) or tempo (in BPM), while other features, such as “Acousticness”, “Energy” or “Valence” are aggregated scores based on algorithmic computations implemented by [Spotify \(2021a\)](#). All of the provided audio features are: “Acousticness”, “Danceability”, “Duration”, “Energy”, “Instrumentalness”, “Key”, “Liveness”, “Loudness”, “Mode”, “Speechiness”, “Tempo”, and “Valence”. While there is some research on mood-orientated music selection and playlist curation on Spotify (e.g., [Eriksson et al., 2019](#); [Luck, 2016](#)), there are, to our knowledge, only few studies using the free aggregated audio features by Spotify for investigating music listening behavior (e.g., [Heggli et al., 2021](#)), whereas within the scope of music information retrieval audio features are often used for content/emotion-based music recommender systems (e.g., [Deng, 2014](#)).

## Statistical Analyses

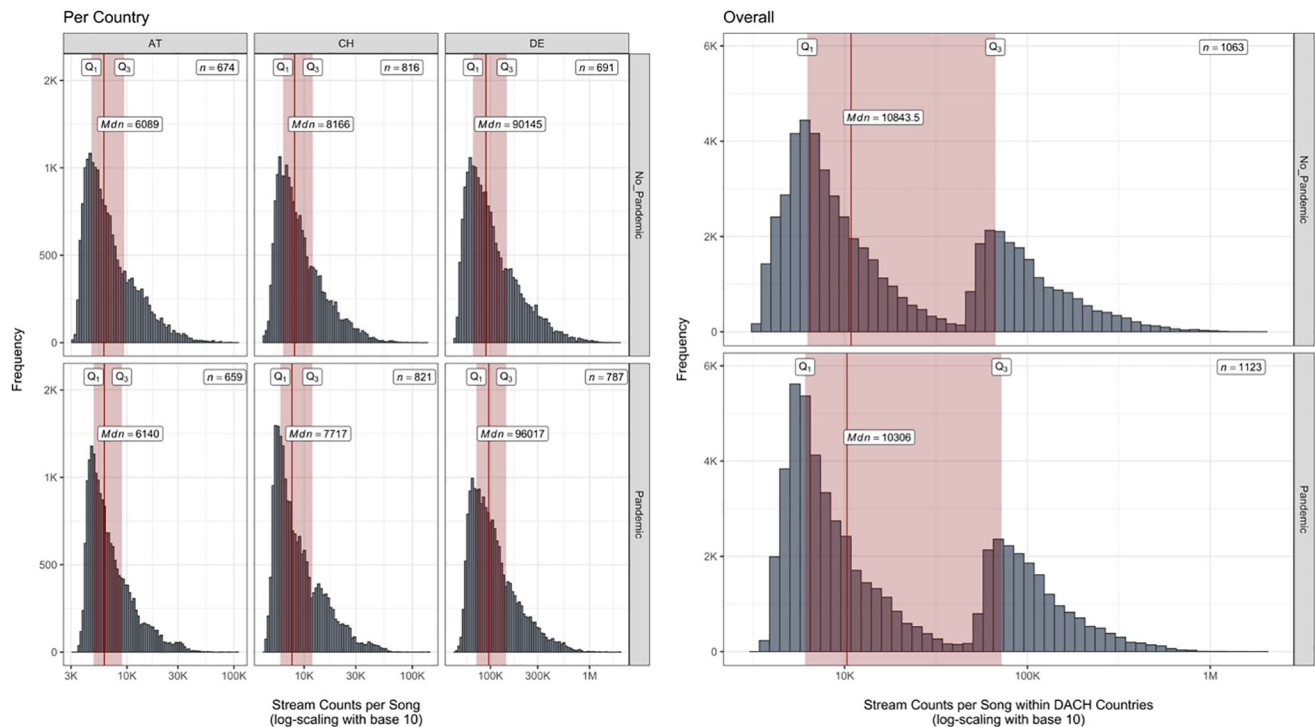
### Exploratory Analyses

For exploring the distributions of the raw (i.e., non-summarized) overall stream counts to understand the properties of the data, it is helpful to plot histograms. Theoretically, a high frequency of low stream counts and a lower frequency of higher stream counts seems more likely than a symmetric (normal) distribution of the stream counts in question. This is because few songs typically reach peak stream counts, while most songs are streamed to a moderate extent—especially those that are listed in such top chart lists. This also matters for recommender systems (see, e.g., [Gorakala & Usuelli, 2015](#), and [Deng, 2014](#)).

The following histograms show highly right-skewed, or left-leaning, frequencies of those stream counts of songs per country and, accordingly, also for all DACH countries together per period. That is, the stream counts per song of the upper 25% quartile (above the third quartile) are higher than in the respective interquartile range (see labels and the red-shaded rectangles in [Figure 1](#)). Hence, the exploratory analysis reveals that stream counts are not normally distributed and the higher stream counts of songs above the third quartile ( $Q_3$ ) are far outside the respective medians. Specifically, when taking a look at the overall histograms we see that smaller frequencies of stream counts in Switzerland and Austria yield bimodal distributions (or two combined unimodal distributions) of the overall data when they are combined with the stream counts of Germany. These findings make it reasonable to summarize the stream counts for each song with medians instead of means (i.e., ensuring robustness).

**Figure 1**

Comparison of Stream Counts of Daily Top 200 Spotify Charts Before and During the Pandemic per Country and Across All DACH Countries



Note. A log-scaling with base 10 was applied to the x-axis for visual purposes, mainly, to avoid a heavy tail of the higher stream counts.

## Toward a Classification Model

To address our first hypothesis, we reduce the dimensionality of the mood-related audio features by using a clustering approach. This makes the mood-related audio features more interpretable, which, in turn, means that we can implement the cluster assignments in question as additional input variables into our classification model.

### K-Means Clustering

We considered commonly used algorithms for structuring the given audio features to have more interpretable information (i.e., reducing the dimensions of a dataset). Partitioning a dataset into fewer dimensions essentially means identifying groups with similar within-characteristics compared to other groups with other within-characteristics (see James et al., 2015). To explore such groups in a given dataset without first reducing dimensionality, we can use one of the most straightforward and effective clustering techniques: The *k*-means algorithm. Once we have determined the number of distinct clusters/groups (in fact, the “*k*” in the name denotes the number of assumed clusters) and normalized the values of the variables that should be clustered, we can run this algorithm. The crucial point of this clustering strategy is that an identified cluster is characterized by the average of all data points that should belong to this cluster because of their similarities. That is, each cluster possesses a center, a centroid, that corresponds to the mean of its assigned data points. Several variations of this strategy exist: They use different distance measures, which account for the mentioned with-

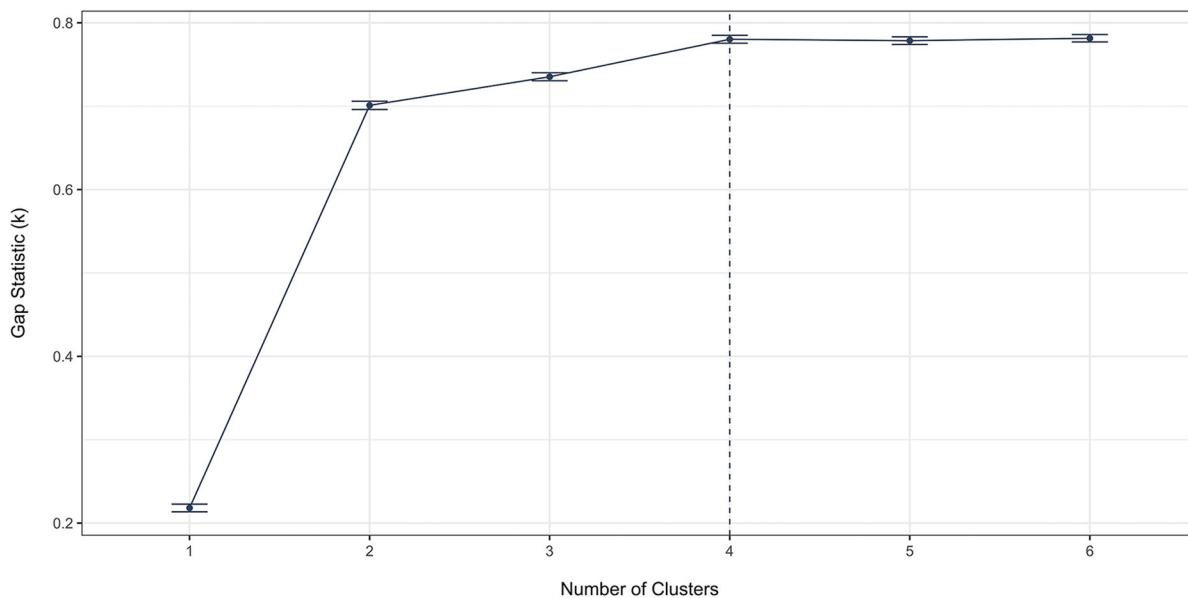
in-cluster-similarities, and each has its respective (dis)advantage and goal (for a discussion of the four most used variations, see [Morissette & Chartier, 2013](#)). Even if there are several variations of this dimensionality reduction strategy with different distant measures, their basic idea boils down to (see [Hartigan & Wong, 1979](#)): Once the number of centroids is determined, for example, by using the gap statistic (see [Tibshirani et al., 2001](#)) and theoretical considerations,  $k$  centers (centroids) are randomly and iteratively assigned to each observation by minimizing the sum of the squared Euclidean distances (or another distance metric) between observations and their centroids. The algorithm terminates as soon as these centroid assignments no longer change during the iterations. Since the centroids are the means of the observations assigned to their cluster and our case includes the tonal modality of the songs (i.e., 0 and 1), we basically weigh these modes to have not only, for instance, two clusters that mainly distinguish between higher and lower valence levels, but also clusters that also account for the different modes. This is, in fact, a trade-off decision between theoretical and data-driven considerations. Since Hartigan and Wong's algorithm ensures that cluster assignments will be stable once the iterations are completed, it also accounts for cluster observations that are closer to other cluster centroids as it aims to minimize the sum of squared errors. Accordingly, observations could be assigned to a cluster whose centroid is farther than other potential centroids, provided such assignments would reduce the sum of squared error (cf. Lloyd's version: [Morissette & Chartier, 2013](#), also [James et al., 2015](#)). Thus, in our case, it appears reasonable to stick with Hartigan and Wong's solution, especially as it empirically outperforms Lloyd's version ([Slonim et al., 2013](#)). However, before implementing an assumed number of clusters, we should validate the optimal number of clusters by testing different  $k$ -values on at least a subset of the entire dataset. For doing data-driven considerations, we use the above-mentioned gap statistic method proposed by [Tibshirani and colleagues \(2001\)](#). Essentially, this method formalizes the well-known elbow/silhouette heuristics (cf. scree plot method) for estimating the optimal number of clusters by using the rationale of a Monte Carlo simulation ([Efron & Tibshirani, 1993](#)). The gap statistic's advantage over elbow/silhouette heuristics is that the total within-cluster variation (characteristics) for different  $k$ -values (centroids) is taken into account as a reference distribution of the data (i.e., the null hypothesis that the data do not have distinguishable clusters) when determining the optimal  $k$ -value. The reference distribution, on the other hand, is sampled by using a Monte Carlo simulation (i.e., bootstrapping). Once the Monte Carlo simulation is done, the final value of optimal clusters is selected according to the maximal gap statistic with regard to the reference distribution.

Following this procedure, we choose [Spotify's \(2021a\)](#) mood-related audio features. We also include songs' tonal modality as a dummy-coded variable, and tempo (BPM) since many well-reviewed theoretical and empirical findings strongly support the assumption that these variables influence listeners' mood ([Sloboda, 2010](#)). Once we have min-max-normalized and rescaled all considered features to the same scale between 0 and 1, we apply the gap statistic method to identify the optimal number of clusters. This procedure partitions the entire dataset of all distinct songs, in our case, into four clusters as can be seen in [Figure 2](#) and [Table 1](#).



**Figure 2**

Optimal Clusters With Error Bars According to the Gap Statistic Method

**Table 1**

*K*-Means Cluster Solution on Min-Max-Normalized and Rescaled Mood-Related Audio Features

Cluster	<i>n</i>	Mode	Danceability	Energy	Loudness (rescaled)	Valence	Tempo (rescaled)
1	26,990	1	.672	.572	.738	.335	.120
2	32,626	0	.754	.716	.812	.678	.121
3	26,129	1	.738	.722	.823	.640	.119
4	29,453	0	.679	.605	.748	.350	.120

Finally, we run [Hartigan and Wong's \(1979\)](#) *k*-means algorithm on the selected audio features of the entire dataset with  $k_{\text{Gap Statistic}} = 4$ , 50 random initializations, and 100 iterations.

Although these clusters are initially meaningful in terms of a high compactness within and high distinguishability between clusters (based on a high between sum of squares [BSS] to total sum of squares [TSS] ratio—in our case:  $\text{BSS/TSS} = 83.3\%$ ), these clusters are still less informative regarding their underlying and assumed mood dimensions since they are unlabeled. To give these clusters informative and interpretable names, we should take a closer look at their specific means while considering the range of their normalized scale (i.e., a value of .5 indicates a constant value for the variables Danceability, Energy, Loudness (rescaled), and Valence). We can state, for example, that “Valence” is represented by two values that indicate both a high (i.e., positive;  $.678 > .5$  and  $.640 > .5$ ) and a low (i.e., negative;  $.335 < .5$  and  $.350 < .5$ ) characteristic for each tonal modality (major = 1, minor = 0) according to Spotify's characterization of these mood-related variables. In this vein, we finally can decide how to name these clusters so that they are conceptually informative and interpretable.

Interpreting these cluster centroids according to the arousal-valence-circumplex model ([Russell, 1980](#), also [Scherer, 2005](#)) and the theories mentioned above, we can state that if a cluster has a high ( $> .5$ ) “Valence”,

“Danceability”, “Energy” and “Loudness” value, this cluster can be characterized by a *higher arousal-potential* and a *positive emotionality* that music listeners may associate with such qualities of those audio features. Since this particular cluster (*higher arousal-potential* with *positive emotionality*) consists of songs whose tonality should be major, we can support the potential positiveness that listeners may also experience based on the tonality of the songs that should belong to this cluster (see e.g., [Parncutt, 2014](#), also [Athanasopoulos et al., 2021](#), also [Sievers et al., 2013](#)). Hence, we can name and characterize the other clusters according to their values on the arousal and valence levels and update the previous [Table 1](#) as shown in [Table 2](#).

**Table 2**

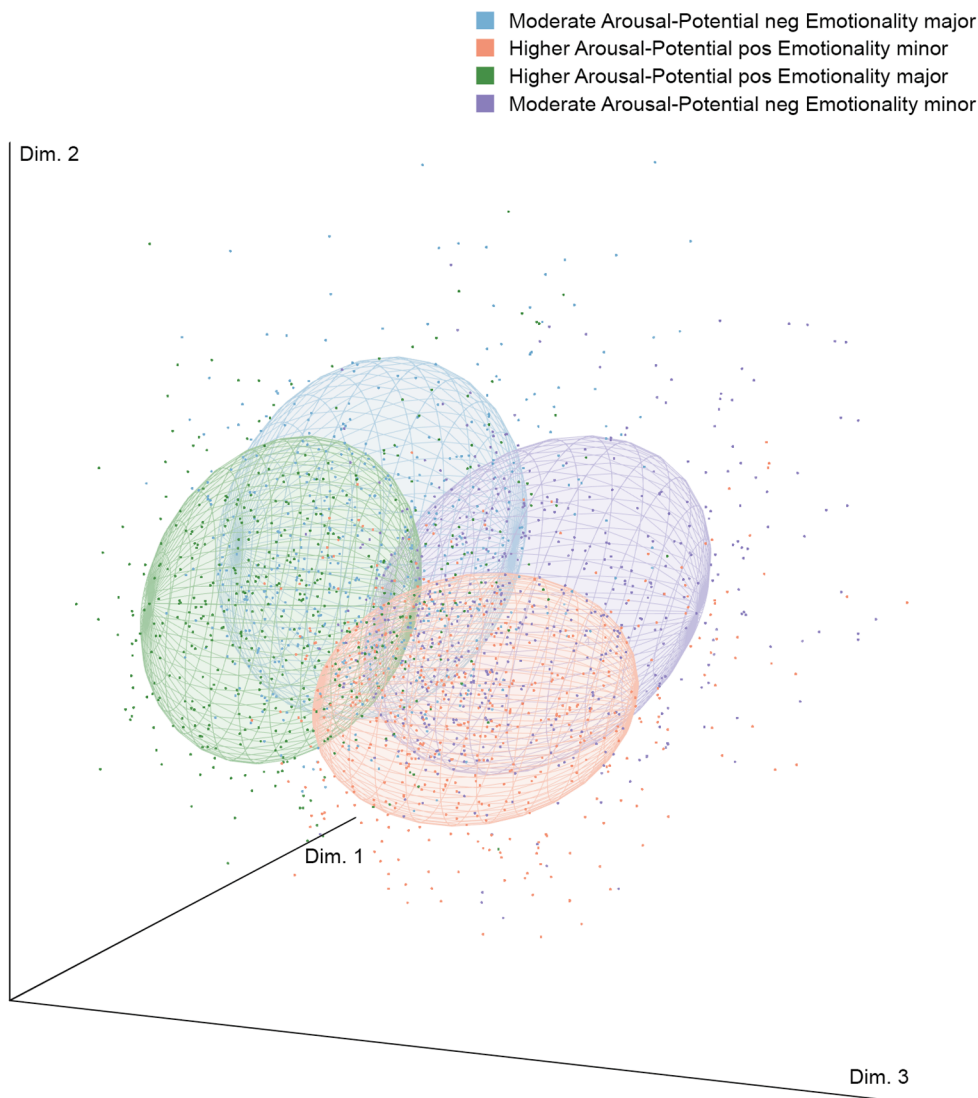
*K-Means Cluster Solution on Min-Max-Normalized and Rescaled Mood-Related Audio Features With Additional Cluster Characterizations*

Cluster/Cluster names	<i>n</i>	Mode	Danceability	Energy	Loudness (rescaled)	Valence	Tempo (rescaled)
<b>1</b> Moderate Arousal- Potential with negative Emotionality (major)	26,990	1	.672	.572	.738	.335	.120
<b>2</b> Higher Arousal- Potential with positive Emotionality (minor)	32,626	0	.754	.716	.812	.678	.121
<b>3</b> Higher Arousal- Potential with positive Emotionality (major)	26,129	1	.738	.722	.823	.640	.119
<b>4</b> Moderate Arousal- Potential with negative Emotionality (minor)	29,453	0	.679	.605	.748	.350	.120

A scatterplot is helpful for visually assessing the compactness and separation of these clusters. As [Figure 3](#) shows, the different clusters are distinguishable when they are plotted in a three-dimensional space. The two clusters with a higher arousal-potential (bright red- and dark green-colored) are more compact than those with moderate arousal-potential, as these show greater dispersion (i.e., purple- and blue-colored).

**Figure 3**

3D Scatterplot With Ellipsoids of the K-Means Cluster Solution Across All DACH Countries



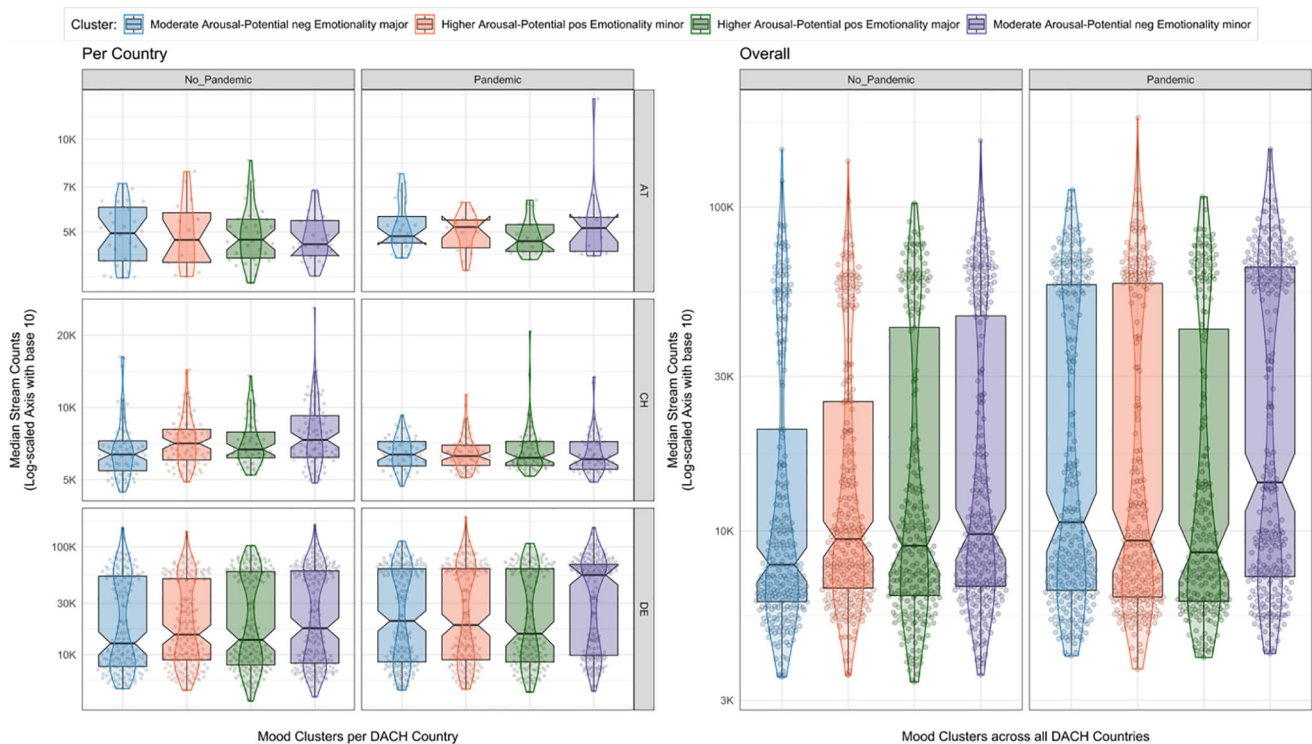
*Note.* Ellipsoids show at a coverage level of 68.27% (i.e., area within the first *SD* when a bivariate normal distribution is assumed) how concentrated the respective clusters are. Dimension reduction for visual purposes was conducted by using principal component analysis.

To account for research hypothesis  $H_1$ , we need to test whether and how these cluster assignments differ between the two periods in question across all DACH countries and between each DACH country. So, to explore any meaningful differences regarding the most streamed cluster per country and per period, we can plot the clusters within each DACH country and between those two periods against their stream counts while taking the skewness of the stream counts (see Figure 1) into account by summarizing on median stream counts. At first glance, Figure 4 shows that the most boxplot notches overlap. Based on that, we can informally conclude that any potential differences in the streamed clusters per country are rather small, if even present, regarding their effect size ( $r \leq .10$ ). Interestingly, the median stream counts of the purple-colored mood cluster “Moderate Arousal-Potential with negative Emotionality (minor)” are higher during the pandemic across all

countries and per country, except for Switzerland. Here we see rather the opposite since the songs belonging to the cluster “Moderate Arousal-Potential with negative Emotionality (minor)” were streamed in Switzerland during the pre-pandemic period more frequently than during the pandemic. To test this visual impression statistically, we run pairwise comparisons via the Dunn test with Holm correction for each country and for both periods across all countries with an alpha-level of 5%.

**Figure 4**

*Combined Box, Violin, and Scatter Plots of the K-Means Cluster Solution for Each DACH Country and Across All DACH Countries Against Their Median Stream Counts Before and During the Pandemic*



As expected from the visual information in Figure 4, there are statistically significant differences in the median stream count per song in a cluster observable, both between two DACH countries and across all DACH countries between both periods. According to the Dunn test with Holm correction for the median stream counts per song per DACH country between both periods, we can observe few relevant difference effects: Indeed, in Switzerland the central tendencies of the median stream counts per song belonging to the purple-colored mood cluster Moderate Arousal-Potential with negative Emotionality (minor) differ to a moderate extent ( $Mdn_{No\_Pandemic} = 7,328.5$ ,  $Mdn_{Pandemic} = 6,085$ ,  $z = -3.69$ ,  $p_{adj} = .002$ ,  $r = .328$ ,  $n_{No\_Pandemic} = 76$ ,  $n_{Pandemic} = 51$ ). That is, people in Switzerland streamed more of the songs that belong to this cluster during the pre-pandemic period. Furthermore, we can also observe a small difference effect regarding the bright red-colored cluster “Higher Arousal-Potential with positive Emotionality (minor)” in Switzerland. During the pandemic, the songs belonging to this clusters were less streamed compared to the pre-pandemic period ( $Mdn_{No\_Pandemic} = 7,084$ ,  $Mdn_{Pandemic} = 6,275$ ,  $z = -2.86$ ,  $p_{adj} = .0416$ ,  $r = .241$ ,  $n_{No\_Pandemic} = 79$ ,  $n_{Pandemic} = 63$ ). The only other relevant difference effect, although a small one, is observable in one cluster in Germany. Here, people streamed the songs belonging to the purple-colored cluster Moderate Arousal-Potential with negative Emotionality (minor)

more often during the pandemic ( $Mdn_{No\_Pandemic} = 17,591.5$ ,  $Mdn_{Pandemic} = 54,730$ ,  $z = 3.75$ ,  $p_{adj} = .002$ ,  $r = .185$ ,  $n_{No\_Pandemic} = 190$ ,  $n_{Pandemic} = 217$ ).

When it comes to differences across all DACH countries between both periods, we can only state small difference effects in the clusters Moderate Arousal-Potential with negative Emotionality (major) ( $Mdn_{No\_Pandemic} = 7,874.5$ ,  $Mdn_{Pandemic} = 10,639$ ,  $z = 3.53$ ,  $p_{adj} = .002$ ,  $r = .154$ ,  $n_{No\_Pandemic} = 250$ ,  $n_{Pandemic} = 275$ ) and Moderate Arousal-Potential with negative Emotionality (minor) ( $Mdn_{No\_Pandemic} = 9,786.5$ ,  $Mdn_{Pandemic} = 14,124$ ,  $z = 3.55$ ,  $p_{adj} = .002$ ,  $r = .148$ ,  $n_{No\_Pandemic} = 288$ ,  $n_{Pandemic} = 283$ ). Based on these findings, we can now address the second hypothesis by building a binary classifier.

### Building a Support-Vector Machine Binary Classifier

To test hypothesis  $H_2$ , we build a binary classifier and change with this the perspective in that we aim to identify and classify given information. That is, the two periods, as factors, pose now the dependent variable, whereas the mood clusters as factors together with rescaled stream counts and chart positions as well as the variables “Acousticness”, “Speechiness”, “Liveness”, “Instrumentalness”, the duration of the songs, and the DACH countries as factors pose now the independent or input variables.

Simple classifications tasks within the scope of machine learning, such as binary classifications (cf. logistic regression), are helpful in deciding to which class an observation belongs in not seen data (i.e., the validation and/or the test dataset) based on learned/modelled structures of a training dataset. Different algorithms have various uses for solving such tasks (see e.g., naïve Bayes,  $k$ -nearest neighbor, or random forest classifiers in James et al., 2015). However, when it comes to multiple input variables (i.e., high-dimensional space) and complex (i.e., non-linear or overlapping) cases regarding which label (i.e., pre-pandemic vs. pandemic) an observation belongs to, fewer algorithms are suited to handle this task well. Considering these preconditions, we choose the support-vector machines (SVM) algorithm, because it is particularly strong in solving high-dimensional and non-linear problems (Rhys, 2020). While the fundamental idea of this algorithm is straightforward, the mathematical background requires more explanation (for a detailed introduction to SVM and their mathematical background, see, e.g., Campbell & Ying, 2011; Hastie et al., 2009; James et al., 2015; for the original algorithm proposal: Boser et al., 1992). The main reason why SVMs often outperform other (multi-label) classifiers is that the algorithm includes a so-called *kernel trick*. This trick essentially transforms the data based on a (non)linear kernel function that adds an extra dimension to the data (i.e., hyperplane) whose position is *supported* by the data points that touch the margin of this hyperplane. The hyperplane allows actual overlapping observations to be separated in a higher dimension, such that either some misclassifications are accepted for ensuring greater robustness (i.e., soft margin classification), or that misclassifications are not tolerated (i.e., hard/maximal margin classification; prone to overfitting). So, it appears reasonable to not always seek the maximal margin and the perfect fit that would yield correct classifications, both on the correct side of the hyperplane and the margin. That is, we also have to choose whether a soft or a hard/maximal margin classifier should be used as an additional cost hyperparameter (C) that penalizes incorrect observations inside the margin and boundary with narrower margins. Finally, to control how much influence an observation has on the position of the hyperplane, we use an additional hyperparameter ( $\gamma$ ) to validate how granular the decision boundaries should be. Before running the SVM algorithm at all, these hyperparameters need to be cross-validated so that the optimal values can be used during model training once the entire dataset is partitioned into training and test sets, such that classifications (or predictions) can be made based on the (learned/modelled) structures within the training data.

Since we have two balanced classes (i.e., the periods), we split the data structure by considering the so-called Pareto principle such that 20% of the entire data were assigned to the test dataset and 80% of the data were used for training and cross-validation purposes. Furthermore, we aggregated the datasets so that the training and test data consist of the same track IDs and the same countries.

After partitioning the entire dataset into training and test sets, we took a smaller random sample (20%) of the training dataset (due to computational reasons) to finally run a random grid search for five-fold cross-validated hyperparameters (Rhys, 2020, regarding the concept of  $k$ -fold cross-validation see, e.g., Witten et al., 2011) for a radial kernel function within the range of  $0.5 \leq \gamma \leq 5$  and  $10^{-1} \leq C \leq 10^4$  (this entire cross-validation took nonetheless more than two and a half hours although multiple cores were parallelized). Building on the outcomes of this five-fold cross-validation of the hyperparameters, we trained the binary SVM classifier with a radial kernel function with the best cross-validated fit of  $C = 100$  and  $\gamma = 2$ . Once our SVM model was trained, we applied it on the test set to classify the two periods in question. The results show that our model can indeed classify the two periods with a high overall accuracy (i.e., macro-F1 measure) across all classes (ACC = 97.87%, 95% CI [.977, .981]). The pre-pandemic period was used as a reference class which means that the model was trained to classify true positive/correct observations of the pre-pandemic period. Although such a high accuracy poses a desired outcome of our model training, we do not know at this stage which independent or input variables had to what extent influence in classifying the pre-pandemic period that correctly. In other words: Why should we trust our model?

**Interpreting SVM classifications** — Since the SVM algorithm is based on complex inner structures, the interpretability of its outcome is not a trivial task—not to speak of when we deal with relatively large datasets. To overcome such an obstacle, we can examine different aspects that ensure or facilitate interpretability on different levels. On a more general level we can, for instance, extract different degrees of importance of the used independent variables to classify (or predict) the dependent variable of interest (Fisher et al., 2019). On a more specific level, it is also possible to investigate, for example, a partial dependence of a certain independent variable regarding its probability to classify (or predict) the outcome variable. This is particularly useful when interpreting marginal effects of an independent variable on the dependent variable: The smaller the marginal effects on the dependent variable are, the less important they are regarding their classification or prediction impact (Greenwell et al., 2018; see also Boehmke & Greenwell, 2020, and Molnar, 2019, where interpretable machine learning methods are introduced and discussed).

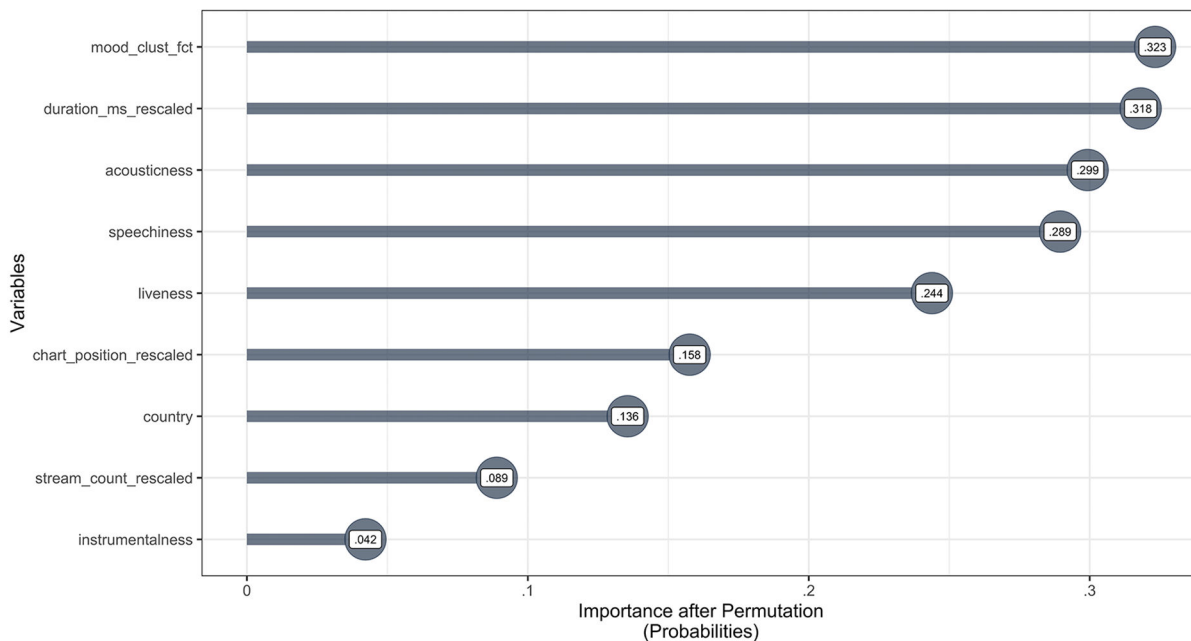
When considering such approaches to ensure interpretability of our built model, we can apply them accordingly so we can explain how our independent variables influence the true correct/positive classification of observations that belong to the pre-pandemic period (and accordingly to the pandemic period). Specifically, we can use a permutation approach to extract the importance of the used independent variables. The main idea of this approach revolves around the error that the model will have during its classification or prediction if the values of the independent variables in the training set were permuted. For if the values of the independent variables are permuted, their relation to the dependent variable is practically destroyed. Now, if the difference in the probability to correctly classify (or predict) the dependent variable between the baseline classification and the permuted version shows that the model error is higher (relative to the other independent variables), we can conclude that the independent variable with the highest probability of making an error if it is ignored after permutation is the most important independent variable to correctly classify (or predict) the reference class of

the dependent variable (Boehmke & Greenwell, 2020). Accordingly, the independent variables with the lowest probability after their permutation are less important to classify the dependent variable correctly.

After carrying out such a permutation of all input variables of the entire training set with five Monte Carlo simulations (i.e., the independent variables of 92,229 observations were, first, five times permuted, then, the results were averaged), we can plot the results of the probabilities for classifying the pandemic period incorrectly if the respective independent variables are ignored relative to the other variables as shown in Figure 5.

**Figure 5**

*Permutation-Based Variable Importance (Independent Variables) for the SVM Model (Training Set)*

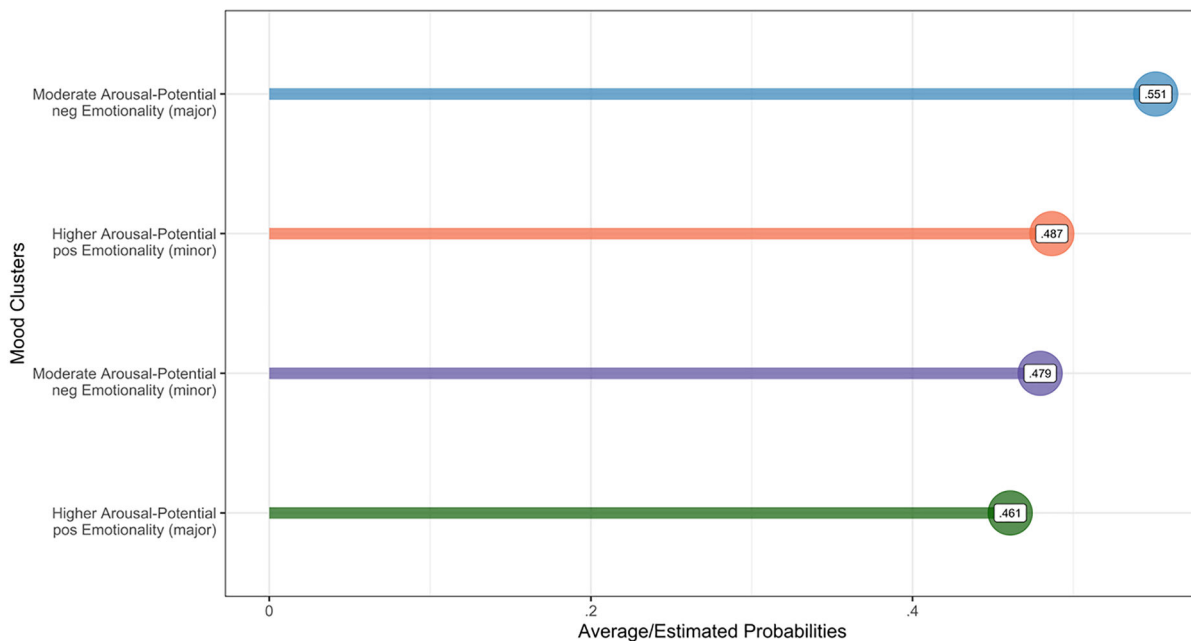


As we can see, our previously clustered mood-related audio features are the most important independent variable (factor) in classifying the pandemic period. That is, the probability for a classification error increases by 32.3% if this factor will be ignored relative to the other independent variables. We see, by using this permutation-based variable importance approach, we already can provide more concrete interpretability since we can explain to what extent the respective independent variables influence the true correct/positive classification of the dependent variable. However, we can even improve the interpretability if we consider the partial dependence of the distinct mood clusters within the variable mood\_clust\_fct since we now know that this variable (factor) is the most important one. By doing so, the values of all four distinct mood clusters are once replaced with each cluster, then the classification of the dependent variable of interest follows. Once this iterative process is done, all classifications are averaged.

Resulting from this procedure, we see in Figure 6 that the average/estimated probabilities of the distinct clusters to classify the dependent variable, the pandemic period in our case, indeed marginally differ.

**Figure 6**

*Partial Dependence Plot of the Mood Clusters With Averaged Probabilities Regarding the Classification of The Pandemic Period*



Furthermore, we see that all clusters have a similar effect on the model classification of the pandemic period. In particular, we see that the songs, that belong to the cluster Moderate Arousal-Potential with negative Emotionality (major), show the highest probability for classifying the pandemic period relative to the other clusters (this goes in line with the observed difference effect, see Figure 4), whereas the songs belonging to the cluster “Higher Arousal-Potential with positive Emotionality (major)” have the lowest probability, although similarly low as both clusters in the middle.

After carrying out these methods to gain more insights, we could indeed extract useful information on both levels. On the more general level, we now know which independent variables contribute to what extent to the classification task. Moreover, we know that our identified mood clusters are the most important variable in our model. On the other hand, we also could zoom in the mood cluster variable (factor) to see how the different clusters have their marginal effects in classifying the pandemic period. That is, these post-hoc methods can ensure more interpretability so that even highly complex supervised machine learning algorithms/models do not remain non-understandable black box models (provided the model is well-trained and cross-validated).

## Results

Examining music-listening behavior during the COVID-19 pandemic in the DACH countries by using open data from the streaming service provider Spotify yields results that support our research hypotheses, even with some caveats. Our work supports  $H_1$  (distinct mood-related clusters can be identified based on the given audio features that reflect music-listening behavior) in that we could indeed identify mood-related clusters with a



BSS/TSS ratio of 83.3% that could represent music-listening behavior during the pandemic as a proxy in virtue of the different audio feature qualities.

Furthermore, we could find statistically significant differences with small to moderate effects ( $.149 \leq r \leq .328$ ) regarding the respective cluster stream counts within a country as well as across all countries between both periods. Hence, we can support  $H_1$  not only in terms of the overall identified clusters, but also with respect to significantly different cluster stream counts within two countries and across all DACH countries between the pre-pandemic and the pandemic period.

On the other hand, we also have evidence to support  $H_2$  (periods can be classified based on the mood-related clusters, the other available audio features, and the DACH countries) based on the outcome of our built binary SVM classifier that yielded a high overall accuracy (ACC = 97.87%, 95% CI [.977, .981]). For a tidy overview of the model's performance metrics, we can consult the confusion matrix to evaluate our model's accuracy by using the harmonic mean of precision (ratio of correct classified observations and the total of all classified observations) and recall (proportion of correct classified observations to all observations that should be classified as correct because they possess the characteristic or the label to be classified). This is especially desirable, as we cannot visualize the final classification results and the decision boundaries because of the high dimensionality of the data (for the so-called *curse of dimensionality*, see Bellman, 1957). Furthermore, a principal component analysis for dimensionality reduction would not yield the desired interpretability, unlike the visualization of the cluster solution, since the different input variables in question would not be distinguishable, while the decision boundaries of our trained model would be obsolete at the same time. However, based on the variable importance and the partial dependence of the most important variable to classify the (pre-)pandemic period, we could develop an understanding of how the independent variables have effects on the classification of the two periods in question. So, to have now the tidy overview of the model's performance, a confusion matrix provides further useful insights.

The correct classification rates for both periods are high as shown by the columns precision and recall of Table 3. Specifically, our model classifies almost all observations correctly; however, the slightly higher recall than precision in the pandemic period indicates that our model also misclassifies some observations as pandemic-related even though they are not (i.e., false correct classifications). Recalling that the blue-colored mood cluster Moderate Arousal-Potential with negative Emotionality (major) has the highest probability relative to the other clusters in classifying the pandemic period, we can at this point infer that false correct classifications of the pandemic period are at least correlated with this cluster. The opposite is true regarding the pre-pandemic-related observations: Our model correctly classifies some pre-pandemic-related observations, while it misses others (lower recall than precision, i.e., false incorrect classifications; for a discussion of these classification measures, see, e.g., James et al., 2015). However, as we deal with balanced data (prevalence = 50%), such misclassifications are also balanced across both classes (see F1-score) so that the overall accuracy provides in our case indeed a suitable evaluation metric.

**Table 3***SVM Classification Confusion Matrix*

Period (classes)	Observations (song IDs)		%		
	Correctly classified	Actual	Precision	Recall	F1
No Pandemic	11,236	11,487	97.93	97.81	97.87
Pandemic	11,244	11,482	97.82	97.92	97.87

All in all, we can therefore state that our model performs quite well in classifying the observations of each period to their respective class based on the used input variables.

## Discussion

Since we clustered Spotify's mood-related audio features together with the tonal modalities that Spotify assigns to the songs, we have, in principle, a framework for making further assumptions about the possible underlying music-listening behavior regarding Spotify's top 200 chart songs during the first wave of the COVID-19 pandemic in DACH countries and the reference period in 2019. In this vein, we can state that the four identified clusters may represent emotional dimensions according to the arousal-valence-circumplex model (Russell, 1980, Scherer, 2005) that can be experienced by potential listeners whose levels of music-related (psycho)social states and traits—such as musical preferences (Rentfrow & Gosling, 2003), personality traits (Costa & McCrae, 1980, also Gosling et al. 2003), cognitive styles of music listening (Kreutz et al., 2008), or musical sophistication (Müllensiefen et al., 2014)—are represented on the levels of those audio features. As we could find statistically meaningful differences in the stream counts of the identified clusters within two countries and across all DACH countries between both periods, we can state that the identified clusters were streamed differently in both periods. This is insofar interesting as we can conclude the following: Since across all DACH countries the clusters with moderate arousal and negative emotional potentials, both in minor and major, were streamed during the pandemic more often, we can tentatively assume that music listeners could cope with pandemic's stress by listening to the songs belonging to these clusters—possibly according to the iso-principle (see Schramm, 2005). However, we have to be careful with such assumptions as such effects might only be correlative and not causal, especially as we do not have survey data at hand, but only proxy variables (see the identified mood cluster) that could support such assumptions.

Furthermore, since our binary SVM classifier could classify each period based on the mood clusters, the remaining audio features, and the DACH countries, the results indicate that both periods show distinct profiles, which is why it was possible to classify the observations to the periods. Thus, we can answer our research question based on  $H_1$  and  $H_2$  as follows:

$H_{1*}$ : The dimensionality of Spotify's mood-related audio features could be reduced to fewer clusters. The statistically significant small to moderate difference effects ( $.149 \leq r \leq .328$ ) in the median stream counts per song within a cluster per country between both periods as well as across all DACH countries support the hypothesized difference effects.

$H_{2*}$ : The mood-related clusters could be implemented in a classification task such that each period was classified with high proportions of precision and recall and a high overall accuracy (ACC = 97.87%).

This means, in turn, that each period shows a distinct profile in terms of the mood clusters, the used audio features of the track IDs and the grouping factor of the DACH countries.

We could thus extract mood clusters that explain (i.e., BSS/TSS ratio) the partitions according to the mood-related variables to a great extent (83.3%). Furthermore, we could classify the pre-pandemic and pandemic period based on these mood clusters in addition with the remaining audio features and the DACH countries as factors quite well by using an interpretable binary SVM classifier (ACC = 97.87%, 95% CI [.977, .981]).

Although we were able to answer our research question, our approach has some limitations, which should be addressed and discussed, so that they can be considered in possible follow-up research projects.

## Limitations

First, since Spotify is not remotely transparent about how it determines these audio features, we should remain skeptical when interpreting these results in terms of the correctness of Spotify's assignments. Some of Spotify's tonal modality assignments are clearly incorrect (e.g., the song "Blinding Lights" by the artist The Weeknd is listed as major or 1 although listening to the song clearly reveals it is written in minor). This could be regarded as a drawback since our cluster assignments depend on Spotify's assignments. That is, even if our clustering approach could identify distinct clusters with a high BSS/TSS ratio, unsupervised clustering methods, such as our used *k*-means algorithm, heavily depend on the data preprocessing and especially the distant measures being considered. As we know that Spotify's assignments are at most approximately true or correct, any results based on Spotify's audio features must be interpreted with caution (see also Heggli et al., 2021). Nevertheless, since Spotify does provide a trove of open data, it is still worth scrutinizing them at least to have reference values for further research: For example, it would be interesting to examine how listeners with different (psycho)social state and trait conditions indeed experience the levels of Spotify's mood-related audio features for certain songs, so that reasonable conclusions can be reached about how different kinds of music were listened to as well as why and how music listeners selected the songs in question (see Section 1).

Second, a period- and country-specific characterization of how people coped with the pandemic-related stress can only be provided when people are surveyed regarding their music-listening behavior (e.g., Fink et al., 2021; Granot et al., 2021). Combining such insights with our data-driven approach or more advanced approaches (by comparing results of different algorithms with each other) could yield further interesting results about the actual music-listening behavior that is surveyed and the assumed mood-related levels.

Third, it would be interesting to analyze the sentiment qualities of the streamed song lyrics to see how listener's experienced emotional states go in line with the qualities of the audio features and the lyrics of the streamed songs (for a description of the iso/catharsis- versus compensation-principle while listening to music, see, e.g., Schramm, 2005).

Fourth, we can only state that the extracted mood clusters can represent the music-listening behavior simply in virtue of the clusters that are extracted from audio features of songs that were daily streamed during the pandemic and its reference period. However, since we cannot assume that the quality (i.e., correctness of the assignments) of Spotify's provided data is perfect, any conclusion of how the clusters indeed reflect any coping-strategies is fairly limited. However, these clusters can be used as reference values for further research.

Following investigations may wish to adjust the audio features with features extracted using music information retrieval (e.g., [Lartillot et al., 2008](#)).

Finally, these findings are based on the overall most listened songs on Spotify and may not be representative of all listeners in the DACH countries. People may have turned to certain music for nostalgic reasons (e.g., [Yeung, 2020](#)) or listened to specific music in their own genre. Yet, using the top songs enlightens us to the behavior of a great fraction of listeners in central European German-speaking countries. In this vein, future studies may want to apply these approaches to analyze how songs from specific genres or times were listened to before and/or during a time of crisis.

## Conclusion

This contribution adds to the growing body of research on music consumption before and during the pandemic by showing how open data can be used to characterize music-listening behavior in German-speaking countries. Although Spotify's aggregated data is constrained, we observe that audio features referring to emotional and arousing states are still useful for summarizing music listening into distinct mood-related clusters. Although our findings cannot explain the actual reasons and motivations behind why people streamed the chart songs, our findings can describe the characteristics of these songs in times of a ubiquitous crisis, e.g., the COVID-19 pandemic, based on the employed theoretical framework in comparison to a reference period.

This approach illustrates how combining data-driven analyses with theoretical psychological concepts and considerations can help to classify and to describe music-listening behavior. We attempted to apply interpretable machine learning techniques on open data to spark an interest in exploring and answering research questions at the intersection of music psychology and data science.

By making our coding scripts and data accessible, we hope to encourage future research in this field. This type of approach is especially desirable when it is hard to establish personal contact with potential participants, experts, or collaborators, as well as when access to big datasets is increasingly possible.

## Funding

Nicolas Ruth's contribution to this study has been funded by a Feodor Lynen Fellowship from the Alexander von Humboldt Foundation.

## Competing Interests

The authors have declared that no competing interests exist.

## Acknowledgments

The authors thank two anonymous reviewers for their critical reading and helpful comments, and Felix Bernoulli from the graphics department of the MPI for Empirical Aesthetics for rendering Figure 3 in high resolution.

## Data Availability

Data for this article is freely available (see the [Supplementary Materials](#) section).

## Supplementary Materials

For this article the following Supplementary Materials are available via PsychArchives (for access see [Index of Supplementary Materials](#) below):

- Retrieved and aggregated datasets for the entire study.
- The codebook, that contains the outputs of the respective code chunks so that it is not necessary to run the entire script to get the outputs of the data analyses and (raw) visualizations.
- A commented R script (web-scraping procedures and data analyses) with relative file paths is available to ensure reproducibility.

### Index of Supplementary Materials

Kalustian, K. K., & Ruth, N. (2021a). *Supplementary materials to: "Evacuate the dancefloor": Exploring and classifying spotify music listening before and during the COVID-19 pandemic in DACH countries* [Datasets, codebook]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.5020>

Kalustian, K. K., & Ruth, N. (2021b). *Supplementary materials to: "Evacuate the dancefloor": Exploring and classifying spotify music listening before and during the COVID-19 pandemic in DACH countries* [R script]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.5018>

## References

- Antonini Philippe, R., Schiavio, A., & Biasutti, M. (2020). Adaptation and destabilization of interpersonal relationships in sport and music during the COVID-19 lockdown. *Heliyon*, 6(10), Article e05212. <https://doi.org/10.1016/j.heliyon.2020.e05212>
- Athanasopoulos, G., Eerola, T., Lahdelma, I., & Kaliakatsos-Papakostas, M. (2021). Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *PLOS ONE*, 16(1), Article e0244964. <https://doi.org/10.1371/journal.pone.0244964>
- Bäuerle, A., Teufel, M., Musche, V., Weismüller, B., Kohler, H., Hetkamp, M., Dörrie, N., Schweda, A., & Skoda, E.-M. (2020). Increased generalized anxiety, depression and distress during the COVID-19 pandemic: A cross-sectional study in Germany. *Journal of Public Health*, Advance online publication. <https://doi.org/10.1093/pubmed/fdaa106>
- Bellman, R. (1957). *Dynamic programming*. Princeton University Press.
- Boehmke, B., & Greenwell, B. (2020). *Hands-on machine learning with R*. Github. <https://bradleyboehmke.github.io/HOML/>
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. A. (1992). Training algorithm for optimal margin classifiers. In D. Haussler (Ed.), *Proceedings of the 5th annual ACM workshop on computational learning theory* (pp. 144–145). Association for Computing Machinery.
- Campbell, C., & Ying, Y. (2011). Learning with support vector machines. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(1), 1-95. <https://doi.org/10.2200/S00324ED1V01Y201102AIM010>

- Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 38(4), 668-678. <https://doi.org/10.1037/0022-3514.38.4.668>
- Delsing, M. J. M. H., ter Bogt, T. F. M., Engels, R. C. M. E., & Meeus, W. H. J. (2008). Adolescents' music preferences and personality characteristics. *European Journal of Personality*, 22(2), 109-130. <https://doi.org/10.1002/per.665>
- Deng, J. (2014). *Emotion-based music retrieval and recommendation* (Publication No. 82) [Doctoral thesis, Hong Kong Baptist University]. Open Access Theses and Dissertations. [https://repository.hkbu.edu.hk/etd\\_oa/82](https://repository.hkbu.edu.hk/etd_oa/82)
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall. <https://doi.org/10.1007/978-1-4899-4541-9>
- Entringer, T. M., & Kröger, H. (2020). *Einsam, aber resilient: Die Menschen haben den Lockdown besser verkraftet als vermutet* [Lonely but resilient: People coped better than expected during lockdown]. Retrieved July 27, 2021 from <http://hdl.handle.net/10419/222876>
- Eriksson, M., Fleischer, R., Johansson, A., Snickars, P., & Vonderau, P. (2019). *Spotify teardown: Inside the black box of streaming music*. MIT Press.
- Feldmann, A., Gasser, O., Lichtblau, F., Pujol, E., Poese, I., Dietzel, C., Wagner, D., Wichtlhuber, M., Tapiador, J., Vallina-Rodriguez, N., Hohlfeld, O., & Smaragdakis, G. (2020). The lockdown effect: Implications of the COVID-19 pandemic on internet traffic. In ACM (Ed.), *Proceedings of the ACM internet measurement conference* (pp. 1–18). ACM Digital Library. <https://doi.org/10.1145/3419394.3423658>
- Fink, L., Warrenburg, L., Howlin, C., Randall, W., Hansen, N., & Wald-Fuhrmann, M. (2021). Viral tunes: Changes in musical behaviours and interest in coronamusic predict socio-emotional coping during COVID-19 lockdown. *Humanities and Social Sciences Communications*, 8, Article 648013. <https://doi.org/10.1057/s41599-021-00858-y>
- Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177), 1-81.
- Giordano, F., Scarlata, E., Baroni, M., Gentile, E., Puntillo, F., Brienza, N., & Gesualdo, L. (2020). Receptive music therapy to reduce stress and improve wellbeing in Italian clinical staff involved in COVID-19 pandemic: A preliminary study. *The Arts in Psychotherapy*, 70, Article 101688. <https://doi.org/10.1016/j.aip.2020.101688>
- Gorakala, S. K., & Usulli, M. (2015). *Building a recommendation system with R*. Packt.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the big five personality domains. *Journal of Research in Personality*, 37(6), 504-528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Granot, R., Spitz, D. H., Cherki Boaz, R., Loui, P., Timmers, R., Schaefer, R. S., Vuoskoski, J. K., Cárdenas-Soler, R.-N., Soares-Quadros, J. F., Li, S., Lega, C., La Rocca, S., Martínez, I., Tanco, M., Marchiano, M., Martínez-Castilla, P., Pérez-Acosta, G., Martínez-Ezquerro, J. D., Gutiérrez-Blasco, I. M., . . . Israel, S. (2021). "Help! I Need Somebody": Music as a global resource for obtaining wellbeing goals in times of crisis. *Frontiers in Psychology*, 12, Article 648013. <https://doi.org/10.3389/fpsyg.2021.648013>
- Greasley, A. E., & Lamont, A. (2011). Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae*, 15(1), 45-71. <https://doi.org/10.1177/1029864910393417>

- Greb, F., Schlotz, W., & Steffens, J. (2018). Personal and situational influences on the functions of music listening. *Psychology of Music, 46*(6), 763-794. <https://doi.org/10.1177/0305735617724883>
- Greenwell, B., Boehmke, B. C., & McCarthy, A. J. (2018). *A simple and effective model-based variable importance measure*. arXiv. <https://arxiv.org/abs/1805.04755>
- Hampton-Sosa, W. (2017). The impact of creativity and community facilitation on music streaming adoption and digital piracy. *Computers in Human Behavior, 69*, 444-453. <https://doi.org/10.1016/j.chb.2016.11.055>
- Hartigan, J., & Wong, M. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C, Applied Statistics, 28*(1), 100-108. <https://doi.org/10.2307/2346830>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Heggli, O. A., Stupacher, J., & Vuust, P. (2021). *Diurnal fluctuations in musical preference*. PsyArXiv. <https://doi.org/10.31234/osf.io/6e4yw>
- Henning, F., & Ruth, N. (2020). Save your artist! Der Einfluss moralischer Appelle von Musikschaaffenden auf die Akzeptanz von kostenpflichtigen Musikstreamingdiensten [Save your artist! The impact of musicians' moral appeal on acceptance of paid music streaming services]. *Jahrbuch Musikpsychologie, 29*, Article e48. <https://doi.org/10.5964/jbdgm.2019v29.48>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). *An introduction to statistical learning*. Springer.
- Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *Public Opinion Quarterly, 37*(4), 509-523. <https://www.jstor.org/stable/2747854https://doi.org/10.1086/268109>
- King, A. (2020). *Spotify says mental health is one of its fastest-growing 'genres' — Listening time up 50% in 2020*. Retrieved February 1, 2021 from <https://www.digitalmusicnews.com/2020/10/09/spotify-mental-health-playlists-listening-time/>
- Kreutz, G., Schubert, E., & Mitchell, L. A. (2008). Cognitive styles of music listening. *Music Perception, 26*(1), 57-73. <https://doi.org/10.1525/mp.2008.26.1.57>
- Krueger, J. (2009). Enacting musical experience. *Journal of Consciousness Studies, 16*(2-3), 98-123.
- Lades, L. K., Laffan, K., Daly, M., & Delaney, L. (2020). Daily emotional well-being during the COVID-19 pandemic. *British Journal of Health Psychology, 25*(4), 902-911. <https://doi.org/10.1111/bjhp.12450>
- Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications: Studies in classification, data analysis, and knowledge organization* (pp. 261-268). Springer. [https://doi.org/10.1007/978-3-540-78246-9\\_31](https://doi.org/10.1007/978-3-540-78246-9_31)
- Lee, D., Baker, W., & Haywood, N. (2020). *Coronavirus, the cultural catalyst*. Retrieved from <https://wim.hypotheses.org/1302>
- Lehman, E. T. (2020). "Washing hands, reaching out": Popular music, digital leisure and touch during the COVID-19 pandemic. *Leisure Sciences*, Advance online publication. <https://doi.org/10.1080/01490400.2020.1774013>

- Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology*, 102(1), 108-134. <https://doi.org/10.1348/000712610X506831>
- Luck, G. (2016). The psychology of streaming: Exploring music listeners' motivations to favour access over ownership. *International Journal of Music Business Research*, 5(2), 46-61. Retrieved May 10, 2021 from <https://musicbusinessresearch.files.wordpress.com/2012/04/volume-5-no-2-october-2016-luck2.pdf>
- Mastnak, W. (2020). Psychopathological problems related to the COVID-19 pandemic and possible prevention with music therapy. *Acta Paediatrica*, 109(8), 1516-1518. <https://doi.org/10.1111/apa.15346>
- Molnar, C. (2019). *Interpretable machine learning: A guide for making black box models explainable*. GitHub. <https://christophm.github.io/interpretable-ml-book/>
- Morissette, L., & Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 9(1), 15-24. <https://doi.org/10.20982/tqmp.09.1.p015>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, 9(2), Article e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Newen, A., De Bruin, L., & Gallagher, S. (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.
- North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of music in everyday life. *Music Perception*, 22(1), 41-77. <https://doi.org/10.1525/mp.2004.22.1.41>
- Palamar, J. J., & Acosta, P. (2021). Virtual raves and happy hours during COVID-19: New drug use contexts for electronic dance music partygoers. *The International Journal on Drug Policy*, 93, Article 102904. <https://doi.org/10.1016/j.drugpo.2020.102904>
- Parncutt, R. (2014). The emotional connotations of major versus minor tonality: One or more origins? *Musicae Scientiae*, 18(3), 324-353. <https://doi.org/10.1177/1029864914542842>
- Pettijohn, T. F., & Sacco, D. F., Jr. (2009). Tough times, meaningful music, mature performers: Popular Billboard songs and performer preferences across social and economic conditions in the USA. *Psychology of Music*, 37(2), 155-179. <https://doi.org/10.1177/0305735608094512>
- Petzold, M. B., Bendau, A., Plag, J., Pyrkosch, L., Mascarell Maricic, L., Betzler, F., Rogoll, J., Große, J., & Ströhle, A. (2020). Risk, resilience, psychological distress, and anxiety at the beginning of the COVID-19 pandemic in Germany. *Brain and Behavior*, 10(9), Article e01745. <https://doi.org/10.1002/brb3.1745>
- Randall, W. M., & Rickard, N. S. (2017). Reasons for personal music listening: A mobile experience sampling study of emotional outcomes. *Psychology of Music*, 45(4), 479-495. <https://doi.org/10.1177/0305735616666939>
- Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6), 1236-1256. <https://doi.org/10.1037/0022-3514.84.6.1236>
- Rhys, H. (2020). *Machine learning with R, the tidyverse, and mlr*. Manning.



- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. <https://doi.org/10.1037/h0077714>
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Sciences Information. Information Sur les Sciences Sociales*, 44(4), 695-729. <https://doi.org/10.1177/0539018405058216>
- Schiavio, A., van der Schyff, D., Cespedes-Guevara, J., & Reybrouck, M. (2017). Enacting musical emotions. sense-making, dynamic systems, and the embodied mind. *Phenomenology and the Cognitive Sciences*, 16, 785-809. <https://doi.org/10.1007/s11097-016-9477-8>
- Schramm, H. (2005). *Mood management durch Musik: Die alltägliche Nutzung von Musik zur Regulierung von Stimmungen* [Mood management with music: The everyday use of music for mood regulation]. H. von Halem.
- Seetharaman, P. (2020). Business models shifts: Impact of COVID-19. *International Journal of Information Management*, 54, Article 102173. <https://doi.org/10.1016/j.ijinfomgt.2020.102173>
- Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), 70-75. <https://doi.org/10.1073/pnas.1209023110>
- Sim, J., Cho, D., Hwang, Y., & Telang, R. (2020). Virus shook the streaming star: Estimating the COVID-19 impact on music consumption. *SSRN*, Article 3649085. <https://doi.org/10.2139/ssrn.3649085>
- Skandland, M. S. (2011). Use of MP3 players as a coping resource. *Music and Arts in Action*, 3(2), 15–33. Retrieved July 27, 2021 from <http://hdl.handle.net/10036/3964/>
- Sloboda, J. A. (2010). Music in everyday life: The role of emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Series in affective science. Handbook of music and emotion: Theory, research, applications* (pp. 493–514). Oxford University Press.
- Slonim, N., Aharoni, E., & Crammer, K. (2013). Hartigan's k-means versus Lloyd's k-means: Is it time for a change? In F. Rossi (Ed.), *Proceedings of the twenty-third international joint conference on artificial intelligence* (pp. 1677–1684). AAAI Press.
- Small, C. (1998). *Musicking: The meanings of performing and listening*. Wesleyan University Press.
- Sony Music. (2021). *Sony ab jetzt in Berlin* [Sony from now on in Berlin]. Retrieved July 27, 2021 from <https://www.sonymusic.de/sony-music-ab-jetzt-in-berlin/>
- Spotify. (2021a). *Discover Spotify's features*. Retrieved February 1, 2021 from <https://developer.spotify.com/discover/>
- Spotify. (2021b). *Spotify Programm EQUAL startet mit Zoe Wees als erster Künstlerin* [Spotify program EQUAL launches with Zoe Wees as the first artist]. Retrieved July 27, 2021 from [https://spotify\\_presse.prowly.com/138158-spotify-programm-equal-startet-mit-zoe-wees-als-erster-kunstlerin-des-monats](https://spotify_presse.prowly.com/138158-spotify-programm-equal-startet-mit-zoe-wees-als-erster-kunstlerin-des-monats)
- Statista. (2020). *Marktanteile der einzelnen Anbieter an den zahlenden Abonnenten von Musikstreaming weltweit im 1. Quartal 2020* [Market shares of individual providers of paid music streaming subscribers worldwide during the 1st quarter of 2020]. Retrieved July 27, 2021 from <https://de.statista.com/statistik/daten/studie/671214/umfrage/marktanteile-der-musikstreaming-anbieter-weltweit/>

- Thompson, C., Antal, D., Parry, J., Phipps, D., & Wolff, T. (2021). *Spotifyfy: R wrapper for the 'Spotify' web API*. GitHub. <https://doi.org/10.5281/zenodo.4946780>.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series A*, 63(2), 411-423. <https://doi.org/10.1111/1467-9868.00293>
- Universal Music. (2021). *Universal Music sichert Zukunft der EMI in Deutschland* [Universal Music ensures future for EMI in Germany]. Retrieved July 27, 2021 from <https://www.universal-music.de/company/presse/universal-music-sichert-zukunft-der-emi-in-deutschland-211223>
- Vandenberg, F., Berghman, M., & Schaap, J. (2020). The “lonely raver”: Music livestreams during COVID-19 as a hotline to collective consciousness? *European Societies*; Advance online publication. <https://doi.org/10.1080/14616696.2020.1818271>
- van der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, A. (2018). Musical creativity and the embodied mind: Exploring the possibilities of 4E cognition and dynamical systems theory. *Musicae Scientiae*, 1, 1-18. <https://doi.org/10.1177/2059204318792319>
- van Goethem, A., & Sloboda, J. (2011). The functions of music for affect regulation. *Musicae Scientiae*, 15(2), 208-228. <https://doi.org/10.1177/1029864911401174>
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Burlington.
- Yeung, T. Y. C. (2020). Did the COVID-19 pandemic trigger nostalgia? Evidence of music consumption on Spotify. *SSRN*, Article 3678606. <https://doi.org/10.2139/ssrn.3678606>
- Zillmann, D. (1988). Mood management: Using entertainment to full advantage. In L. Donohew, H. E. Sypher, & E. T. Higgins (Eds.), *Communication, social cognition, and affect* (pp. 147–171). Lawrence Erlbaum Associates.
- Zillmann, D. (2000). Mood management in the context of selective exposure theory. *Annals of the International Communication Association*, 23(1), 103-123. <https://doi.org/10.1080/23808985.2000.11678971>