
Comparison - Pitch detection.

16 September 2019

Context:

Detect pitch of latest samples from a continuous incoming buffer.

Method:

Once the appropriate syllable is detected, a chunk of signal is selected and processed.

To select:

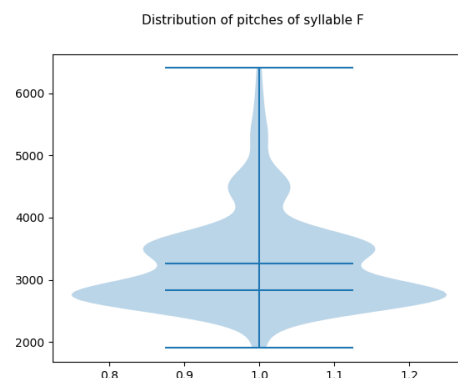
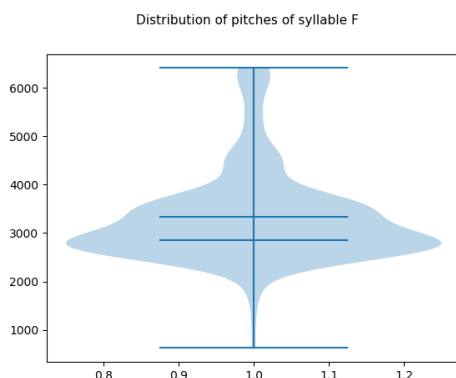
- Either 256 or 512 samples are chunked off for processing.
- These samples are selected either from the onset of syllable or with a delay of 100 or 200 samples.

Processing:

- Pitch is calculated by constructing 10 FFTs.
- Each FFT is constructed by using N , $N-2$, $N-4$,... samples.
- The frequency at the highest peak of the FFT in a given frequency range (with a 30% error margin) is taken to be the pitch.
- For e.g. Here, 600Hz - 1200Hz is considered for syllable F.

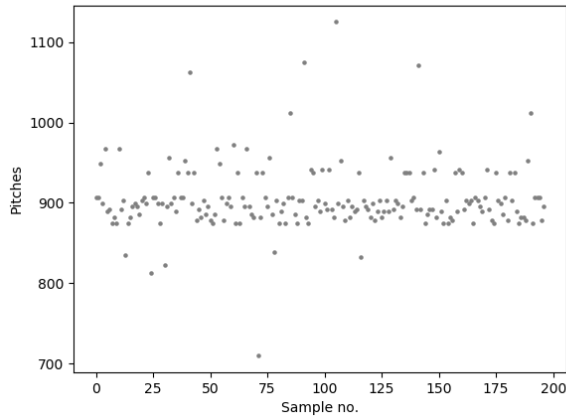
Comparison:

1. Here, the entire range of the FFT is considered when detecting the highest peak. On the left figure, some of the syllables have been manually removed due to erroneous identification. Only syllable F is considered.

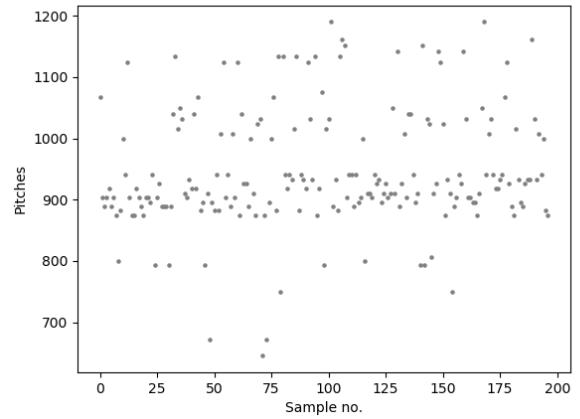


2. The pitch range is restricted to between 600Hz - 1200Hz. Distribution of pitches of syllable F is shown here when 512 samples are considered vs. 256 samples. These samples are selected from the onset of the syllable with 0, 100 and 200 samples lag.

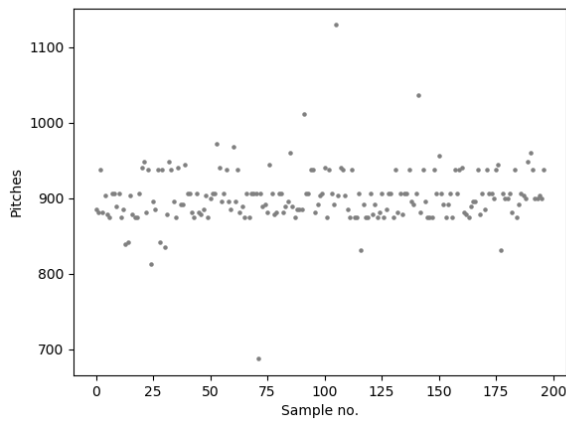
Distribution of pitches with N=512, lag=0.



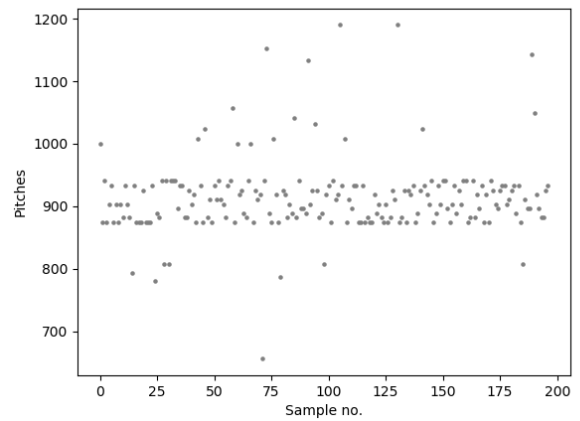
Distribution of pitches with N=256, lag=0.



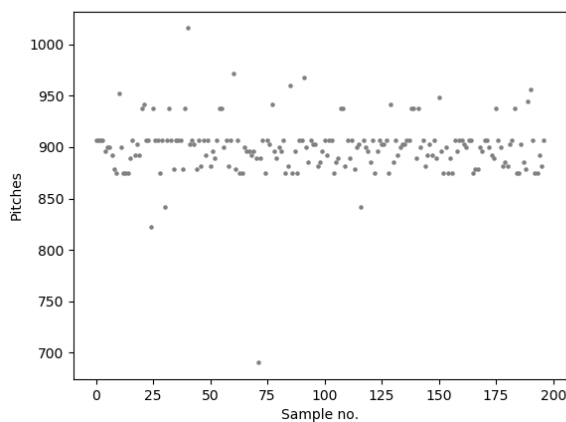
Distribution of pitches with N=512, lag=100.



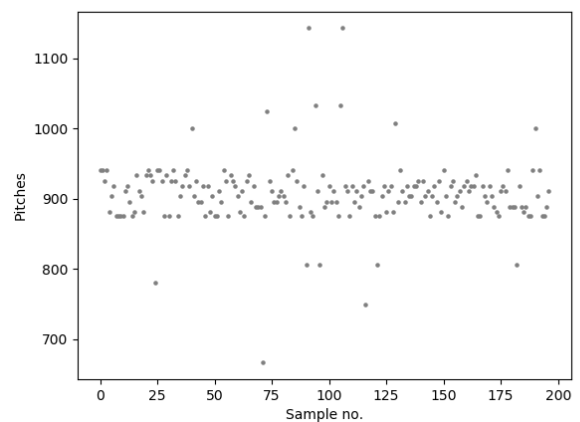
Distribution of pitches with N=256, lag=100.



Distribution of pitches with N=512, lag=200.



Distribution of pitches with N=256, lag=200.

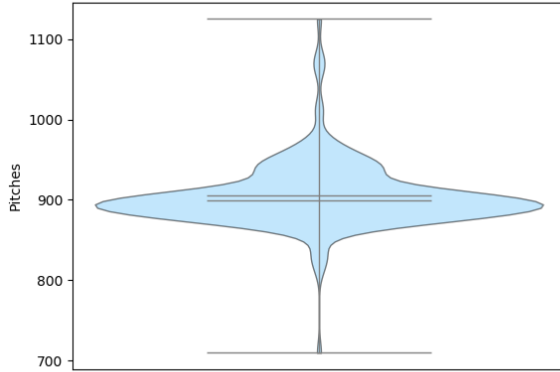


Resolution: 6.3507 Hz

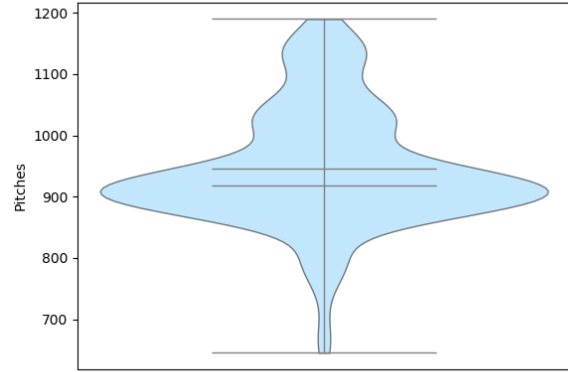
Resolution: 12.9101 Hz

3. Histograms for same as above.

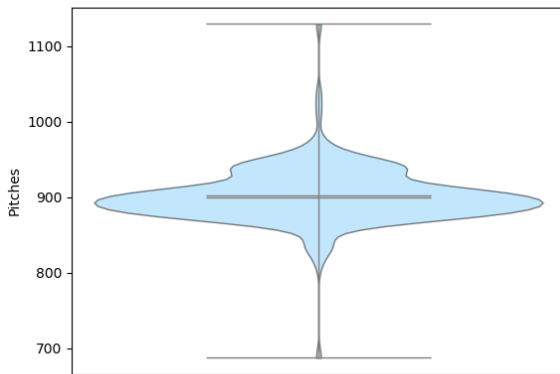
Histogram with N=512, lag=0.



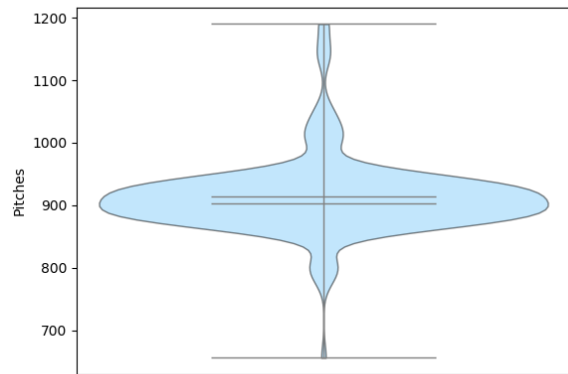
Histogram with N=256, lag=0.



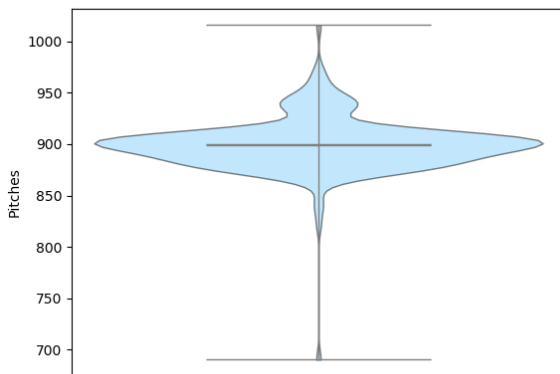
Histogram with N=512, lag=100.



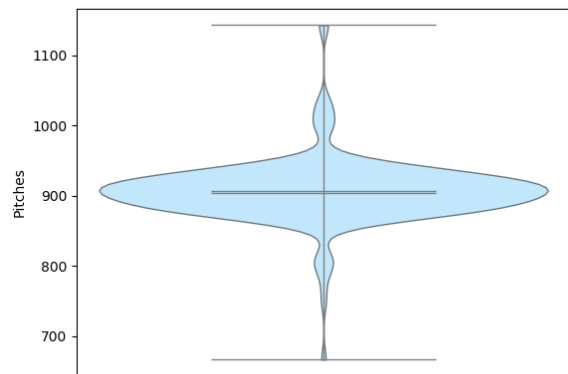
Histogram with N=256, lag=100.



Histogram with N=512, lag=200.



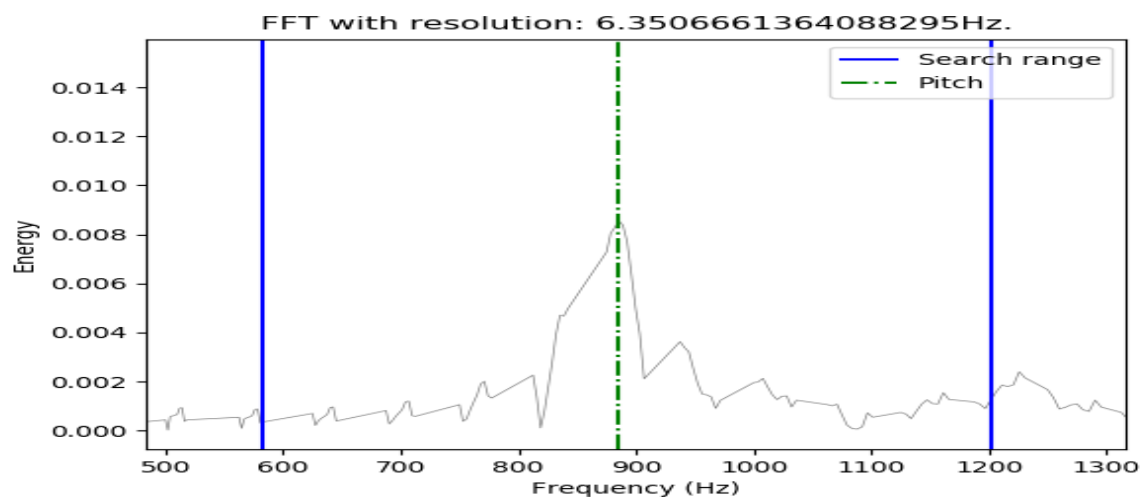
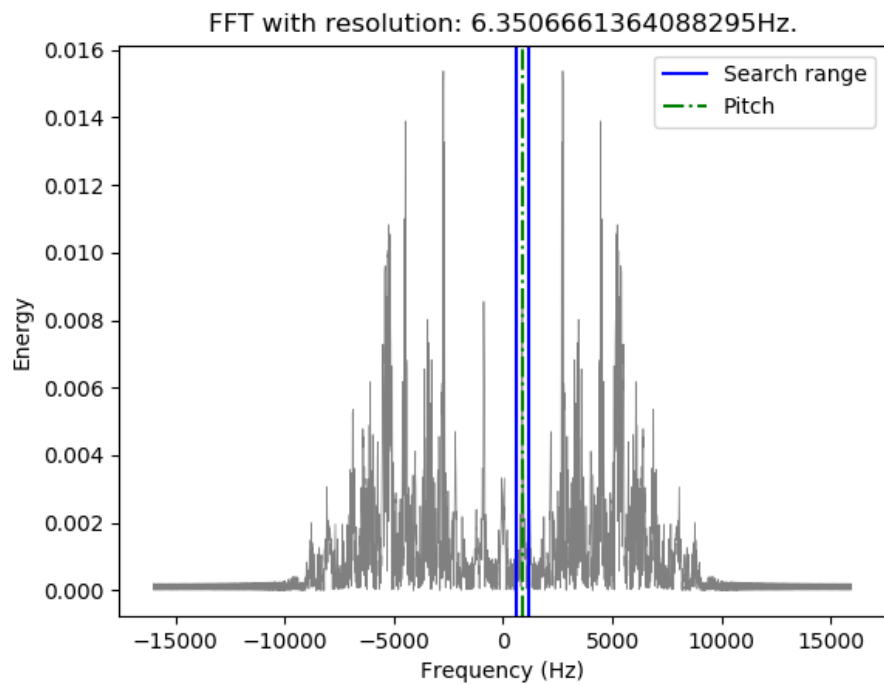
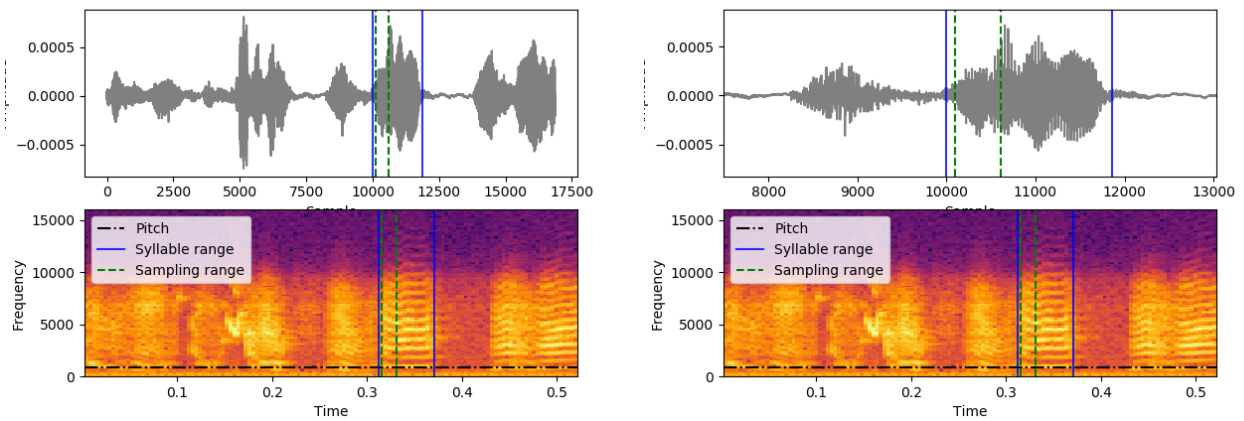
Histogram with N=256, lag=200.



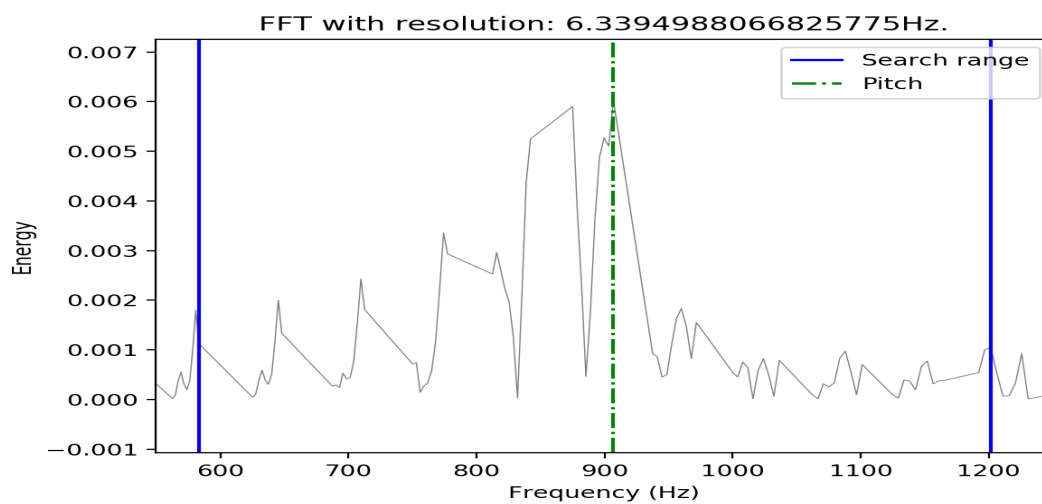
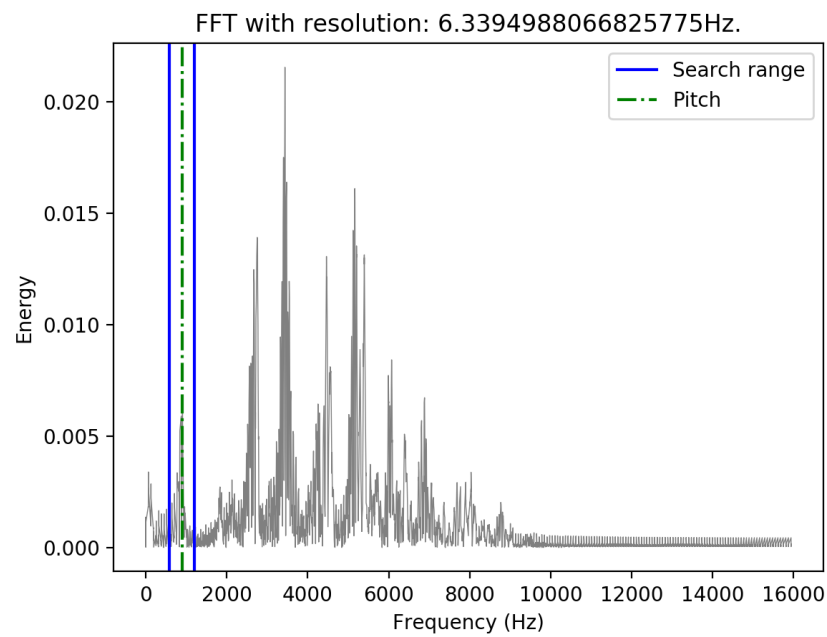
Resolution: 6.3507 Hz

Resolution: 12.9101 Hz

4. Sample syllable and FFT. $N = 512$; Lag = 100. Shows the onset of syllable vs start of sampling.

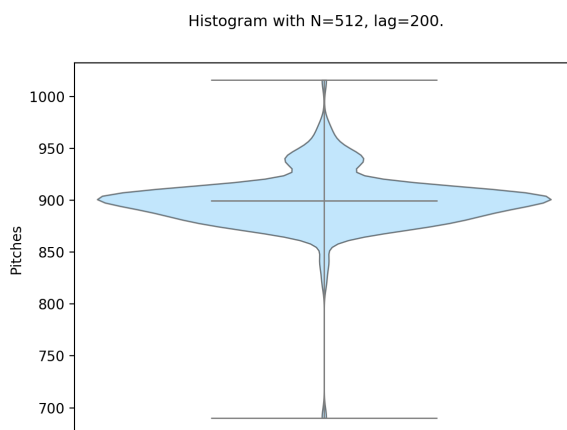
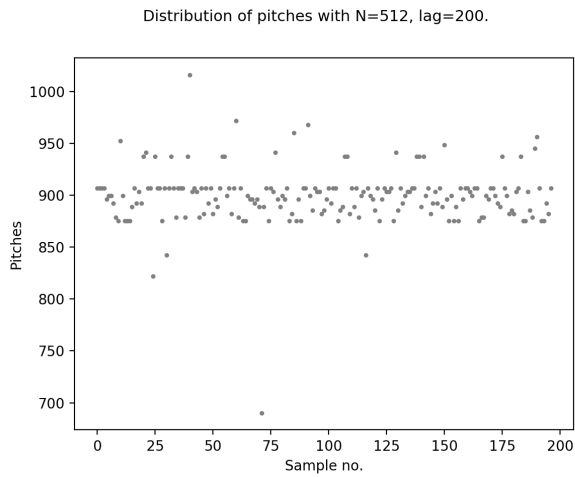


5. Sample FFT using C++. $N = 512$; Lag = 200.



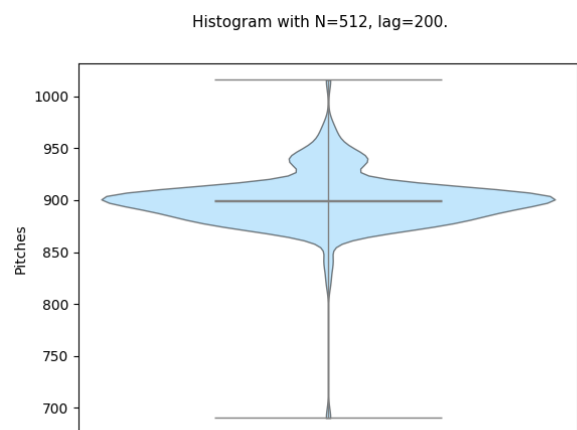
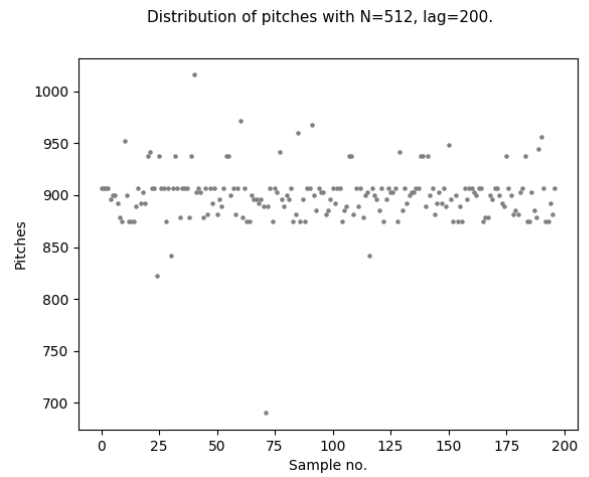
6. Pitch distribution using C++ vs. Python. $N = 512$; Lag = 200.

C++



Resolution: 6.3395 Hz

Python



Resolution: 6.3507 Hz