The mel frequency scale and coefficients

1

The human auditory system doesn't interpret pitch in a linear manner. The human interpretation of the pitch reises with the frequency, which in some applications may be a unwanted feature. To compensate for this the mel-scale was delevoped. The mel-scale was developed by experimenting with the human ears interpretation of a pitch in 1940's. The sole purpose of the experiment were to describe the human auditory system on a linear scale. The experiment showed that the pitch is lineary perceived in the frequencyrange 0-1000hz. Above 1000 hz, the scale becomes logaritmic. An approximated formular widely used for mel-scale is shown below:

$$F_{mel} = \frac{1000}{\log(2)} \cdot \left[1 + \frac{F_{Hz}}{1000}\right] \tag{1.1}$$

, where F_{mel} is the resulting frequency on the mel-scale measured in mels and F_{Hz} is the normal frequency measured in Hz. This is plotted in figure 1.1.

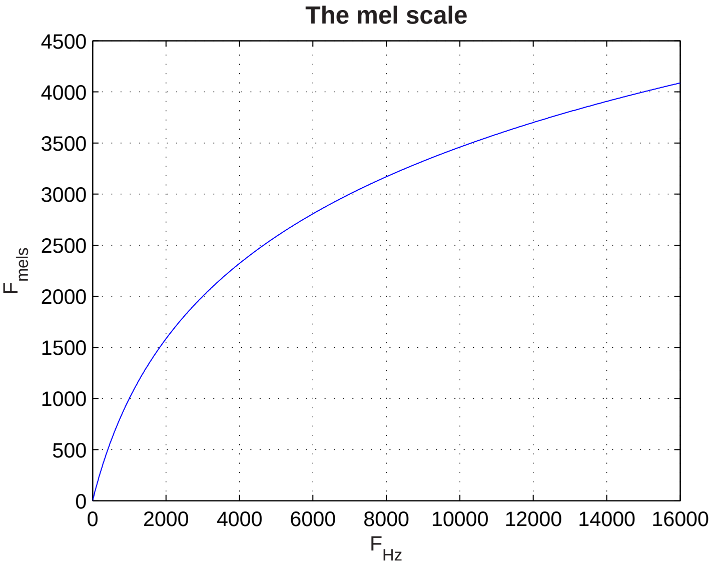

Figure 1.1: Relationship between the frequency scale and mel-scale.

With the mel-scale applied, coefficients from a LPC will be concentrated in the lower frequencies and only around the area perceived by humans as the pitch, which may result in a more precise description of a signal, seen from the preception of the human auditory system.

This is allthough not proved and it is only suggested that the mel-scale may have this effect. The mel-scale is, regardless of what have been said above, a widely used and effective scale within speech regonistion, in which a speaker need not to be identified, only understood.

1.1 MFCC

Mel Frequency Cepstral Coefficients (MFCC) is usaly derived using a filterbank, this is illustrated in figure 1.2. It has been found that the energy in a critical band of a particular frequency influence the human auditory systems perception. This critical band bandwidth varies with the frequency, where it is linear below 1 kHz and logaritmic above. Combining this with the mel scale, the distributions of these critical bands becomes linear.

The critical band is a bandpass filter, adjusted around the center frequency. Below 1 kHz critical bands are placed linear around 100, 200, ... 1000 Hz. Above 1 kHz these bands are placed with the mel-scale. In the calculation of the MFCC's the total energy in each critical band is used, by the use of equation 1.2.

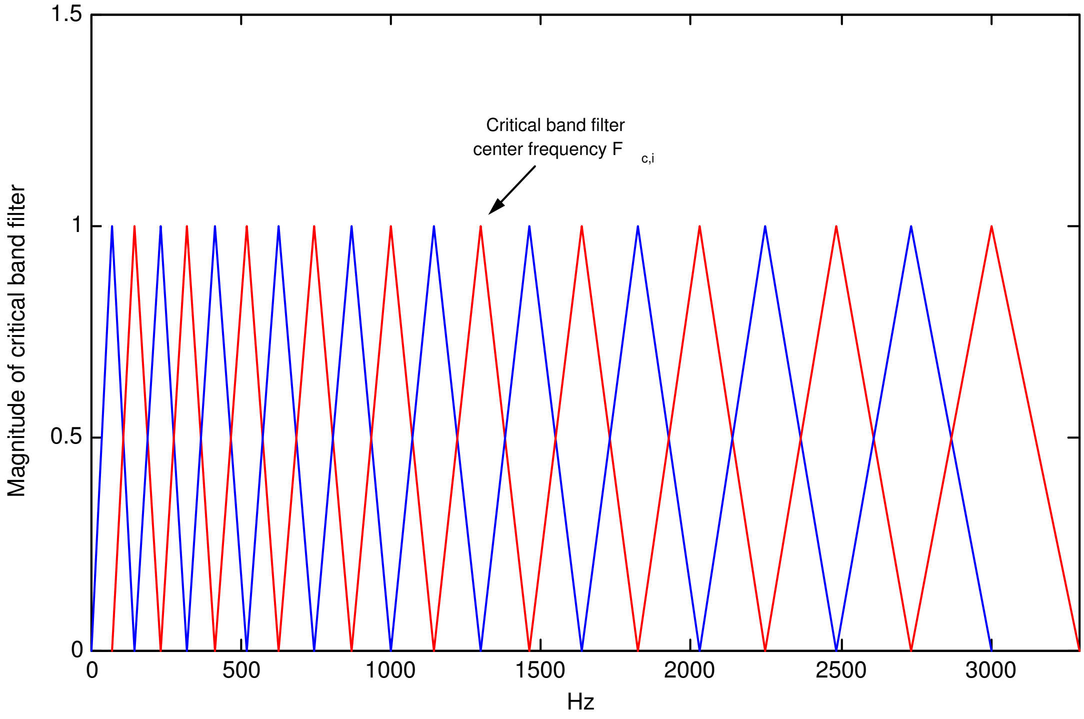

Figure 1.2: Mel scale filterbank. Each peak is the center frequency in the critical band.

$$Y\left(i\right) = \sum_{k=0}^{N/2} \log\left|s(n)\right| \cdot H_i\left(k \cdot \frac{2\pi}{N'}\right) \tag{1.2}$$

, where $Y(i)$ is the total energi in the critical band, N is the framelength, $S(n)$ is DFT signal for which the MFCC's is calculated, $H_i()$ is the critical band filter at the i'th coefficient and N' is the number of points used in the short term DFT (with zero padding).

Next we need to compute the actual IDTF to get the coefficients. For this we must handle each critical band individualy, which is done here:

$$\widetilde{Y}(k) = \begin{cases} Y(i) & ,k = k_i \\ 0 & ,otherk \in [0, N'-1] \end{cases}$$

(1.3)

So the final cepstrum can be derived by:

$$c_s(n) = \frac{1}{N'} \cdot \sum_{k=0}^{N'-1} \widetilde{Y}(k) e^{jk(2\pi/N')n} \tag{1.4}$$

If the real cepstrum is used, the sequence $\widetilde{Y}(k)$ is symmetrical (even) about the critical band center frequency, being N'/2. The equation 1.4 can therefore be reduced to this:

$$c_s(n) = \frac{2}{N'} \cdot \sum_{i=1,2,\ldots,N_{cb}} \widetilde{Y}\left(k_i\right) \cdot \cos\left(k_i \cdot \frac{2\pi}{N'} n\right) \tag{1.5}$$

, where N_{cb} is the number of critical bands.