BME 544 Assignment 5

Kevin Xue

20814292

Estimating Fundamental Frequency, F_0

Method 1: ACF (Autocorrelation function)

This method finds the fundamental frequency, F0, by calculating the ACF of the signal sliced by StartIdx and StopIdx, then restricting the lag window for search to be 3.33 ms - 13.3 ms (prior knowledge of F0 mentioned in assignment instructions), and selecting the lag corresponding to the max ACF value, and converting it to a frequency value with units of Hz.

Method 2: PSD (Power spectral density)

This method finds the fundamental frequency, F0, by calculating the PSD of the signal sliced by StartIdx and StopIdx, then restricting the frequency window for search to be 75 Hz - 300 Hz (corresponds to the same lag window in the ACF method), and selecting the frequency corresponding to the max PSD function value.

Results

Table 1: Mean and Standard deviation of F0 estimate for method 1 vs 2, subject key (me, aka Kevin) = GD391P

	Method 1: ACF		**Method 2: PSD**	
Word	**F0 Mean**	**F0 Standard Deviation**	**F0 Mean**	**F0 Standard Deviation**
had	90.10	1.39	97.66	0.00
hawed	258.66	3.90	286.46	11.28
head	223.11	41.40	195.31	0.00
heed	178.33	68.00	214.84	0.00
hid	165.46	57.00	240.89	45.11
whod	220.19	106.69	292.97	0.00

Table 2: Mean and Standard deviation of F0 estimate for method 1 vs 2, subject key = 200008

	Method 1: ACF		**Method 2: PSD**	
Word	**F0 Mean**	**F0 Standard Deviation**	**F0 Mean**	**F0 Standard Deviation**
had	216.41	13.99	240.89	22.55
hawed	109.75	32.57	279.95	22.55
head	214.44	7.09	227.86	11.28
heed	222.52	9.90	227.86	11.28
hid	217.60	8.38	221.35	11.28
whod	246.81	18.15	253.91	19.53

F0 Estimation Precision/Accuracy Factors

Although ACF has multiple modes (normalized, unbiased), it made no difference on the mean and standard deviation of F0 estimates.

The relevant parameters used in Method 2 (PSD) are `segment_length`, `overlap`, and `nfft`. The default set of parameters being used are:

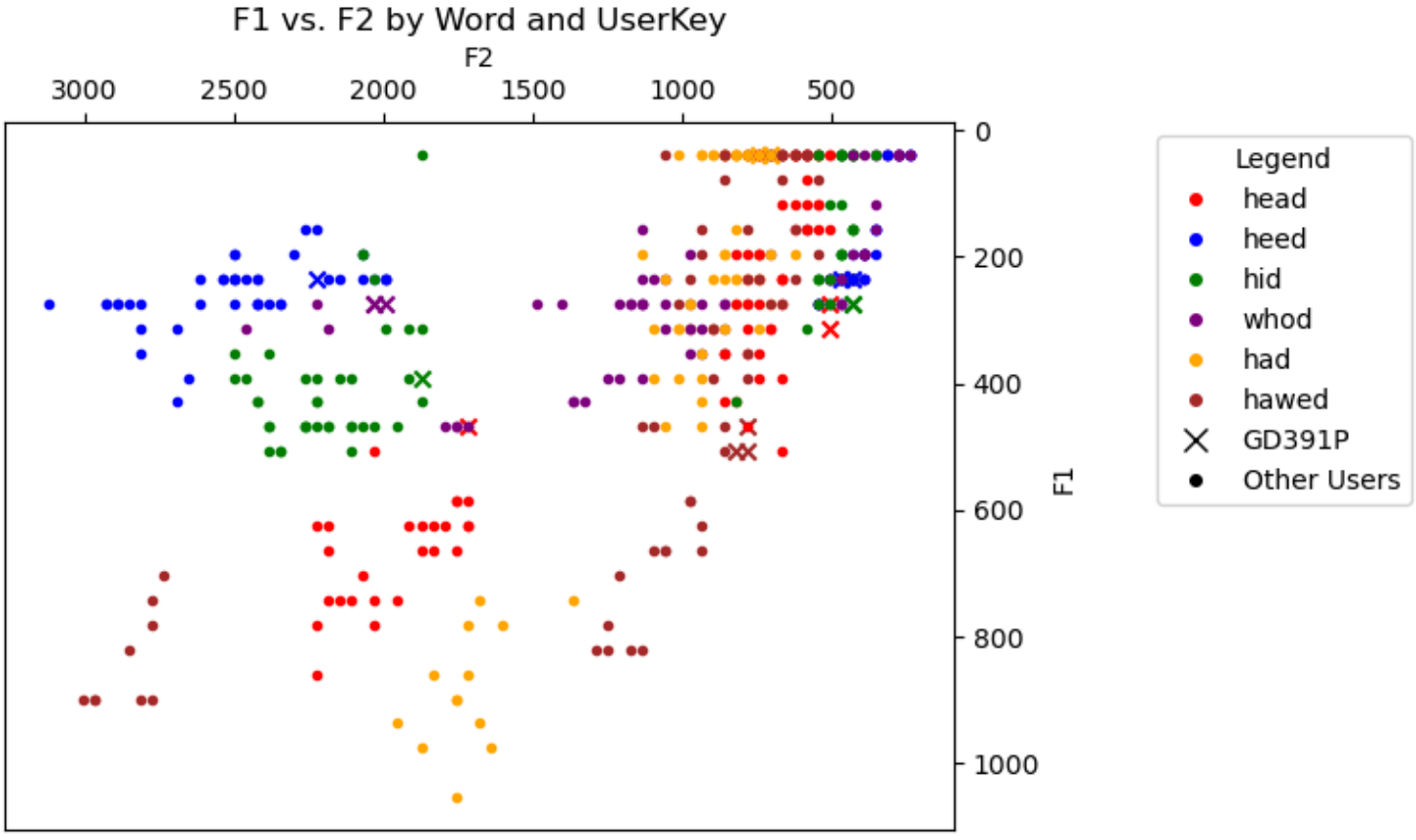
- `segment_length` = 256
- `overlap` = 128
- `nfft` = 512

Given the prior assumption that short segments of 25 - 50 ms can be considered quasi-steady state (mentioned in assignment instructions), the chosen `segment_length` is 256. The `overlap` of 128 and `nfft` of 512 was optimized to reduce the standard deviation of the estimates.

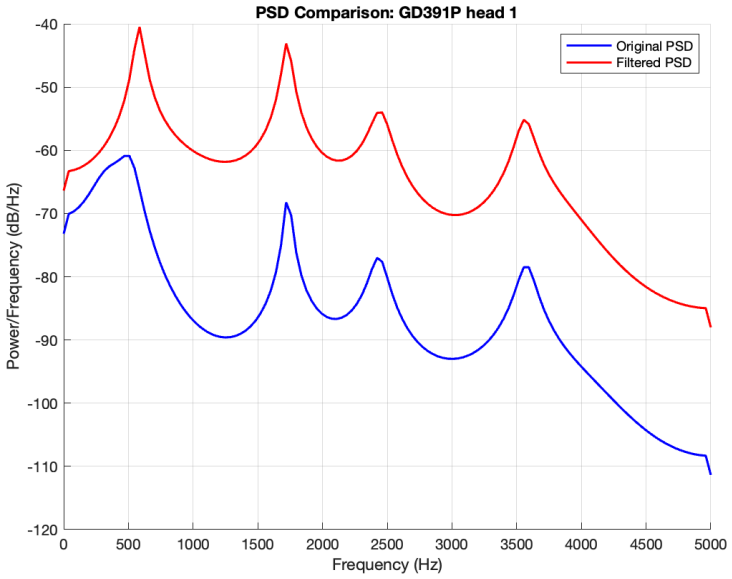
Shorter segments and increased overlap would decrease frequency resolution, but increase averaging, leading to the smoothing out of noise. Since `nfft` > `segment_length`, zero padding is applied, increasing frequency resolution.

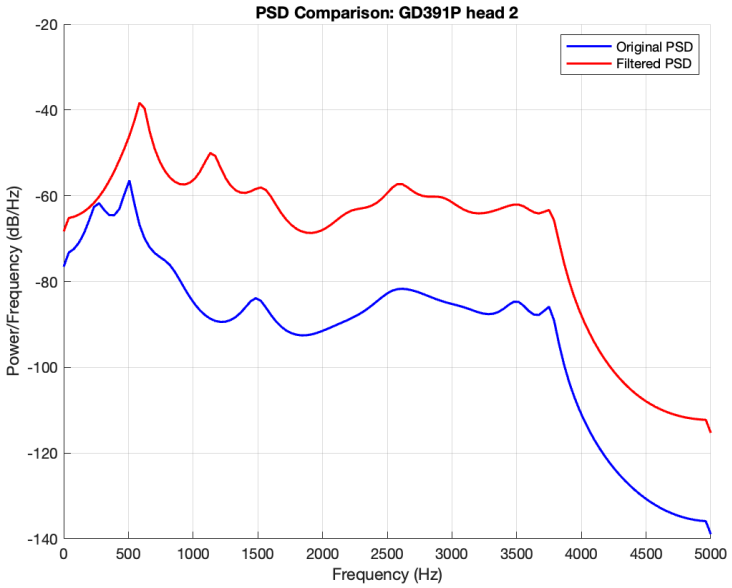
Estimating Formant Frequencies

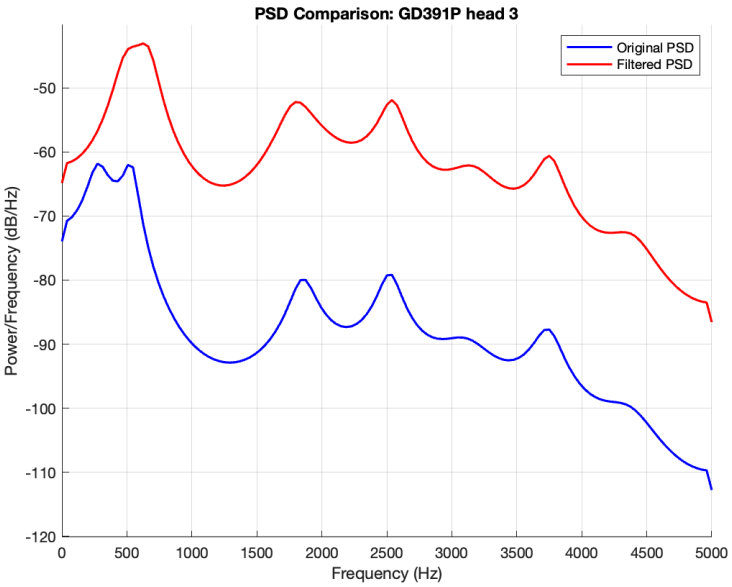
One way to estimate an appropriate model order of the IIR filter is by analyzing the PSD filter at various filter orders, and seeing at which minimum model order does a peak finding algorithm converge on its output within a certain tolerance. Additionally, designating the median model order across the three trials of a specific subject speaking a specific word will prevent overfitting of the IIR filter, allowing for better generalization when filtering new datasets.



Formant Shifting







Vocoding

	Recorded Speech	**Broadband Source**
Window Type	Hamming	Hamming
Segment Length	256	256
Overlap	128	128