

УДК 004.8

**Сулейманов Є.С., аспірант спеціальності 121 Інженерія програмного забезпечення**  
**Науковий керівник: Байбуз О.Г., доктор технічних наук, професор, завідувач**  
**кафедри інженерії програмного забезпечення та інформаційних технологій**  
(Дніпровський національний університет імені Олеся Гончара, м. Дніпро, Україна)

## ЦИФРОВІ ОЗНАКИ МУЗИЧНИХ КОМПОЗИЦІЙ ДЛЯ НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ

Штучні нейронні мережі демонструють високу ефективність у задачах обробки музичного контенту. Залежно від типу архітектури та поставлених цілей, для навчання нейронні мережі можуть потребувати різноманітних представлень вхідних даних.

Розглянемо наступні категорії музичних ознак.

### 1. Часові ознаки

Waveform – звукова хвиля – числове представлення звукового сигналу, що показує його амплітуду в залежності від часу. Характеризується бітовою розрядністю амплітуди та частотою дискретизації, що відображає кількість вимірювань за одиницю часу.

Zero-Crossing Rate (ZCR) – частота перетину нуля – вимірює кількість разів, коли сигнал перетинає рівень нульової амплітуди за одиницю часу. Цей параметр дає змогу оцінити змінність або динамічність сигналу, оскільки високий ZCR зазвичай вказує на шум або сигнали зі швидкими змінами, в той час як низький ZCR властивий гармонічним або повільним сигналам, як музика або мова.

Root Mean Square Energy (RMSE) – середньоквадратична енергія сигналу – використовується для вимірювання енергії сигналу на основі середнього квадратичного значення амплітуд.

Short-Time Energy (STE) – енергія на короткому проміжку часу – дозволяє вивчити зміни енергії сигналу в часі, оскільки сигнал може мати різну енергію в різних фрагментах.

Envelope – огинаюча амплітуди – описує загальний контур зміни амплітуди звукового сигналу з часом і є корисною для аналізу модуляцій та динаміки звуку [1].

### 2. Спектральні ознаки

Spectrogram – спектрограма – часово-частотне представлення звукового сигналу, яке показує, як змінюється розподіл енергії або амплітуди по частотах із плином часу. Отримується шляхом застосування короточасного перетворення Фур'є (STFT) до сигналу, що дозволяє розбити його на короткі фрейми та обчислити спектр для кожного фрейму.

Spectral Centroid – спектральний центроїд – описує "центр мас" спектру. Використовується для оцінки того, де в спектрі розміщена більша частина енергії звуку. Якщо він знаходиться ближче до високих частот, сигнал буде здаватися яскравішим або світлішим (наприклад, для струнних інструментів). Якщо ж він ближче до низьких частот – сигнал буде сприйматися як темніший (наприклад, для басових інструментів).

Spectral Bandwidth – ширина спектра – міра розкиду частот, описує, наскільки розподілені частоти звуку в спектрі, тобто, наскільки "широкий" або "вузький" спектр. Ця характеристика є важливою для визначення текстури звуку та його яскравості.

Spectral Rolloff – спектральний ролоф – вимірює поріг частоти, нижче якої знаходиться певний відсоток енергії спектра (зазвичай обирають 85%). Цей параметр часто використовується для визначення тональної структури звуку: тобто, чи є сигнал тональним (як музика) чи нетональним (наприклад, шум або розмова).

Spectral Flux – спектральний потік – вимірює швидкість зміни енергії в спектрі аудіосигналу між двома послідовними часовими фреймами, допомагає виявити різкі зміни в аудіо, такі як удари, перехід між різними звуками або зміни в темпі музики.

Spectral Contrast – спектральний контраст – вимірює різницю між амплітудами високих і низьких частот спектра. Характеризує текстуру звуку: чи є звук м'яким (як голос або інструмент) чи жорстким (як шум або електронний звук).

Spectral Flatness – спектральна пласкість – вимірює, наскільки спектр звуку є "пласким" або "шумним". Вказує на те, чи є звук більш подібним до гармонійного сигналу (музика) або до шуму (негармонійний сигнал).

### 3. Тембральні ознаки

Mel-Frequency Cepstral Coefficients (MFCCs) – мел-частотні кепстральні коефіцієнти – набір коефіцієнтів, що моделюють тембральні особливості сигналу. Є представленням спектру звукового сигналу, яке зменшує кількість вимірюваних параметрів та враховує особливості сприйняття звуків людським слухом.

Chroma Features (Chromagram) – хроматичні характеристики (хромограма) – це представлення спектральної інформації, яке фокусується на інтервалах між нотами та їх гармонічними відносинами, і ігнорує точні частоти. Вони використовують 12 півтонів на октаву, що відображають музичні ноти, це дозволяє представити спектр музики за допомогою 12 бінів, які відображають рівні енергії для кожної з нот.

Tonnetz (Tonal Centroid Features) – тональний центроїд – це набір акустичних характеристик, які використовуються для аналізу гармонії та тональності музичних композицій. Він представляє музичний сигнал через чотири основні параметри, які описують гармонічні відносини між нотами в межах музичної тональності [2].

### 4. Ритмічні ознаки

Tempo (BPM – beats per minute) – темп (швидкість ударів на хвилину) – вимірює швидкість виконання музики. Окреслює ритмічний та динамічний характер твору.

Beat Tracking – відстеження ударів – встановлює основні ритмічні удари у музичному сигналі.

Rhythm Patterns – ритмічні шаблони – виявлення повторюваних ритмічних патернів у часі.

Onset Detection – визначення початку нових звуків – ідентифікує різку зміну енергії або спектральних характеристик, що сигналізує про початок нової ноти, удару або звукового фрагмента [3].

Налаштування параметрів для розрахунку оглянутих ознак має суттєве значення для досягнення ефективності навчання, тож експерт з нейронних мереж має робити вибір на підставі особистого досвіду, враховуючи складність та специфіку музичних творів, характер завдання, обрану архітектуру мережі, існуючі апаратні, програмні та часові обмеження.

#### Список використаних джерел:

1. Colin Raffel. Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching, Columbia University, 2016, 222 p.
2. Hugo Fastl and Eberhard Zwicker, Psychoacoustics. Facts and Models, Springer Berlin, 2007, 463 p.
3. J.-P. Briot, G. Hadjeres and F.-D. Pachet, Deep Learning Techniques for Music Generation, Computational Synthesis and Creative Systems, Springer, 2019, 310 p.