TISMIR

Raveform: A Dataset of Metrical and Functional Structure Annotations for EDM Tracks in DJ Mixes

DATASET ARTICLE

ubiquity press

TAEJUN KIM
JONGSOO KIM
HYUNGYU KIM
JUHAN NAM

*Author affiliations can be found in the back matter of this article

ABSTRACT

Research on disc jockey (DJ) mixes remains limited in the field of music information retrieval (MIR) despite the long history and cultural significance of DJing. This is largely due to the lack of comprehensive datasets and studies on DJ mixes. In this article, we introduce Raveform, a novel dataset containing 4,902 DJ mix links and 56,873 individual track links, including 1,423 tracks with detailed structural annotations manually annotated by three domain experts. To ensure consistency, we introduce an electronic dance music (EDM)-specific structural vocabulary and propose a framework that reconciles the subjective perspectives of the domain experts through standardized annotation guidelines. Using the dataset, we analyze the structural characteristics of DJ mixes in terms of segment length, flow, and sub-band energy. Furthermore, we train and evaluate deep learning models on metrical and functional structure analysis tasks. We show that the proposed dataset tailored to DJ mixes is effective for the tasks, especially when combined with existing resources. The Raveform dataset and associated models are publicly available at https://mir-aidj.github.io/raveform/.

CORRESPONDING AUTHOR:
Juhan Nam
KAIST, 291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea
juhan.nam@kaist.ac.kr

KEYWORDS:
DJ, dataset, music structure analysis, structural annotations, metrical analysis, beat, downbeat, computational musicology

TO CITE THIS ARTICLE:
Kim, T., Kim, J., Kim, H., & Nam, J. (2026). Raveform: A Dataset of Metrical and Functional Structure Annotations for EDM Tracks in DJ Mixes. *Transactions of the International Society for Music Information Retrieval*, 9(1), 131–143.
DOI: https://doi.org/10.5334/tismir.288

1 INTRODUCTION

The term 'disc jockey (DJ)' first appeared in print in a 1941 issue of *Variety* magazine, originally referring to radio hosts who played recorded music (Fisher, 2009). Since then, the role of the DJ has evolved significantly, with performances now taking place in clubs, festivals, livestreams, and podcasts. Despite its long history and cultural significance, DJing remains underexplored in the field of music information retrieval (MIR), particularly in relation to understanding or supporting its creative practices. This research gap may result from the lack of dedicated datasets and analytical tools tailored for DJ-related studies.

The DJing process can be viewed as the seamless concatenation of multiple tracks, typically involving three steps: (1) track selection, (2) mix point selection, and (3) DJ mixing. Track selection refers to choosing which tracks to play and in what order. Mix point selection involves determining where transitions occur between tracks.[1] DJ mixing consists of applying audio effects, such as equalization, filtering, and beatmatching, to blend the two tracks smoothly. It is important to note that, in practice, these steps are often interdependent, requiring DJs to consider multiple aspects simultaneously.

In MIR, various tasks align with the three steps of DJing and can support their understanding or automation. Track selection (step 1) relates to music recommendation (Van den Oord et al., 2013), music auto-tagging (Kim et al., 2018; Lee et al., 2017), playlist generation (Flexer et al., 2008), and sequencing (Bittner et al., 2017). To determine which tracks a DJ played and in what order, track identification via audio fingerprinting is commonly used (Cano et al., 2005; Sonnleitner et al., 2016; Wang, 2006).

Mix point selection (step 2) has received the least attention among the three DJing steps, both in industry and academia. Although some commercial DJ applications offer features that attempt automatic mix point selection, their performance remains generally unsatisfactory. To the best of our knowledge, the industry has yet to systematically explore the creative decision-making process behind mix point selection. Similarly, academic interest in this area has been limited, with only a few studies addressing the problem (see Section 2 for details). Understanding and automating mix point selection is inherently tied to music structure analysis, as segment boundaries frequently serve as strong candidates for mix points. While structure analysis itself has been actively studied, there is a notable lack of publicly available datasets specifically designed for the structure analysis of electronic dance music (EDM), a broad family of percussive, electronically produced dance music created primarily for nightclubs, raves, and festivals. We posit that this persistent lack of attention to mix point selection is largely due to the scarcity of analytical tools—particularly structure analysis models—and the limited availability of relevant datasets.

DJ mixing (step 3) has received significantly less academic attention than track selection (step 1). Research on automating DJ mixing is scarce, with Chen et al. (2022) being one of the few studies addressing this topic. In contrast, the industry has made several attempts to implement automated DJ mixing through features commonly referred to as AutoMix, which are integrated into commercial DJ applications.[2,3,4] However, these industry efforts focus on functionality rather than on understanding the underlying creative processes involved in DJ mixing. In academia, only a few studies have analyzed real-world DJ performances to investigate how human DJs control mixers during transitions (Kim et al., 2021, 2022). In addition, Schwarz and Fourer (2019) explored synthetic datasets and proposed analytical methods related to DJ mixing.

Figure 1 provides an overview of these components in the context of a DJ mix, and Table 1 summarizes the main files in the dataset and the information each of them contains.

In this article, we address the aforementioned limitations by introducing Raveform, a comprehensive dataset designed to support research on DJing and EDM structure analysis. Raveform contains metadata and track lists for 4,902 DJ mixes, links to 56,873 tracks featured in these mixes, beat sequences for all mixes and tracks, mix-to-track alignment files with estimated mix-in and mix-out positions, and structural annotations for a subset of 1,423 tracks. The structural annotations include tempo, beats, downbeats, functional segment boundaries, and segment labels. To support consistent and meaningful annotation, we propose a specialized vocabulary and a set of annotation guidelines tailored to EDM, emphasizing an energy-centered perspective that reflects how DJs perceive musical structure. These resources were developed through iterative cross-validation and discussion among three domain experts. Figure 1 provides an overview of how these components relate in a DJ mix, and Table 1 summarizes the main dataset files. Using this dataset, we train and evaluate structure analysis models specifically designed for EDM; both the dataset and the models will be publicly released to facilitate future research.

2 RELATED WORKS

2.1 DJ MIX DATASETS

Only a few datasets containing DJ mixes are currently available. Sonnleitner et al. (2016) introduced two datasets for evaluating audio fingerprinting in DJ mixes and publicly released one of them. This dataset includes track lists and the corresponding tracks played in the mixes, similar to some aspects of our proposed dataset. However, its scale is limited to only 10 DJ mixes and 118 tracks, making it insufficient for large-scale analysis.

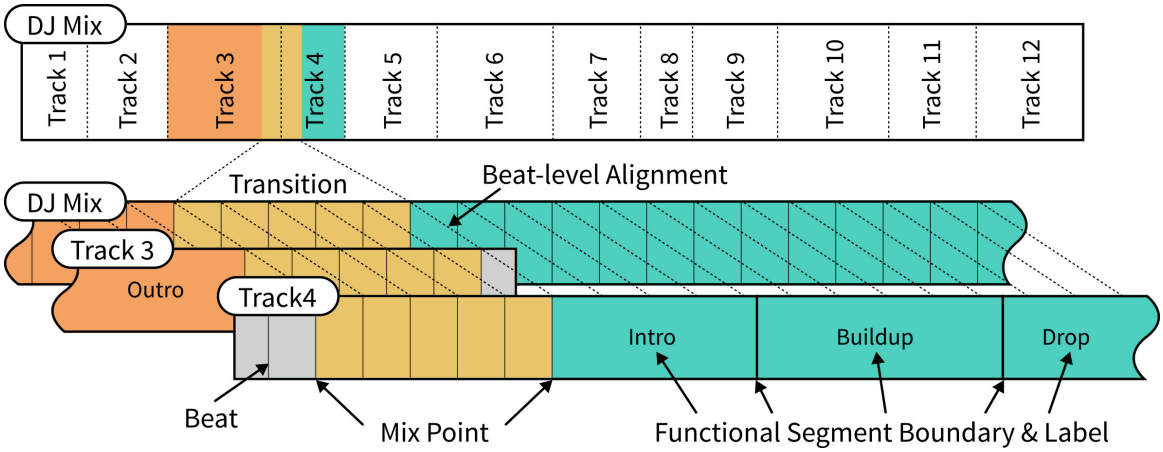


Figure 1 An illustration of components of the Raveform dataset. A DJ mix includes a track list, provides beats of the mix and track and has alignments between the DJ mix and each track at the beat level. Also, mix points are extracted based on the alignment. A subset of tracks are manually annotated structures: beats, downbeats, and functional segment boundaries with labels.

File	Content
`mixes.jsonl`	Metadata for DJ mixes, including track lists.
`tracks.jsonl`	Metadata for individual tracks that appear in the mixes.
`alignments/` `└ *.jsonl`	Mix-to-track alignments for each mix, including estimated mix points.

`beats/`
`└ mixes/*.json`
`└ tracks/*.json`
	Beats estimated for DJ mixes and tracks.

`structures/`
`└ beats/*.csv`
`└ segments.json`
	Human-annotated structure labels for a subset of tracks.

Table 1 Summary of the dataset components.

Another dataset, UnmixDB, was introduced by Schwarz and Fourer (2019). Although it also consists of DJ mixes and associated tracks, the mixes are automatically generated, which limits their suitability for musicological studies that aim to analyze human DJ behavior.

2.2 MIX/CUE POINT DATASETS

The term 'cue point' is inherently ambiguous. While it generally refers to a notable position within a track, a 'mix point' specifically denotes the start or end of a transition between two tracks, resulting in four distinct positions per transition. In this article, we distinguish clearly between these two terms. Prior research on mix/cue point detection is related to three main tasks: (1) mix segmentation that divides a DJ mix into its constituent tracks, (2) mix/cue point extraction that identifies such points given a DJ mix and track list, and (3) mix/cue point estimation that automatically predicts transition points for a given track. Studies on mix segmentation (Glazyrin, 2014; Scarfe et al., 2014) can detect transitions between tracks but are limited in semantic interpretation due to the lack of full track context. In cue point extraction, Schwarz and Fourer (2019) proposed a method evaluated only on synthetic data. Kim et al. (2020) developed a method for mix point extraction using a real-world dataset (1.5k DJ mixes, 15k tracks), but the dataset is not publicly available and lacks structural annotations, limiting deeper analysis. For cue point estimation, several heuristic approaches exist (Schwarz et al., 2018; Veire and De Bie, 2018), but they are not data-driven. Recently, Argüello et al. (2024) have released a dataset containing 4,710 tracks and 21k cue points (on average, 4.6 per track) curated from four professional DJs, along with a data-driven estimation model. However, the inherent ambiguity of the term 'cue point' remains problematic. Annotators often apply inconsistent criteria, and a cue point generally indicates a single position rather than a transition region. For example, it is unclear whether a cue point marks the beginning of a fade-out or the exact moment playback should stop. Moreover, DJ transitions are context-dependent, considering both the previous and next tracks. Yet, to our knowledge, no existing work explicitly models this relational aspect of mix point selection.

2.3 MUSIC STRUCTURE DATASETS

Various datasets have been developed for music structure analysis. A comprehensive overview is available in the work of Nieto et al. (2020), which reviews eight such datasets. Among these, the two most widely used are SALAMI (Smith et al., 2011) and the Harmonix Set (Nieto et al., 2019). SALAMI is the largest publicly available dataset for structure analysis, comprising 1,359 tracks across diverse genres such as classical, jazz, pop, world, and live recordings. It provides hierarchical segment annotations but lacks beat and downbeat information. Of the tracks, 1,048 (77%) were annotated by two independent annotators, while the remaining

335 (23%) were annotated by a single annotator. An annotator's guide was provided, along with a controlled vocabulary for functional segment labels (e.g., intro, verse, chorus). The dataset has been used to study inter-annotator agreement, highlighting the subjectivity involved in perceiving musical structure. The Harmonix set, in contrast, includes annotations for segments, beats, and downbeats, making it the largest dataset with all three types of annotations (912 tracks). This dataset primarily consists of Western popular music, including songs from the dance/electronic, hip hop, country, and rock genres. It was originally created to support rhythm-action (beat-matching) games. Each track is annotated by a single annotator. Although no formal annotation guide has been published, the vocabulary used for segment labels is diverse and often inconsistent, with many terms appearing only once or twice, suggesting the absence of a standardized label set.

In this study, we introduce the Raveform dataset, which includes beat, downbeat, and structural annotations specifically curated for EDM tracks. While several public datasets and models exist for music structure analysis, their applicability to EDM and DJ mixes remains limited. For example, despite the rhythmic regularity of EDM, existing beat and downbeat tracking models often exhibit degraded performance due to the lack of genre-specific training data (see the cross-dataset evaluation in Table 4). Furthermore, in analyses of Western popular music, song structure is often described in terms of verse–chorus forms, where the repetition and variation of lyrics play a central role in distinguishing sections (e.g., verses, choruses, bridges). This lyric-centric perspective works well for much Western popular music but is less suitable for EDM, where tracks are frequently instrumental and structural perception is driven more by changes in rhythm, texture, and energy than by textual form. To address this gap, our annotation methodology adopts an **energy-centered** perspective, focusing on perceived changes in musical intensity and function—an approach more aligned with how DJs interpret track structure during performance and mixing.

Our annotation methodology also diverges from those adopted in existing datasets. While independent annotations by multiple annotators are valuable for exploring the subjectivity of music structure perception, they often result in inconsistencies within the dataset. These inconsistencies can undermine the reliability and interpretability of structure analysis models trained on such data. For instance, annotators may disagree on whether a chorus segment begins with the vocal pickup before the downbeat or exactly at the downbeat. A model trained on conflicting annotations may adopt inconsistent boundary definitions, making its behavior difficult to predict or interpret. To address this issue, Raveform employs a collaborative annotation process in which each track is reviewed by multiple annotators, but the final output is a single, consolidated annotation. Annotators cross-validated each other's work and engaged in iterative discussions to resolve ambiguities and define consistent principles for segmentation and labeling. Through this process, we developed a domain-specific vocabulary and a detailed set of annotation guidelines, ensuring both internal consistency and relevance to the DJing context.

3 THE VOCABULARY FOR EDM STRUCTURES

This section introduces the vocabulary established for segment labels used to describe EDM structures. Figure 2 shows a representative structural pattern commonly observed in the proposed dataset, with darker regions indicating higher energy; see Section 5.2 for details on how this pattern was extracted. A summary of the vocabulary and their key characteristics is presented in Table 2, while detailed definitions are provided in the following text. Most terms are widely used among DJs and EDM producers, though *ambient-intro* and *ambient-outro* are less common. Each term is described from two perspectives: *energy* and *tension*.

Our goal in this section, and in this work more generally, is not to propose a universal theory of EDM form but rather to formalize a practical vocabulary for this dataset. The scope of the definitions is limited to patterns that we repeatedly observe in our corpus of EDM tracks appearing in DJ mixes, which we then generalize into a small set of reusable segment functions. The labels are therefore author-defined conventions, but we deliberately ground them in real-world DJ practice and common terminology, drawing both on our own experience and on practitioner literature, such as Broughton and Brewster (2003).

Energy refers to a combination of loudness, rhythmic density, frequency density, and instrumentation. Segments with high energy are characterized by high volume, dense rhythmic activity, broad frequency coverage, and a rich instrumental texture. Conversely, segments with low energy exhibit reduced loudness, sparser rhythms, limited frequency usage, and fewer instrumental layers.

Tension is distinct from energy. While similar to the concept of tension in jazz, where it creates anticipation for a subsequent release, in EDM, tension is shaped by

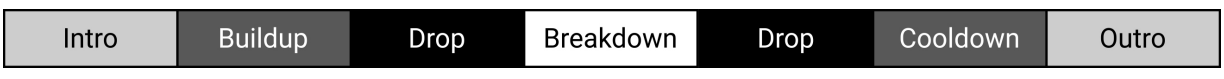


Figure 2 A representative structural pattern commonly observed in the proposed dataset, with darker regions indicating higher energy.

Name	Characteristic Summary
Intro	Appears at the beginning. Primarily consists of percussive sounds, designed to facilitate seamless mixing for DJs.
Buildup	Typically precedes a breakdown or a drop. Gradually increases energy by introducing musical elements one by one.
Breakdown	Appears before a drop. Characterized by a sudden drop in energy to heighten contrast with the upcoming drop.
Drop	The main section of the track, conveying its core musical idea. It is the most energetic and danceable part, usually featuring all instruments together.
Cooldown	Follows a drop and precedes a breakdown or outro. Gradually reduces energy, functioning as the opposite of a buildup.
Bridge	Often occurs between two breakdowns before the final drop. Offers contrast and builds anticipation for the final drop. Rare in this dataset.
Outro	Occurs at the end of the track. Functions as the opposite of an intro, yet similarly features percussive sounds to aid in mixing.
Ambient-intro	An intro section without percussive elements. Beats are nearly unrecognizable; the section consists mainly of melodic, harmonic, or ambient textures.
Ambient-outro	An outro section without percussive elements. Sonically similar to an ambient-intro.

Table 2 A summary of the vocabulary for segment labels.

different musical features. Specifically, it is influenced by rhythmic unpredictability, sharp or distorted high-frequency sounds, loudness, instrumentation, and the distribution of energy across the frequency spectrum, rather than by harmonic progression. Tension typically reaches its peak in breakdowns and is released in the following drop. Due to the diverse production styles in EDM, producers employ various techniques to create tension, making it a more nuanced and multifaceted concept than energy. High-tension segments often exhibit unpredictable rhythms, dissonant or harsh timbres, reduced presence of low frequencies, and piercing sounds in the mid-to-high frequency range. Notably, silence itself can also induce tension. In certain subgenres, such as drum and bass, tension may be heightened by removing drums and retaining only the bass elements.

3.1 FUNCTIONS OF SEGMENT LABELS

To complement the textual definitions below, we also provide audio examples for each segment function on the companion website.[5]

Intro. An *intro* appears at the beginning of a track and is specifically designed by producers to facilitate seamless mixing by DJs. Its primary function is to enable beat-matching with the currently playing track. To serve this purpose, it typically consists of percussive elements with straightforward rhythmic patterns and minimal harmonic content, thereby avoiding potential key clashes. Common rhythmic components include four-on-the-floor kicks (or bass drum hits) and/or backbeat snares. In EDM, which predominantly uses a $\frac{4}{4}$ time signature, 'four-on-the-floor' refers to a pattern where a kick drum is played on every beat of a bar (beats 1, 2, 3, and 4), while the backbeat snare pattern places snare hits on beats 2 and 4. In terms of energy, the intro generally has higher energy than a breakdown but lower than that of a buildup or a drop. Its energy profile is typically steady or gradually increasing. Tension is usually low and stable throughout the segment, contributing to the intro's unobtrusive character—an intentional design choice that allows it to blend smoothly with other tracks during a DJ mix.

Buildup. A *buildup* typically follows the intro and precedes either a breakdown or a drop. Its primary function is to progressively increase both energy and tension, although its energy level remains lower than that of the drop. This escalation is often achieved through the gradual addition of instruments. Not all tracks contain a distinct buildup; its presence depends on the producer's structural strategy. Tension can be developed within the buildup, the breakdown, or distributed across both, reflecting the producer's approach.

Breakdown. A *breakdown* typically precedes a drop and serves as a crucial transitional segment characterized by a sharp reduction in energy and a progressive buildup of tension. It is often the segment with the lowest energy in the entire track. Initially, a breakdown may exhibit both low energy and low tension; however, as it unfolds, the tension intensifies, while energy remains minimal. This contrast amplifies the perceived impact of the forthcoming drop by creating a heightened sense of anticipation. To achieve this effect, producers frequently remove low-frequency elements such as kick drums and basslines. Nonetheless, strategies for building tension in breakdowns vary widely. Some breakdowns consist solely of ambient textures, while others may incorporate sharp hi-hats with complex rhythmic patterns.

Drop. The *drop* represents the core of an EDM track, where the primary musical idea is fully realized. Typically emerging after a breakdown or buildup, it marks

the point of maximal energy and the release of accumulated tension. Drops are characterized by a dense sonic texture, often featuring the simultaneous presence of drums, bass, melodic elements, and various effects. Later drops within a track frequently exhibit greater energy than earlier ones, contributing to a sense of escalation across the track's progression. Structurally, drops are usually followed by either a breakdown or a cooldown.

Cooldown. A *cooldown* segment follows a drop and precedes either a breakdown or an outro. Functionally, it serves as the inverse of a buildup, gradually reducing energy and tension. By providing a smoother transition, cooldowns help avoid abrupt shifts in intensity, particularly when moving toward structurally calmer sections. As with buildups, cooldowns are not universally present and are selectively employed by producers based on the desired flow and emotional contour of the track.

Bridge. As illustrated in Figure 9, a *bridge* segment appears infrequently in this dataset. When it does occur, it is typically situated near the middle of a track following a breakdown and preceding either another breakdown or a drop—and generally appears only once per track. Its role is analogous to that of a bridge in popular music: to introduce contrast, disrupt repetitive patterns, and re-engage the listener's attention. Musically, it often features distinct rhythmic, harmonic, or timbral elements that differentiate it from the surrounding sections.

Outro. An *outro* appears at the end of a track and shares many sonic characteristics with an *intro*, typically featuring percussive sounds and straightforward beats. Functionally, it serves the same purpose as an intro, facilitating smooth mixing for DJs. Unlike the intro, however, the energy in an outro usually remains steady or decreases slightly. Similar to the intro, the tension remains stable and low, ensuring that the segment blends seamlessly when transitioning to another track.

Ambient-Intro. An *ambient-intro* opens a track in a manner similar to an intro but is distinguished by the absence of percussive elements. Instead, it consists primarily of harmonic, melodic, and/or ambient sounds. This absence of rhythmic cues makes beat-matching nearly impossible for DJs, requiring alternative mixing strategies. Although the term *ambient-intro* is not as widely used as other functional segment labels, we introduce it here to capture this segment's distinctive sonic and functional characteristics.

Ambient-outro. An *ambient-outro*, located at the end of a track, is characterized by the absence of percussive elements. Like the ambient-intro, this lack of rhythmic structure makes traditional DJ mixing techniques such as beat-matching difficult. We introduce the term *ambient-outro*—in parallel with *ambient-intro*—to distinguish it from the more typical *outro*, which contains percussive elements.

3.2 CRITERIA FOR RESOLVING AMBIGUITY

Even though the functions of segment labels are described in the previous section, some ambiguity still remains in segmentation and labeling. This section outlines the criteria used to resolve such cases.

Regularity. EDM tracks often exhibit a regular phrase structure, with changes tending to occur at multiples of four or eight measures. Annotators used this regularity as a prior when placing boundaries: when several musically plausible options existed, they preferred boundaries that aligned with these phrase lengths. However, regularity was treated as a soft guideline rather than a hard constraint. When a clear musical change (e.g., in energy, rhythm, harmony, or instrumentation) occurred off the expected grid, boundaries were placed at the actual change, even if this resulted in sections of unusual length.

Consecutive segments with identical functions. Long passages with the same function—such as an extended drop—may span many measures; in our corpus, we even encountered drops lasting on the order of 80–90 measures. Such passages can either be treated as a single segment or split into multiple consecutive segments with the same label. We opt to divide these sections only when there is a clear musical change, such as a shift in energy, rhythm, chord progression, or instrumentation. Gradual evolutions are not used as dividing cues since identifying a precise boundary in such cases is inherently ambiguous. As a consequence, some long drops remain as a single segment, while others are split into several drops when discrete structural events occur. This policy reflects the balance between using regular phrase lengths as a guideline and maintaining consistent, musically motivated criteria across tracks. An example of a long drop that is split into three segments is provided as a listening example in the 'Straightforward Example' section of the companion website.

Drop-like intro. Some tracks exhibit high energy right from the beginning, resembling a drop that lasts from the start to the middle until reaching a breakdown. In such cases, this section is classified as the intro, as it is likely the point where a DJ would perform beat-matching.

Breakdown onset. In some tracks, the onset of a breakdown is ambiguous as the majority of instruments fade out gradually, ultimately reaching the lowest energy level. In such cases, the start is identified at the point where the energy begins to decrease, following the regularity rule previously described.

Cooldown and outro. When a track gradually removes instruments over a long period, it becomes difficult to distinguish between cooldown and outro segments. For example, the lead synth may fade out first, followed by the bassline after eight measures, and then the snares after another eight measures. While the changes in energy are noticeable, labeling becomes ambiguous. In such cases, we defined a segment as an outro when percussion elements dominate the texture. This allows for

sequences like cooldown–cooldown–outro–outro, based on the sonic characteristics.

Excluded tracks. Out of 2,000 tracks, 577 were excluded for various reasons. This was primarily due to the noisy nature of crowdsourced data, as track links were annotated by users and audio files were collected from YouTube. Consequently, we discarded tracks that were not the original versions. This includes tracks with fading starts, short versions (or radio edits), and those created by clipping from a DJ mix. Additionally, we removed experimental EDM tracks that were intentionally made with ambiguous structures. Audio examples of such excluded tracks are provided in the 'Ambiguous Example' section of the companion website.

4 DATASET AND STRUCTURE-ANNOTATION PROCESS

This section presents Raveform, a comprehensive dataset designed to support the analysis and automation of DJing, particularly in the context of EDM.

4.1 OVERVIEW

Table 3 summarizes key statistics of the dataset. The number of played tracks refers to the total instances of tracks played across all DJ mixes, including repeated appearances of the same track in different mixes. In contrast, the number of unique tracks counts each track only once, regardless of how often it appears. As a result, the number of played tracks exceeds the number of unique tracks. The number of transitions is smaller than the number of played tracks minus one because we only count a transition when two consecutive tracks in a mix are both successfully identified with available audio and (as detailed in Section 4.2) when their audio segments overlap after mix-to-track alignment. Out of 56,873 unique tracks, a subset of 1,423 tracks has been manually annotated with structural information by human experts. These annotated tracks represent the most frequently played selections across the mixes. Each annotation includes tempo, beat and downbeat positions, and functional segment boundaries with labels.

The number of mixes	4,902
The number of unique tracks	56,873
The number of played tracks	73,505
The number of tracks with structural annotations	1,423
The total length of mixes (in hours)	6,522
The number of available transitions	53,780

Table 3 Summary statistics of the *Raveform* dataset.

Figure 3 shows the genre distribution of the DJ mixes, while Figure 4 displays the genre distribution of the annotated tracks. Techno is the most represented genre in the dataset, followed by trance, drum and bass, and house, among others. Figure 5 presents the beats per minute (BPM) distribution of the annotated tracks. Tempo in our annotations follows a common EDM convention: when a stable four-on-the-floor kick pattern is present, we set the tempo so that each kick corresponds to one beat. The most common BPM is 128, which aligns with typical tempos in genres such as house and techno. A secondary concentration appears in the 170–175 BPM range, primarily due to the presence of drum and bass tracks.

4.2 DATA-COLLECTION PROCESS

All data, except for the structural annotations, were sourced from *MixesDB*,[6] a user-contributed database for

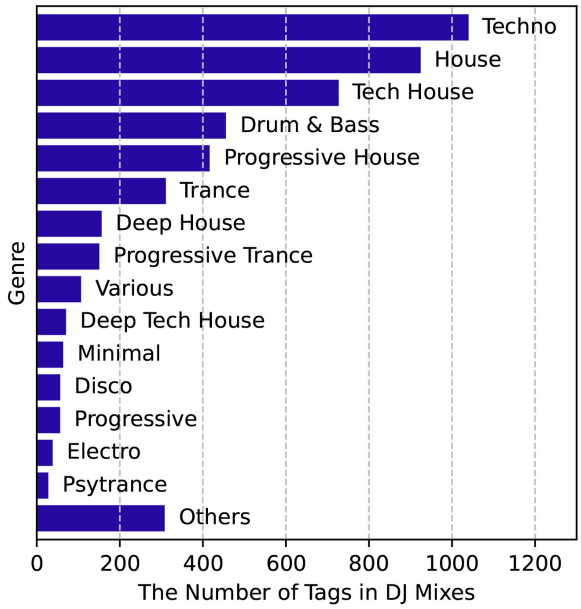


Figure 3 The genre distribution of the DJ mixes.

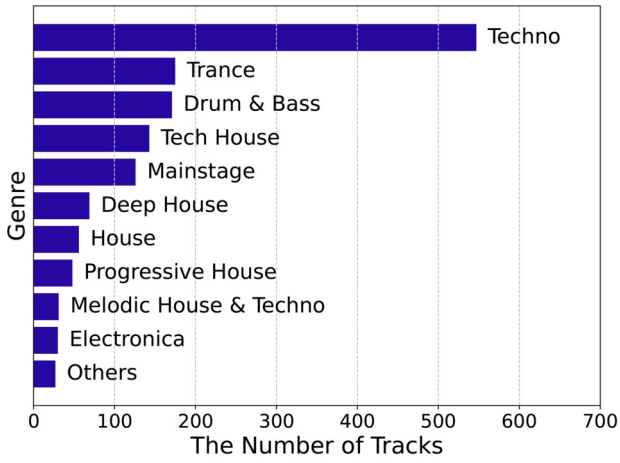


Figure 4 The genre distribution of the structure-annotated tracks.

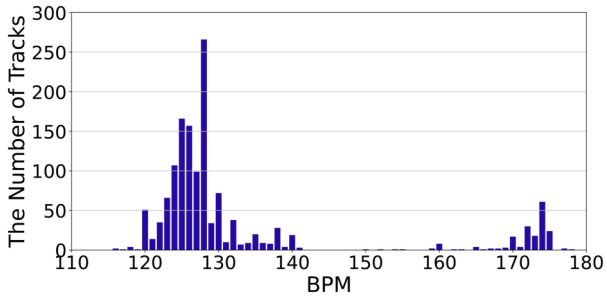


Figure 5 The tempo distribution of the structure-annotated tracks in beats per minute.

DJ mixes that provides metadata, audio links to the mixes, and links to the tracks played. To ensure accurate track identification, we used the site's filtering feature. However, it is important to acknowledge that the MixesDB dataset contains inherent inaccuracies in its annotations. For instance, a track listed in the track list may not be the exact version played by the DJ. Additionally, dance music often includes various versions of a single track, such as short, long, or remix versions, which may not align with the version listed in the dataset. Figure 6 illustrates the distribution of the fraction of identified tracks per DJ mix; 342 mixes (7%) are fully identified.

To obtain time-synchronized relations between mixes and tracks and to further filter out mismatched items, we estimate beat-level mix-to-track alignments for all candidate mix–track pairs using the method of Kim et al. (2020). Each alignment specifies, at the beat level, which beat index in the track corresponds to which beat index in the mix and thereby yields estimated mix-in and mix-out positions for that track within the mix. We use the fraction of correctly aligned beats (*match rate*) as a quality measure and discard mix–track pairs whose match rate is lower than 0.1, treating them as misidentified or nonmatching tracks. Figure 7 shows the distribution of the match rate over all aligned track instances. For the dataset statistics reported in Table 3, we retain only alignments whose match rate is at least 0.1. The process of collecting the structural annotations will be detailed in the following subsection.

4.3 STRUCTURE-ANNOTATION PROCEDURE

Music structure is subjective by its nature (Flexer and Grill, 2016; Nieto et al., 2020). As discussed in Section 2, our approach aims to incorporate perspectives from multiple annotators while establishing consistent annotation criteria. In essence, the goal is to make the inherently subjective judgments in structure annotation consistently applicable across different tracks. This subsection describes the annotation procedure and our efforts to standardize the annotation criteria.

Three experts participated in the annotation: two with DJing experience and one with a background in music production. The annotation process was organized

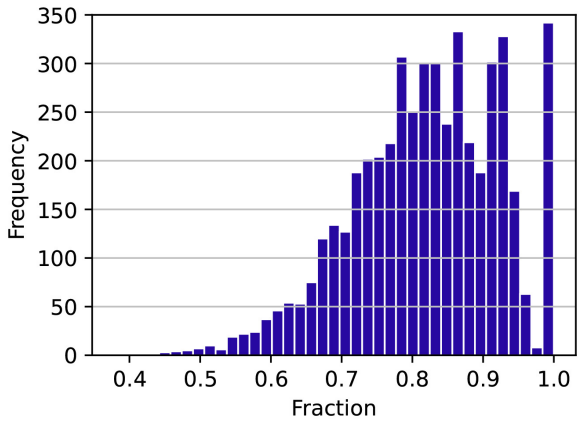


Figure 6 The distribution of the fraction of identified tracks per DJ mix.

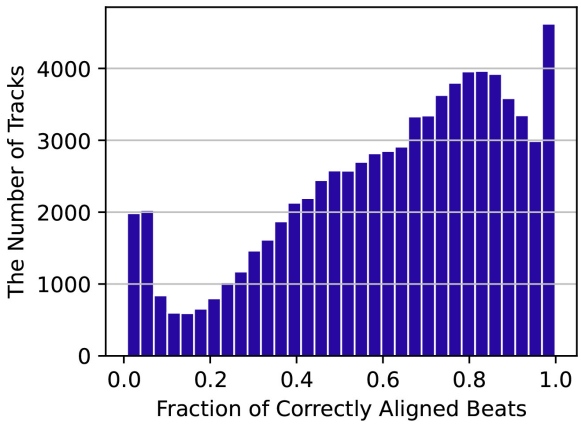


Figure 7 The distribution of the fraction of correctly aligned beats.

into three phases. First, an *annotator* marked segment boundaries and assigned labels for each track. These initial annotations were then reviewed by a *verifier*. Finally, a *final judge* conducted a last round of validation. During the initial phase, annotators were encouraged to flag tracks with ambiguous structures. In the validation phases, reviewers marked any annotations they disagreed with. Weekly meetings were held throughout the process, during which annotators listened to ambiguous tracks together and refined the annotation criteria through discussion. The roles were rotated regularly among the annotators to ensure balanced participation.

The annotation process spanned 7 months. In the early stages, frequent changes to the segment label vocabulary necessitated multiple restarts. Over time, as the annotators collaboratively listened to and discussed ambiguous cases, disagreements became less frequent. Most disagreements were resolved through discussion, as annotators came to understand and appreciate each other's perspectives. Often, disagreements arose from focusing on different musical elements—one annotator might focus on an added bassline, while another might pay more attention to the introduction of drums when

identifying segment boundaries. In such cases, both perspectives were typically valid, and the choice came down to selecting the more consistent or useful interpretation. To guide such decisions, annotators considered questions such as: Which perspective is more helpful for DJs? Is this information necessary for DJing? Which view better reflects energy-centric transitions? Can this criterion be applied consistently across other tracks?

In many cases, the challenge was not disagreement but rather the intrinsic ambiguity of a track's structure. For example, in EDM, particularly in repetitive and hypnotic techno, musical elements often fade in and out gradually. Although one can clearly identify a drop and an outro, it may be difficult to pinpoint the exact boundary between them due to the smooth, continuous decrease in energy and loudness. Capturing such ambiguity in a consistent and descriptive manner is inherently difficult. Therefore, collaborative listening and discussion were essential components of the annotation process.

Annotations were created using Rekordbox,[7] a DJ software platform. Segment boundaries and labels were recorded using the memory cue feature, while beats and downbeats were annotated with the help of Rekordbox's beat and downbeat tracking feature ('track analysis'). In our experience, this analysis works reliably for tracks with approximately constant tempo and four-on-the-floor drum patterns, which covers the majority of our corpus, although downbeat positions can occasionally be offset. For tracks with substantial tempo changes or more irregular rhythms, the estimated beat grid can become unstable. For every annotated track, we inspected the beat grid, corrected local beat/downbeat errors when necessary, and excluded tracks whose automatic estimates were too inaccurate to be corrected with reasonable effort. Annotations were then exported as XML files and processed using Python. Track genres were assigned by referencing the genre tags listed for each track on Beatport.[8]

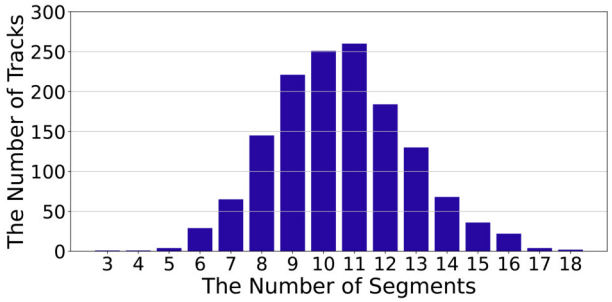


Figure 8 The distribution of the number of segments in the structure-annotated tracks.

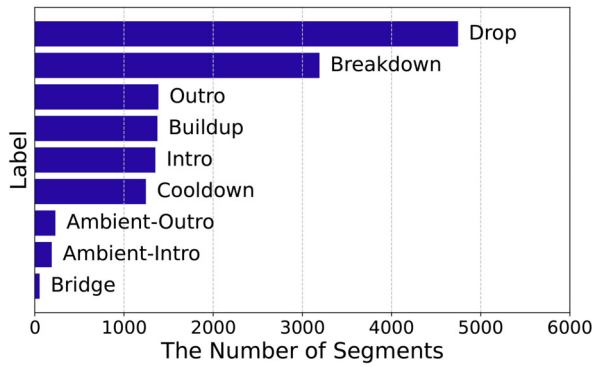


Figure 9 The number of appearing segments of each segment label.

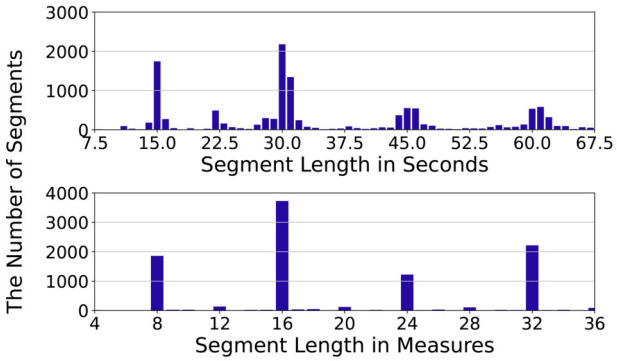


Figure 10 The distribution of segment length in seconds **(top)** and measures **(bottom)**.

5 STRUCTURE-ANNOTATION ANALYSIS

This section presents statistical analyses of segments and structure labels derived from the annotated EDM tracks.

5.1 SEGMENT-LEVEL STATISTICS

Figure 8 illustrates the distribution of the number of segments per track. Most tracks contain between 8 and 13 segments, a range that closely aligns with the segment distribution observed in the Harmonix set (Nieto et al., 2019). Figure 9 shows the distribution of segment labels across the dataset. Given that the dataset includes 1,423 tracks, labels such as 'Ambient-Intro,' 'Ambient-Outro,' and 'Bridge' appear infrequently, occurring far less often than the total number of tracks. In contrast, segments such as 'Intro,' 'Outro,' 'Buildup,' and 'Cooldown' typically appear once per track, while 'Drop' and 'Breakdown' tend to occur multiple times within a single track. Finally, Figure 10 presents the distribution of segment lengths, with the top portion showing lengths in seconds and the bottom portion showing such in measures. The most common segment length is 30 seconds or 16 measures. This peak corresponds to tracks with the most prevalent tempo of 128 BPM, where a 16-measure segment typically spans 30 seconds.

5.2 QUANTITATIVE ANALYSIS

Figure 11 illustrates the common structure of EDM tracks, acquired by aggregating the structural annotations. To analyze common structural patterns, we first normalized

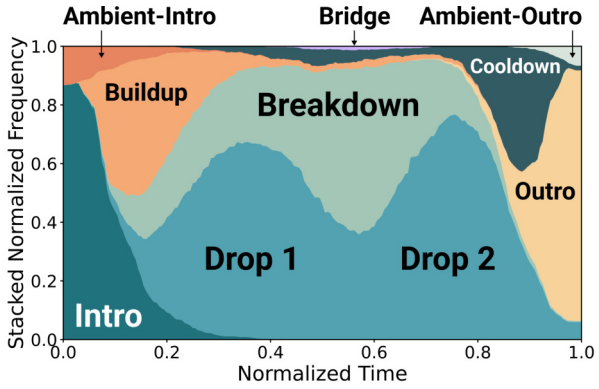


Figure 11 The common structure of EDM tracks.

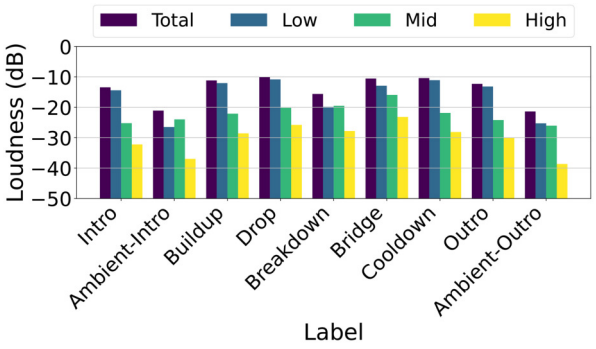


Figure 12 Average loudness (dB) values across frequency bands for each segment label.

the lengths of all tracks to a standard unit length. Then, at each point along this normalized timeline, we calculated the proportion of tracks assigned to each segment label. This approach allowed us to visualize how often each segment label occurs at each point in a typical EDM track. The visualization reveals a typical structure in EDM tracks: intro–buildup–drop–breakdown–drop–cooldown–outro.

Figure 12 illustrates the average loudness (dB) values across frequency bands for each segment label. As we defined the energy characteristics of the labels, the quantitative measure of energy, represented by dB, aligns with this definition. The drop possesses the highest energy, while segments such as the ambient-intro and ambient-outro, which lack many percussive instruments, exhibit the lowest energy. Similarly, the breakdown segment, characterized by the absence of many instruments, also displays low energy. Additionally, the energy increases progressively from intro to buildup to drop and conversely decreases from drop through cooldown to outro.

6 AUTOMATED EDM STRUCTURE ANALYSIS

The purpose of this section is to demonstrate the significance of the proposed dataset for automated metrical and functional structure analysis. Metrical structure analysis involves beat and downbeat tracking, while functional structure analysis includes segmenting and labeling musical tracks into the EDM structure vocabulary. This section trains models to (1) provide a benchmark of structure analysis on the dataset and (2) quantitatively show that EDM has different musical characteristics compared to Western popular music via cross-dataset evaluation. Additionally, the performance of a widely used publicly available beat and downbeat tracking model, Madmom (Böck et al., 2016a), is evaluated on both the proposed dataset and the popular music dataset.

6.1 EXPERIMENTAL SETUP

6.1.1 Model architecture

All models except Madmom have the same architecture as the All-In-One structure analysis model (Kim and Nam, 2023). It takes source-separated spectrograms as inputs, processes the inputs using dilated neighborhood attention (DiNA) transformers (Hassani and Shi, 2022), and applies postprocessing at the end. The source separation allows the model to analyze input tracks in an arrangement view. The DiNA adopts a self-attention mechanism similar to the famous transformer (Vaswani et al., 2017), but only the nearest neighboring values participate in the computation, akin to convolution. The model features exponentially growing dilations, resulting in a receptive field of about 82 seconds. For comprehensive architectural details, please refer to the original paper (Kim and Nam, 2023).

6.1.2 Evaluation metrics

We evaluate beat and downbeat tracking with three conventional metrics: F1, CMLt, and AMLt (Davies et al., 2009). The beat/downbeat F1 score is computed from precision and recall with a fixed tolerance of ±70 ms around each annotated beat. CMLt ('Correct Metrical Level, total') measures the proportion of beats that are correctly tracked at the intended metrical level, using a tempo-relative tolerance window of ±17.5% of the local beat period instead of a fixed time window. AMLt ('Allowed Metrical Levels, total') relaxes this constraint by also giving credit to predictions that are metrically consistent but at simple metrical ambiguities, such as double/half tempo or phase-shifted variants.

For structure, we use one boundary-based metric and two frame-clustering metrics, all of which are conventional in music structure analysis (Nieto et al., 2020). HR.5F is a boundary F1 score with a ±0.5-second tolerance window around each reference boundary. PWF is a pairwise frame-clustering F-score: over all pairs of time frames, it measures how often the system correctly decides whether the two frames belong to the same segment or to different segments. Sf is a normalized conditional entropy score that can be viewed as a probabilistic analog of PWF: instead of counting hard same/different decisions for frame pairs, it measures how consistently

the frames of each reference segment are assigned to predicted segments and penalizes cases where a single reference segment is fragmented across many predicted segments.

6.1.3 Cross-dataset evaluation

The *Harmonix set* (Nieto et al., 2019), which comprises 912 Western popular music tracks, is used for cross-dataset evaluation. Following convention, scores are reported using eight-fold cross-validation. We train models on each of the two datasets and report scores on the test set from the same dataset as the training set and on the cross-dataset. Additionally, to validate the model's capability to learn multiple genres, we train models using both the Harmonix set and the Raveform dataset and evaluate them on each dataset. Lastly, the performance of Madmom, which was trained on 10 datasets covering various genres except EDM (Böck et al., 2016b), on both datasets is reported. Note that Madmom only performs beat and downbeat tracking and does not include functional structure analysis.

6.2 RESULTS

Table 4 summarizes the cross-dataset evaluation results of metrical and functional structure analysis. The top four rows correspond to the test set from the Harmonix set, while the bottom four rows pertain to the Raveform test set. For each test set, the first row reports the performance of the model trained on the training split of the same dataset. The second row shows the model trained on the cross-dataset training split; the third row corresponds to the model trained using both datasets; and the fourth row presents the results of Madmom, which has a different architecture and was trained on 10 datasets, excluding both the Harmonix set and Raveform.

Overall, the performance on Raveform is higher than that on the Harmonix set, suggesting that Raveform provides more straightforward learning targets. In particular, scores for segmentation and labeling are consistently higher under the same experimental conditions, except for the dataset. Given the subjective nature of segment boundaries and labels, this suggests that Raveform's annotations are more consistent. As expected in cross-dataset evaluation, performance degrades when models are tested on a dataset different from the one they were trained on. Notably, the performance drop is more pronounced for downbeats, segments, and segment labels that involve more abstraction compared to beats. Models trained on both datasets exhibit comparable performance to models trained and evaluated on the same dataset. Madmom shows competitive results in beat tracking but lags significantly in downbeat tracking.

7 CONCLUSIONS

We introduce Raveform, a novel dataset comprising 4,902 DJ mix links and 56,873 track links, of which 1,423 tracks include structural annotations provided by three domain experts. These annotations are intended to support the development of EDM structure-analysis models; however, because music structure is inherently subjective, models trained on inconsistently annotated data will likewise produce inconsistent predictions. To address this challenge, we propose a framework that captures

Training Set (Genre)	Model	Beat			Downbeat			Segment	Label	
		F1	CMLt	AMLt	F1	CMLt	AMLt	HR.5F	PWF	Sf
*Tested on **Harmonix set** (Popular Music)*										
Harmonix (Pop)	All-In-One	**0.958**	**0.913**	**0.964**	0.915	0.873	0.932	**0.660**	**0.738**	**0.769**
Raveform (EDM)	All-In-One	0.810	0.622	0.856	0.727	0.591	0.792	0.509	0.533*	0.543*
Both (Pop+EDM)	All-In-One	0.953	0.895	0.963	**0.921**	**0.876**	**0.939**	0.659	0.720	0.751
10 Datasets (Various)	Madmom	0.941	0.859	0.955	0.805	0.756	0.882	–	–	–
*Tested on **Raveform** (EDM)*										
Raveform (EDM)	All-In-One	**0.991**	**0.985**	**0.991**	0.965	0.964	**0.971**	**0.835**	**0.847**	**0.890**
Harmonix (Pop)	All-In-One	0.930	0.890	0.918	0.753	0.746	0.812	0.635	0.542*	0.643*
Both (Pop+EDM)	All-In-One	0.990	**0.985**	**0.989**	**0.967**	**0.965**	**0.971**	**0.835**	0.842	**0.890**
10 Datasets (Various)	Madmom	0.947	0.930	0.938	0.669	0.678	0.792	–	–	–

Table 4 Cross-dataset evaluation results of metrical and functional structure analysis tested on Harmonix set and *Raveform*. Boldface indicates the best performance for each metric on each test set. ∗denotes scores achieved with label mappings (see Section 6.1.3 for details).

the subjective perspectives of multiple annotators while enforcing consistent annotation criteria, and we also present an EDM-specific structural vocabulary to guide the annotation process. Through quantitative evaluation, we demonstrate that existing datasets and state-of-the-art structure-analysis models perform poorly on EDM tracks, underscoring the necessity of Raveform. However, it is important to note that this dataset is limited to specific EDM subgenres—for example, it does not include disco—and that the proposed vocabulary reflects subjective choices that may not generalize universally. We anticipate that Raveform will catalyze future research in DJ and EDM structure analysis.

ACKNOWLEDGMENTS

The authors would like to thank Neutune[9] for providing the computing resources used in this research. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2019-NR042825) and by the Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (No. RS-2021-KC000084).

COMPETING INTERESTS

The authors have no competing interests to declare.

NOTES

1. A mix point refers to the position where a track starts or ends during a transition, totaling four points per transition. While the term 'cue point' is more commonly used, it has varying meanings depending on context, generally referring to any significant moment in a track. For clarity, we use 'mix point' instead of 'cue point.'
2. Offtrack—https://www.offtrack.com/
3. Djay Pro—https://www.algoriddim.com/djay-pro-mac
4. Spotify Automix—https://community.spotify.com/t5/FAQs/Automix-Overview/ta-p/5257278
5. https://raveform.netlify.app
6. https://www.mixesdb.com
7. https://rekordbox.com/
8. https://beatport.com/
9. https://neutune.com/

AUTHOR AFFILIATIONS

Taejun Kim
Neutune, 524, Bongeunsa-ro, Gangnam-gu, Seoul, Republic of Korea

Jongsoo Kim
KAIST, 291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea

Hyungyu Kim
KAIST, 291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea

Juhan Nam
KAIST, 291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea

REFERENCES

Argüello, G., Lanzendörfer, L. A., and Wattenhofer, R. (2024). Cue point estimation using object detection. In *International Society of Music Information Retrieval Conference (ISMIR)*.

Bittner, R. M., Gu, M., Hernandez, G., Humphrey, E. J., Jehan, T., McCurry, H., and Montecchio, N. (2017). Automatic playlist sequencing and transitions. In *International Society of Music Information Retrieval Conference (ISMIR)* (pp. 442–448).

Böck, S., Korzeniowski, F., Schlüter, J., Krebs, F., and Widmer, G. (2016a). madmom: A new Python audio and music signal processing library. In *International Conference on Multimedia* (pp. 1174–1178).

Böck, S., Krebs, F., and Widmer, G. (2016b). Joint beat and downbeat tracking with recurrent neural networks. In *International Society of Music Information Retrieval Conference (ISMIR)* (pp. 255–261).

Broughton, F., and Brewster, B. (2003). *How to DJ right: The art and science of playing records*. Grove Press.

Cano, P., Batlle, E., Kalker, T., and Haitsma, J. (2005). A review of audio fingerprinting. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, *41*, 271–284.

Chen, B.-Y., Hsu, W.-H., Liao, W.-H., Ramírez, M. A. M., Mitsufuji, Y., and Yang, Y.-H. (2022). Automatic DJ transitions with differentiable audio effects and generative adversarial networks. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Davies, M. E., Degara, N., and Plumbley, M. D. (2009). Evaluation methods for musical audio beat tracking algorithms. *Queen Mary University of London, Centre for Digital Music, Tech. Rep. C4DM-TR-09-06*.

Fisher, M. (2009). *Something in the air: Radio, rock, and the revolution that shaped a generation*. Random House.

Flexer, A., and Grill, T. (2016). The problem of limited inter-rater agreement in modelling music similarity. *Journal of New Music Research*, *45*(3), 239–251.

Flexer, A., Schnitzer, D., Gasser, M., and Widmer, G. (2008). Playlist generation using start and end songs. In *International Society of Music Information Retrieval Conference (ISMIR)* (Vol. 8, pp. 173–178).

Glazyrin, N. (2014). Towards automatic content-based separation of DJ mixes into single tracks. In *Proceedings International Society for Music Information Retrieval Conference (ISMIR)* (pp. 149–154). https://github.com/nglazyrin/MixSplitter

Hassani, A., and Shi, H. (2022). Dilated neighborhood attention transformer. *arXiv:2209.15001*.

Kim, T., and Nam, J. (2023). All-in-one metrical and functional structure analysis with neighborhood attentions on demixed audio. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*.

Kim, T., Choi, M., Sacks, E., Yang, Y.-H., and Nam, J. (2020). A computational analysis of real-world DJ mixes using

mix-to-track subsequence alignment. In *Proceedings International Society for Music Information Retrieval Conference (ISMIR)* (pp. 764–770).

Kim, T., Lee, J., and Nam, J. (2018). Sample-level CNN architectures for music auto-tagging using raw waveforms. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 366–370). IEEE.

Kim, T., Yang, Y.-H., and Nam, J. (2021). Reverse engineering the transition regions of real-world DJ mixes using sub-band analysis with convex optimization. In *New Interfaces for Musical Expression (NIME)*.

Kim, T., Yang, Y.-H., and Nam, J. (2022). Joint estimation of fader and equalizer gains of DJ mixers using convex optimization. In *International Conference on Digital Audio Effects (DAFx)* (pp. 312–319).

Lee, J., Park, J., Kim, K. L., and Nam, J. (2017). Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms. In *Sound and Music Computing Conference (SMC)*.

Nieto, O., McCallum, M., Davies, M. E., Robertson, A., Stark, A. M., and Egozy, E. (2019). The Harmonix set: Beats, downbeats, and functional segment annotations of western popular music. In *International Society of Music Information Retrieval Conference (ISMIR)* (pp. 565–572).

Nieto, O., Mysore, G. J., Wang, C.-I., Smith, J. B., Schlüter, J., Grill, T., and McFee, B. (2020). Audiobased music structure analysis: Current trends, open challenges, and applications. *Transactions of the International Society for Music Information*, *3*(1).

Scarfe, T., Koolen, W., and Kalnishkan, Y. (2014). Segmentation of electronic dance music. *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications*, *22*(3), 4. https://github.com/ecsplendid/DanceMusicSegmentation

Schwarz, D., and Fourer, D. (2019). Methods and datasets for DJ-mix reverse engineering. In *Proceedings International Symposium on Computer Music Multidisciplinary Research (CMMR)* (pp. 426–437).

Schwarz, D., Schindler, D., and Spadavecchia, S. (2018). A heuristic algorithm for DJ cue point estimation. In *Proceedings Sound and Music Computing (SMC) Conference*.

Smith, J. B. L., Burgoyne, J. A., Fujinaga, I., De Roure, D., and Downie, J. S. (2011). Design and creation of a large-scale database of structural annotations. In *International Society of Music Information Retrieval Conference (ISMIR)* (Vol. 11, pp. 555–560).

Sonnleitner, R., Arzt, A., and Widmer, G. (2016). Landmark-based audio fingerprinting for DJ mix monitoring. In *Proceedings International Society for Music Information Retrieval Conference (ISMIR)* (pp. 185–191).

Van den Oord, A., Dieleman, S., and Schrauwen, B. (2013). Deep content-based music recommendation. In *The Conference on Neural Information Processing Systems (NIPS)* (p. 26).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *The Conference on Neural Information Processing Systems (NIPS)* (p. 30).

Veire, L. V., and De Bie, T. (2018). From raw audio to a seamless mix: Creating an automated DJ system for drum and bass. *EURASIP Journal on Audio, Speech, and Music Processing*, *2018*(1), 13.

Wang, A. (2006). The Shazam music recognition service. *Communications of the ACM*, *49*(8), 44–48.

TO CITE THIS ARTICLE:
Kim, T., Kim, J., Kim, H., & Nam, J. (2026). Raveform: A Dataset of Metrical and Functional Structure Annotations for EDM Tracks in DJ Mixes. *Transactions of the International Society for Music Information Retrieval, 9*(1), 131–143. DOI: https://doi.org/10.5334/tismir.288

Submitted: 10 June 2025 **Accepted:** 9 January 2026 **Published:** 20 April 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/.

Transactions of the International Society for Music Information Retrieval is a peer-reviewed open access journal published by Ubiquity Press.