Material and methods

We have downloaded classical music archive from … and ….

All midi files were transformed into comma separated data formats using midicsv (<http://www.fourmilab.ch/webtools/midicsv/>). The transformed data containing columns indicating the start, end and note in a piece of music. This study focused on the notes themselves, thus the track, instrutments and other information of the data were not used in further analysis. We processed the note data using a customized python script (github link) to generate 168 informative variables (Table 1) which captures some of the general features of this pies of music.

To obtain these variables, we have divided a piece of music into bins. A bin is defined as a time segment in the music where there is no change to any notes. Within each bin we calculate the number of occurance of a note (128 variables), the number of occurance of a note modulated by 12 (12 variables), and all possible pairwise distances between notes modulated by 12 (12 variables). We also calculated pairwise distances between notes between two adjacent bins modulated by 12 (12 variables). These variables were summed across an entire piece of music, then nomalized by minus the mean and divided by the standard deviation of each category of variable. We also calculated the mean and standard deviation of notes and bin durations across the entire piece of music (4 variables).

Table 1. Variables generated from midi files for analysis

|  |  |  |
| --- | --- | --- |
| Variable type | No. of variables | Interpretation |
| Frequency of modulated distances within bin | 12 | All pairwise distances between notes within a bin, modulated by 12, summed across an entire piece of music then normalized |
| Frequency of modulated distances between bin | 12 | All pairwise distances between notes in one bin versus its next bin, modulated by 12, summed across an entire piece of music then normalized |
| Frequency of modulated notes within bin | 12 | All notes in a bin modulated by 12, summed across an entire piece of music then normalized |
| Frequency of unmodulated notes within bin | 128 | All notes in a bin, summed across an entire piece of music then normalized |
| Bin durations across the entire music | 2 | Mean and standard deviation of bin durations |
| Notes across the entire music | 2 | Mean and standard deviation of all notes |