Digital signal processing of different types of sound TSRT78

Real-Axel
Liu ID: *reaxeRR*
00001224-6785

Imaginär-Axel
Liu ID: *imaxeII*
12345678-123i

***Abstract*—The laboration treated the subject of digital signal processing, it was performed in order to study signals from different types of sound and learn how to manipulate them in various situations. It showed that a parametric model can be used to simulate and imitate sound with the use of simple methods and algorithms.**

***Index Terms*—digital signal processing, TSRT78**

I. Introduction

The laboration consisted of three parts where the first task was to record and analyse a whistle to determine its frequency purity. Thereafter the sound of two vowels were recorded and then recreated using parametric models constructed using the original data. Lastly a sentence was recorded, encoded using a simple algorithm to compress it and then decoded again to analyse the result.

II. Task 1 Whistle

A ten second long whistle was recorded, with a sample frequency of 8000 Hertz, and represented in MATLAB as a vector. The most consistent two seconds of the recording was then chosen for the analysis of its frequency spectrum. To determine the purity of the signal, its energy had to be calculated for the whole spectrum and for the dominant frequencies. A value of the signal purity can the be described as equation 1.

$$\frac{E_{dom.freq}}{E_{tot}} \tag{1}$$

The energy was calculated in both the time domain and the frequency domain. When calculating the energy in the time domain, the resulting purity was 0.936 and in the frequency domain 0.912. With the purity of the signal known it was then compared to the purity resulting from a AR-model of second order. An AR-model of second order is defined as equation 2 and letting $n = 2$, $a1$ and $a2$ are the model parameters.

$$T(q) = \frac{1}{1 + a_1 q^{-1} + \ ... \ + a_n q^{-n}} \tag{2}$$

An estimate of the model parameters was calculated using equation 3 where y is the signal, $\phi(t)$ is equal to equation 4 and $\hat{\theta}$ is equal to equation 5.

$$\hat{\theta} = (\sum_{t=1}^{N} \phi(t)\phi(t)^T)^{-1} (\sum_{t=1}^{N} \phi(t)y(t)) \tag{3}$$

$$\phi(t) = (-y(t-1) \ -y(t-2) \ ... \ -y(t-n))^T \tag{4}$$

$$\hat{\theta} = (a_1 \, a_2 \, ... \, a_n)^T \tag{5}$$

When filtering a periodic signal with the AR model a new signal was obtained which purity was given by calculating the poles distance from the unit circle. The distance was calculated to be 0.0021 and the poles position can be seen in figure 1.

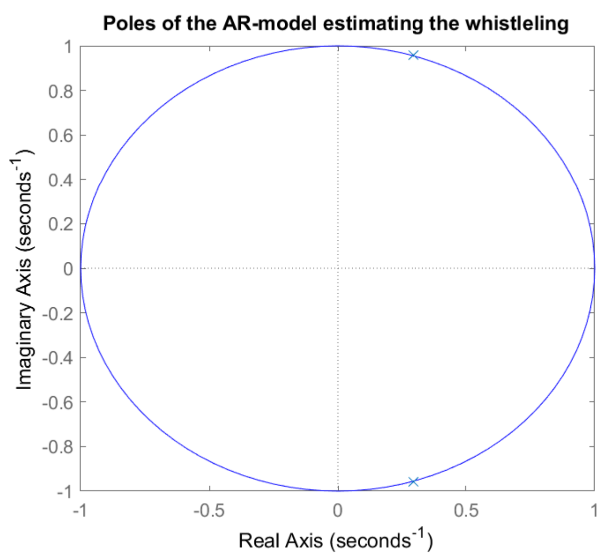

Figure 1. Poles of the AR-model compared to the unit circle.

The reason why an AR-model of second order is suitable in this case is that the whistle should have one peak. This peak can be fairly well modelled with the AR-model consisting of two complex conjugate poles at the dominant frequency. Lastly the spectrum of the whistle was calculated using periodogram and a Fourier transform of the model to determine the dominant frequency which is seen in figure 2.

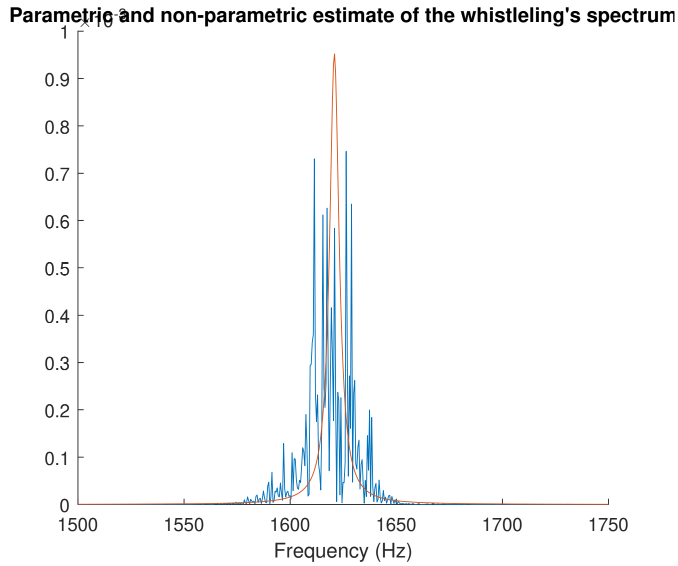

Figure 2. Orange is the spectrum of the filter and blue is the periodogram. Dominant frequency at 1620 Hertz for parametric and between 1610-1630 Hertz for the periodogram.

III. Task 2 Vowels

The objective for this task was to record two vowels for two seconds each and then construct an AR-model of appropriate order for them. The chosen vowels were O and Y recorded with a sampling frequency of 8000 Hertz. To determine the appropriate order for the AR-models, analysis of the loss function and Akaike's information criterion A for different orders were made. The loss function and Akaike's information criterion A is defined by equation 6, where ϵ is the defined by equation 7, and equation 8. By minimising the value of the loss function and Akaike's information criterion A, the error of the model is reduced.

$$V_N(\hat{\theta}_N^{(n)}) = \frac{1}{N}\sum_{t=1}^{N}\epsilon(t,\hat{\theta}_N^{(n)})^2 \tag{6}$$

$$\epsilon[k] = y[k] - \hat{y}[k,\theta] \tag{7}$$

$$U_N(n) = V_N(\hat{\theta}_N^{(n)})(1+\frac{2n}{N}) \tag{8}$$

The result of the calculations is seen in figure 3 and figure 4.

A significant drop is seen at order two and five for O and at order six and seven for Y which was the orders used in the calculations. To determine the performance of the models they were validated by comparing the spectrums and analysing the residuals. From the original signal, a spectrum was calculated using Blackman-Tukey's method which was then compared to the spectrum of the filter. Blackman-Tukey's method is given by equation 9 where T is the sampling time.

$$\hat{\hat{\Phi}} = \frac{1}{NT}|T\sum_{k=0}^{N-1}y[k]e^{-ikT\omega}|^2 \tag{9}$$

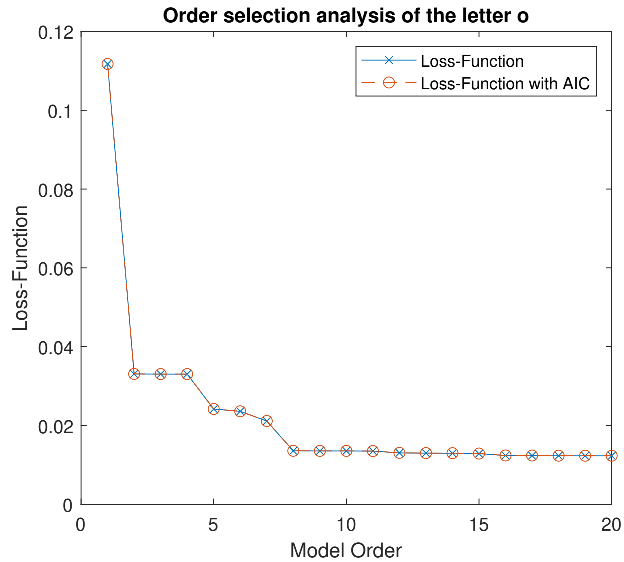

Figure 3. Values of different orders of AR-model for vowel O.

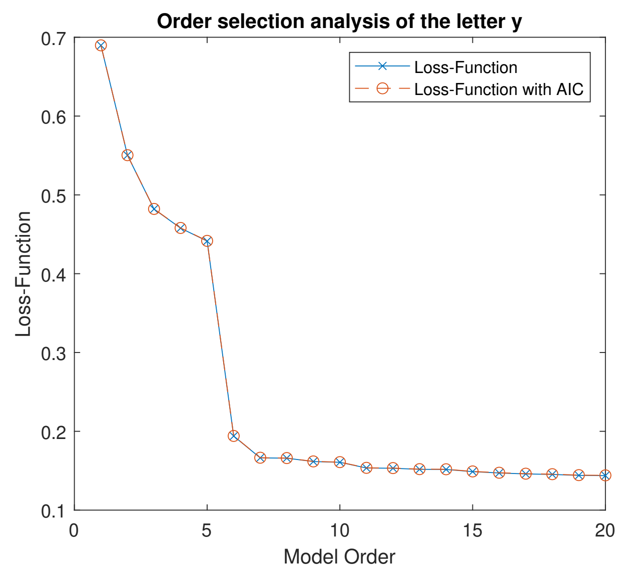

Figure 4. Values of different orders of AR-model for vowel Y.

When comparing the residuals, their covariance should resemble that of white noise's auto correlation i.e. the signals variance in zero delay and zero for every other delay. The residual is defined as the error between the signal and the predictor as seen in equation 7. If this is the case, it means that the error is not correlated with the signal and the model models every part of the system. The auto correlation can be calculated using equation 10.

$$R_{\epsilon\epsilon}[k] = E\{\epsilon[n]\epsilon[n+k]\} \tag{10}$$

Validation of the models showed that they are good approximations and the result of the validation is seen in figure 5 and figure 6 for O and figure 7 and figure 8 for Y.

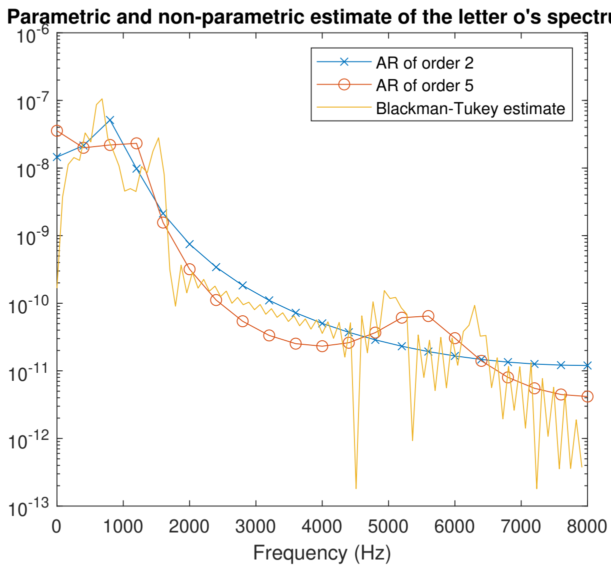

Figure 5. Approximated spectrum of O.

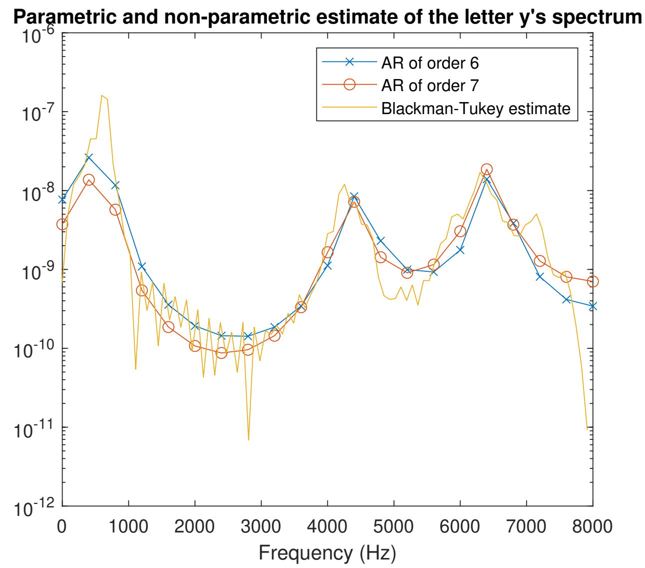

Figure 7. Approximated spectrum of Y.

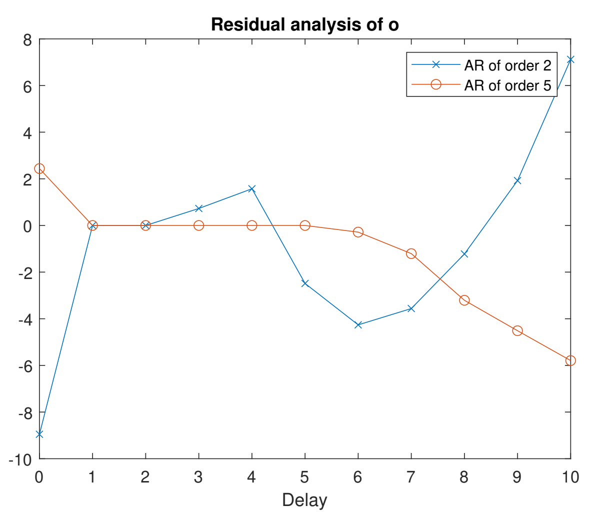

Figure 6. Covariance for the residual.

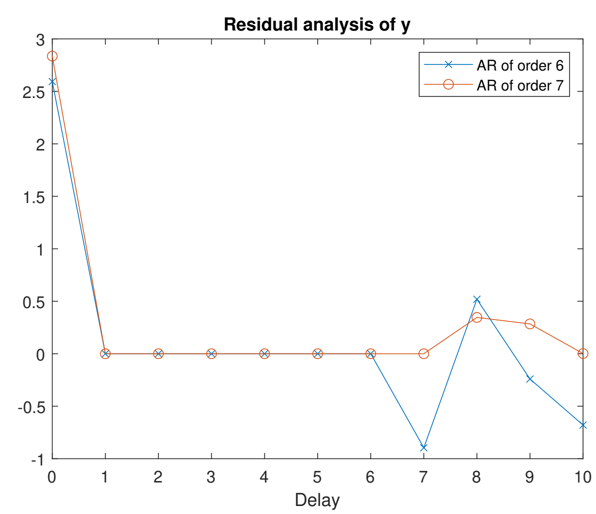

Figure 8. Covariance for the residual.

IV. Task 3 GSM

GSM is a system for wireless communication, it includes a system for encoding sound signals. The third task consists of replicating that encoding on a recorded sentence of two seconds, with sampling frequency 8000 Hertz. The signal is first divided into parts of 160 samples and every such part is estimated using an AR-model of order eight as defined in equation 2 when $n = 8$. The parameters of the AR-model are calculated as equations 3 through 5. To recreate the original sound pulse trains are used as input to the AR-models, where the amplitudes, $\sqrt{A}$, are chosen as the maximum value of the residuals' covariances and the periods D corresponds to the time delays for the same maximums, only values with a delay of 19 samples or more are considered. Equations for residual covariance are found in 7 and 10, A and D in 11 and 12, while the resulting pulse train is displayed in equation 13

$$A = \max_{k>19}(R_{\epsilon\epsilon}[k]) \tag{11}$$

$$D = \arg\max_{k>19}(R_{\epsilon\epsilon}[k]) \tag{12}$$

$$X_{train}[k] = \sqrt{A}\sum_{i=-\infty}^{\infty}\delta(k - iD) \tag{13}$$

Thereafter the different pulse trains are filtered with respective segment and combined back into one long signal, which can be analysed by simply listening to it.

V. Conclusions

Calculations of the harmonic distortion resulted in very similar results in both the time and frequency domain which is to be expected as they theoretically are equivalent. When calculating the harmonic distortion for the AR-model using the distance of the poles from the unit circle it showed a great result. A rather big decrease in harmonic distortion compared to the previous distortion. This meant that the model fits very well and this was confirmed when comparing the spectrum of the model and the original signal. The models spectrum is one peak at the dominant frequency of the signal which is very similar to the estimated spectrum of the signal.

One extra model of higher order was also used for both O and Y. These extra models was chosen because they were the second best choice with regards to model complexity and model error. For O the different model orders differ in some regards. Most obvious is how poorly the second order AR-models residual analysis compares to the fifth order. This shows that an order of two is somewhat low for an accurate model. For Y two very similar model orders were tested in order six and seven. Because of this they also show very similar performance when validating the models. The big difference here is that the sixth order models residuals covariance becomes more fluctuating for higher values compared to the seventh order.

At the end of task two and three, the sound produced from the models were listened to and gave overall good results. The vowels could be recognised as the original ones and GSM gave a really good result where the whole sentence could clearly be heard.

In task three the original sound was sent as a vector containing 16000 values but by using GSM the data that needs to be sent has been dramatically decreased. Now the data that needs to be sent is the 8 parameter values and the values A and D for all 100 segments. This results in 1000 values that need to be sent compared to 16000, a decrease in data sent by 93.75%.

An additional test in the third task was to change the order of the AR-model to analyse the difference in the output. Using lower orders than eight resulted in more noise and at order one even rendering the original sentence inaudible. Higher orders resulted in slightly clearer sound, but not a significant improvement, and when the order rose above 30 some high pitch noise could be heard. When using a constant amplitude of one for the pulse, i.e. $A = 1$, there was a significant increase of noise when the volume of the actual sentence dropped. This is because A varies to strengthen the parts of the signal where the model and input are consistent, which is not the case for the noise.

VI. Appendix

```
%% LABORATION 1 TSRT78
%% task 1
load C:\Users\timaw\OneDrive\Dokument\Skolarbete\TSRT78\vissling.mat
fs = 8000;
% 1
y=y(16000:32000);
energyt = sum(y.*conj(y))

% 2
Ft = fft(y);
energyf = 1/(length(Ft))*sum(Ft.*conj(Ft))


%% 3
xdomfreq = (3170:3270);

energydomfreqf = 2/(length(Ft))*sum(Ft(xdomfreq).*conj(Ft(xdomfreq)));
distt = 1 - energydomfreqf/energyf


[a,b]=butter(8,[3150/8000,3290/8000]);
yfreqdom = filtfilt(a,b,y);
energyfreqdomt = sum(yfreqdom.*conj(yfreqdom));
distf = 1 - energyfreqdomt/energyt
%% 4
[th,P,lam,epsi] = sig2ar(y,2);

figure(1)
hold on
plot(cos(0:pi/50:2*pi), sin(0:pi/50:2*pi),'b');
H = tf([1],[th(2) th(1) 1]);
pzmap(H)
title("Poles of the AR-model estimating the whistleling")
xlabel("Real Axis")
ylabel("Imaginary Axis")
hold off
[p,z] = pzmap(H);

distAR = sqrt(p(1)*p(2)) - 1

%% 5
[Ft, f] = sig2periodogram(y,1/fs);

fnorm = 2*f./length(y);
specpara = abs(1./(1+th(1)*exp(-1i*2*pi*fnorm)+th(2)*exp(-1i*4*pi*fnorm)));

figure(2)
hold on;
plot(f,Ft)
plot(f,((1/fs)*specpara).^2)

title("Parametric and non-parametric estimate of the whistleling's spectrum")
xlabel("Frequency (Hz)")
xlim([1500,1750])
```

```
hold off;

%% task 2
o_best2sec = 0;
y_best2sec = 0;
load C:\Users\timaw\OneDrive\Dokument\Skolarbete\TSRT78\Vowels.mat
o = o_best2sec;
y = y_best2sec;
oest = o_best2sec(1:10000);
yest = y_best2sec(1:10000);
oval = o_best2sec(10001:16001);
yval = y_best2sec(10001:16001);
fs = 8000;
N = 16000;

[oW,oUaic,oUbic]=arorder(o,20);

[yW,yUaic,yUbic]=arorder(y,20);
figure(3)
plot(1:20,oW,'-x',1:20,oUaic,'--o')
title("Order selection analysis of the letter o")
xlabel("Model Order")
ylabel("Loss-Function")
legend("Loss-Function","Loss-Function with AIC")
figure(4)
title("DFT of the vowel: y's spectrum")
plot(1:20,yW,'-x',1:20,yUaic,'--o')
title("Order selection analysis of the letter y")
xlabel("Model Order")
ylabel("Loss-Function")
legend("Loss-Function","Loss-Function with AIC")
%%
[yth6,yP6,ylam6,yepsi6] = sig2ar(yest,6);
[oth2,oP2,olam2,oepsi2] = sig2ar(oest,2);
[yth7,yP7,ylam7,yepsi7] = sig2ar(yest,7);
[oth5,oP5,olam5,oepsi5] = sig2ar(oest,5);
f=0:0.05:1;
yPhi6 = ylam6/fs*abs(freqz(1,[1;yth6],pi*f)).^2;
oPhi2 = olam2/fs*abs(freqz(1,[1;oth2],pi*f)).^2;
yPhi7 = ylam7/fs*abs(freqz(1,[1;yth7],pi*f)).^2;
oPhi5 = olam5/fs*abs(freqz(1,[1;oth5],pi*f)).^2;
[ybtPhi,yf] = sig2blackmantukey(y,30,1/fs);
[obtPhi,of] = sig2blackmantukey(o,30,1/fs);
figure(5)
semilogy(fs*f,yPhi6,'-x',fs*f,yPhi7,'-o',yf,ybtPhi)
title("Parametric and non-parametric estimate of the letter y's spectrum")
xlabel("Frequency (Hz)")
legend("AR of order 6","AR of order 7","Blackman-Tukey estimate")
figure(6)
semilogy(fs*f,oPhi2,'-x',fs*f,oPhi5,'-o',of,obtPhi)
title("Parametric and non-parametric estimate of the letter o's spectrum")
xlabel("Frequency (Hz)")
legend("AR of order 2","AR of order 5","Blackman-Tukey estimate")


[y6Re,k]=sig2crosscovfun(yepsi6,yest);
```

```
y7Re=sig2crosscovfun(yepsi7,yest);
o2Re=sig2crosscovfun(oepsi2,oest);
o5Re=sig2crosscovfun(oepsi5,oest);
y6Re=y6Re/y6Re(1);
y7Re=y7Re/y7Re(1);
o2Re=o2Re/o2Re(1);
o5Re=o5Re/o5Re(1);
figure(7)
plot(k,y6Re,'-x',k,y7Re,'-o')
title("Residual analysis of y")
xlabel("Frequency(Hz)")
legend("AR of order 6","AR of order 7")
xlim([0,10])
figure(8)
plot(k,o2Re,'-x',k,o5Re,'-o')
title("Residual analysis of o")
xlabel("frequency(Hz)")
legend("AR of order 2","AR of order 5")
xlim([0,10])

%%
t = 0 : 1/fs : 2;
d = 0 : 1/105 : 2;
sig = pulstran(t,d,@rectpuls,1/fs);

osound = filtfilt(1,[1;oth5],sig);
soundsc(osound)
%%
ysound = filtfilt(1,[1;yth7],sig);
soundsc(ysound)

%% task 3
load C:\Users\timaw\OneDrive\Dokument\Skolarbete\TSRT78\mustard.mat
fs = 8000;
N = 16000;
out = zeros(16000,1);

K = 160;
k = 1:K;
for J=1:99
    cur_data = mustard(k+J*K);
    detrend(cur_data);
    [th,P,lam,epsi] = sig2ar(cur_data,8);
    a = flipud([1;th]);
    [z,p,c] = tf2zp(1,a');
    for j = 1:length(p)
        if abs(p(j)) > 1
            p(j) = 1/abs(p(j))*exp(1i*angle(p(j)));
        end
    end
    [b,a] = zp2tf(z,p,c);

    e = filter(a,1,cur_data);
    r=covf(e,100);
    [A,D] = max((r(20:100)'));
    A = sqrt(A);
```

```
    D = D + 18;

    d = 1 : D : K;
    ehat = A*pulstran(k,d,@rectpuls,1/fs);

    yhat = filter(1,a,ehat);
    out((k)+J*K)=yhat;
end

soundsc(out)
```