

Aufgabe 2: Mel-Skalierte Dreiecksfilterbank

Das FFT-Spektrum eines Signals hat eine lineare Frequenzskalierung, die in Aufgabe 1.5 berechnet wurde. Um die menschliche Wahrnehmung zu imitieren, wird häufig eine gehör-ähnliche Frequenzskalierung bevorzugt.

Diese kann man beispielsweise mit einer Mel-Filterbank berechnen. Dazu berechnet man zunächst das Kurzzeitspektrogramm wie in Aufgabe 1.4 und summiert dann jeweils innerhalb von einem Filterband alle Spektralanteile gewichtet auf. Dazu wird als Gewichtungsfunktion oft eine Dreiecksfunktion verwendet. Abbildung 1 zeigt das Aussehen einer typischen Dreiecksfilterbank. Die Frequenz-Stützstellen $f[0]$ bis $f[\text{NUMCHANS} + 1]$ werden hier so berechnet, dass sie

- im Mel-Frequenzbereich angenähert äquidistant sind und
- genau mit Frequenzstützstellen im Fourierbereich übereinstimmen (**NUMCHANS**: Anzahl der Mel-Filter, im Beispiel ist **NUMCHANS** = 6).

Die Gewichte der Dreiecksfilter sollten so bestimmt werden, dass sie

- bei der Mittenfrequenz maximal sind,
- bei den Mittenfrequenzen der angrenzenden Bänder Null sind und
- das jedes von Ihnen die gleiche Fläche hat (und so bei weißen, also spektral flachen, Signalen die gleiche Bandenergie ergeben würde).

In dieser Übung soll die Funktion **ComputeFeatures** aus Aufgabe 1 um die Mel-Filterbank-Analyse erweitert werden. Dazu sind folgende Funktionen nützlich:

- 2.1** Eine Funktion, die eine gegebene Anzahl von **NUMCHANS** + 2 viele Punkte im linearen Frequenzbereich berechnet, die auf der MEL-Skala äquidistant sind. Für die Mel-Skalierung gilt:

$$\text{Mel}(f) = 2595 \log_{10}(1 + f/700)$$

Ein dafür nützlicher Funktionskopf könnte folgendermaßen aussehen:

```
freqs = melfreqs(fmin, fmax, numfreqs)
```

Hilfreiche Funktionen: `log10`

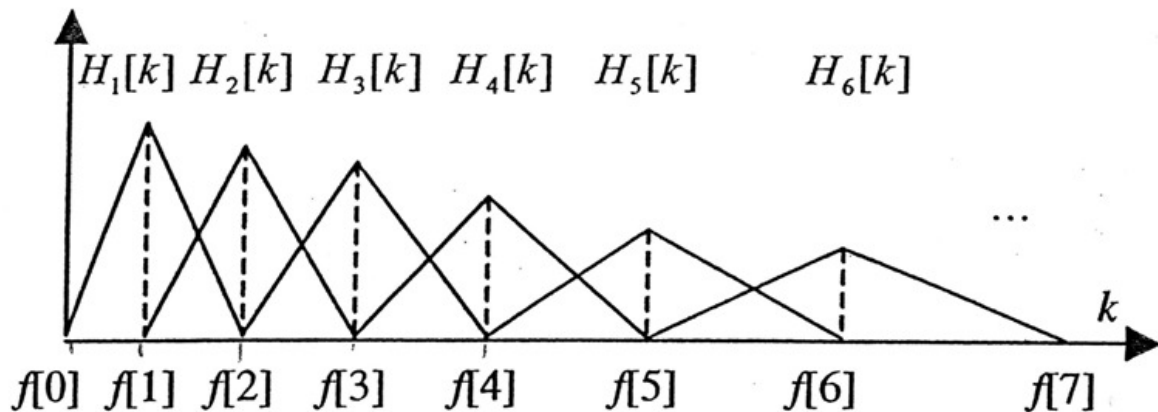


Abbildung 1: Dreiecksfilterbank, aus: HUANG, ACERO, HON: *Spoken Language Processing*

- 2.2** Eine Funktion, die NUMCHANS viele Dreiecksfilter im Frequenzbereich berechnet. Deren Mittenfrequenzen sollten so nah wie möglich an den in Aufgabe 2.1 gewählten Mel-Frequenzen liegen.

Legen Sie dazu zuerst die Mittenfrequenzen aller Mel-Filter-Bänder, $f[i]$, durch Wahl der nächstgelegenen linearen Frequenzstützstelle fest. Wählen Sie bitte $f[0] = 1$, und $f[\text{NUMCHANS} + 1] = \text{NFFT}/2 + 1$. Diese beiden Filterausgänge werden später nicht benötigt. Hierbei sind $f[i]$ der FFT-Index des i 'ten Mel-Bandes und NFFT die FFT-Fensterbreite.

Anschließend lassen sich die Dreiecksfilter bestimmen. Eine Berechnungsvorschrift dafür ist:

Für alle $m = 1$ bis NUMCHANS und $k = 1$ bis NFFT / 2 + 1:

$$H_m[k] = \begin{cases} 0 & \text{falls } k < f[m-1] \\ \frac{2(k - f[m-1])}{(f[m+1] - f[m-1])(f[m] - f[m-1])} & \text{falls } f[m-1] \leq k < f[m] \\ \frac{2(f[m+1] - k)}{(f[m+1] - f[m-1])(f[m+1] - f[m])} & \text{falls } f[m] \leq k \leq f[m+1] \\ 0 & \text{falls } k > f[m+1] \end{cases}$$

2.3 Plotten Sie die Filterparameter aus Aufgabe 2.2 über der linearen Frequenz.

2.4 Erweitern Sie Ihre Funktion `computeFeatures` aus Aufgabe 1 um die Mel-Filterbank-Analyse. Diese soll nach der Spektralanalyse erfolgen und nur dann ausgeführt werden, wenn `TARGETKIND = MELSPEC` angegeben ist.

Dabei erhalten Sie die Sprach-Features im mel-skalierten Spektralbereich aus der Filtermatrix $H_m[k]$ und dem linearen Spektrum $S_{\text{LIN}}[k]$ mittels

$$S_{\text{MEL}}[m, \tau] = \sum_{k=1}^{\text{NFFT}/2+1} H_m[k] |S_{\text{LIN}}[k, \tau]|$$

Zum Testen finden Sie ein Konfigurationsfile `configMELSPEC.con` mit dem unten stehenden Inhalt im zip-File. Schreiben Sie bitte eine Hauptfunktion `Uebung2.m`, die die erweiterte Merkmalsextraktion für die Audiodatei aus Übung 1, `13zz637a.wav`, mit der neuen Konfigurationsdatei testet.

```
# coding parameters 'configMELSPEC' by D.K.
TARGETKIND = MELSPEC
TARGETRATE = 100000.0
SAVECOMPRESSED = F
SAVEWITHCRC = F
SOURCEFORMAT = WAV
WINDOWSIZE = 250000.0
USEHAMMING = T
PREEMCOEF = 0.97
NUMCHANS = 24
ENORMALISE = T
```