Audio Features*

Yi-Chieh Wu

Abstract

The audio features used to characterize the sound signal and classify the sample by instrument.

How do we decide what parts of the spectrum are important? The CUIDADO project(2) (p. 2) provided a set of 72 audio features, and research1 has shown that some of the features are more important in capturing the signal characteristics. We therefore decided to implement a small subset of these features:

Cepstral Features

- Mel-Frequency Cepstrum Coefficients (MFCC), k = 2:13

Spectral Features

- Slope
- Roll-Off
- Centroid
- Spread
- Skew
- Kurtosis
- Odd-to-Even Harmonic Energy Ratio (OER)
- Tristimulus

1 Definitions

Cepstral coefficients have received a great deal of attention in the speech processing community, as they try to extract the characteristics of the filter and model it independently of the signal being produced. This is ideal, as the filter in our case is the instrument that we are trying to recognize. We work on a Mel scale because it more accurately models how the human auditory system perceives different frequencies, i.e. it gives more weight to changes at low frequencies as humans are more adept at distinguishing low frequency changes.

The centroid correlates to the "brightness" of the sound and is often higher than expected due to the energy from harmonics above the fundamental frequency. The spread, skew, and kurtosis are based on the 2nd, 3rd, and 4th moments and, along with the slope, help portray spectral shape.

Odd-to-even harmonic energy ratio simply determines whether a sound consists primarily of odd harmonic energy, of even harmonic energy, or whether the harmonic energy is equally spread.

*Version 1.3: Dec 14, 2005 9:30 pm -0600

†http://creativecommons.org/licenses/by/2.0/

The tristimulus measure energy as well and were introduced as the timbre equivalent to the color attributes of vision. Like the OER, it provides clues regarding the distribution of harmonic energy, this time focusing on low, mid, and high harmonics rather than odd and even harmonics. This gives more weight to the first few harmonics, which are perceptually more important.

2 How We Chose Features

MFCC have shown to work very well in monophonic environments, as they capture the shape of the spectrum very effectively. Unfortunately, they are of less use in polyphonic recordings, as the MFCC captures the shape of a spectrum calculated from multiple sources. Most of the work we have seen on this subject uses MFCC regardless, however. They are particularly useful if only one instrument is playing or is relatively quite salient.

Most wind instruments have their harmonics evenly spread among the odd and even indices, but the clarinet is distinct in that it produces spectra consisting predominantly of odd ratios, with very little even harmonics appearing at all. This makes sense from a physics standpoint, as when played, the clarinet becomes a closed cylinder at one end, therefore allowing only the odd harmonics to resonate. This feature was thus chosen primarily with clarinet classification in mind.

We chose the roll-off and tristimulus as our energy measures, as they were both easy to implement and judged to be important(1) (p. 2). Finally, the first four spectral moments and the spectral slope, in both perceptual and spectral models, were shown to be the top ten most important features in the same study and were therefore some of the first features added to our classification system. We note that we had hoped to implement a perceptual model and thereby nearly double our features, but we could not find an accurate filter model for the mid-ear and thus decided to forgo any features based on perceptual modeling.

For further discussion of these features, along with explicit mathematical formulas, please refer to (1) (p. 2).

3 References

1. A.A. Livshin and X. Rodet. "Musical Instrument Identification in Continuous Recordings," in Proc. of the 7th Int. Conference on Digital Audio Effects, Naples, Italy, October 5-8, 2004.
2. G. Peeters. "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," 2003. URL: http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatur