EE269
Signal Processing and Quantization for Machine Learning

Lecture 3 Part II

Instructor : Mert Pilanci

Stanford University

Outline

- Spectral Descriptors
- Examples

Spectral Descriptors

- spectral centroid
- spectral spread
- spectral skewness
- spectral kurtosis
- spectral entropy
- spectral flatness
- spectral crest
- spectral flux
- spectral slope

  ⋮

Applications of Spectral Descriptors

- Speaker identification and recognition
- Acoustic scene recognition
- Instrument recognition
- Music genre classification
- Mood recognition
- Voice activity detection

⋮

Spectral Centroid

- spectral centroid μ_1 is the frequency-weighted sum normalized by the unweighted sum

$$\mu_1 = \frac{\sum_{k=b_1}^{b_2} f_k s_k}{\sum_{k=b_1}^{b_2} s_k}$$

f_k is the frequency in Hz corresponding to bin k

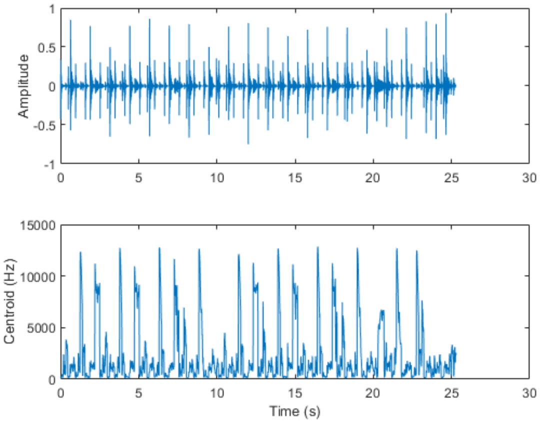
s_k is the spectral value at bin k

magnitude spectrum $|X[k]|$ and power spectrum $|X[k]|^2$ are commonly used

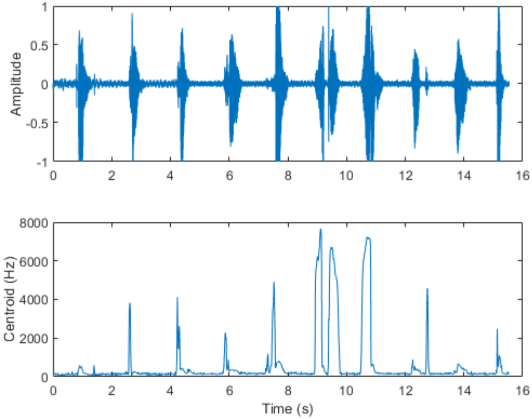
b_1 and b_2 are band edges, over which to calculate the centroid

Spectral Centroid - Example 1

- represents *center of gravity* of the spectrum, and indicates *brightness*. It is commonly used in music analysis and genre classification.
- for example, the jumps in the centroid corresponding to high hat hits



Spectral Centroid - Example 2



Spectral Spread

- spectral spread μ_2 is the standard deviation around the spectral centroid μ_1
- represents instantaneous bandwidth

$$\mu_2 = \sqrt{\frac{\sum_{k=b_1}^{b_2} (f_k - \mu_1)^2 s_k}{\sum_{k=b_1}^{b_2} s_k}}$$

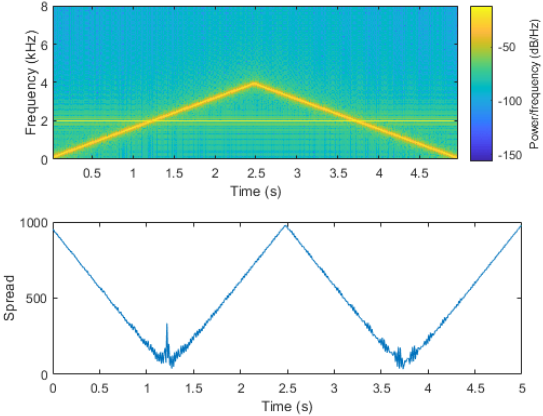
f_k is the frequency in Hz corresponding to bin k

s_k is the spectral value at bin k

b_1 and b_2 are band edges, over which to calculate the spread

μ_1 is the spectral centroid

Spectral Spread - Example



Spectral Kurtosis

- spectral kurtosis μ_4 is the fourth order moment

  measures flatness, or non-Gaussianity of the spectrum around the centroid

  $$\mu_4 = \frac{\sum_{k=b_1}^{b_2} (f_k - \mu_1)^4 s_k}{(\mu_2)^4 \sum_{k=b_1}^{b_2} s_k}$$

  f_k is the frequency in Hz corresponding to bin k

  s_k is the spectral value at bin k

  b_1 and b_2 are band edges, over which to calculate the kurtosis

  μ_1 is the spectral centroid

  μ_2 is the spectral spread
- Kurtosis tells you whether large deviations happen more often or less often than in a Gaussian with the same variance
- X ~Gaussian has $\mathrm{Kurtosis}(X) = \frac{\mathbb{E}[X^4]}{(\mathbb{E}[X^2])^2} = 3$

Numerical examples

$$\text{Kurtosis}(X) = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^4}{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^2}$$

1. Spiky / Heavy-tailed signal

$$x = (0, 0, 0, 0, 0, 0, 10, -10)$$

$$\frac{1}{8}\sum x_i^2 = 25, \qquad \frac{1}{8}\sum x_i^4 = 2500$$

$$\text{Kurtosis}(X) = \frac{2500}{25^2} = 4 \quad (> 3)$$

Rare large outliers increase kurtosis.

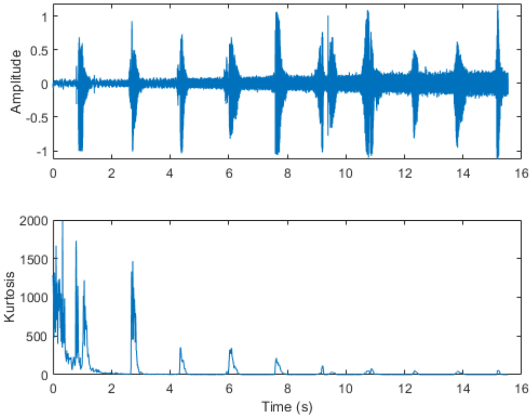
2. Bounded / Light-tailed signal

$$x = (-1, -1, -1, -1, 1, 1, 1, 1)$$

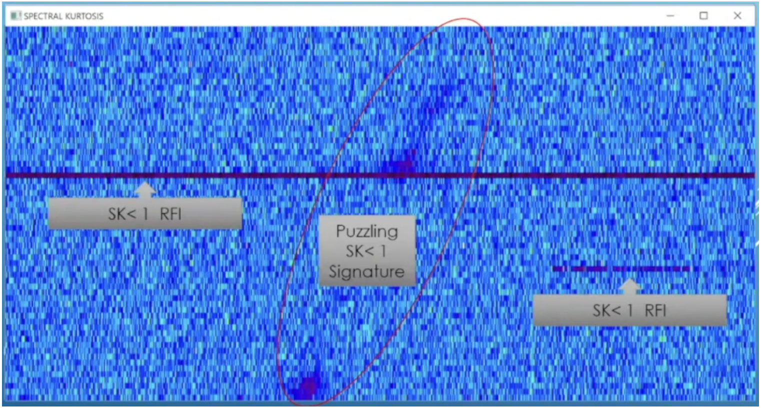
$$\frac{1}{8}\sum x_i^2 = 1, \qquad \frac{1}{8}\sum x_i^4 = 1$$

$$\text{Kurtosis}(X) = 1 \quad (< 3)$$

Spectral Kurtosis - Example



Spectral Kurtosis - Deep Space Signals



slide credit: Gelu Nita

Spectral Entropy

- spectral entropy represents the peakiness of the spectrum

  measure of disorder

$$\text{entropy} = \frac{-\sum_{k=b_1}^{b_k} s_k \log s_k}{\log(b_2 - b_1)}$$

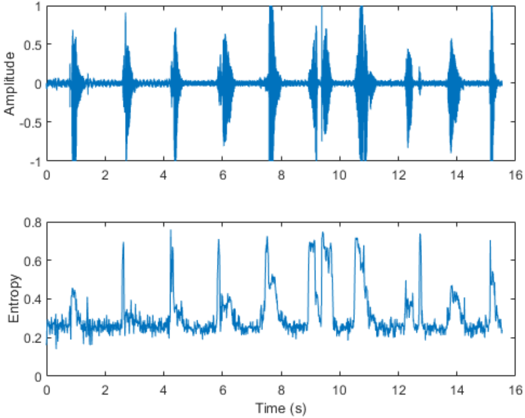
  f_k is the frequency in Hz corresponding to bin k

  s_k is the spectral value at bin k

  b_1 and b_2 are band edges, over which to calculate the entropy

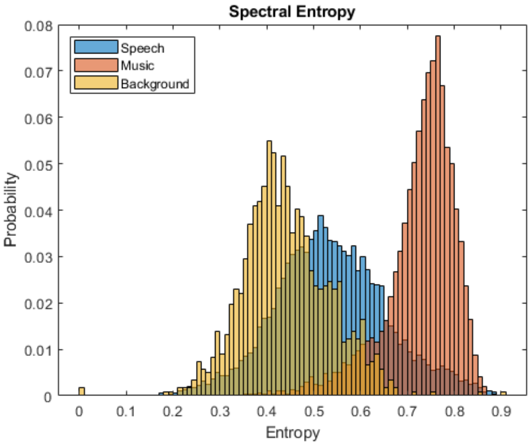
Spectral Entropy - Example 1

- Spectral entropy has been used successfully in voiced/unvoiced decisions for speech recognition



Spectral Entropy - Example 2

- Spectral entropy has also been used to discriminate between speech and music



Spectral Flux

- spectral flux represents the variability of the spectrum over time

$$\text{flux} = \left(\sum_{k=b_1}^{b_2} |s_k(t) - s_k(t-1)|^p \right)^{\frac{1}{p}}$$

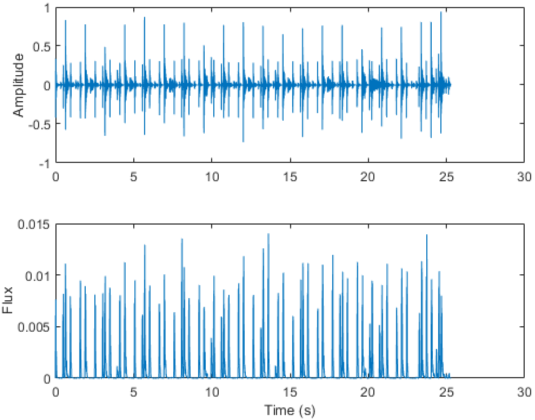
  s_k is the spectral value at bin k

  b_1 and b_2 are band edges, over which to calculate the spectral flux

- p is the norm type, e.g., $p = 1$ or $p = 2$

Spectral Flux - Example

- For example, the beats in the drum track correspond to high spectral flux



References

- https://www.mathworks.com/help/audio/ug/spectral-descriptors.html
- Murthy, H.a., F. Beaufays, L.p. Heck, and M. Weintraub. "Robust Text-Independent Speaker Identification over Telephone Channels." IEEE Transactions on Speech and Audio Processing. Vol. 7, Issue 5, 1999, pp. 554–568.
- Peeters, G. "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project." Technical Report; IRCAM: Paris, France, 2004.
- Grey, John M., and John W. Gordon. "Perceptual Effects of Spectral Modifications on Musical Timbres." The Journal of the Acoustical Society of America. Vol. 63, Issue 5, 1978, pp. 1493–1500.
- S. Zhang, Y. Guo, and Q. Zhang, "Robust Voice Activity Detection Feature Design Based on Spectral Kurtosis." First International Workshop on Education Technology and Computer Science, 2009, pp. 269–272.
- Hansen, John H. L., and Sanjay Patil. "Speech Under Stress: