The Mel Spectrogram

A Spectrogram on an AUDITORY FREQUENCY SCALE

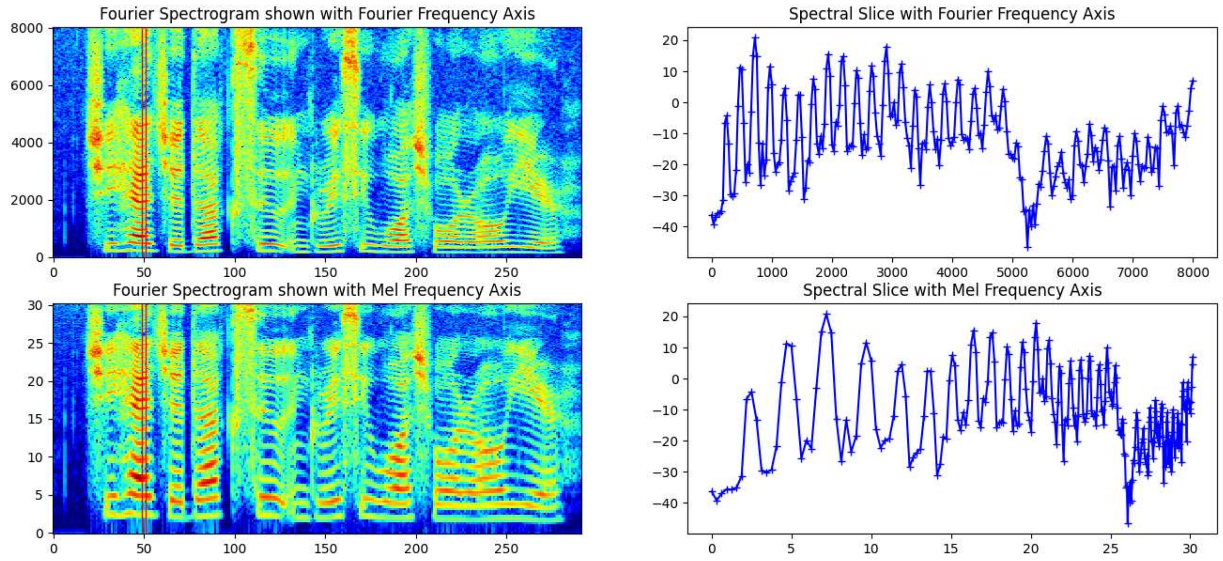
1. The MEL Spectrogram

MEL scale

The mel scale provides a mapping between Hz and mel and is an estimate of the non linear frequency axis of human hearing. There is not such a thing as **the mel scale** written in stone as any formula is **estimated** on te basis human physiology and perception. There are a number of almost equivalent transformation in use in the speech community. Our software uses the librosa default which is the same as the 'Slaney' mapping shown below. Conceptually it is fine to think as follows about the mel scale is: *linear below 1kHz, logarithmic above it*.

Displaying a spectrogram using a warped frequency scale

We can simply display a Fourier spectrogram on the mel-scale by redefining the frequency axis.
This definitely gives us a good impression of the warping realized by using the mel scale. However this overlooks one key aspect of the reason why we do this mel transformation, i.e. that frequency resolution (expressed in Hz) gets worse with high frequencies, alternatively said resolution should be constant along the mel scale.



2. Computing a MEL Spectrogram using a MEL Filterbank

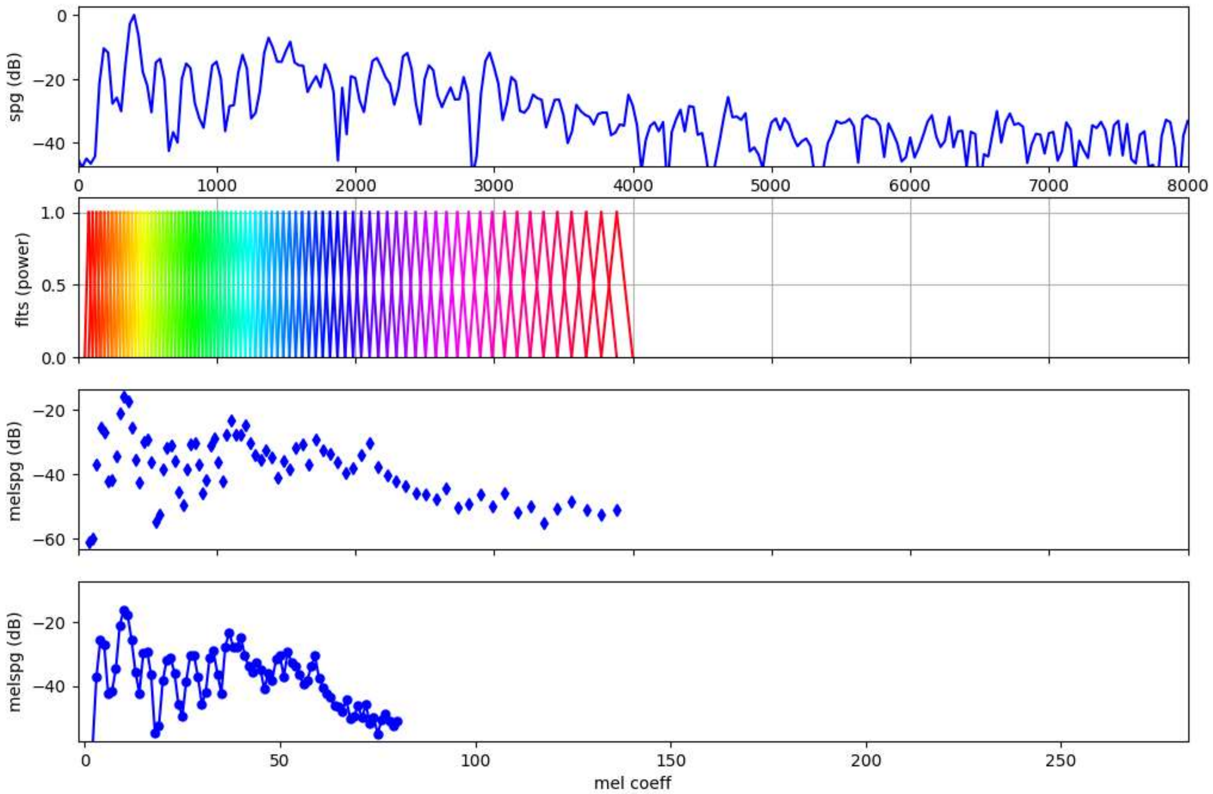
Motivation

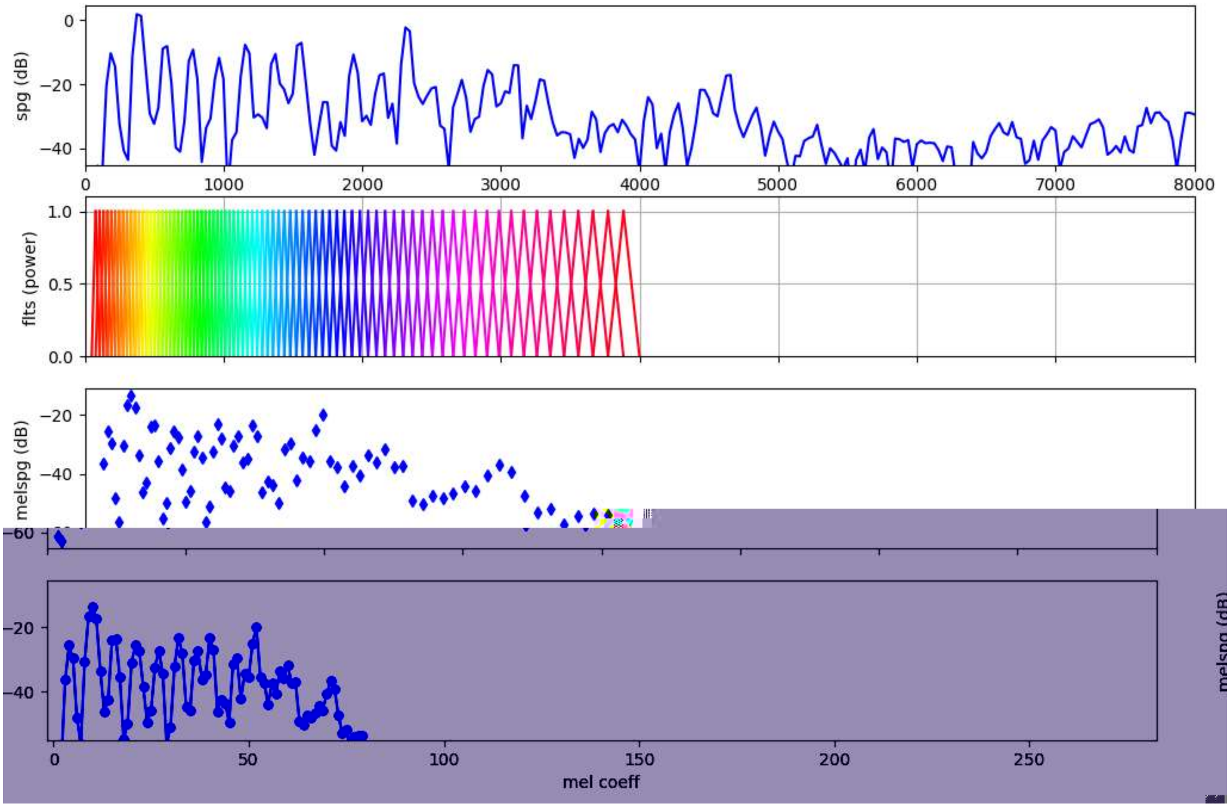
The whole mel-scale concept derives from the observation that perceptual frequency resolution is uniform on the mel scale. Hence we should have uniform sampling of our spectral representation on the mel scale. How can we achieve this ? We need to RESAMPLE and at the same time implement this frequency dependent resolution. There is no straightforward mathematical solution to this. In the speech community it has become common practice to achieve this required resampling by passing the Fourier spectrum through a MEL Filterbank.

MEL filterbank

A **mel filterbank** constitutes a set of filters that have the same shape and that are uniformly spaced on the mel scale. By using such a **mel filterbank** we can transform a Fourier spectrum into a **mel spectrum** with uniform sampling on the mel axis. The Hz-to-mel scale is only one aspect of the design such filterbank. We also need to specify spectral resolution on the mel scale, i.e. we need to choose the number of filters and their bandwidth. By predefining a mild overlap of 50% (standard accepted procedure) we conveniently eliminate the bandwidth as a parameter. The resolution is thus defined by number of filters in the targeted frequency range.

Note that a filter in a mel filterbank is not a traditional filter. Each filter constitutes a set of interpolation weights over the power spectrum.





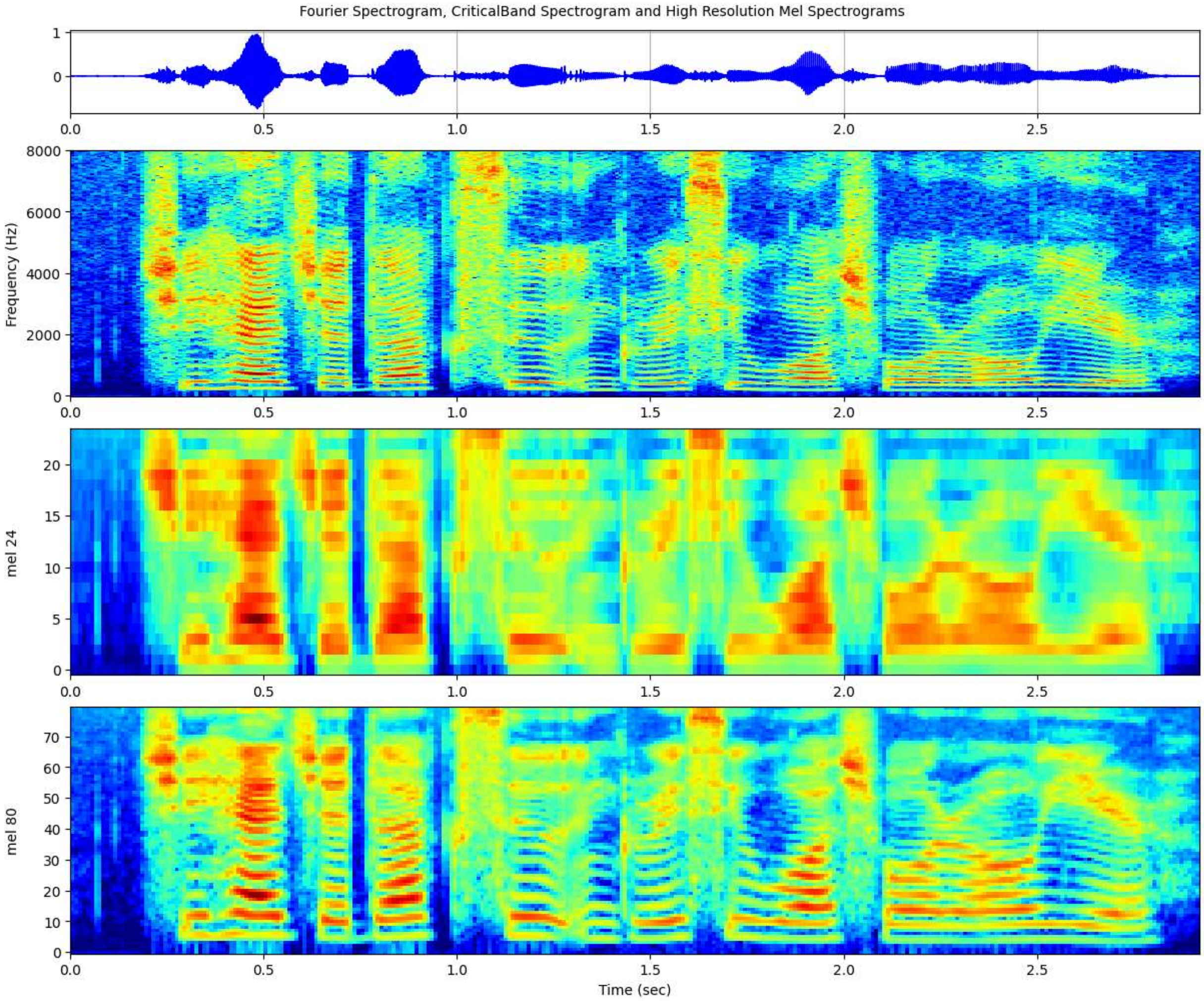
3. Critical Band and High Resolution Mel Spectrograms

The resolution of a mel spectrogram is defined by the number of mel filterbanks in the targeted frequency range. For sampling rates of 8kHz we tend to target the fully frequency range [50Hz, 4kHz], for higher sampling rates we tend to target a frequency range of [50Hz, 6500Hz] roughly as this encompass pretty well the relevant speech frequency range. In terms of resolution we grosso modo distinguish two scenarios:

- a **high resolution** mel filterbank: this filterbank performs in first instance spectral warping and intends to preserve as much detail as possible. Hence resolution in the low frequency zone should be maintained, while the wider filters at high frequencies will inherently imply smoothing. High resolution mel filterbanks are the preferred choice these days for most applications.
  => Recommended number of channels: *64(8kHz), 80(16kHz)*

- a **critical band** mel filterbank: the bands are roughly "one mel" wide; according to psychoacoustics spacing of the filters should be at omst one mel, hence it is the minimal sized filterbank. Such filterbank will preserve spectral envelope well, but [redacted] critical factor in the design.
  => Recommended number of channels: *20(8kHz), 24(16kHz)*

Below Critical Band and High Resolution mel spectrograms are shown side by side.


Fourier Spectrogram, CriticalBand Spectrogram and High Resolution Mel Spectrograms
Frequency (Hz)
mel 24
mel 80
Time (sec)