Midterm Lab Summary – Pitch Detection by Correlogram

Yiran Luo

Mar 3, 2016

A. Implementation Design

In this lab, we are tasked to find out the fundamental frequency (F0) by plotting correlgorams of two English vowels. Given all the conditions and all the Meddis hair cell model (i.e. function r) correlation values provided, from the original autocorrelation formula (Eq. 4, Wang and Brown, 1999):

$$\mathrm{A(i,j,\tau)} = \sum_{k=0}^{K-1} r(i,j-k)r(i,j-k-\tau)w(k)$$

we are essentially plotting the correlogram at the final "slice" where *j* is at time frame 325 with a constant window size *w*. So the formula is transcribed into:

$$\mathrm{A(i,j=325,\tau)} \sim C \sum_{k=0}^{199} r(i,j-k)r(i,j-k-\tau)$$

where C is a constant scale factor for plotting the proportional autocorrelation values.

For the plots in detail, the correlogram for each channel (i.e. a wave line) is placed at above the channel's center frequency. And the plot's Y-axis is converted into log-scale in order to show equal spacing in between the wave lines.

The source code is in Python and is also available at: { HYPERLINK "https://github.com/fpsluozi/CSE5539CALabs/tree/master/Midtermlab" } . NumPy and Matplotlib are used for generating the correlograms.

B. Plot Showcases and Analysis

In this part, we are showcasing the correlograms for /ar0/ and /er4/, as well as their pooled correlograms.

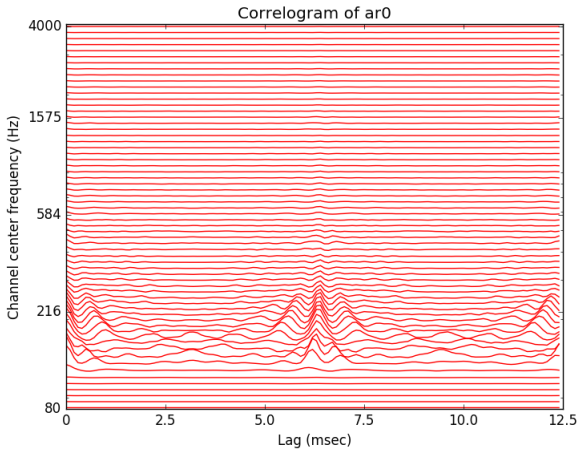

Fig 1. Correlogram of /ar0/

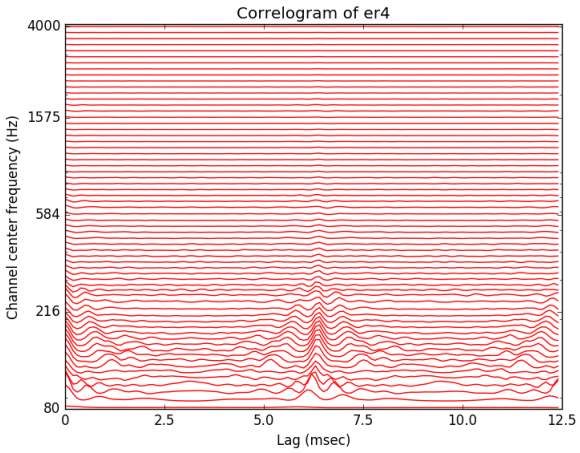

Fig 2. Correlogram of /er4/

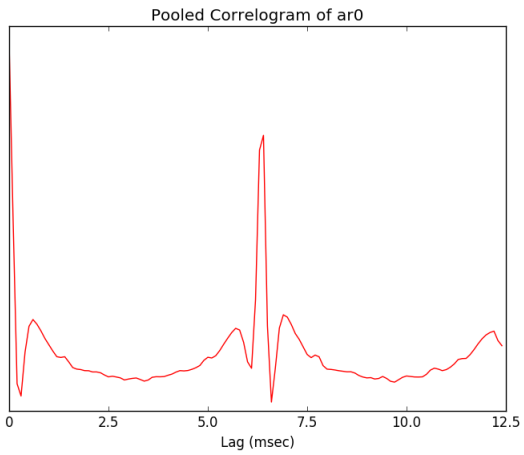

Fig 3. Pooled Correlogram of /ar0/

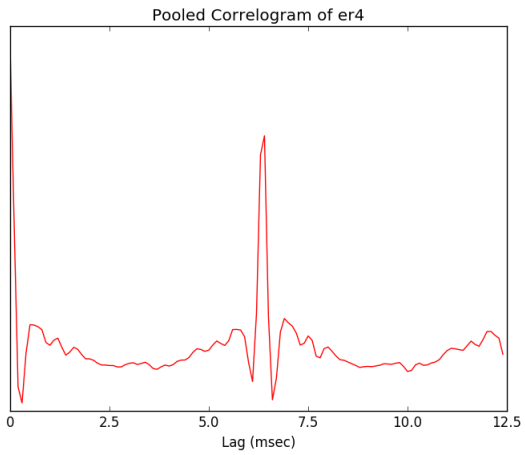

Fig 4. Pooled Correlogram of /er4/

For the pitch detection, we will be looking at at which lag frame, other than zero lag, the correlation is at the peak. As a result, in /ar0/ and /er4/, the maximum correlation is both reached at the lag frame indexed 64 (which can also be acquired by running the source code). This indicates the F0 is at around:

$$F_0 = \frac{1}{64 \times 0.0001 sec} \approx 156 Hz$$

This value is thus verified falling in the expected range between 80Hz and 222Hz.

C. Afterthoughts and Conclusion

We have hence found the consistency that the two English vowels embed the same fundamental frequency via peaks in correlograms. By experimenting in person, we have learned that this method introduced is both practical and intuitive.

On the other hand, the biggest difficulty we have encountered so far is not about calculation the autocorrelations, but about how to plot the correlograms. Initially, there were a lot of confusion trying to understand the meaning of each data point on the correlogram until we found out the correlation values are scalable. Once this is understood, the plotting from then on went smooth and eventually we managed to draw plots mimicking the original ones in the paper.