BRUTE-FORCING MUSIC

Kosobutski V.A.

Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus

Churzina E.A. – scientific supervisor, lecturer

This article is about applying music and mathematics theory to generate a database of melodies using appropriate software to influence the current copyright system. This should help small independent songwriters to get protection from unfounded lawsuits. Also, at the current rate of music creation, we will soon exhaust every melody that can be used in popular songs, which will lead to the increase of the probability of being sued, posing a threat to both small and big musicians.

Current copyright laws are flawed and uncertain: they do permit independent creation, but most of the times are in favor of subconscious infringement. There is no need for two melodies to be an exact match to start a lawsuit, which leads to the limitation upon songwriters' freedom to create music.

Melody is defined by the correlations between frequencies of consecutive notes. Ratio for an octave is 2:1, octave consists of 12 semitones. Semitone ratio is $\sqrt[12]{2} \approx 1.05946$. Western music tradition defines standard pitch A440, or A4, which corresponds to an audio frequency of 440 Hz. Frequencies of all the other notes are calculated by multiplication of the standard pitch by semitone ratio required amount of times [1]. This means melodies are not affected by transpositions (shifting a melody by any interval will not change it).

The equation for the number of melodies:

$$N_m = R^L \quad (1),$$

where N_m is a number of melodies, R – range of notes, L – melody's length.

In theory, the number tends towards infinity but in practice is relatively small. There are 16.777.216 diatonic melodies that span over an octave and are 12 notes long. There are only 6.651 melodies that use 3 degrees of a scale and are 8 notes long. Or 243 melodies that use 3 degrees and are 5 notes long. Giving a monopoly on one of those 243 for 95 years, plus the lifetime, for one person might be harmful to society. The probability of two people accidentally choosing the same melody is increasing due to a large amount of small independent songwriters.

All these melodies can be represented as sequences of numbers where each number stands for a note's scale degree. In C major scale sequence 1-3-6-5 stands for C-E-A-G (Do-Mi-La-Sol). Under copyright laws, numbers are facts, and facts either have thin copyright or no copyright. This means melodies should be licensed under "Creative Commons Zero" license [2] (CC0 in the following) and all cases over a melody alone should be dismissed.

All that must be done to gain a right over intellectual property is to have it stored on a fixed tangible medium. This creates a possibility of going through all the permutations by generating them using MIDI software. After melodies were written down on a medium, CC0 is applied, and there is no way for any musician to sue over them.

During the last year, two people – Damien Riehl and Noah Rubin – were working on this problem. They've created a database of 337 billion melodies, and source code, written in Rust, for generating more is available in the repository on GitHub [3]. Already done:

major and minor scale, octave, length – 12;
major and minor scale, 13 pitches, length – 10;
chromatic scale, octave, length – 10.

$$8^{12} + 8^{12} + 13^{10} + 12^{10} = 337.214.809.545 \quad (2).$$

This database covers most of the popular songs but extending it requires exponentially more computing power, unavailable for an independent individual. It is not possible to generate every melody to copyright all of them with CC0, but what was done might be enough for a proof of concept. Work on this subject focuses on showing the absurdity of the current copyright system. Its aim is not to adjust to it but to change. At this stage, what is done is enough to protect more than 90% of existing songs.

One more problem is still unsolved: you can't overwrite copyright licenses of already existing melodies. They remained unaffected by this work. And will for at least 95 years since their creation. It is another reason to stop extending the database and start working directly with legal institutions.

The mentioned database does not take rhythm into account. Rhythm doesn't always matter in courts, but it's unpredictable and depends only on the civil judge's will while he can be not educated enough in the field of music. Note of length n can be broken down in p(n) different ways called partitions. If order matters, and it does, the sum becomes a composition. Then for every composition, the amount of possible melodies is:

$$N_m = R^{n_c} \quad (3),$$

where n_c stands for the amount of notes in the melody (numbers in a composition).

Using the binomial coefficient, it is possible to calculate how many times will any fixed n_c occur in the compositions:

$$N_r(n_c) = \binom{L-1}{n_c-1}; \; N_m(n_c) = \binom{L-1}{n_c-1} R^{n_c} \quad (4),(5),$$

where $N_r(n_c)$ is a number of possible rhythmic patterns with n_c elements, and $N_m(n_c)$ is a number of possible melodies.

The overall number of melodies for any positive integers L and R and for mentioned previously cases are:

$$N_m = \sum_{n_c=1}^{L} \binom{L-1}{n_c-1} R^{n_c} \quad (6).$$

$$N = \sum_{n_c=1}^{12} \binom{11}{n_c-1} 8^{n_c} = 251.048.476.872 \approx 8^{12} * 3,653 \quad (7),$$

for any diatonic scale, octave, length – 12.

$$N = \sum_{n_c=1}^{10} \binom{9}{n_c-1} 13^{n_c} = 268.593.608.192 \approx 13^{10} * 1,948 \quad (8),$$

for any diatonic scale, 13 pitches, length – 10.

$$N = \sum_{n_c=1}^{10} \binom{9}{n_c-1} 12^{n_c} = 127.253.992.476 \approx 12^{10} * 2,055 \quad (9),$$

for the chromatic scale, octave, length – 10.

Why this will not lead to plagiarism encouragement? Chord progressions will remain copyrightable. To define the general solution for given L and R, every solution for a rhythmic pattern (4) should include combinations of a set with R elements grouped into subsets with $d = \overline{1, R}$ elements. If there was a restriction on using only chords, then subsets should be with $d = \overline{2, R}$ elements. Then the amount of the combinations should be taken to the power of the amount of the notes in the melody:

$$N_C(n_c) = \left(\sum_{d=1}^{R} \binom{R}{d}\right)^{n_c} = (2^R - 1)^{n_c} \qquad (10),$$

where $N_C(n_c)$ stands for the number of possible chords and single notes for the melody with n(c) notes.

Merging equations (4) and (10) will provide the general solution for given L and R:

$$N_C = \sum_{n_c = 1}^{L} \left((2^R - 1)^{n_c} \binom{L-1}{n_c-1}\right) \qquad (11).$$

For clarity there are values of N_C and their ratios to related N_m:

$$N_C = 28.762 \approx N_m * 37{,}45;\ N_C = 1.065.151.889.408 \approx N_m * 7{,}61 * 10^5 \qquad (12),(13),$$

for R=3, L=5 and R=5, L=8 correspondingly.

References:
1. A Smoother Pebble: Mathematical Explorations / Donald C. Benson / Oxford University Press, 2003 – p. 54-56.
2. Creative Commons. 2020. CC0 - Creative Commons. Available at: https://creativecommons.org/share-your-work/public-domain/cc0/. [Accessed 15 April 2020].
3. All the Music LLC – Helping Songwriters Make All of Their Music. Available at: http://allthemusic.info/. [Accessed 15 April 2020].
4. A Smoother Pebble: Mathematical Explorations / Donald C. Benson / Oxford University Press, 2003 – p. 49-54.