***CHROMA FEATURES:***

In Western [music](https://en.wikipedia.org/wiki/Music), the term ***chroma feature*** closely relates to the **twelve** different [pitch](https://en.wikipedia.org/wiki/Pitch_classes)\* classes\*\*.

Chroma-based features, which are also referred to as "[pitch class profiles](https://en.wikipedia.org/wiki/Harmonic_pitch_class_profiles)", are a powerful tool for analyzing music whose pitches can be meaningfully categorized (often into twelve categories) and whose tuning approximates to the [equal-tempered scale](https://en.wikipedia.org/wiki/Equal_temperament). One main property of chroma features is that they capture harmonic and melodic characteristics of music, while being robust to changes in timbre and instrumentation.

\*Pitch: Pitch is a perceptual property of sounds that allows their ordering on a frequency-related scale, or more commonly, pitch is the quality that makes it possible to judge sounds as "higher" and "lower" in the sense associated with musical melodies. Pitch can be determined only in sounds that have a frequency that is clear and stable enough to distinguish from noise. Pitch is a major auditory attribute of musical tones, along with duration, loudness, and timbre.

Pitch may be quantified as a frequency, but pitch is not a purely objective physical property; it is a subjective psychoacoustical attribute of sound. Historically, the study of pitch and pitch perception has been a central problem in psychoacoustics, and has been instrumental in forming and testing theories of sound representation, processing, and perception in the auditory system.

\*\*Pitch Classes: In music, a pitch class (p.c. or pc) is a set of all pitches that are a whole number of octaves apart, e.g., the pitch class C consists of the Cs in all octaves. "The pitch class C stands for all possible Cs, in whatever octave position.” Important to musical set theory, a pitch class is "all pitches related to each other by octave, enharmonic equivalence, or both." Thus, using scientific pitch notation, the pitch class "C" is the set {Cn : n is an integer} = {..., C−2, C−1, C0, C1, C2, C3 ...}.

**CHROMA CENS:** **Chroma energy normalized statistics** (CENS)

The main idea of CENS features is that taking statistics over large windows smooths local deviations in tempo, articulation, and musical ornaments such as trills and arpeggiated chords. CENS are best used for tasks such as audio matching and similarity.

<https://musicinformationretrieval.com/chroma.html#:~:text=Chroma%20energy%20normalized%20statistics%20(CENS,as%20audio%20matching%20and%20similarity>. 🡪 es python

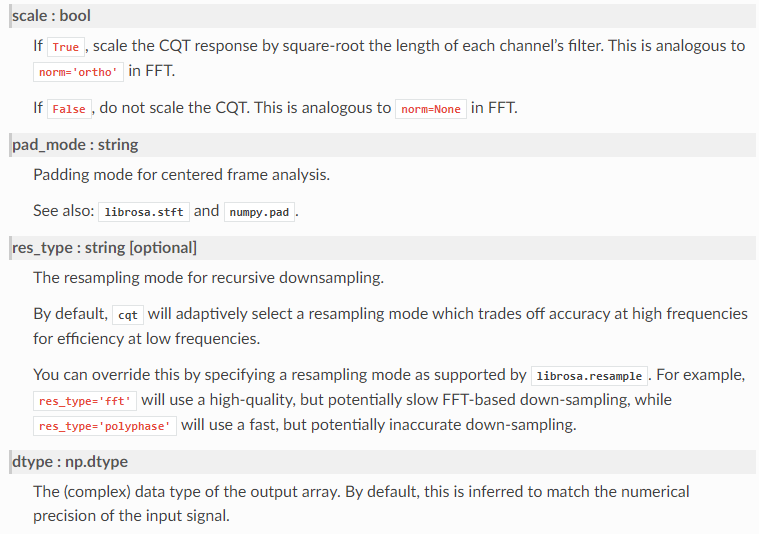
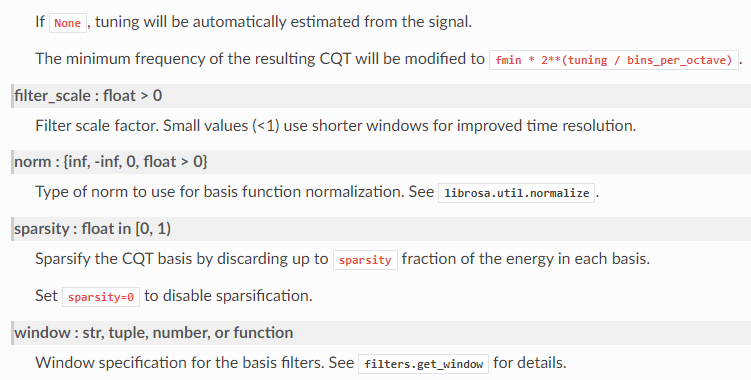
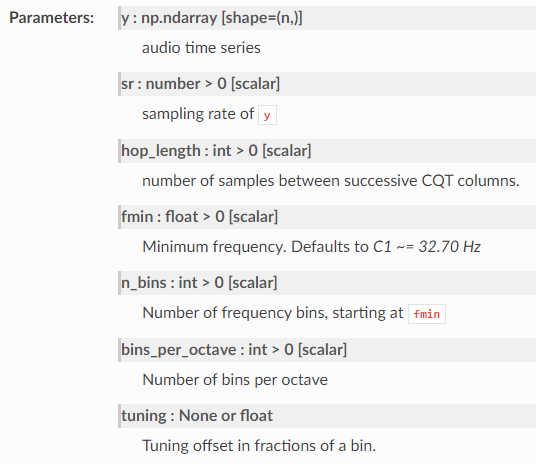
<https://www.audiolabs-erlangen.de/resources/MIR/FMP/C7/C7S2_CENS.html> 🡪 example

**CHROMA STFT: Short-time Fourier transform** (**STFT**)

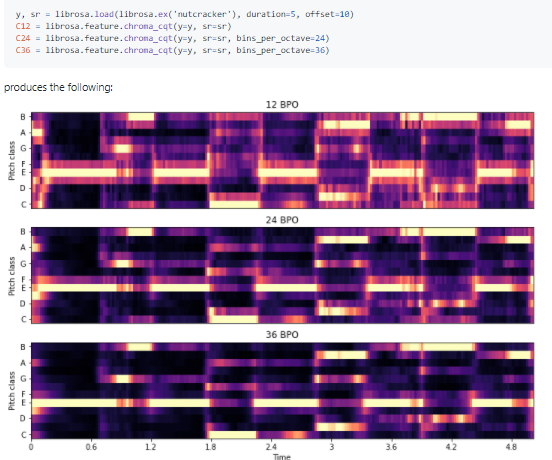
STFT is a [Fourier-related transform](https://en.wikipedia.org/wiki/List_of_Fourier-related_transforms) used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time.[[1]](https://en.wikipedia.org/wiki/Short-time_Fourier_transform#cite_note-1) In practice, the procedure for computing STFTs is to divide a longer time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment. This reveals the Fourier spectrum on each shorter segment. One then usually plots the changing spectra as a function of time, known as a [spectrogram](https://en.wikipedia.org/wiki/Spectrogram) or waterfall plot, such as commonly used in [Software Defined Radio](https://en.wikipedia.org/wiki/Software_Defined_Radio) based spectrum displays. Full bandwidth displays covering the whole range of an SDR commonly use FFTs with 2^24 points on desktop computers.

**CHROMA CQT: constant-Q transform**, simply known as **CQT** transforms a data series to the frequency domain.

**librosa.cqt=** Compute the constant-Q transform of an audio signal. This implementation is based on the recursive sub-sampling method:   
EXAMPLE: https://librosa.org/doc/main/generated/librosa.cqt.html#librosa.cqt



The chroma\_cqt function has a default of 12 bins per octave in its underlying CQT call. This produces a pretty noisy chroma feature, which can be significantly cleaned up by increasing the CQT resolution, as documented in our chroma example notebook. For example:



Altro esempio🡪<https://librosa.org/doc/main/generated/librosa.feature.chroma_cqt.html>

-From a forum:

* STFT is a standard short-time fourier transform.  The signal is broken into sliding windows and each window is transformed by the FFT
* CQT differs from STFT in two major ways: 1) the frequency scale is logarithmic, not linear.  2) the window length is different for each frequency: low frequencies use long windows, high frequencies use short windows.
* chroma\_stft uses the STFT to generate a spectrogram, and then projects that spectrum down to a single-octave (chroma) representation.
* chroma\_cqt is similar, but uses the CQT instead of the STFT.

In general, I recommend chroma\_cqt over chroma\_stft.

More on: <https://librosa.org/doc/0.8.0/generated/librosa.cqt.html>

**MFCC: mel-frequency cepstrum** (**MFC**)MFCC is a representation of the short-term [power spectrum](https://en.wikipedia.org/wiki/Power_spectrum) of a sound, based on a [linear cosine transform](https://en.wikipedia.org/wiki/Cosine_transform) of a [log power spectrum](https://en.wikipedia.org/wiki/Power_spectrum) on a [nonlinear](https://en.wikipedia.org/wiki/Nonlinear_system) [mel scale](https://en.wikipedia.org/wiki/Mel_scale" \o "Mel scale) of frequency  
🡪 cepstrum is the information of rate of change in spectral bands

**Mel-frequency cepstral coefficients** (**MFCCs**) are coefficients that collectively make up an MFC. They are derived from a type of [cepstral](https://en.wikipedia.org/wiki/Cepstrum) representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the [cepstrum](https://en.wikipedia.org/wiki/Cepstrum" \o "Cepstrum) and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in [audio compression](https://en.wikipedia.org/wiki/Data_compression#Audio).

MFCCs are commonly derived as follows:

1. Take the [Fourier transform](https://en.wikipedia.org/wiki/Fourier_transform) of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the [mel scale](https://en.wikipedia.org/wiki/Mel_scale" \o "Mel scale), using [triangular overlapping windows](https://en.wikipedia.org/wiki/Window_function#Triangular_window).
3. Take the [logs](https://en.wikipedia.org/wiki/Logarithm) of the powers at each of the mel frequencies.
4. Take the [discrete cosine transform](https://en.wikipedia.org/wiki/Discrete_cosine_transform) of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

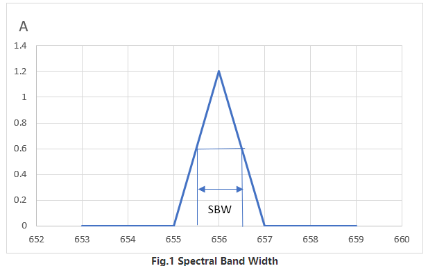
* MFCCs are commonly used as [features](https://en.wikipedia.org/wiki/Features_(pattern_recognition)) in [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition) systems, such as the systems which can automatically recognize numbers spoken into a telephone.
* MFCCs are also increasingly finding uses in [music information retrieval](https://en.wikipedia.org/wiki/Music_information_retrieval) applications such as **genre classification**, audio similarity measures, etc.
* MFCC values are not very robust in the presence of additive noise, and so it is common to normalise their values in speech recognition systems to lessen the influence of noise. Some researchers propose modifications to the basic MFCC algorithm to improve robustness, such as by raising the log-mel-amplitudes to a suitable power (around 2 or 3) before taking the DCT ([Discrete Cosine Transform](https://en.wikipedia.org/wiki/Discrete_cosine_transform)), which reduces the influence of low-energy components.
* Any sound generated by humans is determined by the shape of their vocal tract (including tongue, teeth, etc). If this shape can be determined correctly, any sound produced can be accurately represented. The envelope of the time power spectrum of the speech signal is representative of the vocal tract and MFCC (which is nothing but the coefficients that make up the Mel-frequency cepstrum) accurately represents this envelope.
* Tutorial for the use of MFCC: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>

**RMSE:** (Root Mean Squared Error)

Measure of goodnees of a model, in predictive terms.

[**https://musicinformationretrieval.com/energy.html**](https://musicinformationretrieval.com/energy.html) **--> use of RMSE**

***SPECTRAL ANALYSIS:* Spectral analysis** is a technique that allows us to discover underlying periodicities. To perform **spectral analysis**, we first must transform data from time domain to frequency domain

**SPECTRAL BAND WIDTH**: is an important specification for UV-VIS Spectrophotometer, it affects instruments resolution and test error. The spectral bandwidth is defined as the band width of light at one-half the peak maximum, as shown in Fig.1. The spectral bandwidth of a spectrophotometer is related to the physical slit-width and optical dispersion of the monochromator system.

Bandwidth is the difference between the upper and lower [frequencies](https://en.wikipedia.org/wiki/Frequency) in a continuous [band of frequencies](https://en.wikipedia.org/wiki/Frequency_band). It is typically measured in [hertz](https://en.wikipedia.org/wiki/Hertz), and depending on context, may specifically refer to [*passband*](https://en.wikipedia.org/wiki/Passband)*bandwidth* or [*baseband*](https://en.wikipedia.org/wiki/Baseband)*bandwidth*. Passband bandwidth is the difference between the upper and lower [cutoff frequencies](https://en.wikipedia.org/wiki/Cutoff_frequencies" \o "Cutoff frequencies) of, for example, a [band-pass filter](https://en.wikipedia.org/wiki/Band-pass_filter), a [communication channel](https://en.wikipedia.org/wiki/Communication_channel), or a [signal spectrum](https://en.wikipedia.org/wiki/Signal_spectrum). Baseband bandwidth applies to a [low-pass filter](https://en.wikipedia.org/wiki/Low-pass_filter) or baseband signal; the bandwidth is equal to its upper cutoff frequency.

Bandwidth in hertz is a central concept in many fields, including [electronics](https://en.wikipedia.org/wiki/Electronics), [information theory](https://en.wikipedia.org/wiki/Information_theory), [digital communications](https://en.wikipedia.org/wiki/Digital_communication), [radio communications](https://en.wikipedia.org/wiki/Radio_communication), [signal processing](https://en.wikipedia.org/wiki/Signal_processing), and [spectroscopy](https://en.wikipedia.org/wiki/Spectroscopy) and is one of the determinants of the capacity of a given [communication channel](https://en.wikipedia.org/wiki/Communication_channel).  
🡪<http://man.hubwiz.com/docset/LibROSA.docset/Contents/Resources/Documents/generated/librosa.feature.spectral_bandwidth.html> example librosa

**SPECTRAL CENTROID:**   
The **spectral centroid** is a measure used in [digital signal processing](https://en.wikipedia.org/wiki/Digital_signal_processing) to characterise a [spectrum](https://en.wikipedia.org/wiki/Spectrum). It indicates where the [center of mass](https://en.wikipedia.org/wiki/Center_of_mass" \o "Center of mass) of the spectrum is located. Perceptually, it has a robust connection with the impression of [brightness of a sound](https://en.wikipedia.org/wiki/Brightness_(sound)). Because the spectral centroid is a good predictor of the "brightness" of a sound, it is widely used in digital audio and music processing as an automatic measure of musical [timbre](https://en.wikipedia.org/wiki/Timbre)

indicates at which frequency the energy of a spectrum is centered upon.

This is like a weighted mean:

where S(k)S(k) is the spectral magnitude at frequency bin kk, f(k)f(k) is the frequency at bin kk.

[**librosa.feature.spectral\_centroid**](https://librosa.github.io/librosa/generated/librosa.feature.spectral_centroid.html#librosa.feature.spectral_centroid) computes the spectral centroid for each frame in a signal in pyhton

<https://musicinformationretrieval.com/spectral_features.html> example here

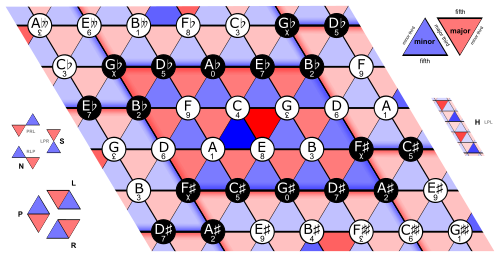
**SPECTTRAL CONTRAST:**

**Spectral contrast** is defined as the decibel difference between peaks and valleys in the spectrum.

**Spectral contrast** is defined as the level difference between peaks and valleys in the **spectrum**. In CIs, **spectral contrast** is degraded because of the limited number of stimulation electrodes and overlapping electric fields activating the nervous system through the bony structure of the cochlea

**SPECTRAL ROLLOFF:**

The **roll**-**off** frequency is defined as the frequency under which some percentage (cutoff) of the total energy of the spectrum is contained. The **roll**-**off** frequency can be used to distinguish between harmonic (below **roll**-**off**) and noisy sounds (above **roll**-**off**).

**TONNETZ:**

Tonnetz is a conceptual [lattice](https://en.wikipedia.org/wiki/Lattice_(music)) [diagram](https://en.wikipedia.org/wiki/Diagram) representing tonal space first described by [Leonhard Euler](https://en.wikipedia.org/wiki/Leonhard_Euler) in 1739. Various visual representations of the *Tonnetz* can be used to show [traditional harmonic relationships](https://en.wikipedia.org/wiki/Tonality) in European classical music.

In general, the Tonnetz is a pictorial representation of the notes in the plane that reveal affinities and structures between notes and on concrete music pieces.

*-Using tonnetz tone to understand jazz armony* **🡪** [**https://jazz-library.com/articles/tonnetz/**](https://jazz-library.com/articles/tonnetz/ù)

# *-Exploring Musical Structure Using Tonnetz Lattice Geometry and LSTMs 🡪* <https://link.springer.com/chapter/10.1007/978-3-030-50417-5_31>

[**https://thermopedia.com/content/1141/**](https://thermopedia.com/content/1141/) **🡪 time series**