Bias beyond Borders: Global Inequalities in AI-Generated Music

Ahmet Solak **Florian Grötschla** **Luca A. Lanzendörfer** **Roger Wattenhofer**

ETH Zurich

{asolak, fgroetschla, lanzendoerfer, wattenhofer}@ethz.ch

Abstract

While recent years have seen remarkable progress in music generation models, research on their biases across countries, languages, cultures, and musical genres remains underexplored. This gap is compounded by the lack of datasets and benchmarks that capture the global diversity of music. To address these challenges, we introduce GlobalDISCO, a large-scale dataset consisting of 73k music tracks generated by state-of-the-art commercial generative music models, along with paired links to 93k reference tracks in LAION-DISCO-12M. The dataset spans 147 languages and includes musical style prompts extracted from MusicBrainz and Wikipedia. The dataset is globally balanced, representing musical styles from artists across 79 countries and five continents. Our evaluation reveals large disparities in music quality and alignment with reference music between high-resource and low-resource regions. Furthermore, we find marked differences in model performance between mainstream and geographically niche genres, including cases where models generate music for regional genres that more closely align with the distribution of mainstream styles.

1 Introduction

In terms of quality and performance, the music generation field has seen remarkable progress in recent years, with commercial systems achieving exceptional results, even outperforming real music in large-scale human evaluation studies [3]. However, despite music being a universal human experience found in all cultures around the world [5], recent studies have highlighted a significant lack of intercultural and multilingual datasets in music generation research [14]. These findings, combined with the rapid progress of generative models, further underscore the urgent need for resources that allow the evaluation of potential biases and weaknesses in these models. In other domains, biases across world regions and cultures have been more widely researched, with benchmarks and datasets released that aim to address intercultural biases in both the image [6] and language domains [16, 19]. To the best of our knowledge, the only publicly available multilingual generated music dataset [13] contains only music in 3 different languages. In comparison, the recently published BLEND [16] and CVQA [19] benchmarks for large language models and multimodal large language models have a global coverage, with BLEnD covering 13 languages and 16 different countries and CVQA covering 31 languages and 30 different countries.

To address these challenges in the music generation field, we present the GlobalDISCO dataset, which is designed to evaluate biases and diversity in music generation. GlobalDISCO consists of 93k real and 73k generated music from 79 countries, across five continents and 147 languages. It is also larger in scale and includes music from more commercial platforms compared to the current largest synthetic music dataset, SONICS [18], which consists of 49k English music tracks generated with Suno [20] and Udio [22] models. The tracks in GlobalDISCO are generated with four state-of-the-art

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Music.

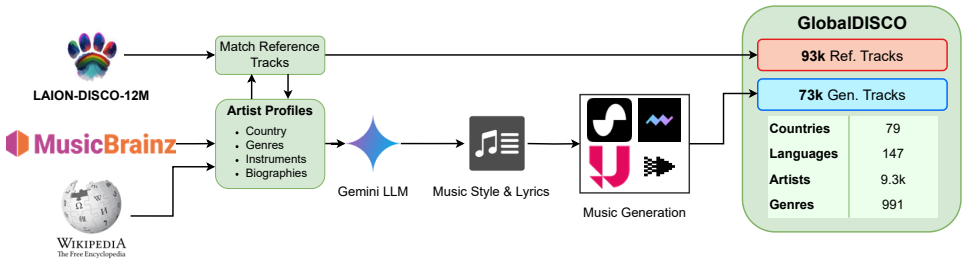

Figure 1: Pipeline of data collection and audio generation for GlobalDISCO. We gather artist information from MusicBrainz and Wikipedia, match it with reference tracks from LAION-DISCO-12M, and construct artist profiles based on this information. These profiles are then used to generate music using state-of-the-art music generation models, resulting in a globally diverse dataset of both generated tracks and reference tracks.

commercial models: Udio [22], Suno [20], Mureka [15], and Riffusion [17]. Their performances and biases are explored across geographical regions and genres to provide a representative evaluation of the current capabilities and limitations of available music generation systems.

Analyzing GlobalDISCO, we find that state-of-the-art music generation models are highly biased across both world regions and genres, and that they generate music much more out-of-distribution for lower-resource regions and genres compared to higher-resource regions. Furthermore, when instructed to generate music for certain regional genres, the models often produce music that more closely aligns with the distribution of mainstream genres. By releasing GlobalDISCO as a public resource, we aim to support the research community in identifying and addressing biases in music generation and to promote greater global diversity in future model development.

2 Methodology

Dataset construction. To construct the GlobalDISCO dataset, we begin by collecting artist entries from MusicBrainz, selecting those that include information about the artist's geographical area, as well as links to additional biographical pages. This initial step gives us 148k artist profiles. For artists without Wikipedia articles linked directly from their MusicBrainz pages, we perform supplementary searches using artist names on English Wikipedia. We select articles that match the MusicBrainz profiles by name and at least two additional attributes, such as area or genre tags. We enrich this artist metadata with reference tracks from the LAION-DISCO-12M [12, 11] dataset by matching artist and channel names, as well as verifying discography overlap when there is more than a single match. We retain the top-10 most viewed tracks per artist, with view numbers taken from LAION-DISCO-12M. To focus on vocal music, we exclude artists associated only with instrumental genres (i.e., the classical or electronic genre remains in the dataset only because they occurred together with vocal genres in artist metadata). From the 34k artists that fulfill this criteria we ensure a balanced global representation by selecting up to a threshold $t = 374$ artists per country, where the value of t is determined via binary search to yield a dataset with at least 10k artists. For each artist, we construct a profile using the collected metadata and generate musical style descriptions and synthetic lyrics for the artists using Gemini [21]. Examples of artist profiles and LLM system prompts are shown in Appendix A and Appendix B, respectively. Using these prompts and lyrics, we generate music with four state-of-the-art music generation models: Suno (v4) [20], Udio (v1.5 Allegro) [22], Mureka (v6) [15], and Riffusion (FUZZ 0.8) [17]. All four models are commercial black boxes that take musical styles and lyrics as textual input to generate music tracks.

The final dataset includes 9.3k artists, for which all models were successfully able to generate music and reference tracks were available in LAION-DISCO-12M. 79 countries across 5 continents are represented in the dataset, with a minimum of 10 artists per country. A world map showing all countries is presented in Appendix C. We identify more than 991 genres using the MusicBrainz list of genre tags[1] as well as 147 different languages among our generated lyrics using the GlotLID language identification model [7]. From all languages in the dataset, 18 languages have more than

[1] `https://musicbrainz.org/genres`

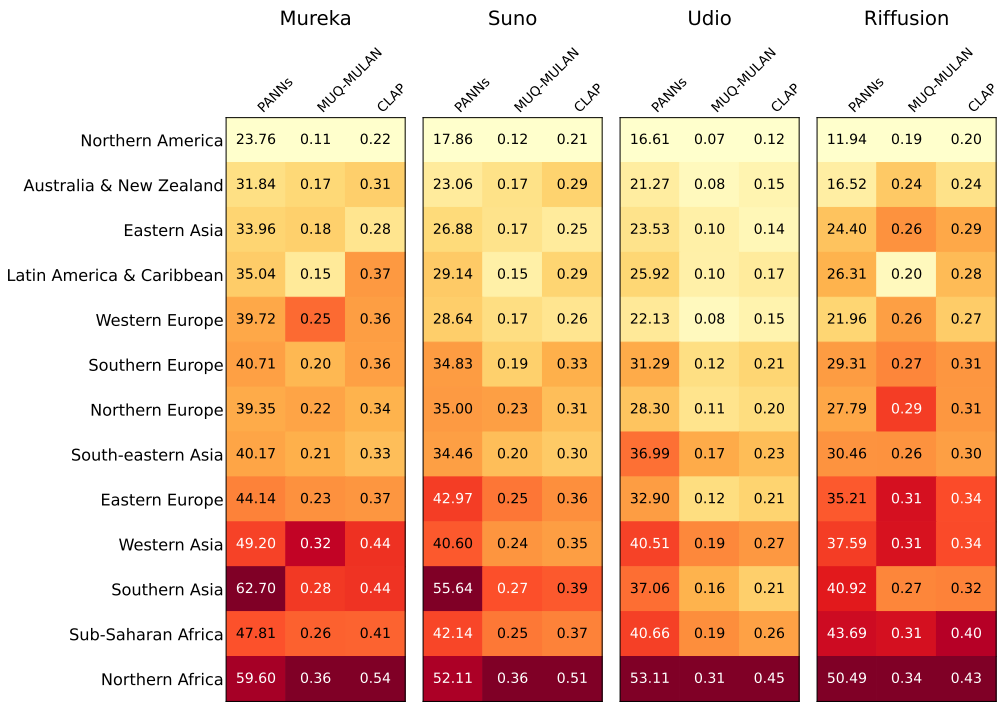

Figure 2: Mean FAD scores (lower is better), averaged across the countries for world sub-regions. The regions are ordered by the mean z-scored FAD scores across embeddings. We find that the distributions between generated and reference tracks varies greatly, with generated music distributions from low-resource regions being significantly different compared to their reference counterparts.

100 different artists associated with each language. A high-level overview of the entire data collection, data processing, and music generation pipeline is shown in Fig. 1. For more details on the frequency and percentage of all 147 languages, please refer to Appendix D.

Evaluation. To evaluate generated and reference music, we choose a number of audio embedding models. We use the PANNs [9] and CLAP [25] audio embedding models, which have shown good alignment with human preference in prior work [3, 4]. For CLAP we choose the `music_audioset_epoch_15_esc_90.14` checkpoint. As CLAP takes 10 second audio inputs, we compute embeddings for 10-second windows across the tracks with 1-second hops and then take the mean of those embeddings as the final CLAP embedding per track. Additionally, we consider MUQ-MULAN [26], which reports state-of-the-art results on music tagging tasks. Using these embedding models, we use the Frechet Audio Distance (FAD) [8] and the Kernel Audio Distrance (KAD) [1] metrics. FAD compares evaluation and reference audio sets by comparing their multivariate Gaussian distributions. KAD is a more recently proposed distribution-free alternative, which is based on the Maximum Mean Discrepancy [2]. For the kernel function in KAD we use the Gaussian radial basis function kernel, as proposed by the authors [1]. For both FAD and KAD lower scores are better.

3 Results

We first explore the difference in music generation quality across different world sub-regions, as defined by the UN M49 Standard [23]. For the PANNs, CLAP, and MUQ-MULAN embedding models, we present FAD scores between generated and reference tracks for the 12 world regions present in GlobalDISCO, shown as a heatmap in Fig. 2. We show the same analysis with KAD in Appendix E. The results clearly show how model performance varies significantly across higher- and lower-resource regions. For all music generation models, world regions from the continent of Africa, as well as Southern and Western Asia, generate music that is considerably more out-of-distribution compared to higher-resource regions. At the other end, Northern America, which is likely the highest-resource region in terms of available music datasets, shows excellent results across the board.

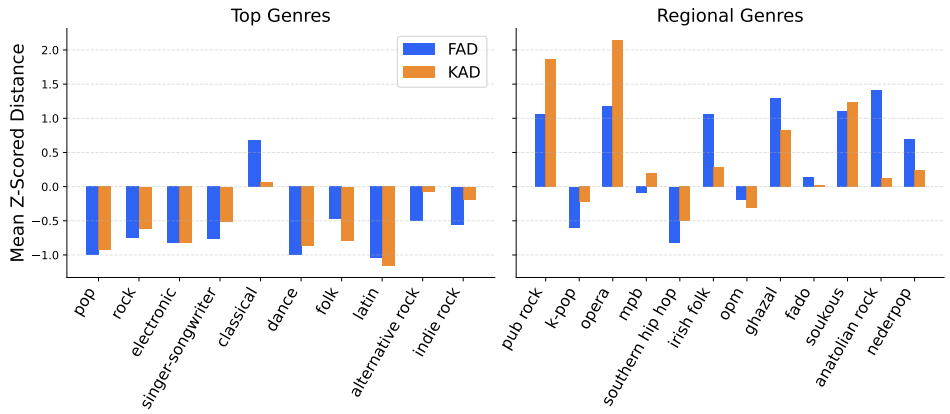

Figure 3: Mean Z-scored FAD and KAD scores across the top ten most popular (left) and top regional music genres (right). Scores are first normalized (Z-scored) across genres for each combination of music generation and embedding model, and then averaged across all such combinations. Current state-of-the-art models seem to have difficulty generating audio for more regional genres, with most exhibiting higher (worse) FAD and KAD scores compared to mainstream genres.

In Fig. 3, we show the mean normalized (z-scored) FAD and KAD scores across popular genres and regional genres, averaged across all music generation models. The ten most popular genres are selected by frequency in the dataset, whereas the regional genres are selected using a tf-idf-like method that gives higher value to genres more frequent in a small number of countries. The methodology and the tf-idf-like scores are described in more detail in Appendix F. We find large differences between popular and regional genres, as well as between more modern and traditional genres such as classical and opera.

We further explore the qualitative distribution of the most difficult regional genre, opera, which has the lowest average FAD and KAD scores among all popular and regional genres considered in Fig. 3.

In Fig. 4, we show t-SNE [24] plots for generated and reference tracks for opera, as well as for pop, the most frequent genre in the dataset. The figures show that some models, such as Suno, seem to generate tracks more in-distribution with real pop music compared to the reference music of their own genre, whereas Udio does better at generating music within distribution of the same reference genre. These qualitative observations are reflected in quantitative results: for Suno, the FAD and KAD scores between generated opera and real pop music (0.35 and 15.94) are notably lower than those between generated and real opera (0.48 and 36.93). Mureka shows a similar trend, with FAD/KAD scores of $0.43/22.11$ for opera–pop and $0.53/42.24$ for opera–opera comparisons. More detailed FAD and KAD scores comparing generated regional genre tracks with reference mainstream genre tracks are presented in Appendix G.

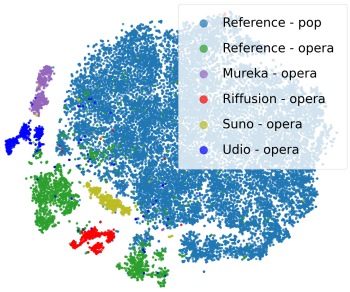

Figure 4: T-SNE plot of generated and reference opera tracks as well as pop reference music, embedded with the CLAP model. Generated opera songs from Suno and Mureka lie closer to reference pop tracks than to reference opera tracks.

Conclusion. In this work we presented GlobalDISCO, a large-scale generated music dataset encompassing musical traditions from around the world aimed at exploring the potential biases in music generation models and addressing the lack of large, multicultural, and multilingual datasets in the generative music domain. Our findings reveal substantial disparities in the ability of models to generate music from lower-resource regions, such as Northern Africa, Sub-Saharan Africa, and Southern Asia. We also observe genre-specific biases, which affect not only lower-resource and regional genres, such as soukous and ghazal, but also more traditional European genres, including classical and opera. As generated music continues to grow in popularity and quality, our results highlight clear biases against lower-resource musical traditions and the need to address these biases to preserve global musical diversity.

References

[1] Yoonjin Chung, Pilsun Eu, Junwon Lee, Keunwoo Choi, Juhan Nam, and Ben Sangbae Chon. Kad: No more fad! an effective and efficient evaluation metric for audio generation, 2025. URL https://arxiv.org/abs/2502.15602.

[2] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL http://jmlr.org/papers/v13/gretton12a.html.

[3] Florian Grötschla, Ahmet Solak, Luca A Lanzendörfer, and Roger Wattenhofer. Benchmarking music generation models and metrics via human preference studies. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2025.

[4] Azalea Gui, Hannes Gamper, Sebastian Braun, and Dimitra Emmanouilidou. Adapting frechet audio distance for generative music evaluation, 2024. URL https://arxiv.org/abs/2311.01616.

[5] Donald A Hodges. *Music in the human experience: An introduction to music psychology*. Routledge, 2019.

[6] Nithish Kannen, Arif Ahmad, Marco Andreetto, Vinodkumar Prabhakaran, Utsav Prabhu, Adji Bousso Dieng, Pushpak Bhattacharyya, and Shachi Dave. Beyond aesthetics: Cultural competence in text-to-image models, 2025. URL https://arxiv.org/abs/2407.06863.

[7] Amir Hossein Kargaran, Ayyoob Imani, François Yvon, and Hinrich Schütze. GlotLID: Language identification for low-resource languages. In *The 2023 Conference on Empirical Methods in Natural Language Processing*, 2023. URL https://openreview.net/forum?id=dl4e3EBz5j.

[8] Kevin Kilgour, Mauricio Zuluaga, Dominik Roblek, and Matthew Sharifi. Fréchet audio distance: A metric for evaluating music enhancement algorithms, 2019. URL https://arxiv.org/abs/1812.08466.

[9] Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, and Mark D Plumbley. Panns: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2880–2894, 2020.

[10] Yanis Labrak, Markus Frohmann, Gabriel Meseguer-Brocal, and Elena V. Epure. Synthetic lyrics detection across languages and genres, 2025. URL https://arxiv.org/abs/2406.15231.

[11] LAION e.V. Laion-disco-12m. https://huggingface.co/datasets/laion/LAION-DISCO-12M, 2024. Dataset of 12 million YouTube music links with metadata; Apache-2.0 license.

[12] Luca A. Lanzendörfer, Florian Grötschla, Emil Funke, and Roger Wattenhofer. Disco-10m: A large-scale music dataset, 2023. URL https://arxiv.org/abs/2306.13512.

[13] Yupei Li, Hanqian Li, Lucia Specia, and Björn W. Schuller. M6: Multi-generator, multi-domain, multi-lingual and cultural, multi-genres, multi-instrument machine-generated music detection databases, 2024. URL https://arxiv.org/abs/2412.06001.

[14] Atharva Mehta, Shivam Chauhan, Amirbek Djanibekov, Atharva Kulkarni, Gus Xia, and Monojit Choudhury. Music for all: Exploring multicultural representations in music generation models, 2025. URL https://arxiv.org/abs/2502.07328.

[15] Mureka. Mureka. https://www.mureka.ai/, 2025. Accessed: April-May, 2025.

[16] Junho Myung, Nayeon Lee, Yi Zhou, Jiho Jin, Rifki Afina Putri, Dimosthenis Antypas, Hsuvas Borkakoty, Eunsu Kim, Carla Perez-Almendros, Abinew Ali Ayele, Víctor Gutiérrez-Basulto, Yazmín Ibáñez-García, Hwaran Lee, Shamsuddeen Hassan Muhammad, Kiwoong Park, Anar Sabuhi Rzayev, Nina White, Seid Muhie Yimam, Mohammad Taher Pilehvar, Nedjma Ousidhoum, Jose Camacho-Collados, and Alice Oh. Blend: A benchmark for llms on everyday knowledge in diverse cultures and languages, 2025. URL https://arxiv.org/abs/2406.09948.

[17] Producer.ai. Producer.ai. https://www.producer.ai/, 2025. Accessed: April-August, 2025.

[18] Md Awsafur Rahman, Zaber Ibn Abdul Hakim, Najibul Haque Sarker, Bishmoy Paul, and Shaikh Anowarul Fattah. Sonics: Synthetic or not – identifying counterfeit songs, 2025. URL https://arxiv.org/abs/2408.14080.

[19] David Romero et al. Cvqa: Culturally-diverse multilingual visual question answering benchmark, 2024. URL https://arxiv.org/abs/2406.05967.

[20] Suno. Suno. https://suno.com/, 2025. Accessed: April-May, 2025.

[21] Gemini Team et al. Gemini: A family of highly capable multimodal models, 2025. URL https://arxiv.org/abs/2312.11805.

[22] Udio. Udio. https://www.udio.com/, 2025. Accessed: April-May, 2025.

[23] United Nations Statistics Division. Standard country or area codes for statistical use (m49). https://unstats.un.org/unsd/methodology/m49/. Accessed 27 Aug 2025. Online version of United Nations publication Series M, No. 49.

[24] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL http://jmlr.org/papers/v9/vandermaaten08a.html.

[25] Yusong Wu*, Ke Chen*, Tianyu Zhang*, Yuchen Hui*, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2023.

[26] Haina Zhu, Yizhi Zhou, Hangting Chen, Jianwei Yu, Ziyang Ma, Rongzhi Gu, Yi Luo, Wei Tan, and Xie Chen. Muq: Self-supervised music representation learning with mel residual vector quantization, 2025. URL https://arxiv.org/abs/2501.01108.

A Artist Profiles

Fig. 5 shows a sample artist profile with information gathered from MusicBrainz and Wikipedia which was processed to separate different sections such as genre and instruments, as well as relevant biographical information.

Artist Profile

Artist Name: [Artist Name]
Country: türkiye
Genres: pop; rock, Turkish folk music , Sufi music , Arabesque music , Anatolian Rock
Active Dates: [Active Dates]
Biography Language: English
Biography: [Artist Name] was a Turkish pop and rock band consisted of members [Members]. While many of their songs poke fun at common Turkish types or satirise prejudice and corruption ...

Figure 5: An artist profile constructed with information gathered from MusicBrainz and Wikipedia. The artist's name (in this case, a band) and the names of its members, as well as the active dates, are replaced with placeholders.

B Musical Style and Lyric Generation Prompts

Fig. 6 shows the system prompt to generate musical styles with the Google Gemini LLM, using artist profiles of the form presented in Appendix A. Fig. 7 shows a system prompt for few shot lyrics generation, where we adopt the methodology of previous works [18] to use real lyrics and then generate synthetic lyrics from up to three real samples with few shot inference [10]. For artists without sample lyrics, we use the system prompt in Fig. 8 to generate lyrics based on artist profiles.

C Global Coverage of GlobalDISCO

A world map showing the countries included in GlobalDISCO is presented in Fig. 9.

D Languages in GlobalDISCO

Table 1 shows a comprehensive listing of languages by count and frequency that were found with the GlotLID model[2] [7] among generated lyrics in GlobalDISCO.

Table 1: Counts and percentages of languages in lyrics across GlobalDISCO.

Language	Count	%	Language	Count	%
English	3675	39.41	Duala	2	0.02
Spanish	1368	14.67	Old Norse	2	0.02
Jamaican Patois	395	4.24	Dogri	2	0.02
Portuguese	382	4.10	Pedi	2	0.02
French	339	3.64	Papiamento	2	0.02
German	300	3.22	Yoruba	2	0.02
Japanese	215	2.31	Kriol	2	0.02
Italian	213	2.28	Zande	2	0.02
Finnish	145	1.55	Haitian	2	0.02
Polish	136	1.46	Galician	2	0.02
Turkish	131	1.40	Ainu (Japan)	2	0.02
Russian	124	1.33	Sindhi	2	0.02

[2] https://huggingface.co/cis-lmu/glotlid

Language	Count	%	Language	Count	%
Korean	123	1.32	Tunisian Arabic	1	0.01
Hindi	115	1.23	North Levantine Arabic	1	0.01
Dutch	114	1.22	Luba-Lulua	1	0.01
Indonesian	113	1.21	Makhuwa-Meetto	1	0.01
Mandarin Chinese	112	1.20	Dadibi	1	0.01
Danish	112	1.20	Nyishi	1	0.01
Swedish	88	0.94	Tosk Albanian	1	0.01
Neapolitan	86	0.92	Zyphe Chin	1	0.01
Czech	83	0.89	Gujarati	1	0.01
Hungarian	64	0.69	Eastern Oromo	1	0.01
Montenegrin	52	0.56	Fijian	1	0.01
Filipino	50	0.54	Wolof	1	0.01
Nigerian Pidgin	48	0.51	Mogofin	1	0.01
Vietnamese	40	0.43	Carpathian Romani	1	0.01
Norwegian Bokmål	37	0.40	Northern Rengma Naga	1	0.01
Urdu	33	0.35	Ata Manobo	1	0.01
Persian	31	0.33	Standard Arabic	1	0.01
Bulgarian	28	0.30	Ligurian	1	0.01
Ukrainian	28	0.30	Sicilian	1	0.01
Panjabi	28	0.30	Guianese Creole French	1	0.01
Hebrew	28	0.30	Sidamo	1	0.01
Romanian	26	0.28	Buhutu	1	0.01
Zulu	25	0.27	Makasar	1	0.01
Tamil	25	0.27	Kölsch	1	0.01
Lithuanian	22	0.24	Rarotongan	1	0.01
Egyptian Arabic	21	0.23	Shona	1	0.01
Afrikaans	20	0.21	Gitonga	1	0.01
Slovak	17	0.18	Tswana	1	0.01
Chakavian	16	0.17	Ladino	1	0.01
Bengali	15	0.16	Tsonga	1	0.01
Modern Greek (1453-)	14	0.15	Chakma	1	0.01
Icelandic	14	0.15	Eastern Khumi Chin	1	0.01
Standard Estonian	14	0.15	Scottish Gaelic	1	0.01
Slovenian	13	0.14	Batak Toba	1	0.01
Latin	12	0.13	Maori	1	0.01
Fiji Hindi	12	0.13	Chavacano	1	0.01
North Azerbaijani	11	0.12	Vlax Romani	1	0.01
Western Panjabi	9	0.10	Northern Kurdish	1	0.01
Telugu	9	0.10	Javanese	1	0.01
Kabuverdianu	8	0.09	Dagbani	1	0.01
Standard Latvian	8	0.09	Interlingue	1	0.01
Thai	8	0.09	Bosnian	1	0.01
Swiss German	7	0.08	Croatian	1	0.01
Ogea	6	0.06	Yue Chinese	1	0.01
Catalan	6	0.06	Standard Malay	1	0.01
Norwegian Nynorsk	6	0.06	Southern Balochi	1	0.01
Marathi	5	0.05	Southern Sami	1	0.01
Malayalam	5	0.05	Hausa	1	0.01
Lingala	5	0.05	Iyo	1	0.01
Basque	4	0.04	Morisyen	1	0.01
Bavarian	4	0.04	Mün Chin	1	0.01
Twi	4	0.04	Võro	1	0.01
Angika	4	0.04	Sranan Tongo	1	0.01
Southern Sotho	4	0.04	Aekyom	1	0.01
Xhosa	4	0.04	Rundi	1	0.01
Kannada	4	0.04	Mangareva	1	0.01
Bambara	3	0.03	Welsh	1	0.01
Yawa	3	0.03	Suena	1	0.01

Language	Count	%	Language	Count	%
Irish	3	0.03	Betawi	1	0.01
Sardinian	3	0.03	Breton	1	0.01
Igbo	3	0.03	Swahili	1	0.01
Kabyle	2	0.02			

E KAD Scores across World Sub-Regions

Analogously to Fig. 2, we present KAD scores of music from world sub-regions per music generation model and embedding model in Fig. 10.

F TF-IDF-like Score and Regional Genre Selection

For a genre g in country c, we define its tf-idf-like value as:

$$\mathrm{TFIDF}'_{g,c} = \frac{\mathrm{count}(g,c)}{|\{a \in A \mid \mathrm{country}(a) = c\}|} \cdot \frac{1}{\mathrm{countries}(g)} \tag{1}$$

Where A is the set of artists in GlobalDISCO, $\mathrm{countries}(g)$ is the number of countries that include artists with genre g in their musical style prompt, $\mathrm{count}(g,c)$ is the number of artists with genre g for country c, and $country(a)$ is the country of artist a. Table 2 shows a table of the top-3 genres with the highest tf-idf-like scores per world sub-region.

Musical Style System Prompt

You are a system that receives an artist profile and generates a concise, descriptive prompt suitable for guiding a music generation model. The goal is to encapsulate the artist's musical style in great detail using reliable information.

Please follow these rules:

- Use a combination of tags and free-form text.
- Prioritize information from the following sections (in order): 'Genres', 'Musical Styles', 'Instruments', and 'Biography (X)'.
- Extract only genre or style-relevant content from the section 'Biography (Y)'
- The '(Z) Language' field indicates the language of the 'Biography (Z)' section.
- If there is not enough style-relevant information in the above sections:
  - You may include the artist's nationality or country if it is relevant to their musical style.
  - If the artist performs vocals, you may mention their gender and vocal role if the gender is clearly stated in the artist profile.
- Do not invent or infer any information not present in the artist profile.
- The final prompt must be under 200 characters, but should also include enough detail to meaningfully describe the artist's musical style. Avoid overly short or vague responses.
- Do not include any artist, track, or album names.
- Do not reference influences or similarities to other artists or individuals.
- Do not repeat the exact same genres multiple times.

Figure 6: System prompt for describing an artist's musical style.

Few Shot Lyrics System Prompt

You will be given up to 3 example lyrics from an artist. Generate new lyrics in the style of that artist.

Please follow these rules:

- Use the same language (or languages) as the example lyrics.
- The generated lyrics should be structured similar to the example lyrics in optional section headers like [Verse], [Chorus] and [Bridge] enclosed in square brackts, with the same terminology and language used in the example lyrics.
- Do NOT include the artist's name or copy any content from the examples.
- The generated lyrics should be suitable in length for a 2-minute track. Note that the example lyrics may come from longer songs.
- Generate exactly 24 lines of lyrics, not counting section headers such as [Verse], [Chorus], or [Bridge].
- Only output the structured lyrics. Do not include explanations, thoughts, or additional instructions.

Figure 7: System prompt to generate lyrics in an artist's style with up to 3-shot learning with example lyrics.

Artist Profile Lyrics System Prompt

You are a system that receives an artist profile and generates lyrics in the style of that artist.

Please follow these rules:

- The fields 'Genres', 'Musical Styles', 'Instruments' will give you some general knowledge on how the lyrics could be.
- You may ONLY use parts of 'Biography (X)' and 'Biography (Y)' that give you an idea on the topics and contents of the lyrics.
- The '(Z) Language' field only indicates the language of the 'Biography (Z)' section. It does not have to define the language for the generated lyrics.
- Use the same language that the artist would most likely sing in.
- The generated lyrics should be structured in optional section headers like [Verse], [Chorus] and [Bridge] enclosed in square brackets in the same language the lyrics are in.
- Do NOT include the artist's name or copy any content from real lyrics.
- The generated lyrics should be suitable in length for a 2-minute track. Note that the example lyrics may come from longer songs.
- Generate exactly 24 lines of lyrics, not counting section headers such as [Verse], [Chorus], or [Bridge].
- Only output the structured lyrics. Do not include explanations, thoughts, or additional instructions.

Figure 8: System prompt to generate lyrics in an artist's style, based entirely on the artist's profile.

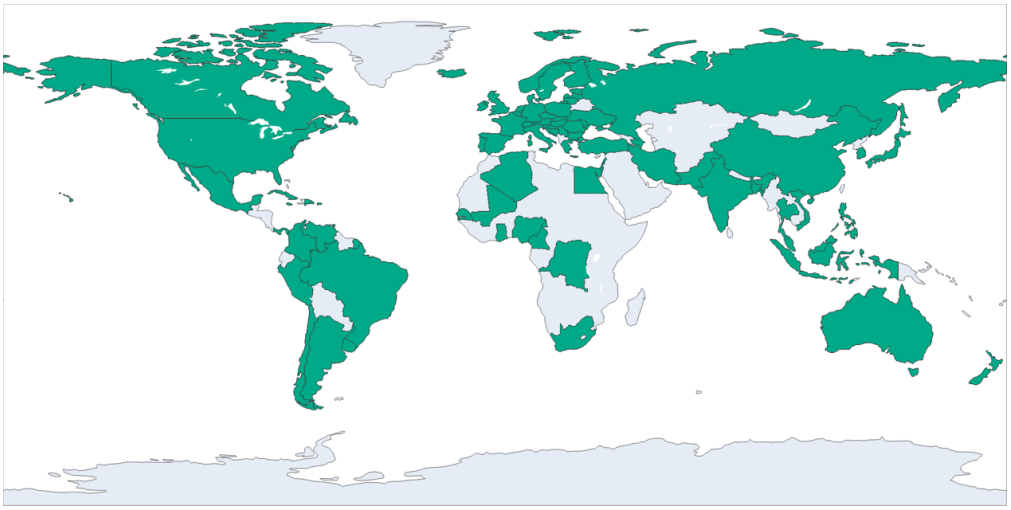

Figure 9: World map with all 79 countries represented in GlobalDISCO shown in green. Each country includes more than 10 reference artists in the dataset.

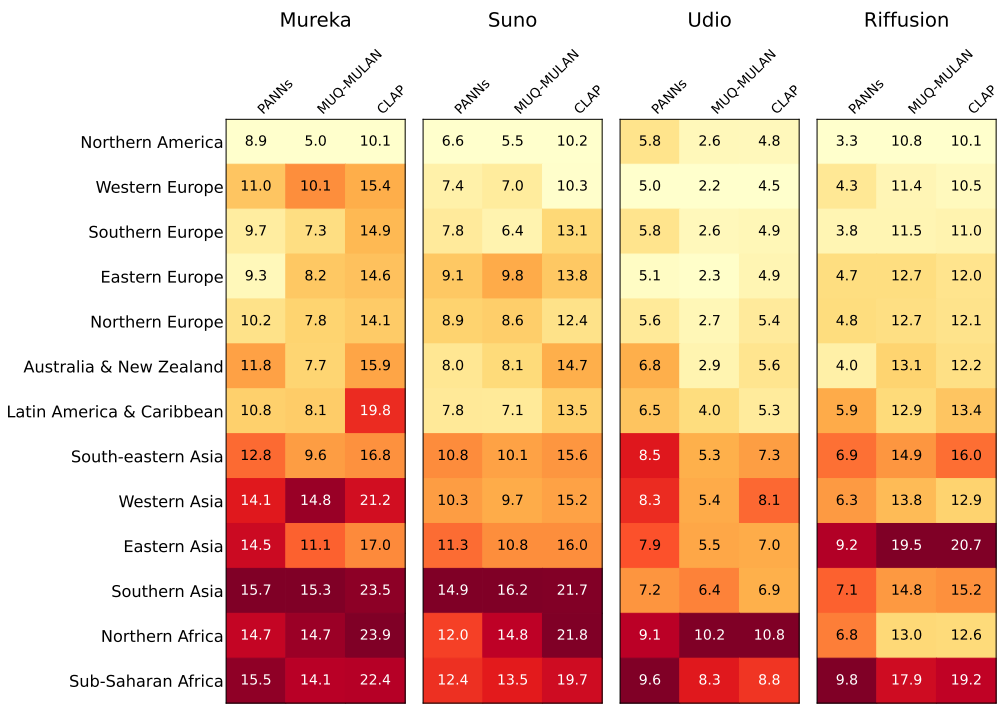

Figure 10: Mean KAD scores (lower is better), averaged across the countries for world sub-regions. The regions are ordered by the mean z-scored KAD scores across embeddings and music generation models.

Table 2: Top-3 regional genres and their country with the highest tf-idf-like scores per region.

Region	Country	Genre	Score
Australia & New Zealand	australia	pub rock	0.0071
	australia	rock	0.0049
	new zealand	rock	0.0047
Eastern Asia	korea, republic of	k-pop	0.3306
	hong kong	cantopop	0.2172
	japan	j-rock	0.1209
Eastern Europe	bulgaria	pop	0.0072
	hungary	classical	0.0068
	hungary	opera	0.0060
Latin America & Caribbean	brazil	mpb	0.3191
	jamaica	rocksteady	0.0780
	colombia	vallenato	0.0675
Northern America	united states	southern hip hop	0.0351
	united states	east coast hip hop	0.0256
	united states	contemporary country	0.0080
Northern Europe	ireland	irish folk	0.0750
	denmark	dansktop	0.0732
	sweden	dansband	0.0197
South-eastern Asia	philippines	opm	0.2821
	viet nam	v-pop	0.2429
	viet nam	ballad	0.0128
Southern Asia	pakistan	ghazal	0.0980
	india	filmi	0.0836
	india	hindustani classical	0.0601
Southern Europe	portugal	fado	0.3243
	spain	flamenco	0.0177
	italy	italo-disco	0.0149
Sub-Saharan Africa	dr congo	soukous	0.3846
	south africa	kwaito	0.0794
	nigeria	r&b	0.0068
Western Asia	türkiye	anatolian rock	0.0977
	türkiye	arabesk	0.0752
	türkiye	pop	0.0050
Western Europe	netherlands	nederpop	0.0490
	france	chanson française	0.0269
	netherlands	levenslied	0.0163

G Comparison of Generated Regional Music with Reference Mainstream Genres

In Fig. 11 and Fig. 12 we compare generated music of 6 regional genres with reference music of the same genre, as well as the two genres with the highest counts in the dataset, pop and rock. The 6 regional genres were selected as difficult genres worth exploring, as they had the lowest mean FAD and KAD scores in Fig. 3.

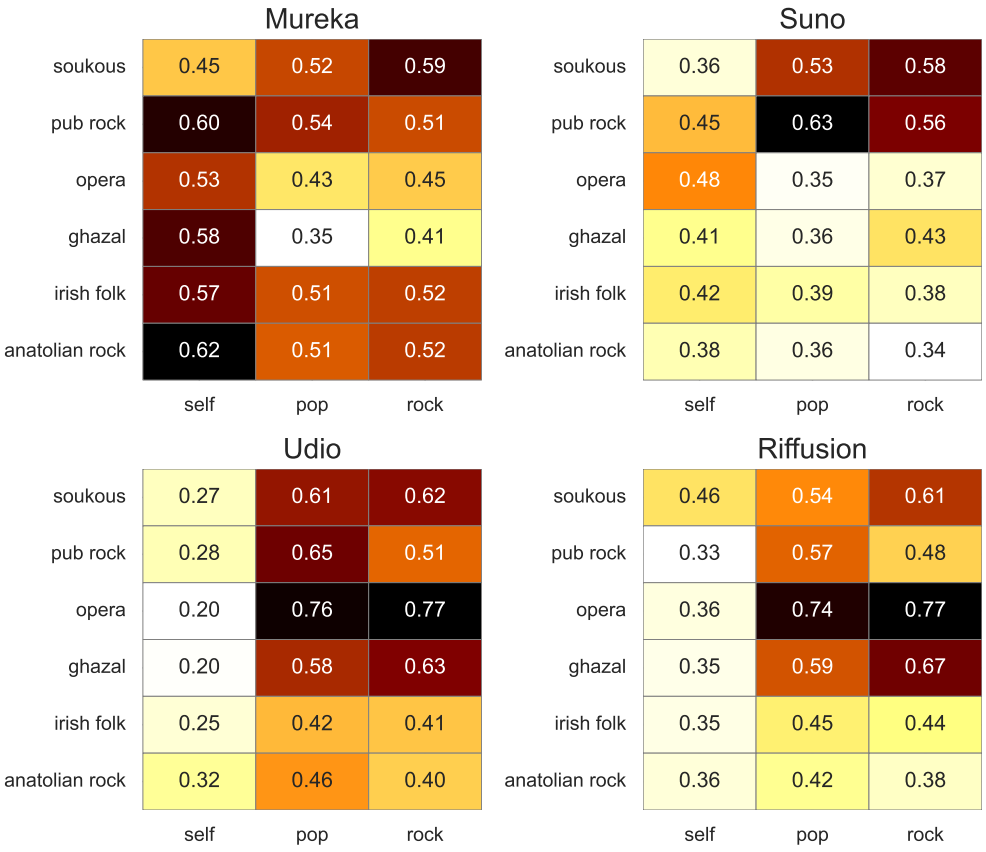

Figure 11: FAD Scores (lower is better) between generated regional genres (Y-axis) and reference genres (X-axis). Generated music is compared to reference music of the same genre (Self) as well as the top two genres by frequency in the dataset, pop and rock. All results are computed with the CLAP embedding model.

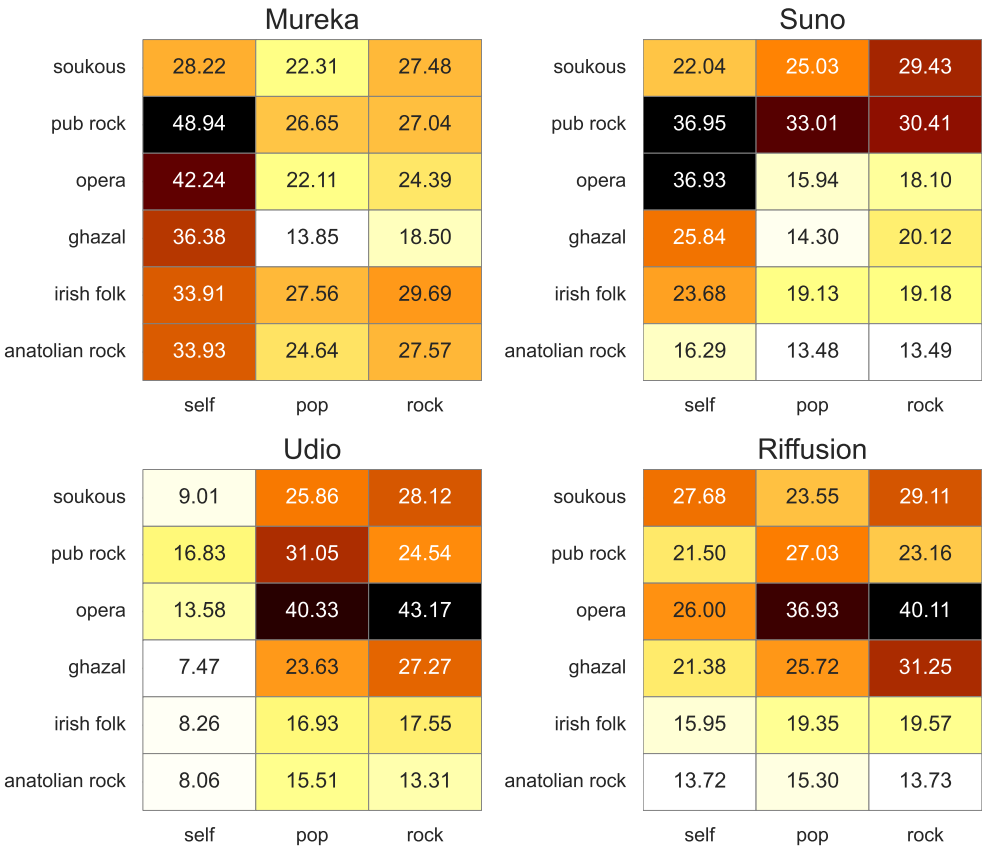

Figure 12: KAD Scores (lower is better) between generated regional genres (Y-axis) and reference genres (X-axis). Generated music is compared to reference music of the same genre (Self) as well as the top two genres by frequency in the dataset, pop and rock. All results are computed with the CLAP embedding model.