Automatic Chord Extraction

Author: Jason Leung

Instructor: Dr. Ichiro Fujinaga

March 26, 2014
1 Introduction

In contrast to many musical traditions of the world, the Western tradition places great emphasis on the palette of polyphonic structures – what is known harmony. Individual pitches (notes) are strung together – sounding either simultaneously or in close succession – and are perceived as one harmonic unit: a chord.

For this reason, automatic musical chord recognition by computers was seen as a task of polyphonic transcription and symbolic inference – a system will extract all the individual notes from an input signal, and then piece those notes back together into a chord (see Figure 1; Fujishima 1999). However, this method suffers from recognition errors at the first stage, as notes are often obscured by noise or overlapping harmonics (especially in consonant harmonies).

2 Feature Vectors

2.1 Pitch Class Profiles (PCP)

Thus, Fujishima (1999) found that a more robust approach is to treat chords as a single entity rather than a sequence of individual notes. By borrowing the concept of pitch classes from musical set theory, Fujishima recast chords as having a pitch class profile (PCP), where a pitch class is the set of all notes that are an octave apart. For example, the pitch class “C” is the set

\[ \{C_n : n \in \mathbb{Z}\} = \{\ldots, C_1, C_2, C_3, C_4, \ldots\}. \]

A PCP measures the signal’s spectral energy sums up the energy across the pitch classes to form a 12-dimensional feature vector (see Figure 2).
Figure 2: (Top) spectra of the notes C, E, and G, (middle) their combined spectra when played together as a C major triad in root position, and (bottom) the pitch class profile (PCP) of the same triad formed by adding together the spectral energies in each pitch class. (Image by J. P. Bello)

2.2 Tonal Centroid

More recently, Harte et al. (2006) takes a different approach by taking inspiration from Euler’s *Tonnetz* (1739), where Western harmony is conceived as a network of thirds and fifths. Harte et al. breaks down a chord into its position along the Circle of Fifths, Circle of Major Thirds, and Circle of Minor Thirds (see Figure 3) to come up with a 6-dimensional vector to describe the chord. One advantage this has over PCPs is that chord changes are represented as a movement of the centroid in a 6-dimensional space, and as such, more easily detectable.
Figure 3: Visualizing the 6-D Tonal Space as three circles. The tonal centroid for chord A major (pitch classes 9, 1, 4 – corresponding to pitches A, C♯, and E) is shown at point A. (Harte et al. 2006)

3 Chord Recognition

The matter of chord recognition becomes just one of pattern matching. A number of ways have been devised to do this: binary template matching (used by both Fujishima 1999 and Harte et al. 2006), as well as statistical methods – single Gaussian models (Sheh & Ellis 2003; Bello & Pickens 2005) and Gaussian mixture models (GMMs; Burgoyne et al. 2005; Reed et al. 2009) with hidden Markov Models (HMMs), and, more recently, networks of HMMs (Lee & Slaney 2008; Khadkevich & Omologo 2009).

3.1 Hidden Markov Models

The appeal of the hidden Markov models (HMMs) is that it has found great success in speech and handwriting recognition systems, so there is a lot of hope that it could work for music as well.

As review, a HMM consists of hidden and observed states, and obeys the Markov property – the next state is dependent only on the current state. The parameters in an HMM are: the initial state, the state space, and the transition probabilities between states. The whole problem is one of estimating these parameters, with the aim of finding the hidden states.

The problem can be simplified somewhat by using knowledge of music theory.
Bello & Pickens (2005) use the Circle of Fifths to fix their transition probabilities, ranking each chord in terms of their harmonic distance from the present chord, while Lee & Slaney (2008) use annotated data from real music to make their models more realistic—in particular, they used 765 Classical pieces and 158 Beatles songs to determine the likelihood of a certain chord transition. In both studies, the Viterbi algorithm was then applied to find the most-likely chord sequence in the song.

4 Fine Tuning

There are a number of ways to further improve the method. One big concern is in reducing spectral leakage due to windowing. With only 12 bins in the PCP, spectral energy can be misplaced if there are any mistuned notes, or if the recording is using a different tuning or temperament system. Purwins et al. (2000) suggest using 36 bins/octave instead, equivalent to 3 bins/semitone. This has the effect of narrowing the bins and focusing more energy in the correct places, but unfortunately, Fourier theory imposes a tradeoff in time-frequency resolution. Harte et al. (2006) perform a constant Q transform across 5 octaves, using 36 bins/octave, and end up with 743 ms-long frames—a very long time for musical signals. Consequently, they decide to overlap the frames using a hop size of 1/8 the frame length.

Knowing the key can also make the model more accurate, as it allows us to use different HMMs with different chord transition probabilities—based on the fact that chords have different harmonic functions in different keys. Therefore, Lee & Slaney (2008) created 24 different HMMs—one for each major and minor key, and applied the Viterbi algorithm through each to find the most likely path; the model with the maximum likelihood is the perceived key of the song (see Figure 4).

5 Future Work

From thereon, future work needs to account for more chords—all the studies up until this point have only considered major, minor, and diminished triads; four-note chords, such as the dominant seventh, have not been assessed despite their indispensability in Western harmony. There also needs to be a method to deal with modulations and key changes, because as mentioned earlier, the same chord functions differently harmonically in different keys, thus altering its transition probabilities.
Beat-synchronous analysis should also make the models more accurate (Lee & Slaney 2008) as chord changes tend to happen on the beat, whereas right now, they are occurring in the middle of frames.

Higher-order HMMs are also a possibility, whereby looking several steps further back in time, a better harmonic model can be constructed that will improve the estimate of transition probabilities. One avenue of research is in language modelling, where the idea of the n-gram is borrowed from speech recognition in hopes that it could also help recognize musical patterns (Khadkevich & Omologo 2009; Yoshii & Goto 2011).

However, a big challenge hindering these efforts is the lack of good annotated training data. McVicar et al. (2011) propose a method to clean up online databases for use, but it is not perfect since a large portion of that data is user-submitted and so contains numerous errors – due either to misidentification of musical elements, or to artistic embellishment.

6 Conclusion

As can be seen, numerous methods have been devised to automatically extract chord from audio data. Each method was largely influenced by their author’s subjective understanding of music theory, and at the present, it is difficult to determine which
A chord extraction algorithm is of great interest to the developers of music recommendation systems, as the harmonic content of songs has largely been unexplored systematically. Songs can be categorized by harmonic content, helping identify tracks with similar styles, as well as their covers by other artists.

It has only been 15 years since Fujishima’s paper was published, and computer technology has vastly improved since then, inspiring new methods and techniques. Thus, it is difficult to speculate what future technologies can bring.

References


