8.1 Introduction:

The goal of this chapter is to create a non-linear model for a sound signal. We have a time series ($X$) of the vowel "a". It is a male voice. 44000 Hz is the Sampling
frequency of this signal. Hence the sampling interval \((h)\) of this signal is \(1/44000\) second. 264688 is the length of the signal. First we do a Fourier transform of the data series \(X\). We find that the first dominating peak is at the bin no.764. Therefore the basic frequency of this voice is 127 Hz. and the corresponding angular frequency \(\omega\) is 797.96 Hz. So we take 797.96 Hz as the initial guess for \(\omega\). We divide the given data series into overlapping sections. In each section the parameters of our model change. Therefore in each section we come out with different equations of the system which predict the data locally.

\[
\text{Figure 8.2: Fourier transform of the signal}
\]

\[8.2 \text{ Equation of the Model:}
\]

We propose a non-linear model for human voice. We take a section of the given data set which consists of 2000 points. We create a data vector out of these data points \((x_0, x_1, \ldots, x_{1999})\) and denote it by \(S\). Let us consider the following equation

\[
S_n = A_0 + \sum_{m=0}^{M} [A_m (Z1^n Z2^{n^2} Z3^{n^3})^m + \overline{A}_m (\overline{Z1}^n \overline{Z2}^{n^2} \overline{Z3}^{n^3})^m] \quad \ldots (1)
\]

where

\[Z1 = e^{i\omega h}, Z2 = e^{ih^2\alpha/2}, Z3 = e^{ih^3\beta/6}\]. Here \(A, \overline{Z1}, \overline{Z2}\) and \(\overline{Z3}\) are the complex conjugates of \(A, Z1, Z2\) and \(Z3\) respectively. Here \(\omega\) is the sampling angular
8.3 Calculating Amplitudes from the initial guesses of Angular Frequency ($\omega$):

Initially we take $\omega = 797.96$ which we have calculated from FFT, $\alpha = 1$ and $\beta = 1$. So we know the initial values of $Z_1$, $Z_2$ and $Z_3$. Now we compute a matrix $P$ which has $N$ rows and $2M$ columns. Here $N = 2000$ and $M = 80$. The 0th column of $P$ consists of the constant vector 1. Any element of $P$ which is from the 1st column to the $(M-1)$th column is of the form $(Z_1^n Z_2^{n^2} Z_3^{n^3})^m$ and any element which is from the $M$th column to the $(2M-1)$th column is of the form $(Z_1^n Z_2^{n^2} Z_3^{n^3})^m$.

We take the generalized inverse (ref. Appendix) of $P$ and compute $GI(P) \times S$. $GI(P) \times S$ gives us the amplitude vector $A$.

8.4 Approximation of Amplitudes and Angular Frequency ($\omega$) by Iteration:

In the previous section we have discussed how we find out the amplitude vector from the initial guesses of $Z_1$, $Z_2$ and $Z_3$. We use this vector in the equation (1) and calculate the prediction, say, Pred. We subtract this Pred from the data vector and find the residue $R$.

Now we try to find a better approximation for the amplitude vector $A$ and $Z_1$, $Z_2$, $Z_3$. We use Taylor series expansion for this approximation. If any function $f(x)$ is differentiable at the point ‘$a$’, then the value of $f$ is given in a small neighbourhood of ‘$a$’ by the Taylor series formula

$$f(a + h) = f(a) + h(df/dx) + h^2/2!(d^2 f/dx^2) + ..... + h^n/d^n x^n + ....$$

Now if $h$ is sufficiently small we can ignore the higher terms of $h$ and rewrite the above formula as
Chapter 8. Prediction of Sound Signal based on a Nonlinear Model

$$f(a + h) = f(a) + h(df/dx)$$

If $f$ is a function of multiple variables, say 3, then its Taylor series expansion will be

$$f(a + h, b + k, c + l) = f(a, b, c) + h(\partial f/\partial x) + k(\partial f/\partial y) + l(\partial f/\partial z)$$

if we ignore the higher terms of $h, k$ and $m$. Similarly we can write the Taylor series expansion of $F(Z1, Z2, Z3)$ as:

$$F(Z1 + h, Z2 + k, Z3 + l) = F(Z1, Z2, Z3) + h(\partial F/\partial Z1) + k(\partial F/\partial Z2) + l(\partial F/\partial Z3)$$

where;

$$\frac{\partial F}{\partial Z1} = \sum_{m=0}^{M} [mn A_m (Z1^{mn-1} Z2^{mn^2} Z3^{mn^3})] + \sum_{m=0}^{M} [mn A_m (Z1^{mn-1} Z2^{mn^2} Z3^{mn^3})]$$

$$\frac{\partial F}{\partial Z2} = \sum_{m=0}^{M} [mn^2 A_m (Z1^{mn} Z2^{mn^2-1} Z3^{mn^3})] + \sum_{m=0}^{M} [mn^2 A_m (Z1^{mn} Z2^{mn^2-1} Z3^{mn^3})]$$

$$\frac{\partial F}{\partial Z3} = \sum_{m=0}^{M} [mn^3 A_m (Z1^{mn} Z2^{mn^2} Z3^{mn^3-1})] + \sum_{m=0}^{M} [mn^3 A_m (Z1^{mn} Z2^{mn^2} Z3^{mn^3-1})]$$

We create an ExtendedP matrix which has $(2M + 6)$ columns. Any nth element in the $2M$th, $(2M + 1)$th, $(2M + 2)$th, $(2M + 3)$th, $(2M + 4)$th and $(2M + 5)$th columns is of the following form respectively

$$\sum_{m=0}^{M} [mn A_m (Z1^{mn-1} Z2^{mn^2} Z3^{mn^3})],$$

$$\sum_{m=0}^{M} [mn A_m (Z1^{mn-1} Z2^{mn^2} Z3^{mn^3})],$$

$$\sum_{m=0}^{M} [mn^2 A_m (Z1^{mn} Z2^{mn^2-1} Z3^{mn^3})],$$

$$\sum_{m=0}^{M} [mn^3 A_m (Z1^{mn} Z2^{mn^2} Z3^{mn^3-1})],$$
8.4. Approximation of Amplitudes and Angular Frequency ($\omega$) by Iteration:

\[
\sum_{m=0}^{M} [mn^2 A_m (Z1^{mn} Z2^{mn^2} Z3^{mn^3})],
\]

\[
\sum_{m=0}^{M} [mn^3 A_m (Z1^{mn} Z2^{mn^2} Z3^{mn^3} - 1)],
\]

\[
\sum_{m=0}^{M} [mn^3 A_m (\overline{Z1}^{mn} \overline{Z2}^{mn^2} \overline{Z3}^{mn^3} - 1)].
\]

We take the generalized inverse of this $Extended P$ and multiply it with the residue vector $R$. Let us denote $(Extended P)^* R$ as $C$. $C$ is a column vector with $(2M + 6)$ elements. We add the first $2M$ elements of $C$ with the corresponding elements of the $A$ vector to get the modified amplitudes. We add $C_{2M}$, $C_{2M+2}$, $C_{2M+4}$ with $Z1$, $Z2$ and $Z3$ respectively to get the modified $Z1$, $Z2$ and $Z3$.

With this new set of $Z1$, $Z2$, $Z3$ and the $A$ vector, we repeat the same process again and again and finally we come to the prediction. Our final prediction is shown in figure 8.3 and figure 8.4. Our final prediction seems to be quite good.

![predicted signal vs actual signal](image)

**Figure 8.3:** predicted and the actual signals

We took 2000 data, but our model predicts 2100 data points. So it does a lot of data compression. Final $Z1$, $Z2$ and $Z3$ are $0.9998366 + 0.018248107i$, $0.999999 + 0.0000001i$ and $i$ respectively. From the value of $Z1$ we find that our modified $\omega$ is $802.959 - 0.137i$. 
8.5 Manifold Structure of the Signal:

We break down a human voice signal into overlapping sets. Each of these sets can be seen as an open subset of the whole manifold. We would show that this manifold is locally 2 dimensional and on each of these charts we would develop a dynamics. We assume that within each subsection $\omega$, $\alpha$ and $\beta$ do not change. We can show this by specific example. We chose a subsection with a length of 200000 points, when "a" was pronounced. By using that we determined the coefficients for the instantaneous frequency by using iteration. Once these are determined, within that chart, we know the expression for local frequency. Now we embedd the data in 1000 dimensions by creating Takens vectors in $\mathbb{R}^{1000}$. We want to project that into 2D. By using the same technique described in section 8.3 we find an amplitude vector $A$ from the local frequencies and a data set of 1000 points. Hence for each Takens vector we find a corresponding amplitude vector $A = (A_1, A_2, ..., A_M)$. By a simple calculation one can show that for any $m$, $A_m = C_m/A_1^m$ and each $C_m$ is constant for a single chart. Hence equation (1) can be written only in terms of $A_1$ and from any local $A_1$ and $Z1, Z2, Z3$ one can transform back to the corresponding Takens vector which consists of 1000 data points. We have a collection of local $A_1$. By plotting the real and the imaginary part of the local $A_1$ we find that it is S1 (nearly a circle). Now we try to see the dynamics in terms of the local $A_1$. If we plot the real part of local $A_1$ against time it is nearly a straight line. If we plot the imaginary part...
of local $A_1$ against time it is also a straight line with a shift in the middle. Hence we get a simple dynamics in each chart of the manifold. When the overlapping area comes we continue in the same way as described above and we get simple 2 dimensional dynamics for that chart.

8.6 Conclusion:

In this chapter we have proposed a nonlinear model for human voice which can predict data locally in a small section. As we take different sections, the parameters $\omega$, $\alpha$ and $\beta$ change. The main significance of this model is that it can do data compression. If we take a section of 2000 data points, our model can predict 2100 data points. Another important aspect of this model is that we can predict a better approximation of $\omega$ in each section. This model can be used for voice modulation and voice recognition.
Figure 8.6: plotting real and imaginary part of local A1

Figure 8.7: left: real part of A1 against time, right: imaginary part of A1 against time