Speech Signal Processing
Exercise 5 — Cepstral Techniques

Timo Gerkmann, Kristina Tesch

Download the archive "Exercise5.zip" from *moodle*. Load the file "filter-data.txt" into a numpy array, which contains some vocal tract filter coefficients (stored as $\mathbf{a} = [a_0\ a_1\ \dots a_{M-1}]$). The sampling frequency is $f_s = 16000$ Hz.

1. Generate a pulse-sequence with a length of 32 ms that represents a voiced excitation source signal with a fundamental frequency of 100 Hz, i.e. $\mathbf{u} = [1\,0\dots0\,1\,0\dots0\,1\,0\dots]$.

2. Create an artificial speech sound $\mathbf{s}$ by filtering the generated excitation source $\mathbf{u}$ with the provided vocal tract filter coefficients $\mathbf{a}$ using the function `lfilter(1, a, signal)`. Then, apply a Hann window $\mathbf{w}$ to $\mathbf{s}$ and $\mathbf{u}$. Plot the amplitude spectrum and the log amplitude spectrum of the vocal tract filter $\mathbf{h}$ (defined by the coefficients $\mathbf{a}$), of the windowed excitation source $\mathbf{u_w} = \mathbf{w} \otimes \mathbf{u}$, as well as of the windowed filtered signal $\mathbf{s_w} = \mathbf{w} \otimes \mathbf{s}$, where $\otimes$ denotes element-wise multiplication. To obtain the spectra, for $\mathbf{u_w}$ and $\mathbf{s_w}$ you can simply use `rfft`, while for $\mathbf{h}$, `H = freqz(b,a,N,whole=True)` might be a good choice, where N is the length of the signal segment in samples.

3. Compute and plot:

   - $c_s[n]$: real cepstrum of the windowed filtered signal $\mathbf{s_w}$.
   - $c_u[n]$: real cepstrum of the windowed excitation source $\mathbf{u_w}$.
   - $c_h[n]$: real cepstrum of the filter $\mathbf{h}$.

   (Hint: if the power of your signal for certain frequency bins is 0, computing log(0) returns `-inf`. To overcome this limitation, you can add the value `np.finfo(np.float64).eps` to your sequences prior to computing the logarithm if necessary.)

4. Answer the following questions:

   a) Is there any symmetry within the real cepstrum of the signals? Why?

   b) Why is the real (and also the complex) cepstrum of any real-valued time-domain signal real-valued?

   c) For voiced sounds, a cepstral peak at a distinct position can be observed. Why is that so? Explain how the position of the peak is related to the fundamental frequency.

5. Using $c_s[n]$, estimate the fundamental frequency of your excitation source and compare it to the value you used in point 1. For this, find the cepstral peak that corresponds to the fundamental frequency. Unfortunately, the values of the cepstrum in the lowest quefrencies are often much larger than the peak that we are searching for. Thus you can set the first cepstral coefficients to `-inf` before searching for the peak. In this way you can also effectively limit your maximum search to a quefrency region relevant for human fundamental frequencies. The found peak position can finally be translated into the fundamental frequency.

6. In speech processing, the actual filter and excitation source are unknown. Based on the theory that the first cepstral coefficients of $c_s[n]$ correspond to the vocal tract $h[n]$ while the remaining ones correspond to the excitation $u[n]$, write a function

   ```
   def reconstruct_filter_and_source(cepstrum: np.ndarray, fs: int, L: float) -> [np.
       ndarray (H_est), np.ndarray (U_est)]
   ```

   with arguments:

`cepstrum`	Vector containing the cepstral coefficients of the filtered signal $c_s[n]$.
`fs`	Sampling rate in Hz.
`L`	The cut-off quefrency in *ms*. The cepstral coefficients of $c_s[n]$ up to L correspond to the filter.
`H_est`	Vector containing the reconstructed spectrum of the filter using the first L coefficients in $c_s[n]$
`U_est`	Vector with the reconstructed spectrum of the excitation source using the remaining coefficients in $c_s[n]$

This function applies cepstral liftering to $c_s[n]$ to estimate the spectrum of the filter and the spectrum of the excitation. (Hint: take care that you don't destroy the symmetry of c_s[n]!)

a) Reconstruct the spectrum of the filter and excitation using $L \in \{0.5\,\text{ms},\ 1\,\text{ms},\ 1.5\,\text{ms},\ 2\,\text{ms},\ 10\,\text{ms},\ 20\,\text{ms}\}$ and compare them with the original spectra, H and U, that you have previously computed. Which value of L gives the best reconstruction? Explain what you observe and support your observations using appropriate plots.

b) Name at least one other possibility to obtain an estimate of the vocal tract transfer function H.

7. Load the file *speech1.wav*. Split the signal into segments of 32 ms overlapping by 16 ms.

a) Plot the logarithmic spectrogram as well as the real cepstrogram of the speech signal using `matplotlib.pyplot.imshow`.

b) Finally, plot the results after liftering (for your optimal value of L), that is the estimated spectral envelope H_{est} and the spectrogram of the estimated excitation U_{est}.