

Gender Bias and Sonic Gender Performance in AI-Generated Music: Representation, Expression, and Perception

Minyeong Bae¹, Hyunjae Kim², Kyung Myun Lee²

¹Ewha Womans University

²Korea Advanced Institute of Science and Technology (KAIST)
minyoung_bae@ewhain.net, present@kaist.ac.kr, kmlee2@kaist.ac.kr

Abstract

Gender bias in generative AI has typically been studied in text and image domains, but music-generating AI poses a different challenge: in music, gender can be expressed through sound, style, genre, and multimodal presentation as well as explicit representation. Through a case study of Suno, this study examines gender bias not only in vocal-gender representation, but also in generated musical expression and listener perception. First, generated vocal music shows a more extreme male-dominant pattern across genres than real-world music metadata. To further examine how gender is musically realized, we generated non-vocal instrumental music using gendered prompts. Gendered bias appeared through distributed stylistic cues, including genre descriptions, title words, cover images, audio features, and symbolic MIDI features. In genre-controlled generations, masculine-prompted outputs were often farther from neutral outputs than feminine-prompted outputs were, suggesting a stronger stylistic distinction for masculinity. Listeners' gendered impressions were shaped largely by mood and culturally familiar genre associations. Together, these findings suggest that gender bias in music-generating AI appears across multiple layers, involving how music and gender are repeatedly connected, expressed, and perceived throughout generation and listening.

Introduction

Generative AI is rapidly expanding into artistic creation, including music generation. Recent systems have lowered the barrier to music making and enabled users with little technical training to generate songs with ease. At the same time, however, AI systems trained on human-produced data may reflect, reproduce, or reinforce social biases embedded in existing cultural materials (Jain et al. 2022; Maluleke et al. 2022; Zhou et al. 2024; O'Connor and Liu 2024).

A growing body of work has shown that generative AI can exhibit gender, racial, and cultural bias. Existing research has focused predominantly on text- and image-based systems (O'Connor and Liu 2024; Naik and Nushi 2023; Nicolas and Caliskan 2024), while auditory domains remain comparatively underexplored (Agnew et al. 2025). Even within music-generating AI, recent studies have focused primarily on cultural and regional bias, such as the underrep-

resentation of non-Western music, representational imbalance across genres, and limits in cross-cultural adaptability (Mehta, Chauhan, and Choudhury 2024; Mehta et al. 2025; Bryan-Kinns and Li 2024; Solak et al. 2025), while gender-related bias has received little attention. This gap is important because music is not a neutral medium: voice, genre, style, and expression all carry culturally learned gender associations (Kim 2022; McClary 2002) that may be reproduced through AI-generated outputs.

In the sociotechnical process of AI music generation and reception, gender is not only a question of which vocal impressions appear more often, but also of how it is organized, expressed, and repeatedly performed through genre, musical style, mood, multimodal presentation, and listeners' responses, all of which are shaped by historically formed gender associations, stereotypes, conventions of gender performance, biased datasets, and unequal industry structures.

To empirically examine this perspective, we conduct a multi-part case study of the commercial music-generation system Suno, organized around four sequential questions. First, does the system tend to select different gendered vocal impressions across genres? Study 1 examines which gendered vocal impressions Suno tends to generate in response to genre prompts. Second, moving from how much each gender is represented to how gender is expressed, Study 2A investigates how gender-only prompts such as *feminine* and *masculine* are translated into musical, textual, and visual outputs without genre control. Third, do these gendered musical expressions persist when genre is controlled? Study 2B compares audio features, MIDI features, and foundation-model embeddings under genre-controlled conditions to test for asymmetric patterns of musical expression. Finally, how do human listeners perceive and interpret these generated gendered expressions? Study 2C examines how listeners perceive and interpret the outputs from Study 2B.

Across these studies, we report three main findings. First, male-perceived vocals were generated more frequently across genres, and AI-generated outputs exhibited a more extreme male-dominant pattern than the genre-level MusicBrainz reference. Second, gender-related prompts produced systematic gendered differences across musical features and accompanying non-musical content, including titles, descriptions, and cover images. These differences persisted under genre-controlled conditions, where masculine-

prompted outputs often showed stronger stylistic displacement from neutral outputs than feminine-prompted outputs did. Third, listeners interpreted the gender of AI-generated music through mood, perceived genre impressions, and culturally learned genre–gender associations.

These findings raise broader ethical and philosophical questions about AI gender performance. Since generative AI is likely to become more widely used in artistic and cultural production, future research should ask how such AI gender performance takes shape, whose gendered conventions it reproduces or transforms, and how it may reconfigure human gender perception and social understandings of gender. This connected process, from AI music generation to human listening, is therefore not only a site for observing gender bias, but also a case for examining how generative AI constructs, performs, and expresses gender, and how humans perceive it.

Related Work

Gender Bias in Generative AI Beyond Representation Counts

Recent research has shown that generative AI can reflect and reproduce social biases embedded in training data and deployment contexts. Gender bias has been widely documented in language (Gross 2023; O'Connor and Liu 2024; Prates, Avelar, and Lamb 2020; Hafner, Valdivia, and Rocher 2025) and image generation (Locke and Hodgdon 2025; Chauhan et al. 2024; Zhou et al. 2024; Gorska and Jemielniak 2023; Sun et al. 2024; Naik and Nushi 2023), often through unequal representation and stereotypical associations in generated outputs. Much of this work, however, has focused on representational imbalance and attribute co-occurrence.

Music-generating AI requires a broader analytic lens, because generated outputs may become associated with gender in multiple ways: through vocals, genre, musical expression, style, and multimodal elements such as textual descriptions, song covers, and lyrics. Existing work on AI music generation has primarily examined Western-centric training data and broader issues of cultural, regional, and genre representation (Mehta, Chauhan, and Choudhury 2024; Mehta et al. 2025; Bryan-Kinns and Li 2024), leaving gender-related patterns comparatively underexplored. Accordingly, it is necessary first to examine whether gender bias exists in music-generating AI, and further to investigate this issue from multiple perspectives.

Gendered Meaning in Music: Performance, Listening, and Interpretation

Genre-Gender Associations in Music Prior work in feminist musicology and sound studies has shown that gendered meaning in music is culturally constructed rather than inherent to sound itself (Kim 2022; McClary 2002; Werner 2019). Listeners interpret musical features through socially and historically shaped frameworks, and gendered perception depends not only on acoustic properties but also on expectation and context (Sergeant and Himonides 2014, 2016).

Genre plays a central role in this process, since it is associated not only with musical form but also with particular affects, identities, and stereotypes (Susino and Schubert 2019). As a result, genre and its associated musical characteristics can shape whether music is interpreted as masculine or feminine (Werner 2019; Sergeant and Himonides 2016). These findings suggest that AI-generated music may be interpreted through pre-existing genre–gender schemas rather than through sound alone.

Gendered Voice and Vocal Perception Voice is one of the most salient cues through which gender is invoked in music, yet vocal gender is not determined by biological factors alone. Rather, it is shaped by social norms, technological mediation, and cultural expectations (Kim 2022). Prior research has shown that voices are culturally associated with gendered social roles, affects, and values, including care, authority, and trust (Carson 1994; Wallis 2018; Phan 2017; Costa 2018; Weidman 2014). These findings suggest that vocal gender in AI-generated music should be understood not as a neutral acoustic property, but as a socially constructed category that is continuously produced through cultural interpretation and listening.

From Music Metadata and Industry Bias to Music-Generating AI Research

Gender bias in music is embedded not only in musical content, but also in the structures through which music is produced, distributed, and categorized, including metadata, industry organization (Wang and Horvát 2019), and platform recommendation (Epps-Darling, Bouyer, and Cramer 2020). These patterns matter for music-generating AI because such AI is trained on musical ecosystems in which gendered distributions and categorizations are already embedded. At the same time, music-generating AI may do more than simply inherit these imbalances: it may also recombine, amplify, or reshape them through prompt-conditioned generation and multimodal output.

This perspective can be extended through Butler's theory of gender performativity, which understands gender not as a fixed essence, but as something constituted through repeated acts, norms, and interpretive practices (Butler 2002, 2011). Recent work has applied this perspective to AI, arguing that bias is not only a matter of unequal representation, but also a process through which AI systems, through repeated patterns of gendered performance, reproduce and entrench gendered norms (Nino and Lisi 2024; Hafner, Valdivia, and Rocher 2025; Gross 2023). Building on this line of work, the present study uses gender performativity as an analytic lens for examining how AI-generated music organizes gendered associations across voice, sound, genre, text, image, and listening.

Study 1: Vocal-Gender Selection Across Genres

Stimulus Generation with Suno To examine gender bias in AI-generated music, we used Suno, a publicly accessible music generation platform that allows controlled prompt-based generation. We adopted the ten-genre taxonomy from

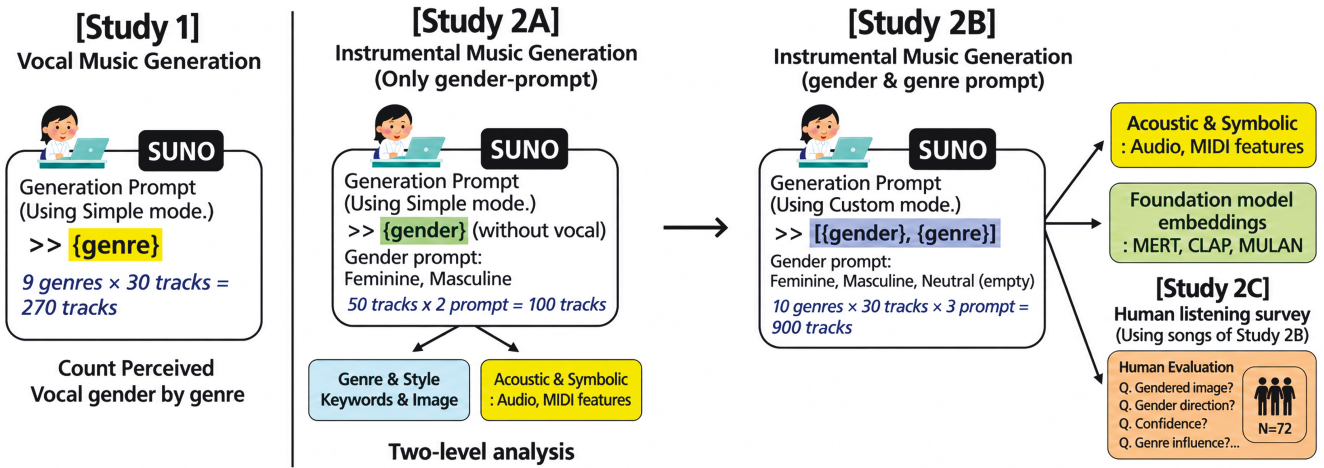


Figure 1: Overview of the study design and analytic pipeline. Study 1 analyzes vocal-gender selection in Suno-generated music across genre prompts. Study 2 examines instrumental generation under gender-only prompting (Study 2A), genre-controlled gender prompting (Study 2B), and human listening evaluation (Study 2C). Across these studies, we analyze gender in AI-generated music at the levels of vocal representation, musical and multimodal expression, and listener perception.

GTZAN (Tzanetakis and Cook 2002), a widely used benchmark dataset for music genre classification in music information retrieval, as a genre framework. The taxonomy includes blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. We excluded classical from Study 1 due to its limited relevance for vocal-gender analysis.

For each of the nine remaining genres, 30 tracks were generated using Suno v5 in Simple mode, resulting in a total of 270 tracks. Each generation used a prompt of the form “{genre}” with the vocal option enabled, without additional modifiers such as mood, artist references, instruments, or stylistic descriptors, in order to isolate genre-conditioned generation behavior. To reduce local repetition effects, the same prompt was not used more than three times consecutively (Figure 1).

To examine genre-dependent patterns in perceived vocal-gender presentation, we analyzed the distribution of gendered vocal impressions within each genre. The analysis focused on solo-vocal outputs; duet or group-vocal tracks were excluded. The first author listened to all tracks and classified each as male-presenting, female-presenting, or ambiguous based on the overall gender impression conveyed when the first clearly audible solo lead vocal entered, rather than on any single predefined acoustic feature. System-provided descriptions were used as supplementary information when available and were consistent with the listening judgments in all such cases. Because this was a single-coder analysis, inter-annotator agreement was not calculated. Ambiguous cases were excluded from the main comparative analysis.

MusicBrainz Reference Metadata To assess whether Suno’s vocal-gender selection bias reflected or amplified real-world genre patterns, we used MusicBrainz¹ metadata as a genre-level reference (Swartz 2002). Using the

same genre framework (excluding classical), we collected metadata for the nine genres from 1980 to 2025 and retained only person-type artists with male or female gender metadata, excluding groups and other non-person artist types.

The reference distribution was constructed at the unique-artist level. Because perceived vocal-gender presentation and artist-level gender metadata are not fully equivalent categories, MusicBrainz was used only as an approximate genre-level reference rather than a direct ground-truth counterpart.

Results: Vocal-Gender Imbalance Relative to a Metadata Reference According to the MusicBrainz-based reference distribution, male individual artists were generally more prevalent in all nine genres, while pop, disco, and jazz showed the highest female-artist proportions (Figure 2A). Suno showed the same broad genre pattern: female-presenting vocals appeared only in pop, disco, and jazz, whereas all 30 samples in each of the remaining six genres contained only male-presenting vocals (Figure 2B). These results suggest that Suno broadly follows the male-dominant genre-level reference pattern while further amplifying it in genres with lower female-artist representation.

Study 2A: Gender-Only Prompt Realization in Instrumental Music

Method

Generation Setup For Study 2A, we generated 50 instrumental tracks for each of the feminine and masculine prompt conditions in Suno v5 Simple Mode, yielding 100 tracks in total. Each generation used a prompt of the form “{gender}”, where {gender} was either feminine or masculine, with the instrumental option enabled and no additional genre, mood, artist, instrument, or stylistic mod-

¹<http://musicbrainz.org/>

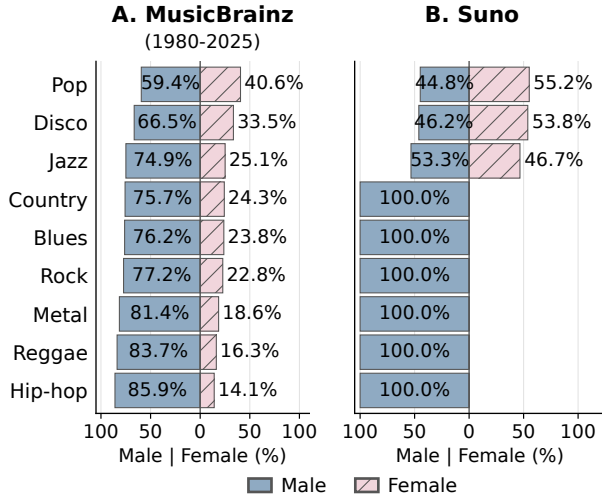


Figure 2: Comparison of MusicBrainz artist-gender metadata and perceived vocal-gender distributions in Suno-generated tracks across the nine GTZAN-based genres. **(A)** Individual-artist gender distributions from MusicBrainz metadata (1980–2025). **(B)** Perceived solo-vocal gender distributions in Suno-generated tracks for each genre. Male-female duets, group vocals, and tracks with unclear vocal gender were excluded from the Suno analysis.

ifiers. All generated tracks were used for audio and MIDI analyses without exclusion.

For non-musical modality analyses, including titles, genre/style descriptions, and cover images, we used a curated subset of 25 generations per prompt condition, because Suno produced two audio tracks with the same accompanying textual and visual information (Figure 1).

Musical Features

Audio Features We analyzed 84 track-level scalar audio variables derived from Essentia’s MusicExtractor outputs (Bogdanov et al. 2013). For descriptive interpretation, we organized them into five study-defined families: spectral shape and variation, energy and dynamics, rhythm and groove, pitch and tonality, and additional timbral descriptors (Figure 3A). The grouping was informed by Essentia descriptor definitions and the primary acoustic or musical property represented by each measure, rather than by an official Essentia taxonomy. Representative descriptors included spectral complexity, spectral flux, spectral spread, spectral centroid, spectral entropy, low-band energy, pitch salience, key- and chord-related measures, danceability, onset rate, and tempo-related statistics.

We then compared the feminine and masculine conditions feature by feature. For each descriptor, we computed Welch’s two-sample t -test and applied false discovery rate (FDR) correction for multiple comparisons. We additionally report Cohen’s d as an effect-size estimate and used Mann-Whitney tests as a robustness check. Figure 3 shows the significant features with the largest absolute effect sizes and

descriptively summarizes all significant features by family.

MIDI Features To obtain an approximate symbolic representation of the generated audio, we transcribed the tracks to MIDI using TransKun (Yan and Duan 2024), an automatic piano transcription model. We then extracted 225 symbolic features using jSymbolic (McKay, Cumming, and Fujinaga 2018), which broadly organizes MIDI features into pitch statistics, melody/horizontal intervals, chords/vertical intervals, rhythm, instrumentation, texture, and dynamics. Statistical comparisons followed the audio-feature procedure, and significant descriptors were summarized using these categories (Figure 3B).

Non-Musical Elements

Textual Descriptions: Title and Genre/Style For each instrumental track generated in Suno v5 Simple Mode, we collected the title and the short genre/style description automatically provided by the system alongside the audio output. From these descriptions, we extracted genre-related keywords and normalized them into comparable genre categories. We then counted the frequency of each genre category for each prompt condition in order to compare how Suno conceptualized genre in response to gender prompts.

Using the same descriptions from which genre labels were extracted, we additionally analyzed recurring instrumentation- and style-related expressions across the feminine and masculine conditions. Closely related lexical variants were grouped only when they pointed to the same instrument family or highly similar musical cue (e.g., guitar/guitars, bells/chimes, stomp/stomping). The keys/piano/chords category was treated as a harmonic-accompaniment category.

Song Cover Image We also examined cover images as part of the model’s broader multimodal framing of each output. Cover images were compared qualitatively rather than through formal feature extraction. For figure presentation, representative examples were selected within each prompt condition to reduce near-duplicate or visually overlapping motifs and to highlight broader differences in visual framing.

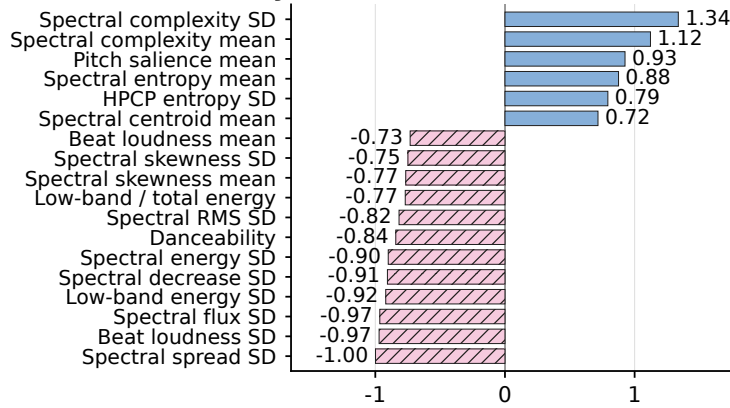
Results

Musical Analysis

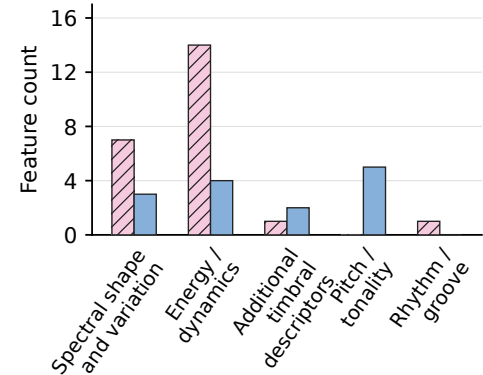
Audio Features Figure 3A shows the statistically significant audio features that differed between the feminine and masculine prompt conditions. Masculine tracks were higher in spectral complexity, pitch salience, spectral entropy, and spectral centroid, whereas feminine tracks showed higher danceability and beat loudness, a higher low-band-to-total energy ratio, and greater within-track variability in low-band energy, spectral flux, and spectral spread. These contrasts indicate significant differences in spectral profile, pitch prominence, low-band emphasis, and rhythmic-accent-related features across the two prompt conditions.

MIDI Features Figure 3B shows the statistically significant symbolic features that differed between the feminine and masculine prompt conditions. Masculine tracks showed

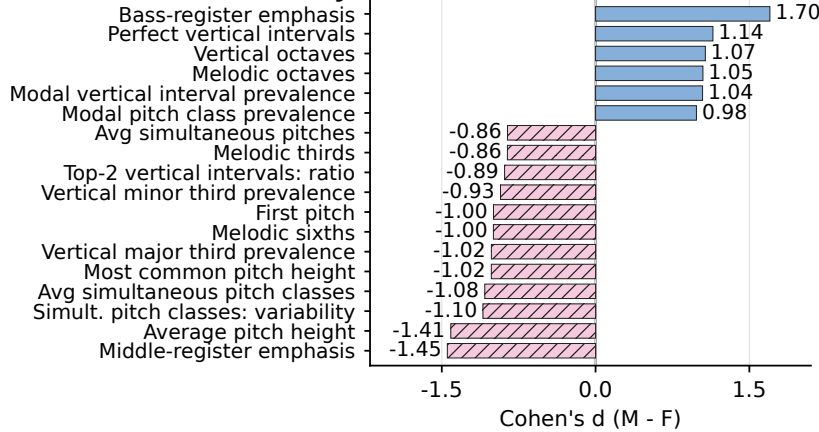
A. Audio features analysis



Audio Feature Family



B. MIDI features analysis



MIDI Feature Family

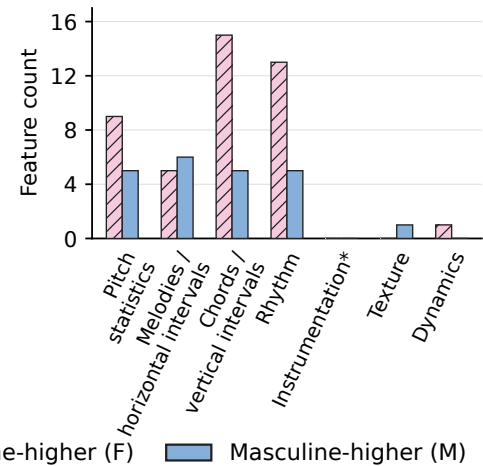


Figure 3: Significant acoustic and symbolic differences between feminine and masculine Suno outputs. (A) Audio variables derived from Essentia outputs; (B) symbolic MIDI descriptors analyzed with jSymbolic. Left panels show significant features with the largest absolute Cohen's d (*masculine* – *feminine*); positive values indicate masculine-higher features and negative values feminine-higher features. Right panels descriptively summarize significant features by category; Instrumentation* indicates that no descriptors in this category yielded valid Welch tests.

stronger bass-register emphasis and greater octave- and perfect-interval-based organization, while feminine tracks showed stronger middle-register emphasis, higher average pitch, and greater richness and variability in simultaneous pitch-class organization. In other words, the two prompt conditions differed significantly not only in which pitch ranges were emphasized, but also in how notes were distributed and combined in the MIDI-derived symbolic representations.

Textual Descriptions: Title and Genre/Style

Genre Discrimination As shown in Figure 4A, Suno's descriptions associated the feminine prompt primarily with a concentrated Pop-R&B frame, whereas the masculine prompt was distributed more broadly across Pop-rock, Pop/Alt-pop, and Alt-rock/Indie rock categories. In other words, feminine outputs were described within a relatively narrow genre frame, while masculine outputs were framed

through a more heterogeneous set of pop-, rock-, and alternative-oriented cues. This pattern suggests that gender prompting in Suno was accompanied by systematic genre framing rather than by a purely abstract stylistic distinction.

Title and Style-Description Keywords Title words differed clearly across the two prompt conditions. Feminine titles were more likely to evoke softness, beauty, and delicate materiality (e.g., silk, velvet, lipstick, pink, glass), whereas masculine titles more often invoked hardness, bodily action, and industrial materiality (e.g., iron, hands, steel, bones, collar) (Figure 4D). The keyword-level description analysis also revealed clear asymmetries in instrumentation and style framing. Feminine outputs were more often described through keys/piano/chords and bells/chimes, whereas masculine outputs were more often framed through guitar/guitars, palm-muted guitar, and rubbery bass (Figure 4C). In style terms, feminine tracks were more likely to be described

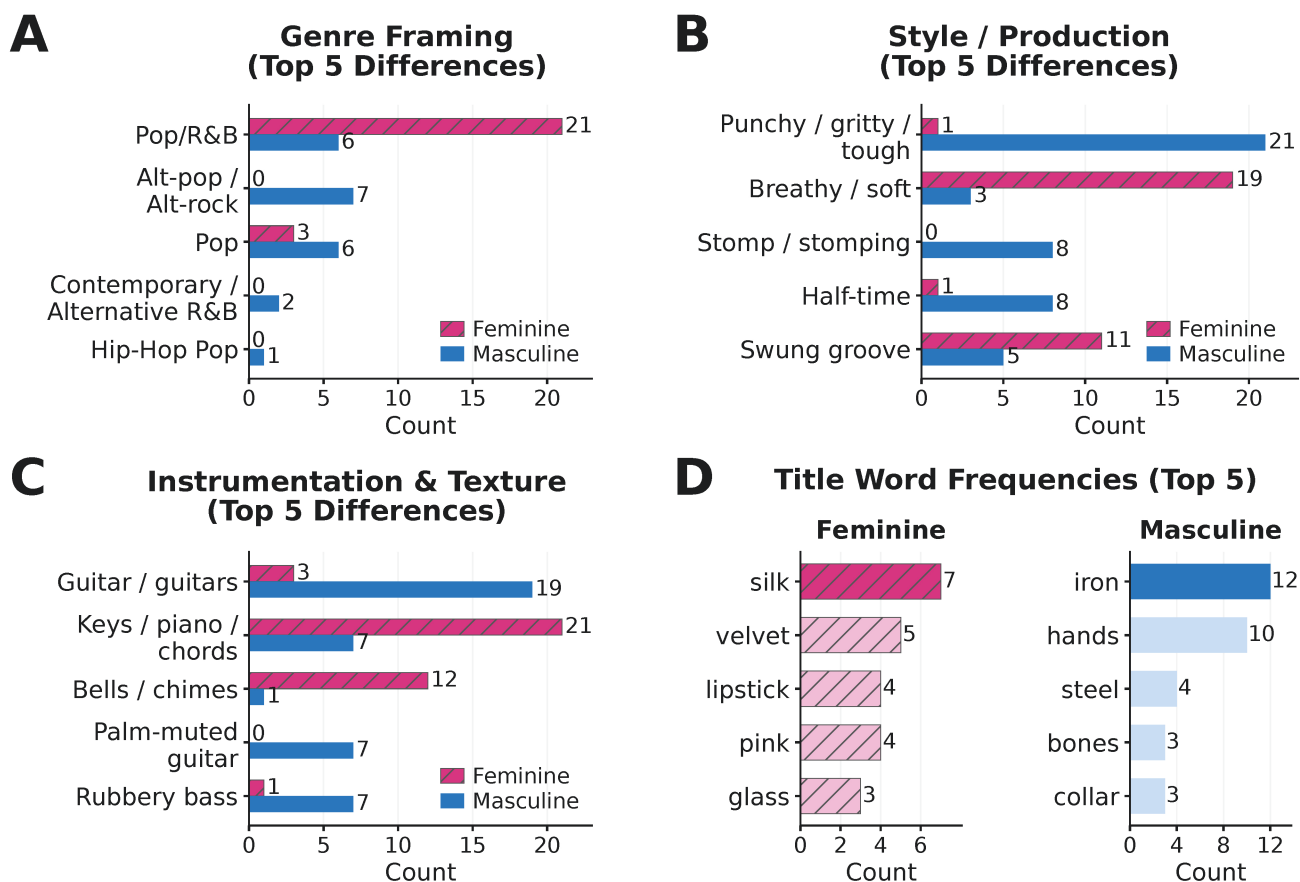


Figure 4: Textual framing differences between feminine and masculine Suno outputs. (A) shows genre-framing counts across Suno-provided descriptions. (B) and (C) show the style/production and instrumentation/texture descriptors with the largest absolute count differences between conditions. (D) shows the top title-word frequencies by prompt condition. All counts are based on document-frequency hits within the retained 25-track subset for each condition. Pink indicates feminine outputs and blue indicates masculine outputs.

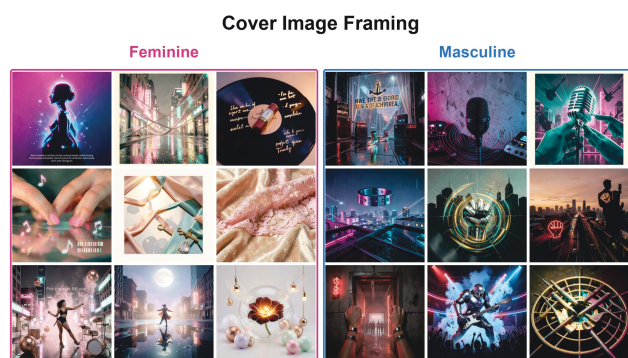


Figure 5: Suno-generated cover images for feminine and masculine instrumental tracks. Representative cover images are grouped by prompt condition, with feminine outputs on the left and masculine outputs on the right. Repeated or visually overlapping motifs were condensed into a smaller set of representative examples to highlight broader differences in visual framing across the two conditions.

as breathy/soft and swung groove, while masculine tracks were more often associated with punchy/gritty/tough, stomp/stomping, and half-time (Figure 4B). These patterns suggest that Suno differentiated the two prompt conditions not only through genre labels, but also through distinct instrumentation and production cues in its own descriptive language.

Image Suno’s automatically generated cover images also showed qualitative differences across prompt conditions. Feminine outputs were often paired with images highlighting conventionally feminized visual cues, such as close-up body imagery, beauty products including lipstick, soft fabrics such as velvet, and pink-dominant color palettes (Figure 5). In contrast, masculine outputs were more often paired with images featuring iron-related objects, machines or robotic forms, darker backgrounds, and depictions of male artists (Figure 5).

A few exceptions were also observed. Some masculine covers depicted steam irons, apparently interpreting the

word “iron” literally, while some feminine covers depicted male artists. Even so, the broader pattern indicates that Suno framed the two prompt conditions through distinct and stereotypically gendered visual imagery.

Study 2B: Gender-Genre Prompt Realization in Instrumental Music

Because the outputs in Study 2A were not genre-controlled, some of the observed differences may have reflected genre-linked conventions rather than gender-related prompting alone. Study 2B therefore introduces a genre-controlled instrumental setting to test whether gender-related prompts still produce systematic differences and asymmetries when genre is held constant.

We analyzed the generated music across three complementary representation spaces: audio features, symbolic MIDI features, and foundation-model embeddings. Audio features capture acoustic properties of the generated tracks, symbolic MIDI features provide approximate note-level representations, and foundation-model embeddings provide higher-level learned representations of musical and stylistic similarity. Using these representations, we conducted a distance-based analysis that compares how far feminine- and masculine-prompted outputs diverge from genre-matched neutral outputs. In this study, we do not treat these features as direct measures of gender. Instead, we use them to examine whether gender-related prompts produce systematic differences in musical expression under controlled genre conditions.

Method

Generation Setup Here, we used the same GTZAN-based ten-genre framework (Tzanetakis and Cook 2002) as in Study 1, with classical included. All tracks were generated using Suno v5 Advanced mode with prompts of the form “{gender}, {genre}”, while the neutral baseline used genre-only prompts. We generated 30 tracks for each genre-condition combination, yielding 900 tracks in total (10 genres $\times$ 3 conditions $\times$ 30 tracks).

Audio and MIDI Features For Study 2B, we applied the same audio- and symbolic-feature pipelines used in Study 2A to the genre-controlled instrumental outputs. We analyzed the same 84 scalar audio and track-level variables derived from Essentia’s MusicExtractor outputs (Bogdanov et al. 2013) and the same 225 symbolic features extracted with jSymbolic (McKay, Cumming, and Fujinaga 2018) from MIDI transcriptions produced with TransKun (Yan and Duan 2024). Dimensions that were constant across all 900 tracks were excluded, resulting in 83 audio and 175 MIDI dimensions for the distance-based analysis.

Foundation Model Embedding Analysis Because Suno is a closed commercial system, we cannot directly inspect the model representations that shape its outputs. As an indirect way to examine how gender-related structure may be reflected in learned audio representations, we analyzed the generated excerpts using embeddings from three representative audio and music foundation models: CLAP (Elizalde et al. 2023), MERT (Li et al. 2023), and MuLan (Huang et al.

2022). Unlike handcrafted acoustic descriptors, these embeddings capture higher-level regularities in audio, including semantic, stylistic, and musical information learned from large-scale data. Using multiple models allowed us to test whether the observed gender-related tendencies were consistent across learned audio representations. For each generated track, we extracted an embedding vector from each foundation model: 512, 1024, and 512 dimensions for CLAP, MERT, and MuLan, respectively.

Distance Metrics To quantify the relative displacement of feminine- and masculine-prompted outputs from the neutral baseline, we computed centroid distances from the neutral condition to each gendered condition for each genre and analysis domain. Audio and MIDI features were z-score standardized across all 900 tracks prior to centroid computation, whereas foundation-model embeddings were analyzed in their original embedding spaces. Centroids were computed by averaging the 30 tracks in each genre-condition group, and cosine distance was used for all representations.

The normalized contrast was defined as:

$$\Delta_g = \frac{d(N, M) - d(N, F)}{r(N)},$$

where N , M , and F denote the neutral, masculine, and feminine conditions, respectively; $d(\cdot, \cdot)$ denotes cosine distance between centroids; and $r(N)$ denotes within-neutral variability, computed as the mean cosine distance of neutral tracks from their centroid. Positive values indicate greater masculine displacement from the neutral baseline, whereas negative values indicate greater feminine displacement (Figure 6).

Statistical uncertainty was assessed using 1,000 bootstrap resamples within each genre-condition group, with centroids, neutral variability, and contrasts recomputed at each iteration. Pointwise 95% percentile confidence intervals were estimated from the resulting distributions. CLAP, MERT, and MuLan contrasts were combined using an equal-weight mean, and the overall mean was likewise averaged equally across genres; these averages were recomputed within each bootstrap iteration before estimating percentile confidence intervals.

For visualization only, t-SNE was applied to centroid cosine-distance matrices using the same preprocessing, with CLAP, MERT, and MuLan distance matrices equally averaged for the embedding panel.

Results

Audio Feature Space In the audio feature space, the results showed a predominantly positive but genre-dependent pattern. Positive contrasts with 95% bootstrap confidence intervals entirely above zero were observed for disco, jazz, blues, reggae, hip-hop, and classical, indicating greater masculine displacement from the neutral baseline in those genres. By contrast, rock showed a negative contrast with its confidence interval entirely below zero, indicating greater feminine displacement. Pop, country, and metal had positive contrasts, but their confidence intervals included zero. Across 1,000 bootstrap resamples, the proportion yielding

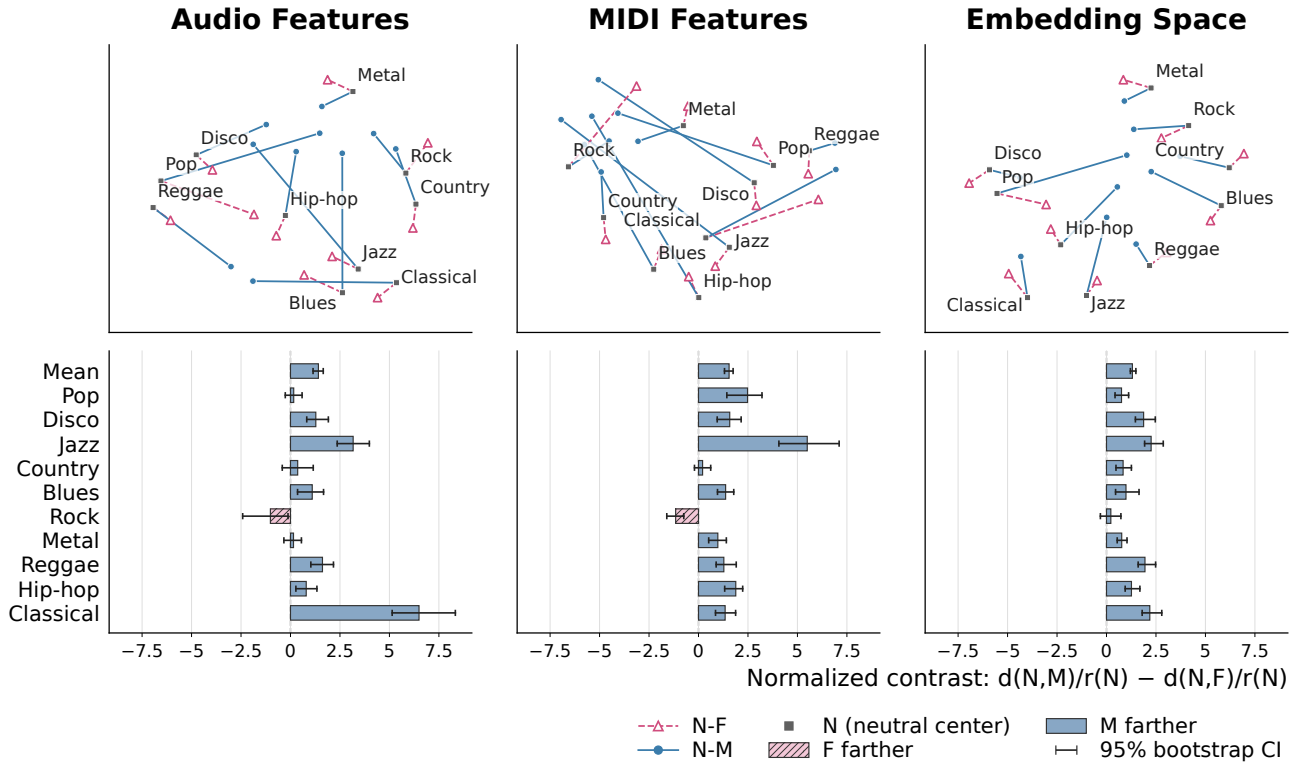


Figure 6: **Normalized Gender Contrast Relative to the Neutral Baseline.** The top row shows centroid-level t-SNE visualizations of neutral (N), feminine (F), and masculine (M) conditions for audio features, MIDI features, and foundation-model embeddings. Audio and MIDI representations were feature-wise standardized, whereas foundation-model embeddings were analyzed in their original spaces. Dashed and solid lines indicate N-F and N-M relations, respectively. The bottom row shows the normalized contrast, $d(N, M)/r(N) - d(N, F)/r(N)$, where $d(\cdot, \cdot)$ is cosine distance and $r(N)$ denotes within-neutral variability. Positive values indicate greater masculine displacement and negative values greater feminine displacement. Foundation-model contrasts were averaged equally across CLAP, MERT, and MuLan. Error bars show 95% bootstrap confidence intervals. t-SNE is shown for visualization only.

a positive contrast was 100% for disco, jazz, reggae, and classical, 99.8% for blues, and 99.9% for hip-hop. Overall, the audio analysis suggested greater masculine displacement across most genres, with substantial genre-specific variation and an opposite pattern in rock.

MIDI Feature Space We next examined symbolic MIDI features using the same normalized contrast measure (Figure 6). Consistent with the predominantly positive audio results, the MIDI feature space showed positive contrasts in nine of the ten genres, indicating that masculine-prompted outputs generally showed greater displacement from the neutral baseline than feminine-prompted outputs. Across 1,000 bootstrap resamples, the proportion yielding a positive contrast was 100% for disco, jazz, blues, metal, reggae, hip-hop, and classical, and 99.9% for pop. Country remained comparatively ambiguous (84.4%), with its confidence interval including zero, whereas rock showed greater feminine displacement (0% positive bootstrap contrasts), with its confidence interval entirely below zero. The centroid-level t-SNE visualizations are presented only as illustrative summaries;

the contrasts were calculated in the high-dimensional feature space.

Foundation Model Embeddings We finally examined foundation-model embedding spaces using raw embeddings, cosine distance, and the same normalized contrast measure (Figure 6). The genre-averaged contrast was positive for each of the three models, although the direction and magnitude varied across models and genres. In the equal-weight average across CLAP, MERT, and MuLan, contrasts were positive in all ten genres, with 95% bootstrap confidence intervals entirely above zero in nine of the ten genres. Rock was comparatively weak, with its confidence interval including zero. We therefore interpret the averaged representation panel as a descriptive summary of an overall tendency toward greater masculine displacement, while the accompanying t-SNE visualization is presented only as an illustrative view of the centroid relationships.

Study 2C: Human Listening Survey with Instrumental Music

Method

Participants To examine how gendered musical expression in AI-generated music was perceived by human listeners, we conducted an online survey-based listening experiment. A total of 72 Korean participants completed the survey. One participant was excluded from the quantitative analyses due to an irregularity in the recorded stimulus allocation, leaving 71 participants.

Stimuli and Survey Design Survey stimuli were drawn from the instrumental dataset used in Study 2B. For each genre-condition pair, two tracks were selected based on maximal pairwise cosine distance in embedding space, averaged across CLAP, MERT, and MuLan, in order to sample relatively distinct outputs within each pair. This yielded 60 stimuli in total (10 genres $\times$ 3 conditions $\times$ 2 tracks). Each stimulus was a 10-second excerpt from the beginning of the track. Participants were not given prompt or genre labels. Each participant listened to 30 excerpts drawn from five genres, with different genre subsets assigned across participants and stimuli presented in randomized order.

The survey included structured-response and open-ended items. Q1 asked whether a gendered impression immediately came to mind. Q1-1 asked participants to indicate the direction of an impression that arose, whereas Q1-2 asked those for whom none immediately arose to imagine one while listening; both used the same 5-point feminine-to-masculine scale. The responses were combined into a perceived gender-direction variable. Q2 measured confidence, Q3 asked which cue participants considered most influential in shaping their judgment, Q4 asked which genre was closest to the presented excerpt through fixed-response and free-text fields, and Q5 measured how strongly this genre impression influenced their judgment. Open-ended questions elicited participants' general criteria for judging gendered imagery and their views on whether music can have gendered qualities. The survey was designed to examine whether listeners' gendered impressions varied across gender-prompt conditions within the same genre, rather than to establish a consensus gender label for each excerpt. Accordingly, conventional inter-rater agreement was not calculated; analyses instead focused on response distributions across prompt conditions and genres.

Results

Quantitative Results Among 1,505 valid Q1 responses from 71 participants, "somewhat" was the most frequent response (62.7%), indicating that participants generally perceived some degree of gendered imagery. Gender-direction analyses included 1,109 valid responses contributed by 70 of these participants. When condition-level proportions were averaged equally across the three prompt conditions, gender-direction judgments leaned masculine: 62.5% fell on the masculine side, compared with 28.5% on the feminine side and 9.0% at the neutral midpoint. Although this pattern was less extreme than the male-dominant vocal representation observed in Study 1, it nevertheless indicates

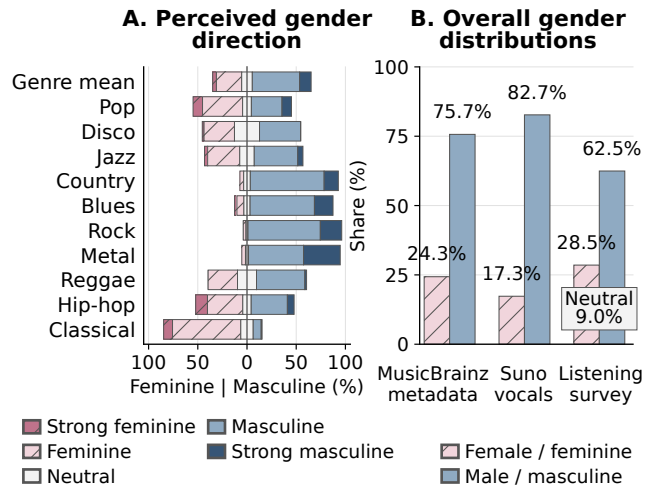


Figure 7: Survey-based gender judgments and comparison with MusicBrainz artist-gender metadata and Suno vocal-gender distributions. **(A)** Genre-wise distribution of perceived gender impressions (Q1-1/Q1-2). **(B)** Overall gender distributions in MusicBrainz metadata, Suno-generated vocals, and survey judgments of instrumental tracks, with survey percentages macro-averaged across the three prompt conditions.

a masculine-leaning tendency in listeners' gender perceptions (Figure 7B). Perceived gender direction also varied by genre: rock and metal were strongly masculine-leaning (95.2% and 93.2%, respectively), whereas classical was clearly feminine-leaning (78.4%; Figure 7A).

In Q3, participants most often selected mood and genre feel as the cues that shaped their judgments. This genre-based interpretation was also reflected in the Q4 free-text responses. Among stimuli whose actual genre was not rock, rock impressions occurred more often under the masculine prompt condition than under the neutral or feminine conditions. Conversely, among stimuli whose actual genre was not pop, pop impressions occurred more often under the feminine condition than under the neutral or masculine conditions. Participants also reported that genre impression played a meaningful role in shaping their judgments, with option 4 ("some influence") as the most frequent response (48.4%).

Qualitative Feedback from Listeners The first author reviewed open-ended responses at the participant level and inductively coded recurring, non-mutually exclusive themes, allowing multiple themes per participant. As this was a single-coder analysis, intercoder agreement was not calculated. The survey examined how perceived gender varied across prompt conditions and genres rather than seeking consensus labels for individual excerpts; structured responses were therefore analyzed as response distributions rather than through inter-rater agreement.

The coded responses included cultural learning or bias (about 40% of all participants), mood or affect (about 35%),

and genre associations. Together with the cue and genre responses, these findings suggest that participants interpreted the excerpts through culturally familiar genre-based schemas. This highlights the role of cultural interpretation in listeners' perceptions of AI-generated music.

Discussion

This study does not seek to define what “women’s music” or “men’s music” is, nor to determine what music is inherently feminine or masculine. Instead, our aim is to examine how a music-generating AI system associates femininity and masculinity with particular vocal impressions, musical features, styles, genres, textual descriptions, and visual elements. Grounded in feminist and philosophical attempts to move away from essentialist understandings of gender, and particularly in Butler’s theory of gender performativity (Butler 2002), we approach gender as constituted through repeated acts, conventions, and interpretive practices rather than as something that preexists its expression. From this perspective, the process from input prompts through AI music generation to human listening can be understood as a site of AI sonic gender performance. The studies examine three connected stages of this process: vocal representation, instrumental and multimodal expression, and human interpretation.

Study 1 examined the first stage of this process: how gendered vocal representation varied across genres. Male-presenting vocal impressions were generated more frequently across several genres, and this male-dominant tendency was stronger than that observed in the genre-level MusicBrainz reference metadata. These findings suggest that music-generating AI may amplify, rather than merely reflect, existing representational patterns by repeatedly associating particular genres with male-presenting voices.

Shifting from vocal representation to instrumental expression, Study 2 examined how Suno expressed femininity and masculinity when gender terms were used as explicit prompt conditions. In Study 2A, gender-only prompts produced systematic differences in genre tendencies, musical features, style keywords, titles, and cover images. These differences persisted under genre control in Study 2B: across audio, MIDI, and foundation-model representations, masculine-prompted outputs generally showed greater displacement from neutral outputs than feminine-prompted outputs did, although the magnitude and direction varied across genres. This pattern suggests an asymmetry in how masculine and feminine prompts were represented in the generated music.

Study 2C then examined how these outputs were perceived by human listeners. Listeners’ gender judgments relied primarily on mood and genre impressions, with rhythm and tempo also contributing to their judgments. These responses suggest that listeners drew on culturally familiar genre associations, particularly between rock and masculinity and between pop and femininity, even when the stimuli belonged to other genres.

Taken together, these findings show that the gendering of AI-generated music unfolds across three linked processes: representational allocation in vocal generation, stylistic and

multimodal expression in instrumental generation, and cultural interpretation in human listening. Across these processes, gender bias emerges as a distributed pattern spanning vocal presentation, musical features and representations, textual and visual framing, and listener perception, rather than as an imbalance confined to a single type of output.

This distributed form of bias becomes particularly important as music-generating AI is more widely used, because it may extend beyond unequal outputs to shape how gender is repeatedly performed through music. Moreover, because numerous artists have long used music to disturb, subvert, and freely express gender norms (Geffen 2020), we should consider whether music-generating AI can also enable diverse and experimental forms of gender performance, or whether it instead further confines the relationship between gender and music to a narrow range of familiar and recognizable conventions.

Finally, the role of culturally learned genre associations in Study 2C highlights how AI-generated gendered expression is mediated through human interpretation. This raises broader questions about whose gendered conventions generative AI learns and expresses, how these performances interact with social norms, and how they may reshape the ways humans hear, see, imagine, and perform gender. Music-generating AI thus provides a case through which to raise broader ethical and philosophical questions about gender performance and bias across generative AI systems.

Limitations

This study examines a single, closed, and evolving commercial system, Suno, and therefore cannot generalize its findings to other models or determine whether the observed patterns arise from its training data, prompt-processing mechanisms, or other components of the generation pipeline. Future work should compare systems and versions to distinguish shared from model-specific patterns of gender association and bias. The use of deliberately minimal prompts enabled controlled comparisons, but such prompts do not fully reflect real-world music-generation practices, where users typically combine genre, mood, style, and contextual descriptors. The findings may therefore have limited validity in more realistic prompting contexts. The MIDI analyses also rely on a piano-transcription model applied to multi-instrument audio and may therefore reflect transcription errors or domain mismatch.

The perceptual study was limited by 10-second excerpts and a relatively homogeneous Korean listener sample, while its English-only prompts, Western-centered genres, and MusicBrainz metadata may reflect cultural and representational biases. Future work should examine multilingual prompts, diverse musical traditions, and listeners across regions.

Ethical Statement

The listening survey was voluntary, and personally identifying information was excluded from the analysis. Although feminine and masculine prompts were used to examine dominant gender associations, these binary categories

were treated as analytic prompts rather than as essential properties of music or identity. Gender identities and expressions beyond this binary were outside the scope of the present analysis, but their exclusion should not be understood as denying their existence or relevance. Future research should examine how music-generating AI represents and expresses nonbinary, transgender, and other queer experiences of gender. Similarly, figure colors distinguish analytic categories only and do not imply gender-essentialist associations. Generative AI assisted with code generation and minor wording edits.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2023R1A2C100475512).

References

- Agnew, W.; Barnett, J.; Chu, A.; Hong, R.; Feffer, M.; Net-zorg, R.; Jiang, H. H.; Awumey, E.; and Das, S. 2025. Sound Check: Auditing Recent Audio Dataset Practices. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 8, 26–40.
- Bogdanov, D.; Wack, N.; Gómez, E.; Gulati, S.; Herrera, P.; Mayor, O.; Roma, G.; Salamon, J.; Zapata, J. R.; and Serra, X. 2013. ESSENTIA: An Audio Analysis Library for Music Information Retrieval. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 493–498.
- Bryan-Kinns, N.; and Li, Z. 2024. Reducing Barriers to the Use of Marginalised Music Genres in AI. *arXiv preprint arXiv:2407.13439*.
- Butler, J. 2002. *Gender trouble*. routledge.
- Butler, J. 2011. *Bodies that matter: On the discursive limits of sex*. routledge.
- Carson, A. 1994. The gender of sound. *Thamyris*, 1: 10–31.
- Chauhan, A.; Anand, T.; Jauhari, T.; Shah, A.; Singh, R.; Rajaram, A.; and Vanga, R. 2024. Identifying race and gender bias in stable diffusion ai image generation. In *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*, 1–6. IEEE.
- Costa, P. 2018. Conversing with personal digital assistants: On gender and artificial intelligence. *Journal of Science and Technology of the Arts*, 10(3): 59–72.
- Elizalde, B.; Deshmukh, S.; Al Ismail, M.; and Wang, H. 2023. CLAP:learning audio concepts from natural language supervision. In *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.
- Epps-Darling, A.; Bouyer, R. T.; and Cramer, H. 2020. Artist gender representation in music streaming. In *ISMIR*, 248–254.
- Geffen, S. 2020. *Glitter up the dark: How pop music broke the binary*. University of Texas Press.
- Gorska, A. M.; and Jemielniak, D. 2023. The invisible women: uncovering gender bias in AI-generated images of professionals. *Feminist Media Studies*, 23(8): 4370–4375.
- Gross, N. 2023. What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sciences*, 12(8): 435.
- Hafner, F. S.; Valdivia, A.; and Rocher, L. 2025. Gender Trouble in Language Models: An Empirical Audit Guided by Gender Performativity Theory. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 1677–1695.
- Huang, Q.; Jansen, A.; Lee, J.; Ganti, R.; Li, J. Y.; and Ellis, D. P. 2022. MuLan: A joint embedding of music audio and natural language. *arXiv preprint arXiv:2208.12415*.
- Jain, N.; Olmo, A.; Sengupta, S.; Manikonda, L.; and Kambhampati, S. 2022. Imperfect ImaGANation: Implications of GANs exacerbating biases on facial data augmentation and snapchat face lenses. *Artificial Intelligence*, 304: 103652.
- Kim, K. H. 2022. Sound and Gender: Gendered Aspects of Sound, Music, Cultural Practice and Discourse. *Eumak Nondan*, 47: 27–56.
- Li, Y.; Yuan, R.; Zhang, G.; Ma, Y.; Chen, X.; Yin, H.; Xiao, C.; Lin, C.; Ragni, A.; Benetos, E.; et al. 2023. MERT: Acoustic music understanding model with large-scale self-supervised training. *arXiv preprint arXiv:2306.00107*.
- Locke, L. G.; and Hodgdon, G. 2025. Gender bias in visual generative artificial intelligence systems and the socialization of AI. *AI & SOCIETY*, 40(4): 2229–2236.
- Maluleke, V. H.; Thakkar, N.; Brooks, T.; Weber, E.; Darrell, T.; Efros, A. A.; Kanazawa, A.; and Guillory, D. 2022. Studying bias in gans through the lens of race. In *European Conference on Computer Vision*, 344–360. Springer.
- McClary, S. 2002. *Feminine endings: Music, gender, and sexuality*. U of Minnesota Press.
- McKay, C.; Cumming, J.; and Fujinaga, I. 2018. JSYMBOLIC 2.2: Extracting Features from Symbolic Music for use in Musicological and MIR Research. In *ISMIR*, 348–354.
- Mehta, A.; Chauhan, S.; and Choudhury, M. 2024. Missing Melodies: AI Music Generation and its “Nearly” Complete Omission of the Global South. *arXiv preprint arXiv:2412.04100*.
- Mehta, A.; Chauhan, S.; Djanibekov, A.; Kulkarni, A.; Xia, G.; and Choudhury, M. 2025. Music for all: representational bias and cross-cultural adaptability of music generation models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, 4569–4585.
- Naik, R.; and Nushi, B. 2023. Social biases through the text-to-image generation lens. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 786–808.
- Nicolas, G.; and Caliskan, A. 2024. Directionality and representativeness are differentiable components of stereotypes in large language models. *PNAS nexus*, 3(11): pgae493.
- Nino, G.; and Lisi, F. A. 2024. Exploring the Question of Bias in AI through a Gender Performative Approach. *Sexuality and Gender Studies Journal*, 2(2): 14–31.
- O’Connor, S.; and Liu, H. 2024. Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. *AI & SOCIETY*, 39(4): 2045–2057.

- Phan, T. 2017. The materiality of the digital and the gendered voice of Siri. *Transformations*, 29: 23–33.
- Prates, M. O.; Avelar, P. H.; and Lamb, L. C. 2020. Assessing gender bias in machine translation: a case study with google translate. *Neural Computing and Applications*, 32(10): 6363–6381.
- Sergeant, D. C.; and Himonides, E. 2014. Gender and the performance of music. *Frontiers in psychology*, 5: 276.
- Sergeant, D. C.; and Himonides, E. 2016. Gender and music composition: A study of music, and the gendering of meanings. *Frontiers in psychology*, 7: 411.
- Solak, A.; Grötschla, F.; Lanzendörfer, L. A.; and Wattenhofer, R. 2025. Bias beyond Borders: Global Inequalities in AI-Generated Music. *arXiv preprint arXiv:2510.01963*.
- Sun, L.; Wei, M.; Sun, Y.; Suh, Y. J.; Shen, L.; and Yang, S. 2024. Smiling women pitching down: auditing representational and presentational gender biases in image-generative AI. *Journal of Computer-Mediated Communication*, 29(1): zmad045.
- Susino, M.; and Schubert, E. 2019. Cultural stereotyping of emotional responses to music genre. *Psychology of Music*, 47(3): 342–357.
- Swartz, A. 2002. MusicBrainz: a semantic Web service. *IEEE Intelligent Systems*, 17(1): 76–77.
- Tzanetakis, G.; and Cook, P. 2002. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5): 293–302.
- Wallis, C. 2018. Gender and the Telephonic Voice. In *The Routledge Companion to Sound Studies*, 329–337. Routledge.
- Wang, Y.; and Horvát, E.-Á. 2019. Gender differences in the global music industry: Evidence from musicbrainz and the echo nest. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 517–526.
- Weidman, A. 2014. Anthropology and voice. *Annual Review of Anthropology*, 43(1): 37–51.
- Werner, A. 2019. What does gender have to do with music, anyway?: Mapping the relation between music and gender. *Per Musi*, (39): 1–11.
- Yan, Y.; and Duan, Z. 2024. Scoring time intervals using non-hierarchical transformer for automatic piano transcription. *arXiv preprint arXiv:2404.09466*.
- Zhou, M.; Abhishek, V.; Derdenger, T.; Kim, J.; and Srinivasan, K. 2024. Bias in generative AI. *arXiv preprint arXiv:2403.02726*.