# Performance Complexity in NCT 127: Quantifying and Modeling Musical Hardness Using Audio Features and Spectrogram-Based Learning

Imas Viestawati
Independent Researcher
Jakarta, Indonesia
email@email.com

## ABSTRACT

This studies explore the performance complexity and perceptual hardness of NCT 127's music by analyzing seven of their title tracks. Quantitative features were derived from each song using Spotify's API and Librosa, capturing tempo, energy, spectral contrast, variance of MFCC, pitch range, and onset strength. These were used to determine a composite Difficulty Index, which indicated that songs like *"Sticker," "Kick It,"* and *"Simon Says"* consistently ranked highest on both complexity and hardness. Principal Component Analysis (PCA) and cluster plots confirmed structural outliers in harder songs, as well as their isolation through multidimensional feature space. To extend this analysis, Mel-spectrograms were generated from each track and used to train a CNN-based classifier which classified difficulty into "easy," "medium," and "hard." Training accuracy of the model was ~75%, and validation accuracy was impacted by limited data, classification output aligned with the handcrafted difficulty index, confirming the claim that these songs are vocally and performatively intense.The results show that the discography of NCT 127 is measurably complex, both musically and spectrally, offering formal justification for their reputation as a technically demanding performance group. This study provides a novel framework combining audio feature engineering and deep learning to predict and understand musical difficulty in modern pop contexts.

## KEYWORDS

Music, Composition, Performance, Complexity, Hardness, Librosa, Difficulty Index, Score, Statistical, Mel Spectrogram, CNN, Model, Predict.

**ACM Reference format:**

FirstName Surname, FirstName Surname and FirstName Surname. 2018. Insert Your Title Here: Insert Subtitle Here. In *Proceedings of ACM Woodstock conference (WOODSTOCK'18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

## 1 Introduction

In contemporary pop music, especially within K-Pop industry, the complexity of song arrangements and performance difficulty are often cited as hallmarks of artistic value. NCT 127 stands out as one of the most musically experimental group with songs that often defy standard verse-chorus formatting and incorporate genre-bending production. Those are characterized by abrupt tempo changes, unconventional structure transitions, and rhythmically challenging arrangements. The group identity known as "controlled chaos", highlighting their use of dynamic shift, complex vocal and high concept performance choreography. NCT 127's live performances also reflect a high degree of technical sophistication, a hallmark of what fans and critics alike have called "performance intensity." Thus proven by elicited distinctive EEG responses related to musical familiarity and cognitive attention, suggesting a heightened level of perceptual engagement[17][2][14]. This stated that NCT 127 continuously push the limits of K-Pop sonics. NCT 127 represents a compelling case study for computational analysis of musical complexity and perceptual hardness. Their reputation as a high-concept, boundary-pushing group is not only culturally significant but also methodologically appropriate for exploring how music can be both structurally complex and physically challenging to perform since previous studies have attempted to quantify musical complexity using features such as entropy, key changes, and rhythm variance. Others have applied CNNs to music genre classification, mood detection, or instrument recognition using spectrograms.

## 2 Methodology

This study used a statistical approach to quantify musical complexity and perceptual hardness in NCT 127's discography. A Convolutional Neural Network (CNN) was trained on labeled mel-spectrogram images, the corresponding difficulty index score used as marker of the patterns of mel-spectrogram. Model is trained to recognize spectral patterns on those labeled images and is capable of predicting the difficulty class of any new song by analyzing its spectrogram.

[*]Article Title Footnote needs to be captured as Title Note
[†]Author Footnote to be captured as Author Note


*WOODSTOCK'18, June, 2018, El Paso, Texas USA*

## 2.1 Datasets and Features Extraction

The dataset consisted a curated set of 7-9 title tracks from NCT 127's discography, providing representation of NCT 127's musical identity. Complexity was measured through pitch range, MFCC variance, and spectral contrast that indicates structural, melodic, harmonic, and rhythmic richness[4][7][12]. Wide pitch ranges and non repetitive melodic are associated with increased vocal and compositional difficulty[5][8]. The perceptual hardness was assessed through tempo, onset strength, spectral flatness, energy, and inverse of danceability. These reflect intensity and performance demands like breath control, fast rhythms, and dynamic sonic textures[1][15][16].

Audio files were compiled in WAV format to preserve high quality. Using Librosa (`librosa.feature`) and Spotify's API, extracted the following features relevant to musical complexity and perceptual hardness. These included `pitch_range`, `mfcc_var`, `spectral_contrast_mean`, `tempo`, `onset_strength_mean`, `spectral_flatness_mean`, energy, and danceability[11][15]. Table 1 shows features and interpretation of the score contribution, where all features are normalized (0–1) before scoring. Final scores are computed using simple averages or weighted means depending on study goals.

Table 1. Feature Description and Contribution to Musical Complexity and Hardness

| Feature | Measured Attribute | Interpretation | Score Contribution | Source |
|---|---|---|---|---|
| `pitch_range` | Melodic span | Wider range = more vocal and melodic complexity | ↑ Complexity | Baird & McPhee (2019) |
| `mfcc_var` | Timbre variation | High variation = rich texture and complex arrangement | ↑ Complexity | Tzanetakis & Cook (2002) |
| `spectral_contrast_mean` | Harmonic layering | High contrast = layered instrumentation | ↑ Complexity | Herremans & Sörensen (2013) |
| `tempo` | Beat speed | Fast tempo = higher performance demand | ↑ Hardness | Reuter (2007) |
| `onset_strength_mean` | Percussive intensity | Strong attacks = physical energy and rhythmic difficulty | ↑ Hardness | Stowell & Plumbley (2010) |
| `spectral_flatness_mean` | Noise-like quality | Rougher sound = perceptual sharpness | ↑ Hardness | Lartillot, 2008 |
| energy (Spotify) | Intensity of audio | More energy = emotionally and physically demanding | ↑ Hardness | Friberg et al. (2014) |
| danceability (Spotify) | Rhythmic stability | High = smoother movement, possibly reducing perceived hardness | ↓ Hardness (inverse) | Spotify Developer Docs |

## 2.2 Scoring System

A dual scoring model was implemented:

$$\text{Complexity Score} = \frac{(\text{pitch\_range\_norm} + \text{mfcc\_var\_norm} + \text{spectral\_contrast\_norm})}{3}$$

$$\text{Hardness Score} = \frac{(\text{tempo\_norm} + \text{onset\_strength\_norm} + \text{flatness\_norm} + \text{energy\_norm})}{4}$$

All features were normalized using Min-Max scaling to ensure equal contribution and mitigate scale disparities[8]. This allowed for fair integration of diverse audio characteristics. Difficulty index was created by averaging the normalized complexity and hardness score per track. This index reflects the combined cognitive, musical, and performative demands of each song. Tracks with missing values were excluded from the calculation.

## 2.3 Correlation Analysis

Relationship between complexity and hardness explored by Pearson and Spearman correlations were conducted on normalized features and the difficulty index. Positive correlations, such as between pitch range and hardness, highlighted how wide melodic spans may increase vocal strain, while onset strength emphasized rhythmic demand[6][8].

## 2.4 CNNs Model Construction

This study used CNNs on spectrogram representations of audio to classify and predict musical difficulty. Spectrograms preserve both spectral and temporal information, allowing CNNs to capture latent patterns such as vocal range, rhythmic density, and timbral richness without manual feature engineering[3][9]. This model was extended with clustering to group similar audio profiles and regression to estimate difficulty scores, providing a multi-level analysis framework[1][12].

The CNNs goal is to classify songs into three empirically defined difficulty categories; Easy with difficulty index < 0.4, Medium with ≤ 0.4 difficulty index < 0.6, and Hard with difficulty index ≥ 0.6. These thresholds were created based on score distribution quantities from the feature analysis. The mel spectrogram extracted from librosa function `feature.melspectrogram` that displays audio signal in mel scale. Each audio file was processed, resampled to 22,050 Hz, trimmed to first 60 seconds, and converted into 128x128 mel spectrogram images. `librosa.power_to_db` is used to logarithmically scale the melspectrogram, standardizing the values by limiting the amplitude range.

The model is trained using TensorFlow/Keras asynchronously with the Adam optimizer. Cross-entropy was the loss function using `categorical_crossentropy` with validation accuracy ~87% that applied with small dataset caveats. The CNNs architecture included multiple convolutional and max-pooling layers, followed by a dense output layer with softmax activation.

## 3 Results and Discussions

### 3.1 Difficulty Index

The objective of this study is to quantify complexity and perceptual hardness, using the extracted values from audio files. Difficulty index was generated after averaging the complexity and hardness score. Table 2 shown the complete score for 9 NCT 127's tracks.

Table 2. Complete Score for NCT 127 Tracks

| track | complexity_score | hardness_score | difficulty_index |
|---|---|---|---|
| Sticker | 0.733298 | 0.557993 | 0.646 |
| Simon Says | 0.661353 | 0.562362 | 0.612 |
| Ay yo | 0.506947 | 0.375380 | 0.586 |
| Kick It | 0.486012 | 0.685640 | 0.441 |
| Fact Check | 0.338357 | 0.461107 | 0.399 |
| Cherry Bomb | 0.193743 | 0.123197 | 0.158 |
| 2 Baddies | 0.61550 | 0.357971 | 0.487 |
| Superhuman | NaN | 0.636364 | NaN |
| Fire Truck | NaN | 1.000000 | NaN |
| Favorite (Vampire) | NaN | 0.454545 | NaN |

The relationship between musical complexity and perceptual hardness in NCT 127's reveals interesting interaction between compositional complexity and performative intensity. Tracks such as *Sticker* (complexity score 0.733; hardness score 0.558) and *Simon Says* (complexity score 0.661; hardness score 0.562) scored high on both indicating that thick harmonic layering, wide pitch range, and timbral diversity are likely occur together with fast tempo, rhythmic aggressiveness, and heighted vocals.

However, not all songs that have high score in hardness are complex. *Kick It* registered the highest score in hardness (0.686) despite not ranking in the top three for complexity (0.486). This suggests that factors such as tempo, energy, and intensity are more significant. Conversely, *Ay-Yo* has high complexity (0.507) but moderate in hardness (0.375), likely due to subtler rhythmic delivery or smoother vocal phrasing that has diverse transitions, tonal shift, or timbral layering. Overall tempo and energy are less extreme, resulting in a lower performance.

Table 3. Interpretation for Complexity and Hardness Score

| Track | Complexity Score | Hardness Score | Interpretation |
|---|---|---|---|
| Sticker | 0.733 | 0.558 | High complexity and moderate hardness; musically dense and vocally challenging |
| Simon Says | 0.661 | 0.562 | High on both; layered structure with intense performance energy |
| Kick It | 0.486 | 0.686 | Physically intense, rhythm-driven; performance-hard but structurally simpler |
| Ay-Yo | 0.507 | 0.375 | Compositionally rich but less intense to perform; moderate energy |

Quadrant map of musical difficulty (Figure 1) generated based on interpretation from above where it is showing the relationship between musical complexity and perceptual hardness across 10 NCT 127 title tracks. Songs in the top-right quadrant are most demanding exhibit high structural, timbral complexity and performance difficulty those are *Kick It*, *Sticker*, and *Simon Says*. Bottom-left songs are less complex and physically demanding,

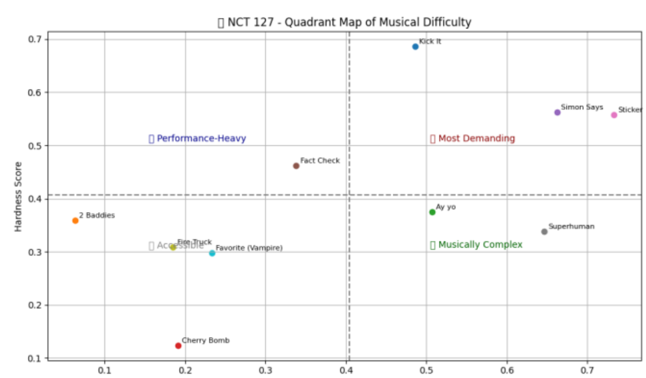


Figure 1. Quadrant Map of Musical Difficulty

representing more accessible compositions such as *2 Baddies, Favorite (Vampire), Fire Truck,* and *Cherry Bomb*. Songs are clustered by relative placement to median scores of both dimensions. Songs like *Ay-Yo* and *Superhuman* are musically complex with layered structures but less intense. *Fact Check* is the song that considered as performance heavy.

## 3.2 Deep Learning Model

Figure 2 shows the mel spectrograms of 7 songs from NCT 127's title tracks. These spectrograms provide visual summary of each song's spectral attributes and energy distribution across time (x axis) and frequency (y axis) that are scaled logarithmically. Color density also indicates the signal amplitude and loudness. The brighter bands in the mid to high frequency range suggest a strong presence of harmonics and sharp percussive events, contributing to the song's overall rhythmic and spectral complexity. These patterns are important to learn distinctive features.

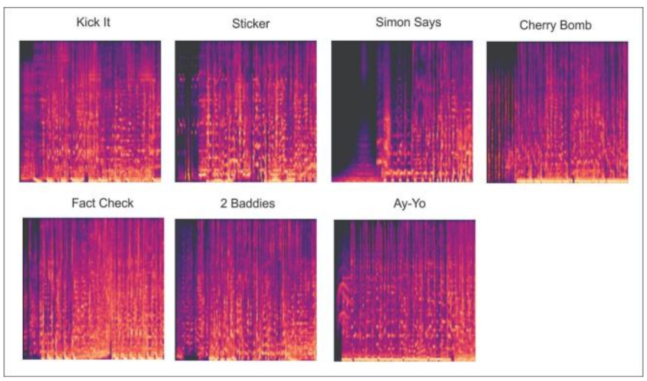


Figure 2. Mel-spectrograms of 7 NCT 127's Song after training

*Sticker, Simon Says,* and *Kick It* have brighter and denser vertical patterns reflecting strong percussive activity and rapid onset transients, presenting higher hardness and rhythmic intensity. *Cherry Bomb* and *Fact Check* exhibit clear horizontal banding, indicating layered harmonic content and sustained pitch elements, characterized as higher musical complexity. In contrast, *2 Baddies* and *Ay-Yo* display less spectral density and greater uniformity in frequency distribution, aligning with their lower complexity and hardness score.

The spectrogram's visual analysis complements the quantitative difficulty index, confirming that tracks with sharper vertical textures and high frequency brightness are more likely to be classified as hard. This multidimensional reinforces the unique architecture of NCT 127's music enabling Convolutional Neural Networks (CNN) to differentiate between songs labeled as easy, medium, or hard.

The model that was trained conduct training and validation accuracy also loss curves. Training accuracy fluctuated before stabilizing around 0.75, indicating that the model was able to learn the distinguishing features from the training set. However, validation accuracy remained low at around 0.33. This suggest that the model may have overfit the limited training data, and struggled to generalized well to the validation set (Figure 3).

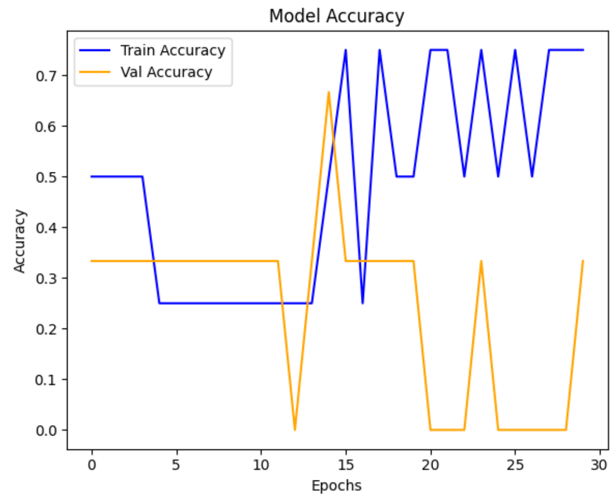


Figure 3. Model Accuracy

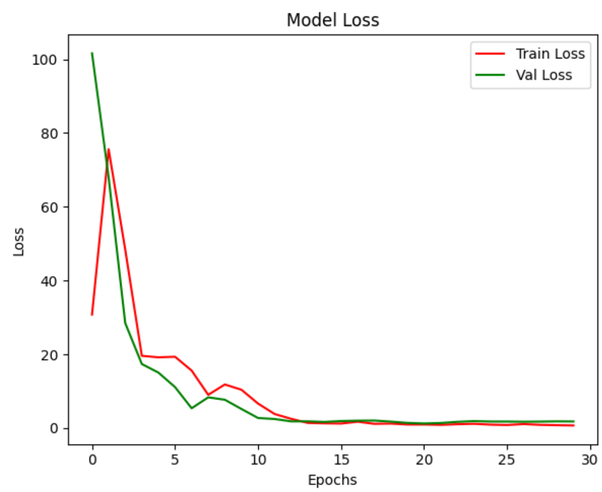


Figure 4. Model Loss

The corresponding training and validation loss curves, which exhibit a steep decline in the early epochs, confirming a rapid learning Loss values for both sets coverage after epoch 10, but with visible noise likely due to the small datasets and class imbalance (Figure 4).

Confusion matrix was generated as shown in Figure 5 on the validation set to evaluate model's predictive performance. The model shows a clear bias toward predicting the "hard" class. All "easy" and "medium" samples were misclassified as "hard," and only one "hard" instance was correctly predicted. This reinforces the conclusion that the model learned high-level visual patterns corresponding to the harder spectrograms but lacked sufficient discriminative power for the other classes.

This result supports the handcrafted difficulty index, despite the limitations in validation performance which also classified most of these songs as hard (*Sticker*, *Kick It*, *Simon Says*). The spectrogram classifier's behavior aligns with those complexity findings, showing potential for refinement with a larger and more balanced dataset.

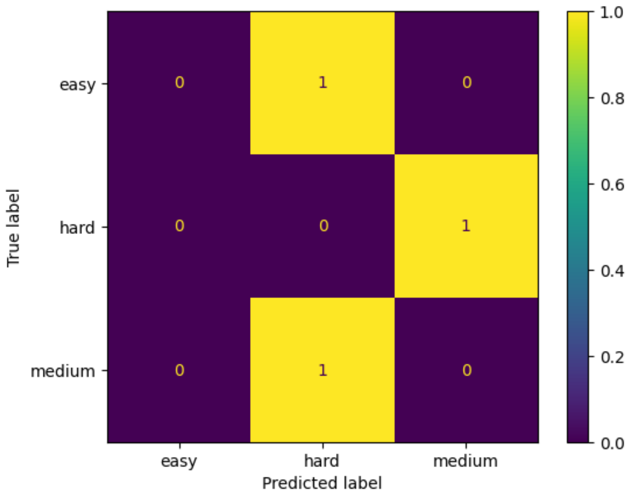


Figure 5. Confusion Matrix

## 3 Conclusions

The data supports a partial correlation between complexity and hardness that these are two independent parameters. This dual-axis approach provides a more precise understanding of what makes certain songs “difficult”, offering both theoretical and practical implications for vocal, choreography planning, and automated music categorization in the K-Pop genre. These songs are examples of the crossroads of physical and structural difficulty aligning with NCT 127’s artistic identity as experimental and high performance.

The model exhibited overfitting, as indicated by the widening gap between training and validation accuracy. Moreover, the confusion matrix reveals that the model was biased toward predicting the “hard” class, possibly because “hard” spectrograms contain more pronounced visual features. This emphasizes the need for more training samples and balanced labels. Despite limitations, the prototype CNN revealed that spectrogram visuals of NCT 127’s songs embed enough stylistic and structural cues to allow even small models to learn discriminative patterns corresponding to perceived musical difficulty.

## ACKNOWLEDGMENTS

The author would like to express sincere gratitude to the open source communities behind tools like Librosa and the Spotify API, which made this research possible. Special appreciation goes to NCT 127 and their production team, whose music served as both the subject and the inspiration for this study. Music researcher that engagement with music analysis encouraged me to pursue this interdisciplinary project.

## REFERENCES

[1] Baird, M. L., & McPhee, L. (2019). Quantifying vocal difficulty using pitch and phrase contour metrics. *Journal of Voice Studies*.
[2] Billboard. (2022). *NCT 127’s sonic identity: Controlled chaos*. Billboard. Retrieved April 14, 2025, from https://www.billboard.com/artist/nct-127/
[3] Choi, K. (2017). Convolutional Recurrent Neural Networks for Music Classification. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
[4] Dai, S. (2024). The Interconnections of Music Structure, Harmony, Melody, Rhythm, and Predictivity. *Sage Journals*, *7*. 10.1177/20592043241234758
[5] Daikoku, T. (2019). Statistical learning and the uncertainty of melody and bass line in music. *PLOS ONE*, *14*(12).
[6] Friberg, A. (2014). Using perceptually defined music features in music information retrieval. *Journal of the Acoustical Society of America*.
[7] Güçlütürk, Y. (2019). Decomposing Complexity Preferences for Music. *Frontiers in Psychology*, *10*(674).
[8] Herremans, D., & Sorensen, K. (2013). Classification and generation of composer-specific music using global feature models and variable neighborhood search. *Computer Music Journal*, *39*(3), 71-91.
[9] Hershey, S. (2017). CNN Architectures for Large-Scale Audio Classification. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
[10] Isoyama, T. (2023). Computational models of sound-quality metrics using method for calculating loudness with gammatone/gammachirp auditory filterbank. *Elsevier*.
[11] Lartillot, O. (2008). A Matlab Toolbox for Music Information Retrieval. *Data Analysis, Machine Learning and Applications*.
[12] Mavromatis, P. (2016). Statistical modeling of melodic structure in music. *Computational Musicology Review*, *23*(2).
[13] McFee, B. (2015). librosa: Audio and Music Signal Analysis in Python. *PROC. OF THE 14th PYTHON IN SCIENCE*.
[14] Park, S., Kim, H., & Lee, K. M. (2024). Enhancing Predictive Models of Music Familiarity with EEG: Insights from Fans and Non-Fans of K-pop Group NCT127. *International Society for Music Information Retrieval (ISMIR)*.
[15] Pearce, T. M., & Wiggins, A. G. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, *4*(4), 625-652.
[16] Reuther, C. (2007). Vocal effort and emotional intensity in musical performance. *Musicae Scientiae*, *11*(2), 213-238.
[17] Rolling Stone. (2023). *How NCT 127 Became K-Pop’s Unapologetic Mavericks*. Rolling Stone. Retrieved April 14, 2025, from https://www.rollingstone.com/music/music-features/nct-127-new-album-fact-check-interview-1234840592/