Source codes for Chroma/Chord-based Large-Scale Audio Indexing/Hashing are released now, together with some example datasets of music songs. More datasets can be provided under request. You can directly run programs by following the instructions in “readme.txt” and can freely modify programs to do any scientific experiments. The basic descriptions are listed below:

1. Indexing Based on Low-Level Music Features
   Based on investigating the statistical representation of acoustic short-time sequential correlations, a multi-probe histogram (MPH) is computed from each audio track, to provide a more adequate balance between scalability, robustness and discrimination ability. The idea of locality sensitive hashing (LSH) was applied to compute MPH from a sequence of chroma features. The major MPH bins of an audio track are in the top-\(n\) major MPH bins of its variants with a high probability. Based on the analysis of the order statistics (OS) of MPH bins, an adapted LSH approach is suggested to map MPHs to hash values. ([Codes and Data](#))

2. Indexing Based on Mid-Level Music Attributes
   Chord progressions (CPs) are exploited to realize accurate and meaningful summarization of music content and efficient organization of the database. The SVM\(^{hmm}\) model was adopted, SVM for per-feature chord recognition, and HMM for CP recognition. Through a modified Viterbi algorithm, \(N\)-best CPs are locally probed to generate a simple and descriptive chord progression histogram (CPH). Organizing songs in the layered tree-structure further helps alleviate the potential imbalance among buckets. ([Codes and Data](#))