Feature Descriptions and Their Groupings

This document serves as a supplementary material for the research paper **"Semantic-Aware Interpretable Multimodal Music Auto-Tagging"** where the first section details all the features used in the study , their explanation and the second presents the features contained in each of the categories presented in the study.

1. Features Description

On this section we present the features' explanations which will be arranged based on their method of extraction.

Signal Processing

1. **Danceability**: Measures how suitable the music is for dancing. High values mean the music is very danceable with a strong, steady beat. Low values mean it's less suitable for dancing.
2. **Loudness**: The overall volume of the music. High values mean the music is loud. Low values mean the music is quiet.
3. **Chord Change Rate**: How often the chords change in the music. High values mean chords change frequently. Low values mean chords change less often.
4. **Dynamic Complexity**: The amount of variation in the music's volume. High values mean the volume changes a lot. Low values mean the volume stays more constant.
5. **Zero Crossing Rate**: How often the sound wave crosses the zero line. High values mean the music has more noise or sharp sounds. Low values mean the music is smoother.
6. **Chords Number Rate**: The proportion of different chords in the music. High values mean there are many different chords. Low values mean there are fewer different chords.
7. **Pitch Salience**: How clear the main pitch or note is. High values mean the notes are very clear and easy to hear. Low values mean the notes are less clear.
8. **Spectral Centroid**: Indicates whether the music sounds bright and sharp or dull and deep. High values mean the music sounds bright and sharp. Low values mean the music sounds dull and deep.
9. **Spectral Complexity**: The number of different tones or frequencies in the music. High values mean there are many different tones. Low values mean there are fewer tones.
10. **Spectral Decrease**: How quickly the higher frequencies fade away. High values mean the high sounds fade quickly. Low values mean the high sounds fade more slowly.
11. **Spectral Entropy**: Measures how varied the sounds are. High values mean the music has many different sounds. Low values mean the music has more distinct, fewer sounds.
12. **Spectral Flux (high, high middle, middle, low middle, low, mean)**: The amount of change in the sound over time. High values mean the sound changes a lot. Low values mean the sound is more consistent.
13. **Spectral Kurtosis**: Detects unusual noises in the music. High values mean there are strange or unusual noises. Low values mean the music is smoother.

14. **Spectral Roll Off**: The frequency below which most of the sound energy is found. High values mean the energy is concentrated in higher frequencies. Low values mean the energy is in lower frequencies.
15. **Spectral Spread**: The range of frequencies in the music. High values mean the music has a wide range of frequencies (noisy). Low values mean the music has a narrow range of frequencies (clear tones).
16. **Onset Rate**: The number of distinct notes or sounds per minute. High values mean there are many notes. Low values mean there are fewer notes.
17. **Spectral RMS**: The average loudness of the music. High values mean the music is loud on average. Low values mean the music is quiet on average.
18. **Spectral Skewness**: Measures energy around specific frequencies. High values mean there is a lot of energy around certain frequencies (like speech). Low values mean the energy is more evenly spread.
19. **Spectral Energybands (high, high middle, middle, low middle, low, mean)**: The amount of energy in specific frequency ranges. High values mean there is a lot of energy in certain ranges. Low values mean there is less energy.
20. **BPM (Beats per Minute)**: The speed of the music in beats per minute. High values mean the music is fast. Low values mean the music is slow.
21. **Pulse Clarity (high, high middle, middle, low middle, low, mean):** How easily in a given musical piece, or a particular moment during that piece, listeners can perceive the underlying rhythmic or metrical pulsation.
22. **Attack Slope (high, high middle, middle):** The attack phase is the period where the sound moves from silence (or a low level) to its full volume. The attack slope controls how quickly or slowly this increase happens.
23. **Attack Time (high, high middle, middle, low middle, low, mean):** Attack time is the amount of time it takes for the effect to reach its full impact after the input signal surpasses a specified threshold.
24. **Spectral Flatness (high, high middle, middle, low middle, low, mean):** Spectral flatness is typically measured in decibels, and provides a way to quantify how much a sound resembles a pure tone, as opposed to being noise-like.
25. **Key Strength**: Confidence that the key is the correct one.
26. **Chord Strength**: Confidence that the chord is the correct one.
27. **Chroma**: Represent the intensity of different pitches in music, capturing the harmonic and melodic aspects. 12 features in total.
28. **MFCC**: 13-in-total coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip.
29. **Entropy Clarity (high, high middle, middle, low middle, low, mean)**: Refers to the entropy of Pulse Clarity.
30. **Meter**: Refers to the meter of music.

Deep Neural Network (mid-level):

1. **Melodiousness/Melody**: How much does the music make you want to sing along?
2. **Articulation**: Are there more short, sharp sounds in the music?
3. **Rhythmic Complexity**: Is it hard to tap along to the music or figure out the beat?
4. **Rhythmic Stability**: How easy is it to march in time with the music?
5. **Dissonance**: Does the music has dissonant notes?

6. **Tonal Stability/Atonality**: How easy is it to identify the main key or notes in the music? Are there many key changes?
7. **Modality/Mode**: Does the music sound like it has more minor chords, does the music sound more “sad”?

Symbolic Knowledge (harmonic):

1. **Dominant**: This chord creates a sense of tension, often leading to a resolving chord.
2. **Subdominant**: This chord sounds peaceful and often moves to a tonic chord.
3. **Tonic**: This is the home chord that follows a dominant chord.
4. **Major Dominant**: A major dominant chord that creates a lot of tension.
5. **Aeolian**: Minor resolution, dark and melancholic dominant function.
6. **Mixolydian**: Modal and open-ended dominant function.
7. **Minor Fourth:** Smooth and expressive subdominant function.
8. **n-grams**: Contains in total 312 that correspond to bigrams and trigrams, as combinations of *Dominant, Subdominant, Tonic, Major Dominant, Aeolian, Mixolydian* and *Minor Fourth.*

Lyrical:

1. **Alliteration:** The occurrence of the same letter or sound at the beginning of adjacent or closely connected words.
2. **Assonance:** Resemblance of sound between syllables of nearby words, arising particularly from the rhyming of two or more stressed vowels, but not consonants.
3. **Consonance:** When two words have the same consonant sound following different vowel sounds.
4. **Rhyme Density:** How frequently rhymes occur in a set of lyrics. We extract the last two letters of each word (the syllables proxy), we want to calculate how often these "syllables" (last two letters) appear in the lyrics.
5. **Syllables per Word:** Average number of syllables per word.
6. **Total Syllables**: Total number all syllables found in the lyrics.
7. **Repetition Rate:** Represents the fraction (or rate) of words in the text that are repetitions. A higher repetition rate indicates a greater proportion of words are repeated within the text**.**
8. **Lexical Diversity:** Ratio of unique words to all words in a lyric.
9. **Words per Second:** Number of words per second.

2. Feature Groupings

User-Friendly

1. Brightness, Sharpness:
   - Spectral Spread
   - Spectral Centroid
   - Spectral Rolloff
   - Zero Crossing Rate
   - Loudness
   - Pitch Salience
   - Spectral Rms
   - Spectral Flux (high, high middle)

2. Danceability, Rhythm:
   - Danceability
   - Onset Rate
   - Rhythm Stability
   - Pulse Clarity (high, high middle, middle, low middle, low, mean)

3. Tension, Complexity:
   - Dynamic Complexity
   - Spectral Entropy
   - Spectral Complexity
   - Chord Change Rate
   - Dissonance
   - Chord Number Rate
   - Dominant
   - Articulation
   - Rhythm Complexity
   - BPM
   - Major Dominant
   - Mixolydian

4. Acoustic Smoothness:
   - Spectral Decrease
   - Atonality
   - Mode
   - Melody
   - Subdominant
   - Spectral Energyband (high, middle high, middle low, low)

- Spectral Flatness (high, high middle, middle, low middle, low, mean)
- Spectral Kurtosis
- Attack Time (high, high middle, middle, low middle, low, mean)
- Spectral Flux (middle, low middle, low, mean)
- Aelian
- Minor Fourth

5. Lyrical:
    - Alliteration
    - Assonance
    - Consonance
    - Rhyme Density
    - Syllables per Word
    - Total Syllables
    - Repetition Rate
    - Lexical Diversity
    - Words per Sec

<u>Domain-Expert:</u>

1. Spectral:
    - Spectral Spread
    - Spectral Centroid
    - Spectral Rolloff
    - Spectral Rms
    - Spectral Flux (mean)
    - Spectral Flatness (mean)
    - Spectral Kurtosis
    - Spectral Decrease
    - Spectral Entropy
    - Spectral Complexity

2. Harmonic:
    - Atonality
    - Mode
    - Pitch Salience
    - Melody
    - Subdominant
    - Dominant
    - Mixolydian

- Aeolian
- Dissonance
- Major Dominant
- Minor Fourth

3. Rhythmic:
    - Danceability
    - Pulse Clarity
    - Rhythm Complexity
    - Rhythm Stability

4. Sound Shaping:
    - Dynamic Complexity
    - Chord Change Rate
    - Chords Number Rate
    - Attack Time (mean)
    - Zero Crossing Rate
    - Articulation
    - Onset Rate
    - Loudness

5. Lyrical
    - Alliteration
    - Assonance
    - Consonance
    - Rhyme Density
    - Syllables per Word
    - Total Syllables
    - Repetition Rate
    - Lexical Diversity
    - Words per Sec

All-Features:

This category incorporates all the features presented in Section 1, with groups being the method of extraction as stated.

1. Signal Processing
2. Deep Neural Network (mid-level)
3. Symbolic Knowledge (harmonic)
4. Lyrical