Cultural differences in music features across Taiwanese, Japanese, and American markets

Kongmeng Liew [Corresp., 1, 2], Yukiko Uchida [3], Igor de Almeida [4]

1 Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma City, Nara Prefecture, Japan

2 Graduate School of Human and Environmental Studies, Kyoto University, Kyoto City, Kyoto Prefecture, Japan

3 Kokoro Research Center, Kyoto University, Kyoto City, Kyoto Prefecture, Japan

4 Institute of Liberal Arts, Otemon Gakuin University, Ibaraki City, Osaka Prefecture, Japan

Corresponding Author: Kongmeng Liew
Email address: liew.kongmeng@is.naist.jp

Background. Preferences for music can be represented through music features. The widespread prevalence of music streaming has allowed for music feature information to be consolidated by service providers like Spotify. In this paper, we demonstrate that machine learning classification on cultural market membership (Taiwanese, Japanese, American) by music features reveals variations in popular music across these markets.

Methods. We present an exploratory analysis of 1.08 million songs centred on Taiwanese, Japanese and American markets. We use both multiclass classification models (Gradient Boosted Decision Trees (GBDT) and Multilayer Perceptron (MLP)), and binary classification models, and interpret their results using variable importance measures and Partial Dependence Plots. To ensure the reliability of our interpretations, we conducted a follow-up study comparing Top-50 playlists from Taiwan, Japan, and the US on identified variables of importance.

Results. The multiclass models achieved moderate classification accuracy (GBDT = 0.69, MLP = 0.66). Accuracy scores for binary classification models ranged between 0.71 to 0.81. Model interpretation revealed music features of greatest importance: Overall, popular music in Taiwan was characterised by high acousticness, American music was characterized high speechiness, and Japanese music was characterized by high energy features. A follow-up study using Top-50 charts found similarly significant differences between cultures for these three features.

Conclusion. We demonstrate that machine learning can reveal both the magnitude of differences in music preference across Taiwanese, Japanese, and American markets, and where these preferences are different. While this paper is limited to Spotify data, it underscores the potential contribution of machine learning in exploratory approaches to research on cultural differences.

1 # Cultural Differences in Music Features Across Taiwanese,
2 # Japanese, and American Markets

4 Kongmeng Liew[1,2], Yukiko Uchida[3], Igor de Almeida[4]

6 [1] Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma,
7 Nara Prefecture, Japan
8 [2] Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Kyoto
9 Prefecture, Japan
10 [3] Kokoro Research Center, Kyoto University, Kyoto, Kyoto Prefecture, Japan
11 [4] Institute of Liberal Arts, Otemon Gakuin University, Ibaraki, Osaka Prefecture, Japan

13 Corresponding Author:
14 Kongmeng Liew[1]
15 8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan
16 Email address: liew.kongmeng@is.naist.jp

17   **Abstract**

19   **Background.** Preferences for music can be represented through music features. The widespread

20   prevalence of music streaming has allowed for music feature information to be consolidated by

21   service providers like Spotify. In this paper, we demonstrate that machine learning classification

22   on cultural market membership (Taiwanese, Japanese, American) by music features reveals

23   variations in popular music across these markets.

25   **Methods.** We present an exploratory analysis of 1.08 million songs centred on Taiwanese,

26   Japanese and American markets. We use both multiclass classification models (Gradient Boosted

27   Decision Trees (GBDT) and Multilayer Perceptron (MLP)), and binary classification models,

28   and interpret their results using variable importance measures and Partial Dependence Plots. To

29   ensure the reliability of our interpretations, we conducted a follow-up study comparing Top-50

30   playlists from Taiwan, Japan, and the US on identified variables of importance.

31

32   **Results.** The multiclass models achieved moderate classification accuracy (GBDT = 0.69, MLP

33   = 0.66). Accuracy scores for binary classification models ranged between 0.71 to 0.81. Model

34   interpretation revealed music features of greatest importance: Overall, popular music in Taiwan

35   was characterised by high acousticness, American music was characterized high speechiness,

36   and Japanese music was characterized by high energy features. A follow-up study using Top-50

37   charts found similarly significant differences between cultures for these three features.

38

39   **Conclusion.** We demonstrate that machine learning can reveal both the magnitude of differences

40   in music preference across Taiwanese, Japanese, and American markets, and where these

41   preferences are different. While this paper is limited to Spotify data, it underscores the potential

42   contribution of machine learning in exploratory approaches to research on cultural differences.

43

44   **Introduction**

45   With 219 million active listeners a month, and a presence in over 60 countries, Spotify is

46   one of the largest music streaming service providers in the world (as of 2020, Schwind et al.,

47   2020). To facilitate such a service, they maintain a database of music features for all songs in

48   their service, that is made publicly accessible through the Spotify API (Application Programme

49   Interface). This database contains a wealth of meta and music feature data, that researchers have

50   been using to research human behaviour and engagement with music. For example, Park and

51   colleagues (Park et al., 2019) analysed data from 1 million individuals across 51 countries and

52   uncovered consistent patterns of music preferences across day-night cycles: relaxing music was

53   more commonly played at night, and energetic music during the day. Pérez-Verdejo and

54   colleagues (2020) found that popular hit songs in Mexico shared many similarities to global hit

55   songs. Spotify's music features have also been used to examine songs in clinical and therapeutic

56   settings, with Howlin and Rooney (2020) finding that songs used in previous pain management

57   research, if chosen by the patient, tended to have high energy, danceability, and lower

58   instrumentalness features, than experimenter-chosen songs.

59   In this paper, we use Spotify features to examine how cultures differ in music

60   preferences, through a bottom-up, data-driven analysis of music features across cultures. Here,

61   we quantify music preference through music features, following past research (e.g., Fricke et al.,

62   2019). Music plays a huge role in human society, be it in emotion regulation or for social

63   displays of identity and social bonding (Groarke & Hogan, 2018; Dunbar, 2012). These are often

64   embedded in the cultural norms and traditions of the listener. In other words, understanding how

65   music differs across cultures may reflect corresponding cultural differences in the sociocultural

66   context that shape our individual preferences towards certain types of music over others. Thus,

67   understanding how music differs between these cultural markets, e.g., by examining their

68   features, may then shed light on possible cultural differences, which can guide follow-up

69   research by generating novel hypotheses, or in supporting various theories on cultural differences

70   aside from music (see below).

71   To achieve this aim, we rely on machine learning classification of songs based on cultural

72   markets (i.e., culture of origin), as interpretation of these models may reveal insight into how

73   (and which) features differ between cultures. We utilize data seeded originally on Taiwanese,

74   American and Japanese Top-50 lists. This arguably aligns more towards a sociolingual

75   distinction of cultural membership, and not a country-based sampling commonly used in culture

76   research. Our reasoning was that this more closely resembles the differentiated cultural markets

77   that artists of a certain language operate in, that often transcend country boundaries. For

78   example, popular Chinese music meant for a Chinese cultural market often subsumes artists from

79   wider Chinese cultural origins (in countries and territories like Taiwan, Hong Kong, Singapore,

80   and Malaysia; Fung, 2013; Moskowitz, 2009). As such, in this paper, these were considered to
81   belong to a 'Chinese' cultural market (including music from language subtypes and dialects).
82   Japanese music was similarly treated as belonging to a 'Japanese' market, and music from
83   Western (Anglo-European) cultural origins (e.g., the US, UK, Canada, and Australia) were
84   considered as belonging to a 'Western' cultural market.
85   Past approaches towards music preference through Spotify have largely focused on
86   curated lists of Top-50 or Top-200 popular songs (e.g., Pérez-Verdejo et al., 2020; Febrirautami,
87   Surjandari & Laoh, 2019). While such lists are often diluted by the inclusion of 'global' hit
88   songs, they nevertheless provide a window to examine culturally based music preferences.
89   Accordingly, we also conduct a follow-up study using these Top-50 lists from Taiwan, Japan,
90   and the US to ensure the reliability of our interpretations (from the classification model).
91

Cultural differences in music preference
93   Typically, most of the research in cultural differences in the psychological literature has
94   come from top-down, theoretical approaches. These have been instrumental in shaping the field,
95   by increasing awareness of systematic ways by which people from cultures are different. One of
96   the most instrumental differences is in the independence and interdependence of the self (Markus
97   & Kitayama, 1991; 2010). Westerners generally tend to be independent, in that they prioritise the
98   autonomy and uniqueness of internal (self) attributes. In contrast, East Asians generally tend to
99   be interdependent, where their concept of self is intricately linked to close social relationships.
100   This has been shown to have implications on music preferences through differences in
101   desirability of emotions. For example, Westerners generally tend to view happiness as a positive
102   and internal hedonic experience to be maximised where possible (Joshanloo & Weijers, 2014).
103   As such, Western music preference tends towards high-arousal music, possibly in the search of
104   strong 'happiness' experiences. East Asians, however, view happiness as a positive feeling
105   associated with harmonious social relationships, in contrast to the hedonic, high-arousal
106   definition in Western contexts (citation). Consistently, East Asian preferences for music do not
107   have this high-arousal component, and is more calm, subdued, and relaxing (Tsai, 2007; Uchida
108   & Kitayama, 2009; Park et al., 2019).
109   However, such theoretically based analyses may overrepresent Western cultures in
110   research and literature. Consequently, cultural differences within similar, non-Westernised
111   spheres are not well understood, due to the lack of pre-existing theory. For example, Chinese and
112   Japanese cultures are often grouped together in cross-cultural research as a representation of East
113   Asian collectivism, that functions as a comparative antithesis to Western findings (e.g., Heine &
114   Hamamura, 2007). Yet, research has also uncovered differences between China and Japan that
115   cannot be explained by these theories (Muthukrishna et al., 2020). As such, cultural differences
116   within East Asia are not well understood in the psychological literature, and few theories exist to
117   offer predictions on differences in music preference between these cultures.
118   Our solution was to examine music as cultural products from the bottom-up. Doing so
119   would reduce the effect of experimenter bias in guiding theory formation and interpretation,

120   when examining a wide database of music features. Cultural products are behavioural
121   manifestations of culture that embody the shared values and collective aesthetics of a society
122   (Morling & Lamoreaux, 2008; Lamoreaux & Morling, 2012; Smith et al., 2013). This implies
123   that music consumption behaviour underscores culturally based attitudes, cognitions and
124   emotions that afford preferences for certain congruent types of music. For example, in a cross-
125   cultural comparison between Brazil and Japan, de Almeida and Uchida (2018) found that
126   Brazilian song lyrics contained higher frequencies of positive emotion words and lower
127   frequencies of neutral words than Japanese lyrics. This was consistent and reflective of their
128   respective cultural emphases on emotion expressions (see Triandis et al., 1984; Uchida &
129   Kitayama, 2009), and showed that comparing music 'products' elucidated differences between
130   the collective shared values of different cultures. Past research on cultural products have relied
131   on both popularity lists (charts; e.g., Askin & Mauskapf, 2017), and on artifacts produced by a
132   culture (e.g., Tweets: Golder & Macy, 2011; newspaper articles: Bardi, Calogero, & Mullen,
133   2008). We utilize both methods, and propose that examining cultural differences in 'music'
134   products on a large-scale may provide potential insight into the sociocultural circumstances that
135   give rise to these differences.

136

The Present Research

138   We adopted a data-driven, bottom-up approach to explore music preferences between
139   cultures/industries through musical features for this study through machine learning. I.e., we first
140   train a multiclass model to classify songs as belonging to (originating from) Chinese
141   (Taiwanese), Japanese, or Western (American) markets. This is to establish the presence and
142   magnitude of discernible cultural differences in music features. Next, we decompose the model
143   by training 3 binary machine learning classifiers to classify songs as belonging to one culture or
144   another. By applying model interpretation techniques on these models (such as Partial
145   Dependence Plots [PDPs]), we aim to discover the specific difference in preferred musical
146   features between Chinese (Taiwanese)-Japanese markets, Chinese (Taiwanese)-Western
147   (American) markets, and Japanese-Western (American) markets. We aimed to include as many
148   songs as possible that were produced from these respective culture-based music industries to
149   observe systematic trends and differences from as wide a range of musical styles and genres
150   within these industries as possible. Finally, we examine the generalizability of these
151   interpretations by conducting a follow-up study on Top-50 songs from Taiwan, Japan, and the
152   US. If the identified features of difference present in songs produced by a cultural market were
153   indeed representative of cultural differences in music preferences, we expect that these features
154   should also show consistent differences for their respective Top-50 (popular) songs.

155

Materials & Methods
Overview

158   To explore differences between cultures, we used machine learning to classify a database
159   of Chinese, Japanese, and English songs into their respective cultural (linguistic) markets. As this

160 was a multiclass classification problem, we conducted the analysis twice using gradient boosted
161 decision trees (GBDTs) and artificial neural networks (multi-layer perceptron, MLP) that are
162 inherently capable of multiclass classification. This was also to examine the consistency in
163 results between two differing methods of analysis, and strengthen the reliability of the analyses.
164 To infer the features that accounted for cultural differences, we use model interpretation
165 techniques, namely relative feature importance (RFI; Friedman, 2001), permutational feature
166 importance (PFI; Fisher, Rudin, & Dominici, 2018), and partial dependence plots (PDPs,
167 Friedman, 2001), to examine and visualise the relationships between the feature and its influence
168 on the probability of classification.
169

170 **Data mining**
171   We accessed the Spotify Application Programme Interface (API) through the 'spotifyr'
172 wrapper (Thompson, Parry, & Wolff, 2019) in R, to obtain song-level music feature information
173 from Chinese, Japanese and English artists from the Spotify database. This was through a
174 pseudo-snowball sampling method: we relied on Spotify's recommendation systems (the
175 'get_related_artists' function) to recommend artists related to those in the official Spotify Top-50
176 chart playlists for Taiwan, Japan, and the US respectively and created a list of artists per country.
177 We then used the same method to obtain another list of recommended artists to these respective
178 'lists', for up to 6 iterations, in order to obtain comparable sample sizes between these three
179 markets. We also excluded all non-Chinese, non-Japanese, and non-English (language) artists
180 from the respective list. This was through an examination of the associated genres for each artist,
181 which often contained hints to their cultural origins (e.g., J-pop, J-rock, Mandopop). Artists that
182 did not have listed genres were checked manually by the researchers. This resulted in a final
183 N(artists) = 10259 (Japanese = 2587; English = 2466; Chinese = 5206). All song-level feature
184 information for all artists were then obtained from the Spotify database. Duplicates (such as the
185 same song being rereleased in compilation albums) were removed, for a total of N(songs) =
186 1810210 (Japanese = 646440; Chinese = 360101; English = 803669). To ensure class balances,
187 we randomly downsampled the Japanese and English samples to match the Chinese sample,
188 resulting in a final N(songs) = 1080303.
189

190 **Data handling and analysis**
191   Except for 'key' and 'time signature', all Spotify features were inputted as features in the
192 classification models. These were: 'danceability', 'energy', 'loudness', 'speechiness',
193 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo', 'duration' (ms), and 'mode'. A
194 list of definitions for these features is available in Table 1. These were to classify songs
195 according to their cultural membership (Chinese, English, or Japanese), as the outcome variable.
196 The data was split into a training and testing set along a 3:1 ratio. Parameters for the GBDT
197 model and weights for the MLP model were tuned through 5-fold cross validation on the training
198 set. We also examined RFI scores for each model. For GBDT, this was a measure of the
199 proportion that a feature was selected for stratification in each iterative tree, and for MLP, this

200 was based on PFI, which measures the resultant error of a model when each feature is iteratively
201 shuffled – the greater the error, the larger the influence a feature exerts on the outcome variable
202 (Fisher, Rudin, & Dominici, 2018; Molnar, 2019). We then simplified the classification problem
203 by splitting it into 3 separate binary classifications: Japanese-Chinese, Japanese-English, and
204 Chinese-English. GBDTs and MLP models were conducted for these three comparisons, and in
205 addition to RFI measures, we visualised the effect of each variable using PDPs. These show the
206 averaged marginal effect of a feature on the outcome variable in a machine learning model, and
207 is useful to glean an understanding of the nature of the relationship between these variables.
208 PDPs were conducted through the 'pdp' package (Greenwell, 2017), and PFIs were conducted
209 through the 'iml' package (Molnar, Bischl, & Casalicchi, 2019). Machine learning was
210 conducted through the 'gbm' package Greenwell et al., 2019) for GBDTs, and the 'nnet' package
211 (Venables & Ripley, 2002) for MLPs, via the 'caret' wrapper (Kuhn, 2019) in R (R Core Team,
212 2019). All R scripts used for data mining and analysis are available in our OSF repository (*Note*
213 *to reviewers: anonymous review link:*
214 *https://osf.io/d3cky/?view_only=cc89c024cedb4bfd8401544032e505a1*).

215

216 **Results**
217 **Descriptives**
218 Table 2 reports the descriptive medians, lower/upper quantiles, and missing data for each feature
219 per culture. The full list of artists, genres, and songs are available in our OSF repository (*Note to*
220 *reviewers: anonymous review link:*
221 *https://osf.io/d3cky/?view_only=cc89c024cedb4bfd8401544032e505a1*). Additionally, we note
222 that while our database of songs spans as early as the 1950s, most of the songs in our database
223 were from the mid-2000s to 2020 (see Figure 1).

224

225 **Multiclass classification (Chinese-Japanese-English)**
226   For the GBDT model, the parameter tuning resulted in N(trees) = 150, interaction depth =
227 3, alongside default parameters of shrinkage = 0.1, and number of minimum observations per
228 node = 10. The GBDT achieved a classification accuracy of 0.682, 95%CI (0.680, 0.683),
229 significantly above the no information rate (NIR) of 0.333, $p < .0001$. Aside from the input and
230 output layers, we used a MLP model consisting of 1 hidden layer with 5 nodes. The MLP model
231 achieved a slightly lower accuracy score of 0.660, 95%CI (0.659,0.662), but was still
232 significantly above the NIR of 0.333, $p < .0001$. RFIs for both models are reported in Table 3.

233

234 **Binary classifications**
235   We first unpack the Chinese-Japanese model: For the GBDT model, the parameter tuning
236 resulted in N(trees) = 150, interaction depth = 3, and no changes were made to the other default
237 parameters (as above). The GBDT achieved a classification accuracy of 0.784, 95%CI (0.783,
238 0.786), AUC = 0.865, significantly above the no information rate (NIR) of 0.500, $p < .0001$. The
239 MLP model achieved a comparable accuracy score of 0.766, 95%CI (0.764,0.768), AUC =

240   0.844, significantly above the NIR of 0.500, $p < .0001$. Next, the Chinese-English model: For the
241   GBDT model, the parameter tuning resulted in N(trees) = 150, interaction depth = 3, and no
242   changes were made to the other default parameters. The GBDT achieved a classification
243   accuracy of 0.807, 95%CI (0.805, 0.809), AUC = 0.885, significantly above the no information
244   rate (NIR) of 0.500, $p < .0001$. The MLP model achieved a comparable accuracy score of 0.803,
245   95%CI (0.801,0.805), AUC = 0.880, significantly above the NIR of 0.500, $p < .0001$. Finally, the
246   Japanese-English model: For the GBDT model, the parameter tuning resulted in N(trees) = 150,
247   interaction depth = 3, and no changes were made to the other default parameters. The GBDT
248   achieved a classification accuracy of 0.713, 95%CI (0.711, 0.715), AUC = 0.797, significantly
249   above the no information rate (NIR) of 0.500, $p < .0001$. The MLP model achieved a comparable
250   accuracy score of 0.709, 95%CI (0.707,0.711), AUC = 0.791, significantly above the NIR of
251   0.500, $p < .0001$. All RFIs and PFIs are reported in Table 4. Additionally, the 2 most important
252   features are visualised by PDPs in Figure 2. A visual inspection of the PDPs suggests that
253   English music is higher than both Japanese and Chinese music in speechiness, Chinese music is
254   higher than both Japanese and English music in acousticness, and Japanese music is higher than
255   English and Chinese music in energy. In comparing Japanese and Chinese music, we note that
256   acousticness and energy were also present, but were identified only in the GBDT model. In
257   contrast, the MLP model identified loudness and instrumentalness as higher in Japanese music
258   than Chinese music.
259   Overall, this suggests that, unlike English-Japanese or English-Chinese comparisons
260   which were markedly different on a few main features, the differences between Chinese and
261   Japanese music were spread widely across the various features. Consequently, despite relying on
262   different 'important variables', both the MLP and GBDT managed to achieve a comparably high
263   classification accuracy, with the GBDT outperforming the MLP for all classification tasks
264   (results from DeLong's tests are available on our OSF repository).
265
266   **Additional Analyses**
267   We also visualised the changes in features over time for speechiness, acousticness,
268   energy, instrumentalness, and loudness, from 2000 to 2020. Feature information for songs before
269   2000 were excluded due to the markedly smaller sample. Other than instrumentalness, which
270   showed a notable decrease over time in Japanese songs, the remaining four features showed
271   stability over time. This suggests that the differences in preference highlighted by the RFIs, PFIs
272   and PDPs could indicate long term cultural preferences for music.
273
274   **Follow-up Study**
275   We obtained a second round of data (approximately one year later) from Top-50 lists for
276   Japan, Taiwan and the USA. Focusing on the identified features of speechiness, acousticness,
277   energy, instrumentalness, and loudness from the previous study, Kruskal-Wallis tests revealed a
278   significant effect of energy on speechiness ($\chi^2(2) = 30.5$, $p < .001$), acousticness ($\chi^2(2) = 24.5$, p
279    $< .001$), energy ($\chi^2(2) = 21.0$, $p < .001$), and loudness ($\chi^2(2) = 33.5$, $p < .001$), but no significant

280   effect was observed for instrumentalness ($\chi^2(2) = 1.4$, $p = .49$). For speechiness, post-hoc Dwass-
281   Steel-Critchlow-Fligner pairwise comparisons revealed that USA was significantly higher than
282   Taiwan (W = 7.34, $p < .001$) and Japan (W = 5.91, $p < .001$), but no significant difference was
283   observed between Taiwan and Japan (W = -1.48, $p = .55$). For acousticness, Taiwan was
284   significantly higher than Japan (W = 5.97, $p < .001$) and the USA (W = 6.11, $p < .001$), but no
285   significant difference was observed between the USA and Japan (W = 0.53, $p = .93$). For energy,
286   Japan was significantly higher than Taiwan (W = 6.01, $p < .001$), and the USA (W = 4.35, $p =$
287   .006), but no significant difference was observed between Taiwan and the USA (W = 2.88, $p =$
288   .103). For loudness, Japan was significantly higher than Taiwan (W = 7.44, $p < .001$) and the
289   USA (W = 6.36, $p < .001$), but no significant difference was observed between Taiwan and the
290   USA (W = 2.20, $p = .27$). Finally, for instrumentalness, no significant difference was observed
291   between Japan and Taiwan (W = 0.61, $p = .90$), Japan and the USA (W = 1.67, $p = .48$), or
292   Taiwan and the USA (W = 1.01, $p = .75$).
293   In short, with the exception of instrumentalness, Top-50 playlists obtained one year later
294   nevertheless demonstrate strong consistency with the earlier results. American Top-50 songs are
295   higher than both Japanese and Taiwanese Top-50 songs in speechiness, Taiwanese Top-50 songs
296   are higher than both Japanese and American Top-50 songs in acousticness, and Japanese Top-50
297   songs are higher than American and Taiwanese music in energy and loudness. However,
298   instrumentalness, that was originally identified as a variable of importance for the MLP model,
299   did not consistently differ between cultures. Indeed, Figure 3 shows that instrumentalness in
300   Chinese music is inconsistent, with strong fluctuations depending on year. More research is
301   needed to determine if instrumentalness is indeed a preferred feature in Taiwanese markets or
302   merely a passing trend.
303
304   **Discussion**
305   Across the multiclass and subsequent binary classification tasks, both the GBDT and
306   MLP models were able to consistently classify songs by cultural market with moderately high
307   accuracy, and the GBDT was often marginally better than the MLP for this purpose. This
308   suggests that the patterns of difference between cultural markets were robust enough to be
309   detected by two different algorithms. A comparison of accuracy scores suggested that the
310   difference between Chinese and Japanese music afforded higher accuracy to the models, than
311   Japanese and English differences. While this could be for several reasons, we speculate a
312   possibility in that music preferences between Japanese and Chinese cultures differed greater that
313   Japanese and English differences. Such a reasoning would support growing calls for
314   decentralisation and internationalisation of psychological research (Heine & Ruby, 2010;
315   Heinrich, Heine, & Norenzayan, 2010; Cheung, 2012), in showing that Japanese and Chinese
316   speaking cultures, sometimes thought to be homogenous in cross-cultural research, may actually
317   be more different that previously assumed.
318   A visual inspection of the PDPs and feature importance scores provide some indicator of
319   where these differences lie. Apart from instrumentalness, all identified feature importance

320 proved to be different across (geographical) cultures in similar directions in the follow-up study.
321 This implies that cultural music preferences are reflected both in the music produced by a culture
322 for their respective industry or market, as well as the overall music preferences by
323 geographically bound members of that culture. While this paper does not empirically explore the
324 underlying cultural mechanisms that may account for these differences, this is nevertheless a
325 starting point for future research to continue from, and we speculate on some interpretations for
326 these results. Western music similarly differed from Chinese and Japanese music through higher
327 speechiness. One explanation could be prosodic bias, in that normal spoken Mandarin Chinese
328 inherently contains more pitch movements than English (Hirst, 2013), and consequently, what
329 may be perceived as 'speech-like' by Chinese listeners may not correspond to high speechiness
330 scores. However, this can also be explained through previous research on emotion-arousal
331 preferences in the Western and East-Asian contexts. The high speechiness score in English music
332 could indicate larger preferences for hip-hop and rap music. Rap-music has seen a dramatic
333 increase in popularity in Western markets from the 1980s (Mauch et al., 2015), and has been
334 shown to express and embody high-arousal emotions like anger (e.g., Hakvoort, 2015), and its
335 relative popularity in Anglo-American cultures could be representative on cultural preferences
336 towards these high arousal emotions described earlier (Tsai, 2007), compared to Japanese and
337 Chinese cultures.
338   One feature that differentiated Chinese from English and Japanese music was high
339 acousticness. This points to lower use of electronic instruments in the production process, and
340 may suggest a preference for more organic, natural sounds in Chinese music. Energy appeared to
341 be more important in Japanese than English or Chinese music. We posit that energy preferences
342 in Japan (defined by Spotify as a combination of loudness, complexity, timbre, dynamic range,
343 and noise) could be due to remnants of traditional music aesthetics, that overlap considerably
344 with energy definitions (e.g., beauty in noise/simplicity in complexity: *sawari*, *wabi-sabi*, see
345 Deva, 1999; Anderson, 2014; Okuno, 2015). On the surface, this could be similarly concluded
346 from increased loudness features in Japanese over Chinese music (from the follow-up study), but
347 the U-shaped relationship between energy and Japanese/Chinese music classification seen in
348 Figure 2 suggests a deeper nuance that requires further research.
349   Finally, we consider the strengths and limitations our exploratory approach. Comparing
350 music features offer a greater insight into behavioural and consumption patterns of music
351 preference across cultural spheres. In doing so, we uncover systematic differences between
352 groups that, while being consistent with previous literature, also offer new insight into how
353 cultures differ, that future research can build from in understanding societies. Unfortunately, we
354 were unable to eliminate certain sample biases from our dataset: we assumed our Chinese data to
355 be representative of Chinese music in general, but Spotify is not (as of 2021) active in China
356 despite the inclusion of several mainland Chinese artists in the database. Instead, our findings
357 represented Chinese-speaking listeners in Taiwan and Hong Kong, along with possibly Malaysia
358 and Singapore, who may have differing values and preferences from mainland Chinese listeners,
359 particularly given differences in demographics of users and variation in dialect. Moreover, using

360   the Spotify API limited our selection of music features to those available in the API. Future
361   studies could examine music features through publicly available software (e.g., MIRtoolbox;
362   Lartillot, Toiviainen, & Eerola, 2008) that have both greater amounts of features and more
363   transparent documentation.
364   On the other hand, our strengths include our comparisons of features, as opposed to
365   genre, that allowed for validity in comparing cultures because of universality in the perceptual
366   properties of music (Savage et al., 2015). This enabled us to conclude that any differences in
367   music features would be due to preference for those features. By contrast, comparing preferences
368   by genre differences across cultures could have introduced confounds to the investigation, as
369   genre is not homogenous across cultures (see Bennett, 1999).
370

Conclusions

372   In sum, we demonstrated the variability of music preferences across Chinese (Taiwan),
373   Japanese, and Western (American) cultural markets, and identified the features that best account
374   for these differences. In particular, Chinese music was marked by high acousticness, Anglo-
375   American Western music was marked by high speechiness, and Japanese music appeared to be
376   marked by high energy. While we speculated on some reasons why this would be so, future
377   research is needed to validate these theories to develop a holistic understanding of popular music
378   preferences in Chinese and Japanese cultures. As music is an integral part of human society and
379   culture, understanding the mechanisms by which we prefer different *types* of music may also
380   shed light on the aspects of human society and experience that correspond to these differences.
381   Our paper also demonstrates the potential uses of machine learning and other computer
382   science methods in cross-cultural research. Given the advent of digital and online media, these
383   repositories of cultural products may hold valuable insight into the diversity of humanity.
384   Computer science as a field has utilised these kinds of data to great effect, be it in developing
385   recommendation systems, or in predicting consumer behaviour, and we hope to demonstrate that
386   these same data and methods can also contribute towards research on society and culture. While
387   a common argument has been that machine learning emphasises prediction, whereas social
388   scientific research prefers interpretation, we show that the two goals are not mutually exclusive.
389   As demonstrated through the use of model interpretation techniques like RFIs and PDPs,
390   supplementing commonly-used prediction focused models in computer science with explanation
391   and model interpretation techniques enables big data and machine learning to offer an efficient
392   and viable means for the empirical analysis of sociocultural phenomenon. At the same time,
393   these methods are more objective, and hold less bias than commonly used methods like self-
394   reports. Particularly for cross-cultural research, future directions could also apply this
395   methodology to identify new avenues of cultural differences in other mediums, as an additional
396   analysis tool. Additionally, computer scientists and engineers could also benefit from this
397   knowledge, as such analyses of sociocultural phenomenon could also aid with the fine-tuning of
398   weights in more opaque deep learning models, such as in recommendation systems used by
399   streaming companies when targeting users from different cultures.

400

References

Anderson, C. (2014). *The use of traditional Japanese music as an inspiration for modern saxophone compositions: An interpretive guide to Joji Yuasa's Not I but the wind' and Masakazu Natsuda's West, or Evening Song in Autumn* [doctoral dissertation]. http://hdl.handle.net/2142/49546, 2014.

Askin, A., Mauskapf, M. (2017). What makes popular cultural popular? Product features and optimal differentiation in music. *American Sociological Review*, *82*(5), 910-944. https://doi.org/10.1177/0003122417728662

Bardi. A., Calogero, R. M., & Mullen, B. (2008). A new archival approach to the study of values and value--Behavior relations: Validation of the value lexicon. *Journal of Applied Psychology*, *93*(3), 483–497. https://doi.org/10.1037/0021-9010.93.3.483

Bennett, A. (1999). Hip hop am Main: the localization of rap music and hip hop culture. *Media, Culture & Society, 21*(1),f 77-91, 1999.

Cheung, F. M. (2012). Mainstreaming culture in psychology. *American Psychologist, 67*(8), 721-730. https://doi.org/10.1037/a0029876.

de Almeida, I., & Uchida, Y. (2018). Examining affective valence in Japanese and Brazilian cultural products: An analysis on emotional words in song lyrics and news articles. *Psychologia, 61*(3), 174-184. https://doi.org/10.2117/psysoc.2019-A103.

Deva, D. (1999). Underlying socio-cultural aspects and aesthetic principles that determine musical theory and practice in the musical traditions of China and Japan. *Renaissance Artists and Writers Assocation.*

Dunbar, R. I. M. (2012). On the evolutionary function of song and dance. In N. Bannan (Ed.), *Music, language and human evolution* (pp. 201– 214). Oxford, England: Oxford University Press.

Febirautami, L. R., Surjandari, I., & Laoh, E. (2019). Determining characteristics of popular local songs in Indonesia's music market. In *Proceedings of the 5th International Conference on Information Science and Control Engineering*, Zhengzhou, China.

Fisher, A., Rudin, C., & Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research, 20*(177), 1-81.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine, *The Annals of Statistics, 29*(5), 1189-1232.

Fricke, K. R., Greenberg, D. M., Rentfrow, P. J., & Herzberg, P. Y. (2019). Measuring musical preferences from listening behavior: Data from one million people and 200,000 songs. *Psychology of Music*. https://doi.org/10.1177/0305735619868280

Golder, S., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878-1881. https://doi.org/10.1126/science.1202775

439 Greenwell, B. M. (2017). pdp: An R Package for constructing partial dependence plots. *The R*
440   *Journal, 9*(1), 421-436. https://journal.r-project.org/archive/2017/RJ-2017-
441   016/index.html.
442 Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers, *gbm: Generalized Boosted*
443   *Regression models.* R package version 2.1.5. https://CRAN.R-project.org/package=gbm
444 Groarke, J. M., & Hogan, M. J. (2018). Development and psychometric evaluation of the
445   Adaptive Functions of Music Listening Scale. *Frontiers in Psychology, 9,*
446   516. https://doi.org/10.3389/fpsyg.2018.00516
447 Hakvoort, L. (2015). Rap music therapy in forensic psychiatry: Emphasis on the musical
448   approach to rap. *Music Therapy Perspectives, 33*(2), 184-192.
449   https://doi.org/10.1093/mtp/miv003.
450 Heine, S. J., & Hamamura, T. (2007). In search of East Asian self-enhancement. *Personality and*
451   *Social Psychology Review, 11*, 4-27. https://doi.org/10.1177/1088868306294587.
452 Heine, S., & Ruby, M. B. (2010). Cultural psychology. *Wiley Interdisciplinary Reviews, 1*(2),
453   254-266. https://doi.org/10.1002/wcs.7.
454 Henrich, J., Heine, S., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466,* 29.
455   https://doi.org/10.1038/466029a.
456 Hirst, D. (2013). Melody metrics for prosodic typology: comparing English, French and Chinese.
457   In *Interspeech. Annual Conference of the International Speech Communication*
458   *Association.*, Lyon, France.
459 Howlin, C., & Rooney, B. (2020). Patients choose music with high energy, danceability, and
460   lyrics in analgesic music listening interventions. *Psychology of Music.*
461   https://doi.org/10.1177/0305735620907155.
462 Joshanloo M., & Weijers, D. (2014). Aversion to happiness across cultures: A review of where
463   and why people are averse to happiness. *Journal of Happiness Studies, 15*, 717-735.
464   https://doi.org/10.1007/s10902-013-9489-9.
465 Kuhn, M. (2019). *caret: Classification and regression training.* R package version 6.0-82.
466   https://CRAN.R-project.org/package=caret
467 Lamoreaux, M., & Morling, B. (2012). Outside the head and outside individualism-collectivism:
468   Further meta-analyses of cultural products. Journal of Cross-Cultural Psychology, 43(2),
469   299-327.
470 Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information
471   retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), Data
472   analysis, machine learning, and applications: Studies in classification, data analysis, and
473   knowledge organization (pp. 261–268). Berlin, Germany: Springer.
474 Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion,
475   and motivation. *Psychological Review, 98*(2), 224-253.
476 Markus, H. R., & Kitayama, S. (2010). Cultures and selves: A cycle of mutual constitution.
477   *Perspectives on Psychological Science, 5*(4), 420-430.
478   https://doi.org/10.1177/1745691610375557.

479 Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular
480   music: USA 1960-2010. *Royal Society Open Science*, *2*(5), 150081.
481   https://doi.org/10.1098/rsos.150081

482 Molnar, C. (2019). Interpretable Machine Learning. A Guide for Making Black Box Models
483   Explainable. https://christophm.github.io/interpretable-ml-book/.

484 Molnar, C., Bischl, B., & Casalicchio, G. (2019). iml: An R package for interpretable machine
485   learning. *Journal of Open Statistical Software, 3*(26), 786. doi: 10.21105/joss.00786.

486 Morling, B. & Lamoreaux, M. (2008). Measuring culture outside the head: A meta-analysis of
487   individualism-collectivism in cultural products. *Personality and Social Psychology*
488   *Review, 12*, 199-221. https://doi.org/10.1177/1088868308318260.

489 Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., &
490   Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD)
491   psychology: Measuring and mapping scales of cultural and psychological distance.
492   *Psychological Science, 31*(6), 678-701. https://doi.org/10.1177/0956797620916782.

493 Okuno, T. (2015). Roots of Cool Japan: From the Japanese raditional Edo culture to Anime and
494   Manga," *Kwansei Gakuin University Social Sciences Review, 19*, 1-7.

495 Park, M., Thom, J., Mennicken, S., Cramer, H., & Macy, M. (2019). Global music streaming
496   data reveal diurnal and seasonal patterns of affective preference. *Nature Human*
497   *Behaviour, 3,* 230-236. https://doi.org/10.1038/s41562-018-0508-z.

498 Pérez-Verdejo, J. M., Piña-García, C. A., Ojeda, M. M., Rivera-Lara, A., & Méndez-Morales L.
499   (2020). The rhythm of Mexico: an exploratory data analysis of Spotify's top 50. *Journal*
500   *of Computation Social Science.* https://doi.org/10.1007/s42001-020-00070-z.

501 R Core Team, (2019). *R: A language and environment for statistical computing.* R Foundation
502   for Statistical Computing: Vienna, Austria.

503 Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the
504   structures and functions of human music. *Proceedings of the National Academy of*
505   *Sciences, 112* (29), 8987-8992. https://doi.org/10.1073/pnas.1414495112.

506 Schwind, A., Haberzettl, L., Wamser, F., & Hoßfeld, T. (2019). QoE analysis of Spotify audio
507   streaming and app browsing. In *Internet-QoE'19: Proceedings of the 4th Internet-QoE*
508   *Workshop on QoE-based Analysis and Management of Data Communication Networks*.

509 Smith, P. B., Fischer, R., Vignoles, V. L., & Bond, M. H. (2013). Understanding social
510   psychology across cultures: Engaging with others in a changing world. Sage.

511 Thompson, C., Parry, J., & Wolff, T. (2019). *spotifyr: R wrapper for the 'Spotify' web API,* R
512   package version 2.1.1. http://github.com/charlie86/spotifyr.

513 Triandis, H., Marin, G., Lisansky, J., & Betancourt, H. (1984). Simpatía as a cultural script of
514   Hispanics. *Journal of Personality and Social Psychology, 47*, 1363-1375.

515 Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on*
516   *Psychological Science, 2*(3), 242-259. https://doi.org/10.1111/j.1745-6916.2007.00043.x.

517 Uchida, Y., & Kitayama, S. (2009). Happiness and unhappiness in East and West: Themes and
518   variations. *Emotion, 9*(4), 441-456. https://doi.org/10.1037/a0015634.

519   Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S, New York:
520   Springer.

A list of song-level audio features obtained from Spotify for our analyses

Audio features refer to the music features as listed on the Spotify API, followed by a brief description of each feature. More information on the features are available at:

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

Table 1:

A list of song-level audio features obtained from Spotify for our analyses

Audio feature	Description
Duration	The duration of the music in milliseconds (ms).
Mode	If the melody of a track is in a major or minor key.
Acousticness	A confidence measure on whether a song is acoustic.
Danceability	The suitability of a song for dancing. This is based on several musical features, such as tempo, rhythmic stability, regularity, and beat strength.
Energy	A measure of the intensity and activity of a song as perceptually lound, fast, or noisy. This is based on several musical and spectral features, such as dynamic range, loudness, timbre, onset rate, and entropy.
Instrumentalness	A confidence measure of whether a song contains no vocals.
Liveness	A confidence measure on the presence of audiences in the recording.
Loudness	The overall intensity of the song in decibels (dBFS).
Speechiness	A confidence measure on the presence of spoken words in a song.
Valence	An estimate of whether a song conveys positive or negative affect.
Tempo	The estimated main tempo of a song.

1 (https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/)

Medians (L/U quantiles) and missing data for musical features (excluding mode), and release year

Table 2:
Medians (L/U quantiles) and missing data for musical features (excluding mode), and release year.

Feature	Chinese		Japanese		Western	
	Median (L/U)	Missing	Median (L/U)	Missing	Median (L/U)	Missing
Danceability	0.56 (0.46/0.66)	3	0.56 (0.45/0.67)	4	0.66 (0.55/0.76)	1
Energy	0.46 (0.34/0.63)	3	0.76 (0.51/0.90)	4	0.69 (0.54/0.82)	0
Loudness	-8.9 (-11.1/-6.9)	3	-6.2 (-9.3/-4.3)	4	-6.8 (-8.9/-5.2)	0
Speechiness	0.04 (0.03/0.05)	3	0.05 (0.04/0.09)	4	0.08 (0.04/0.24)	0
Acousticness	0.54 (0.21/0.76)	3	0.11 (0.01/0.49)	4	0.09 (0.02/0.29)	1
Instrumentalness	1.4E-6 (0.00/1.1E-4)	3	1.0E-4 (0.00/0.36)	4	0.00 (0.00/0.0006)	1
Liveness	0.13 (0.10/0.21)	3	0.14 (0.10/0.29)	4	0.14 (0.10/0.29)	1
Valence	0.37 (0.24/0.57)	3	0.52 (0.31, 0.71)	4	0.51 (0.33/0.69)	1
Tempo	122.7 (100.0/138.1)	3	123.7 (99.0/144.0)	4	119.7 (96.6.136.2)	0
Duration (ms)	240213 (205586/272586)	3	236840 (187266/281573)	4	215640 (184727/250693)	0
Release year	2011 (2005/2015)	148	2015 (2010/2018)	100	2016 (2012/2018)	0

Comparison of feature importance between the GBDT and MLP multiclass models

Table 3:
Comparison of feature importance measures.

Feature	RFI (GBDT)	PFI (MLP)
Speechiness	24.3	1.26
Loudness	7.4	1.15
Instrumentalness	15.6	1.19
Acousticness	16.5	1.14
Energy	17.5	1.12
Mode name	0.1	1.00
Duration	8.4	1.12
Danceability	5.9	1.06
Valence	1.8	1.03
Tempo	2.2	1.01
Liveness	0.2	1.00

Comparison of feature importance measures for respective binomial classification models

Note: While scales between RFI and PFI are not equivalent, both measure model-specific feature importance relative to other features: the higher the score, the larger the importance within the model. Features with highest importance are in bold. PFIs were reported with 2 decimal places, but we used 3 decimal places for PFIs denoted by '^'. This was to identify the 2nd most important feature for the PDP.

Table 4:
Comparison of feature importance measures for respective binomial classification models.
Note: While scales between RFI and PFI are not equivalent, both measure model-specific feature
importance relative to other features: the higher the score, the larger the importance within the model.
Features with highest importance are in bold. PFIs were reported with 2 decimal places, but we used 3
decimal places for PFIs denoted by '^'. This was to identify the 2nd most important feature for the PDP.

Feature	Chinese-Japanese		Chinese-English		Japanese-English	
	RFI (GBDT)	PFI (MLP)	RFI (GBDT)	PFI (MLP)	RFI (GBDT)	PFI (MLP)
Speechiness	4.8	1.05	**46.6**	**1.67**	**39.6**	**1.22**
Loudness	14.1	**1.46**	3.0	1.19	4.7	1.04
Instrumentalness	26.0	1.22	2.8	1.08	8.5	1.105^
Acousticness	31.1	1.16	34.1	1.35	2.4	1.04
Energy	**32.9**	1.07	4.7	1.09	16.0	1.114^
Mode name	0.0	1.00	0.0	1.01	0.2	1.00
Duration	9.3	1.08	4.3	1.07	12.9	1.02
Danceability	1.2	1.01	1.6	1.06	10.4	1.110^
Valence	0.7	1.06	1.2	1.02	3.3	1.003
Tempo	1.4	1.01	1.5	1.02	1.9	1.00
Liveness	0.0	1.00	0.2	1.00	0.1	1.00

Figure 1

Number of songs (in our data) by year for Japanese (JP), English (US), and Chinese (ZH) medium songs
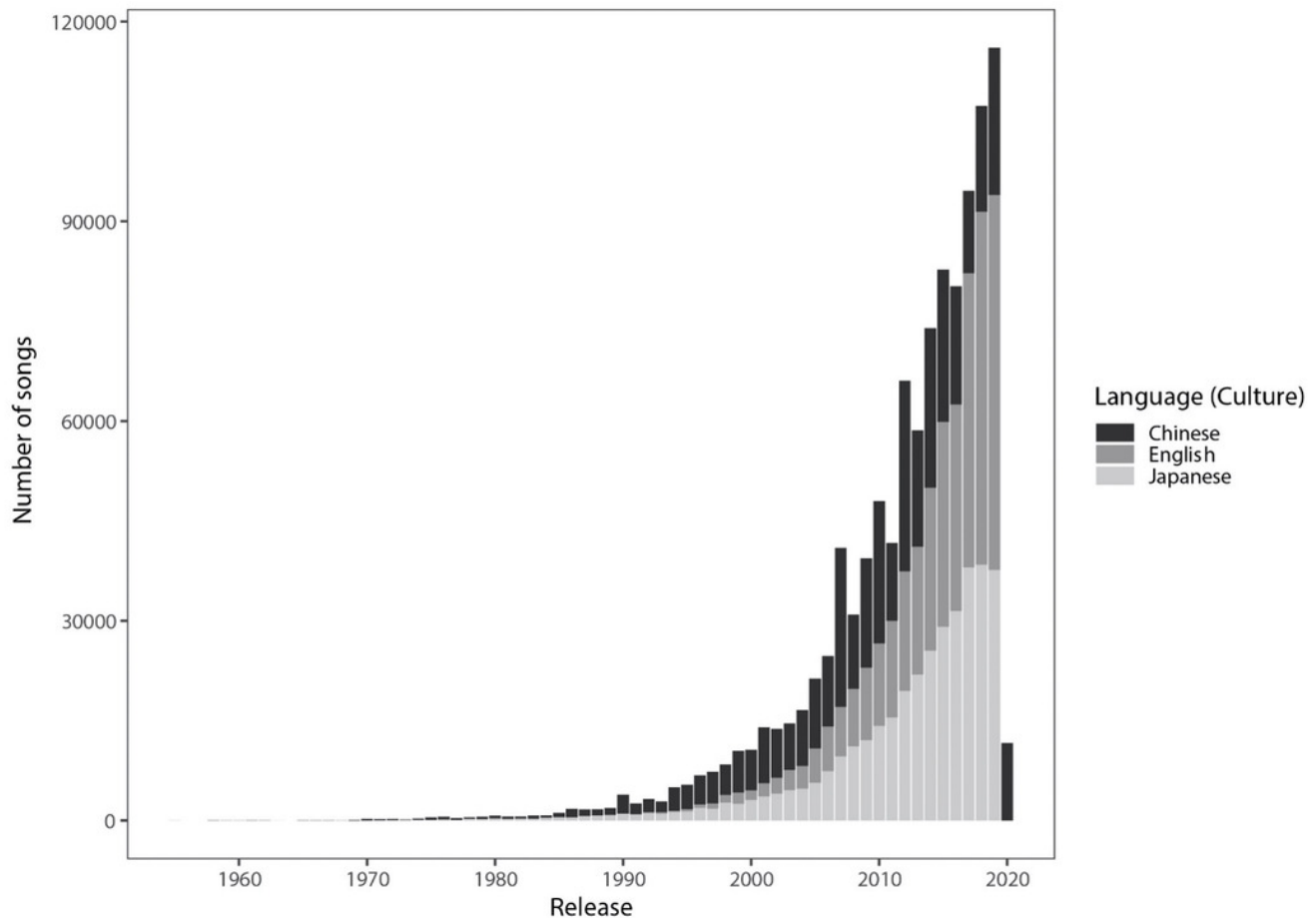
Figure 1: Number of songs (in our data) by year for Japanese (JP), English (US), and Chinese (ZH) medium songs.

Figure 2

PDPs of top 2 most important features in each model on the probability of classification

The positive class is indicated at the top of the Y-axis. For example, in the Japanese-English GBDT model (top right) for the speechiness feature, the decreasing trend indicates that the higher the speechiness score, the lower the probability of classification (of a song) as being Japanese (i.e., higher probability of being English), in a fairly linear fashion.
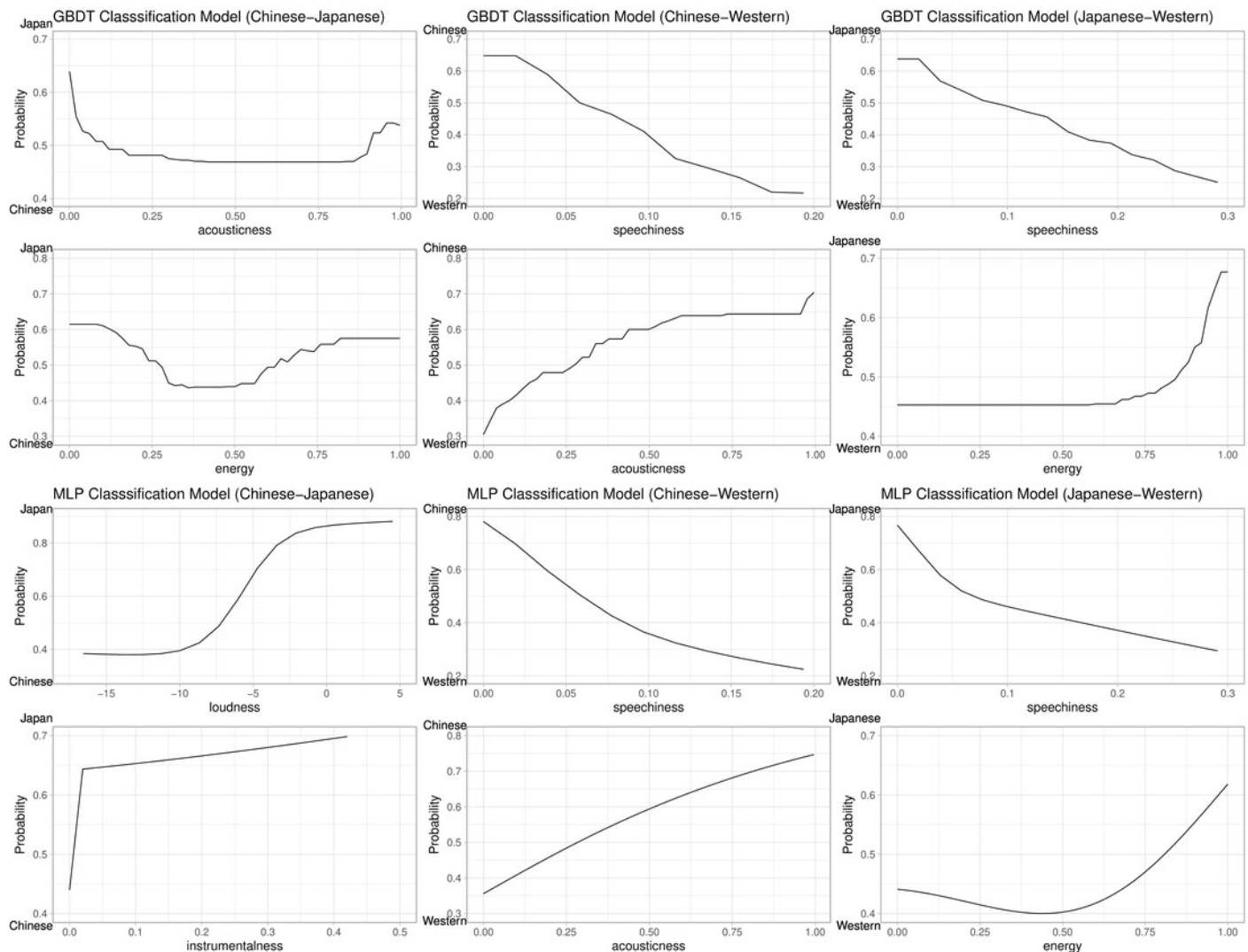
Figure 3

Variation and stability of features (medians) across cultures from 2000 to 2020

Note that scales differ bewteen features.