(ii) Explain the following terms:

a. LP residual

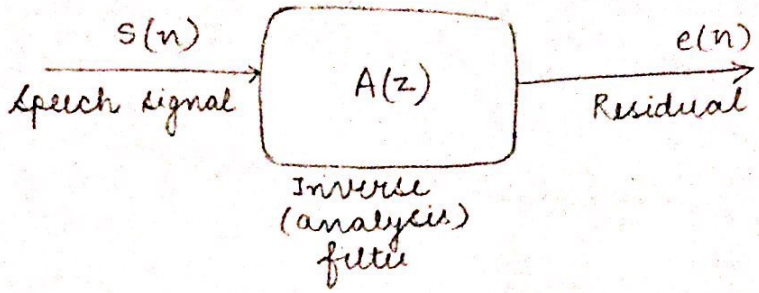
The LP residual is the error between the speech signal and its predicted value and is given as —

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^{p} a_k s(n-k)$$

where, $\hat{s}(n)$ is the predicted value of $s(n)$

The LP residual represents the excitations for production of speech. The residual is

typically a series of pulses, when derived from voiced speech or noise-like, when derived from unvoiced speech.



(b) Properties of Toeplitz matrix

1. A matrix ~~which~~ in which each descending diagonal from left to right is constant. The $n \times n$ matrix A of the form

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \cdots & \vdots \\ a_2 & a_1 & \ddots & \ddots & a_{-2} \\ \vdots & \ddots & \ddots & a_1 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_0 \end{bmatrix}$$

is a Toeplitz matrix. If the i, j element of A is denoted A_{ij} then we have

$$A_{ij} = A_{i+1\, j+1} = a_{i-j}$$

2. A Toeplitz matrix is not necessarily square
3. Toeplitz matrices are persymmetric. Symmetric Toeplitz matrices are both centrosymmetric and bicymmetric.
4. Toeplitz matrices commute asymptotically This means they diagonalise in the same basis when the row and column dimension tends to infinity.

5. Two toeplitz matrices may be added in $O(n)$ time and multiplied in $O(n^2)$ time.

(c) **Prediction Error:**

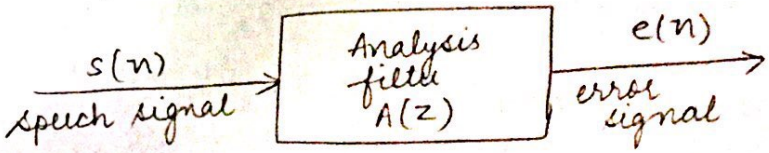
The prediction error, $e(n)$, can be viewed as the output of the prediction ~~fitter~~ error filter $A(z)$, where $H(z)$ is the optimal linear predictor

$$A(z) = \frac{1}{H(z)} \quad \text{and}$$

$$e(n) = s(n) + \sum_{k=1}^{p} a_k s(n-k) \longrightarrow (1)$$

i.e. It is the difference between the input speech and the estimated speech.

In eq. (1), the gain G is ignored to allow the parameterisation to be independent of the signal intensity.



(d) **Pole - Formant Relationship:**

Pole angles in the LP spectrum have information about the formant values of the speech signal, The pole formant relationship can be analysed by plotting the LP spectrum with pole angles on the frequency axis. The poles indicate the permanent peaks. In general, one can obtain one formant value using 2 poles.

(e) Normalised error: Normalised error helps in determining the optimal number of parameters to be used in the model spectrum. It is the ratio of minimum error to the energy in the signal R(0). It is equal to the normalised frequency of the model spectrum

$$V_i = \frac{E_i}{R(0)} = 1 + \sum_{k=1}^{i} a_k r(k)$$

for $i \geq 0$ $\quad 0 \leq V_i \leq 1$

The final normalised error V_p is —

$$V_p = \prod_{i=1}^{p} (1 - k_i^2)$$

The intermediate quantities k_i, $1 \leq i \leq p$ are known as reflection coefficients.

Question 2

2. LP spectrum is the graph plotted with the pole angles of LP analysis located on the normalised frequency axis (x-axis) and the amplitude/magnitude (dB) on the y-axis.

LP spectrum is defined as $\frac{1}{1 - A(z)}$

where

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \quad (G=1)$$

where a_k is the k^{th} LP coefficient and p is the order of the LPC predictor.

Here, the pole values are one of the roots of $\{1-A(z)\}$. This is what is plotted on the graph.

1. Solve $1-A(z)$ to obtain poles. Here $1-A(z)$ becomes a companion matrix, used to find eigen values (roots of $1-A(z)$). Use QR algorithm for this.
2. Draw the graph plotting the LP spectrum.
3. Estimate formants and nulls. — arrange positive angles in ascending order (omitting negative ones since they are conjugate pairs, symmetric) Compute the magnitude response using $H(w)$

$$= \prod_{i=1}^{4} \sqrt{1+r_i^2 - 2r_i \cos\phi}$$

for given angle 'w' — $\phi = \theta - w$. Compute forward and backward slopes of neighbouring angles (m_1, m_2) as

$m_1 = H(\theta_i + \delta w) - H(\theta_i)$
$m_2 = H(\theta_{i+1}) - H(\theta_{i+1} - \delta w)$. If $m_1 < 0$ & $m_2 > 0$ a null is assumed b/w the 2 angles and these two poles are treated as independent formants. Else, magnitudes are computed — if $|H(\theta_i) - H(\theta_{i+1})| < 3dB$ both poles are formants.

4. Finally, estimated formants locations and no. of poles for each formant are used to compute the bandwidths of the formants and finally the frequency response of the desired post-filter

3. There are two widely used methods for estimating the LP coefficients:

(i) Autocorrelation

(ii) Covariance.

Both methods choose the short term filter coefficients a_k in such a way that the energy in the error signal (residual) is minimised. For speech processing tasks, the autocorrelation method is almost exclusively used because of its computational efficiency and inherent stability whereas the covariance method does not guarantee the stability of the all-pole LP synthesis filter. The auto-correlation method of computing LP coefficients is as follows:

First, speech signal $s(n)$ is multiplied by a window $w(n)$ to get the windowed speech segment $s_w(n)$. Normally, a Hamming or Hanning window is used. The windowed speech signal is expressed as

$$s_w(n) = s(n)\, w(n)$$

The next step is to minimise the energy in the residual signal. The residual energy E_p is defined as

$$E_p = \sum_{-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} \left(s_w(n) + \sum_{k=1}^{p} a_k s_w(n-k) \right)^2$$

The values of a_k that minimise E_p are found by setting the partial derivatives of the energy E_p with respect to the LP coefficient parameters

equal to zero.

$$\frac{\partial E_p}{\partial a_k} = 0 \quad , \qquad 1 \le k \le p$$

This results in the following 'p' linear equations for the 'p' unknown parameters $a_1, \ldots, a_p$

$$\sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} s_w(n-i)\, s_w(n-k)$$

$$= \sum_{n-\infty}^{\infty} s_w(n-i)\, s_w(n) \quad , \quad 1 \le i \le p \qquad \text{—— (1)}$$

This linear equation can be expressed in terms of the autocorrelation function. This is because the autocorrelation function of the windowed segment $s_w(n)$ is defined as

$$R_s(i) = \sum_{n=-\infty}^{\infty} s_w(n)\, s_w(n+i) \quad , \quad 1 \le i \le p \qquad \text{—— (2)}$$

Exploiting the fact that the autocorrelation function is an even function i.e. $R_s(i) = R_s(-i)$.

By substituting the values from (2) in (1), we get —

$$\sum_{k=1}^{p} R_s(|i-k|)\, a_k = -R_s(i) \qquad 1 \le i \le p$$

These set of 'p' linear equations can be represented in the following matrix form as

$$\begin{bmatrix} R_s(0) & R_s(1) & \cdots & R_s(p-1) \\ R_s(1) & R_s(0) & & R_s(p-2) \\ \vdots & \vdots & & \vdots \\ R_s(p-1) & R_s(p-2) & & R_s(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_s(1) \\ R_s(2) \\ \vdots \\ R_s(p) \end{bmatrix}$$

This can be summarised using vector-matrix notation as —

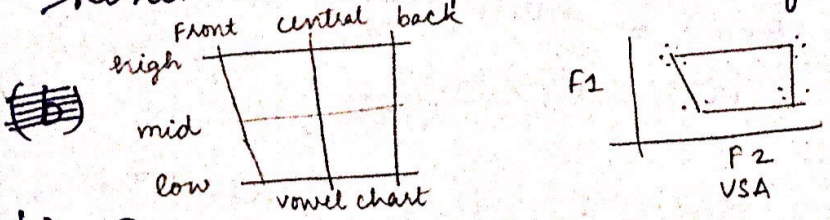
$$R_s \, a = -r_s$$

where the pxp matrix R_s is known as the autocorrelation matrix. The resulting matrix is a Toeplitz matrix where all elements along a given diagonal are equal. This allows the linear equations to be solved by the ~~Levis~~ Levinson-Durbin algorithm. Because of the Toeplitz structure of R_s, A(z) is minimum phase. At the synthesis filter $H(z) = \frac{1}{A(z)}$, the zeros of A(z) become the poles of H(z). Thus the minimum phase of A(z) guarantees the stability of H(z).

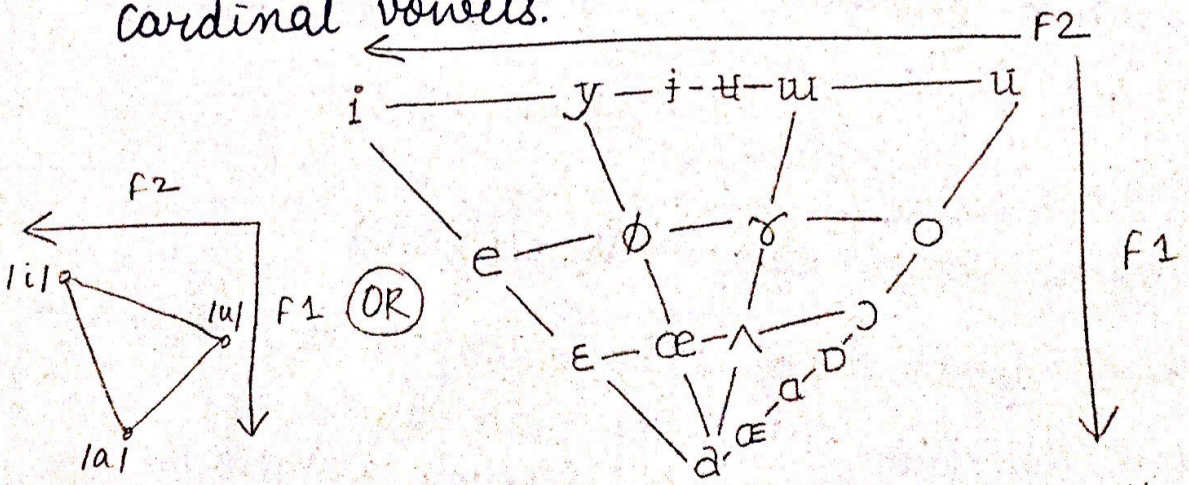
4. (a) Define Vowel Space Area (VSA)

Vowel Space Area (VSA) refers to the two-dimensional area bounded by lines connecting first and second formant frequency coordinates (F1/F2) of vowels. The vowel space illustration provides a graphical method of showing where a speech sound, such as vowel, is located in both "acoustic" and "articulatory" space. A typical computation involves making static measurements of the F1/F2 values for each of the four corner vowels (or three -point vowels (for triangle) /a, i, u/) at 50%.

vowel duration, for several productions of each vowel. The mean F1/F2 value for each of the four corner vowels is then used to compute the area of the quadrilateral (or △) formed by the corner vowels. Since frequencies of the first and second formants roughly relate to the size and shape of the cavities created by jaw opening (F1) and tongue position (F2), the VSA is an acoustic proxy for the kinematic displacements of the articulators.



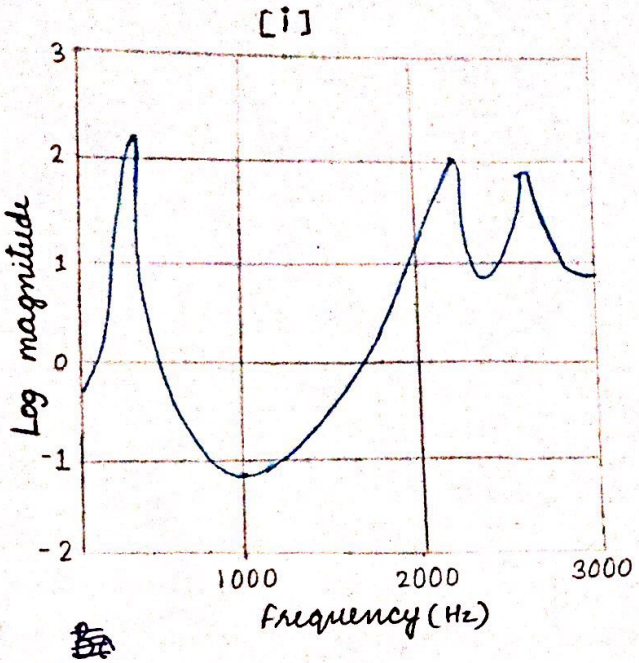
(b) Draw a vowel triangle by considering the cardinal vowels.

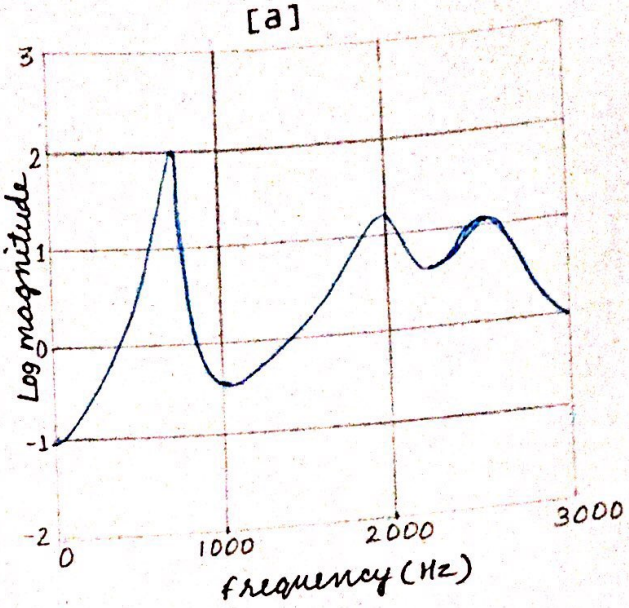


Vowel triangle considering 18 cardinal vowels

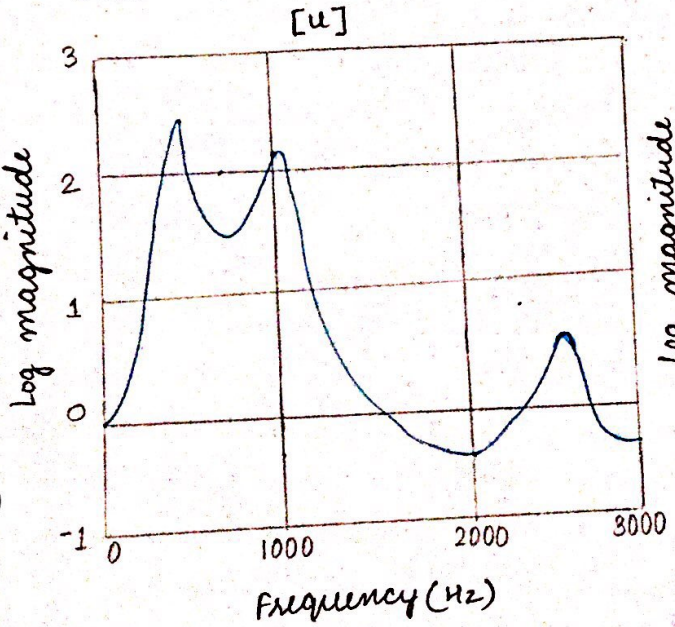
(c) Draw the LP spectrum for any two front and back vowels and explain how the vowels are characterised according to their formants.

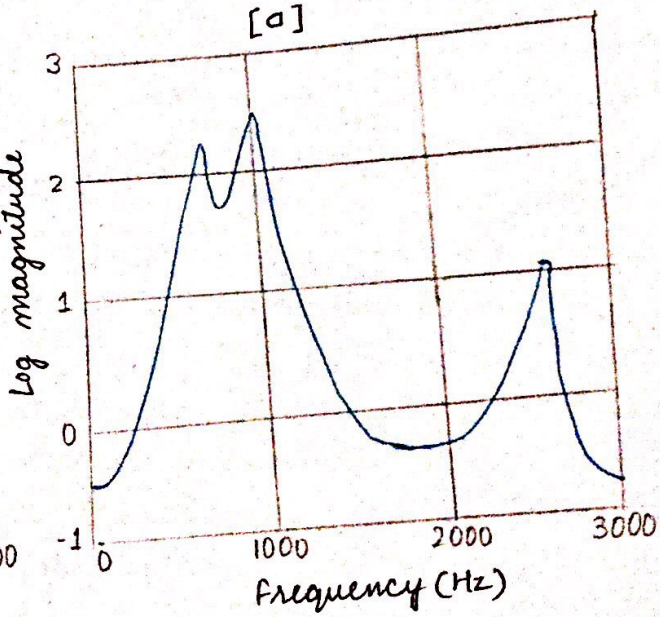
Front Vowels :





Back Vowels :





The frequency of the first formant is mostly charachterised by the height of the tongue body.

High F_1 = low vowel and Low F_1 = high vowel

The frequency of the second formant is mostly charachterised by the frontness / backness of tongue.

High F_2 = front vowel and Low F_2 = back vowel