



Kris Shaffer

1.2K Followers · About [Follow](#)



What is computational musicology?



[Kris Shaffer](#) Jan 20, 2016 · 4 min read

Computational musicology is a young, growing field — one in which I find myself increasingly involved. In addition to [The Lieder Project](#), I will be teaching my course on computational music analysis again this summer, and two students and I will be giving a presentation on it for the College of Music colloquium series later this spring. Since it is a young field, and draws on so many disciplines, I often find myself asked just what computational musicology is. As far as I know, there isn't a public-facing introduction

to the field. So with both my upcoming class and these informal interactions in mind, I thought I'd write one. So here goes...

In a nutshell, *computational musicology* — more-or-less synonymous with *computational music theory* or *music informatics* — is the use of computational methods and statistics to analyze musical structures (notes, chords, rhythms, etc., and patterns thereof). This combination of computation, statistics, and a domain of knowledge makes computational musicology a form of *data science*. However, due to the music theoretical aspect, computational musicology sits firmly within the *digital humanities* and focuses on the same kinds of questions as traditional humanities research. (Though, because of the differing methods, the specific questions are often different.) Like other digital humanists, computational musicologists use (and make) digital tools to explore, ask, and answer questions about human artifacts — in this case, the structural elements of musical works and (meta)data about human interaction with music.

Types of computational musicology projects

Corpus studies. Corpus studies are possibly the most common type of project in computational musicology. A corpus study uses software to analyze statistical patterns in a large collection — corpus — of musical works. It is, essentially, descriptive statistics for musical data. Like text-based corpus studies, musical corpus studies often use n-gram and cluster analysis methods. Unlike text-based corpus studies, musical corpus studies often involve Markov models — probability analyses for progressions in time, such as how likely is G C-major chord to progress to a D-minor chord in a piece in the key of A minor.

Modeling. More recently, computational musicologists have been employing more advanced statistical methods to uncover underlying functions and patterns that contribute to the “surface” musical features of pieces in their corpora. This may involve Bayesian statistics, and more specifically, a hidden Markov model: for example, what are the underlying chordal *functions* in a chord progression, and do the traditional T/S/D functions explain the actual chord progressions observed in a corpus.

Music encoding. In order to perform a computational analysis, musical data must be encoded digitally in a way that makes sense both for the music and for the analysis to be performed. This can be a complicated and subjective task for musical artifacts.

Music information retrieval. Encoding musical data by hand is incredibly time-consuming. Music informational retrieval (MIR) is an attempt to automate that process, accurately extracting information from musical scores and, especially, audio. MIR draws on electrical engineering, machine learning, and digital signal processing (DSP).

Example projects and resources

That's just a brief overview. For more details, check out some of the following projects, resources, and articles.

Burgoyne, John Ashley. 2011. "Stochastic Processes & Database-Driven Musicology." Ph.D. diss., McGill University.

Music Genre and Spotify Metadata

De Clercq, Trevor and David Temperley. 2011. "A corpus analysis of rock harmony." In *Popular Music* 30/1, pp. 47–70.

Yim, Gary. 2012. "Affordant Harmony in Popular Music: Do Physical Attributes of the Guitar Influence Chord Sequences?" In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music., July 23–28, 2012.

The Lieder Project. A collaborative, computational research project looking at the relationship between the sounds of poetry and the structure of the music it is set. David Lonowski (CU–Boulder), Jordan Pyle (CU–Boulder), Stephen Rodgers (U Oregon), Kris Shaffer (CU–Boulder), Leigh VanHandel (Michigan State U).

CorpusMusic group on GitHub

Python Hidden Markov Model – Michael Hamilton, Colorado State University (the basis for unpublished HMM musical studies by Christopher William White and Ian Quinn, presented at the Society for Music Theory)

The McGill Billboard Project – a large data set of harmonic annotations of songs from several decades of the Billboard Top 100 list.

"A Corpus Study of Rock Music" – David Temperley and Trevor de Clercq's data and programs for analyzing the harmonic progressions in a corpus of 200 songs from Rolling Stone magazine's "Greatest [Rock] Songs of All Time."

The Million Song Dataset — “a freely-available collection of audio features and metadata for a million contemporary popular music tracks” (Columbia Univ.).

Empirical Musicology Review.

Originally published at kris.shaffermusic.com.

Data Science

Machine Learning

Music Theory

[About](#) [Help](#) [Legal](#)

Get the Medium app

