SGN-24006 Analysis of Audio, Speech and Music Signals

Exercise 2

22.3.2017

1 Homework to solve 1p

(Exceptionally, this week’s homework is a Matlab question.)

Mel-scale is a perceptually-motivated frequency scale, defined as

$$f_{Mel} = \frac{2595 \log_{10}(1 + \frac{f_{Hz}}{700})}{2595}$$

Conversion from Mel-units back to Hertz-units can be done using

$$f_{Hz} = 700 \times (10^{f_{Mel}/2595} - 1);$$

The above conversions are implemented in

http://www.cs.tut.fi/sgn24006/exercises/E2/hz2mel.m
http://www.cs.tut.fi/sgn24006/exercises/E2/mel2hz.m

Question: Your task is to distribute 5 frequency bands between 100 Hz and 2000 Hz, so that the bandwidths of the 5 bands are equally large in Mel-frequency units. In other words, the subbands should be uniformly distributed on the Mel-frequency scale. Write the Matlab commands that calculate the boundaries of the 5 subbands in Hz units.

You can check your answer by verifying that 1) Lower boundary $L_1$ of subband 1 is 100 Hz. 2) Higher boundary $H_5$ of subband 5 is 2000 Hz. 3) The width in Mel-units, $hz2mel(H_b)-hz2mel(L_b)$ is the same for all bands $b = 1, 2, \ldots, 5$. Note that higher boundary of band $b$ equals the lower boundary of band $b + 1$.

2 Matlab 3p

Sinusoidal analysis and illustration of the results. See pages 24-39 on lecture slides "Mid-level representations for audio content analysis”


Sinusoidal model is a parametric representation, where the periodic components of the sound are represented with sinusoids having time-varying amplitudes, frequencies, and phases. The task in this exercise is to implement the analysis part of the sinusoidal model.

a) Apply windowing using 24 ms Hanning window on some input signal and compute its complex spectrum using FFT. You may increase the FFT
length for example to 4096 (help fft) to improve the frequency resolution. Retain only the positive frequency bins of the spectrum.

b) From the magnitude spectrum (help abs) of each analysis frame, pick the N highest maxima that are at least 50 Hz apart from each other. Use the pickpeak.m function

http://www.cs.tut.fi/~sgn24006/exercises/E2/pickpeak.m
(Note: remember to translate the 50 Hz distance to frequency-bin units.)

c) For each found maximum, determine its frequency \( f_n^k \), amplitude \( a_n^k \) and phase \( \varphi_n^k \), where \( n = 1, \ldots, N \) is the index of the maximum and \( k \) is frame index. Frequency corresponds to the sinusoidal location in the spectrum (convert from bin index to frequency). Amplitude is obtained by picking the magnitude spectrum value at the maximum and by scaling it with the function

\[
c = \frac{2}{\sum_{t=1}^{T} w_t},
\]

where \( w_t \) is the window function value at index \( t \) and \( T \) is frame length. Phases are picked directly from the phase spectrum (help angle) at the positions of the found maxima.

d) If there are less than \( N \) local maxima in the magnitude spectrum, use zero values for the frequencies, amplitudes and phases of the remaining sinusoids.

e) Illustrate the sinusoids that you found by plotting them on a graph, where x-axis has the frame index and y-index has the sinusoid frequency. Indicate each found frequency in each frame with a dot. Plot a similar graph about the amplitudes of the sinusoids. Visualize the results only for the first 100 frames.

For testing the analysis, use the signals:

One sinusoid, amplitude 1, frequency 305 Hz, \( N=1 \)

Two sinusoids, amplitudes 0.5 and 0.25, frequencies 400 Hz and 500 Hz, \( N=2 \)

Violin, example plot for checking, \( N=25 \).
Kuva 1: Example graph for problem 1 e), testsignal3.wav