The MEL Scale

Frequency Sensitivity in Auditory Perception

The frequency range of the human ear spans from 20Hz to 20kHz. Many properties of our hearing system are easier understood in the frequency domain than in the time domain which makes Fourier spectrograms a favorite representation of acoustic signals.

However, Fourier analysis ignores important characteristics of [illegible] sensitivity:

- frequencies close to each interfere a lot, while frequencies far apart are quite independent
- the frequency range over which interference is prominent is small at low frequencies and much higher at high frequencies

Based on diverse perceptual experiments it was therefore suggested to model the human hearing system using a set of non-uniform auditory (band-pass) filters. The bandwidth of these filters is called the *critical bandwidth*.

Auditory Filters in the cochlea

Processing in the cochlea is easiest understood as an auditory filterbank. Individual auditory nerve fibers are sensitive only to a narrow frequency range centered around the *characteristic frequency* of the corresponding nerve fiber. Both the density of these filters and their bandwidth follow the critical band characteristics (linear at low, logarithmic at high frequencies)

1. Mel Scale Approximations

From the above it is obvious that there is not such a thing as **THE** auditory frequency scale as different experiments lead to slightly different scales: **MEL** scale, **BARK** scale, the **ERB** scale, **1/3th octave filterbanks**, ...

Putting small differences aside, it is often sufficient to reason as follows:

- linear up to 1kHz with a bandwidth of 100Hz
- logartihmic above 1 kHz
- resulting in 24 critical bands spanning the auditory frequency range

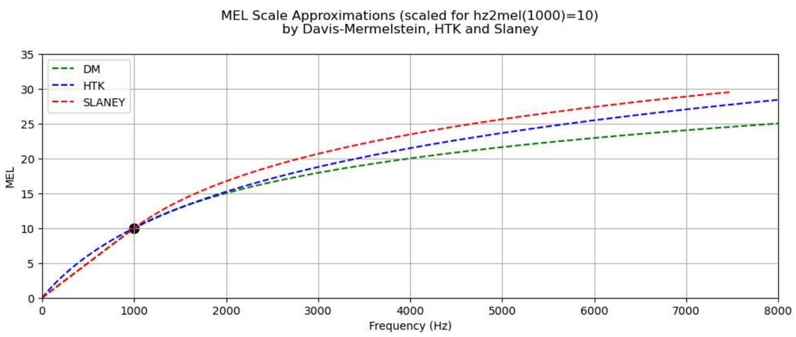
In the sequel we will use the term **mel-scale** as a generic term, and in order to compare different implementations we rescale such that frequency is mapped to a number corresponding for critical band number, in particular we always map 1000.0Hz each to 10.0 mel. (Note: this is our own way of standardizing different mappings; there is no established standard)

The functions mel2hz() and hz2mel() implement a few of these mel-scale variants:

- DM(Davis and Mermelstein): as defined in the 1980 paper in which they introduce the concept of mel cepstra, one of the oldest handy approximations
- HTK: as used in the HTK and KALDI [illegible] Fant in 1959
- Slaney: as used in the MATLAB Auditory Toolbox and also used as default in the librosa library.

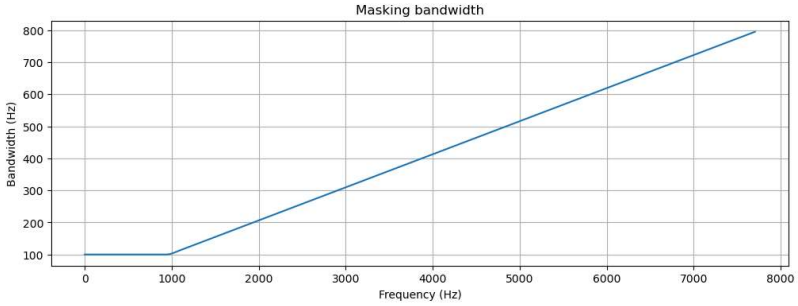
In Python packages librosa and torchaudio both the Slaney and HTK scales are implemented

All 'mel'-centered code is grouped in *mel.py* in this folder



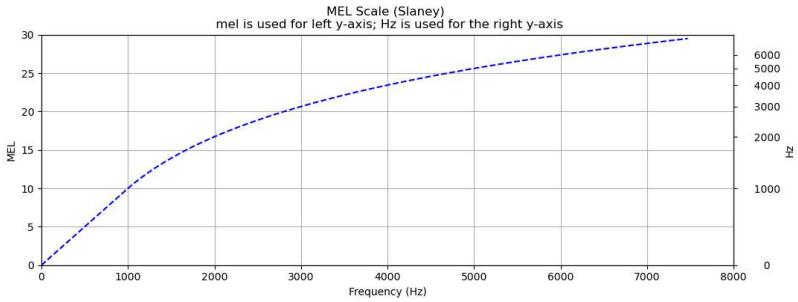
2. Mel scale and Equivalent Bandwidth

Another way to understand the mel scale is by thinking in terms of equivalent perceptual bandwidths; i.e. the range over which there is strong interference. Below 1kHz this bandwidth is estimated to be roughly 100Hz and at higher frequencies this bandwidth linearly increases.



Labeling a mel axis with 'mel' or 'Hz' ??

Because the lack of standardization of a mel scale, we see in practice that people use 'Hz' as labels on the frequency axis and that the non linear behavior is obvious from the tick values



2. MEL FILTERBANK

The primary usage of the mel-scale in speech processing is to map of a **Fourier spectrum** to a **mel spectrum** , using a **mel filterbank**.

Typically these filterbanks are designed such that overlap between adjacent bands is minimal. The filtering itself is done by summing up the powers of Fourier spectral coefficients within a band, weighted by the filter shape. The most common design choice is to use triangular shaped filters along the mel scale with 50% overlap.

Further Notes:

- It is also well known that the data at the fringes of the frequency range is not be very reliable. At very low frequencies there may only be 50Hz hum from electrical equipment and at high frequencies the content is highly dependent on the sampling rate and the anti-aliasing filtering that was used. Therefore it is common practice to throw away one or several of the lowest and highest bands in speech recognition systems or alternatively to limit the frequency range of the filterbank.
- In the filterbank design routines below we use the librosa() functions for mel filterbank design. Arguments *fmin* and *fmax* specify the range (

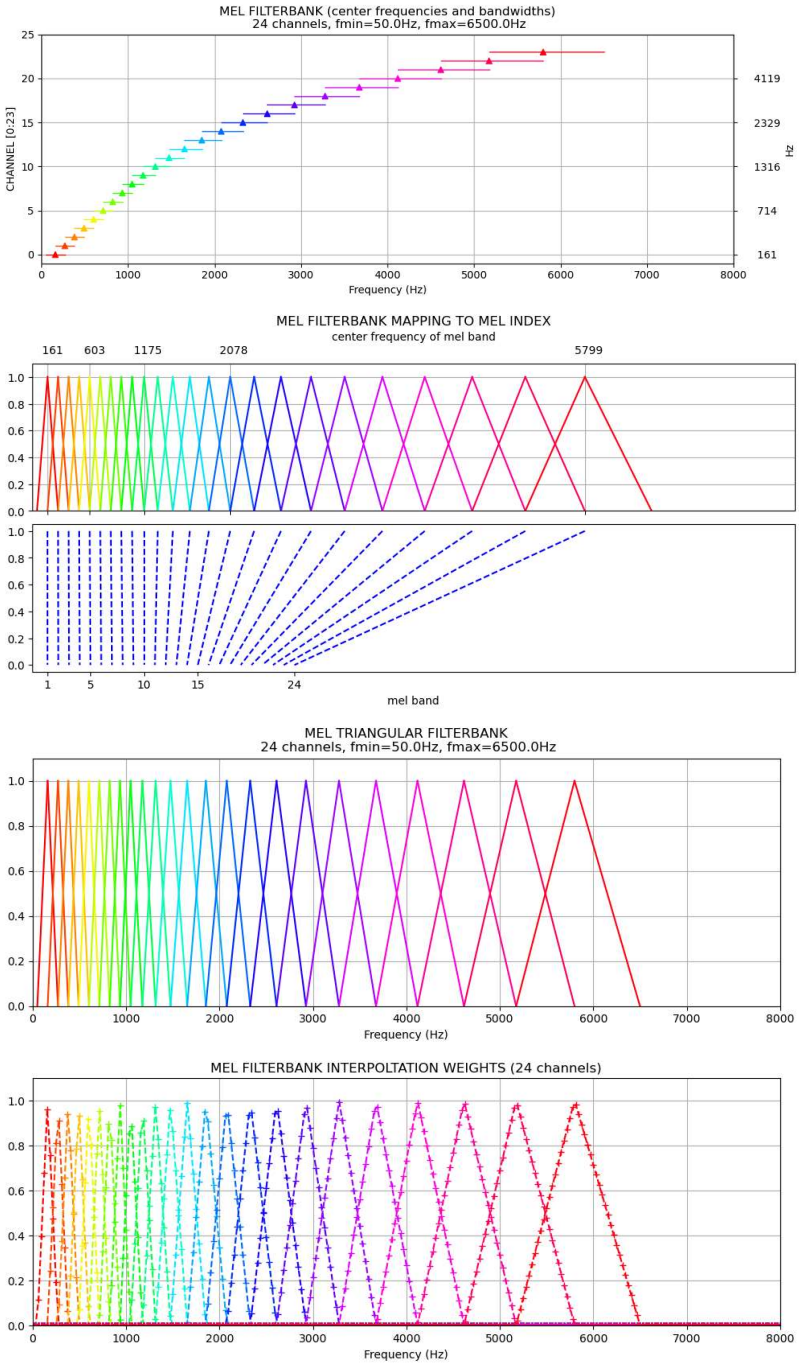
A critically spaced mel filterbank

The first filterbank that we design use a filter spacing (and bandwidth) of roughly 1 mel. This gives us 24 channels with a range of 0 - 6300 Hz. If we use a sampling rate of 8kHz then the upper 4 channels fall outside the usable frequency range, thus maintaining 20 channels. Also the first channel (0-100Hz) is likely to be useless for speech recognition applications as the content is too unpredictable. Alternatively we could set the fmin to 50 or 100Hz.

The figures show (1) designed centerfrequencies and bandwidth
(2) designed filterbank (conceptual) (3) actual interpolation weights to apply to an FFT spectrum to compute a mel spectrum and simulate the filterbank above

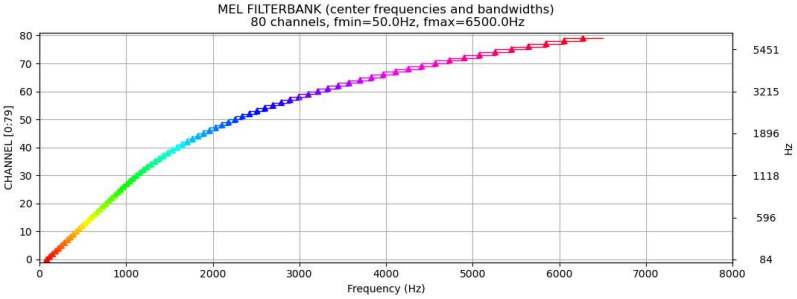
TASK:

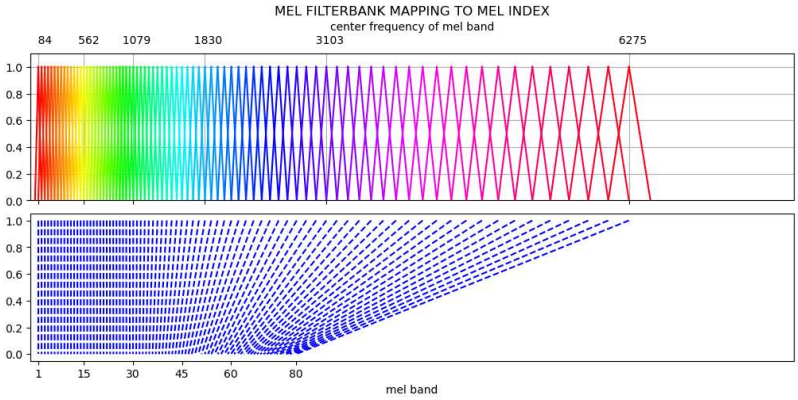
Design a filterbank for a sampling rate of 8kHz , using the full bandwidth and 20 channels. Verify that this yields the same as taking the first 20 channels from the originally designed filterbank

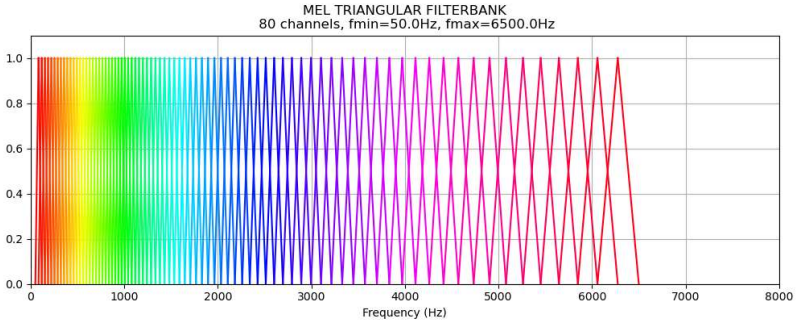


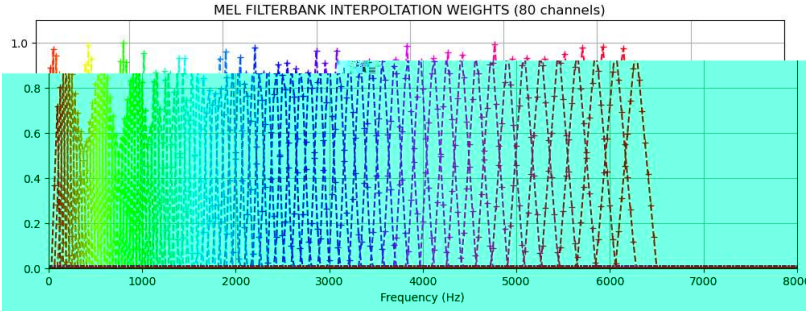
A high resolution mel filterbank

A high resolution mel filterbank uses filter spacings and bandwidths that are significantly higher resolution than 1 mel. It seems to make little sense to go beyond the resolution of our Fourier spectrum. With some fair assumptions about sampling rate and FFT parameters (256 pt FFT for 8kHz sampling and 512pt FFT for 16kHz sampling) we can conclude that *80 channels* is a reasonable number of channels (upper limit) for a high resolution filterbank. In the design below we set fmin to 50Hz and fmax to 6500Hz.


MEL FILTERBANK (center frequencies and bandwidths)
80 channels, fmin=50.0Hz, fmax=6500.0Hz
CHANNEL [0:79]
Frequency (Hz)
Hz
5451
3215
1896
1118
596
84


MEL FILTERBANK MAPPING TO MEL INDEX
center frequency of mel band
84
562
1079
1830
3103
6275
mel band


MEL TRIANGULAR FILTERBANK
80 channels, fmin=50.0Hz, fmax=6500.0Hz
Frequency (Hz)


MEL FILTERBANK INTERPOLTATION WEIGHTS (80 channels)
Frequency (Hz)