# LPC for Formant Analysis of Concurrent Vowels


Taru Kapoor[a] Kshitij Khandelwal[a] Anubhav Sachan[a]

[a] Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani



***Abstract*— This report focuses on understanding the effects of noise on the formant representations of both single and concurrent vowels. The dataset used for this purpose included values of single and concurrent vowels at two fundamental frequencies, 100 Hz and 126 Hz respectively. With the help of Linear Predictive Coding (LPC), the formant frequencies were first found for these vowels. Following this, using various speech signals [6], Speech Spectrum Shaped Noise was generated. After this, the noise was added to the vowels for 3 different SNR values. The formant frequencies were then again estimated for these corrupted signals having different SNR levels. An analysis has been carried out to investigate the effects of the different SNR levels on the formants. Further, an attempt has been made to understand which vowels (both single and concurrent) are more susceptible to noise.**



***Keywords—LPC, Formants, Concurrent Vowels, Speech Spectrum Shaped Noise, SNR.***


## I. INTRODUCTION

The general character of the speech signal varies at the phoneme rate, which is on the order of 10 phonemes per second, while the detailed time variations of the speech waveform are at a much higher rate. That is, the changes in vocal tract configuration occur relatively slowly compared to the detailed time variation of the speech signal. The sounds created in the vocal tract are shaped in the frequency domain by the frequency response of the vocal tract. The resonance frequencies resulting from a particular configuration of the articulators are instrumental in forming the sound corresponding to a given phoneme. These resonance frequencies are called the formant frequencies of the sound. On analyzing the frequency spectrum of a speech signal, various peaks can be observed. The formant frequencies correspond to local maxima in the spectrum. To identify these formants, linear predictive coding proves extremely useful. On carrying out LPC analysis for a speech signal, a prediction polynomial $A(z)$ in $z^{-1}$ is obtained. In the pole-zero plot of $A(z)$, the zeros lying close to or on the unit circle correspond to the formant frequencies of the speech signal. Figure 1 shows the frequency spectrum of a speech signal along with the LPC spectrum of the signal. The peaks in the LPC spectrum correspond to the formant frequencies.

The report has been divided into IV sections. Section II presents the methodology used with brief explanations of the theory behind the report. Section III presents the results and discussion on the observations made. The final Section IV contains the conclusions, limitations and future scope of the project.

## II. METHODOLOGY

A dataset of single and concurrent vowels [6] was used. The generation and analysis was done in three stages – finding the formant frequencies with the help of LPC; generation of Speech Spectrum Shaped Noise; addition of Speech Spectrum Shaped Noise at different SNR values to the formant representation of the signals; analyzing the effects that the different SNR levels of noise have on the formant representation of both single and concurrent vowels.

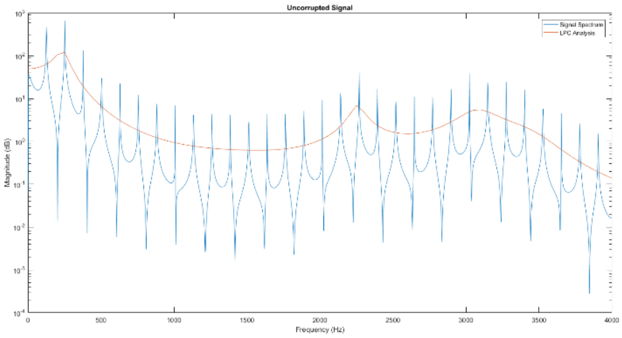


**Fig. 1: LPC analysis of a speech signal (in dB)**

### 1.LINEAR PREDICTIVE CODING [1]

For LPC analysis, the sampled speech signal is modeled as the output of a linear, slowly time-varying system excited by either quasi-periodic impulses (during voiced speech), or random noise (during unvoiced speech). Over short time intervals, the linear system is described by an all-pole function of the form:

$$H(z) = \frac{S(z)}{E(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

The major advantage of this model is that the gain parameter, $G$, and the filter coefficients $a_k$ can be estimated in a very straightforward and computationally efficient manner by the method of linear predictive analysis. The speech samples s[n] are related to the excitation e[n] by the difference equation:

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + Ge[n]$$

A linear predictor with prediction coefficients, αk, is defined as a system whose output is

$$\tilde{s}[n] = \sum_{k=1}^{p} \alpha_k s[n-k]$$

The prediction error is defined as

$$d[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^{p} \alpha_k s[n-k]$$

Hence, the prediction error sequence is the output of an FIR linear system whose system function is

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = \frac{D(z)}{S(z)}$$

If the speech signal obeys the model mentioned above exactly, and if $\alpha_k = a_k$, then $d[n] = Ge[n]$. Thus, the prediction error filter, *A(z)*, will be an inverse filter for the system, *H(z)*, i.e.,

$$H(z) = \frac{G}{A(z)}$$

The set of predictor coefficients $\{\alpha_k\}$ are to be found from the speech signal directly to obtain a useful estimate of the time-varying vocal tract system. This must be done to minimize the mean-squared prediction error over a short segment of the signal. The short-time average prediction error is defined as

$$E_{\hat{n}} = \langle d_{\hat{n}}^2[m] \rangle = \left\langle \left( s_{\hat{n}}[m] - \sum_{k=1}^{p} \alpha_k s_{\hat{n}}[m-k] \right)^2 \right\rangle$$

Where $s_{\hat{n}}[m]$ is a segment of speech that has been selected in a neighborhood of the analysis time, i.e.

$$s_{\hat{n}}[m] = s[m+\hat{n}] \qquad -M_1 \le m \le M_2$$

The notation < > denotes averaging over a finite number of samples. The values of $\alpha_k$ that minimize $E_{\hat{n}}$ can be found by setting $\partial E_{\hat{n}} \,/\, \partial \alpha_i$, for i = 1,2, . . . p, thereby obtaining the equation

$$\sum_{k=1}^{p} \alpha_k \langle s_{\hat{n}}[m-i] s_{\hat{n}}[m-k] \rangle = \langle s_{\hat{n}}[m-i] s_{\hat{n}}[m] \rangle$$

Where i = 1,2, ……. p.

Defining,

$$\varphi_{\hat{n}}[i,k] = \langle s_{\hat{n}}[m-i] s_{\hat{n}}[m-k] \rangle$$

Then the minimization condition can be written as

$$\sum_{k=1}^{p} \alpha_k \varphi_{\hat{n}}[i,k] = \varphi_{\hat{n}}[i,0] \qquad 1 \le i \le p$$

In matrix form, the above equation can be written as

$$\mathbf{\Phi}\alpha = \psi$$

The above matrix equation can be solved to find the vector $\alpha = \{\alpha_k\}$.

The minimum mean-squared prediction error can be shown to be

$$E_{\hat{n}} = \varphi_{\hat{n}}[0,0] - \sum_{k=1}^{p} \alpha_k \varphi_{\hat{n}}[0,k]$$

### *A. Computing the Predictor Coefficients using Autocorrelation Method*

There are two methods of linear predictive analysis – the Covariance method and Autocorrelation Method. The authors have used the Autocorrelation method for this project.

In the autocorrelation method, the analysis segment $s_{\hat{n}}[m]$ is defined as

$$s_{\hat{n}}[m] = \begin{cases} s[n+m]w[m] & -M_1 \le m \le M_2 \\ 0 & otherwise \end{cases}$$

where the analysis window w[m] is used to taper the edges of the segment to zero. The window is defined such that analysis segment is zero outside the range $-M_1 \le m \le M_2$. Hence, $d_{\hat{n}}[m]$ is non-zero only in the range $-M_1 \le m \le M_2$.

Therefore,

$$E_{\hat{n}} = \sum_{m=-M_1}^{M_2+p} \left( d_{\hat{n}}^2[m] \right) = \sum_{m=-\infty}^{\infty} \left( d_{\hat{n}}^2[m] \right)$$

Using this,

$$\varphi_{\hat{n}}[i,k] = \sum_{m=-\infty}^{\infty} s_{\hat{n}}[m] s_{\hat{n}}[m+|i-k|] = \phi_{\hat{n}}[|i-k|]$$

Thus, the resulting set of equations for the optimum predictor coefficients is therefore:

$$\sum_{k=1}^{p} \alpha_k \phi_{\hat{n}}[|i-k|] = \phi_{\hat{n}}[i] \qquad 1 \le i \le p$$

Some of the properties implied by the mathematical structure of the equations yielded by the autocorrelation method are

1. The mean-squared error is greater than 0. With this method, it is theoretically impossible for the error to be exactly zero because there will always be atleast one sample at the beginning and one at the end of the prediction error sequence that will be nonzero.
2. The matrix $\mathbf{\Phi}$ is a symmetric positive-definitive Toeplitz matrix.
3. The roots of the prediction error filter *A(z)* are guaranteed to lie within the unit circle of the z-plane so that the vocal tract model filter is guaranteed to be stable.
4. The equations can be solved efficiently by using the Levinson-Durbin algorithm.

### *B. The Levinson-Durbin Recursion*

The matrix equation $\mathbf{\Phi}\alpha = \psi$ can be written as

$$\begin{bmatrix} \phi[0] & \phi[1] & \dots & \phi[p-1] \\ \phi[1] & \phi[0] & \dots & \phi[p-2] \\ \dots & \dots & \dots & \dots \\ \phi[p-1] & \phi[p-2] & \dots & \phi[0] \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \phi[1] \\ \phi[2] \\ \dots \\ \phi[p] \end{bmatrix}$$

The vector $\psi$ has almost the same values as the matrix $\mathbf{\Phi}$. Due to this special structure, it is possible to design a recursive algorithm for inverting the matrix $\mathbf{\Phi}$.
The prediction error system function satisfies

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1})$$

*C. Roots of the Prediction Error System Function*

The system function of the prediction error filter is a polynomial in $z^{-1}$ and can be expressed in terms of its zeros as

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = \prod_{K=1}^{p} (1 - z_k z^{-1})$$

The zeros of A(z) are the poles of H(z). Figure 2 shows an example of the roots (marked with ×) of a $12^{th}$ order predictor. Note that eight (four complex conjugate pairs) of the roots are close to the unit circle. These are the poles of H(z) that model the formant resonances. The remaining four roots lie well within the unit circle, which means that they only provide for the overall spectral shaping.

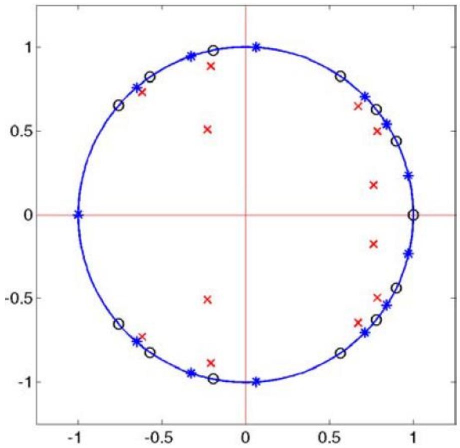


**Fig. 2: Poles of H(z) (zeros of A(z)) marked with × and LSP roots marked with * and o**

## 2. Speech Spectrum Shaped Noise

For our analysis, it was required to generate speech spectrum shaped noise. For this purpose, an input of 30 different speech samples was taken. [5] These speech samples were randomized and added together to generate the speech spectrum shaped noise. [3] (Fig. 4)

The sampling frequency of this noise is 10000 Hz; same as the sampling frequency of the audio files of the single and concurrent vowels, which were down sampled from 100000 Hz to 10000 Hz. This noise was then added to the vowels for 3 different SNR values i.e. 5 dB, 15 dB and 30 dB. Following this, LPC analysis was carried out on these corrupted signals to observe the effects of different noise levels on the formants.

The following figure (Fig. 3) shows the results before and after addition of Speech Spectrum Shaped Noise.

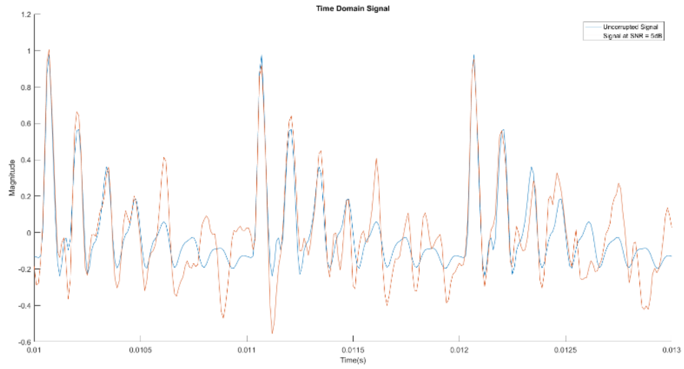


**Fig. 3: Time domain representation of 'ae' at a fundamental frequency of 100 Hz before and after corruption with Speech Spectrum Shaped Noise at an SNR of 5 dB**

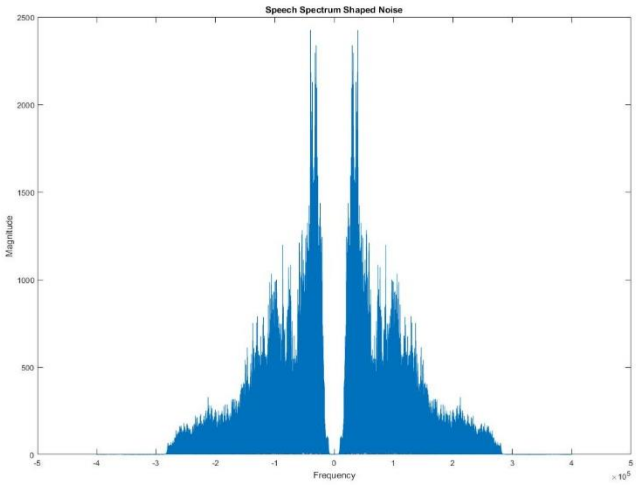


**Fig. 4: Speech Spectrum Shaped Noise**

## 3. Analysis Of Noise Effects On Single Vowels

Each of the single vowels considered for the analysis were with two different fundamental frequencies i.e. 100 Hz and 126 Hz. The vowels analyzed were AE, AH, EE, ER and OO. The analysis was carried out for two filters orders; 15 and 31. The frequencies highlighted in green are the formant frequencies. They were identified based on the following two criteria:

i. Bandwidth of the peak should be less than 100 Hz.
ii. If two formants are close (around 150 Hz), the one with the greater magnitude was chosen.

An explanation for the use of the aforementioned methods for thresholding has been provided in the discussion section of this report.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| AH 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 758.308 | 1051.41 | 1578.56 | 2954.03 | 3368.4 | 3991.47 | 5 |
| | Bandwidth (Hz) | 75.8437 | 309.633 | 26.5389 | 30.0061 | 506.862 | 48.4127 | 76.6419 | 837.536 | 255. |
| | Magnitude (dB) | 0.94098 | 0.90818 | 0.72513 | 0.963 | 0.9672 | 0.90909 | 0.67767 | 0.52891 | 0.34 |
| AE 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 756.431 | 1449.36 | 1545.87 | 2449.03 | 3364.26 | 3859.19 | 5 |
| | Bandwidth (Hz) | 80.8557 | 252.971 | 25.5179 | 31.5901 | 397.143 | 40.9124 | 77.0077 | 569.525 | 251. |
| | Magnitude (dB) | 0.90776 | 0.72922 | 0.94989 | 0.48886 | 0.96108 | 0.96844 | 0.6071 | 0.90339 | 0.72 |
| EE 100 | Root of Predictor Polynomial (Hz) | 85.6042 | 225.804 | 1501.08 | 2254.89 | 3049.57 | 3347.97 | 4769.57 | 5000 | |
| | Bandwidth (Hz) | 165.949 | 31.2696 | 303.13 | 29.0996 | 46.9049 | 72.3504 | 361.609 | 899.187 | |
| | Magnitude (dB) | 0.96147 | 0.81177 | 0.96409 | 0.68323 | 0.94276 | 0.91309 | 0.63482 | 0.32305 | |
| ER 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 446.567 | 1152.19 | 1204.09 | 2673.06 | 3463.62 | 4242.92 | 5 |
| | Bandwidth (Hz) | 68.8219 | 377.822 | 29.6242 | 63.4871 | 27.4865 | 151.097 | 154.718 | 160.378 | 161. |
| | Magnitude (dB) | 0.81598 | 0.81747 | 0.82331 | 0.82706 | 0.96605 | 0.92332 | 0.96346 | 0.91715 | 0.62 |
| OO 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 235.564 | 835.961 | 2116.01 | 2823.3 | 3501.72 | 4260.19 | 5 |
| | Bandwidth (Hz) | 70.7196 | 847.478 | 23.9209 | 41.4528 | 166.232 | 260.245 | 240.988 | 233.808 | 233. |
| | Magnitude (dB) | 0.7454 | 0.74542 | 0.73872 | 0.72106 | 0.81148 | 0.94924 | 0.97039 | 0.91497 | 0.34 |
| AH 126 | Root of Predictor Polynomial (Hz) | 135.174 | 757.026 | 1044.53 | 1671.38 | 2942.82 | 3362.06 | 3561.11 | 5000 | |
| | Bandwidth (Hz) | 121.394 | 15.2535 | 25.2213 | 306.126 | 55.0286 | 93.692 | 377.119 | 197.283 | |
| | Magnitude (dB) | 0.78043 | 0.88893 | 0.93319 | 0.62257 | 0.68066 | 0.9688 | 0.98101 | 0.85852 | |
| AE 126 | Root of Predictor Polynomial (Hz) | 159.271 | 756.848 | 1440.5 | 1562.49 | 2443.27 | 3362.5 | 3440.44 | 5000 | |
| | Bandwidth (Hz) | 107.322 | 14.3035 | 39.33 | 234.592 | 36.6378 | 172.744 | 130.202 | 550.763 | |
| | Magnitude (dB) | 0.955 | 0.84907 | 0.80487 | 0.50052 | 0.95178 | 0.74468 | 0.98219 | 0.87383 | |
| EE 126 | Root of Predictor Polynomial (Hz) | 173.111 | 239.095 | 1422.71 | 2253.98 | 3071.74 | 3160.55 | 3469.99 | 5000 | |
| | Bandwidth (Hz) | 136.305 | 18.4433 | 275.578 | 20.7085 | 60.7803 | 190.307 | 105.047 | 386.436 | |
| | Magnitude (dB) | 0.97431 | 0.61532 | 0.87634 | 0.92647 | 0.7873 | 0.7073 | 0.97709 | 0.84258 | |
| ER 126 | Root of Predictor Polynomial (Hz) | 197.482 | 481.693 | 1129.9 | 1274.9 | 2399.85 | 3300.43 | 4214.4 | 5000 | |
| | Bandwidth (Hz) | 91.8932 | 35.2765 | 19.4647 | 34.2533 | 191.097 | 161.07 | 173.443 | 231.532 | |
| | Magnitude (dB) | 0.74755 | 0.80416 | 0.81676 | 0.78652 | 0.95787 | 0.97584 | 0.95664 | 0.89094 | |
| OO 126 | Root of Predictor Polynomial (Hz) | 79.1043 | 240.469 | 863.053 | 1588.14 | 2249.58 | 3380.51 | 4262.31 | 5000 | |
| | Bandwidth (Hz) | 170.003 | 17.0434 | 26.685 | 401.621 | 45.5967 | 117.349 | 134.134 | 168.247 | |
| | Magnitude (dB) | 0.80943 | 0.84488 | 0.86289 | 0.94431 | 0.60369 | 0.96702 | 0.97881 | 0.80765 | |

**Fig. 5: Formants for uncorrupted signals with LPC filter order 15**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| AH 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 757.323 | 1048.32 | 1781.46 | 2957.96 | 3376.44 | 3900.26 | 5000 |
| | Bandwidth (Hz) | 84.8736 | 241.807 | 25.248 | 28.6257 | 402.797 | 50.4446 | 76.1317 | 884.976 | 251.49 |
| | Magnitude (dB) | 0.93858 | 0.90876 | 0.72904 | 0.96467 | 0.96877 | 0.89884 | 0.73796 | 0.6028 | 0.32887 |
| AE 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 756.403 | 1448.53 | 1543.16 | 2449.93 | 3363.63 | 3749.33 | 5000 |
| | Bandwidth (Hz) | 82.5844 | 244.804 | 25.4656 | 31.7132 | 391.919 | 41.4174 | 77.7146 | 537.344 | 239.725 |
| | Magnitude (dB) | 0.90696 | 0.7399 | 0.94928 | 0.50903 | 0.96093 | 0.96851 | 0.6111 | 0.90142 | 0.73519 |
| EE 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 227.908 | 1427.3 | 2257.36 | 3044.99 | 3342.67 | 4347.06 | 5000 |
| | Bandwidth (Hz) | 123.641 | 241.656 | 29.6872 | 195.324 | 28.8165 | 47.1547 | 70.2028 | 534.454 | 271.203 |
| | Magnitude (dB) | 0.96444 | 0.78235 | 0.96338 | 0.8561 | 0.7381 | 0.94247 | 0.91556 | 0.7112 | 0.51088 |
| ER 100 | Root of Predictor Polynomial (Hz) | 0 | 0 | 438.73 | 1132.78 | 1194.95 | 2601.43 | 3295.33 | 4208.44 | 5000 |
| | Bandwidth (Hz) | 56.4659 | 849.992 | 27.6367 | 101.297 | 20.6698 | 175.357 | 186.288 | 181.19 | 183.705 |
| | Magnitude (dB) | 0.79386 | 0.79637 | 0.79129 | 0.80223 | 0.9315 | 0.96587 | 0.97436 | 0.88048 | 0.34365 |
| OO 100 | Root of Predictor Polynomial (Hz) | 0 | 235.939 | 824.427 | 1736.7 | 2444.57 | 3141.66 | 3992.12 | 4672.54 | 0 |
| | Bandwidth (Hz) | 62.2813 | 22.0891 | 48.7092 | 180.788 | 210.738 | 198.046 | 246.318 | 274.716 | 0 |
| | Magnitude (dB) | 0.94063 | 0.97262 | 0.92472 | 0.79677 | 0.76734 | 0.77968 | 0.73379 | 0.70807 | 0 |
| AH 126 | Root of Predictor Polynomial (Hz) | 136.6 | 756.797 | 1044.27 | 1746.15 | 2944.16 | 3364.53 | 3632.81 | 5000 | 0 |
| | Bandwidth (Hz) | 122.411 | 15.6834 | 26.2955 | 277.483 | 56.8629 | 91.3027 | 444.592 | 205.922 | 0 |
| | Magnitude (dB) | 0.9675 | 0.98048 | 0.85742 | 0.70561 | 0.93104 | 0.8916 | 0.772 | 0.57196 | 0 |
| AE 126 | Root of Predictor Polynomial (Hz) | 153.482 | 757.073 | 1441.35 | 1585.6 | 2446.74 | 3338.3 | 3531.6 | 5000 | 0 |
| | Bandwidth (Hz) | 112.938 | 14.845 | 37.4844 | 255.079 | 37.6808 | 125.57 | 162.016 | 906.905 | 0 |
| | Magnitude (dB) | 0.81579 | 0.85402 | 0.95375 | 0.95399 | 0.98152 | 0.86769 | 0.72576 | 0.31993 | 0 |
| EE 126 | Root of Predictor Polynomial (Hz) | 48.4169 | 240.494 | 1374.42 | 2254.3 | 3070.66 | 3166.36 | 3465.32 | 5000 | 0 |
| | Bandwidth (Hz) | 165.731 | 16.5158 | 169.964 | 21.9549 | 60.379 | 192.532 | 105.457 | 310.436 | 0 |
| | Magnitude (dB) | 0.97279 | 0.67699 | 0.87588 | 0.92693 | 0.7851 | 0.80768 | 0.97946 | 0.81199 | 0 |
| ER 126 | Root of Predictor Polynomial (Hz) | 115.353 | 460.345 | 1145.34 | 1269.77 | 2466.41 | 3139.2 | 4297.3 | 5000 | 0 |
| | Bandwidth (Hz) | 145.008 | 37.1928 | 22.8144 | 57.1553 | 132.04 | 93.6879 | 198.639 | 346.214 | 0 |
| | Magnitude (dB) | 0.88893 | 0.7791 | 0.64722 | 0.84711 | 0.9307 | 0.97174 | 0.95434 | 0.83342 | 0 |
| OO 126 | Root of Predictor Polynomial (Hz) | 0 | 0 | 243.158 | 845.124 | 1696.48 | 2474.22 | 3153.65 | 4483.94 | 5000 |
| | Bandwidth (Hz) | 72.9993 | 1628.69 | 12.8574 | 43.6856 | 148.006 | 140.973 | 102.415 | 191.827 | 811.828 |
| | Magnitude (dB) | 0.7858 | 0.87924 | 0.83765 | 0.83028 | 0.94658 | 0.98397 | 0.91235 | 0.36053 | 0.12916 |

**Fig. 6: Formants for corrupted signals (SNR=30 dB) with LPC filter order 15**

Fig. 7 and Fig. 8 show the above analysis for one of the single vowels – AE, at a fundamental frequency of 100 Hz, for an LPC order of 15, with and without noise at 15 dB SNR, respectively.

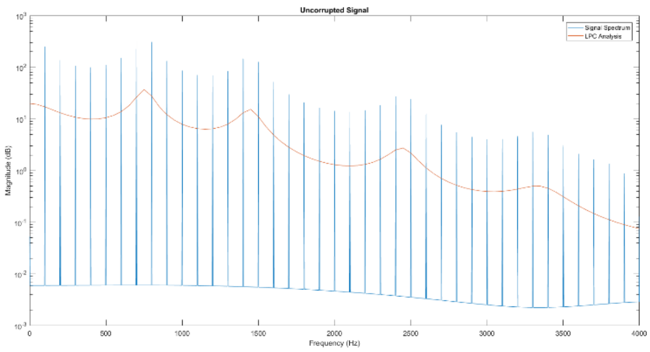


**Fig. 7: Analysis of uncorrupted 'ee', for a fundamental frequency of 100 Hz with LPC filter order 15**

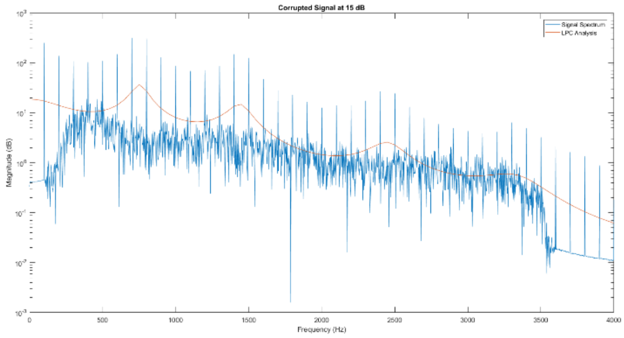


**Fig. 8: Analysis of 'ee' corrupted with Speech Shaped Spectrum Noise at an SNR of 15 dB, for a fundamental frequency of 100 Hz with LPC filter order 15**

### 4. Analysis Of Noise Effects On Concurrent Vowels

The concurrent vowels analyzed were AH-AE, EE-AH, EE-OO at fundamental frequencies of 100 Hz (both) and 100 Hz + 126 Hz respectively. The analysis was carried out for filter order of 31 only. The formant frequencies were identified based on the following two criteria:

i. Bandwidth of the peak should be less than 50 Hz.

ii. If two formants are close (around 150 Hz), the one with the greater magnitude was chosen.

An explanation for the use of the aforementioned methods for thresholding and the results of using the same have been comprehensively discussed in the following section.

## III. Results And Discussion

### A. Single Vowels

The analysis of single vowels showed that order 15 LPC filter was more suitable as compared to order 31. [4] The higher order LPC filter returned intermediate local maxima which could be falsely interpreted as formant frequencies. The roots of the prediction polynomial were estimated. However, to truly interpret a certain root as the formant frequency, two methods were used, namely thresholding based on bandwidth and magnitude. The formants are sharp peaks in the frequency spectrum. Hence the ones with a lower bandwidth and hence higher sharpness were used to determine the formants. After the bandwidth thresholding, magnitude thresholding was performed, wherein if two peaks occurred close to each other, (in a range of ≈150 Hz), the one with the higher magnitude was chosen as the Formant frequency.

| Uncorrupted | F0 | F1 | F2 | F3 |
|---|---|---|---|---|
| "ah" 100Hz | 758.3 | 1051.4 | 2954.0 | 3368.4 |
| "ah" 126Hz | 757.0 | 1044.5 | 2942.8 | 3362.1 |
| "ae" 100Hz | 756.4 | 1449.4 | 2449.0 | 3364.3 |
| "ae" 126Hz | 756.8 | 1440.5 | 2443.3 | - |
| "ee" 100Hz | 225.8 | 2254.9 | 3049.6 | 3348.0 |
| "ee" 126Hz | 239.1 | 2254.0 | 3071.7 | - |
| "er" 100Hz | 446.6 | 1204.1 | 2673.1 | - |
| "er" 126Hz | 481.7 | 1274.9 | - | - |
| "oo" 100Hz | 235.6 | 836.0 | - | - |
| "oo" 126Hz | 240.5 | 863.1 | 2249.6 | - |

**Table. 1: Results for Uncorrupted Signals at LPC order of 15**

| SNR 30 dB | F0 | F1 | F2 | F3 |
|---|---|---|---|---|
| "ah" 100Hz | 757.3 | 1048.3 | 2958.0 | 3376.4 |
| "ah" 126Hz | 756.8 | 1044.3 | 2944.2 | 3364.5 |
| "ae" 100Hz | 756.4 | 1448.5 | 2449.9 | 3363.6 |
| "ae" 126Hz | 757.1 | 1441.3 | 2446.7 | - |
| "ee" 100Hz | 227.9 | 2257.4 | 3045.0 | 3342.7 |
| "ee" 126Hz | 240.5 | 2254.3 | 3070.7 | - |
| "er" 100Hz | 438.7 | 1194.9 | - | - |
| "er" 126Hz | 460.3 | 1269.8 | 3139.2 | - |
| "oo" 100Hz | 235.9 | 824.4 | - | - |
| "oo" 126Hz | 243.2 | 845.1 | - | - |

**Table 2: Results for SNR value of 30 dB at LPC order of 15**

| SNR 15 dB | F0 | F1 | F2 | F3 |
|---|---|---|---|---|
| "ah" 100Hz | 751.9 | 1041.1 | 3421.8 | - |
| "ah" 126Hz | 755.5 | 1047.4 | 3284.4 | - |
| "ae" 100Hz | 755.4 | 1443.7 | 2455.3 | 3345.7 |
| "ae" 126Hz | 759.2 | 1442.7 | 2480.7 | 3238.9 |
| "ee" 100Hz | 217.9 | 2269.3 | 3062.2 | 3376.9 |
| "ee" 126Hz | 233.2 | 2275.6 | 3068.9 | 3411.8 |
| "er" 100Hz | 442.1 | 1176.0 | 3228.7 | - |
| "er" 126Hz | 432.7 | 1167.7 | 2705.7 | 3236.3 |
| "oo" 100Hz | 198.3 | - | 3183.3 | - |
| "oo" 126Hz | 231.9 | 2729.5 | 3251.4 | - |

**Table 3: Results for SNR value of 15 dB at LPC order of 15**

| SNR 5 dB | F0 | F1 | F2 | F3 |
|---|---|---|---|---|
| "ah" 100Hz | 784.0 | - | 3350.1 | - |
| "ah" 126Hz | 808.4 | - | 3259.0 | - |
| "ae" 100Hz | 755.0 | 1427.1 | 2567.5 | 3216.2 |
| "ae" 126Hz | 299.4 | 756.1 | 1435.6 | 3307.5 |
| "ee" 100Hz | - | - | 3065.7 | 3378.8 |
| "ee" 126Hz | 256.2 | - | 3287.8 | - |
| "er" 100Hz | 450.5 | 1172.9 | 1726.6 | 2864.9 |
| "er" 126Hz | 435.6 | 1149.8 | 2731.1 | 3251.0 |
| "oo" 100Hz | 233.4 | 2779.3 | 3276.7 | - |
| "oo" 126Hz | 259.3 | 2723.3 | 3248.6 | - |

**Table 4: Results for SNR value of 5 dB at LPC order of 15**

The results show that the formant frequencies of a certain vowel with different fundamental frequencies (F0) have approximately the same Formant frequencies, as expected. Vowels like ‘ae’ and ‘er’ have stronger and more prominent formant peaks as compared to ‘oo’. (Figs. 7, 10 and 11) From the tables, it’s clear that certain vowels are more susceptible to noise as compared to the others.

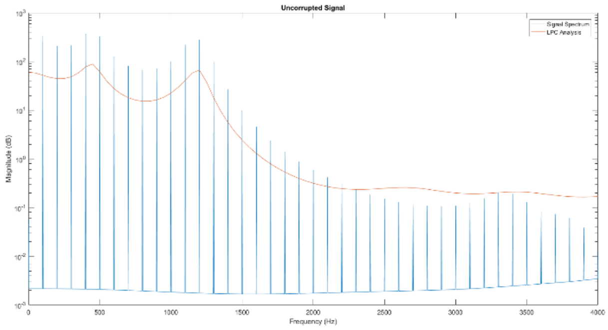


**Fig. 10: Frequency Spectrum of uncorrupted ‘er’ at a fundamental frequency of 100 Hz with LPC filter order 15 (More prominent formant peaks are seen)**

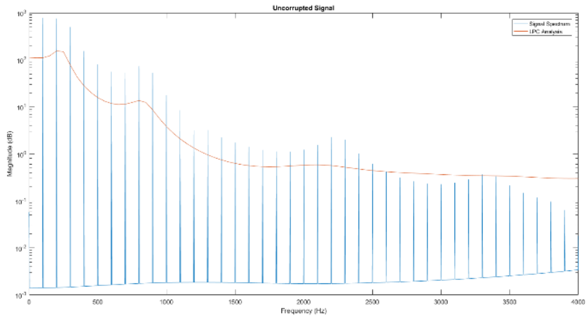


**Fig. 11: Frequency Spectrum of uncorrupted ‘oo’ at a fundamental frequency of 100 Hz with LPC filter order 15 (Less prominent formant peaks are seen)**

However, vowels with lesser fundamental frequency (100 Hz) are more immune to noise. ‘er’ and ‘ae’ be recognized at an SNR of 5dB (Figs. 12 and 13), unlike all the other samples. However, ‘oo’ couldn’t be recognized at an SNR of 15dB (Fig. 14). To identify single vowels, it is sufficient to test their formant frequencies F1 and F2 and the difference between the two [7].

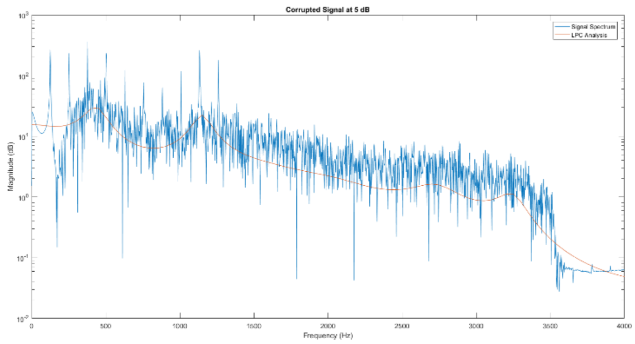


**Fig. 12: Frequency Spectrum of ‘er’ corrupted with noise of 5dB at a fundamental frequency of 100 Hz with LPC filter order 15 (More immune to noise)**

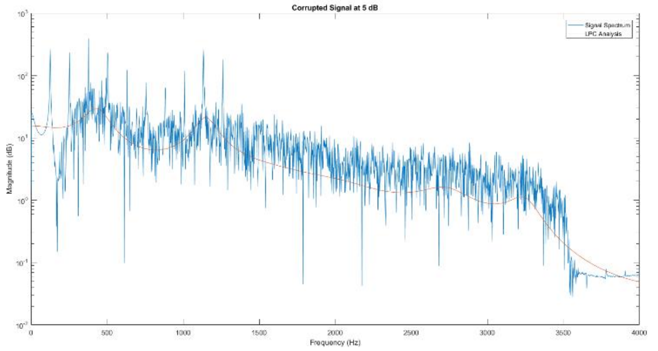


**Fig. 13: Frequency Spectrum of 'er' corrupted with noise of 5dB at a fundamental frequency of 126 Hz with LPC filter order 15 (Less immune to noise)**

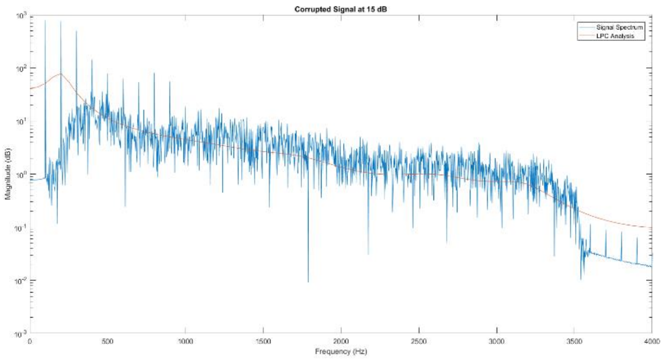


**Fig. 14: Frequency Spectrum of 'oo' corrupted with noise of 15dB at a fundamental frequency of 100 Hz with LPC filter order 15 (Less immune to noise)**

### *B. Concurrent Vowels*

A similar analysis was done for Concurrent Vowels. However, it was observed that the filter order had to be increased from 15 as used for Single vowels. The filter order had to be chosen in such a way so that it estimates the formant frequencies which correspond to the formant frequencies of the constituting vowels. Hence, LPC of order 31 was sufficient for the above case.

The bandwidth thresholding was performed for a bandwidth for <50Hz and magnitude thresholding was done for a range of 150 Hz.

| Uncorrupted | F'1 | F'2 | F'3 | F'4 | F'5 | F'6 |
|---|---|---|---|---|---|---|
| "ah" 100 Hz "ae" 100 Hz | 765.3 | 982.7 | 1446.8 | 2475.9 | 2998.6 | 3356.6 |
| "ee" 100 Hz "ah" 100 Hz | 169.2 | 728.8 | 888.7 | 1072.0 | 2332.1 | 2974.6 |
| "ee" 100 Hz "oo" 100 Hz | 141.6 | 296.9 | 850.9 | 2259.0 | 3072.4 | |
| "ah" 100 Hz "ae" 126 Hz | 760.5 | 1016.1 | 1501.9 | 2482.3 | 2984.0 | 3381.0 |
| "ee" 100 Hz "ah" 126 Hz | 174.0 | 753.4 | 996.7 | 2276.6 | 2967.6 | 3135.3 |
| "ee" 100 Hz "oo" 126 Hz | 273.3 | 876.8 | 2296.2 | 2996.1 | 3145.9 | 3366.8 |

**Table 5: Results for uncorrupted signals at LPC order of 31**

| SNR 30dB | F'1 | F'2 | F'3 | F'4 | F'5 | F'6 |
|---|---|---|---|---|---|---|
| "ah" 100 Hz "ae" 100 Hz | 762.8 | 983.9 | 1449.1 | 2478.6 | 3001.2 | - |
| "ee" 100 Hz "ah" 100 Hz | 175.4 | 738.5 | 1081.7 | 2250.1 | 3007.7 | 3392.6 |
| "ee" 100 Hz "oo" 100 Hz | 285.3 | 824.5 | 2269.2 | 3368.7 | - | - |
| "ah" 100 Hz "ae" 126 Hz | 760.0 | 1017.1 | 1502.3 | 2488.0 | 2968.0 | 3392.6 |
| "ee" 100 Hz "ah" 126 Hz | 162.8 | 751.9 | 1001.5 | 2265.9 | 3006.2 | 3395.2 |
| "ee" 100 Hz "oo" 126 Hz | 272.4 | 872.1 | 2280.2 | 3013.6 | 3381.6 | 3924.5 |

**Table 6: Results for SNR value of 30 dB at LPC order of 31**

| SNR 15dB | F'1 | F'2 | F'3 | F'4 | F'5 | F'6 |
|---|---|---|---|---|---|---|
| "ah" 100 Hz "ae" 100 Hz | 755.5 | 984.4 | 1449.9 | - | - | - |
| "ee" 100 Hz "ah" 100 Hz | 155.6 | 751.6 | 1007.5 | 2261.3 | 3389.7 | - |
| "ee" 100 Hz "oo" 100 Hz | 153.8 | 758.1 | 1017.1 | 1491.3 | 3380.4 | - |
| "ah" 100 Hz "ae" 126 Hz | 146.2 | 301.1 | 2262.9 | - | - | - |
| "ee" 100 Hz "ah" 126 Hz | 161.4 | 748.9 | 1060.1 | 2267.0 | - | - |
| "ee" 100 Hz "oo" 126 Hz | 158.0 | 273.4 | 2264.7 | 3028.0 | - | - |

**Table 7: Results for SNR value of 15 dB at LPC order of 31**

| SNR 5dB | F'1 | F'2 | F'3 | F'4 | F'5 | F'6 |
|---|---|---|---|---|---|---|
| "ah" 100 Hz "ae" 100 Hz | 444.0 | 752.5 | 992.1 | - | - | - |
| "ee" 100 Hz "ah" 100 Hz | 156.0 | 411.4 | 750.4 | 1010.7 | 3376.1 | - |
| "ee" 100 Hz "oo" 100 Hz | 448.0 | 756.7 | 1028.4 | 3373.2 | - | - |
| "ah" 100 Hz "ae" 126 Hz | 163.8 | 3362.2 | - | - | - | - |
| "ee" 100 Hz "ah" 126 Hz | 422.9 | 761.7 | 1032.0 | - | - | - |
| "ee" 100 Hz "oo" 126 Hz | 199.5 | 3359.7 | - | - | - | - |

**Table 8: Results for SNR value of 30 dB at LPC order of 31**

The results obtained show that formant frequencies obtained for concurrent vowels do not necessarily correspond to the formant frequencies of the constituting vowels. This can be attributed to the following reasons:

a. The formant frequencies of the two vowels are same or very close to each other (≈50 Hz). The case of 'ah' (100 Hz) and 'ae' (100 Hz) can be considered here. Their F1 frequencies are 758 Hz and 756 Hz, which shows up as F'1 (765.3 Hz) in the formant representation of their concurrent vowel. (Fig. 15)

b. In the case of vowels with same fundamental frequencies, the harmonics, situated at integral multiples of the fundamental frequencies, superimpose to give maxima, which does not necessarily correspond to any formant frequency of any of the constituting vowel. Eg. F'3 of 'ee'(100 Hz) and 'ah'(100 Hz).

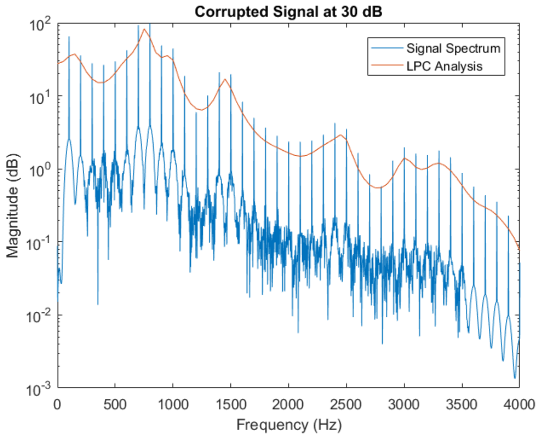


**Fig. 15: Frequency Spectrum of 'ah-ae' corrupted with noise of 30 dB at a fundamental frequency of 100 Hz (both) with LPC filter order 31 (F1 frequencies are very close to each other which shows up at F'1)**

The concurrent vowels are much more prone to noise as can be seen from the attached figure. Also the vowels with different fundamental frequencies (F0) are less prone to noise as compared to the ones with same fundamental frequency [8]. (Figs. 15 and 16).

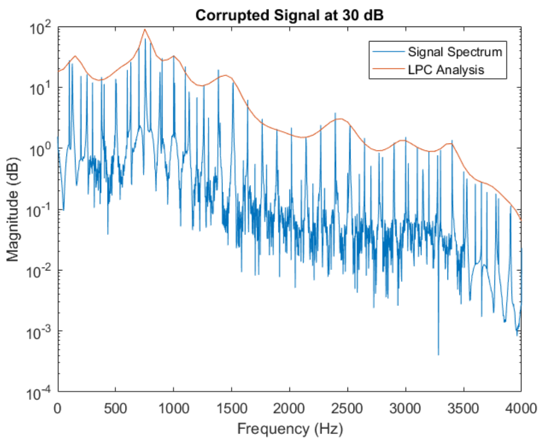


**Fig. 16: Frequency Spectrum of 'ah-ae' corrupted with noise of 30 dB at a fundamental frequency of 100 Hz (ah) and 126 Hz (ae) with LPC filter order 31 (Less prone to noise)**

## IV. Conclusion and Scope For Further Research

From the analyses above it the authors conclude that LPC analysis isn't very efficient in identifying the formants in case of concurrent vowels. However, the analysis works well enough for single vowels. Another method to identify concurrent vowels is with the help of a neural cancellation filters which are capable of segregating weak targets from competing harmonic backgrounds. [9]

Secondly, the authors conclude that for identifying the formants, bandwidth thresholding (100 Hz for single vowels and 50 Hz for concurrent vowels) is necessary. This however is ineffective for higher formants (F3, F4 etc.) because their bandwidth is larger than that of F1 and F2. To account for this problem, dynamic thresholding can be used.

The autocorrelation method was used for finding out the predictor polynomial coefficients. These coefficients can be calculated using the Covariance method.


## Acknowledgment

The authors thank Dr. Ananthakrishna Chintanpalli, Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani for providing us an opportunity to undertake this project work. The authors also thank him for imparting the signal processing knowledge necessary for this endeavor and for his guidance and clarifications.

The authors would also like to express their sincere gratitude to Mr. Harshavardhan S, for his support and guidance.


## References


[1] L.Rabiner and R.Schafer, Introduction to Digital Speech Processing. Hanover, Now Publishers, 2007.

[2] Shodhganga.inflibnet.ac.in, 2017. [Online]. Available: http://shodhganga.inflibnet.ac.in/bitstream/10603/50843/9/09_chapter%202.pdf [Accessed: 15- Nov- 2017]

[3] "Speech Spectrum Shaped Noise - File Exchange - MATLAB Central", In.mathworks.com, 2017. [Online]. Available: https://in.mathworks.com/matlabcentral/fileexchange/55701-speech-spectrum-shaped-noise [Accessed: 17- Nov- 2017]

[4] R. Snell and F. Milinazzo, "Formant location from LPC analysis data", IEEE Transactions on Speech and Audio Processing, vol. 1, no. 2, pp. 129-134, 1993.

[5] "NOIZEUS: Noisy speech corpus - Univ. Texas-Dallas", Ecs.utdallas.edu, 2017. [Online]. Available: http://ecs.utdallas.edu/loizou/speech/noizeus/ [Accessed: 17- Nov- 2017].

[6] T. Kapoor, K. Khandelwal and A. Sachan, "khandelwalkshitij/LPC_Formant_Analysis", GitHub, 2017. [Online]. Available: https://github.com/khandelwalkshitij/LPC_Formant_Analysis

[7] Catford, J.C. (1988) A Practical Introduction to Phonetics, Oxford University Press, p. 161. ISBN 978-0198242178

[8] Assmann, P. F. (1996). Modeling the perception of concurrent vowels: Role of formant transitions. The Journal of the Acoustical Society of America, 100(2, Pt 1), 1141-1152. [Online] http://dx.doi.org/10.1121/1.416299

[9] Cheveigné, A. (1997). Concurrent vowel identification. III. A neural model of harmonic interference cancellation. The Journal of the Acoustical Society of America, 101, 2857. [Online] https://doi.org/10.1121/1.419480

APPENDIX:

## A. Code for plotting the Graphs:

```
close all;
clear all;

%SSN Generation
%*********************
[noise
Fs]=audioread('SpeechSpectrumShapedNoise.wav');
noise=noise(4000:7999);
%*********************
[x11 Fs]=audioread('ah_ae_0.wav');
[x21 Fs]=audioread('ah_ae_4.wav');
[x31 Fs]=audioread('ee_ah_0.wav');
[x41 Fs]=audioread('ee_ah_4.wav');
[x51 Fs]=audioread('ee_oo_0.wav');
[x12 Fs]=audioread('ee_oo_4.wav');
audio=[x11, x21, x31, x41, x51, x12];

for j=1:6
    a=audio(:,j);
    x=decimate(a,10);
    audio_uncorrupted(:,j)=x;
    SNR=5;
    rms_noise=rms(x)*10^(-0.05*SNR);
    noise=(rms_noise/rms(noise)).*noise;
    x_corrupted=x+noise;
    audio_corrupted(:,j)=x_corrupted;
end
    Fs=Fs*0.1;
     figure;
     hold on;
     t=1/Fs:1/Fs:length(x)/Fs;
     plot(t,x);
     plot(t,x_corrupted);
     xlabel('Time(s)');
     xlim([0.1 0.13])
     ylabel('Magnitude');
     title('Time Domain Signal');
     legend('Uncorrupted Signal','Signal at SNR =
5dB')
     f=linspace(0,Fs,length(x));
 for j=1:6;
     y=fft(audio_uncorrupted(:,j));
     figure;
     semilogy(f,(abs(y)));
     xlabel('Frequency (Hz)');
     ylabel('Magnitude (dB)');
     hold;
     a=lpc(audio_uncorrupted(:,j),31);
     b=1;
     [h,w]=freqz(b,a,100);
     w=w*Fs/(2*pi); % w is assigned as formant
frequency
     semilogy(w,(abs(h)));
      xlim([0 4000]);
      legend('Signal Spectrum','LPC Analysis')
      title('Uncorrupted Signal');
 end
for j=1:6;
    y=fft(audio_corrupted(:,j));
    figure;
    semilogy(f,(abs(y)));
    xlabel('Frequency (Hz)');
    ylabel('Magnitude (dB)');
    hold;
    a=lpc(audio_corrupted(:,j),31);
    b=1;
    [h,w]=freqz(b,a,100);
    w=w*Fs/(2*pi); % w is assigned as formant
frequency
    semilogy(w,(abs(h)));
    xlim([0 4000]);
    legend('Signal Spectrum','LPC Analysis')
    title('Corrupted Signal at 5 dB');
end
```

## B. Code for Generating the Roots of LPC Prediction Polynomial

```
close all;
clear all;

%SSN Generation
%*********************


[noise
Fs]=audioread('SpeechSpectrumShapedNoise.wav');
noise=noise(4000:7999);
%*********************
[x11 fs]=audioread('ah_100.wav');
[x21 fs]=audioread('ae_100.wav');
[x31 fs]=audioread('ee_100.wav');
[x41 fs]=audioread('er_100.wav');
[x51 fs]=audioread('oo_100.wav');
[x12 fs]=audioread('ah_126.wav');
[x22 fs]=audioread('ae_126.wav');
[x32 fs]=audioread('ee_126.wav');
[x42 fs]=audioread('er_126.wav');
[x52 fs]=audioread('oo_126.wav');
disp(fs);
audio=[x11, x21, x31, x41, x51, x12, x22, x32, x42,
x52];
audio_corrupted=zeros(4000,10);
audio_uncorrupted=zeros(4000,10);
for j=1:10
    a=audio(:,j);
    x=decimate(a,10);
    audio_uncorrupted(:,j)=x;
    SNR=5;
    rms_noise=rms(x)*10^(-0.05*SNR);
    noise=(rms_noise/rms(noise)).*noise;
    x_corrupted=x+noise;
    audio_corrupted(:,j)=x_corrupted;
end
    f=linspace(0,Fs,length(x));
    OUTPUT_lpc=zeros(18,9);
    for j=1:10;
    disp(j);
    a=lpc(audio_uncorrupted(:,j),15);
    b=1;
    [h,w]=freqz(b,a,100);
    w=w*Fs/(2*pi); % w is assigned as formant
frequency
    rts=roots(a);
    rts = rts(imag(rts)>=0);
    angz = atan2(imag(rts),real(rts));
    [frqs,indices] = sort(angz.*(Fs/(2*pi)));
    bw = -1/2*(Fs/(2*pi))*log(abs(rts(indices)));
    nn = 1;
    OUTPUT_lpc((j-1)*3+1,1:length(frqs))=frqs;
    OUTPUT_lpc((j-1)*3+2,1:length(bw))=bw;
    OUTPUT_lpc((j-1)*3+3,1:length(rts))=abs(rts);
    end
```