COGNITIVE APPROACHES TO MUSIC PREFERENCE USING LARGE DATASETS

Bridging the gap: cognitive approaches to musical preference using large datasets

Michael Barone

July 18, 2017

McMaster University MASTER OF SCIENCE (2017) Hamilton, Ontario (Psychology, Neuroscience, & Behaviour)

TITLE: Bridging the gap: cognitive approaches to music preference using large datasets
AUTHOR: Michael David Barone, B.A. (McMaster University)
SUPERVISOR: Professor Matthew H. Woolhouse
NUMBER OF PAGES: vii, 49

This thesis examines whether research from cognitive psychology can be used to inform and predict behaviours germane to computational music analysis including genre choice, music feature preference, and consumption patterns from data provided by digital-music platforms. Specific topics of focus include: information integrity and consistency of large datasets, whether signal processing algorithms can be used to assess music preference across multiple genres, and the degree to which consumption behaviours can be derived and validated using more traditional experimental paradigms. Results suggest that psychologically motivated research can provide useful insights and metrics in the computationally focused area of global music consumption behaviour and digital music analysis. Limitations that remain within this interdisciplinary approach are addressed by providing refined analysis techniques for future work.

Abstract

Using a large dataset of digital music downloads, this thesis examines the extent to which cognitive-psychology research can generate and predict user behaviours relevant to the distinct fields of computer science and music perception. Three distinct topics are explored. Topic one describes the current difficulties with using large digital music resources for cognitive research and provides a solution by linking metadata through a complex validation process. Topic two uses this enriched information to explore the extent to which extracted acoustic features influence genre preferences considering personality, and mood research; analysis suggests acoustic features which are pronounced in an individual's preferred genre influence choice when selecting less-preferred genres. Topic three examines whether metrics of music listening behaviour can be derived and validated by social psychological research; results support the notion that user behaviours can be derived and validated using an informed psychological background, and may be more useful than acoustic features for a variety of computational music tasks. A primary motivation for this thesis was to approach interdisciplinary music research in two ways: (1) utilize a shared understanding of statistical learning as a theoretical framework underpinning for prediction and interpretation; and (2) by providing resources, and approaches to analysis of "big data" which are experimentally valid, and psychologically useful. The unique strengths of this interdisciplinary approach, and the weaknesses that remain, are then addressed by discussing refined analyses and future directions.

Acknowledgements

Contents

List of Figures

List of Tables

Glossary

API Application Program Interface: A set of subroutine definitions, protocols, and tools for building application software.. 2, 8–13, 27, 31, 32

C$_{CV}$ Cardinality Consistency value: A validation metric which considers the number of elements in a set.. 9–11

CV Consistency Value: A value produced by GRAIL's validation procedures. 9, 10

GRAIL General Recorded Audio Identity Linker: An open-access API designed to check and validate music metadata.. vii, 7–13, 27, 30, 31, 33, 45

HDI Human Development Index: A composite statistic of life expectancy, education, and per capita income indicators.. vii, 25–30, 32

HDR Human Development Report: An annual report by the United Nations which publishes statistics on global indicators of health, education, income, and equality. 3

iHDI Inequality-adjusted Human Development Index: A derivative of the HDI that considers resource disparity across all people within a region.. vii, 26–29, 32

ISRC International Standard Recording Code: A code for uniquely identifying sound recordings and music video recordings. The code was developed by the recording industry.. 11–13, 30

LD Levenschtein Distance: A method to calculate text similarity. 10

O_{CV} Overall Consistency Value: A validation metric derived from an aggregation of all other consistency checks.. 9, 10

P_{CV} Position Consistency Value: A validation metric which considers track positioning.. 9–11

SM_{CV} SM_{CV}: String-matching Consistency Value: A validation metric which considers string similarity.. 9–11

UN United Nations: An intergovernmental organization to promote international co-operation and to create and maintain international order.. 3, 25, 32

1   Introduction

Understanding the origins of human music behaviour has been a historic curiosity of the life sciences since Darwin [1, 2]; however, despite this longstanding interest conclusive understanding of music cognition and preference remains obscure [3, 4]. New approaches to cognitive science come at a time when digital music services such as Apple, Google, Amazon, and Spotify are seeing unprecedented growth within the music industry; company evaluations are higher than the yearly sales for US market sales for all other media combined.[1] This success comes during an intense period of technological development on the "The Internet of Things" [5]: the collection of, and access to, data on any conceivable topic. The manifestation of this technological renaissance within music is best embodied by the computationally focused field of Music Information Retrieval (MIR) [6]; MIR research has been principally responsible for the generation of publicly available databases containing metrics for virtually the entire corpus of recorded music. Although the MIR community is keen to highlight the potential benefits of informed psychological perspectives of music processing by humans [6, 7], surprisingly, relatively little psychology research exists which utilizes these data sources [8]. Sparsity of interdisciplinary research may be, in part, due to several conceptual and methodological contentions that are controversial across the two fields [9]. However, using MIR techniques may be the most appropriate current method available to address unsolved questions with respect to human preferences and it is incumbent upon psychologists to recognize [9, 8] and contribute [10, 11] to these approaches in order produce meaningful results which align closely with experimental support.

The recent surge of open-access data provided by digital-music streaming services may address lacunae in music cognition research and provide valuable contributions to computational analysis of music; MIR research has voiced the need for more contextual user metadata grouped into 4 general categories: demographics, response characteristics, needs/uses, and search behaviours [12]. This thesis examines differences in search behaviours and demographics with respect to genre preference using digital download data from 17 million users from over 36 countries enriched with extracted acoustic features for millions of tracks. The gap between music cognition and MIR research is addressed by developing and testing cognitively meaningful user metadata using an interdisciplinary approach: Cognitively-based Music Information Retrieval (CogMIR). Although CogMIR is nascent, combining advanced computational skills common in MIR with necessary knowledge of the neurological processing (and bias) of the human perceptual system is crucial for understanding global music behaviour with the highest degree of external validity. The theoretical background considers shared research areas between MIR and music cognition–using statistical learning and emotional affect–to inform methodology and interpretation. How these seemingly disparate fields can become more interdisciplinary is utilized as the foundation for the methodology used to investigate the relationship between acoustic features found in a user's preferred genres with features when seeking out lesser-preferred, secondary music genres. Furthermore, it supports the notion of behavioural measures of music behaviours motivated by cognitive research for recommendation, classification, and extraction tasks [10].

MIR seeks to derive meaningful features from music automatically [6], and this focus has advanced research relevant to music cognition specific to this thesis; acoustic feature extraction [13, 14] is germane to auditory signal processing [15, 16]; personalized music recommendation [17, 18] can be informed by mood and personality research [19, 20, 21]; and user behaviour analysis [12] has implications for cross-cultural social psychology [22, 23]. In particular, a significant portion of MIR research relies on the extensive use of signal processing, in tandem with statistical learning procedures, for classification and prediction of stochastic data [24]. MIR has generated hundreds of feature extractors provided as signal processing libraries [13, 25, 14], including both well-understood features (e.g. beat extraction and tempo, Bark/Mel-band cepstral coefficients and timbre [15]) and unsolved behavioural phenomena (e.g. emotion [26], genre categorization [27], and "danceability" [28]). The expansion of features to include more perceptual categories is surprising given that the psychological validity of extracted audio features is an open question [9, 10, 29]. Research demonstrating the influence of acoustic features upon musical preference, may in part, help to legitimize their perceptual existence; MIR extractors, foregrounded by informed psychological research, present an opportunity for both disciplines to advance the scientific understanding music perception and preference.

Statistical learning, the ability of humans and other animals to acquire implicit knowledge about

---

[1]

the world through the extraction of statistical regularities within their environments [24], has more recently become an area of intense study in cognitive sciences [30]. Cognitive and neurological processes using statistical learning have been modelled in vision [31, 32, 33], haptics [32, 34, 35], language [36], and audition [32, 37]. Neurological evidence for statistical learning has been found when listening to music; levels of activation in reward areas of the brain (e.g. nucleus accumbens) are proportional to the degree to which music is liked [38]; the behavioural effect of music on risk-taking co-varies with brain activation in the amygdala and the dorsal striatum ([39]), key brain regions associated with the calculation and assessment of value. Research demonstrating this phenomenon is robust [40, 16, 41], and neurological evidence supports this notion for music with respect to expectation violation [42]—mismatch negative responses are larger for violations within a musicians preferred genre [43] and N400 violation responses in language tasks [44]. A possible mechanism underpinning music-feature influence is posed with reference to Aniruddh Patel's research involving music, language, and statistical learning (for additional neurological evidence concerning statistical learning of language, see [45, 46]). In order to account for the finding that the prosodies of English and French are reflected in the rhythms and melodies of these countries' respective instrumental music, Patel proposes a "direct-route" hypothesis, in which "statistical learning of the prosodic patterns of speech creates implicit knowledge of rhythmic and melodic patterns in language, which can in turn influence the creation of rhythmic and tonal patterns in music" ([47], pg. 3043. In other words, statistically acquired sound-pattern knowledge from the language domain results in the rhythmic and melodic modification of music. In short, music listening gives rise to distributed neurological operations in which functionally differentiated networks are simultaneously activated. For a review of research relating to the induction of mood through music see [48].

Although acoustic feature extraction is used to varying degrees in both disciplines (e.g. MIR-Toolbox [13] is used extensively in cognitive research), computational analogs for the complex neurological processes rarely align [9, 8]. Bridging the gap by providing music features derived from state-of-the-art learning procedures alone does not entice cognitive research; determining whether a training network learned intended features is difficult to interpret without an informed theoretical framework [24]. To that end, a strong understanding of the statistical learning processes involved in the neural and cognitive processing of auditory stimuli, alongside other foundational areas of psychology (specific to this thesis: personality, mood, affect induction, and social research), are important for modelling the complex system of global music preference using digital platforms. Furthermore, open-access MIR datasets with information useful for controlled cognitive experiments are relatively scarce. A resource which links digital music catalogues to open-access digital databases and evaluates data integrity may encourage traditional psychologists to explore MIR resources and techniques for future study.

This thesis investigates three topics under the framework of psychological research and statistical learning. The first topic (Section 2.2) explores methodological concerns and data quality of large open databases of music consumption downloads. Using a library of ca. 16 million songs available for digital consumption, Section 2.2 also addresses these methodological concerns by validating and linking metadata from a variety of MIR resources, providing the results as a publicly accessible Application Program Interface (API). Informed by personality and mood research, topic two (Section 3.1) utilizes this validated, enriched metadata to explore the degree to which acoustic features in a user's preferred genre influence track/song selection when seeking out non-favourite genres [49]. For example, might someone whose preference is for Metal also gravitate towards relatively dynamic, "high-octane" Country or Blues music (assuming, of course, that those genres are not mutually exclusive; [50]). Although conceptually straightforward, this question addresses active research areas within the fields of music cognition and MIR, and, to some extent, highlights current limitations within both. Firstly, the phenomenon of features within a preferred genre influencing song selection within secondary musical styles falls under the general topic of "cognitive leakiness"—a notion explored in depth in the area of consumer choice and commerce (e.g. [51]), but less so in music perception. Secondly, musical features extracted from audio by necessity utilize MIR techniques [13]. The psychological reality of extracted audio features is an open question [10], and therefore research demonstrating their influence upon musical preference, may in part, help to legitimize their perceptual existence. Topic three (Section 3.2.2) develops two distinct user behaviours relevant to CogMIR research, eclecticism and temporality [23]. The validity of these measures considers social psychological research regarding cross-cultural consumption [22, 23] i.e., that general patterns of music consumption and preference are influenced by social and economic

capital and can be predicted using statistics produced by the United Nation's annual Human Development Report.[2] The analysis considers the extent to which measures of human development (e.g. measures of income, education, health, and gender equality) developed and collected in the United Nations (UN) annual Human Development Report (HDR) can be used to explain eclecticism at the country- and user-level. The interpretation and discussion of results supports the theoretical framework first addressed in [22]. The interdisciplinary approach affords several unique advantages because it addresses methodological needs voiced by both MIR (e.g. psychologically relevant models of user behaviours) [8, 12] and social psychological (e.g. limited sample sizes, collection methods, and cultural diversity) communities [52, 53]. Future directions consider new behaviours for MIR classification, and recommendation tasks.

---

[2]http://www.hdr.undp.org/

2 Data Description

2.1 Nokia Music Database

This research utilizes a global music-consumption database as a primary resource for scientific inquiry. Beginning in 2012, a data-sharing agreement was initiated between McMaster University and the Nokia Corporation. The Nokia Corporation provided information pertaining to 1.3 billion song downloads onto mobile phones and personal computers (PC). The Nokia DB contains song downloads from 33 countries and regions from every major continent. In total, the database consists of over 1.36 billion track-downloads from November 2007 to September 2014. Following account activation through mobile phone or PC, users could explore artists and tracks via online stores. Newly purchased Nokia mobile phones included one year of free downloads as part of their activation. Users were able to access approximately 36 million tracks, written and/or performed by over one million artists (see Figures 3 and 4 for country coverage during the first and last years of international consumption).

In order to carry out the research reported in this thesis, service-use activities were stored in a relational database management system and queried using the open-source MySQL 5.1 implementation of SQL [54]. Each download's metadata includes information such as track name, artist, album, anonymous user identifier (ID), date, local time, country, and artist-level genre tags. In total, one of 63 genre tags, supplied by the record label were applied to each Artist. Genres range from mainstream (Country, Pop, Rock) to relatively obscure (Ambient, Flamenco, Khaleeji). Genre information was provided by record labels. In total, one of 63 distinct genre tags were assigned to each of the 1,078,485 artists in the Nokia DB. 36,907,012 tracks were available for download in each region, corresponding to 3,920,965 million releases (albums). The mean and median amount of tracks per artist is 34.2 ($\sigma = 6236.2$)) and 10 respectively. For an exhaustive list of genres, refer to Appendix 7. For additional information and research concerning Nokia DB, see [23, 55, 50, 21, 49].

User information was anonymized prior to data-sharing in order to ensure privacy. The only direct information regarding users contained in the Nokia DB is the timestamp and country or region of download. In order to ensure security for users in sensitive areas in the Middle East, the country-code AE (United Arab Emirates) represents a digital store for countries in close proximity including: Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Palestine, and Qatar. Additionally, the Argentina store includes Chile and Uruguay. 17,182,470 accounts downloaded at least one track during the 7-years of recorded service-availability. Total accounts by country ranged from 77 in Vietnam to 6,946,307 in India. On average, users downloaded 79 ($\sigma^2 = 785.4$) songs, and used the music downloading service for 47 ($\sigma^2 = 85.3$) days. Average retention by country was calculated by taking the average number of days between a user's first and last download. For an exhaustive list of consumption details by country, refer to Table 1.

Preliminary descriptive analyses on total tracks and artists by genre indicate differences in production patterns by content creators. Total number of artists and tracks grouped by genre were counted and converted to proportions; Figure 1 shows the top 10 artists and tracks attributed to a genre in the catalogue. The tenth category, Other, comprises all 53 other genre categories. The majority of tracks in the Nokia DB are Pop, but the majority of artists are Dance suggesting that artists with different genre affiliations release different amounts of tracks.

Next, artists were ranked according to their total download count. Figure 2 displays the top 5000 most downloaded tracks in the Nokia DB, converted to a proportion (y-axis) and ordered by rank (x-axis). The distribution of music downloads ranked by artist is highly skewed to the most popular performers, and is consistent with research from data provided by other digital services [56, 57]. Qualitatively (and represented as a dashed line in Figure 2), the shape of this relationship is similar to a Zipf distribution [58]. Zipf distributions have been found with respect to music classification, physiological responses, aesthetics, and acoustic feature extraction [59, 27]. To the author's knowledge, no research within MIR or music cognition has demonstrated a Zipf distribution in music with respect to consumption.

Nokia Music's service developed over time, and as a result, all 33 regions did not enter into the database simultaneously. The service first opened in the Great Britain in September 2007, with Venezuela joining the service last in September 2013. Figures 3 and 4 are global heat maps, coloured by the amount of yearly downloads for each region. Total downloads are normalized by taking the log of the yearly count in order to visually discern differences in activity. China, India, Finland, Great Britain and Brazil were the most active regions using the service.

Country	Users	Total Downloads	Average Downloads	Retention
IN	6946307	777025337	111.9	43.2
CN	6752297	189959130	28.1	50.4
FI	223519	72861408	326.0	83.7
GB	552784	60331294	109.1	49.7
BR	202117	58217105	288.0	57.8
DE	258500	45206821	174.9	56.3
ID	693333	38038506	54.9	40.8
MX	68178	25887836	379.7	84.9
ZA	155252	17543739	113.0	80.4
AU	142161	15200180	106.9	53.5
SG	50211	12430761	247.6	74.6
RU	182419	7523492	41.2	32.4
IT	229951	7218080	31.4	31.2
CH	34986	5937698	169.7	75.8
SE	48638	4826545	99.2	45.5
TR	81203	4222706	52.0	33.2
ES	94735	3138141	33.1	36.5
AE	58013	2810923	48.5	24.0
CL	10447	2656640	254.3	74.0
NL	47183	2612654	55.4	47.5
AT	19765	2494002	126.2	69.8
FR	153150	1615792	10.6	26.4
MY	5567	821133	147.5	50.6
PL	51308	780930	15.2	37.8
US	38160	726936	19.0	32.1
NO	34094	639335	18.8	43.8
IE	19323	477753	24.7	45.6
PT	27228	312090	11.5	29.1

Table 1: Descriptive statics of users for each country within Nokia DB. Columns from left-to-right for each country include: a two-letter ISO-003 country code, number of unique accounts, total number of downloads, average downloads per account, and retention (average number of days for an active account on the service). Countries with less than 1000 users (CA,VN,TH, SA, and AR) have been excluded from the table.

Of the ca. 36 million tracks available in the Nokia DB catalogue, 9,036,675 tracks were enriched with high-level acoustic features extracted through a variety of signal processing techniques [14, 60], and musicological information annotated by community driven open-access resources including: MusicBrainz [61, 62, 25][3], Spotify,[4] and The Echo Nest,[5]. In the context of this work, high-level features (i.e. Danceability) refer to a single value derived from an entire track, whereas low-level features (i.e. Mel frequency cepstral coefficients) are arrays of values extracted over small windows (typically seconds) of time. High-level acoustic features utilized in this thesis include:

Acousticness. Value representing the probability that a track was created using acoustic instruments, including voice. Float; range, 0-1.

Danceability. A track's "foot-tapping" quality, based on tempo, rhythm stability, beat strength and isochrony. Float; range, 0-1.

Duration. The duration of a track in seconds as calculated by the Spotify analyzer. Float; maximum value, 6060 seconds.

Energy. A perceptual estimation of frenetic activity throughout a track. High-energy tracks have increased entropy, and tend to feel fast, loud, and noisy (e.g. Death Metal). Float; range, 0-1.

---

[3]https://musicbrainz.org/
[4]https://www.spotify.com
[5]https://the.echonest.com/
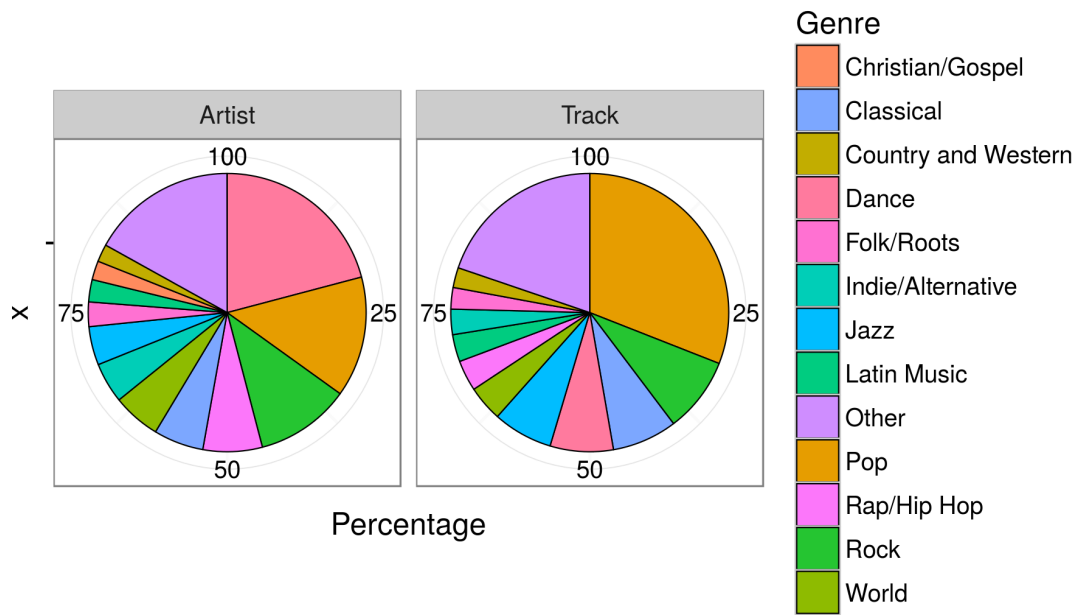
Figure 1: Percentage of artist and tracks by genre. Dance artists are the most represented artist by genre, and Pop is the most represented track by genre. Genres downloaded < 2% are aggregated into the "Other" genre
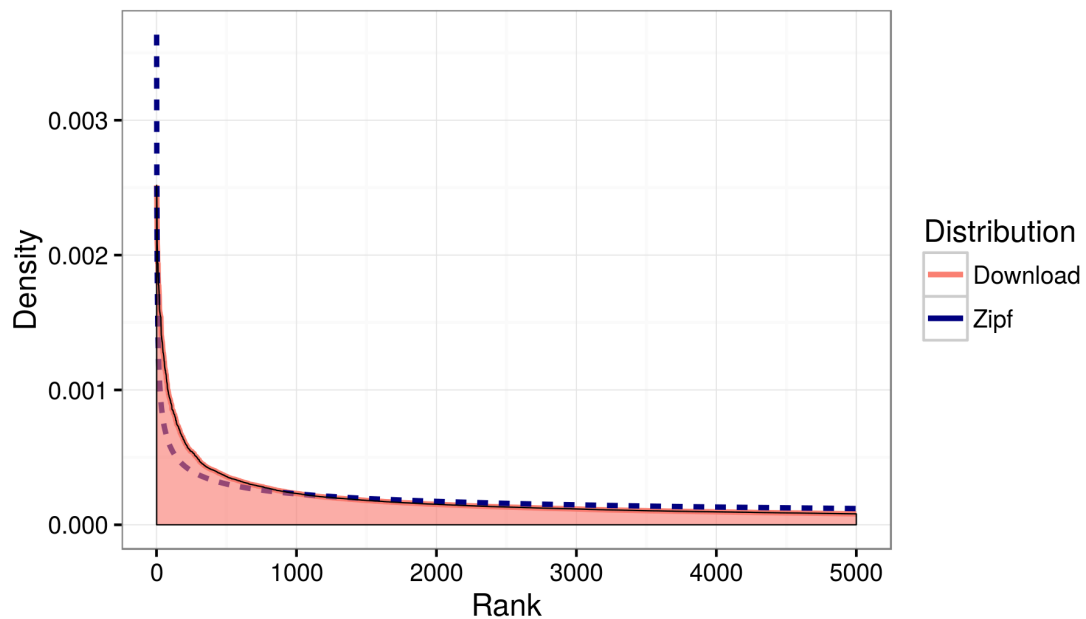
Figure 2: Zipf density distribution and total downloads

Instrumentalness. Value representing the probability that a track was created using only instrumental sounds, as opposed to speech and/or singing. Float; range, 0-1.

Liveness. Value representing the probability that a track was recorded in the presence of an audience rather than in a studio. Float; range, 0-1.

Loudness. The average loudness of a track in decibels. Loudness is the psychological correlate of signal amplitude.

Speechiness. Value representing the presence of spoken words in a track, e.g. talk show, audio book, poetry, rap. Float; range, 0-1.
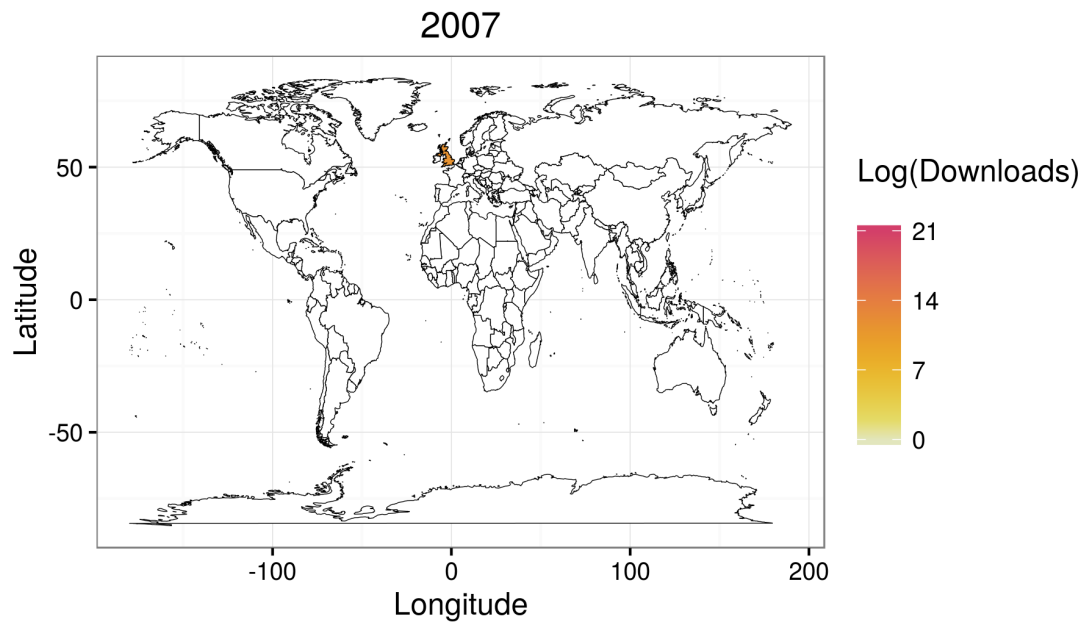
2007

Figure 3: Nokia Consumption Map 2008. Colour intensity illustrate total downloads by country in 2007. Values are converted to log for scaling
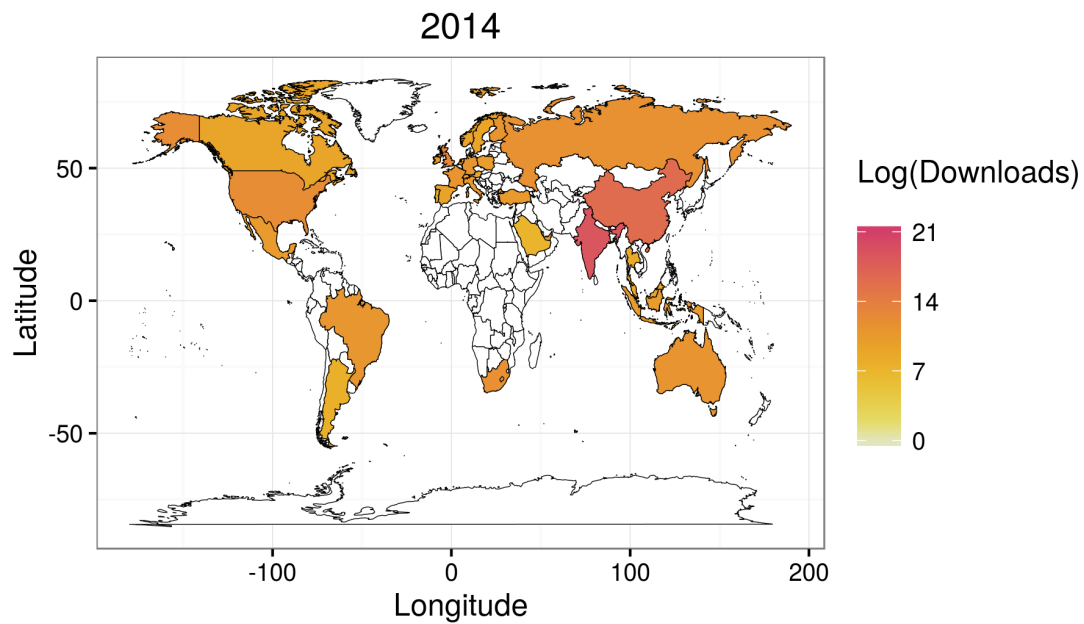
2014

Figure 4: Nokia Consumption Map 2014. Colour intensity illustrate total downloads by country in 2008. Values are converted to log for scaling

Tempo. The estimated tempo of a track in beats per minute. Float; range, 0-294.

Valence. A perceptual estimation of a tracks positive/negative affect, e.g. happy and cheerful, or sad and depressed. Float; range, 0-1.

Enrichment was made possible through a metadata linkage, storage, and sharing process created for this thesis known as the General Recorded Audio Identity Linker (GRAIL). GRAIL represents a significant contribution to CogMIR-related research, and is utilized for all analyses in this thesis. Motivations, linking process, and coverage are now outlined.

2.2 The General Recorded Audio Identity Linker (GRAIL)

2.2.1 Motivation

Inconsistent music metadata is not merely an esoteric problem for MIR. As of 2012, 42% of music classification algorithms use open-access data [63], but resources that verify data accuracy and publish linked music API IDs have been absent from the field. This is important because using a variety of open-access databases for research could be advantageous, but rarely discussed because linking metadata is a complex problem. For example, [64] attempts to link metadata to the Codaich Database[6] and found exact matches for only 22% of popular songs and 0% for classical. [65] using the same classification algorithm across five standard collections found genre identification accuracy to range widely from 48-87%, reflecting significant differences in music metadata. [66] supports the notion of a metadata linking resource, stating that it should, "...be aware of the existence of different schemas to store and exchange music metadata".

Due to services optimizing their data architectures for specific purposes, there is a paucity of consistent music metadata within both academia and industry [66]. Metadata faults can be further compounded by incomplete or inconsistent data management practices [64], and, in part, to the changing availability of data over time. For example, approximately 50% of the Million Song Dataset's [67] linkages to The Echo Nest were no longer accessible immediately prior to its deprecation in 2014,[7] due to the former service not being consistently updated. Arguably, therefore, in order to reduce data inaccuracies that may accrue over time, it is important for MIR to develop systems that automatically cross-validate, check, and update information from multiple resources.

In order to address these issues, a modular linking system capable of being augmented at any time to access services is utilized. Furthermore, the architecture constantly re-validates music information accuracy, and can store results in a relational database system for later use. A resource of linked IDs from digital music services is valuable for MIR for the following reasons: (1) it can support historic research efforts by linking data collected from deprecated services (e.g. Rdio) to operational ones (e.g. Spotify); (2) new research is enabled by linking data from ontologically distinct resources (e.g. IDs for lyric analysis linked to IDs with consumption information [68]); and (3) provide a public resource for cross-validation of metadata provided by music services.[8] This work addresses the complex challenges of music-metadata organization, referred to above, and provides our solution in the form of an public API. To the author's knowledge, GRAIL is the only public resource which includes data integrity metrics across multiple resources.

2.2.2 Web-Crawlers and Workflow
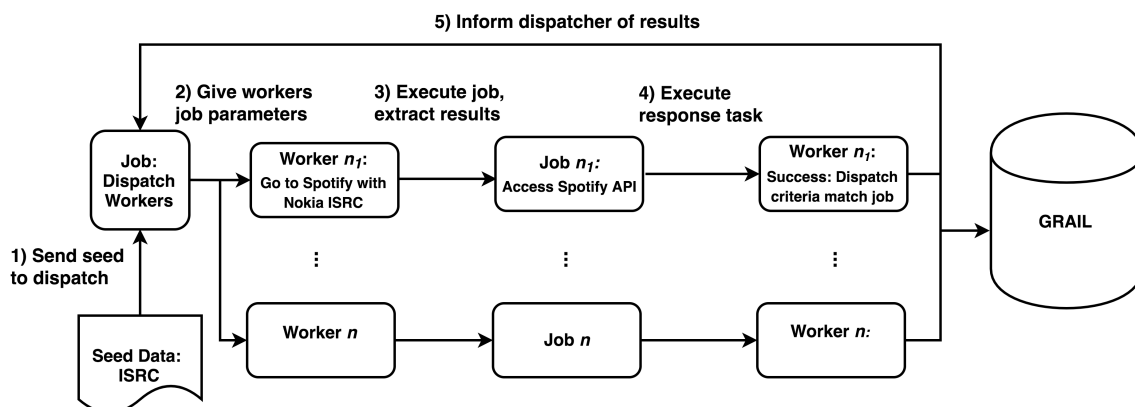
Figure 5: Web crawler process. Workers are provided seed-data including necessary API endpoints, validation procedure, and any necessary data required to execute linkages. Workers access an API, conduct the validation procedure, and update GRAIL's database with new information.

---

[6]http://jmir.sourceforge.net/manuals/jMusicMetaManager_manual/introduction.html
[7]https://techcrunch.com/2014/03/07/spotify-echo-nest-100m/
[8]https://www.berklee.edu/news/fair_music_report

To constantly check and validate API metadata, we developed a web-crawler[9] that asynchronously checks metadata, collects and updates IDs, and cross-validates information from strong linkages to test for data inconsistencies. Figure 6 illustrates a web-crawler job sent to the Spotify API. For this crawl, the scheduler is seeded with Nokia Music metadata, including artist, release, and track information (1). The scheduler spawns a worker, and provides endpoint parameters to crawl Spotify's API (2); e.g. rate-limit restrictions, and what criteria will be employed). The worker executes its task and calculates consistency values (3) (see Section 2.2.3 for validation process). Linked IDs from responses are ingested into GRAIL, alongside calculated consistency values (CV) and timestamp information (4). Lastly, the worker updates GRAIL with job results, and seeds the scheduler with any new information that requires validation based on these results (5). As the crawler verifies any metadata changes, the CVs of the linkages are updated. Some services, such as The Echo Nest, provided additional API IDs; when available, these IDs are also ingested into GRAIL after calculation of consistency.

2.2.3 Consistency Analysis

The metadata linkage analysis is based on three criteria: cardinality (C); string-matching (SM); and position (P). Using these criteria, CVs (C_{CV}, SM_{CV}, and P_{CV}) ranging from zero to one are calculated for artist, release and track IDs. An overall, summary statistic (O_{CV}) is also calculated as a weighted average of C_{CV}, SM_{CV}, and P_{CV}. As APIs do not offer a universal set of metadata or search methods, it is not always possible for every criterion to be used at every comparison level.[10] Table 2 illustrates the criteria ideally used. In the interests of transparency, documentation regarding which criteria are utilized for each API are provided on the GRAIL website.[11] Information will continue to be updated as GRAIL expands to include other services. Table 3 provides a release analysis between two hypothetical albums.

CV	Artist	Release	Track
SM_{CV}	✓	✓	✓
C_{CV}	✓	✓	—
P_{CV}	—	✓	✓
O_{CV}	✓	✓	✓

Table 2: Table showing CVs calculated for artist, release, and track. Overall CVs (O_{CV}) are calculated for each level based on a weighted average.

---

[9]www.github.com/digitalmusiclab/\Gls{GRAIL}_web_crawler
[10]For example, some APIs (e.g. Last.FM, MusixMatch) do not provide track position metadata.
[11]api.digitalmusiclab.org

Order	X Release	Y Release	SM_{CV}	C_{CV}	P_{CV}
1.	im ahab	I am aahb	0.75	0.8	1.0
2.	Island	Blood and thunder	1.0	0.8	0.6
3.	Sea beast	seabeast	0.94	0.8	1.0
4.	blood & thunder	Island	0.875	0.8	0.6
5.	Iron tusk	—	0.38	0.8	0.6

Table 3: Example of release matching process using two APIs X and Y (e.g. Spotify and MusicBrainz). Release and artist names are first compared prior to assigning track consistency values. When releases agree on artist and release strings, track information regarding the releases is compared. In this example, the 3 columns on the left disagree in the following ways: X has a cardinality of 5, whereas Y's cardinality is 4; and track orders 2 and 4 within X are reversed within Y. The 3 columns on the right show CVs calculated per track for each criterion (C_{CV}, SM_{CV} and P_{CV}, described below).

String-Matching Consistency Value ($SM_{\mathbf{CV}}$). Levenshtein distances (LDs) [69] are calculated as ratios for tracks, releases, and artists. LD compares character similarity across two strings and provides a value ranging from 0 to 1. LD is a common procedure in string-matching processes, and has been used previously to link music metadata into MusicBrainz[62]. Track, release, and artist LDs are calculated on a pairwise basis, and expressed as SM_{CV} values (see Table 3, Column 4). For example, in Table 3 the 3rd track under X is "Sea Beast"; whereas under Y it is "seabeast", resulting in a LD/SM_{CV} of 0.94.

Cardinality Consistency Value ($C_{\mathbf{CV}}$). C_{CV} values are calculated by dividing the minimum cardinality by the maximum; see Equation 1. For example, Table 3 compares two releases, X and Y, with different cardinalities (5 versus 4). The C_{CV} column in Table 3 (Column 5) is therefore: $\frac{4}{5} = 0.8$. This value is attributed to each track within both releases. In a similar manner, artist-level cardinality is calculated by comparing the number of distinct tracks attributed to each artist across resources. As Table 2 shows, the cardinality criterion is not used with respect to track matching.

$$C_{CV}(X,Y) = \frac{C\min}{C\max} \tag{1}$$

Position Consistency Value ($P_{\mathbf{CV}}$). When position metadata are available, P_{CV} is calculated for tracks and releases. Similar to the C_{CV}, P_{CV} is calculated for the track with the maximum LD comparison; the linked attribute from the API with the maximum position is used as the denominator against the attribute with the minimum position; see Equation 2. In Table 3, Track 2 in X is to Track 4 in Y, the strings exactly match. The P_{CV} for Track 2 in X is therefore 0.6, calculated by dividing the minimum track position (2) by the maximum (4).

$$P_{CV}(X,Y) = 1 - \frac{|X_{Order} - Y_{Order}|}{C\max} \tag{2}$$

Overall Consistency Value ($O_{\mathbf{CV}}$). O_{CV} values are calculated for artist, release and track by taking a weighted average from all available CVs. As shown in Table 2 Artist O_{CV}s are calculated by taking the weighted average of SM_{CV} and C_{CV}; track O_{CV}s are calculated by taking the weighted average of SM_{CV} and P_{CV}. For a release O_{CV}, a weighted average of all CVs is possible. Using the example in Table 3, the overall release score for this comparison would be: $\mu(SM_{CV}, P_{CV}, C_{CV})$.

Figure 6 is a schematic representation of GRAIL, and the way in which jobs are processed and information is ingested. Jobs are separated into validation steps (numeric circles) and ingestion steps (letter circles), and are described below:
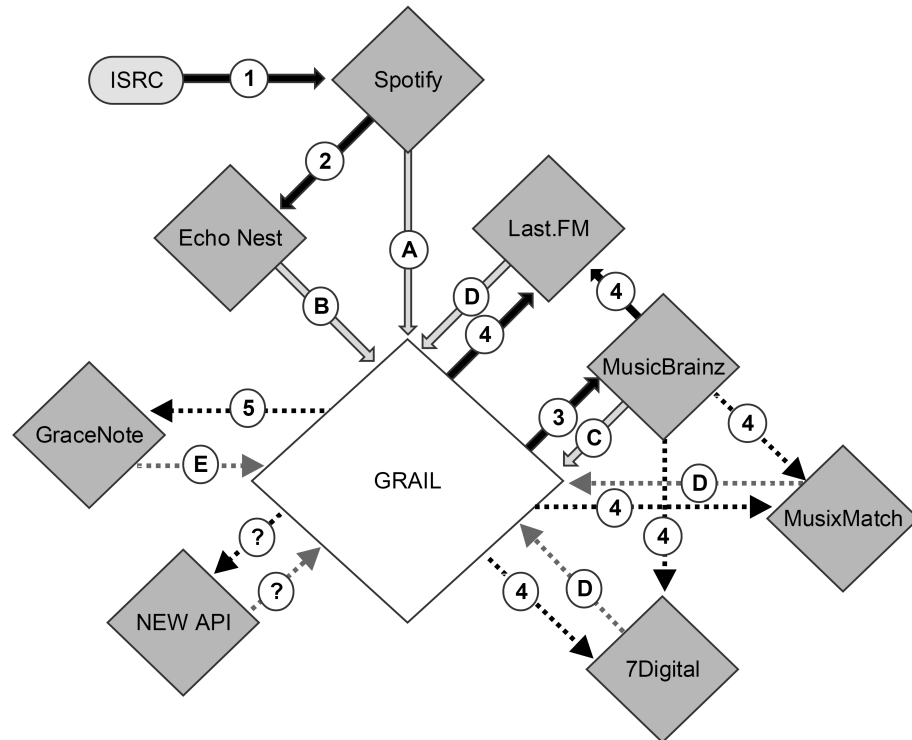
Figure 6: Workflow llustrates how crawlers access targeted APIs (gray diamonds), what validation steps are used to access APIs (black lines), and what metadata is ingested into GRAIL (gray lines). Solid lines represent active GRAIL crawls, and dotted lines represent crawls currently in development.

Validation Steps

Step 1. International Standard Recording Code (ISRC) seed-data from Nokia DB accesses metadata from the Spotify API. Artist, release, and track consistency calculations are conducted using SM_{CV}, P_{CV}, and C_{CV}.

Step 2. Spotify tracks IDs were used to link to Rosetta Stone artist IDs.[12] Due to Spotify's purchase of The Echo Nest in 2014 and subsequent discontinuation, consistency validation between this two services was not possible (or necessary).

Step 3. Nokia Music artist, release, and track string names are used to query the MusicBrainz API. Artist, release, and track consistency calculation using SM_{CV}, P_{CV}, and C_{CV}.

Step 4. Nokia Music artist, release and track string names, and linked MusicBrainz IDs, from ingestion step C are being used to conduct artist, release, and track consistency calculations using SM_{CV}, P_{CV}, and C_{CV}.

Step 5. In due course, Nokia Music artist, release, track string names will be used to access GraceNote's metadata via text search. As above, artist, release, and track consistency calculations are conducted using SM_{CV}, P_{CV}, and C_{CV}.

Ingestion Steps

Step A. Spotify artist, release, and track IDs are ingested into GRAIL. Crawlers check if an ID has changed since last checked. New IDs, including CV values and timestamps, are inserted as additional rows. IDs which have not changed are updated with modified CV values and timestamps.

Step B. The Echo Nest track and artist IDs were ingested into GRAIL, in addition to any artist IDs collected from the Rosetta Stone project.

---

[12]https://musicmachinery.com/2010/02/10/introducing-project-rosetta-stone/

API ID Space	No. Linkages
Spotify track IDs	17.5M
Nokia Music track ID*	11.8M
Echo Nest track ID*	6.5M
MusicBrainz track ID	465K
MSD track ID	521K
MusixMatch track ID	3.8M
LyricFind track ID	1.5M
Spotify release ID	1.5M
Nokia Music release ID*	1.6M
MusicBrainz release ID	523K
Spotify artist ID	6.6M
Nokia Music artist ID*	358K
MusicBrainz artist ID	203K
MusixMatch artist ID	2.1M
Jambase artist ID	860K
OpenAura artist ID	3.7M
SeatGeek artist ID	2.8M
Seatwave artist ID	740K
LyricFind artist ID	1.5M
Twitter artist ID	670K
Facebook artist ID	3.7M
Tumblr artist ID	27K
FMA artist ID	240K
7digital artist ID	5.5M

Table 4: GRAIL linkage results for track, release, and artist IDs. Asterisks indicate discontinued services.

Step C. MusicBrainz artist, release, and track IDs are ingested into GRAIL based on API text searches.

Step D. Artist, release, and track IDs CVs are calculated for Last.FM metadata, and are ingested to GRAIL using linked MusicBrainz IDs and text searches. MusixMatch and 7digital API crawlers are in development using a similar process.

Step E. Artist, release, and track ID CVs will be ingested for new APIs, and any shared metadata that can improve accuracy.

Each API validation conducts consistency analyses using all available metadata shared between resources. As discussed in Section 2.2, resources are not consistent with respect to the information they provide and how it can be accessed. To that end, linkage processes across APIs have different dependencies. For instance, Spotify's coverage of ISRCs are more extensive than in MusicBrainz, track-level requests are made to Spotify first, followed by release- and artist- validation. Whereas in MusicBrainz, the most comprehensive request we can make is at the release-level first, followed by track and artist.

2.2.4 Linkage Results

Table 4 describes the outcome of our linkages to date. In summary, we have ingested millions of IDs from 16 providers at track, release, and artist levels into GRAIL, making it, to our knowledge, possibly the largest set of linked metadata of its type available for public use. Spotify achieves the highest linkage results, because the API can be queried with ISRCs, of which there are a 12 million distinct examples within GRAIL. Given the ongoing activity of the crawlers, we anticipate that the numbers in Table 4, both in terms of linkages and services will expand significantly, other than those services which are discontinued (Table 4, asterisks).

2.2.5 Application Program Interface

All GRAIL linkages are provided for public use through the GRAIL REST API (https://api.digitalmusiclab.org/). Users can query the API using track, release or artist IDs or strings, from

a documented list of services and datasets (refer to Figure 4 for resource availability). Users can request specific IDs in return and filter requested IDs based on CV thresholds, and timestamps (e.g. most recent crawl). Updated documentation relating to validation calculations, resource availability, and API news can be accessed via the website. The code for crawlers, website, and API are available on github at: https://github.com/digitalmusiclab/

GRAIL requires an authentication key in order to be accessed. Authentication keys are generated as part of the free registration process available via the GRAIL website. The terms of service state that this API is for, and by the Creative Commons, designed for use in MIR research. API keys are rate limited based on current server restrictions. Users are restricted from using GRAIL for profit without direct consent from the services within GRAIL and in their applications.

2.2.6 GRAIL Applications

Accessible and validated metadata has important uses for the MIR community and can contribute to innovative research. GRAIL could arguably improve top-performing genre-classifiers which use a diverse combination of user behaviour, lyrics, and cultural information [68] with expanded datasets. Perhaps more directly, psychological work examining the degree to which acoustic features influence musical choices was made possible through GRAIL's linkages to open-access resources [49, 70]. Future work using GRAIL may wish to consider exploiting linkages to lyric databases to examine the relationship between sung words and preference using psycholinguistics and computational natural language processing. In sum, accessibility to already linked metadata without having to readdress data validation makes new research more feasible.

Accurate metadata is also vital for the music industry; a study by Berkley University found that between 20–50 percent of music payments from digital services do not reach the rightsholder.[13] Community-maintained databases [67, 61, 71] provide essential metadata, and attempt to alleviate some of these challenges. However, the quality of data linkages can degrade over time without constant re-validation, as previously discussed. Furthermore, invaluable open resources like The Echo Nest can suddenly deprecate due to corporate buyouts or insufficient funding, arresting global research and development reliant upon it.[14] GRAIL may be a valuable resource, in that it addresses the above issues; as the Berkley study suggests, a community-maintained database linked to ISRCs is a necessary step to reduce data faults and improve non-payment of artists, which GRAIL's linkage process is based on.

Research relevant to this thesis using GRAIL assessing music preference is now discussed. Section 3.1 uses personality and mood research to inform an investigation of user preference and genre using high-level extracted acoustic features provided through GRAIL's linkage to the Echo Nest; Section 3.2 relies on cross-cultural psychology research as a framework to relate measures of user behaviour metrics such as eclecticism and temporality to GRAIL's linkage to yearly United Nations human development data pertaining to health, income, and education. Section 4.1 discusses future applications using GRAIL to refine current analyses and to contribute new areas of research.

3 Analyses

3.1 Cognitive Influence

Similarly to the devices used in our study, which were mobile phones and PCs (see Section 2.1), [72] sought to predict patterns of mobile-phone use from 120 participants' using the Coopersmith self-esteem inventory [73] assessing four of the Big Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism), and self-esteem. Individuals assessed as being neurotic, disagreeable, unconscientious and/or extraverted tended to spend more time messaging using SMS; disagreeable extraverts changed cellphone backgrounds and ringtones more frequently, indicating phone use as a means of stimulation and/or diversion; individuals who scored highly in neuroticism had relatively greater internet use, according to the authors, perhaps in an attempt to overcome loneliness.

In addition to device- and personality-specific research, studies exploring the interconnectedness of various forms of media and the consumption of culture, including music, have been undertaken. [20] assert that the perception of a musical genre depends in part upon the social setting in which it

---

[13]https://www.berklee.edu/news/fair_music_report
[14]https://blog.musicbrainz.org/2016/04/01/deprecating-mbids/

is heard and, by extension, the medium through which it is accessed; in other words, that people's preferences for certain media over others may influence musical categorization. With respect to music listening, personality studies have uncovered multiple associations: openness with Blues and Jazz; conscientiousness with Soul and Funk; extraversion with Pop and Rap, and so on ([74]; see also [20]). Moreover, personality appears not only to influence the extent to which individual genres are chosen, but also whether we possess narrow or wide-ranging music-listening habits [75]. In sum, personality research provides evidence for the existence of an overarching psychological framework in which effects akin to cognitive leakiness may occur [51]. As the research outlined above indicates, personality is a potent phenomenon, suffusing, guiding and shaping our decisions.

While successfully modeling human behavior, some researchers (e.g. [76]) have argued that personality traits may be too general to capture the subtle relationship between personality and online behaviour, including cellphone usage. For example, [77] investigated whether a lower-order, relatively narrow personality facet such as Need for Cognition (NFC) was able to predict online social and information-seeking behaviors. NFC is an individual's predisposition to engage with and enjoy information and cognitive endeavours, e.g. news content, crossword puzzles, Sudoku ([78, 79]). Despite its specificity, as opposed to a broader dimension such as openness, NFC (in addition to other personality-related factors) in the study conducted by [77] was found to correlate positively with Twitter usage, presumably, due to this social networking service's relatively high information content. Those with high ratings for sociability and extraversion appeared to prefer Facebook. In sum, [72] concluded that psychological profiling with respect to established personality dimensions could robustly explain how people chose to use their mobile phones.

In contrast to personality, which is relatively stable over time [80], mood can undergo rapid affective swings [81]. Moreover, while research has tended to concentrate on how music influences or induces mood, particularly with respect to consumer choice (e.g. [82, 83, 84, 85]), the converse is also true: mood influences musical choice [86]. Which is to say, assuming environmental factors and personal histories to be equal, a person's musical preferences do not depend solely on their personality, but, in addition, are subject to spur-of-the-moment choices influenced by mood. Amongst the theoretical models advanced to elucidate the role of mood in decision-making, perhaps the most influential is the Affect Infusion Model (AIM) [19]. In brief, the AIM seeks to explain how mood determines a person's capacity to process information—the importance of mood tends to increase in situations involving heavy cognitive load. As information complexity rises, and redundancy falls, the influence of mood on an individual's evaluations and responses increases, resulting in "intuitive" decision-making. Arguably, therefore, when faced with a plethora of diverse musical artists, tracks and genres, people tend to rely more upon their current mood, in which case the influence of personality may be temporarily reduced or suspended. To the author's knowledge, within the domain of music-preference research, this hypothesis has yet to be directly tested.

Despite this possible lacuna within the experimental literature, paradigms employing music-induced moods have produced results that are consistent with aspects of the AIM model. For example, risk-taking varies when mood is induced through listening to preferred versus disliked music. In a real-money gambling study, in which participants placed bets during either music-liked or disliked trial blocks, [87] found that people's appetites for risk-taking significantly increased when listening to preferred tracks. They conjectured that listening to preferred types of music increases the "marginal utility" of money (i.e. the additional satisfaction someone gains from consuming a good or service; [88, 84, 85]), which, in turn, increases the likelihood of participating in gambling. Although the foregoing covers aspects of decision-making involving music, none of the research and experimental scenarios referred to above necessarily replicate or are fully applicable to the particular issue at hand; namely, the degree to which the features of a person's preferred musical genre influence their choice of tracks or how musical tastes in the other genres they choose to consume. That personality and/or mood are in part responsible for why certain tracks belonging to different genres are preferred would seem to be self-evident—personality and mood underscore multiple aspects of human behavior, and presumably also our musical choices within genres of secondary importance. However, the extent to which personality and/or mood influence secondary musical choices is opaque. Supported by mood, personality, and statistical learning research, Section 3.1 explores the dynamics of musical preference using extracted acoustic features using Nokia's digital music streaming service (see Section 2.1 for detailed description of data).

This topic is divided into three main analyses. Using correlation, Section 3.1.2 identified differences in feature-dispersion patterns of genre defined subgroups of users. Section 3.1.3 2 involved the exhaustive calculation of feature-influence matrices, which, in combination with central-tendency

statistics, were used to detect the influence of main-genre features on those of secondary genres. Lastly, Section 3.1.4 investigated whether users whose main genre strongly dominated their collections exhibited a greater degree of feature influence with respect to secondary genres in comparison to users with relatively heterogeneous collections. Results strongly indicated that specific acoustic features in users' main genres influence the music they listened to when downloading secondary genres; the intensity of this influence appears to be related to the strength of preference for the main genre.

3.1.1 Users as "X-heads"

Due to the anonymous and sparse demographic information regarding our users, the behavioural aspect of our analyses are primarily based on the categorization of users into "X-head" subgroups, where X is the most numerous genre in their song download history. For example, a user with a majority of Metal downloads was classified as a Metal-head; most Classical downloads, a Classical-head and so on.[15]

An additional consequence of this categorization is the ability to determine the relative strength of a user's preference for a particular genre. We calculate a user's general preference for their favourite genre by normalizing their download histories into proportions by genre. For instance, if a user downloaded ten songs with six of the downloads labelled as Jazz, they would be classified as a Jazz-head, and the strength of their Jazz-preference is 0.6.

Calculating this value for every user was aggregated for each X-head subgroup to visualize population distributions for intensity of preference (see Figure 14 for a 4-genre sample). Comparing distribution statistics enables a high-level comparison of preference strength and characteristics across different X-head subgroups (i.e. in general, which communities are more heterogeneous with respect to the other genres they consume). Users with at least 20 downloads and over 7 days of retention (see Section 2.1) were selected for analysis.

3.1.2 Analysis 1: Feature Descriptors

The initial task in Analysis 1 was to identify whether extracted acoustic features (see Section 2.2 for acoustic feature description) for pairs of X-head subgroups are distinct. The reason for this is illustrated in Figures 7 and 8. Figure 7 shows the Energy distributions of tracks belonging to two X-head subgroups: the solid-orange line, D_M, shows the distribution for Dance tracks downloaded by Dance-heads ($Dance_{Main} = D_M$); the solid-blue line, J_M, shows the distribution for Jazz tracks downloaded by Jazz-heads ($Jazz_{Main} = J_M$). Notice that the peak of D_M is to the right, while the peak of J_M is to the left. The two peaks' relative positions indicate that, in general, Dance tracks listened to by Dance-heads have higher Energy than Jazz tracks listened to by Jazz-heads, as calculated by the Spotify analyzer.

Also present within Figure 7 are lines that show Energy distributions belonging to Dance- and Jazz-heads, but for tracks other than their predominant genres: the dotted-orange line, D_O, shows the distribution of non-Dance tracks downloaded by Dance-heads ($Dance_{Other} = D_O$); the dotted-blue line, J_O, shows the distribution of non-Jazz tracks downloaded by Jazz-heads ($Jazz_{Other} = J_O$). Two things are important to note: (1) that the Dance-heads' non-Dance tracks Energy distribution mirrors the distribution of their Dance tracks, e.g. both D_M and D_O peak on the right; and (2) that the Jazz-heads' non-Jazz tracks Energy distribution mirrors the distribution of their Jazz tracks, e.g. both J_M and J_O peak on the left. Which is to say, when Dance-heads download non-Dance tracks, there is a tendency for these tracks to be similar in terms of Energy to Dance tracks. Alternatively put, the generally high Energy of Dance tracks influences the choices Dance-heads make with respect to non-Dance music, while the generally low Energy of Jazz tracks influences the choices Jazz-heads make with respect to non-Jazz music.

The observation above relies upon X-head pairs having dissimilar feature distributions (i.e. lines D_M and J_M), and, in the case of Figure 7, the distribution of D_M being closer to D_O than J_O, and J_M being closer to J_O than D_O. If, however, the distributions of the X-heads' main genres are homologous, as is the case for Bollywood- and Reggae-heads in Figure 8 (solid green and purple lines), then no such pattern of similar/different distributions is possible. Which is to say, distributions where X-heads' main genres are more-or-less similar, are less able to demonstrate musical-feature influence.

---

[15]In rare instances where no genre had an absolute majority, the chronologically earliest downloaded genre is used to determine the user's categorization.

The distributions for all possible X-heads' main genres were correlated with each other in order to identify pairs with dissimilar distributions. This was conducted for all ten features. Using Pearson product-moment correlation, five features yielded no negative coefficients, and were thereby eliminated from the analysis. The five remaining features yielding negative coefficients, usable in the analysis, included Acousticness, Danceability, Energy, Loudness, and Valence. Figure 9 shows the correlation matrix for feature Energy. The two highlighted cells within the matrix relate to the solid lines in Figures 7 and 8, Dance and Jazz, and Bollywood and Reggae respectively. The Dance-Jazz coefficient is negative (reflected in the dissimilar distributions in Figure 7); the Bollywood-Reggae coefficient is positive (reflected in the similar distributions in Figure 8). In the case of Energy, this process yielded 19 X-head pairs suitable for analysis, i.e. 19 cells with negative coefficients.

Following this, for each X-head pair AB, the distribution of A's main genre (e.g. D_M, Figure 7) was correlated with the distribution of A's other music (e.g. D_O). Next, the distribution of A's main genre (e.g. D_M) was correlated with the distribution of B's other music (e.g. J_O). This produced two coefficients. This process was then repeated for B: B's main genre (e.g. J_M) was correlated with the distribution of their other music (e.g. J_O), and the distribution of B's main genre (e.g. J_M) was correlated with the distribution of A's other music (e.g. D_O). A and B together, therefore, produced four coefficients. For each feature, this was repeated for all X-head pairs with negatively correlated distributions, and the resulting coefficients entered into a paired sample t-tests in which "within-group" coefficients (e.g. D_M correlated with D_O) were paired with "between-group" coefficients (e.g. D_M correlated with J_O).

Figure 10 illustrates this process for Energy with respect to Dance- and Jazz-heads. In total, the 19 X-head pairs identified in the Energy correlation matrix in Figure 9 gave rise to a t-test into which 38 pairs were entered. This enabled us to observe whether there was a closer relationship between the features of A's main genre and their other music (Figure 10, red column; e.g. D_M and D_O) than with the features of B's other music (Figure 10, blue column; e.g. D_M and J_O) and vice versa, i.e. whether there was a significant influence of the main genre on music of secondary importance within people's downloads. If there had been no influence, then the distributions of either A or B's other music (i.e. D_O or J_O) would not be expected to show a consistently closer relationship to their respective main genre distributions (i.e. D_M or J_M). The results of this analysis for the five viable features referred to above are now presented.
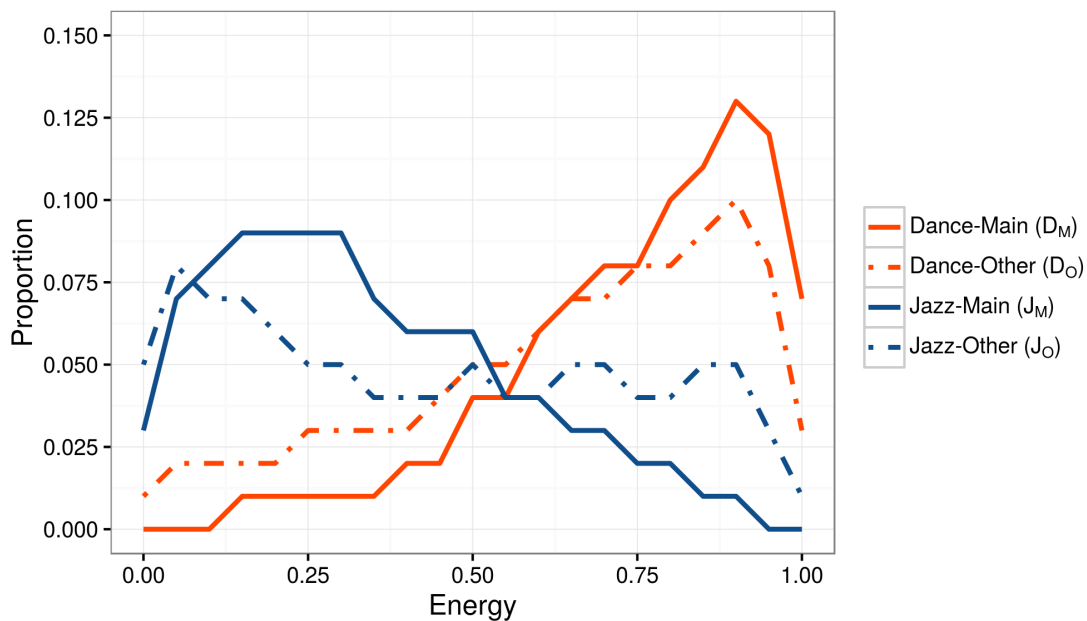
Figure 7: Energy distributions of tracks belonging to Dance-head and Jazz-head subgroups. The orange lines, D_M and D_O, show the distributions of tracks downloaded by Dance-heads; the blue lines, J_M and J_O, show the distributions of tracks downloaded by Jazz-heads.
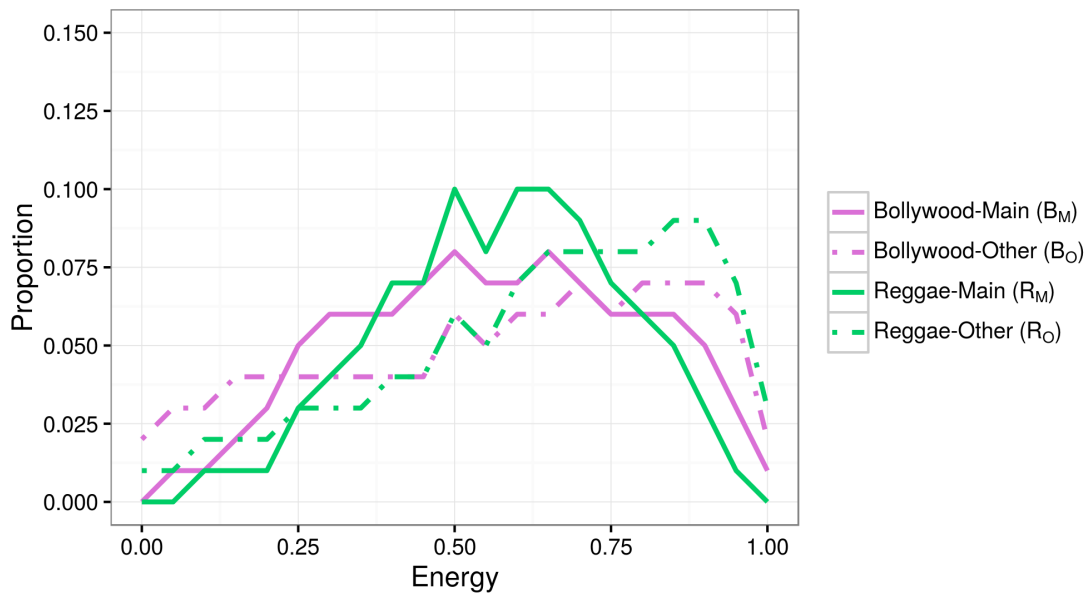
Figure 8: Energy distributions of tracks belonging to Bollywood-head and Reggae-head subgroups. The purple lines, B_M and B_O, show the distributions of tracks downloaded by Bollywood-heads; the green lines, R_M and R_O, show the distributions of tracks downloaded by Reggae-heads.
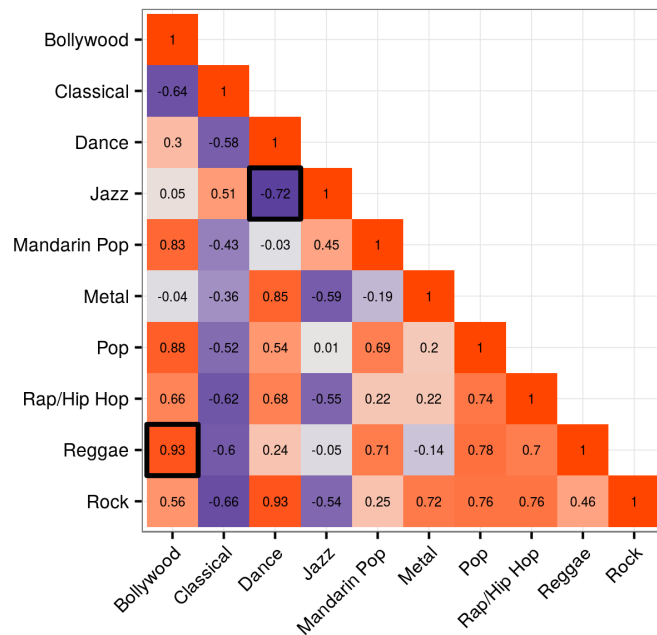
Figure 9: Correlation Matrix for Energy. The two highlighted cells relate to the solid lines in Figures 7 and 8, Dance and Jazz, and Bollywood and Reggae respectively.

As described above, the presence of negatively correlated distributions, shown in Figure 9 with respect to Energy, enabled the influence of five features to be studied using the present methodology. In sum, Acousticness had 12 negatively correlated distributions, Danceability 10, Energy 19, Loudness 5, and Valence 9 (Appendix 7). Figure 10 shows boxplots of the five features within the analysis. The red boxes on the left of each graph represent the within X-head coefficients; blue boxes on the right are the between X-head coefficients. Paired-sample t-tests, conducted to compare the within X-head coefficients and between X-head coefficients, showed the following results (sig. 2-tailed):

X-Head Subgroup	Users	Downloads
Pop	3,215,135	74,421,204 (23.14)
Bollywood	2,134,919	30,340,368 (14.21)
Mandarin Pop	1,944,975	15,165,454 (7.80)
Rock	349,205	23,489,799 (67.27)
Dance	163,388	1,676,634 (10.26)
Rap/Hip Hop	156,384	2,122,862 (13.57)
Metal	94,015	4,513,560 (48.01)
Classical	45,903	2,260,496 (49.25)
Jazz	22,644	1,131,941 (50.00)
Reggae	13,530	240,615 (17.63)

Table 5: Descriptive statistics for X-head subgroups: number of users; number of downloads (average downloads per user).

Acousticness. Significant difference for within (M=0.749, SD=0.170) and between (M=0.250, SD=0.302) X-head coefficients; $t(23)=11.887$, $p<0.0001$.

Danceabilty. Significant difference for within (M=0.608, SD=0.225) and between (M=0.313, SD=0.288) X-head coefficients; $t(19)=8.046$, $p<0.0001$.

Energy. Significant difference for within (M=0.557, SD=0.233) and between (M=-0.174, SD=0.414) X-head coefficients; $t(37)=9.110$, $p<0.0001$.

Loudness. Significant difference for within (M=0.636, SD=0.270) and between (M=0.315, SD=0.301) X-head coefficients; $t(9)=10.656$, $p<0.0001$.

Valence. Significant difference for within (M=0.653, SD=0.113) and between (M=0.223, SD=0.199) X-head coefficients; $t(17)=11.887$, $p<0.0001$.

	Paired sample t-test		
Within X-head subgroup correlation	**Coefficient**	**Coefficient**	**Between X-head subgroup correlation**
Dance-head main (D_M) to Dance-head other (D_O)	.96	-.44	Dance-head main (D_M) to Jazz-head other (J_O)
Jazz-head main (J_M) to Jazz-head other (J_O)	.56	-.72	Jazz-head main (J_M) to Dance-head other (D_O)
. . .	. . .	. . .	. . .

Figure 10: Example of paired sample t-tests with respect to Energy in which within-group coefficients were paired with between-group coefficients. Only the first two from 38 pairs are shown.

The statistics above confirm what is clearly evident in the boxplots in Figure 11: there is a significant difference in the two sets of coefficients for each feature; in general, coefficients for the within condition are greater than the between condition. This is also true for feature Loudness, which had only five X-head pairs with negatively correlated distributions (producing ten pairs of coefficients). In other words, even with a relatively low n, there is a significantly closer relationship between the features of an X-head's main genre and their other music than with the features of a different X-head's other music, i.e. a significant influence of the main genre on music of secondary importance within people's downloads.

3.1.3 Analysis 2: Feature influence Matrices

Despite the results having a clear direction, the current method was not able to test the influence of five of the ten features within the analysis: Duration, Instrumentalness, Liveness, Speechiness,
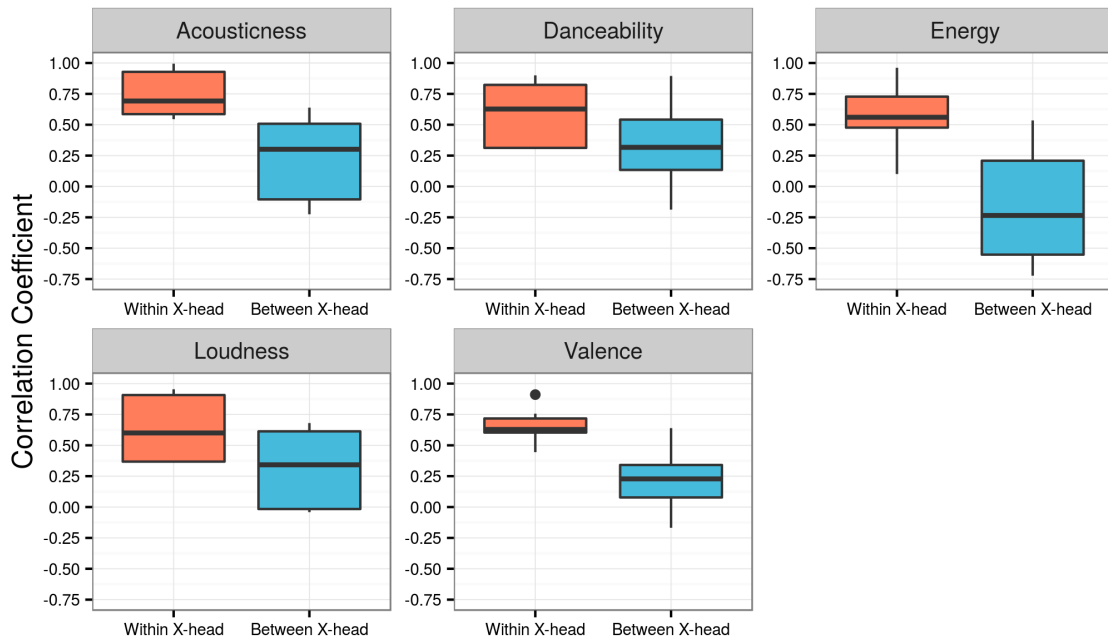
Figure 11: Boxplots of the five features within Analysis 1. Red boxes represent the within X-head coefficients; blue boxes represent the between X-head coefficients.

and Tempo. Although the observed pattern of influence may well extend to these features, this is by no means certain—for cognitive and neurological reasons, this phenomenon may be limited to particular acoustic features; for example, perhaps those that are more closely tied in some way to personality (e.g. [89]). Moreover, Analysis 1 was not able to address whether some X-head subgroups exhibited more influence, or if specific genres were more susceptible to being influenced by other genres more dominant within people's download collections. For example, is it the case that Country is more prone to the influence of negatively valenced or sad music than, say, Metal? Similarly, might Jazz be more immune to the influence of up-tempo music than Bollywood, and what might be the interaction of X-head subgroup on these processes? The aim was not necessarily to explain such patterns, which may well involve a confluence of personality and sociocultural factors, but rather to observe the degree to which they existed within Nokia DB. To this end, we undertook a second analysis in which detailed information relating to each X-head subgroup and our selected ten genres was brought to light.

Each feature-influence matrix, referred to as C, was calculated from two submatrices, A and B. A, a 10 (X-heads) x 10 (genres) submatrix, contained the average feature values of all songs within a genre downloaded by an X-head subgroup (for example, the average value of Valence for all Reggae tracks downloaded by Classical-heads). B, a 10 (genres) x 1 (averages) submatrix, contained the average feature value of each genre downloaded by all users, excluding those made by the main X-head subgroup (for example, the value for Metal music determined by calculating the average Valence of Metal tracks downloaded by everyone except Metal-heads). C, the 10 (X-heads) x 10 (genres) feature-influence matrix, was calculated by subtracting the row values in A (subtrahends) from those in B (minuends), and converting the resulting differences into percentage changes from the averages in B. Formally, the above is given by:

$$F = \{\forall x_g \in F_{xg}, x_g = \{\frac{F_{xg} - P_g}{P_g} * 100\}\} \tag{3}$$

Where:

F = Feature-influence matrix (Matrix C)

x = X-head subgroup

F_{xg} = Average feature value for genre (g) in X-head (x) subgroup (Submatrix A)

P_g = Average feature value for genre (g) for entire population (Submatrix B)

19

x_g = Average feature value for genre (g) listened to by X-head subgroup (x)

We illustrated this process with reference to Submatrices A and B, and Matrix C (the feature-influence matrix), and feature Valence. For clarity, the calculation is simplified to include only three X-head subgroups and genres: Dance, Metal, and Pop.

Submatrix A   Table 6 shows Matrix A: rows (i) represent X-head subgroups; columns (j) represent genres downloaded by each X-head subgroup. For example, average Classical, Dance, and Metal Valence values for Dance-heads ($i = 2, j = (1, 2, 3)$) are $(2, 1) = 0.28, (2, 2) = 0.41$, and $(2, 3) = 0.37$ respectively.

	Classical	Dance	Metal
Classical-heads	0.27	0.43	0.38
Dance-heads	0.28	0.41	0.37
Metal-heads	0.28	0.44	0.35

Table 6: Example of Submatrix A showing the average Valence of three genres downloaded by three X-head subgroups. Rows represent X-heads; columns represent genres.

Submatrix B   Table 7 shows Submatrix B: the columns are genres; the row is the average Valence of each genre, excluding members of that particular X-head subgroup. For example, the average Valence for Metal music downloaded by non-Metal-heads $(1, 3) = 0.39$.

	Classical	Dance	Metal
Population average	0.28	0.47	0.39

Table 7: Example of Submatrix B for feature Valence with three genres. Rows represent genres; column represents the average feature value of a downloaded genre, excluding members of that particular X-head subgroup.

Matrix C (Feature-Influence Matrix)   Table 8 shows Matrix C, generated by subtracting cell i,j in matrix A from cell i,j in Matrix B. We take the percentage change for that feature using the population average for a particular genre in Matrix B (similar results were obtained using population medians as opposed to averages). For example, to calculate cell (2,2) of Matrix C:

$$c_{i,j} = \left(\frac{A(i, j) - B(j)}{B(j)}\right) * 100$$

$$c_{2,2} = \left(\frac{0.41 - .47}{.47}\right) * 100 = \textbf{-12.8}$$

This example indicates that Dance-heads downloaded Dance music that was 12.8% more negatively valenced than the rest of the population downloading Dance.

	Classical	Dance	Metal
Classical-heads	-3.57%	-8.51%	-2.56%
Dance-heads	0.0%	**-12.8%**	-5.12%
Metal-heads	0.0%	-6.1%	-10.26%

Table 8: Example of Matrix C, the Feature-Influence Matrix, showing the percentage Valence change of three genres downloaded by three X-head subgroups.

Figure 12 shows the feature-influence matrix for Acousticness. Cell values indicate the percentage change in Acousticness of genres (columns) downloaded by X-head subgroups (rows), compared to the average Valence of genres downloaded by the overall population. The highlighted diagonal cells (running top-left to bottom right) show X-heads with respect to their main genres. The highlighted column on the right shows the median value of each row, excluding diagonally highlighted cells.
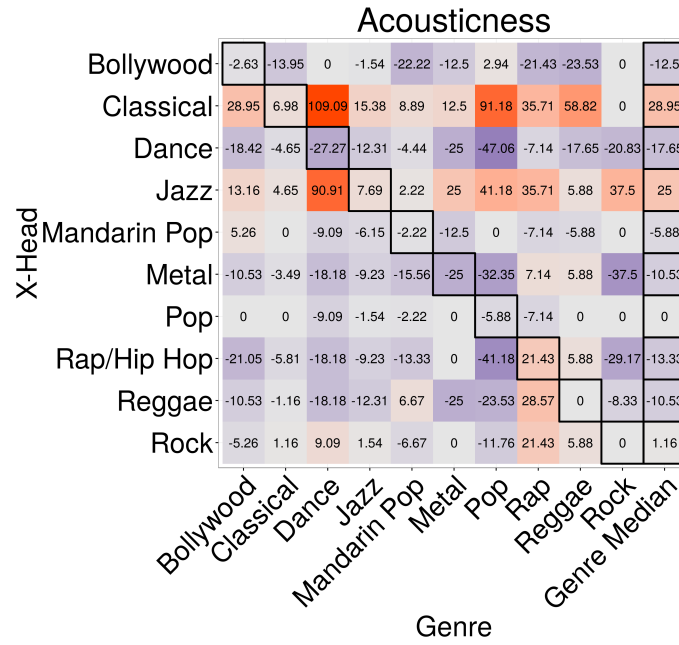
Figure 12: Feature-influence matrix for Acousticness. The highlighted diagonal cells (running top-left to bottom right) show X-heads with respect to their main genres. The highlighted column on the right shows the median value of each row, excluding diagonally highlighted cells.
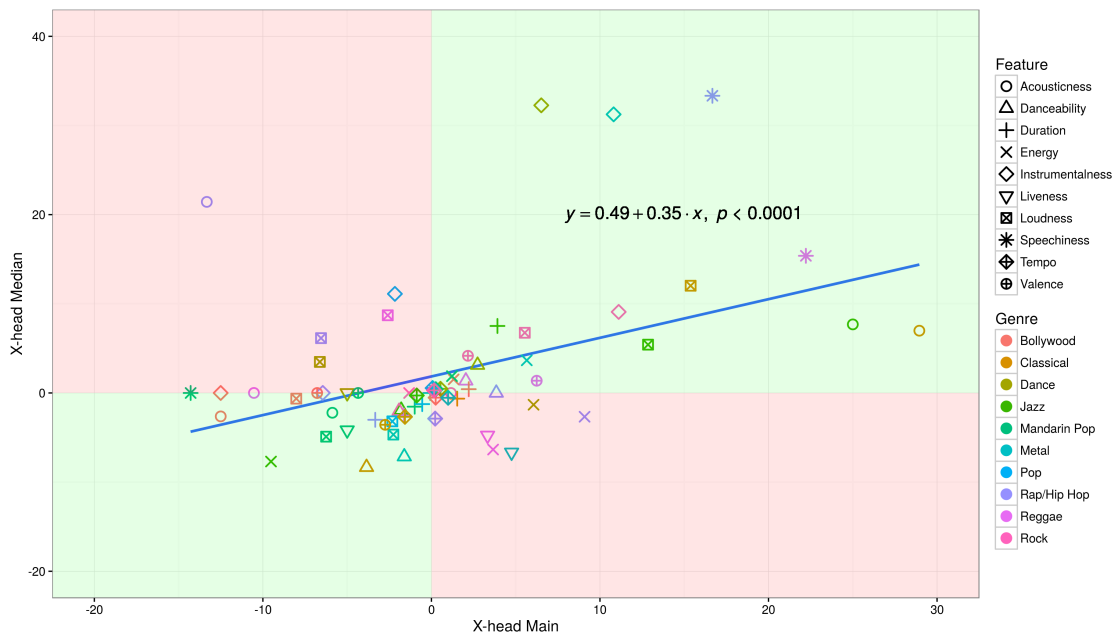
Figure 13: Scatterplot showing the 100 diagonal-to-median cell pairings of the ten feature-influence matrices. Light-green quadrants indicate sign agreement between the row medians and X-heads with respect to their main genres, either positive or negative; pink quadrants indicate sign disagreement.

Of the 100 possible diagonal-to-median cell pairings (10 features x 10 X-heads), the signs of 64 were in agreement; 36 were in disagreement (see Supplementary Material B). These pairings are shown in the scatterplot in Figure 13. Light-green quadrants indicate sign agreement between the row medians and X-heads with respect to their main genres, either positive or negative; pink quadrants indicate sign disagreement. The adjusted r^2 value, 0.1039, gives rise to the following statistic: $r=0.34$, $n=100$, $p<0.0001$. The finding in the feature-influence matrices that 64 diagonal-

Rank	Feature ($n = 90$)
1	Speechiness; $r = 0.52$***
2	Danceability; $r = 0.48$***
3	Loudness; $r = 0.45$***
4	Energy; $r = 0.44$***
5	Acousticness; $r = 0.28$**
6	Tempo; $r = 0.19$˙
7	Duration; $r = 0.17$
8	Valence; $r = 0.14$
9	Liveness; $r = 0.06$
10	Instrumentalness; $r = 0.04$

Table 9: Table showing the ranked order of features. The feature column shows which feature of X-heads' main genres is closest to their other genres (˙ $= p < 0.1$; * $= p < 0.05$; ** $= p < 0.01$; *** $= p < 0.001$).

Rank	X-head ($n = 90$)
1	Metal; $r = 0.56$***
2	Jazz; $r = 0.49$***
3	Dance; $r = 0.45$***
4	Classical; $r = 0.41$***
5	Rap; $r = 0.28$**
6	Rock; $r = 0.26$*
7	Bollywood; $r = 0.14$
8	Reggae; $r = 0.11$
9	Mandarin Pop; $r = 0.09$
10	Pop; $r = 0.08$

Table 10: Table showing the ranked order of X-heads. The X-head column shows the subgroup with the closest relationship between their main genre and other genres, across all features (˙ $= p < 0.1$; * $= p < 0.05$; ** $= p < 0.01$; *** $= p < 0.001$).

to-median cell pairings were in agreement, with 36 in disagreement, strongly suggests that there is a directional relationship, either positive or negative, between the features of X-heads' main genres and those of their other genres. This is confirmed in the scatterplot in Figure 13, and associated correlation statistic ($r = 0.34$), in which there was a significant, positive relationship between the variables.

The feature-influence matrices enabled two further, complementary questions to be explored: (1) which feature of X-heads' main genres most strongly influenced their other genres, and (2) which X-head subgroup's main genre most strongly influenced their other genres across all feature. For example, is the relationship between X-heads' main and other genres stronger for Energy than Danceability? This question was assessed by correlating X-heads' main genres with the nine other genres in each of their download collections. This produced one overall coefficient per feature-influence matrix; the resulting ten coefficients were then ranked in order. The second question asked which X-head subgroup across all features had the closest relationship between their main genre and other genres. For example, do the features of Mandarin Pop-heads' main genre more strongly influence the corresponding features of their other genres than is the case for Reggae-heads? This question was investigated by correlating each X-head's main genre with the nine other genres in each feature-influence matrix. This produced one overall coefficient per X-head subgroup, and, as before, the resulting ten coefficients were ranked in order. The results of these analyses are shown in Tables 9 and 10. Rows represent ranks, either of features or X-heads, columns include Pearson product-moment correlation statistics.

In Table 9, the top-ranked feature was Speechiness ($r = 0.52$)—the presence or absence of spoken words in tracks belonging to main genres appears to have created a preference for similarly "speechy" tracks in X-heads' other genres. Similarly, the Danceability, Loudness and Energy of users' predominant tracks appear to have heavily influenced tracks of secondary importance. At the other end of the spectrum, there was little-to-no relationship tracks between X-heads' main and other genres in terms of Liveness (whether a track was recorded at a live event) and Instrumentalness (whether a track was created using only instrumental sounds).

The top-ranked X-head subgroups were Metal ($r = 0.56$) and Jazz ($r = 0.49$). The dynamic nature of much Metal music seems to have created a musical "fingerprint" that strongly expressed itself in the other genres Metal-heads downloaded. Likewise, Jazz-heads seem compelled to seek out music containing Jazz-like qualities when exploring non-Jazz music. Conversely, Mandarin Pop and Pop's musical features did not significantly influence the other music downloaded by these X-head subgroups, perhaps because the features of these genre are relatively indistinct ($r = 0.09$ and $r = 0.08$ respectively).

3.1.4   Analysis 3: X-head Strength

While Analysis 2 provides a novel method to determine the relative influence of features and general susceptibility to features by X-head in aggregate, what specific features contribute to each

subgroup and the strength of the feature's effect remains unclear. For example, although Table 10 suggests that Metal-heads are most influenced by acoustic features found in Metal music and that speechiness is the most influential feature across subgroups, the degree to which speechiness is influencing Metal-heads (as opposed to loudness or valence) cannot be determined using this analysis. Furthermore, X-head categorization is determined by the majority downloaded genre in a user's collection and neither Analysis 1 or 2 consider the intensity of preference. When listening to non-main genres, what features are most important for an Jazz-head? Are Metal-heads whose download history contains 100% Metal tracks more likely to prefer these high-tempo features than a Metal-head whose history contains 51%? Analysis 3 examines the relationship between an X-heads intensity of genre preference and acoustic features of their selected music.

As described in Section 2.1, user data is anonymous; X-head categories (the majority downloaded genre in a user's collection) were developed in order to group users into meaningful categories for examining music preferences. The degree of this majority (e.g. 80% of downloads for a user corresponding to their X-head category, as opposed to 60%) may be a useful metric for determining "X-head intensity" in a way that builds upon the previous analyses. Based on the 62 genres available in the database (Appendix 7) and the X-head calculation described in Section 3.1.1, the X-head intensity metric can hypothetically range from 0.016 ($\frac{1}{62}$) - 1.00. Figure 14 visualizes X-head intensity distributions for the 20 largest X-head subgroups for users with at least 20 downloads. The distributions qualitatively vary in normality, skewness, and kurtosis. Communities with the most spread distributions statistics (i.e. the lowest average, mode, standard deviation, skew and kurtosis) are considered, in general, more genre inclusive (heterogenous) because the distribution of intensities are less focused on a specific area; whereas the communities with the highest values are considered more genre exclusive (homogenous). The most heterogeneous communities using these distribution metrics for all genres include: Bollywood- ($Kurtosis = -0.900$), Ghazal- ($Skew = 0.694$), Bhojpuri- ($\mu = 0.428$), Khaleeji- ($\sigma = 0.0086$), and Region/Other-heads ($Mode = 0.33$). The most homogenous communities were Italiana- ($Skew = 8.566, Kurtosis = 78.64$), Classic Arab/Tarab- ($\sigma = 0.569$), and Iskelda-heads ($\mu = 0.767$). 41 communities had modes at the maximum value of 1.00, suggesting that distributions are generally non-parametric and overall users tend to download only one genre. For an exhaustive list of values, refer to Table 17 in Appendix Section 7.
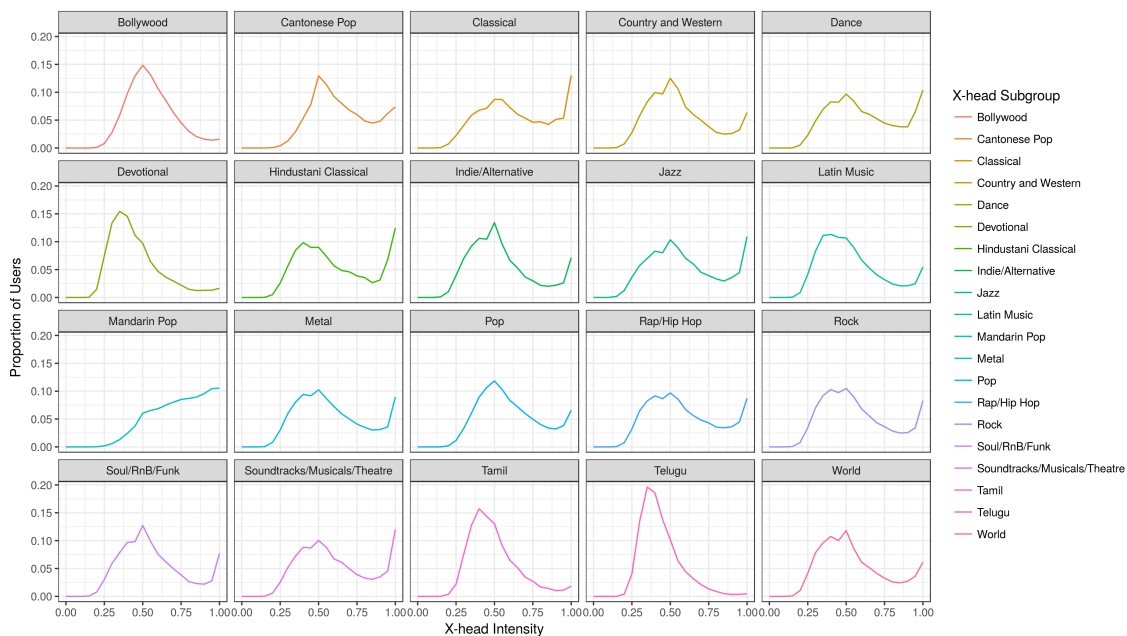
Figure 14: X-head distributions for the top-20 subgroups for users with at least 20 downloads

Because not all X-head subgroups exhibit normal distributions, non-parametric analyses were utilized. User's were aggregated into quartiles based on their intensities (labelled from one-to-four as: weak-heads, medium-weak heads, medium-strong-heads and strong heads) in order to balance unequal distributions across X-head subgroups. For example, users labelled as strong-heads will have an X-head intensity closer to 1.00, whereas weak-heads have intensities closer to the hypothetical minimum (0.016). Due to the population dynamics of this dataset, each subgroup

contains a differing counts of total users, equally distributed across the four categories. Defining X-head intensity as ordinal quartile categories, rather than a continuous rational value is advantageous for two reasons: (1) analyses will not have an over-representation of any particular intensity (i.e. the analysis will not overfit to the mode X-head intensity of 1.00), and (2) it enables the use of two non-parametric tests which, and to what degree, acoustic features are influencing an X-head's preference when seeking out new music.

Jonchkeere-Terpstra trend tests (JT) [90] were used to evaluate whether there is a significant trend in ordered differences among categories. This test determines what features are producing a significant trend across X-head intensity. However, the JT does not provide a measure of effect size. To that end, a Kendall-Tau-β test (KT) [91], a non-parametric test similar to the JT is utilized in order to determine the effect size of specific features for X-head subgroups. Based on the success of analysis in Sections 3.1.2 and 3.1.3, five additional X-head subgroups were included in this analysis: Blues-, Easy Listening/Oldies-, Indie/Alternative-, and World-heads. Because the JT and KT were developed for sample sizes which are orders of magnitude smaller than what is available in Nokia DB, bootstrapping with replacement was utilized. 100 random samples of X-head subgroups were generated. Each subgroup included 4000 users for each subgroup with equally distributed intensities (i.e. 1000 weak, medium-weak, medium-strong and strong users). Two subgroups (Blues and Country) were removed because they did not have at least 2000 users in each intensity category. The results of the 100 random samples of JT and KT tests for all ten features and 12 X-head subgroup were then averaged.

Figure 15 and Table 18 in Appendix 7 outline the averaged values from the 100 bootstrap samples for factors with at least marginally significant JT p-values only. 51 of 120 (14 X-head subgroups, each with 10 features) achieved marginal JT significance, 43 of 120 with significant JT test. The top 3 largest Kendall Tau-β include: Speechiness by Easy Listening-heads ($\tau = -0.148$), Speechiness by Rap-heads ($\tau = 0.124$), and Acousticness by Easy Listening-heads. Figure 15 suggests that X-head's choose new music based on features that are hallmark in their preferred genre, and that they select music with features more similar to their preferred as their preference increases. The results provide additional support for trends found in Tables 9 and 10 – that some acoustic features and X-head subgroups are more sensitive or influenced by preference than others. Easy-listening/Oldies-, Classical-, Metal-, and Dance-heads are generally more influenced by their favourite genre than Pop-, Bollywood-, or Rock-heads; Speechiness, Loudness, Danceability are more often significant than Valence and Duration. The influence of some features and X-heads is different between Analysis 2 and 3. Features Instrumentalness and Valence, and the Reggae-head subgroup should not demonstrate an influence in preference according to Analysis 2, but qualitatively are some of the more represented tests in Analysis 3. Because Analysis 2 does not consider preference intensity, the differences in significance for some X-heads and acoustic features suggest that the degree to which a genre influences feature choice in secondary styles differs across subgroups.

3.2   Ecleticism and Human Development

While mood and personality pertain, to some degree, to an individual's artistic taste, shared demographic factors including culture, education, income, sex, and age also effect musical people's musical choices [22, 92, 93, 94, 95, 96]. Of these, age appears to be the strongest predictor of musical preference [92]; older adolescents also prefer lighter qualities in music compared to younger adolescents [93]. General enjoyment of music from grade one to college drops for a time until rising around age of puberty, following a U-shaped curve across development [97]. With respect to sex, a music-choice study suggests that males prefer music with themes of dominance and independence, whereas females preferred music with relationship and emotion themes [92]. However, the extent to which this research is generalizable is open to debate: almost 30 years has elapsed since [92]'s study, which was based on low-sample surveys with relatively little demographic variation. Furthermore, [94, 96] found that demographic information did not conclusively determine music preferences; two- and three-way interactions were found between age, sex and country, and controlling for these factors significantly reduced the strength of the relationships.

One of the most studied social psychological behaviours related to music preference is a concept known as "cultural omnivorism" (used interchangably with eclecticism), coined by Richard Peterson in the 1990s [22, 98, 99, 100, 101]. Peterson argues that for much of human history, music has been a means of stratifying groups across class and status, thus defining "highbrow" and "lowbrow"
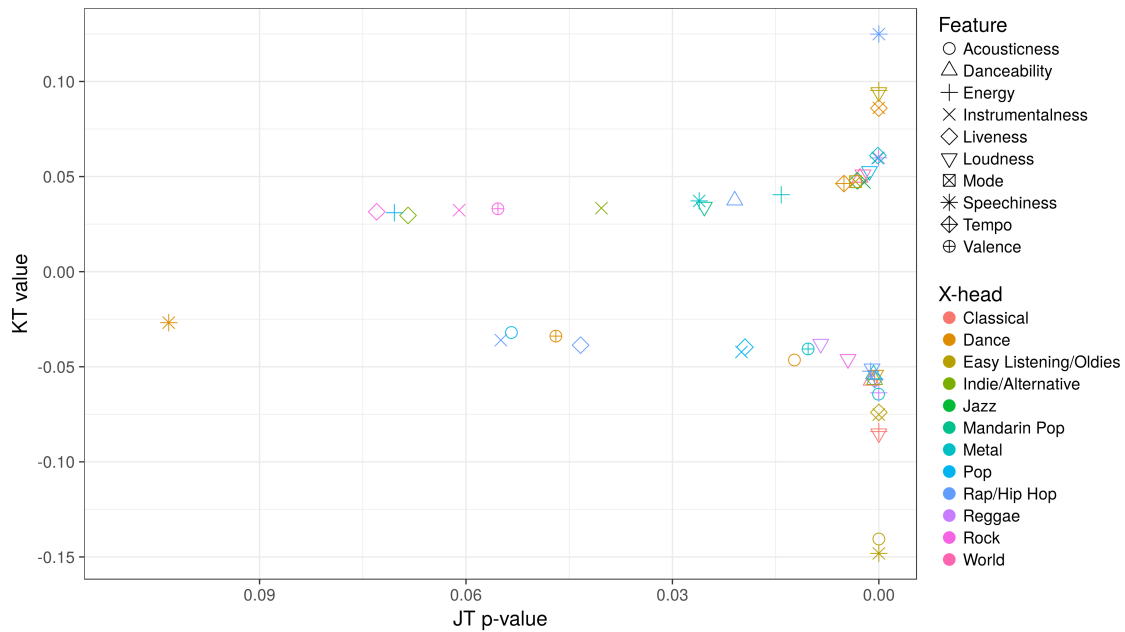
Figure 15: Results of the non-parametric Jonckheere-Terpsra trend test (x-axis) and Kendall Tau-β rank correlation (y-axis). Colours represent X-head tested subgroups and shapes represent tested features.

artistic styles. Traditionally, highbrow music was socially and economically expensive to participate in; instruments and concerts were costly (restricting access to low-class individuals), and learning music theory required education and time investments (restricting access to low-status and low-class individuals). However, as access to music began to become easier through technological advances (e.g. sheet music through printing press, radio, recorded sound, personal recording studios, and the internet), lowbrow consumers were able access and appreciate highbrow art, thereby negating the ability to stratify across class boundaries using music. [22] argues that in order for highbrow to maintain stratification highbrow consumers transitioned from an elitist to eclectic to maintain class and status boundaries [102]. Cross-cultural research generally supports the notion that diverse tastes are influenced by economic and education across many other types of consumption including: visual arts [102], television [103], theatre [104], and restaurants [98].

Other related user behaviours, such as voraciousness [105], support the omnivore hypothesis, and also correlate with personal education and income. Supported by these cross-cultural studies, a testable hypothesis emerges, that is, user behaviours such as possessing eclectic music tastes, should rise as national education and income values improve. Some evidence supports the hypothesis education [106] and income [107] influence music and visual art preferences [104, 101, 102]; however, several criticisms including methodological problems [52, 53] suggest that omnivores may not be as distinctive as previously thought. Relevant methodological issues addressed by this thesis include: measures of education and income are typically collected from participants during manual data collection and are inadequate proxies for general human development; eclecticism behaviour is quantified using survey and census data, which have low sample sizes and can be influenced by selection/experimenter bias; and narrow geographic distributions limit the extension of these findings to primarily European, Caucasian, English-speaking countries.

Using digital music resources, and data provided by globally recognized institutions may be particularly suited to addressing methodological weaknesses, and by extension, the interpretability of results. Past work using Nokia DB [23, 55] found that daily and weekly download dispersion patterns (used interchangeably in this thesis with the term "temporality") correlates with the Human Development Index (HDI) provided in the UN's Human Development Report [108]. The HDI attempts to evaluate general human well-being in a single numerical value. HDI (Equation 4) considers education, income, and health levels per country, aggregated into a geometric mean. From the original "cultural omnivore" research using general indicators of human development, a more specific hypothesis suggests that of the composite indices in the HDI, a main effect for the education and income indices, and user eclecticism and temporality is expected. Some social

psychological research supports this idea: eclecticism correlates with education based on 20-years of consumption information [106], and [107] finds three distinct types of consumption behaviour in the UK, stratified across economic boundaries [107]. Some criticisms regarding the sensitivity of the HDI to regionally resource inequality have been discussed [109, 110]. To that end, this thesis also considers whether derivative HDI metrics such as the inequality-adjusted HDI (IHDI) [111], are more appropriate metrics for validating user behaviours within the theoretical framework of cultural consumption patterns.

$$\sqrt[3]{LEI \cdot EI \cdot II} \tag{4}$$

where:

LEI: Life Expectancy Index. Calculated using mean life-expectancy per country.

EI: Education Index. Calculated using mean years of schooling per country.

II: Income Index. Calculated using Gross National Index (GNI), a measure similar to gross national product.

3.2.1 Analysis 1: Country-level

Global download patterns across the day and week are influenced by economic conditions within a country [23]. However, regularity in consumer patterns (i.e. the consistency of consumption) may be under-reported because users tend to repeat purchasing habits regardless of intent [112]. Extending upon [23], which found predictable patterns of consumption across countries using the Human Development Index, this analysis investigates whether components of the HDI-income, education, or health influence when people download across the world. Using identical methodology as [23], we count the amount of yearly downloads per country across two temporal windows, a 24-hour and 7-day period. Aggregated downloads counts are normalized into proportions and standard deviations are calculated; we refer to these calculations as hourly/weekly dispersions. A high degree of variance indicates that most of the downloads occurred more during a specific hour of the day or day of the week. Because countries entered the service incrementally over time, the Nokia DB was subsampled to include HDI and download information between 2010-2012, when most countries were included in the system. Based on the research of [23] it is expected that the HDI income index should be significantly correlated to dispersion across countries in our sample.

Multiple linear regression analysis compared our temporal dispersion measures to the components of the HDI. Table 11 indicates a significant effect of dispersion to HDI components $(F(3, 81) = 3.25, p = 0.001 **, r^2 = 0.269)$. However, no specific main effects across component or interaction effects were found to contribute significantly to this difference. Although the expected significant main effect between dispersion and the HDI income index was not found, the overall significance of the F-statistic supports the previous work in [23]. The current method may not be able to detect differences among components in a few ways. The sensitivity of the HDI calculation does not account for concentration of income, education, and health resources among a countries entire population [111], potentially impacting the sensitivity of the index. Additionally, other measures of human consumption (i.e. eclecticism) may be more appropriate behaviours to compare with HDI due to the relative larger body of research examining what, (see Section 3.2), as opposed to when, people consume digital music. The inequality-adjusted HDI (IHDI) and metrics for country- and user-level eclecticism are next used to potentially overcome methodological issues from previous analyses.

Country-level Eclecticism. Based on the data available within Nokia DB, this thesis defines eclecticism heuristic as the degree of overall balance in downloaded genres across countries or users, and is formalized in Equation 5. Eclecticism is calculated by counting every genre downloaded within a country. The top 10 downloaded genres are selected, and normalized to a proportion and σ is calculated for this 10 element array. Higher values indicate less eclectic users or countries with less eclecticism, and lower values indicate more eclectic countries. Country-level eclecticism scores were calculated for every year and country in the Nokia DB. Figure 16 illustrates eclecticism as a stacked-bar graph. Each bar is a country, and every colour stacked on a country represents proportion of downloads from one of 10 downloaded genres. For example, the orange bar for IN (India) is Bollywood, which corresponds to approximately 40% of all downloaded music by India

X	Y	Score	p	Adjusted r^2
1) HDI Income Index 2) HDI Education Index 3) HDI Health Index	7-day Periodicity	$F(7, 77) = 6.9$ $t(84) = 0.133$ $t(84) = -0.407$ $t(84) = -0.703$	$p < 0.0001$*** 1) $p < 0.895$ 2) $p = 0.685$ 3) $p = 0.484$	Adjusted: 0.3296
1) HDI Income Index 2) HDI Education Index 3) HDI Health Index	24-Hour Periodicity	$F(7, 77) = 5.995$ $t(84) = 0.552$ $t(84) = .484$ $t(84) = 0.382$	$p < 0.0001$*** 1) $p = 0.583$ 2) $p = 0.484$ 3) $p = 0.704$	Adjusted: 0.2939

Table 11: 24-hour and 7-day to HDI multiple linear regression tables. No significant two- or three-way interactions were found.

in 2011. More eclectic countries have more equal representation of genres (i.e. for a country to be maximally eclectic, all ten bars should be close to 0.1).

First, individual linear regressions are conducted on HDI and closely related statistics found in the HDR (Table 12); additional metrics include school enrolment ratio (i.e. how many students were expected to enrol versus how many did), expected years of schooling for males and females, and acoustic features extracted through GRAIL. To determine if the iHDI is a more sensitive measure for analysis, two multiple linear regressions are conducted using country-level eclecticism with the components of the HDI and IHDI (Table 13). Based on the sociological work described in Section 3.2, it is hypothesized that education and income (indicators of class and status) should be negatively correlated with our measure of country-level eclecticism.

$$\sigma_c = \left\{ p_g = \frac{g_c}{\sum\limits_{g=1}^{10} g_c}; g \in [1, 10] \right\} \tag{5}$$

Where:

g = Genre

u = Country

$\bar{\sigma}_c$ = Standard Deviation of top-10 genre proportions in a given country

p_g = Proportion of downloads (p) for a top-10 genre

g_c = Count of downloaded genre in a country

Table 12 includes all significant individual correlations found during preliminary analysis. Income, education, and all but one additional measure of education are significantly correlated in the direction we expect. Effect sizes are relatively weak ranging from $r^2 = 0.0729 - 1.591$. Significant correlations generally support the hypothesis that education and income influence music preferences. As expected, no correlation was found for eclecticism and HDI's health index. The direction of correlation for gross primary enrolment ratio is opposite to what was expected. This may be due to how enrolment ratios are calculated; determining expected enrolment for initial entry into education is much harder to track than at later years due to parents deciding to delay enrolment or inaccurate birth reporting.

Two multiple linear regressions were run to determine whether IHDI is more sensitive to use with measures of human behaviour like music eclecticism. At the time of collection from the API API, only one year of iHDI data (2012) was available for analysis. Consequently, this analysis compared multiple linear regression models for 2012 HDI and iHDI data only. Overall, iHDI accounted for more than twice as much variance as the HDI model ($r^2 = 0.09961 - 0.1934$, respectively); significant main effects were found for the income and education indices as expected. Significant two-way interactions were found between income/life expectancy indices ($t = -2.166, p = 0.0468$), and education/income indices ($t = 2.22, p = 0.0422$). Marginal two- and three- way interactions were found for education/health ($t = 2.21, p = 0.0992$) and education/health/income ($t = 2.072, p = 0.056$).
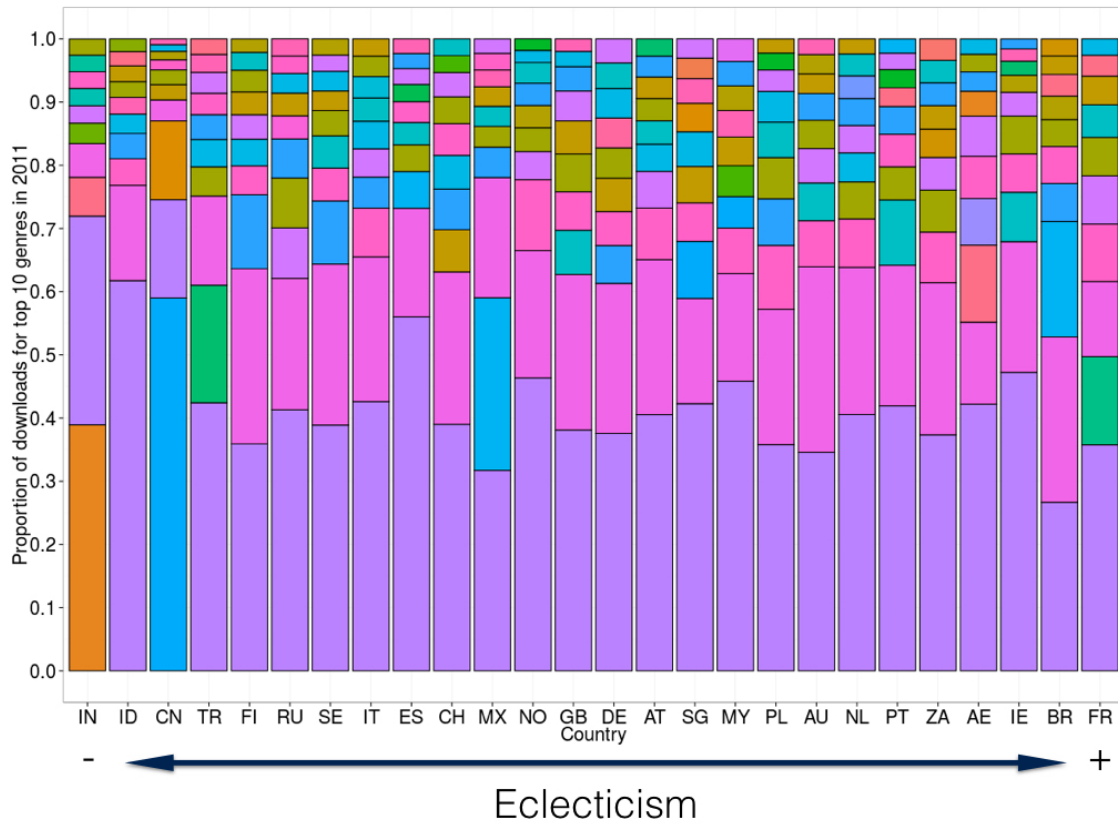
Figure 16: Proportion of top-10 downloaded genres by country in 2011. Genres represented by different colours. Countries are order by least to most eclectic based on the calculation described in Equation 5

X	Y	N	P	R
HDI	Top 10 SD	87	0.0127*	-0.27
Income Index	Top 10 SD	87	0.0093**	-0.28
Education Index	Top 10 SD	87	0.0042**	-0.3
Expected years of schooling; female	Top 10 SD	70	0.0016**	-0.37
Expected years of schooling; males	Top 10 SD	70	0.0018**	-0.37
Gross primary enrolment ratio	Top 10 SD	70	0.0098**	0.3
Gross secondary enrolment ratio	Top 10 SD	69	0.0166*	-0.28
Gross tertiary enrolment ratio	Top 10 SD	69	0.001***	-0.39

Table 12: Individual Linear Regressions for country-level eclecticism measure.

Although significant and marginal interaction effects were found for all indices, these results support past research examining demographic information including education and employment to global music preferences [94, 96]. Although interactions are consistent with previous literature, complex effects make it difficult to confirm the nature of the relationship as described in the hypothesis. In the context of this work, it is unclear which, or how, each index is influencing eclecticism. No significant main effects or interactions were found using HDI, suggesting that the IHDI may be more appropriate to use for analyzing global music preferences. Additional methodological weaknesses should be considered: although the IHDI generally performs better than the HDI for this analysis, our sample size is relatively low and does not span across many years. Another weakness is that country-level eclecticism is very high-level, and does not account for individual user differences. In light of this considerations, a user-level eclecticism calculation is derived and compared to HDI data for regression analysis.

X	Y	Score	p	Adjusted r^2
1) IHDI Income Index 2) IHDI Education Index 3) IHDI Health Index	Country SD	F(3, 24) = 2.998 t(26) = 2.556 t(26) = -2.092 t(26) = -1.002	p < 0.0523 1) p = 0.018* 2) p = 0.04823* 3) p = 0.327	0.1934
1) HDI Income Index 2) HDI Education Index 3) HDI Health Index	Country SD	F(7, 20) = 1.427 t(26) = 0.62 t(26) = 0.162 t(26) = -0.046	p = 0.2494 1) p = 0.542 2) p = 0.912 3) p = 0.964	0.09961

Table 13: Multiple Linear Regression Analysis using HDI to IHDI statistics and country-level eclecticism. Results suggest that iHDI may be a more sensitive statistic for evaluating global consumption trends.

3.2.2 Analysis 2: User-level

User-level eclecticism, Equation 6 was calculated in a manner similar to Equation 5. However, rather than calculating a single σ per country, an exhaustive calculation is conducted for every user. User-level eclecticism is determined by taking a median eclecticism value from every user within a country. Values of zero are added to the the vector when a user has less than 10 genres in their download history.

$$\tilde{\sigma}_c = \left\{ p_g = \frac{g_u}{\sum\limits_{g=1}^{10} g_u}; g \in [1, 10] \right\} \tag{6}$$

Where:

g = Genre

u = User

$\tilde{\sigma}_c$ = Median SD of top-10 genres for all users in a given country

p_g = Proportion of downloads (p) for top-10 genres in a user's collection

g_u = Count of downloaded genre in a user's (u) collection

As previous, individual linear regression analysis was first conducted using the user-level eclecticism measure with the HDI and other closely related statistics; additional metrics are identical to those used in country-level analysis. For a comprehensive list of HDR metrics, see Appendix **??**. Significant correlations are included in Table 14. Individual correlations reveal similar and stronger relationships than found in the country-level analysis of Table 12. Of note, a new significant correlation was found for the valence acoustic feature, and gross primary enrolment ratio is no longer significant. As described in Section 2.1, valence is a perceptual estimation of a tracks positive/negative affect. Lastly, a multiple linear regression was conducted and compared to the results of Table 15. A significant overall effect was found ($F(7, 20) = 2.259, p = 0.00482$), and a significant main effect was found for the health index. However, a significant 2-way interaction was found between health/education indices ($t = -2.257, p = 0.0354*$). The overall effect size using user-level eclecticism accounted for three times more variance than the country-level method.

4 General Discussion

Using an interdisciplinary approach, Section 2.2 enriched and validated a large corpus of digital music metadata for CogMIR research; using this resource, Section 3 examined two topics relevant to MIR and cognitive psychology, which to some extent, addresses methodological concerns within both fields. However, some limitations remain. A common issue with "big data" resources concerns data consistency and reliability. This thesis, as well as a majority of MIR tasks, either rely on

X	Y	N	P	R
HDI	Median SD	28	0.0383*	-0.39
Income Index	Median SD	28	0.0097**	-0.48
Education Index	Median SD	28	0.0401*	-0.39
Combined Enrolment both sexes	Median SD	20	0.0237*	-0.5
Gross Primary Enrolment Ratio	Median SD	21	0.0248*	0.49
Gross Tertiary Enrolment Ratio	Median SD	20	0.0076**	-0.49
Average Valence	Median SD	28	0.0153*	-0.45

Table 14: Individual Linear Regressions for user-level eclecticism measure.

X	**Y**	Score	p	r^2
1) HDI Income Index 2) HDI Education Index 3) HDI Health Index	User SD	$F(7, 20) = 2.259$ $t(27) = 0.955$ $t(27) = 1.962$ $t(27) = 2.118$	$p = 0.04482*$ 1) $p < 0.351$ 2) $p = 0.0638$ 3) $p = 0.0469*$	Adjusted: 0.2919

Table 15: Multiple linear regression model using user-level eclecticism value. A two-way interaction between health and education (t=-2.257 p=0.0354)

public user-contributed efforts [61, 113] or private digital music platforms (see resource coverage in Section 2.2), both of which have been criticized for data reliability [29]. Public resources include inconsistent and/or controversial labeling, are absent of filters, and contain duplicate examples [65]; furthermore, datasets fail to achieve similar classification accuracies using identical classification algorithms: cross-collection accuracies ranged from 6-53% [65]. Private resources also contribute to data faults: TuneCore,[16] an online digital music-distribution service, allegedly issued incorrectly formatted ISRCs without permission from the ISRC. As a result of TuneCore's alleged error, hundreds of thousands of songs were shown as being created in the British overseas territory of the Turks and Caicos Islands (ISO 3166-1-Alpha-2 country code TC). Given that TC has a population of only 33,000, this prodigious musical output is surely questionable and approximately one million ISRCs in GRAIL contain incorrectly assigned country code of TC. This inconsistency does not impact our linking process, but illustrates how industry-standard data can be affected by improper data practices. The intention of GRAIL was to address these issues by including filter procedures into the linking architecture, and motivate future work by making the codebase public; however, it does not fully alleviate the problem. Cooperation from the MIR community in the form of data access as well as technical and monetary support will be vital in order to achieve music-data accuracy.

Other measures of variance such as the Gini coefficient, g, should also be considered for deriving behavioural differences such as eclecticism from users. For example, the Gini coefficient is a measure of inequality traditionally used to evaluate income inequality, but has more recently seen applications in biology [114], genetic expression [115], and social psychology [116]. It is calculated (Equation 7) by taking the mean absolute difference divided by the average, and normalized for scaling. Given that our current version of eclecticism takes the standard deviation of a cumulative distribution of proportions (see Equation 5), the Gini coefficient may be a more sensitive measure for determining inequalities in a user's genre preference and should be considered.

$$g = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n \sum_{i=1}^{n}} \tag{7}$$

The framework used to determine testable user preferences from large datasets considers statistical learning to be a shared theoretical underpinning across topics. An additional benefit was that statistical learning is also a familiar topic within MIR, providing an area of shared research interest that is equally understood in different but important ways. Although statistical learning is discussed as a mechanism for preferences and behaviours, no statistical learning procedure was

---

[16]http://www.tunecore.com/

used during the analysis. This thesis decided to use more traditionally psychological analysis techniques in part due to novel use of large volumes of acoustic features derived from private companies for cognitive research. Some research supports the notion that statistical learning is important in music perception and preference: EEG mismatch negativity is stronger when musicians hear musical violations in their genre of expertise [43, 117]. Although high-level acoustic features were used in this thesis, low-level features extracted in short windows of time are available for each track. Future work should include examining how statistical learning influences the perception and preference of acoustic features in short time intervals.

The algorithms responsible for Spotify's acoustic features are proprietary, and therefore not publically available. Although my primary aim was to investigate and record the presence of musical-feature influence, I was unable to assess in detail which specific acoustic elements were responsible for my findings. Which frequency bands within an X-head's main genre, for example, have in general a greater influence on their other genres? Which components of Acousticness are present throughout an X-head's download collection, and which are specific to the main genre? Moreover, and perhaps of greater import, as mentioned at the outset the psychological reality of acoustic features is, as yet, unquantified [10]. Although a feature like Valence may make "sense" to those who know and love music's emotional power, its interpretation across listeners may be highly divergent. Another example, Danceability, is available in signal processing packages [14, 60] and open-access APIs.[17] It is derived from a combination of beat salience and consistency to derive a floating point for a song between zero and one [28]. However, according to this definition the most danceable song should be a steady, metronomic-like pulse, which clearly does not capture the cognitive nuances of what makes music danceable. Perceptual experiences of music such as Danceability or Valence are undeniably more complex than these heuristic definitions and contribute to the inaccessibility of MIR for music cognition research [9]. Acoustic features which focus on human perception [10] or user behaviour [118], rather than simplified or heuristic representations, are advantageous because they improve classification [119, 68, 120], and provide more nuanced interpretability over acoustically or heuristically defined measures.

Section 3.1.4 found little variation between acoustic features across genres in the Spotify library. Given that hundreds of acoustic features have been derived and used in a variety of MIR tasks, in recent years relatively small gains in accuracy have been made [29], suggesting possible ceiling effect using acoustic-only analysis. As a result, MIR's growing interest in using psychologically derived behaviours with the aim of improving predictive accuracy is not surprising [12]. Many features can be extracted from a piece of recorded sound. However, some acoustic features, like danceability, may not be wholly represented in the acoustic signal. These features can be considered inherently psychological features [10]. The process in Section 3.2 of deriving the psychological feature for eclecticism from consumption and demographic data, and providing an experimental means to determine its efficacy provides a framework to explore other types of consumption behaviours. Future work should include testing new behaviours (four described in Section 4.1) relevant to CogMIR research. Additionally, directly testing the perceptual efficacy of acoustic features in an controlled laboratory should be considered. Appendix 7.5 includes more future work using GRAIL as a research proposal which is beyond the scope of this work's research focus. This work describes a digital journaling application which uses music and natural language processing, and psycholinguistics as a way of informing current mental health states.

4.1 Future Directions

4.1.1 User Acoustic Behaviours

Because there is a low degree of variance across high-level acoustic features for music, user behaviour may be more appropriate for a variety of MIR classification tasks [118]. Diversity can be measured in many ways. It could represent preferences for a variety of genres (i.e. eclecticism), or a preference for music that is more or less mainstream (i.e. mainstreamness [118]). Additionally, temporal patterns of music consumption differ across national and demographic boundaries [121, 23]. Users with differing musical preferences exhibit different patterns of regular downloading based on differences in personality traits such as openness to experience[20]. However, regularity in consumer patterns may be under-reported because users tend to repeat purchasing habits regardless of intent [112]. Global patterns of genre and temporal consumption are relatively unexamined

---

[17]https://developer.spotify.com/web-api/get-audio-features/

within both music cognition and MIR. Alongside our measure of eclecticism (Section 3.2.2), we propose four new behavioural features and outline their calculations: overall regularity, periodic regularity, adventurousness, and volume.

Adventurousness can be defined for each user by considering the average popularity for each track downloaded by a user in the Nokia DB. The total number of times each track was downloaded was pre-computed as a measure of overall song awareness. In general, music consumption generally follows an Zipf-like distribution of growth with respect to downloads and popularity (see Section 2.1); Appendix 7.4.1 describes a method and example for calculating adventurousness. Volume measure of user engagement that considers traditional measures like play count [119], but also takes into consideration aspects of temporality. For example, two users who download 100 songs should not be considered equal consumers if user 1 downloaded 1 song over 100 days or 100 songs or 1 day; Appendix 7.4.2 provides an equation for calculating user volume. Overall regularity measures the degree of day-by-day download activity. Appendix **??** illustrates how overall regularity is calculated for two users. Periodic (7-day) regularity refers to the degree of regular behaviour across the week. Measuring overall regularity may find general patterns of behaviour across a users entire history, but miss regularity in specific, weekly intervals; Appendix 7.4.4 illustrates periodic regularity for one user with three weeks of download behaviour.

5   Closing Remarks

In summary, this thesis explored three topics relevant to cognitive science and MIR research, and to some extent, attempts to address issues with both. The first topic explored the current state of music metadata consistency. Using a large metadata catalogue of digital tracks provided by Nokia, Section 2.2 enriched an already expansive set to include an diverse set of music metadata which enabled the analysis. Although the purpose of this topic was to inform cognitive research, and was primarily methodological in nature, it represents an novel contribution to MIR and future cognitive research by focusing on metadata linking and validation efforts. To encourage further efforts, Section 2.2 presents code-base and linked data as an open-access API.[18]

Topic two regarding acoustic features and genre preference found strong evidence of influence with respect to users' consumption of multiple styles of music; clear relationships emerged between the features of X-heads' main and secondary genres. This effect was found to be stronger for some features than others, most noticeably Speechiness, Danceability, and Loudness, and more pronounced in certain subgroups, such as Metal-heads, Jazz-heads and Dance-heads. While the reasons for differential effects within features and X-heads is unknown, two probable, independent causal mechanisms were suggested to account for general main-to-secondary genre influence. First, personality creates an overarching psychological framework in which certain factors, such as openness and agreeableness, guide musical preference, irrespective of genre; some personality factors may be linked to specific acoustic features. Second, via statistical learning, listeners extract the acoustic regularities of various musical features, which in turn influence the creation of musical preferences beyond established styles and/or genres. Of course, these mechanisms need not be mutually exclusive, but may serve to reinforce one another. Attempts, therefore, to tease apart the effects of personality and statistical learning could prove to be difficult, although paradigms in which these factors are independently manipulated might settle the issue of personality versus statistical learning conclusively.

Topic three explored the extent to which user behaviours defined from social psychological research could be used to determine trends in global music preference. Extending past research examining temporal aspects of human consumption [23] and human development, this work examined whether a new feature, eclecticism, could be derived and validated using a more rigorous hypothesis and analysis. As predicted, significant correlations were found for income and education measures provided by the UN. Furthermore, this analysis determined that user-level eclecticism and inequality-sensitive derivatives of the HDI were significant in a multiple linear regression using country-level eclecticism and the IHDI. Main effects for the multiple linear regression model included income and education, also as predicted. Although the method of deriving eclecticism from digital consumption data in all likelihood will not be the definitive measure over time, the theoretical underpinning and process of validation is novel and contributes to powerful, but relatively unexplored data sources in the cognitive sciences.

---

[18]http://api.digitalmusiclab.org/

The topics explored in this work has a variety of implications future music cognition research. This work uses utilizes enriched metadata from multiple open-access big data research is uncommon in cognitive sciences, but is quickly becoming an attractive alternatives to traditional methods; the amount of data and relative ease of computational analysis enables research to be executed quickly with a high degree of external validity. By addressing significant concerns of data integrity commonly voiced by the psychological community and providing a resource which attempts at alleviate them is novel; GRAIL's validation procedures contribute to the groundwork for comprehensive music metadata collection and cleaning. While it is intuitive that human preference for new genres may be informed by features commonly found in past preferences, little work has been able to address the questions highlighted in Analysis 1 concerning which, and to what degree, these features are motivating music choice. Furthermore, this information informs MIR research regarding which extraction algorithms are sensitive to differences across genres and users. Lastly, the work in Analysis 2 implies that that user features may be more valuable than acoustic features for MIR classification tasks and provides additional legitimacy to topics in social psychological research on eclecticism examining the extent to which socioeconomic factors influence music consumption behaviours across national boundaries.

A primary motivation of this thesis was to undertake interdisciplinary research in MIR and cognitive science, capitalizing on the advantages of each while minimizing the concerns of both. While commonplace for technologically driven research, large and diverse datasets like the Nokia DB are rare within cognitive psychology. The relative rarity of these kinds of resources is, in part, due to methodological concerns voiced by more traditional psychological practices [8], and weak applications of established psychological findings on human music preception [122, 9]. The extent to which the research contributions generated by this thesis address these concerns will be determined over time; it is the author's hope that the resources and methods made public by this endeavour eases future interdisciplinary research initiatives and provides a stepping stone for psychology to enter into CogMIR research.

6   References

[1] C. Darwin, "The descent of man," 1871.

[2] C. Darwin, *The Descent of Man and Selection in Relation to Sex.* Penguin Classics, 1883.

[3] S. Pinker, "How the mind works," *Annals of the New York Academy of Sciences*, vol. 882, no. 1, pp. 119–127, 1999.

[4] S. Mithen, I. Morley, A. Wray, M. Tallerman, and C. Gamble, "The singing neanderthals: the origins of music, language, mind and body," *Cambridge Archaeological Journal*, vol. 16, no. 01, pp. 97–112, 2006.

[5] K. Ashton, "That 'internet of things' thing," *RFiD Journal*, vol. 22, no. 7, pp. 97–114, 2009.

[6] J. S. Downie, "Music information retrieval," *Annu Rev Inform Sci*, vol. 37, no. 1, pp. 295–340, 2003.

[7] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: Current directions and future challenges," *Proceedings of the IEEE*, vol. 96, no. 4, pp. 668–696, 2008.

[8] J.-J. Aucouturier and E. Bigand, "Seven problems that keep mir from attracting the interest of cognition and neuroscience," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 483–497, 2013.

[9] J.-J. Aucouturier and E. Bigand, "Mel cepstrum & ann ova: The difficult dialog between mir and music cognition.," in *ISMIR*, pp. 397–402, 2012.

[10] A. Friberg, E. Schoonderwaldt, A. Hedblad, M. Fabiani, and A. Elowsson, "Using perceptually defined music features in music information retrieval," *arXiv preprint arXiv:1403.7923*, 2014.

[11] K. Siedenburg, I. Fujinaga, and S. McAdams, "A comparison of approaches to timbre descriptors in music information retrieval and music psychology," *Journal of New Music Research*, vol. 45, no. 1, pp. 27–41, 2016.

[12] J. H. Lee and J. S. Downie, "Survey of music information needs, uses, and seeking behaviours: Preliminary findings.," in *ISMIR*, vol. 2004, p. 5th, Citeseer, 2004.

[13] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International Conference on Digital Audio Effects*, pp. 237–244, 2007.

[14] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. R. Zapata, and X. Serra, "Essentia: An audio analysis library for music information retrieval.," in *ISMIR*, pp. 493–498, Citeseer, 2013.

[15] P. Toiviainen, M. Tervaniemi, J. Louhivuori, M. Saher, M. Huotilainen, and R. Näätänen, "Timbre similarity: Convergence of neural, behavioral, and computational approaches," *Music Perception: An Interdisciplinary Journal*, vol. 16, no. 2, pp. 223–241, 1998.

[16] A. D. Patel and J. R. Daniele, "An empirical comparison of rhythm in language and music," *Cognition*, vol. 87, no. 1, pp. B35–B45, 2003.

[17] B. Shao, D. Wang, T. Li, and M. Ogihara, "Music recommendation based on acoustic features and user access patterns," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1602–1611, 2009.

[18] Q. Li, S. H. Myaeng, and B. M. Kim, "A probabilistic music recommender considering user opinions and audio features," *Information processing & management*, vol. 43, no. 2, pp. 473–487, 2007.

[19] J. P. Forgas, "Mood and judgment: The affect infusion model (AIM)," *Psychological Bulletin*, vol. 117, no. 1, p. 39, 1995.

[20] P. J. Rentfrow and S. D. Gosling, "The do re mi's of everyday life: the structure and personality correlates of music preferences.," *Journal of personality and social psychology*, vol. 84, no. 6, p. 1236, 2003.

[21] M. Woolhouse and J. Renwick, "Generalizing case-based analyses in the study of global music consumption," *Digital Studies/Le champ numérique*, 2016.

[22] R. A. Peterson and R. M. Kern, "Changing highbrow taste: From snob to omnivore," *American sociological review*, pp. 900–907, 1996.

[23] M. Woolhouse and J. Bansal, "Work, rest and (press) play: Music consumption as an indicator of human economic development," *The Journal of Interdisciplinary Music Studies*, vol. 7, no. 1&2, pp. 45–71, 2013.

[24] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics Springer, Berlin, 2001.

[25] A. Porter, D. Bogdanov, R. Kaye, R. Tsukanov, and X. Serra, "Acousticbrainz: a community platform for gathering music information obtained from audio," in *Proceedings of the 16th ISMIR Conference*, vol. 787, 2015.

[26] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "Eeg-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[27] G. Tzanetakis, *Music data mining*. CRC Press, 2011.

[28] S. Streich and P. Herrera, "Detrended fluctuation analysis of music signals: Danceability estimation and further semantic characterization," in *Proceedings of the 118th AES Convention*, 2005.

[29] B. Sturm, "The state of the art ten years after a state of the art: Future research in music information retrieval," *Journal of New Music Research*, vol. 43, no. 2, pp. 147–172, 2014.

[30] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, "How to grow a mind: Statistics, structure, and abstraction," *science*, vol. 331, no. 6022, pp. 1279–1285, 2011.

[31] N. Z. Kirkham, J. A. Slemmer, and S. P. Johnson, "Visual statistical learning in infancy: Evidence for a domain general learning mechanism," *Cognition*, vol. 83, no. 2, pp. B35–B42, 2002.

[32] C. M. Conway and M. H. Christiansen, "Modality-constrained statistical learning of tactile, visual, and auditory sequences.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 31, no. 1, p. 24, 2005.

[33] N. B. Turk-Browne, P. J. Isola, B. J. Scholl, and T. A. Treat, "Multidimensional visual statistical learning.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 34, no. 2, p. 399, 2008.

[34] D. Goldreich, "A bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions," *PloS one*, vol. 2, no. 3, p. e333, 2007.

[35] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.

[36] C. M. Conway, A. Bauernschmidt, S. S. Huang, and D. B. Pisoni, "Implicit statistical learning in language processing: Word predictability is the key," *Cognition*, vol. 114, no. 3, pp. 356–371, 2010.

[37] J. R. Saffran, E. K. Johnson, R. N. Aslin, and E. L. Newport, "Statistical learning of tone sequences by human infants and adults," *Cognition*, vol. 70, no. 1, pp. 27–52, 1999.

[38] G. S. Berns, C. M. Capra, S. Moore, and C. Noussair, "Neural mechanisms of the influence of popularity on adolescent ratings of music," *Neuroimage*, vol. 49, no. 3, pp. 2687–2696, 2010.

[39] M.-L. Halko, T. Mäkelä, L. Nummenmaa, Y. Hlushchuk, and M. Schürmann, "Hedonic context modulates risky choices and reward responses in amygdala and dorsal striatum.," *Journal of Neuroscience, Psychology, and Economics*, vol. 8, no. 2, p. 100, 2015.

[40] D. Huron and J. Ollen, "Agogic contrast in French and English themes: Further support for patel and daniele (2003)," *Music Perception*, vol. 21, no. 2, pp. 267–271, 2003.

[41] J. R. Daniele and A. D. Patel, "The interplay of linguistic and historical influences on musical rhythm in different cultures," in *Proceedings of the International Conference on Music Perception and Cognition*, pp. 3–7, International Conference on Music Perception and Cognition, 2004.

[42] M. T. Pearce, M. H. Ruiz, S. Kapasi, G. A. Wiggins, and J. Bhattacharya, "Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation," *NeuroImage*, vol. 50, no. 1, pp. 302–313, 2010.

[43] E. Brattico, K. J. Pallesen, O. Varyagina, C. Bailey, I. Anourova, M. Järvenpää, T. Eerola, and M. Tervaniemi, "Neural discrimination of nonprototypical chords in music experts and laymen: an meg study," *Journal of Cognitive Neuroscience*, vol. 21, no. 11, pp. 2230–2244, 2009.

[44] S. Koelsch, E. Kasper, D. Sammler, K. Schulze, T. Gunter, and A. D. Friederici, "Music, language and meaning: brain signatures of semantic processing," *Nature neuroscience*, vol. 7, no. 3, pp. 302–307, 2004.

[45] M. Cheour, R. Ceponiene, A. Lehtokoski, A. Luuk, J. Allik, K. Alho, and R. Näätänen, "Development of language-specific phoneme representations in the infant brain," *Nature Neuroscience*, vol. 1, no. 5, pp. 351–353, 1998.

[46] M. Rivera-Gaxiola, L. Klarman, A. Garcia-Sierra, and P. K. Kuhl, "Neural patterns to speech and vocabulary growth in american infants," *NeuroReport*, vol. 16, no. 5, pp. 495–498, 2005.

[47] A. D. Patel, J. R. Iversen, and J. C. Rosenberg, "Comparing the rhythm and melody of speech and music: The case of British English and French," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 3034–3047, 2006.

[48] D. Västfjäll, "Emotion induction through music: A review of the musical mood induction procedure," *Musicae Scientiae*, vol. 5, no. 1, pp. 173–211, 2002.

[49] M. Barone, J. Bansal, and M. Woolhouse, "Acoustic features influence musical choices across multiple genres," *Frontiers in Psychology*, 2017.

[50] J. Bansal and M. Woolhouse, "Predictive power of personality on music-genre," in *Proceedings of the International Society for Music Information Retrieval*, pp. 652–658, International Society for Music Information Retrieval, 2015.

[51] J. Rieskamp, J. R. Busemeyer, and B. A. Mellers, "Extending the bounds of rationality: evidence and theories of preferential choice," *Journal of Economic Literature*, vol. 44, no. 3, pp. 631–661, 2006.

[52] R. A. Peterson and A. Simkus, "Seven how musical tastes mark occupational status groups," *Cultivating differences: Symbolic boundaries and the making of inequality*, vol. 152, 1992.

[53] A. Warde, D. Wright, and M. Gayo-Cal, "Understanding cultural omnivorousness: Or, the myth of the cultural omnivore," *Cultural sociology*, vol. 1, no. 2, pp. 143–164, 2007.

[54] J. R. Groff, P. N. Weinberg, *et al.*, *SQL: the complete reference*, vol. 2. McGraw-Hill/Osborne, 2002.

[55] M. Woolhouse, J. Renwick, and D. Tidhar, "Every track you take: Analysing the dynamics of song and genre reception through music downloading," *Digital Studies/Le champ numérique*, 2014.

[56] C. Anderson, *The long tail: How endless choice is creating unlimited demand.* Random House, 2007.

[57] M. Levy and K. Bosteels, "Music recommendation and the long tail," in *1st Workshop On Music Recommendation And Discovery (WOMRAD), ACM RecSys, 2010, Barcelona, Spain*, Citeseer, 2010.

[58] L. Wilson and G. K. Zipf, "Human behavior and the principle of least effort," 1949.

[59] B. Manaris, J. Romero, P. Machado, D. Krehbiel, T. Hirzel, W. Pharr, and R. B. Davis, "Zipf's law, music classification, and aesthetics," *Computer Music Journal*, vol. 29, no. 1, pp. 55–69, 2005.

[60] D. Bogdanov, J. Serrà, N. Wack, and P. Herrera, "From Low-Level to High-Level: Comparative Study of Music Similarity Measures," *2009 11th IEEE International Symposium on Multimedia*, vol. 7, no. 3, pp. 453–458, 2009.

[61] A. Swartz, "Musicbrainz: A semantic web service," *IEEE Intelligent Systems*, vol. 17, no. 1, pp. 76–77, 2002.

[62] G. Vigliensoni, J. Burgoyne, and I. Fujinaga, "Musicbrainz for the world: the chilean experience.," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, pp. 131–136, 2013.

[63] B. L. Sturm, "A survey of evaluation in music genre recognition," in *International Workshop on Adaptive Multimedia Retrieval*, pp. 29–66, Springer, 2012.

[64] B. Angeles, C. McKay, and I. Fujinaga, "Discovering metadata inconsistencies.," in *Proceedings of the 11th International Society for Music Information Retrieval*, pp. 195–200, 2010.

[65] D. Bogdanov, A. Porter, P. Herrera, and X. Serra, "Cross-collection evaluation for music classification tasks," in *Proc. of the Int. Conf. on Music Information Retrieval (ISMIR)*, 2016.

[66] N. Corthaut, S. Govaerts, K. Verbert, and E. Duval, "Connecting the dots: music metadata generation, schemas and applications.," in *Proceedings of the 9th International Conference on Music Information Retrieval*, pp. 249–54, 2008.

[67] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pp. 591–596, 2011.

[68] C. McKay, J. A. Burgoyne, J. Hockman, J. B. Smith, G. Vigliensoni, and I. Fujinaga, "Evaluating the genre classification performance of lyrical features relative to audio, symbolic and cultural features.," in *ISMIR*, pp. 213–218, 2010.

[69] G. Navarro, "A guided tour to approximate string matching," *Association for Computing Machinery: Computing Surveys*, vol. 33, no. 1, pp. 31–88, 2001.

[70] J. Bansal and M. Woolhouse, "Predictive power of personality on music-genre exclusivity.," in *Proceedings of the 15th International Conference on Music Information Retrieval*, pp. 652–658, 2015.

[71] C., Cannam and M., Sandler and M.O., Jewell and C. Rhodes and M. d'Inverno, "Linked data and you: Bringing music research software into the semantic web," *Journal of New Music Research*, vol. 39, no. 4, pp. 313–325, 2010.

[72] S. Butt and J. G. Phillips, "Personality and self reported mobile phone use," *Computers in Human Behavior*, vol. 24, no. 2, pp. 346–360, 2008.

[73] S. Coopersmith, *Coopersmith Self-esteem Inventories: SEI.* Consulting Psychologists Press, 1981.

[74] R. L. Zweigenhaft, "A do re mi encore: A closer look at the personality correlates of music preferences," *Journal of Individual Differences*, vol. 29, no. 1, pp. 45–55, 2008.

[75] D. Rawlings and V. Ciancarelli, "Music preference and the five-factor model of the NEO personality inventory," *Psychology of Music*, vol. 25, no. 2, pp. 120–132, 1997.

[76] C. Ross, E. S. Orr, M. Sisic, J. M. Arseneault, M. G. Simmering, and R. R. Orr, "Personality and motivations associated with facebook use," *Computers in Human Behavior*, vol. 25, no. 2, pp. 578–586, 2009.

[77] D. J. Hughes, M. Rowe, M. Batey, and A. Lee, "A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage," *Computers in Human Behavior*, vol. 28, no. 2, pp. 561–569, 2012.

[78] B. Verplanken, "Need for cognition and external information search: Responses to time pressure during decision-making," *Journal of Research in Personality*, vol. 27, no. 3, pp. 238–252, 1993.

[79] C. P. Haugtvedt, R. E. Petty, and J. T. Cacioppo, "Need for cognition and advertising: Understanding the role of personality variables in consumer behavior," *Journal of Consumer Psychology*, vol. 1, no. 3, pp. 239–260, 1992.

[80] G. R. Leon, B. Gillum, R. Gillum, and M. Gouze, "Personality stability and change over a 30-year period—middle age to old age.," *Journal of Consulting and Clinical Psychology*, vol. 47, no. 3, pp. 517–524, 1979.

[81] J. McFarlane, C. L. Martin, and T. M. Williams, "Mood fluctuations," *Psychology of Women Quarterly*, vol. 12, no. 2, pp. 201–223, 1988.

[82] A. C. North, D. J. Hargreaves, and J. McKendrick, "The influence of in-store music on wine selections.," *Journal of Applied Psychology*, vol. 84, no. 2, p. 271, 1999.

[83] D. Kim and C. S. Areni, "The influence of background music on shopping behavior: Classical versus top-forty music in a wine store," *Advances in Consumer Research*, vol. 20, no. 1, pp. 336–340, 1993.

[84] L. W. Turley and R. E. Milliman, "Atmospheric effects on shopping behavior: a review of the experimental evidence," *Journal of business research*, vol. 49, no. 2, pp. 193–211, 2000.

[85] C. S. Areni and D. Kim, "The influence of background music on shopping behavior: classical versus top-forty music in a wine store," *NA-Advances in Consumer Research Volume 20*, 1993.

[86] R. S. Friedman, E. Gordis, and J. Förster, "Re-exploring the influence of sad mood on music preference," *Media Psychology*, vol. 15, no. 3, pp. 249–266, 2012.

[87] M.-L. Halko and M. Kaustia, "Risk on/risk off: Risk-taking varies with subjectively preferred and disliked music," *PLoS ONE*, vol. 10, no. 8, p. e0135436, 2015.

[88] E. Kauder, *History of Marginal Utility Theory*. Princeton University Press, 2015.

[89] W. McCown, R. Keiser, S. Mulhearn, and D. Williamson, "The role of personality and gender in preference for exaggerated bass in music," *Personality and Individual Differences*, vol. 23, no. 4, pp. 543–547, 1997.

[90] J. Cuzick, "A wilcoxon-type test for trend," *Statistics in medicine*, vol. 4, no. 4, pp. 543–547, 1985.

[91] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[92] P. G. Christenson and J. B. Peterson, "Genre and gender in the structure of music preferences," *Communication Research*, vol. 15, no. 3, pp. 282–301, 1988.

[93] D. F. Roberts and L. Henriksen, "Music listening vs. television viewing among older adolescents.," 1990.

[94] A. LeBlanc, Y. C. Jin, L. Stamou, and J. McCrary, "Effect of age, country, and gender on music listening preferences," *Bulletin of the council for Research in Music Education*, pp. 72–76, 1999.

[95] T. Schäfer and P. Sedlmeier, "From the functions of music to music preference," *Psychology of Music*, vol. 37, no. 3, pp. 279–300, 2009.

[96] A. C. North and J. W. Davidson, "Musical taste, employment, education, and global region," *Scandinavian journal of psychology*, vol. 54, no. 5, pp. 432–441, 2013.

[97] A. LeBlanc, W. L. Sims, C. Siivola, and M. Obert, "Music style preferences of different age listeners," *Journal of Research in Music Education*, vol. 44, no. 1, pp. 49–59, 1996.

[98] A. Warde, L. Martens, and W. Olsen, "Consumption and the problem of variety: cultural omnivorousness, social distinction and dining out," *Sociology*, vol. 33, no. 1, pp. 105–127, 1999.

[99] T. Katz-Gerro, "Highbrow cultural consumption and class distinction in italy, israel, west germany, sweden, and the united states," *Social forces*, vol. 81, no. 1, pp. 207–229, 2002.

[100] T. Katz-Gerro, "Cultural consumption research: review of methodology, theory, and consequence," *International Review of Sociology*, vol. 14, no. 1, pp. 11–29, 2004.

[101] T. W. Chan and J. H. Goldthorpe, "Social stratification and cultural consumption: Music in england," *European sociological review*, vol. 23, no. 1, pp. 1–19, 2007.

[102] T. W. Chan and J. H. Goldthorpe, "The social stratification of theatre, dance and cinema attendance," *Cultural Trends*, vol. 14, no. 3, pp. 193–212, 2005.

[103] O. Lizardo and S. Skiles, "Highbrow omnivorousness on the small screen?: Cultural industry systems and patterns of cultural choice in europe," *Poetics*, vol. 37, no. 1, pp. 1–23, 2009.

[104] T. W. Chan and J. H. Goldthorpe, "Social stratification and cultural consumption: The visual arts in england," *Poetics*, vol. 35, no. 2-3, pp. 168–190, 2007.

[105] O. Sullivan and T. Katz-Gerro, "The omnivore thesis revisited: Voracious cultural consumers," *European sociological review*, vol. 23, no. 2, pp. 123–137, 2007.

[106] J. López-Sintas and T. Katz-Gerro, "From exclusive to inclusive elitists and further: Twenty years of omnivorousness and cultural diversity in arts participation in the usa," *Poetics*, vol. 33, no. 5, pp. 299–319, 2005.

[107] B. Le Roux, H. Rouanet, M. Savage, and A. Warde, "Class and cultural division in the uk," *Sociology*, vol. 42, no. 6, pp. 1049–1071, 2008.

[108] M. Ul Haq, *Reflections on human development*. Oxford University Press, 1995.

[109] O. H. Chowdhury, "Human development index: A critique," *The Bangladesh development studies*, vol. 19, no. 3, pp. 125–127, 1991.

[110] A. D. Sagar and A. Najam, "The human development index: a critical review," *Ecological economics*, vol. 25, no. 3, pp. 249–264, 1998.

[111] D. A. Hicks, "The inequality-adjusted human development index: a constructive proposal," *World development*, vol. 25, no. 8, pp. 1283–1298, 1997.

[112] M. F. Ji and W. Wood, "Purchase and consumption habits: Not necessarily what you intend," *Journal of Consumer Psychology*, vol. 17, no. 4, pp. 261–276, 2007.

[113] B. L. Sturm, "An analysis of the gtzan music genre dataset," in *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pp. 7–12, ACM, 2012.

[114] J. Weiner and O. T. Solbrig, "The meaning and measurement of size hierarchies in plant populations," *Oecologia*, vol. 61, no. 3, pp. 334–336, 1984.

[115] D. L. Kirchman, M. T. Cottrell, and C. Lovejoy, "The structure of bacterial communities in the western arctic ocean as revealed by pyrosequencing of 16s rrna genes," *Environmental Microbiology*, vol. 12, no. 5, pp. 1132–1143, 2010.

[116] K. K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology," *Human-centric Computing and Information Sciences*, vol. 4, no. 1, p. 14, 2014.

[117] M. Tervaniemi, M. Huotilainen, and E. Brattico, "Melodic multi-feature paradigm reveals auditory profiles in music-sound encoding.," *Frontiers in human neuroscience*, vol. 8, no. July, p. 496, 2014.

[118] G. Vigliensoni and I. Fujinaga, "Automatic music recommendation systems: Do demographic, profiling, and contextual features improve their performance?,"

[119] H.-c. Chen and A. L. P. Chen, "A music recommendation system based on music data grouping and user interests," *CIKM '01 Proceedings of the tenth international conference on Information and knowledge management*, pp. 231–238, 2001.

[120] S. Lippens, J. Martens, and T. D. Mulder, "A comparison of human and automatic musical genre classification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings.*, vol. 4, pp. iv–233, IEEE, 2004.

[121] L. Molteni and A. Ordanini, "Consumption patterns, digital technology and music downloading," *Long Range Planning*, vol. 36, no. 4, pp. 389–406, 2003.

[122] C. L. Krumhansl, "An exploratory study of musical emotions and psychophysiology.," *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 51, no. 4, p. 336, 1997.

[123] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *Proc. ISMIR*, pp. 255–266, Citeseer, 2010.

[124] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, 2010.

[125] W. M. Wundt, *Outlines of psychology.* W. Engelmann, 1897.

[126] D. J. Hargreaves and A. C. North, "The functions of music in everyday life: Redefining the social in music psychology," *Psychology of music*, vol. 27, no. 1, pp. 71–83, 1999.

[127] P. Saari and T. Eerola, "The Role of Audio and Tags in Music Mood Prediction: A Study Using Semantic Layer Projection.," *Proceedings of the 14th Intentional Society for Music Information Retrieval Conference*, pp. 201–206, 2013.

[128] E. Lacourse, M. Claes, and M. Villeneuve, "Heavy metal music and adolescent suicidal risk," *Journal of Youth and Adolescence*, vol. 30, no. 3, pp. 321–332, 2001.

[129] P. W. Corrigan, "The impact of stigma on severe mental illness," *Cognitive and Behavioral Practice*, vol. 5, no. 2, pp. 201–222, 1999.

[130] I. Kawachi and L. F. Berkman, "Social ties and mental health," *Journal of Urban Health*, vol. 78, no. 3, pp. 458–467, 2001.

[131] J. Whipple, "Music in intervention for children and adolescents with autism: A meta-analysis," *Journal of Music Therapy*, vol. 41, no. 2, pp. 90–106, 2004.

[132] A. A. Clair and J. Memmott, *Therapeutic uses of music with older adults.* ERIC, 2008.

[133] P. G. Hunter, E. G. Schellenberg, and U. Schimmack, "Mixed affective responses to music with conflicting cues," *Cognition & Emotion*, vol. 22, no. 2, pp. 327–352, 2008.

[134] "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2011.

[135] Y. Hu, X. Chen, and D. Yang, "Lyric-based song emotion detection with affective lexicon and fuzzy clustering method.," in *Proceedings of the 10th Intentional Society for Music Information Retrieval Conference*, pp. 123–128, 2009.

[136] R. Mayer, R. Neumayer, and A. Rauber, "Rhyme and style features for musical genre classification by song lyrics.," in *Proceedings of the 9th Intentional Society for Music Information Retrieval Conference*, pp. 337–342, 2008.

[137] T. Danisman and A. Alpkocak, "Feeler: Emotion classification of text using vector space model," in *Artificial Intelligence and Simulation of Behaviour Convention on Communication, Interaction and Social Intelligence*, vol. 1, p. 53, 2008.

[138] M. Van Zaanen and P. Kanters, "Automatic mood classification using tf* idf based on lyrics.," in *ISMIR*, pp. 75–80, 2010.

[139] C. D. Manning and H. Schütze, *Foundations of statistical natural language processing*, vol. 999. MIT Press, 1999.

[140] T. Joachims, *Learning to classify text using support vector machines: Methods, theory and algorithms.* Kluwer Academic Publishers, 2002.

[141] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 160–167, ACM, 2008.

7   Appendix

7.1   Nokia DB Genre List

Exhaustive Genre List:
African Ethnic/Traditional, Afrikaans, Ambient/New Age, Asian, Australian Music, Bengali, Bhojpuri, Blues, Bollywood, Cancion de autor, Cantonese Pop, Chinese Original, Christian/Gospel, Christmas, Classic Arabic/Tarab, Classical, Comedy, Country and Western, Dance, Dangdut, DeutschPop, Devotional, Easy Listening/Oldies, Electronica, Flamenco, Folk/Roots, French Variety, Ghazal, Greek, Gujarati, Hindustani Classical, Indie/Alternative, Indipop and Remix, Instrumental, Irish Music, Iskelda, Islamiyat, Italiana, Jazz, Keroncong, Khaleeji, Kids/Novelty, Latin Music, Malayalam, Mandarin Pop, Marathi, Metal, Nederlandstalig, Oriental Pop, Other, Pop, Punjabi, Punk, Rap/Hip Hop, Reggae, Regional/Others, Rock, Schläger, Shanson, Soul/R 'n' B/Funk, Soundtracks/Musicals/Theatre, South Devotional, Spanish pop rock, Spoken Word, Tamil, Telugu, and World.

Table 16: Exhaustive list of all genres included in Nokia DB.

7.2   User Distribution Statistics

7.3   Analysis 1: Supplemental Information
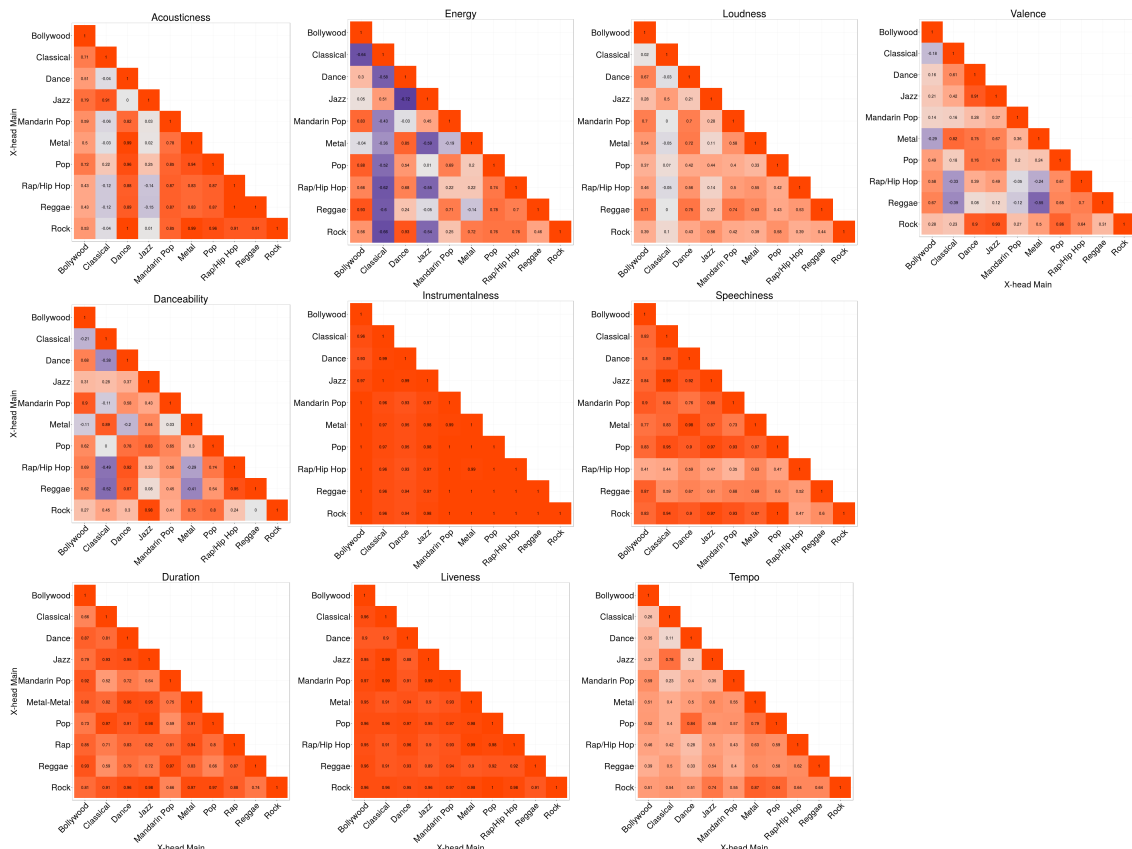
Figure 17: Correlation Matrices for all features
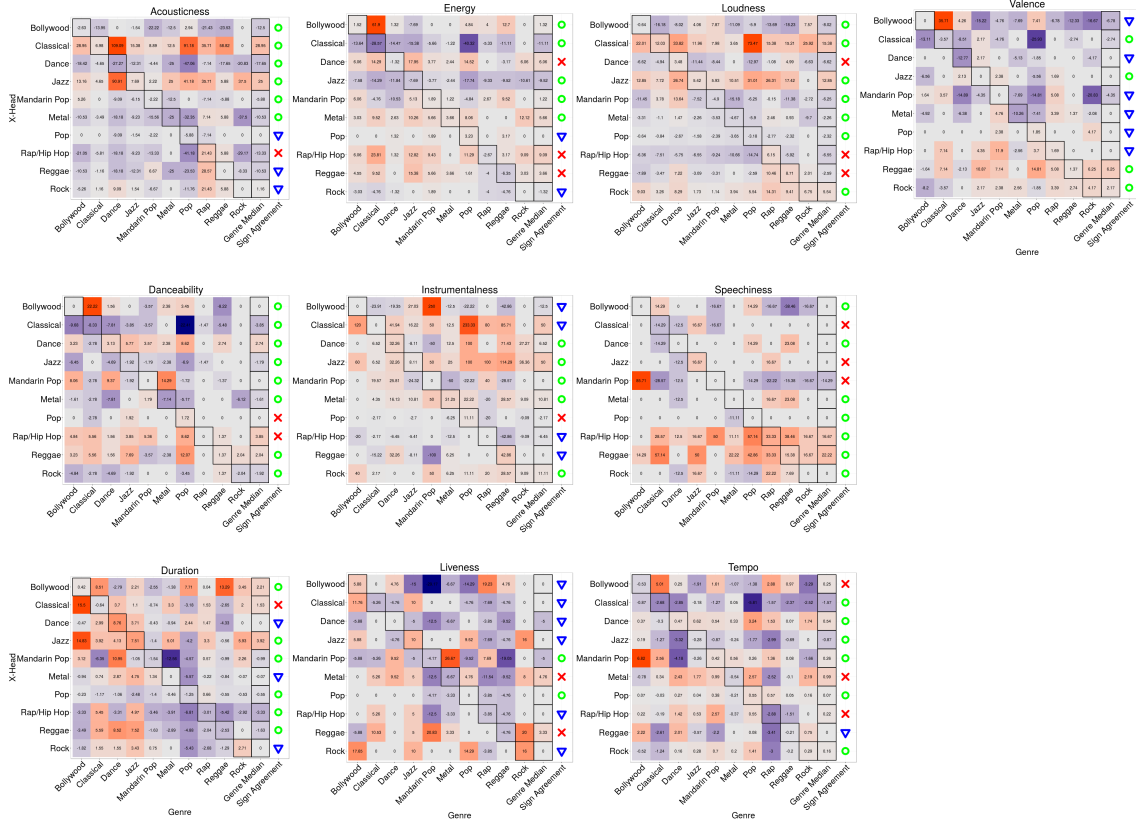
Figure 18: Leaky Features Matrices for all features

7.4 Analysis 2: Supplemental Information

7.4.1 Adventurousness Calculation

In order to scale adventurousness so that highly and lowly popular music does not dramatically influence averaging, the logarithm of each tracks total download count is used. In Table 20, the first track in each user's collection was downloaded 10 million of times would receive an adventurousness score of 7, or 1000 times, receiving a score of 3. User's are considered more adventurous if they explore less popular (in the case of Nokia DB, less downloaded) music. After collecting the log converted total downloads for each track in a user's collection, an average is calculated. Users with high averages are considered less adventurousness and do not range from 0-1, we convert and scale values in equation 8 such that highly adventurous users are closer to 1, and less adventurous users are closer to 0.

$$A_u = 1 - \frac{\bar{A}_{\log t}}{\max(\bar{A}_{\log t})} \tag{8}$$

where:

A_u: User adventurousness value

$\bar{A}_{\log t}$: Average of $\log t$ from each track in user

$\max(\bar{A}_{\log t})$: Maximum average of $\log t$ for any user in the database.

7.4.2 Volume Calculation

Equation 9 is a measure of user volume which considers total amount of downloads by a user over the total days of active service, defined as the amount of days between their first and last download:

$$U_V = \frac{L}{\sum DL} \tag{9}$$

where:

L**:** Total days between first and last download

DL**:** Total downloads in a user's collection

U_V**:** User volume value

7.4.3   Overall Regularity Calculation

For each user, an array is created with values that represent the number of days since last download. The numbers in each array increase as more days go without any download behaviour. In the Table 21 example, User 1 and 2 both download their first and last track on the same day, but User 1 downloads an additional song on day 4, and User 2 downloads 2 songs on the day 2.

$$U_O = \frac{1}{\sigma_{(OR)}} \tag{10}$$

where:

U_O**:** User overall regularity

$\sigma_{(}OR)$**:** SD of downloads across days

With the calculation for $\sigma_{(OR)}$, users with a high degree of regularity have a lower value; the inverse of the standard deviation is used to reverse this relationship for intuitive ease. In the rare case where users are completely regular (i.e. they download the same number of songs across the same interval, creating a SD in the denominator of 0) are assigned a value of 1.

7.4.4   Periodic (7-day) Regularity Calculation

Table 22 illustrates the calculation of periodic regularity for a single user. Each download times-tamp is converted to a day of the week (1-7; representing Monday to Sunday). Each cell in the table represents a users consumption history across day of the week and week number. The standard deviation is calculated for each day of the week, and then averaged across the 7 days. As with overall regularity, high average standard deviations represent low degrees of periodic download behaviour, whereas users with a high degree of regularity would have low values. This calculation is inverted using the same method for overall regularity; Equation 10.

7.5   Stigma Mental Health Application

Although some behavioural measures map well to neural or cognitive models (i.e. Mel frequency cepstral coefficients and timbre[15]), others (i.e. emotion [123]) have not achieved similar progress. This is important for MIR because research that hastily extends contentious cognitive theories of music perception as valid indicators of human behaviour may hinder future work when replication becomes difficult. For example, representations of emotion in MIR may be methodologically inconsistent because of neurological and affective differences between perceived versus felt emotions [122]. Furthermore, models exploring music-induced emotion in MIR have traditionally use two-dimensional arousal-valence models, which may be too reductive for complex behaviours like music perception [124]; three-dimensional models of emotion using tension, arousal, and valence are as early as Wundt [125] and align well with representations of emotion induction in music theory. The relative absence of accurate emotion classification in MIR may be influenced by these inconsistent representations of poorly-understood complex human processes.

Recent advances suggests that music contributes to emotional well-being in three primary ways: regulating mood, reinforcing self identity, and by facilitating social bonding (particularly in the context of dance) [126]. Mood and mood-management disorders constitute a complete category in the Diagnostic and Statistical Manual of Mental Disorders; studies have shown that music-listening is used by adolescents to facilitate emotion regulation [127], and is negatively correlated with suicide-risk in secondary school girls [128]. Individuals with mental health issues frequently perceive themselves as stigmatized, which negatively impacts self-esteem, self-image and behaviour [129]. Social withdrawal is particularly prevalent in individuals with mental-health issues [130]; music-based therapies can improve social behaviour in individuals with autism [131], and help facilitate social integration [132].

However, methodological problems also exist; a common issue is the over reliance on western classical music, handpicked, and generally conveying discrete emotional categories in order to maintain experimental control. Moreover, opposite emotions (e.g. happy, sad) in low-dimension models fail to encapsulate subtle affective experiences; participants attribute *both* labels (happy *and* sad) when listening to musical excerpts with mixed emotional cues [133]. Which is to say, binary emotional categories, such as happy-sad, may not reflect natural listening responses to music. As a result, emotion models using more dimensions, such as arousal, tension, and energy, may be better able to capture the dynamic associations between music, emotion and mood, not least because they would more closely align with musical perceptions [134]. Most importantly is the simple fact that much (classical) music-emotion research lacks sung lyrics, a common feature in other musical styles that can significantly contribute to our emotional responses. Song lyrics are an informative resource in automated classification of song mood [135], and genre [136]. To our knowledge, no detailed research exists which examines the linguistic relationship between song lyrics and the emotions documented during introspective journaling.

When linked to GRAIL (Section 2.2), using a mental-health journaling application such as *Stigma*, detailed qualitative analysis of subjective emotional reports while listening to music becomes possible. *Stigma* is a journaling, mood tracking application in which users record personal, anonymized entries tagged with one of 16 emotion labels (angry, sad, stressed, frustrated, down, lonely, anxious, overwhelmed, tired, okay, calm, good, productive, accomplished, happy, excited, ecstatic); see Figure 1. *Stigma* is highly popular with its users, and is the top rated mental-health application on the Apple store. Without advertisement, *Stigma* has collected 1.5 million mood-tagged journal entries, from ca. 10 thousand monthly active users. *Stigma* also features personality trait questionnaires, which could be used to refine emotional intensity calculations (e.g. the responses of introverts, who tend to understate their emotions, can be subject to a multiplier, for example).

Emotional intensity can be assessed using a variety of techniques. NLP, including word and phrase tokenization, and term frequency-inverse document frequency (TF-IDF) [137]. TF-IDF has been used to classify emotional categories from human-annotated text and song lyrics [138]. Using TF-IDF scores, valence dictionaries can be built to derive an emotional intensity value for specific entries. Machine learning techniques have been employed for classification of text [139]. And since *Stigma's* entries are tagged, it will be possible to assess clusters for parts-of-speech using support vector machines [140] and neural networks with back propagation [141]. Classification confidence from trained models will be used alongside derived measures of intensity to improve performance.

To our knowledge no research initiatives exists in either scope or scale that combine introspective journaling with concurrent music-listening data. As a result, our project will have important

implications for the development of sophisticated recommender systems. Therapeutic applications for at-risk individuals could provide medical practitioners with non-invasive, preventative interventions. Automated analysis of music listening and emotional tags could provide personalized monitoring of mental health. Customized playlists using machine learning to select music which is maximally associated with an individual's positive affect could be generated. Linguistic research will assess the role that lyrics play with respect to emotional states. Lastly, cognitive psychologists and musicologists will be able to explore the relationship between the mood labels used to tag songs, with the mood labels we give to songs when regulated by our emotions.
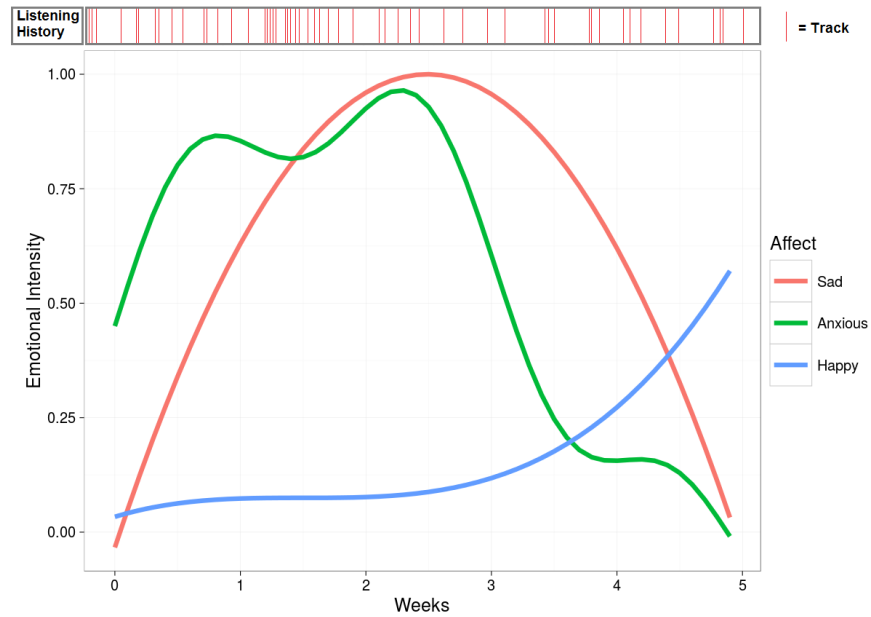

Figure 19: Graph showing hypothetical data relating to changing affect intensity (y-axis) over time (x-axis). Emotional intensity will be ascertained using NLP, and machine learning classification. Red ticks above the line graph represent date/time of specific tracks in a user's listening history. These data will be mapped to the intensities of the emotion tags.

Xhead	Average	Mode	Standard Deviation	Skew	Kurtosis
African Ethnic/Traditional	0.525	1.0	0.01	0.95	0.284
Afrikaans	0.542	1.0	0.01	3.977	27.696
Ambient/New Age	0.505	1.0	0.009	1.697	5.832
Asian	0.448	0.36	0.01	1.144	0.612
Bengali	0.423	0.36	0.01	0.96	-0.223
Bhojpuri	0.406	0.35	0.013	1.328	0.42
Blues	0.522	1.0	0.011	2.785	12.724
Bollywood	0.545	0.5	0.01	0.694	-0.9
Cancion de autor	0.44	1.0	0.031	3.226	10.398
Cantonese Pop	0.617	1.0	0.009	1.667	5.98
Christian/Gospel	0.578	1.0	0.011	5.395	42.001
Christmas	0.507	1.0	0.01	2.029	7.485
Classic Arabic/Tarab	0.465	0.54	0.057	5.541	28.697
Classical	0.582	1.0	0.012	6.487	54.389
Comedy	0.567	1.0	0.011	5.772	46.289
Country and Western	0.524	1.0	0.009	1.742	5.236
Dance	0.555	1.0	0.01	4.926	37.119
Dangdut	0.634	1.0	0.01	3.636	24.565
Deutsch Pop	0.549	1.0	0.013	2.212	6.853
Devotional	0.427	0.38	0.01	0.932	-0.466
Easy Listening/Oldies	0.526	1.0	0.011	4.668	33.138
Electronica	0.519	1.0	0.012	4.948	35.612
Flamenco	0.591	1.0	0.016	5.492	38.939
Folk/Roots	0.539	1.0	0.01	3.42	20.456
French Variations	0.55	1.0	0.018	7.107	60.479
Ghazal	0.476	0.38	0.009	0.566	-0.894
Gujarati	0.438	0.38	0.011	0.978	-0.2
Hindustani Classical	0.511	1.0	0.013	5.872	46.052
Indie/Alternative	0.498	1.0	0.011	2.292	8.402
Indipop and Remix	0.46	0.38	0.01	0.767	-0.424
Instrumental	0.419	0.45	0.032	2.884	6.32
Irish Music	0.53	1.0	0.017	5.702	43.708
Iskelda	0.657	1.0	0.035	3.766	14.635
Islamiyat	0.508	0.48	0.01	1.019	0.651
Italiana	0.617	1.0	0.023	8.565	78.638
Jazz	0.535	1.0	0.011	5.371	40.96
Keroncong	0.525	0.5	0.017	2.77	9.027
Khaleeji	0.504	0.44	0.009	0.653	-0.262
Kids/Novelty	0.536	1.0	0.009	2.602	13.699
Latin Music	0.492	1.0	0.009	1.127	2.133
Malayalam	0.415	0.38	0.013	1.353	0.612
Mandarin Pop	0.73	1.0	0.011	3.336	21.233
Marathi	0.431	0.41	0.012	1.279	0.576
Metal	0.543	1.0	0.01	3.81	25.286
Nederlandstalig	0.603	1.0	0.012	4.87	35.33
Oriental Pop	0.489	0.52	0.01	0.932	-0.391
Pop	0.564	1.0	0.009	1.55	5.603
Punjabi	0.444	0.38	0.01	0.788	-0.691
Rap/Hip Hop	0.539	1.0	0.01	3.822	25.8
Reggae	0.531	1.0	0.012	5.823	46.412
Regional/Others	0.419	0.33	0.011	1.011	-0.288
Rock	0.512	1.0	0.01	3.076	17.932
Schlager	0.596	1.0	0.014	6.304	51.271
Shanson	0.603	1.0	0.012	5.277	40.079
Soul/RnB/Funk	0.514	1.0	0.011	2.714	12.768
Soundtracks/Musicals/Theatre	0.54	1.0	0.013	6.031	48.63
Spoken Word	0.572	1.0	0.014	7.541	66.612
Tamil	0.474	0.38	0.011	0.918	-0.527
Telugu	0.42	0.38	0.013	1.253	0.188
World	0.494	1.0	0.009	1.635	5.235

Table 17: X-head distribution summary statistics

X-head	Feature	JT value	JT p-value	KT value	KT p-value
Rap/Hip Hop	Speechiness	4420833.225	0	0.12491392	5.08E-21
Easy Listening/Oldies	Speechiness	3483741.525	1.43E-30	-0.148123249	2.12E-32
Easy Listening/Oldies	Acousticness	3509958.405	1.18E-27	-0.140552878	2.64E-29
Easy Listening/Oldies	Energy	4326665.495	7.26E-13	0.095272168	1.58E-13
Easy Listening/Oldies	Loudness	4323052.59	1.15E-12	0.094302639	2.49E-13
Classical	Loudness	3702954.72	2.17E-08	-0.085093698	9.04E-09
Classical	Energy	3706545.725	3.03E-08	-0.084104458	1.27E-08
Dance	Liveness	4210263.695	3.41E-08	0.085957604	1.45E-08
Easy Listening/Oldies	Liveness	3457439.455	4.58E-08	-0.073989712	1.87E-08
Easy Listening/Oldies	Instrumentalness	3452329.965	1.19E-07	-0.075137635	5.78E-08
Dance	Instrumentalness	4210658.875	4.61E-07	0.086140214	2.46E-07
Reggae	Energy	3773462.645	1.67E-05	-0.063703114	1.03E-05
Reggae	Speechiness	4201039.445	2.85E-05	0.059831632	1.75E-05
Metal	Acousticness	3769708.685	4.57E-05	-0.06439462	3.09E-05
Metal	Liveness	4162542.96	0.000146414	0.06104769	0.000101317
Metal	Instrumentalness	4158687.675	0.000174479	0.059759437	0.000124271
Dance	Loudness	3802907.725	0.000470196	-0.054276528	0.000346346
Rap/Hip Hop	Tempo	3792420.385	0.000600026	-0.056716182	0.000477608
Indie/Alternative	Danceability	3795041.915	0.000626752	-0.056934399	0.000479848
Metal	Danceability	3809333.595	0.000767723	-0.053938317	0.000585056
Rap/Hip Hop	Loudness	3811668.515	0.000983393	-0.050979958	0.000751357
Classical	Danceability	3799482.445	0.00111005	-0.057267287	0.000914941
Rap/Hip Hop	Energy	3807555.425	0.001194797	-0.052338623	0.000942086
Pop	Loudness	4176554.4	0.001366168	0.052823373	0.001071435
Jazz	Instrumentalness	4138031.56	0.002094058	0.046914829	0.001683323
World	Loudness	4126809.52	0.002310272	0.051064518	0.001893119
World	Danceability	4120797.03	0.002731817	0.049438012	0.002234467
Jazz	Liveness	4140682.89	0.003141327	0.047777196	0.00256336
Easy Listening/Oldies	Mode	1373230.74	0.003389393	0.047338072	0.002594141
Rock	Loudness	3837985.35	0.004482714	-0.046008169	0.003747369
Dance	Tempo	4148610.84	0.005072021	0.046330913	0.00432647
Reggae	Loudness	3862190.335	0.008443676	-0.037985968	0.006949752
Metal	Valence	3855420.78	0.010277592	-0.04063668	0.008966439
Dance	Acousticness	3829994.22	0.012267716	-0.04644522	0.011123995
Metal	Energy	4136202.535	0.014187845	0.040459403	0.0123767751
Pop	Liveness	3333950.37	0.019457481	-0.039706259	0.017301732
Pop	Instrumentalness	3324721.405	0.019963691	-0.042221331	0.018013182
Rap/Hip Hop	Danceability	4118192.445	0.020945384	0.037469885	0.018736737
Mandarin Pop	Loudness	4083001.82	0.025359759	0.034072809	0.022178097
Metal	Speechiness	4124863.955	0.026107106	0.03719727	0.023439102
Indie/Alternative	Instrumentalness	4048017.01	0.04031904	0.033406131	0.036238921
Rap/Hip Hop	Liveness	2392465.12	0.043350052	-0.038637571	0.040514344
Dance	Valence	3873661.13	0.046940941	-0.033883289	0.043581095
Pop	Acousticness	3883316.605	0.053402234	-0.031996299	0.049285072
Rap/Hip Hop	Instrumentalness	2400866.635	0.054940394	-0.036013261	0.051331202
Rock	Valence	4112147.94	0.055357385	0.033061069	0.05216693
Rock	Instrumentalness	3966354.31	0.061004966	0.03232337	0.056820527
Indie/Alternative	Liveness	4035347.78	0.068423033	0.029564819	0.063453665
Pop	Energy	4101348.62	0.070413574	0.03097653	0.066164627
Rock	Liveness	3963854.315	0.072977068	0.031477018	0.068418915
Dance	Speechiness	3898293.015	0.103202653	-0.026821199	0.097596423

Table 18: Jonchkeere and Kendall-Tau-β results. Are based on an average of bootstrap analysis for 100 random samples.

Human Development Statistics Utilized:
Combined gross enrolment in education both sexes, adult literacy rate both sexes, expenditure on public education as percent of GDP, expected years of schooling females, expected years of schooling males, pre-primary enrolment rate, gross primary enrolment ratio,gross secondary enrolment ratio, gross tertiary enrolment ratio, HDI education index, HDI health index,HDI income index„HDI human development index, percent of people with internet access

Table 19: Exhaustive list of all metrics included from HDR.

Log of total downloads for selected track: $\log t$	User 1	User 2	User 3
1	3	7	14
2	1	1	14
3	1	3	14
$\bar{A}\log t$ **per user:**	0.94	2.49	$\max \mu(14)$
Adventurousness per user (A_U)	0.93	0.82	0

Table 20: Table representation of Adventurousness calculation.

Day	User 1	User 2
1	1 (NA)	1 (NA)
2	0 (1)	**2 (1,0)**
3	0 (2)	0 (1)
4	**1 (3)**	0 (2)
5	0 (1)	0 (3)
6	0 (2)	0 (4)
7	**1 (3)**	**1 (5)**
Overall regularity:	$\frac{1}{\sigma(3,3)} = 1$	$\frac{1}{\sigma(1,0,5)} = 0.46$

Table 21: Table representation of overall regularity

Day of Week	Week 1	Week 2	Week 3	SD
1	1	1	1	0
2	0	2	0	0.94
3	3	5	1	1.63
...				...
7	3	3	3	0
$\bar{\sigma_{pr}}$				**0.68**
Periodic Regularity				**1.45**

Table 22: Table representation of Periodic regularity