# A COMPUTATIONAL MODEL FOR PREDICTING PERCEIVED TENSION IN ELECTRONIC DANCE MUSIC BUILDUPS

**Stefanos BILIOUSIS**[1] and **Sofia DAHL**[1]

[1]*Department of Architecture, Design and Media Technology*, **Aalborg University**, Copenhagen, Denmark

## ABSTRACT

This study presents a computational model of perceived musical tension in Electronic Dance Music (EDM) build-up sections, grounded in the Gestalt principles of pitch proximity, timbral similarity, and temporal proximity. The model derives a composite Tension Index from audio features representing pitch, timbre, and temporal density. We evaluated the model in a continuous-rating study in which 14 listeners rated twelve extended excerpts spanning breakdown, build-up, and drop. Continuous listener ratings were compared with the model using Pearson correlation. Accounting for a 2.0 s response lag, the model correlated positively with listener ratings (mean $r = .57$, 95% bootstrap CI $[.51, .62]$, $p < .001$). Inter-rater reliability, assessed per track, was high (ICC(2,$k$) .85 to .95). After removing the shared linear trend, the model retained significant predictive power, although a linear-ramp baseline performed better on average, indicating that substantial track-specific tension remains unexplained. These findings show that Gestalt-based audio features capture a meaningful component of perceived tension in EDM break routines and establish a quantified benchmark for future genre-aware tension models.

## 1. INTRODUCTION

The perceived need for release, referred to as musical tension, is a fundamental dynamic, which is especially crucial and explicit in Electronic Dance Music (EDM) [1, 2]. The genre's producers actively manipulate tension and release with the purpose to elicit a powerful emotional and physiological response on the dance floor [3, 4], as well as reported in laboratory settings [5, 6].

The quintessential structure of modern EDM centers on the large-scale "build-up" and "drop" cycle. The build-up is a dedicated section, usually 8 to 16 bars, that systematically introduces musical elements to heighten anticipation of the subsequent drop, which provides the cathartic release as the main beat and bassline return with full force. The perceived impact of the drop is magnified by the preceding build-up [7]. Motion capture studies have reported more vertical movement and higher level of inter-personal synchronization after the drop [5].

*Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

While studies on musical tension typically have been based on melodic or harmonic consonance, dissonance or roughness [8, 9], EDM typically uses continuous pitch glides or filter effects [2] calling for other theoretical foundations. Music cognition relies on active organizational processes such as Auditory Scene Analysis [10], where events are grouped according to Gestalt principles [10–13] such as Proximity (grouping elements that are close in time or pitch); Similarity (grouping elements that share features like timbre or loudness); and Good continuation (outlining the smoothest, most continuous contour). We here argue that an EDM build-up creates tension by utilizing production techniques to systematically and deliberately challenge these grouping principles, forcing a state of increasing instability and "perceptual strain." In this paper, we present and evaluate a model of perceived tension, based on the Gestalt grouping principles proximity and similarity commonly violated in EDM build-ups.

## 2. RELATED WORK

The idea of tension and release in music, and whether expectations are fulfilled or violated, have been the basis for several well-known theories on music and emotion [14–16]. Drawing on Meyer's earlier application of Gestalt principles to music [14], Narmour's Implication-Realization (I-R) model [15] proposed that listeners form melodic expectations based on registral direction and interval size. Small intervals imply continuation of the current pitch direction, invoking the Gestalt principle of good continuation, whereas large intervals imply a reversal of direction and a return to proximity. When these implications are denied rather than realized, perceptual tension arises. While grounded in Gestalt, the model, as other models departing from discrete pitch sequences in tonal, notated music (c.f. [13, 17, 18]), is less suitable for the continuous changes in timbre, pitch, and tempo used in EDM [2]. Solberg describes how producers employ a standard toolkit to create the characteristic tension-building effects during build-ups [1]:

**Filtering** High-pass filter sweeps progressively remove low frequencies, creating a sense of spectral thinning and anticipation.

**Rhythm** Accelerating snare rolls increase event density, pushing the listener toward the expected climax.

**Pitch** Synthesized rising effects with continuously ascending pitch ("uplift" or "riser") increase arousal and direct attention forward.

**Additional techniques** Increasing reverb decay, momentary silences, or strategic volume drops immediately before the drop to maximize its impact by contrast.

These production choices show how an EDM build-up purposefully establishes and then delays the expected climactic drop event, thereby increasing arousal. The toolkit aligns with Huron's ITPRA (Imagination-Tension-Prediction-Reaction-Appraisal) model [16], describing how listeners' imagination of what will happen in the music, leads to increased tension during the build-up. Depending on whether the prediction made for the upcoming event is right or not, there is a fast reaction followed by a slower appraisal. In the specific case of EDM break sections, the anticipation of the drop triggers an initial release of dopamine in the brain's reward centers [19]. The drop itself serves as the powerful resolution that fulfills this chemically primed expectation, resulting in a euphoric dopamine rush. This effect is further amplified by contrastive valence: the negative, tense state of the build-up makes the subsequent positive release of the drop feel more potent [16] and pleasurable [20, 21]. Although some studies have investigated the psychophysiological responses to EDM by comparing movement [6] or skin conductance [20] with averages of extracted music features such as root mean square amplitude, spectral flux and brightness, we know of no temporal models of tension during EDM build-ups.

Computational models on expectation [18] and tonal tension [17] have explored probabilistic [18] as well as geometric approaches [17, 22, 23]. From the background of information theory, Pearce and Wiggins formulated the cognitive model of musical expectation IDyOM, performing well in experiments of pitch prediction and segmentation [18]. MorpheuS [22] uses a three-dimensional model to represent tonality and dissonance, distance to key and tonal movement between chords in order to generate long-term tension and musical structure. Navarro-Cáceres et al. modeled the instantaneous tonal tension profiles of chord progressions based on tonal distances, dissonance, voice leading and hierarchical tension [23]. These models all have merit, but the focus on harmony and tonality fail to take the continuous processes and textural changes central to EDM into account [2]. Instead timbre may be more suitable. Farbood and Price found that features such as inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness to particularly contributed to perceived tension in short stimuli [24].

Farbood proposed a parametric, temporal model for tension based on harmony, pitch height, melodic expectation, dynamics, onset frequency, tempo, meter, rhythmic regularity, and syncopation. The musical parameters were evaluated within a moving perceptual window and summarized in terms of trend salience over time [25]. By fitting this model to continuous tension ratings for complex Western classical stimuli, Farbood demonstrated that perceived tension can be captured with high correlations when appropriate time scales and feature slopes are taken into account [25], establishing the weighted-sum-of-features paradigm that underlies much subsequent work on computational tension modeling.

More recently, Barchet et al. introduced TenseMusic, an automatic prediction model that brings Farbood's temporal framework into an audio-based MIR setting [26]. Given a musical audio file as the only input, TenseMusic uses state-of-the-art MIR techniques to extract six global descriptors: loudness, pitch height, tonal tension (derived from tonal-space models), roughness, tempo, and onset frequency—and then applies a Lasso-regularized regression on their temporal slopes to predict continuous tension ratings collected for 38 Western classical pieces [26]. Taken together, these approaches share several core assumptions. First, they operationalize musical tension as a scalar time series derived from a small set of audio or symbolic features whose temporal evolution (increases or decreases) is assumed to track listeners' changing sense of tension [25, 26]. Second, the feature sets are largely motivated by broad psychoacoustic and tonal considerations of dynamics, harmonic distance, roughness, tempo, and onset density—rather than by a genre-specific account of how particular musical practices generate tension. Finally, existing models have been developed and validated predominantly on Western classical or tonal art-music corpora, where harmonic motion and dynamic shaping are primary carriers of tension, with comparatively less attention to genres such as EDM where timbral and rhythmic manipulation can dominate while harmony remains static or plays a secondary role [17, 25, 26]. In relation to TenseMusic and Farbood's model, our approach can thus be viewed as a genre and mechanism specific specialization.

## 3. MODEL

### 3.1 Track Selection and Pre-Processing

The development and analysis of the model [1] were conducted on a curated corpus of twelve commercially produced EDM tracks (see Table 1), selected from genres with prominent build-up sections: progressive house, bass house, techno, trance, and big room house. Each track features parts of the buildup leading to a clear 8 or 16-bar build-up and then into the main drop. An example from TrackID 10 is shown in Figure 2, where panels (A) to (E) illustrate the waveform, the three normalized features, and the composite Tension Index.

All excerpts underwent a standardized pre-processing pipeline prior to feature extraction. Each audio file was loaded using the librosa library, converted to a monophonic signal, and resampled to a sampling rate of 48000 Hz. The audio was processed in short, overlapping frames. STFT frame length was set at 4096 samples and hop length at 512 samples. The analysis was performed on short, overlapping frames to create a time-series for each feature.

### 3.2 Pitch Proximity Feature (The Riser)

The violation of pitch proximity, as exemplified by the continuously ascending pitch of a synth riser was tracked using the YIN fundamental frequency $F_0$ estimation algo-

[1] https://github.com/stefbil/tensionmodel.

| TrackID | Artist | Song | Part |
|---|---|---|---|
| 01 | Steve Aoki & Cascada | Everytime we touch | 0:40 - 1:27 |
| 02 | ZERB & The Chainsmokers (feat. Ink) (Argy and Omnia Remix) | Addicted | 1:13 - 2:10 |
| 03 | Argy & MEDUZA (feat. PollyAnna) | Melodia | 1:22 - 2:03 |
| 04 | Crankdat | B.T.W | 1:01 - 1:38 |
| 05 | Crankdat | Back to you | 3:08 - 3:47 |
| 06 | OneRepublic & Alesso | If I lose myself | 2:05 - 2:47 |
| 07 | DubVision | No more | 2:31 - 3:18 |
| 08 | Dimitri Vegas & Like Mike | Ocarina | 2:56 - 3:45 |
| 09 | 22Bullets | Remote Control | 2:22 - 2:56 |
| 10 | Armin van Buuren & Gryffin | What took you so long | 2:38 - 3:36 |
| 11 | Gareth Emery & Standerwick (feat. HALIENE) | Saving Light | 2:44 - 3:28 |
| 12 | Canonblade | Debug | 3:36 - 4:33 |

Table 1. EDM tracks used in analysis and rating experiment.

rithm [27] (panel B in Figure 2) . This implementation specifically targets the most stable tonal element by first applying a harmonic-percussive source separation (HPSS) and running the YIN algorithm only on the resulting harmonic component of the signal. The algorithm estimates $F_0$ for each analysis frame. Unvoiced frames (those with low periodicity) were set to Not-a-Number to be excluded from later calculations. The raw $F_0$ time-series was first smoothed using a NaN-aware moving average filter with reflect padding to reduce spurious fluctuations.

### 3.3 Timbral Similarity Feature (The Filter Sweep)

The violation of timbral similarity caused by the timbral brightening effect of a high-pass filter sweep was quantified by the spectral centroid [28, 29], calculated as the weighted mean of the frequencies present in the spectrum, where the magnitude of each frequency serves as its weight [29]:

$$C = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (1)$$

where $x(n)$ is the magnitude of frequency bin $n$, and $f(n)$ is the center frequency of that bin. The spectral centroid has a strong perceptual correlation with brightness [28, 30]. As a high-pass filter's cutoff frequency rises, low-frequency energy is removed and the centroid shifts upward (panel C in Figure2). Spectral centroid was calculated per frame using and smoothed with a NaN-aware moving average filter.

### 3.4 Temporal Proximity Feature (The Snare Roll)

An increase in temporal proximity, or decrease in Inter-Onset Intervals (IOI) characterizes rhythmic acceleration, such as a snare roll [1]. The feature (panel D in Figure2) is calculated through a multi-step process where precise time points of percussive events in the audio signal are identified with and onset detection function. The time difference between each consecutive pair of detected onsets yield a series of IOIs. The feature is computed as the inverse of the IOI: $IOI^{-1}$. To generate a continuous time-series, this $IOI^{-1}$ value is assigned to all analysis frames that fall within a symmetric time window (e.g., 2.0 seconds, user-configurable) centered on the start of that IOI. This creates a step-function-like series where the rhythmic density value is updated at each new onset. This step-function series is then smoothed with a NaN-aware moving average filter to blend the discrete changes.

### 3.5 Composite Tension Index

The Composite Tension Index (bottom panel in Figure2), is a single time-series, $Tension(t)$, representing the predicted overall tension at each moment in time as a combination of the three feature streams. Before combination, each feature stream (pitch, centroid, $IOI^{-1}$) was independently normalized using a robust percentile-based method, scaling the 5th percentile to 0 and the 95th percentile to 1. This prevents outlier values from dominating the feature's contribution. The final index is computed as a NaN-aware weighted average, not a simple sum. Then, for each time frame $t$, the model calculates the tension as [25]:

$$Tension(t) = \frac{\sum_{i=1}^{3} w_i \cdot F_{i,norm}(t)}{\sum_{i=1}^{3} w_i \cdot M_i(t)} \quad (2)$$

where $F_{i,norm}(t)$ is the normalized value of feature $i$ at time $t$, $w_i$ is its normalized weight, and $M_i(t)$ is a mask that is 1 if the feature is valid (not NaN) at time $t$ and 0 otherwise. For the purposes of having all the features to equally contribute to the final curve, the weights were set to unity ($w_1 = w_2 = w_3 = 1$). The implications and limitations of this assumption are explored in Section 5. The final Tension(t) curve is then typically scaled or normalized for clear visualization.

## 4. EVALUATION

The model's normalized Tension Index curves for twelve extended EDM build-up sections (part of breakdown, full buildup, part of the drop) were evaluated against listeners' ratings of tension for the tracks in Table 1. Our hypothesis was that average rated tension for a track would correlate with that of the normalized Tension index curve.

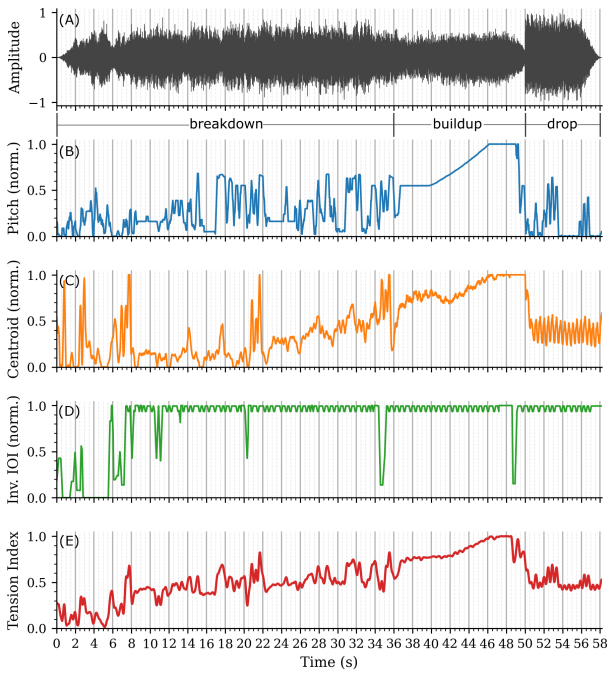



Figure 1. From top to bottom (A): raw waveform amplitude for TrackID 10, (B): normalized pitch feature, (C): normalized spectral centroid feature, (D): normalized inverse inter-onset interval $IOI^{-1}$, (E): the composite musical tension index Tension(t).

A sample of $N = 14$ participants (aged 21 to 56 years) took part in the online rating experiment. Recruitment was conducted through public posts on the Auditory and ISMIR mailing lists and social media. No participants reported hearing impairment, six participants reported music production experience.

### 4.1 Procedure

After following a link to the online listening test participants were informed of the definition of tension as perceived need for release and the use in electronic dance music. A gif illustrated the task of rating perceived tension over time by moving a vertical digital slider up and down. Participants were informed about the test procedure before giving their concent by clicking the "start" button. After demographic questions about musical training, familiarity, age and listening conditions, participants underwent a brief training where they could experience moving the vertical digital slider "UP as tension increases and DOWN as it decreases" for two buildups, seeing their tension curve and having the possibility to redo. After training, participants proceeded with the actual test where they continuously rated the tracks listed in table 1 in random order, receiving visual feedback but without being able to redo the rating. The test took about 15 minutes to complete.

### 4.2 Analysis

Slider positions were logged on every change, resulting in an event-driven record of approximately 495,000 tension values. Each participant-track series was resampled onto a 0.5-s grid using last-value interpolation. The model output was averaged within the same bins and smoothed with a centered 2-s moving average. Each participant's rating series was min-max normalized to $[0, 1]$ before computing the mean listener curve for each track.

Because continuous self-report responses lag the eliciting audio, we estimated a group-level response lag by shifting listener series relative to the model over lags from 0 to 4 s in 0.5-s steps, selecting the lag that maximized the mean Fisher-$z$-transformed correlation across all participant-track series. All subsequent model-listener comparisons were evaluated at this lag.

Inter-rater reliability was quantified per track using ICC(2,$k$) as the primary reliability estimate for the averaged rating curve, with Cronbach's $\alpha$ reported as a complementary measure of internal consistency.

Model-listener agreement was assessed at two levels. For inferential testing, we computed the Pearson correlation between the model Tension Index and each participant's own rating series (138 participant-track correlations), Fisher-$z$ transformed the values, averaged them within participant, and tested the resulting $N = 14$ participant-level means against zero (one-sample $t$-test with 95% bootstrap confidence intervals from 10,000 participant resamples). For descriptive reporting, we additionally computed per-track Pearson $r_{\text{model}}$, Spearman $\rho$, and RMSE between the model curve and the mean listener curve. Reference analysis compared the model against a linear ramp baseline, mismatched model curves (permutation test), and a leave-one-listener-out noise ceiling, repeated after linear detrending (Figure 2).

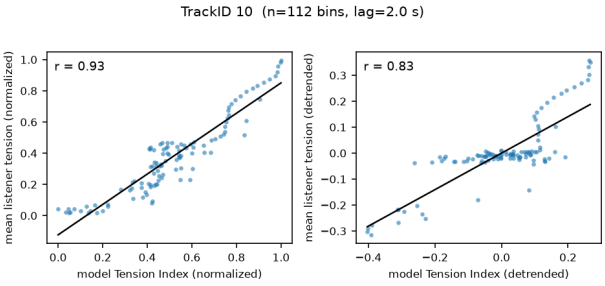



Figure 2. TrackID 10: model Tension Index versus mean listener rating across all time bins in the excerpt. Left panel: normalized values, showing strong agreement between model and listeners ($r = .94$). Right panel: linearly detrended values, isolating local deviations from the global rise ($r = .84$). Dots denote individual 0.5-s bins; lines indicate linear fits.

### 4.3 Results

Despite being a short test three participants ended early (two rated one track and one rated four tracks), resulting in 11-13 raters per track (Table 2). One participant reported EDM listening daily, two weekly, seven monthly, and four rarely or never.

#### 4.3.1 *Inter-rater reliability*

Inter-rater consistency was high (Table 2, middle): ICC(2,$k$) ranged from .85 to .95 and Cronbach's $\alpha$ from .90 to .97, indicating that the aggregated rating curves constitute a reliable perceptual ground truth against which the model can be meaningfully evaluated.

#### 4.3.2 *Response lag*

The cross-correlation analysis yielded a group-level optimal lag of 2.0 s at the 0.5-s bin resolution. Individual tracks exhibited highly distinctive optima (e.g., TrackID 09 improved at a 4-s lag), but since per-track lag selection capitalizes on chance, the group-level lag was adopted throughout.

#### 4.3.3 *Model-listener agreement*

The Tension Index correlated positively with individual ratings (participant-level mean $r = .57$, 95% bootstrap CI $[.51, .62]$, $p < .001$). Per-track model–mean correlations (Table 2) were highest for TrackID 10 ($r_{\text{model}} = .93$) and remained strong for TrackIDs 07 ($r_{\text{model}} = .80$), 04 ($r_{\text{model}} = .73$), 06 ($r_{\text{model}} = .73$), and 09 ($r_{\text{model}} = .73$). Weaker agreement was observed for TrackIDs 02 ($r_{\text{model}} = .37$) and 11 ($r_{\text{model}} = .37$). On TrackID 08, the model underperformed the ramp baseline ($r_{\text{model}} = .59$ vs. $r_{\text{ramp}} = .67$), but it still captures the post-drop tension release that a linear ramp cannot represent.

As Figure 3 illustrates for TrackID 10, the model Tension Index follows the overall trajectory of the mean listener rating but with more pronounced local fluctuations. During the central build-up segment, the model anticipates the listener rise but after the drop, the two curves converge. By comparison, the tension index for TrackID 09 display larger peaks and deviate substantially from the aggregated ratings.

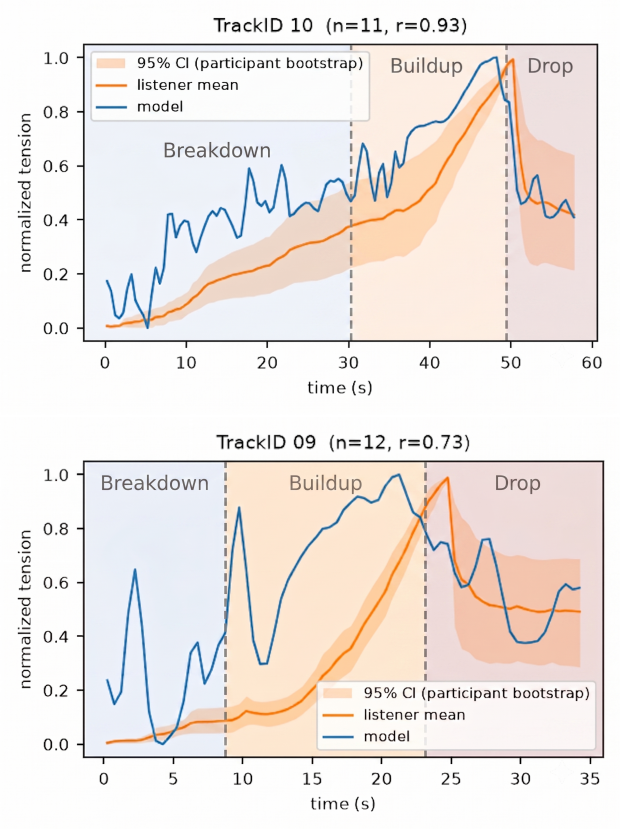



Figure 3. Comparison of the computational tension model output (blue line) against aggregated listener ratings (red line) for TrackID 10 (top) and 9 (bottom). Shading around listener mean curves indicate 95% confidence interval.

## 5. DISCUSSION AND CONCLUSION

The Gestalt-based Composite Tension Index correlated reliably with individual continuous ratings, and inter-rater reliability was high across all twelve excerpts. These results support the premise that pitch ascent, timbral brightening, and rhythmic densification jointly index perceived tension in EDM break sections. At the same time, baseline and detrending analyses qualify how much of this agreement reflects genre-generic structure rather than track-specific dynamics: a linear ramp exceeded the model on average, yet after removing the best-fitting linear trend the model retained significant residual predictive power, while the detrended leave-one-listener-out ceiling remained considerably higher. The track-shuffled permutation test was not significant, indicating that mismatched model curves can still explain a substantial fraction of the ratings. Together, these findings suggest that equal-weight combination of the three features captures both the shared rising arc of EDM build-ups and a smaller but reliable stratum of excerpt-specific tension fluctuations.

Per-track agreement (Table 2) clarifies where the model succeeds and where it diverges from listeners. The strongest case was TrackID 10, where the canonical co-occurrence of a pitch riser, filter sweep, and accelerating percussion aligns closely with the mean rating trajectory and remains high even after detrending. Similar but somewhat lower agreement was observed for TrackIDs 07, 04, 06, and 09, consistent with excerpts in which the three Gestalt violations intensify in parallel toward the drop [1].

The weakest model–listener correspondence occurred for TrackIDs 02 and 11, yet inter-rater reliability remained high, indicating that the poor fit is unlikely to reflect noisy ratings. In both cases the linear ramp correlated more strongly with listeners, suggesting that much of the perceived tension in these excerpts is carried by a steady global rise rather than by the local feature dynamics the model emphasizes. TrackID 02 (a vocal-forward progressive track) and TrackID 11 (a progressive trance track with prominent melodic and harmonic development) may therefore draw on tension cues outside the current feature set, such as lyrical expectancy, harmonic progression, or arrangement-level contrast, that are not fully captured by YIN pitch tracking, spectral centroid, and inverse IOI alone [24–26]. TrackID 11 also yielded the highest RMSE, consistent with larger moment-to-moment discrepancies between model and mean ratings despite monotonic agreement in Spearman $\rho$.

Several tracks illustrate mixed success. TrackID 09 achieved moderate aggregate agreement but, as Figure 3

| Participants | | Inter-rater reliability | | | Model vs. mean rating | | | |
|---|---|---|---|---|---|---|---|---|
| TrackID | $n$ | ICC(2,$k$) | 95% CI | $\alpha$ | $r_{\text{model}}$ | $\rho$ | RMSE | $r_{\text{ramp}}$ |
| 01 | 11 | .85 | [.76, .90] | .91 | .53 | .43 | .23 | .85 |
| 02 | 13 | .95 | [.93, .97] | .96 | .37 | .25 | .26 | .90 |
| 03 | 12 | .91 | [.87, .94] | .95 | .71 | .59 | .19 | .87 |
| 04 | 11 | .90 | [.83, .94] | .94 | .73 | .75 | .20 | .91 |
| 05 | 11 | .93 | [.89, .95] | .95 | .58 | .64 | .27 | .84 |
| 06 | 12 | .95 | [.93, .97] | .97 | .73 | .73 | .30 | .90 |
| 07 | 11 | .92 | [.89, .95] | .95 | .80 | .76 | .21 | .85 |
| 08 | 11 | .94 | [.91, .96] | .96 | .59 | .55 | .24 | .67 |
| 09 | 12 | .94 | [.90, .96] | .95 | .73 | .75 | .29 | .81 |
| 10 | 11 | .91 | [.86, .94] | .94 | .93 | .91 | .17 | .87 |
| 11 | 12 | .90 | [.86, .93] | .93 | .37 | .42 | .31 | .81 |
| 12 | 11 | .92 | [.88, .94] | .95 | .69 | .75 | .19 | .87 |

Table 2. TrackIDs and per track participation $n$ (left). Inter-rater reliability (middle) and descriptive per-track agreement between the model and the mean listener rating (right), evaluated at the group-level 2.0-s response lag.

shows, the model curve exhibits sharper local peaks than the mean listener rating during the central build-up segment. This pattern is compatible with listeners integrating cues over a longer response window: the per-track lag optimum for TrackID 09 was 4.0 s, whereas all reported comparisons use the group-level 2.0-s lag. TrackID 08 underperformed the ramp baseline yet retained strong detrended agreement, reflecting the model's sensitivity to the post-drop tension release that a monotonic ramp cannot represent. TrackID 01 showed the lowest inter-rater reliability together with only moderate model fit, suggesting that both rating dispersion and feature mismatch may contribute. TrackID 12 reached acceptable raw correlation but near-zero detrended agreement, implying that the composite curve tracks the excerpt's overall rise while missing reliable within-excerpt fluctuations; this is consistent with production styles in which rhythmic acceleration is weak or absent and timbral change dominates [2]. Decomposing the composite into its constituents supports these track-level patterns in general terms. The YIN-based pitch feature tracks canonical risers in most excerpts, inverse IOI captures snare-roll acceleration where present, and spectral centroid reflects filter sweeps but with high cross-track variability, as producers implement brightening through noise sweeps, synth modulation, and layered percussion that differ in spectral footprint [13, 24, 28]. Equal weighting is a simplifying assumption: listeners may prioritize dynamics, pitch, or timbre differently depending on arrangement context [25, 26]. Further limitations include the rhythmic descriptor's focus on onset acceleration rather than syncopation or groove disruption, and the treatment of features as independent despite possible masking (e.g., a broadband noise sweep obscuring pitch detail).

In conclusion, the proposed Gestalt-based model demonstrated that simple audio-derived features based on pitch proximity, timbral similarity, and temporal proximity capture a meaningful component of perceived tension in EDM break sections. Although the model significantly predicted listeners' continuous ratings, the remaining unexplained variance highlights opportunities for incorporating additional musical cues in future work.

Future work should estimate perceptually informed feature weights from larger annotated corpora, enrich the rhythmic and timbral representations, and evaluate whether learned combinations retain interpretability within the Gestalt framework. Complementary physiological or movement measures [5, 6, 20] may also help disambiguate cases where listeners agree reliably but the current acoustic features underexplain perceived tension. Extending evaluation across EDM subgenres and performance contexts will further clarify which production techniques the model generalizes to and where genre-specific extensions are required. Code and analysis scripts are available at github [2] .

## 6. AUTHOR CONTRIBUTIONS

The initial concept, implementation model, listening test and most of the analysis was done by author Biliousis supervised by Dahl. Both authors wrote and approved of the final manuscript.

## 7. REFERENCES


[1] R. T. Solberg, "Waiting for the bass to drop: Correlations between intense emotional experiences and production techniques in build-up and drop sections of electronic dance music," *Dancecult: Journal of Electronic Dance Music Culture*, vol. 6, no. 1, pp. 61–82, 2014.

[2] J. W. Smith, "The functions of continuous processes in contemporary electronic dance music," *Music theory online*, vol. 27, no. 2, 2021.

[3] W. Darling, "Creating tension & excitement in electronic dance music," EDM Tips, https://edmtips.com/tension-excitement/.

[4] S. Matla, "The advanced guide to tension and energy in electronic music," EDMProd, 2023, [Online]. Available: https://www.edmprod.com/tension/. Accessed: March 2026.


[2] `https://github.com/stefbil/tensionmodel`.

[5] R. T. Solberg and A. R. Jensenius, "Group behaviour and interpersonal synchronization to electronic dance music," *Musicae Scientiae*, vol. 23, no. 1, pp. 111–134, Mar. 2019.

[6] B. Burger and P. Toiviainen, "Embodiment in Electronic Dance Music: Effects of musical content and structure on body movement," *Musicae Scientiae*, vol. 24, no. 2, pp. 186–205, Jun. 2020.

[7] A. S. Turrell, A. R. Halpern, K. Bannister, D. Chai-Wi-Ting, and A.-H. Javadi, "Building the anticipation: How variation in tension mediates emotions in music," *Music Perception*, vol. 42, no. 3, pp. 256–268, 2025.

[8] M. Costa and M. Nese, "Perceived tension, movement, and pleasantness in harmonic musical intervals and noises," *Music Perception*, vol. 37, no. 4, pp. 298–322, 2020.

[9] I. Lahdelma, T. Eerola, N. Ahmad, J. Armitage, M. Clayton, B. Bhattacharyya, and N. Munsamy, "The perception of tension in harmonic intervals across indian and western listeners: the roles of psychoacoustics and musical expertise," *Journal of New Music Research*, vol. 54, no. 1, pp. 3–14, 2025.

[10] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, 05 1990.

[11] M. Reybrouck, "Gestalt concepts and music: Limitations and possibilities," in *Music, Gestalt, and Computing*, M. Leman, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 57–69.

[12] S. Sanyal *et al.*, "Gestalt phenomenon in music? a neurocognitive physics study with eeg," arXiv preprint arXiv:1703.06491, 2017.

[13] S. McAdams, "Timbre as a Structuring Force in Music," in *Timbre: Acoustics, Perception, and Cognition*, K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, and R. R. Fay, Eds. Cham: Springer International Publishing, 2019, pp. 211–243.

[14] L. Meyer, *Explaining Music: Essays and Explorations*. University of California Press, 1973.

[15] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago, IL: University of Chicago Press, 1990.

[16] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press, 2006.

[17] F. Lerdahl and C. Krumhansl, "Modeling tonal tension," *Music Perception*, vol. 24, pp. 329–366, 04 2007.

[18] M. T. Pearce and G. A. Wiggins, "Auditory Expectation: The Information Dynamics of Music Perception and Cognition," *Topics in Cognitive Science*, vol. 4, no. 4, pp. 625–652, 2012.

[19] V. N. Salimpoor, M. Benovoy, K. Larcher, A. Dagher, and R. J. Zatorre, "Anatomically distinct dopamine release during anticipation and experience of peak emotion to music," *Nature Neuroscience*, vol. 14, no. 2, pp. 257–262, 2011.

[20] R. T. Solberg and N. Dibben, "Peak Experiences with Electronic Dance MusicSubjective Experiences, Physiological Responses, and Musical Characteristics of the Break Routine," *Music Perception*, vol. 36, no. 4, pp. 371–389, Apr. 2019.

[21] R. T. Solberg and A. R. Jensenius, "Pleasurable and Intersubjectively Embodied Experiences of Electronic Dance Music," *Empirical Musicology Review*, vol. 11, no. 3/4, pp. 301–318, 2016, num Pages: 18.

[22] D. Herremans and E. Chew, "MorpheuS: Generating Structured Music with Constrained Patterns and Tension," *IEEE Transactions on Affective Computing*, vol. 10, no. 4, pp. 510–523, Oct. 2019.

[23] M. Navarro-Cáceres, M. Caetano, G. Bernardes, M. Sánchez-Barba, and J. Merchán Sánchez-Jara, "A Computational Model of Tonal Tension Profile of Chord Progressions in the Tonal Interval Space," *Entropy*, vol. 22, no. 11, p. 1291, Nov. 2020.

[24] M. M. Farbood and K. C. Price, "The contribution of timbre attributes to musical tension," *The Journal of the Acoustical Society of America*, vol. 141, no. 1, pp. 419–427, Jan. 2017.

[25] M. M. Farbood, "A parametric, temporal model of musical tension," *Music Perception*, vol. 29, no. 4, pp. 387–428, 2012.

[26] A. V. Barchet, J. M. Rimmele, and C. Pelofi, "Tensemusic: An automatic prediction model for musical tension," *PLOS ONE*, vol. 19, no. 1, p. e0296385, 2024.

[27] A. de Cheveigné and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.

[28] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1493–1500, 1978.

[29] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.

[30] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid?" *Acta Acustica United with Acustica*, vol. 92, no. 5, pp. 820–825, 2006.