

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Do Melody and Rhythm Coevolve?

Permalink

<https://escholarship.org/uc/item/5nd83609>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Lee, Harin

Polak, Rainer

Anglada-Tort, Manuel

et al.

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Do Melody and Rhythm Coevolve?

Harin Lee¹, Rainer Polak², Manuel Anglada-Tort³, Marc Schönwiesner⁴, Minsu Park⁵ & Nori Jacoby⁶

¹University of Cambridge, UK

²RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway

³Department of Psychology, Goldsmiths College, University of London, UK

⁴Department of Life Sciences, Leipzig University, Germany

⁵Division of Social Science, New York University Abu Dhabi, UAE

⁶Department of Psychology, Cornell University, USA

Abstract

Music comprises two core structural components, melody and rhythm, that vary widely across cultures. Whether these components coevolve in a coupled way or follow independent trajectories remains unclear. We introduce a novel computational pipeline to extract vocal melodic pitch-interval and percussive inter-onset timing distributions from 27,628 popular songs across 59 countries, enabling large-scale cross-cultural comparison that bypasses traditional music annotations. Musical similarities between countries aligned with geographic and linguistic relationships, validating our approach. Substantial variation emerged in both melodic and rhythmic structures across countries, yet the diversity of the two components was not significantly correlated, challenging assumptions of coupled evolution. Only rhythmic diversity was significantly associated with ethnic and linguistic heterogeneity, while melodic diversity showed no such association. These findings suggest that melody and rhythm constitute partially independent systems shaped by distinct cultural and evolutionary pressures, rather than components of a single monolithic musical style.

Keywords: music cognition; cultural evolution; cross-cultural; large-scale

Introduction

Songs are present in all human cultures and serve important social and individual functions (Mehr et al., 2019; Savage et al., 2015). Across cultures, songs consistently comprise two core structural components: melody, which organizes pitch, and rhythm, which structures temporal durations (Mehr, 2025). Both components show cross-cultural regularities consistent with shared cognitive constraints, yet they also vary widely both within and between societies, reflecting culturally learned traditions that change over time (Brown et al., 2025; Jacoby et al., 2024; McBride, Phillips, et al., 2025; Ozaki et al., 2024; Passmore et al., 2024). A fundamental open question is whether these cultural changes are coupled: when a musical tradition shifts melodically, does it tend to shift rhythmically as well, or can melody and rhythm follow partially independent evolutionary trajectories?

There are reasons to expect dissociation between these two components. Comparative and evolutionary accounts suggest that rhythm-related capacities may have older phylogenetic roots than pitch-related capacities (Kotz et al., 2018; Patel et al., 2006). Neuro-cognitive work points to at least partially distinct mechanisms for processing temporal versus pitch structure (Bianco et al., 2025; Hádén et al., 2024). If

melody and rhythm are supported by distinct cognitive and neural systems, cultural transmission could act on them at different rates or along different social boundaries. For instance, one might observe high rhythmic similarity within groups but greater melodic diversification, or vice versa. This leads to a concrete testable prediction. Melody and rhythm should not necessarily co-vary in their degree of cross-cultural diversity, each potentially responding to distinct sociodemographic factors

Existing research, however, often treats music as a monolithic entity or focuses on a single component, making it difficult to test whether melody and rhythm change in tandem or independently (Jacoby et al., 2021; Yurdum et al., 2023; Zivic et al., 2013). A recent exception is McBride, Kim, et al., 2025, who studied co-variation of pitch and duration features using transcribed corpora and found correlated entropy between melodic and rhythmic dimensions arising from the same vocal melody. Our work extends this line of inquiry using audio-derived features and a different operationalization of rhythm—percussive timing rather than vocal rhythm.

Scaling this kind of cross-cultural comparison remains challenging. Traditional approaches rely on manual transcription and ethnographic annotation (Lomax, 1978; Mehr et al., 2019; Savage et al., 2015) are often constrained by limited sample size, and it can introduce bias when analysis is filtered through Western-derived musical concepts (Clayton et al., 2003; Daikoku et al., 2022; Rice, 2017). Standard analytical frameworks built on equal temperament, tonal centers, or beat isochrony can misrepresent non-Western traditions (Hannon et al., 2012; Jacoby et al., 2019; Polak et al., 2018). Recent machine learning approaches have improved scalability (Castellanos et al., 2022; Mauch et al., 2015), but often perpetuate similar biases due to training predominantly on Western music datasets (Lee, Çelen, et al., 2025).

Here, we introduce a computational pipeline that extracts melodic and rhythmic distributional profiles directly from raw audio at scale (Figure 1). At its core, the pipeline uses deep learning source separation (Rouard et al., 2023) to decompose each recording into a vocal and a percussive component, then characterizes each through distributions of pitch intervals (melody) and inter-onset timing ratios (rhythm). Crucially, this *within-song* matched measurement means that both components derive from the same recording rather than from different datasets, annotation conventions, or analytical frame-

works. By working with distributional features rather than higher-level constructs, the pipeline minimizes the number of analytical commitments such as key, scale, or meter.

We apply this pipeline to a novel dataset of 27,628 country-exclusive popular songs from 59 countries, sampled from YouTube’s weekly music charts (Figure 1A). By focusing on songs that chart locally rather than internationally, the dataset captures region-specific musical preferences. We then test whether melody and rhythm exhibit coupled versus independent patterns of cultural variation by examining whether the same sociodemographic boundaries constrain both components. We here focus on one specific quantity, distributional diversity, as a measure of cross-cultural variation, while noting that other aspects of melodic and rhythmic structure (e.g., entropy, central tendency, sequential dependencies) remain to be investigated.

Methods

Data collection

Our dataset consists of weekly Top 100 song charts from YouTube Music (<https://charts.youtube.com>) collected between September 2017 and February 2022. The initial dataset contained 1,067,100 entries across the all available 61 countries, comprising 64,945 unique songs. From this full set, we identified songs exclusive to individual countries to capture musical characteristics specific to that country (and not the international mega hits that are popular everywhere). After excluding songs that appeared in multiple countries (26.1%), we retained 47,995 country-exclusive or ‘locally niche’ songs. To balance computational feasibility with comprehensive coverage, we implemented a 1,500-song limit per country. For countries exceeding this threshold, we randomly sampled 1,500 songs, while for others, we took all existing songs. We excluded countries with fewer than 50 songs to achieve enough statistical power. This removed El Salvador ($n=18$) and Costa Rica ($n=24$), resulting in a final set of 59 countries and 27,628 unique songs.

Feature extraction pipeline

Source separation We used the Demucs music source separation tool (version 4) to extract vocal singing and drum parts from the raw audio (Rouard et al., 2023). Demucs is a state-of-the-art model capable of separating vocals, drums, bass, and other accompaniment parts from the song. For processing efficiency, we used the ‘mdx_q’ model variant and segmented each song’s raw audio into 1-minute clips taken from random points. Using 32GB GPU and 64 CPU cores in parallel, we processed all 27,628 audio files in approximately 3 days.

Melodic features We tracked fundamental frequency (f_0) in the source-separated vocal tracks to capture pitch contours and infer the distribution of melodic intervals. f_0 was extracted from the audio recordings using the ‘pyin’ pitch tracking algorithm implemented in the ‘librosa’ Python

package (Mauch & Dixon, 2014). This process yields time-stamped frequency values representing the vocal melody.

Next, we computed melodic intervals from the pitch contours by calculating semitone differences between proximal pitch values. Frequencies were first converted to MIDI pitch values using $m(t) = 12 \log_2(f(t)/440) + 69$, where $f(t)$ is the fundamental frequency at time t . Then we used a sliding time window, where intervals between pitches were analyzed at varying timescales, ranging from 100ms to 2,000ms, with a step size of 10ms. For a given time lag Δ , the melodic interval at time t was computed as:

$$I(t, \Delta) = m(t) - m(t - \Delta)$$

For instance, with a time window of $\Delta = 100$ ms, we computed the difference in semitones between the pitch at each time point and the pitch 100ms earlier; with $\Delta = 2000$ ms, we captured slower melodic movements spanning up to two seconds. The minimum and maximum limits of these time windows were chosen based on literature to encompass both rapid ornamental figures and slower melodic gestures typical of vocal music (London, 2002).

To create a continuous representation, we pooled interval values across all time lags (from 100 ms to 2,000 ms in 10 ms steps) into a single set and then applied kernel density estimation (bandwidth = 0.10 semitones), producing one summary distribution per song. This pooling captures melodic relationships at multiple timescales simultaneously. We acknowledge that this approach yields an approximation of traditional pitch-interval distributions, and that formal evaluation against ground-truth transcribed datasets is an important direction for future work.

Rhythmic features We tracked the timing patterns of percussive onsets in the source-separated drum tracks and computed inter-onset ratios that capture rhythmic structures independent of tempo. The source separation algorithm isolates the rhythmic component into a single track, combining all percussive elements (e.g., hi-hat and bass drum) into a unified stream with audible and typically separated onsets.

First, we extracted note onset times from the unified audio recordings using an onset detection algorithm implemented in the librosa package. This process yields time-stamped markers indicating the start of each musical event. The method relies on sets of heuristics that track strong peaks on the spectral novelty function, rather than being trained on specific music corpora (which could introduce certain cultural biases).

Second, we computed inter-onset ratios from consecutive onset triplets following the approach of Roeske et al., 2020. For each set of three consecutive onsets at times t_1 , t_2 , and t_3 , we calculated the two inter-onset intervals $IOI_1 = t_2 - t_1$ and $IOI_2 = t_3 - t_2$, and then computed the normalized ratio:

$$r = \frac{IOI_1}{IOI_1 + IOI_2}$$

This ratio represents the proportion of the first interval within a two-interval rhythmic cycle, yielding values between 0 and

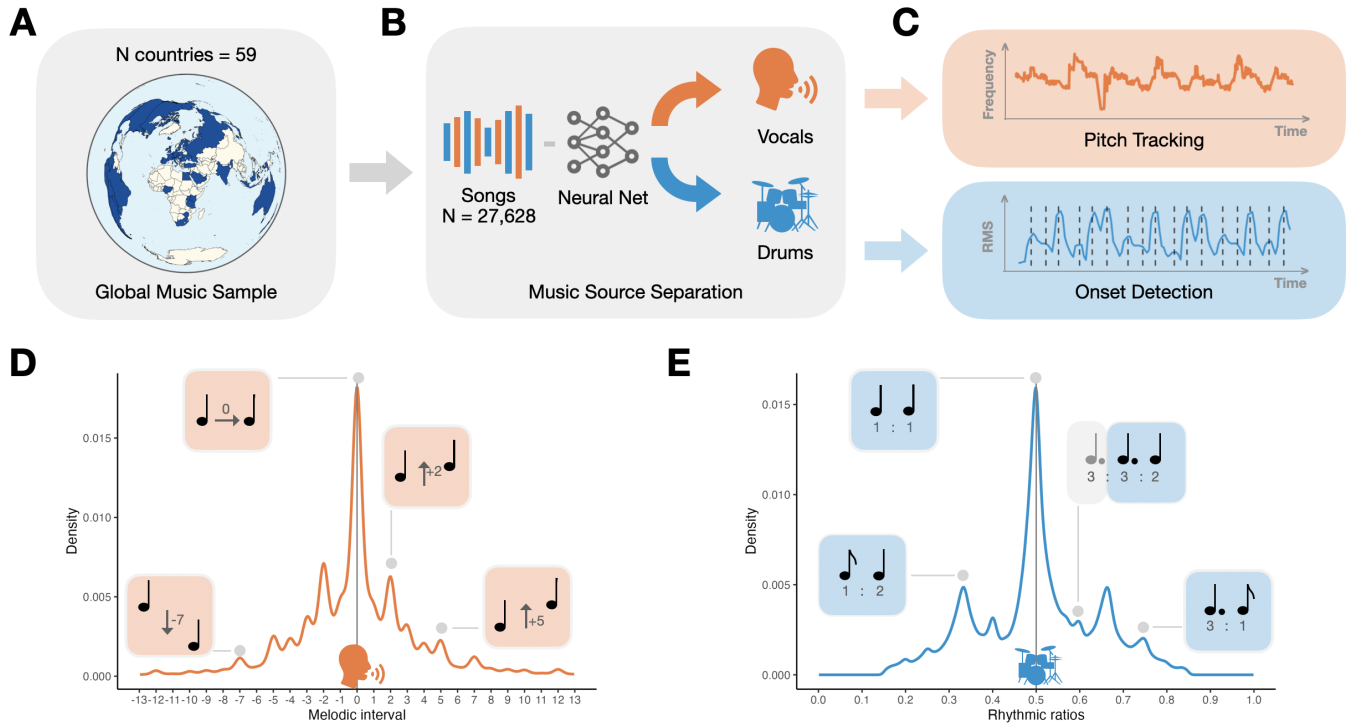


Figure 1: Computational pipeline for extracting melodic and rhythmic characteristics from the raw audio. (A) Global coverage of the dataset comprising 27,628 popular songs sampled from YouTube music charts across 59 countries (‘Data collection’ in Methods). (B) Deep learning-based source separation technique decomposes each song into vocal and drums, enabling independent analysis of melody and rhythm (‘Source separation’ in Methods). (C) Extracted components undergo specialized processing (‘Feature extraction pipeline’ in Methods): fundamental frequency tracking for vocals (orange) and onset detection for drums (blue). (D) Global distribution of melodic intervals shows known characteristic peaks such as the unison (0), major second (± 2), and perfect fifth (± 7), while a characteristic dip at tritone (± 6). (E) Distribution of rhythmic inter-onset ratios reveals preferences for simple integer relationships: isochrony (1 : 1), duple (1 : 2 / 2 : 1), triple (1 : 3 / 3 : 1), and 3 : 2 / 2 : 3 that are elements to forms more complex patterns such as *tresillo* (3 : 3 : 2). Musical notation illustrates some of these melodic intervals and rhythmic ratios at the distribution peaks.

1. Ratios below 0.15 or above 0.85 were excluded to filter out extremely uneven subdivisions likely arising from onset detection errors or grace notes. This ratio-based measure is tempo-invariant, enabling direct comparison across pieces with different tempi.

Finally, just as with melodic distribution, we applied kernel density estimation on the inter-onset ratios (bandwidth = 0.005) and normalized the density profile to sum to one.

Socio-demographic factors

We compared melodic and rhythmic diversity within each country with demographic diversity along four dimensions: ethnicity, language, religion, and genetics. Data for ethnic, linguistic, and religious diversity were derived from Fearon’s Fractionalization index (Fearon, 2003). Fearon’s index quantifies a country’s diversity as the probability (ranging from 0 to 1 with 1 indicating maximum fragmentation) that two randomly selected individuals belong to different groups.

Data for genetic diversity were derived from the GeLaTo database (Barbieri et al., 2022), which compiles original data

collected by Pemberton et al., 2012. We used heterozygosity as our measure of genetic diversity, which measures the proportion of individuals in a given population that carry two different alleles at a given locus.

Results

Global melodic and rhythmic patterns provide face validity

Before examining cross-cultural variation, we first validate our computational pipeline by aggregating melodic and rhythmic distributions across all songs and countries to produce global summaries (Figures 1D&E). The melodic interval distribution (Figure 1D) reproduced known peaks at small intervals and the characteristic tritone dip (Anglada-Tort, Harrison, et al., 2023; Brown et al., 2025; Huron, 2006; Tenney, 1988; Vos & Troost, 1989), while the rhythmic distribution (Figure 1E) showed the expected preferences for isochrony and small-integer ratios (Jacoby et al., 2024; Polak et al., 2018; Roeske et al., 2020). These convergent patterns confirm that the

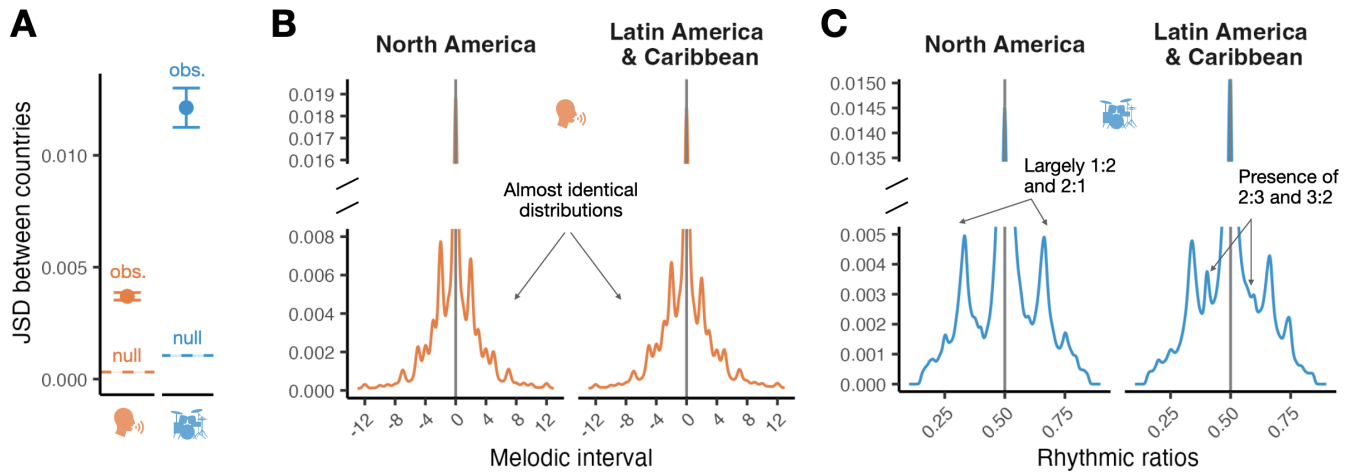


Figure 2: Cross-cultural variation in melodic and rhythmic structure. (A) Observed Jensen-Shannon divergence (JSD) between country-level distributions of melody and rhythm compared against a null model that shuffles country labels across songs while preserving sample sizes. Both domains show significantly greater between-country divergence than expected by chance. Error bars represent 95% CI across bootstrap iterations. (B) Melodic interval distributions for North America and Latin America & Caribbean (shown as example regions) exhibit strikingly similar profiles, whereas (C) rhythmic ratio distributions for the same regions show marked differences. Y-axes are truncated to highlight variation beyond the dominant peaks at unison (melodic interval = 0) and isochrony (rhythmic ratio = 0.5).

pipeline successfully recovers established regularities in both melodic and rhythmic structure, hence providing face validity for the subsequent cross-cultural analyses.

Geolinguistics explains between-country musical similarities

To examine whether melodic and rhythmic patterns demonstrate culturally distinct signatures, we computed Jensen-Shannon divergence (JSD) between all pairs of countries for both melodic intervals and rhythmic ratios, and compared these observed distances to a permutation null model (Figure 2A). The null model shuffles country labels across songs while preserving sample sizes within each country, allowing direct comparison between observed and chance-level between-country distances. Both melody and rhythm showed significantly greater between-country divergence than expected by chance (melody: $z = 10.0$, Cohen's $d = 1.68$, $p < 0.001$; rhythm: $z = 14.0$, $d = 1.28$, $p < 0.001$), confirming that musical features cluster by country origin rather than arising from identical universal distribution.

We further assessed whether the countries that are geographically closer also share similar melodic and rhythmic characteristics. Using geographical regions defined by the World Bank¹, we found that countries within the same geographic region exhibited significantly more similar musical profiles than countries in different regions. Specifically, we compared the JSD values of all same-region country pairs against all different-region country pairs, yielding moderate effect sizes for both melody (Cohen's $d = 0.39$, $p < 0.001$) and

rhythm ($d = 0.58$, $p < 0.001$). We also tested linguistic proximity by computing Spearman correlations between pairwise musical JSD and pairwise linguistic distance (Gurevich et al., 2025). Musically similar countries tend to share more similar spoken languages (melody: $\rho = 0.26$, bootstrapped 95% CI = [0.04, 0.45], $p = 0.016$; rhythm: $\rho = 0.27$ [0.10, 0.44], $p = 0.002$). Together, these expected geolinguistic alignments provide convergent validity that our extracted melodic and rhythmic features capture culturally meaningful variation.

However, regional proximity does not imply uniform similarity across both musical dimensions. As visual examples, Figures 2B&C illustrate regional aggregate profiles comparing North America versus Latin America & the Caribbean, showing that the extracted melodic-interval and rhythmic-ratio distributions yield interpretable, systematically different patterns across regions. For these two regional comparisons, the melodic intervals appear very similar, suggesting no major differences in melodic structure. In contrast, the rhythmic distributions show noticeable differences, namely the prominence of 2:3/3:2 ratios in Latin America & the Caribbean. These ratios form subcomponents of more complex rhythmic patterns, such as the 3:3:2 *tresillo* (Kubik, 1979; Rocamora, 2018), and have also been reported as significant in human perceptual experiments conducted in this region (Jacoby et al., 2024).

Melodic and rhythmic diversity are independent

Having established that the pipeline captures meaningful variance of musical characteristics, we now directly tested whether heterogeneity in one component predicts heterogeneity in the other (coupled evolution) or whether they vary independently.

¹<https://www.worldbank.org>

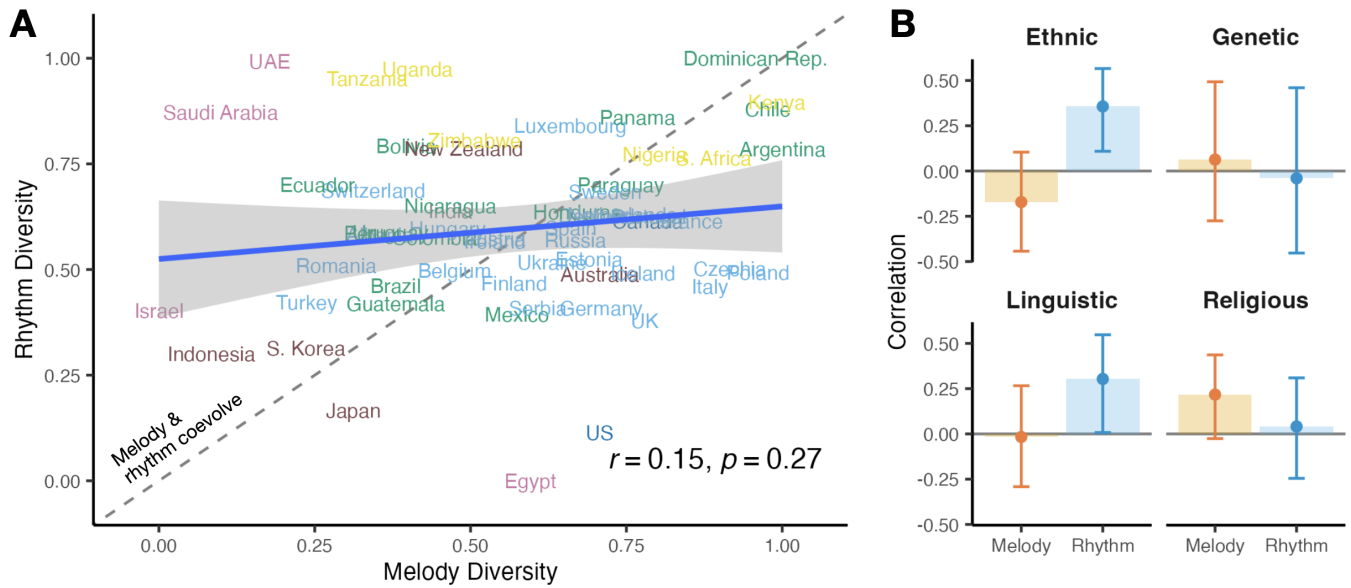


Figure 3: Within-country melodic and rhythmic diversity demonstrate independence. (A) Scatter plot of melodic diversity versus rhythmic diversity across countries, colored by world region, showing no significant relationship between the two ($p = 0.27$). The dashed diagonal line represents perfect correlation, a theoretical scenario in which the two components could have coevolved, resulting in similar levels of diversities. The observed pattern towards being flat indicates independence between these two components. (B) Association between musical diversity and diversity in population demographics (‘Socio-demographic factors’ in Methods), showing countries with diverse mix of ethnic groups and language tend to have more diverse rhythm but not melody. Error bars indicate bootstrapped 95% CI of the mean.

We quantified the diversity of melodic and rhythmic profiles within each country by computing the JSD between all pairs of songs, summarizing by the median (robust to skew and outliers), and normalizing values to range between 0 and 1, where 1 indicates maximum diversity.

Figure 3A plots melodic diversity against rhythmic diversity across all 59 countries and shows no significant association ($p = 0.27$). Under a coupled-change account, countries with high melodic diversity should also exhibit high rhythmic diversity, yielding a positive relationship aligned with the diagonal. Instead, the relationship is effectively flat, indicating that melodic and rhythmic diversity vary independently across cultures. To ensure this pattern is not driven by regional clustering (i.e., Simpson’s paradox), we performed partial correlation by removing regional means, which revealed the same outcome ($p = 0.41$).

Regional structure nevertheless revealed substantial differences in where countries fall in this two-dimensional space. Sub-Saharan African countries cluster toward high rhythmic diversity (mean regional diversity $D = 0.86$, 95% CI = [0.76, 0.96]). Europe and Central Asia occupy an intermediate range for both melodic and rhythmic diversity (melody = 0.64 [0.56, 0.71]; rhythm = 0.55 [0.51, 0.59]). Middle East and North Africa show notably low melodic diversity (0.22 [−0.20, 0.64]), while East Asia and Pacific (particularly Japan, Korea, and Indonesia) appear relatively homogeneous in both dimensions (melody = 0.37 [0.067, 0.66]; rhythm =

0.41 [0.12, 0.71]).

Together, these results indicate that melodic and rhythmic diversity are unevenly distributed across regions, and that high diversity in one component does not imply high diversity in the other. In particular, Sub-Saharan Africa stands out as especially diverse in rhythmic patterns, consistent with ethnomusicological accounts emphasizing the diversity of rhythm in many African musical traditions (Polak, 2025).

Independent socio-demographic constraints

If melody and rhythm evolve independently, they may also respond to different socio-demographic constraints. We tested this by examining whether within-country musical diversity reflects broader patterns of demographic diversity across ethnic, genetic, linguistic, and religious dimensions (‘Socio-demographic factors’ in Methods). Rhythmic diversity showed positive associations with both ethnic diversity ($r = 0.36$, $p = 0.007$, Benjamini-Hochberg $p_{\text{adj}} = 0.054$) and linguistic diversity ($r = 0.30$, $p = 0.044$, $p_{\text{adj}} = 0.18$), suggesting that countries with more heterogeneous populations tend to exhibit greater variety in their rhythmic patterns. In contrast, melodic diversity showed no significant association with any demographic diversity measure (all $p > 0.08$, all $p_{\text{adj}} > 0.21$). Consistent with recent findings by Passmore et al., 2024 showing weak links between musical and genetic similarities, neither melody nor rhythm demonstrated associations with genetic diversity. Together, these correlations

possibly suggest that rhythmic diversity weakly aligns with ethnolinguistic heterogeneity, consistent with findings of perceptual enculturation and niche construction in rhythm (Polak et al., 2026), whereas melodic diversity remains decoupled from these demographic dimensions.

Discussion

Our computational pipeline enabled the large-scale, within-song comparison of melodic and rhythmic variation across cultures. By extracting both components from the same recordings using source separation, we bypassed the scalability limitations and potential biases inherent in manual annotation systems. This methodological advance revealed cross-cultural patterns that would otherwise be difficult to detect through traditional approaches.

We found that melody and rhythm were not significantly correlated in their within-country diversity. For instance, Sub-Saharan Africa's exceptionally high rhythmic diversity combined with moderate melodic diversity, or the near-identical melodic profiles but divergent rhythmic distributions between North America and Latin America & the Caribbean, exemplify how cultural exchange can homogenize one dimension while preserving distinctiveness in another. These patterns suggest that melody and rhythm are not merely different features of a monolithic 'musical style', but rather partially independent systems responding to different evolutionary and cultural pressures.

These results complement the findings of McBride, Kim, et al., 2025, who operationalized both melody and rhythm from the vocal line itself and found correlated entropy between pitch and duration dimensions, suggesting co-evolution. Our finding that vocal melody and percussive rhythm vary independently is not contradictory. The rhythm extracted from drums and the rhythm inherent in vocal phrasing likely reflect different levels of musical organization. In the case of McBride, Kim, et al., 2025, vocal pitch and vocal rhythm are produced by the same "instrument" (the singer) and constrained by the same physiology such as breathing, whereby we would expect some coupling between the two. Percussive rhythm, by contrast, involves separate performers, instruments, and transmission pathways, allowing it to diverge from melodic structure.

Our results point to potential differential mechanisms of cultural transmission. Rhythm appears more tightly coupled to local population structure, with rhythmic diversity tracking ethnolinguistic heterogeneity, a pattern consistent with rhythm serving as a core constituent of cultural practice (Polak et al., 2026) and marker of group identity and social cohesion (Brown et al., 2014; Savage et al., 2020). In contrast, melodic diversity shows no such demographic associations, suggesting that melody may more easily diffuse across cultural boundaries and be more adaptable to change than rhythm (Bomin et al., 2016).

Limitations and Future Works

The reliance on YouTube chart data may underrepresent traditional and non-commercial music, particularly in undersampled regions like Sub-Saharan Africa. Future work should incorporate traditional music repositories to ensure broader cultural coverage (Jacoby et al., 2020). Moreover, like many applications in automatic music analysis, source separation algorithms are primarily trained on Western popular music and may carry cultural biases in distinguishing vocal and drum tracks. For instance, the same algorithm would likely fail to extract timpani lines from Western classical orchestral music. A systematic evaluation of source-separation quality across genres and decades would strengthen confidence in the pipeline; for instance, running a human listener evaluation experiment across a diverse selection of music (Lee, Çelen, et al., 2025; Lee et al., 2021).

Our operational definitions are necessarily selective: we define melody as pitch trajectory (voice only) and rhythm as inter-onset intervals (drums only), thereby excluding contributions from other instruments and overlooking how voice conveys rhythm, drums convey melody, and ensemble interplay creates emergent melodic-rhythmic structures. Similarly, we do not capture how melody is shaped by timbre and instrumentation, or how rhythm is constructed from event duration and pitch-timbre contrasts. While these choices reflect practical constraints, future work could systematically compare alternative operationalization. We also note that melodic pitch-interval distributions span a higher-dimensional space than the bounded rhythmic ratio distributions (which fall between 0 and 1), which may make melodic diversity inherently noisier to estimate from limited samples. Future work should assess the sensitivity of diversity estimates to sample size and alternative summary statistics (e.g., entropy).

Finally, our cross-sectional design captures contemporary patterns but cannot establish causal relationships between socio-demographic factors and musical features. The observed differential associations that, rhythm is linked to ethnolinguistic heterogeneity while melody does not, may reflect independent evolutionary trajectories, but future work using longitudinal datasets or natural experiments (Anglada-Tort, Lee, et al., 2023; Lee, Jacoby, et al., 2025) are needed to test causal hypotheses and trace historical change.

Code and Data Availability

Audio obtained for research analysis purposes cannot be redistributed and thus not made available. Instead, we make available the full metadata of the sampled songs in the study, along with core analysis scripts, and aggregated distributions of melody and rhythm. Publicly available at <https://github.com/harin-git/mel-rhy>.

Acknowledgments

We greatly appreciate the feedback from the reviewers, including John McBride, who provided detailed and constructive comments that substantially improved this manuscript.

References

- Anglada-Tort, M., Harrison, P. M. C., Lee, H., & Jacoby, N. (2023). Large-scale iterated singing experiments reveal oral transmission mechanisms underlying music evolution. *Current Biology*, 0(0). <https://doi.org/10.1016/j.cub.2023.02.070>
- Anglada-Tort, M., Lee, H., Krause, A. E., & North, A. C. (2023). Here comes the sun: Music features of popular songs reflect prevailing weather conditions. *Royal Society Open Science*, 10(5), 221443. <https://doi.org/10.1098/rsos.221443>
- Barbieri, C., Blasi, D. E., Arango-Isaza, E., Sotiropoulos, A. G., Hammarström, H., Wichmann, S., Greenhill, S. J., Gray, R. D., Forkel, R., Bickel, B., & Shimizu, K. K. (2022). A global analysis of matches and mismatches between human genetic and linguistic histories. *Proceedings of the National Academy of Sciences*, 119(47), e2122084119. <https://doi.org/10.1073/pnas.2122084119>
- Bianco, R., Tóth, B., Bigand, F., Nguyen, T., Sziller, I., Háden, G. P., Winkler, I., & Novembre, G. (2025, February). Human newborns form musical predictions based on rhythmic but not melodic structure. <https://doi.org/10.1101/2025.02.19.639016>
- Bomin, S. L., Lécointre, G., & Heyer, E. (2016). The Evolution of Musical Diversity: The Key Role of Vertical Transmission. *PLOS ONE*, 11(3), e0151570. <https://doi.org/10.1371/journal.pone.0151570>
- Brown, S., Phillips, E., Husein, K., & McBride, J. (2025). Musical scales optimize pitch spacing: A global analysis of traditional vocal music. *Humanities and Social Sciences Communications*, 12(1), 546. <https://doi.org/10.1057/s41599-025-04881-1>
- Brown, S., Savage, P. E., Ko, A. M.-S., Stoneking, M., Ko, Y.-C., Loo, J.-H., & Trejaut, J. A. (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society B: Biological Sciences*, 281(1774), 20132072. <https://doi.org/10.1098/rspb.2013.2072>
- Castellanos, F. J., Garrido-Munoz, C., Ríos-Vila, A., & Calvo-Zaragoza, J. (2022). Region-based layout analysis of music score images. *Expert Systems with Applications*, 209, 118211. <https://doi.org/10.1016/j.eswa.2022.118211>
- Clayton, M., Herbert, T., & Middleton, R. (Eds.). (2003). *The Cultural Study of Music: A Critical Introduction*. Routledge.
- Daikoku, H., Ding, S., Sanne, U. S., Benetos, E., Wood, A., Shimozone, T., Fujii, S., & Savage, P. E. (2022). Agreement among human and automated estimates of similarity in a global music sample. *Proc. of the 2022 Folk Music Analysis Workshop*. <https://doi.org/10.31234/osf.io/76fmq>
- Fearon, J. D. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, 8(2), 195–222. <https://doi.org/10.1023/A:1024419522867>
- Gurevich, T., Herman, P. R., Toubal, F., & Yotov, Y. V. (2025). A Dataset on Linguistic Connectivity Across and Within Countries. *Scientific Data*, 12(1), 542. <https://doi.org/10.1038/s41597-025-04692-8>
- Háden, G. P., Bouwer, F. L., Honing, H., & Winkler, I. (2024). Beat processing in newborn infants cannot be explained by statistical learning based on transition probabilities. *Cognition*, 243, 105670. <https://doi.org/10.1016/j.cognition.2023.105670>
- Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology. Human Perception and Performance*, 38(3), 543–548. <https://doi.org/10.1037/a0027225>
- Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT press.
- Jacoby, N., Margulis, E. H., Clayton, M., Hannon, E., Honing, H., Iversen, J., Klein, T. R., Mehr, S. A., Pearson, L., Peretz, I., Perlman, M., Polak, R., Ravignani, A., Savage, P. E., Steingo, G., Stevens, C. J., Trainor, L., Trehub, S., Veal, M., & Wald-Fuhrmann, M. (2020). Cross-Cultural Work in Music Cognition Challenges, Insights, and Recommendations. *Music Perception*, 37(3), 185–195. <https://doi.org/10.1525/mp.2020.37.3.185>
- Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Godoy, R., Undurraga, E. A., Huanca, T., Thalwitzer, T., Doumbia, N., Goldberg, D., Margulis, E. H., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., Konno, R., . . . McDermott, J. H. (2024). Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries. *Nature Human Behaviour*, 1–32. <https://doi.org/10.1038/s41562-023-01800-9>
- Jacoby, N., Polak, R., & London, J. (2021). Extreme precision in rhythmic interaction is enabled by role-optimized sensorimotor coupling: Analysis and modelling of West African drum ensemble music. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1835), 20200331. <https://doi.org/10.1098/rstb.2020.0331>
- Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and Non-universal Features of Musical Pitch Perception Revealed by Singing. *Current Biology*, 29(19), 3229–3243.e12. <https://doi.org/10.1016/j.cub.2019.08.020>
- Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, 22(10), 896–910. <https://doi.org/10.1016/j.tics.2018.08.002>
- Kubik, G. (1979). Angolan traits in black music, games and dances of Brazil. A study of African cultural extensions overseas. *Estudos de Antropologia Cultural. Junta de Investigações Científicas de Ultramar Lisboa*, (10), 1–55.
- Lee, H., Çelen, E., Harrison, P., Anglada-Tort, M., Rijn, P. v., Park, M., Schönwiesner, M., & Jacoby, N. (2025). GlobalMood: A cross-cultural benchmark for music emotion recognition. *The 26th International Society for Music Information Retrieval Conference*. <https://doi.org/10.48550/arXiv.2505.09539>

- Lee, H., Höger, F., Schönwiesner, M., Park, M., & Jacoby, N. (2021). Cross-cultural Mood Perception in Pop Songs and its Alignment with Mood Detection Algorithms. *The 22nd International Society for Music Information Retrieval Conference*, 366–373. <https://doi.org/10.5281/zenodo.5625680>
- Lee, H., Jacoby, N., Hennequin, R., & Moussallam, M. (2025). Mechanisms of cultural diversity in urban populations. *Nature Communications*, 16(1), 5192. <https://doi.org/10.1038/s41467-025-60538-2>
- Lomax, A. (Ed.). (1978). *Folk Song Style and Culture*. Routledge. <https://doi.org/10.4324/9780203791844>
- London, J. (2002). Cognitive Constraints on Metric Systems: Some Observations and Hypotheses. *Music Perception*, 19, 529–550. <https://doi.org/10.1525/mp.2002.19.4.529>
- Mauch, M., & Dixon, S. (2014). PYIN: A fundamental frequency estimator using probabilistic threshold distributions. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 659–663. <https://doi.org/10.1109/ICASSP.2014.6853678>
- Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, 2(5), 150081. <https://doi.org/10.1098/rsos.150081>
- McBride, J. M., Kim, N., Nishikawa, Y., Saadakeev, M., Pearce, M. T., & Tlusty, T. (2025). Information and motor constraints shape melodic diversity across cultures. <https://arxiv.org/abs/2408.12635>
- McBride, J. M., Phillips, E., Savage, P. E., Brown, S., & Tlusty, T. (2025). Melody predominates over harmony in the evolution of musical scales across 96 countries. <https://arxiv.org/abs/2408.12633>
- Mehr, S. A. (2025). Core systems of music perception. *Trends in Cognitive Sciences*, 29(8), 763–777. <https://doi.org/10.1016/j.tics.2025.05.013>
- Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, 366(6468). <https://doi.org/10.1126/science.aax0868>
- Ozaki, Y., Tierney, A., Pfordresher, P. Q., McBride, J. M., Benetos, E., Proutskova, P., Chiba, G., Liu, F., Jacoby, N., Purdy, S. C., Opondo, P., Fitch, W. T., Hegde, S., Rocamora, M., Thorne, R., Nweke, F., Sadaphal, D. P., Sadaphal, P. M., Hadavi, S., ... Savage, P. E. (2024). Globally, songs and instrumental melodies are slower and higher and use more stable pitches than speech: A Registered Report. *Science Advances*, 10(20), eadm9797. <https://doi.org/10.1126/sciadv.adm9797>
- Passmore, S., Wood, A. L. C., Barbieri, C., Shilton, D., Daikoku, H., Atkinson, Q. D., & Savage, P. E. (2024). Global musical diversity is largely independent of linguistic and genetic histories. *Nature Communications*, 15(1), 3964. <https://doi.org/10.1038/s41467-024-48113-7>
- Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, 119(5 Pt 1), 3034–3047. <https://doi.org/10.1121/1.2179657>
- Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., & Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics*, 91(2), 275–292. <https://doi.org/10.1016/j.ajhg.2012.06.014>
- Polak, R., Danielsen, A., Jacoby, N., Pearson, L., Horlor, S., Aubinet, S., & Nozaradan, S. (2026). Culture in psychology and neuroscience: Concepts, relevance, and empirical evidence in rhythm perception. https://doi.org/10.31234/osf.io/sqbzn_v2
- Polak, R. (2025). From Mali to Ghana: Pulsation Non-Isochrony in West African Percussion Music Genres. <https://doi.org/10.5281/zenodo.14675227>
- Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic prototypes across cultures: A comparative study of tapping synchronization. *Music Perception*, 36(1), 1–23. <https://doi.org/10.1525/mp.2018.36.1.1>
- Rice, T. (2017, March). *Modeling Ethnomusicology*. Oxford University Press.
- Rocamora, M. (2018). *Computational methods for percussion music analysis: The afro-uruguayan candombe drumming as a case study* [Doctoral dissertation, Universidad de la República].
- Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current biology: CB*, 30(18), 3544–3555.e6. <https://doi.org/10.1016/j.cub.2020.06.072>
- Rouard, S., Massa, F., & Défossez, A. (2023). Hybrid Transformers for Music Source Separation. *ICASSP 23*.
- Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29), 8987–8992. <https://doi.org/10.1073/pnas.1414495112>
- Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2020). Music as a coevolved system for social bonding. *The Behavioral and Brain Sciences*, 44, e59. <https://doi.org/10.1017/S0140525X20000333>
- Tenney, J. (1988). *A history of 'consonance' and 'dissonance'*. Excelsior Music Publishing Company.
- Vos, P. G., & Troost, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, 6(4), 383–396.
- Yurdum, L., Singh, M., Glowacki, L., Vardy, T., Atkinson, Q. D., Hilton, C. B., Sauter, D., Krasnow, M. M., & Mehr, S. A. (2023). Universal interpretations of vocal music. *Pro-*

ceedings of the National Academy of Sciences, 120(37), e2218593120. <https://doi.org/10.1073/pnas.2218593120>

Zivic, P. H., Shifres, F., & Cecchi, G. A. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences*, 110(24), 10034–10038. <https://doi.org/10.1073/pnas.1222336110>