[Tyler Olivieri]

ECE 3512: Stochastic Processing in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 1912

# Problem Statement

Correlation measures how similar two sets of data are. It can show if the two data sets are independent or dependent. Given an audio signal, 240 samples will be extracted at .9sec and 3 sec (vector x). This set of 240 samples will shifted by k samples (vector y). The correlation coefficient will be taken between vector x and vector y for k ranging from 0 to 512. The covariance is how much the two random variables change together. It is similar to correlation. Then a similar operation will be performed with 240 samples again at the same periods of time. This time, both vectors will be shifting instead of keeping a vector static. This will become a matrix of covariance.

# Approach and Results

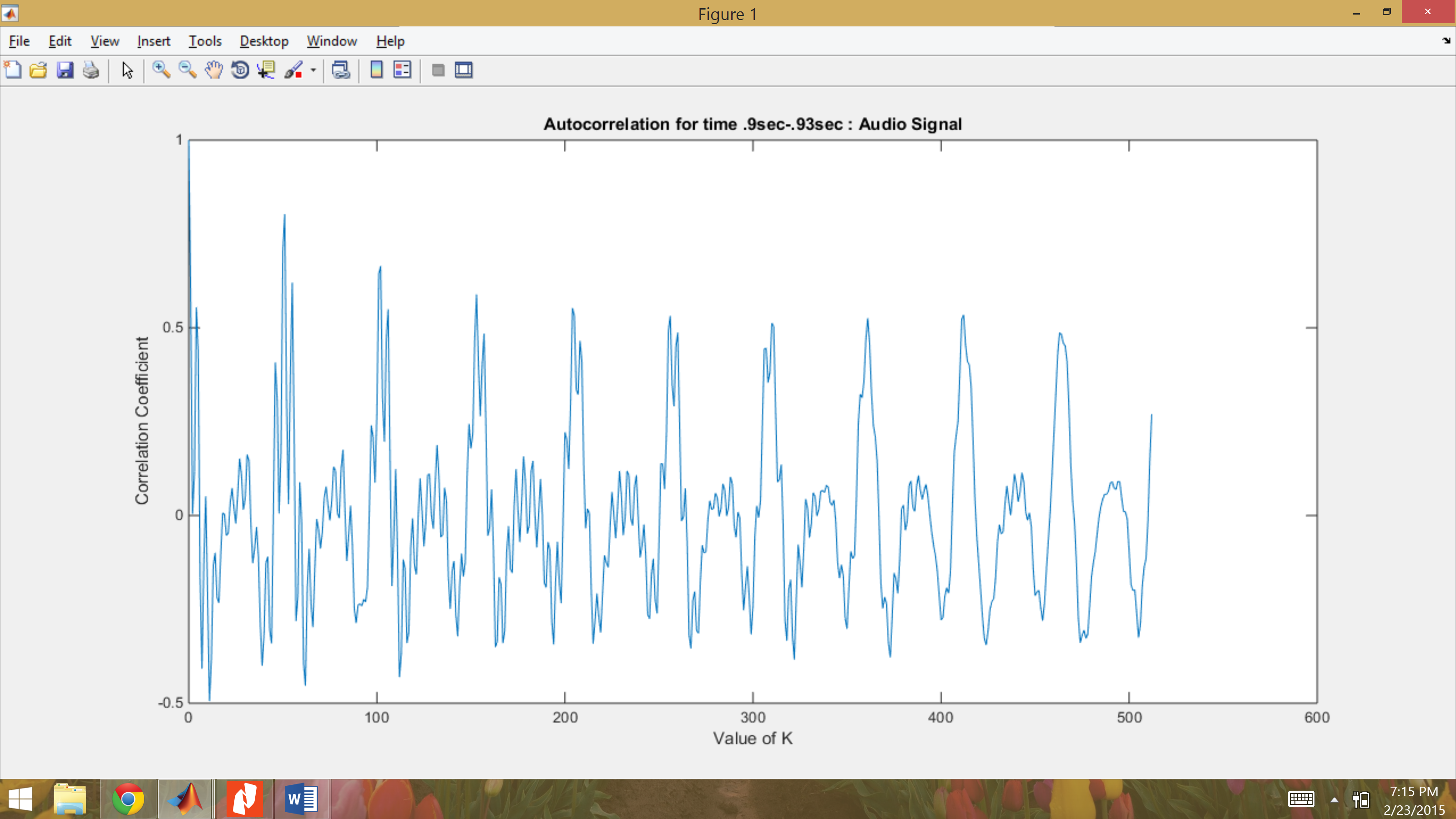


Figure : Autocorrelation for time .9sec-.93sec

The plot of correlation coefficient against value of k(how many samples you are shifting) can be seen in Figure 1 and Figure 2. The only difference is what part of the signal we start at. The correlation coefficient is in-between -1 and 1 which is good. The correlation is also semi-periodic which is a good sign because the audio signal is sums of sine waves so it is also semi-periodic. What I mean by semi-periodic is that the signal is not perfectly periodic with the same values but will be sums of sine waves at different frequencies and it will oscillate negative peak to positive peak etc. It is the nature of an audio signal. So the correlation will approach 0 as the vector shifts but then the correlation will increase again as the vector reaches the next part of the signal that it is similar to, thus raising the correlation coefficient. The correlation coefficient will be negative will the vector is shifted about 180 degrees out of phase the correlation coefficient will be negative. About the same value but negative. When they are “0” degrees in phase they we expect the correlation coefficient to be high. When the signal is not shifted, the correlation coefficient should be 1, which does occur.

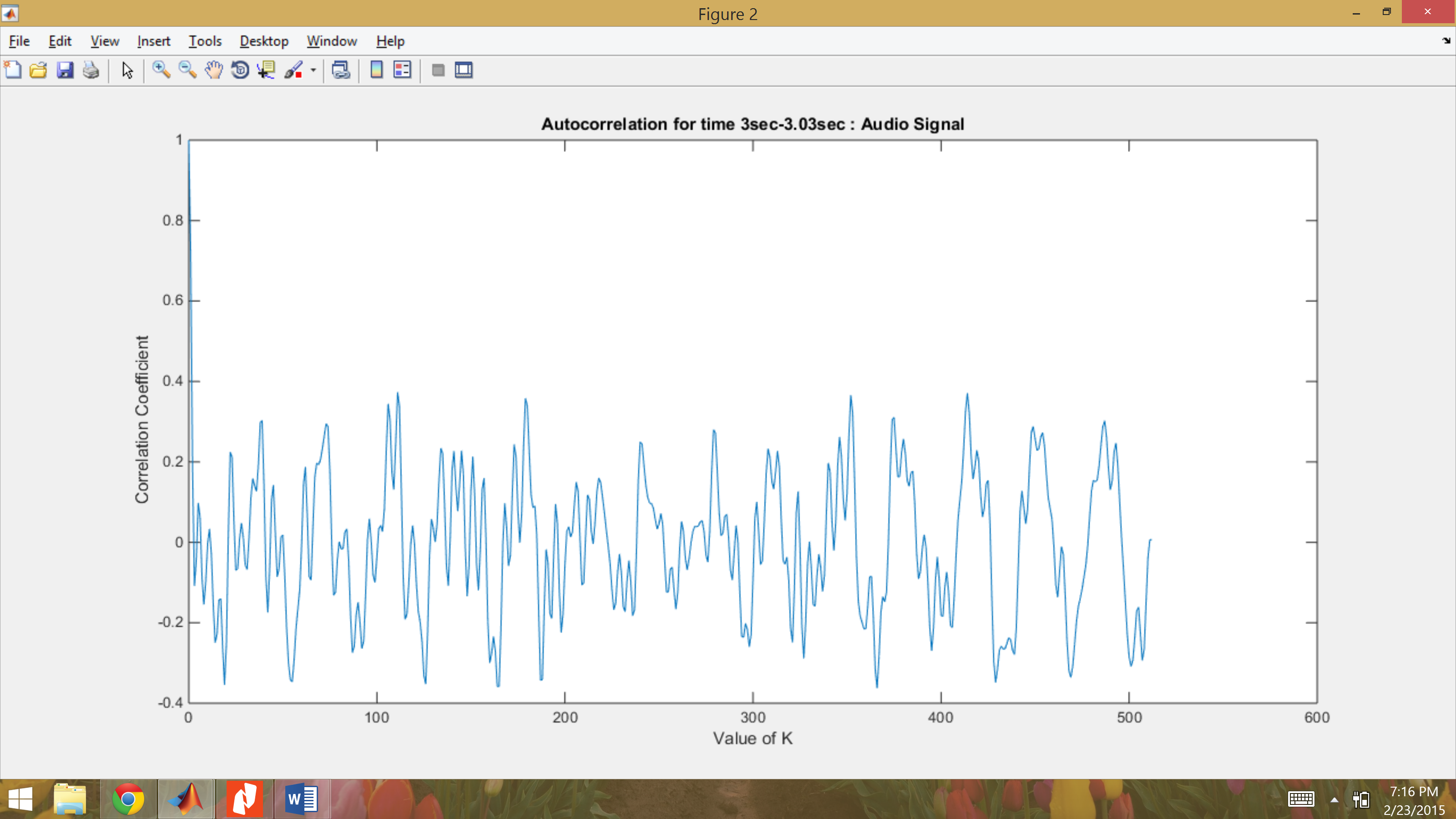


Figure : Autocorrelation for time 3-3.03 sec

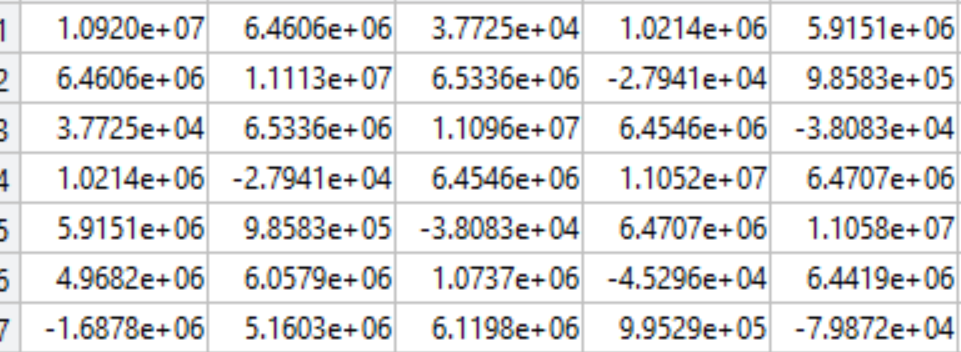


Figure : Covariance matrix( small portion) starting at .9sec

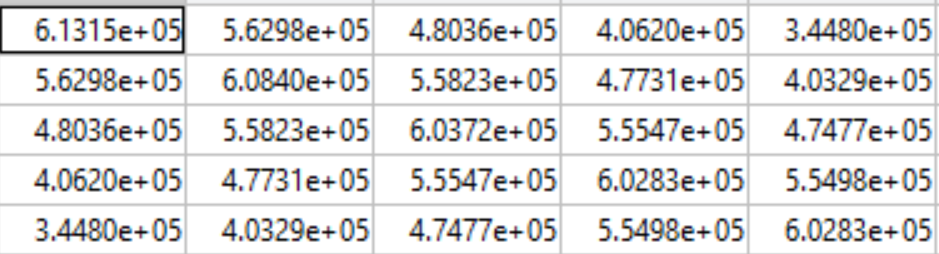


Figure : Covariance matrix (small portion) at 1.1sec

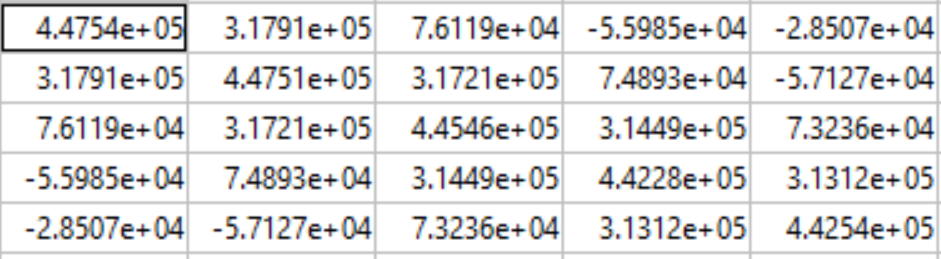


Figure : Covariance matrix (small portion) starting at 3 sec

A snippet of the covariance matrix calculated can be seen in Figure 3, Figure 4, and Figure 5. The only difference in calculation is again what time the calculation starts at in the signal. The covariance for the main diagonal of this matrix, when i = j, is the highest covariance seen. This is expected because when I = j, the two vectors are shifted by the same amount of samples so the correlation is essentially calculated for the two exact same vectors. This also confirms why all the covariances are the same along this diagonal, the same vectors are used for each of these calculations. Any variance in the signal from this original section will decrease the covariance.

# MATLAB Code

In this code, the audio signal is read into MATLAB with the sampling frequency noted. The first sample is found at .9sec. . Then the end sample was found by adding 239. Now, a nested for loop is used to go through all k values and for each k it computes the covariance between vector x and vector y. Vector x is the original section of the signal. Vector y is computed by starting at the starting sample and adding k + an index that increments for every point in the section(D) and then storing this at an index D. The vector y will be a shifted vector of x with the same amount of samples. Then the covariance is calculated using a matlab function. This matlab function returns an identity matrix with the correlation coefficient on the off diagonal. The off diagonal was stored into a variable indexed at k. Then a plot of k vs the correlation coefficient was created. The process was then repeated for a different section of the signal.

Then the covariance will be dealt with. Again, the first sample will be found for the specified time in the audio signal. More nested for loops were used to calcklulate covariance while indexing with I, J, and n. The covariance matrix was indexed with I and J while the n index helped with the summation for each I and J. Covariances for different sections in the signal were calculated for sections of the signal that start at different times.

clc;clear;clf;  
  
%----read in audio signal  
fp =fopen('rec\_01\_speech.raw');  
audio\_signal = fread(fp,inf, 'int16');  
fclose(fp);  
  
%----find sample at .9sec  
fs = 8000;  
sample9start = fs\*.9;  
   
%----- find ending sample and make vector (30ms) total length  
sample9end = sample9start + 240 -1;  
x = audio\_signal([sample9start :sample9end]);  
  
%----- new vector y shifted by k samples from x  
for k = 1:513  
 index = -1;  
 for D = 1:length(x)  
 y(D) = audio\_signal(sample9start + index + k);  
 index= index +1;  
 end  
 P = corrcoef(x,y);  
 R(k) = P(1,2); %corrcoff has a stupid output(identity)  
end  
  
k\_space = 0:1:512;  
  
figure(1);  
plot(k\_space, R)  
title('Autocorrelation for time .9sec-.93sec : Audio Signal');  
xlabel('Value of K');  
ylabel('Correlation Coefficient');  
  
%-------------time 3 seconds  
sample3start = fs\*3;  
  
%----- find ending sample and make vector (30ms) total length  
sample3end = sample3start + 240 -1;  
x2 = audio\_signal([sample3start :sample3end]);  
  
%----- new vector y shifted by k samples from x  
for k = 1:513  
 index = -1;  
 for D = 1:length(x2)  
 y2(D) = audio\_signal(sample3start + index + k);  
 index= index +1;  
 end  
 Q = corrcoef(x2,y2);  
 S(k) = Q(1,2); %corrcoff has a stupid output(identity)  
end  
  
k\_space = 0:1:512;  
  
figure(2);  
plot(k\_space, S)  
title('Autocorrelation for time 3sec-3.03sec : Audio Signal');  
xlabel('Value of K');  
ylabel('Correlation Coefficient');  
  
%----- Part 2  
  
%----find sample at .9sec  
fs = 8000;  
sample9start2 = fs\*.9;  
  
%----- find ending sample and make vector (30ms) total length  
sample9end2 = sample9start + 240 -1;  
x3 = audio\_signal([sample9start2 :sample9end2]);  
  
%----- Calculating Covariance matrix t = .9sec  
for i = 0:15  
 for j = 0:15  
 for n =0:239  
 E(n+1) = audio\_signal(sample9start2 +n -j)\*audio\_signal(sample9start2 + n - i);  
 end  
 Cov(i+1,j+1)= 1/240\*sum(E);  
 end  
end  
  
%----find sample at 1.1sec  
fs = 8000;  
sample11start2 = fs\*1.1;  
  
%----- find ending sample and make vector (30ms) total length  
sample11end2 = sample11start2 + 240 -1;  
x3 = audio\_signal([sample11start2 :sample11end2]);  
  
%----- Calculating Covariance matrix t = 1.1sec  
for i = 0:15  
 for j = 0:15  
 for n =0:239  
 E2(n+1) = audio\_signal(sample11start2 +n -j)\*audio\_signal(sample11start2 + n - i);  
 end  
 Cov2(i+1,j+1)= 1/240\*sum(E2);  
 end  
end  
  
  
%----find sample at 3sec  
fs = 8000;  
sample3start2 = fs\*3;  
  
%----- find ending sample and make vector (30ms) total length  
sample3end2 = sample3start2 + 240 -1;  
x3 = audio\_signal([sample3start2 :sample3end2]);  
  
%----- Calculating Covariance matrix t = 3 sec  
for i = 0:15  
 for j = 0:15  
 for n =0:239  
 E3(n+1) = audio\_signal(sample3start2 +n -j)\*audio\_signal(sample3start2 + n - i);  
 end  
 Cov3(i+1,j+1)= 1/240\*sum(E3);  
 end  
end

# Conclusions

Covariance and the correlation coefficient are very similar. They tell similar things for two sets of data, how similar they are. The correlation coefficient shows if the two data sets are independent or dependent. If the two data sets are independent they will have a correlation coefficient of 0. As the correlation coefficient increases it shows that the data sets are more dependent. The covariance is more of a quantitative analysis of the correlation. It gives more of a quantitative measurement of how correlated the data sets are not bounded by -1 and 1. The covariance is normalized by the number of total samples is used in the calculation. Covariance and the correlation coefficient are used in many linear regression type calculations.