Between Genres, Before Genre

Fabian S. Klinke

2026-08-29

Genre labels as musical controls in Udio

Udio's prompting guide offers `eurodance + black metal` as an example. Two long musical histories appear as controls that users can combine (Udio, n.d.). The prompt does not explain which traits Udio takes from either genre or how listeners will describe the result. Sangheon Park and Claire Arthur's study of 200 Udio songs reveals part of this gap: genre terms appear in about 95 per cent of the prompts and about 61 per cent of the descriptions written after people hear the clips (Park & Arthur, 2026).

The prompt seems to offer a genre between two genres: Udio can produce a sound whose traits point towards several known categories. A genre begins when people accept a recurring set of traits as rules for later music. The provisional term *pre-generic* describes the evidence before that point and makes no prediction about the sound's future. Choosing labels gives users *semantic control* within Udio's *structural control* over musical effects and circulation. The platform can make a musical difference repeatable and visible before musicians and listeners decide what it means. Calling the first unfamiliar output a genre credits the model with shared cultural work and hides the platform's control over the conditions in which genres form.

Genre as a musical control

In Udio, a genre name shapes the music before any sound exists. The guide places genre beside mood, tempo and instrumentation, then encourages users to combine established labels (Udio, n.d.). A genre term moves through the system twice: it first guides generation, then returns as a description or sorting label. Users bring names formed through musical culture, which Udio turns into sound through training data, a learned representation and a generation process that remain hidden. An *affordance* is a possibility for action that a system offers a user. Udio's prompt field makes named genre combinations legible and easy while hiding the relation between each word and its musical effect.

The model stores statistical relations among words and sounds in a *latent space* whose shape follows the material and categories preserved during training (Coelho, 2026). Its opacity forces users to learn its behaviour by prompting, listening and trying again. Thor Magnusson calls digital instruments epistemic tools: their categories and controls contain a theory of music (Magnusson, 2009). Udio turns this theory into a usable interface by treating genre names as units that users can combine without knowing their exact musical effects.

This hidden mapping demands a specific kind of musicianship. Users choose terms, judge their sonic effects and revise the instruction, often through many generations, extensions and rejections (Coelho, 2026). The course frames this as a shift from execution to mediation: users guide a system through language, then direct and select among its proposals instead of specifying every musical event. It calls this *semiotic labour*, a form of musical skill built from signs and judgement. Language gives users access to a mapping the company controls.

Fabbri's genre theory shows what the interface cannot create by itself. He defines a genre as a set of real or possible musical events governed by socially accepted rules, and one event may belong to several genres at once (Fabbri, 1982). Formal and technical rules shape the sound. Meaning, conduct, social identity, economic arrangements and law sustain the genre as a shared practice. Udio can recombine audible conventions from these histories, though an output alone contains only part of the system that gives them force. The same strange mixture may still belong to several existing genres.

Prompt, perception and genre slippage

From prompt to perception

Park and Arthur make the change from words to sound visible without claiming access to Udio's internal model. They pair 200 English-language prompts from public Udio songs with the first 32 seconds of each song. Their 148 listeners write 2,624 descriptions, which the researchers sort into seven categories (Park & Arthur, 2026). Genre appears in about 95 per cent of the prompts and about 61 per cent of the descriptions. Once listeners describe the clips, instrumentation becomes the largest category. Genre carries much of the instruction at the interface, then becomes one part of a wider account of the sound.

The listeners' words show what carries over. Park and Arthur describe a group of related meanings in which a metal prompt may return as rock or punk, or as electric guitar and screaming. Genre produces the strongest match between prompt and description, with story-heavy prompts at the opposite end (Park & Arthur, 2026). A single genre label can suggest instrumentation, rhythm, vocal style and production. It therefore sets clearer expectations of sound than an imagined scene. Listeners hear these traits separately and rebuild the category from them. The course calls this movement *intersemiotic mediation*:

translation between verbal and nonverbal systems of meaning. Prompt language becomes sound and returns as listener language. The category persists even though the words do not repeat exactly.

Two stages of selection narrow the claim: users had already chosen the songs for public sharing, and the researchers removed low-quality or non-musical clips. Mixed prompts and listener recognition of a new category fall outside the study. Within this sample, listeners often hear signs linked to the genre named in the prompt.

Mixed prompts and unstable sound

Coelho addresses mixed prompts through a study of his own practice. After making hundreds of Udio outputs, he selects examples that create a repeated sense of "almostness." One sequence starts with `john oswald plunderphonics experimental, deconstructed, classical music, stockhausen, no vocals` and extends its first 32 seconds several times with the same prompt (Coelho, 2026). He calls the relation among these inseparable terms a "prompt assemblage." His phrase *genre slippage* describes musical signs that appear and fade before one identity settles.

Coelho hears piano textures move towards accordion-like timbres. Brief traces of electric guitar dissolve into woodwind sounds, and MIDI-like passages fade into granular textures (Coelho, 2026). These identities overlap and bleed into each other. The sound keeps referring to familiar instruments and genres without settling into one of them. Coelho maintains continuity across hundreds of possible variations by choosing the terms and deciding which extensions deserve attention.

The method sets a firm limit. Coelho chose the examples because they produced the experience his theory describes, and he listened with the prompts visible. The paper

documents what one trained listener heard. It cannot show how often genre slippage occurs or whether other listeners would hear the same instability.

In an intervention study of MusicGen, changing an instrument works more reliably than changing a genre, and a change aimed at one concept can alter unrelated parts of the audio (Vélez Vásquez et al., 2024). In that model, genre consists of linked musical features that resist separate control. Udio's proprietary model remains outside the experiment. Here, *between* names a relation that listeners hear among labels and sounds. No study has found a geometric midpoint inside Udio's latent space.

Genre formation under platform control

From repeated difference to genre

Fabbri argues that new genres begin inside existing musical systems when a practice breaks an accepted rule and a community adopts the event as a model for later music (Fabbri, 1982). Pre-generic describes the evidence before this acceptance. It includes the sound's cultural history and marks the point at which an audible difference exists while its later social use remains unknown.

Musicians give the difference social force when they carry it into new works. Listeners form shared expectations around its name, while critics or organisers may connect the sound to a scene. Lena and Peterson studied a nonrepresentative sample of sixty commercial genres in the twentieth-century United States, including genres first named by the music industry. Every genre in their sample later passed through a scene phase (Lena & Peterson, 2008). A name can arrive before a scene. Udio could spread a recurring sound under a shared label before musicians organise around it. The platform would take part

in genre formation by making the proposed difference public and visible. Musicians and listeners would still have to make that difference durable through use.

Platform control and circulation

By making songs and organising their public display in the same service, Udio gains some control over later use. Poell, Nieborg and van Dijck use *platformisation* for the way platforms reshape culture through their infrastructure, markets and rules (Poell et al., 2019). The course's concept of the *dispositif* broadens this account by joining the model and interface to company rules, economic incentives, public discourse and user practice. This dispositif shapes what people can make and how their work gains value. Udio's interface defines the available controls and terms of access. Its public pages turn prompts and tags into labels for publishing and finding songs. Casini and colleagues collected 101,953 public Suno and Udio songs over six months in 2024. Genres, qualifiers and instruments dominated their tags (Casini et al., 2025). Their sample contains only songs that users chose to publish. A genre label first helps to generate a song and then helps other people to find it.

Mads Krogh's account of Spotify shows why a platform's categories have cultural power. Spotify turns genres into numerical distances in a large map of musical similarity, then gives those distances practical force through search and recommendation (Krogh, 2023). Krogh studies recommendation, so his account does not describe Udio's generator. Text-to-music places the category earlier in the process: genre shapes a work before it exists and later organises its public life. Udio joins generation, labelling and circulation in one interface whose tags and public pages affect which proposed differences are easy to find and discuss.

Unequal musical vocabularies

Platform power begins with the words that work as controls. Udio's English guide gives genre terms this role (Udio, n.d.). Their musical effects depend on what the model has learned from its data and metadata. In the music-generation datasets that Mehta and colleagues could classify, non-Western genres made up 5.7 per cent of the total hours. They also found uneven results when adapting MusicGen and Mustango to Hindustani classical and Turkish makam music (Mehta et al., 2025). Their study shows unequal representation across the field and provides no evidence about Udio's private training data. Uneven representation can make some genre labels easier for a model to translate into specific musical traits. A limited vocabulary may then appear to be a universal language for musical control.

Georgina Born describes musical action as something spread across many people and tools (Born, 2005). Within this account, a genre label holds histories made by musicians, scenes, media and conflict. Names such as black metal or techno refer to practices that go far beyond their sound (Fabbri, 1982). Udio depends on those histories, on choices made by developers, on the user's prompt and on the listener's judgement. By owning the interface and hiding the link between each word and its musical effect, the company turns a shared cultural vocabulary into a private control system. This arrangement concentrates power over which histories users can combine and which differences the platform can make available for later genre formation.

Before genre

The result of `eurodance + black metal` may be strange and coherent without becoming a genre. Udio has generated an event from overlapping musical histories. Genre begins

later, when people treat a recurring difference as a rule for further music. Pre-generic names this opening stage of a social process and leaves the model's internal geography unspecified.

Generative systems change where that process begins. At the micro level, Udio's affordances turn genre words into musical instructions and make prompting a practice of mediation and selection. At the macro level, the interface belongs to a dispositif of private infrastructure, publication and cultural power. Genre labels have long helped describe and circulate music (Fabbri, 1982; Krogh, 2023). In Udio they also act before composition as instructions. Prompt writers learn to hear and revise their ideas through categories whose musical effects they cannot inspect. Musicians and listeners give a repeated difference its name and meaning through continued use, under conditions of access and visibility that the platform sets. Generative systems can produce many sounds that might become genres. The range of candidates already bears the limits of the data, vocabulary and interface through which they were made.

References

Born, G. (2005). On musical mediation: Ontology, technology and creativity. *Twentieth-Century Music*, *2*(1), 7–36. https://doi.org/10.1017/S147857220500023X

Casini, L., Cros Vila, L., Dalmazzo, D., Kaila, A.-K., & Sturm, B. L. T. (2025). *Data-driven analysis of text-conditioned AI-generated music: A case study with suno and udio*. https://doi.org/10.48550/arXiv.2509.11824

Coelho, G. (2026). Latent music: Emergent sonic forms and sonic liminality in text-to-audio systems. *Proceedings of the International Computer Music Conference*, 122–129. https://icmc2026.ligeti-zentrum.de/wp-content/uploads/2026/05/ICMC2026_proceedings_V2.pdf

Fabbri, F. (1982). A theory of musical genres: Two applications. In D. Horn & P. Tagg (Eds.), *Popular music perspectives* (pp. 52–81). International Association for the Study of Popular Music. https://www.francofabbri.net/files/Testi_per_Studenti/ffabbri81a.pdf

Krogh, M. (2023). Rampant abstraction as a strategy of singularization: Genre on spotify. *Cultural Sociology*, *19*(1), 89–107. https://doi.org/10.1177/17499755231172828

Lena, J. C., & Peterson, R. A. (2008). Classification as culture: Types and trajectories of music genres. *American Sociological Review*, *73*(5), 697–718. https://doi.org/10.1177/000312240807300501

Magnusson, T. (2009). Of epistemic tools: Musical instruments as cognitive extensions. *Organised Sound*, *14*(2), 168–176. https://doi.org/10.1017/S1355771809000272

Mehta, A., Chauhan, S., Djanibekov, A., Kulkarni, A., Xia, G., & Choudhury, M. (2025). Music for all: Representational bias and cross-cultural adaptability of music generation models. *Findings of the Association for Computational Linguistics: NAACL*

2025, 4569–4585. https://doi.org/10.18653/v1/2025.findings-naacl.258

Park, S., & Arthur, C. (2026). From prompting to describing: A cross-cultural study of language for AI-generated music. *Proceedings of the 27th International Society for Music Information Retrieval Conference*. https://doi.org/10.48550/arXiv.2608.06634

Poell, T., Nieborg, D., & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, *8*(4), 1–13. https://doi.org/10.14763/2019.4.1425

Udio. (n.d.). *Prompt like a master*. Retrieved August 11, 2026, from https://help.udio.com/en/articles/10716541-prompt-like-a-master

Vélez Vásquez, M. A., Pouw, C., Burgoyne, J. A., & Zuidema, W. (2024). Exploring the inner mechanisms of large generative music models. *Proceedings of the 25th International Society for Music Information Retrieval Conference*, 791–798. https://zenodo.org/records/14877449/files/000086.pdf