

به نام خدا

دانشگاه خواجه نصیرالدین طوسی
پروژه نهایی پردازش گفتار
دکتر رقیه دوست

محمد جلیلی ترکمانی
شماره دانشجویی : 9523783
نیمسال دوم تحصیلی 98-99

پاسخ سوال 1 :

رابطه بدست آوردن ضرایب کپستروم و فرکانس پیچ :

همانطور که میدانیم، ضرایب کپستروم به این ترتیب بدست می آیند که ابتدا از سیگنال مورد نظر تبدیل DFT گرفته تا به حوزه فرکانس برود. سپس از اندازه مقادیر آن لگاریتم در مبنای ده میگیریم تا بخش های e و h به صورت حاصل جمع از هم جدا شوند. سپس حاصل را تبدیل معکوس میگیریم تا مجدداً به حوزه زمانی (کیوفرنسی) برگردد و ضرایب کپستروم به دست بیاید. در نهایت حاصل مورد نظر را از لیفتر بالاگذر (با مقدار 20) رد میکنیم و اندیس (متغیر iPos در کد) که منجر به ایجاد قله در نمودار آن میشود (نقاط اکسترمم قله ها) را یادداشت میکنیم. سپس فرکانس نمونه برداری را به آن تقسیم میکنیم تا فرکانس پیچ بدست بیاید. از لحاظ ریاضیاتی به شرح زیر است :

$$C[n] = \text{IFFT} \{ \log | \text{FFT} \{ x[n] \} | \}$$

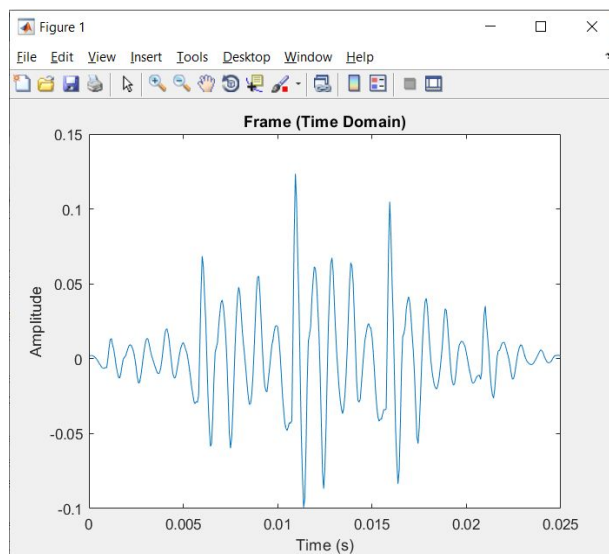
$$C[c] = \text{High pass lifter} (C[n], N = 20)$$

$$\text{Pitch Frequency} = \text{Sampling Frequency} / (\text{First Local maximum extremum index})$$

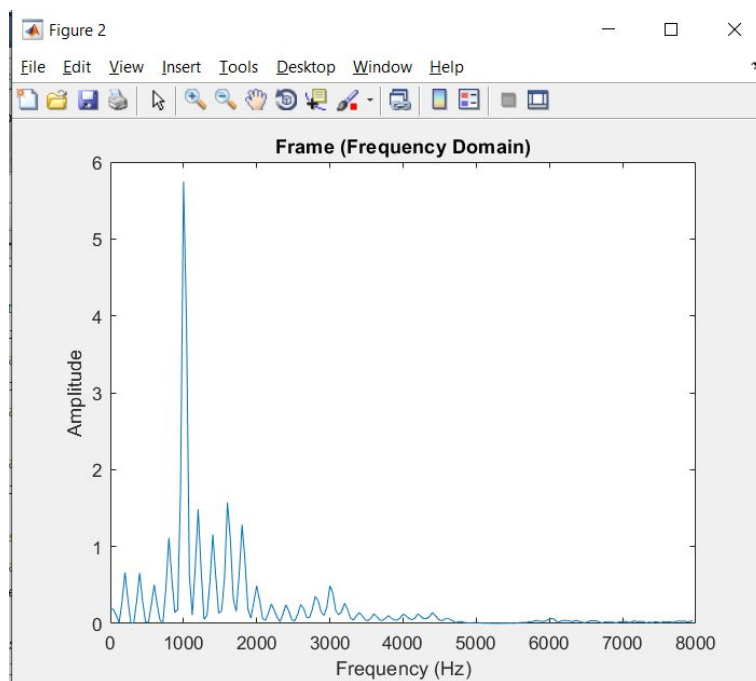
نکات تکمیلی :

1. در کد متلب، شماره فریم حتماً باید از تعداد کل فریم ها کمتر باشد.
2. به جهت عملکرد بهتر، فایل ورودی حتماً شامل فریم واکدار باشد.
3. تصاویر خروجی، برای فریم واج a می باشد.
4. توضیح خط به خط کد در قالب کامنت در کد درج شده است.
5. نموداری که برای مقدار فرکانس پیچ رسم شده، صرفاً مقدار فرکانس پیچ در فریم انتخاب شده را رسم میکند و برای سایر فاقد مقدار (0) است.
6. با توجه به فایل صوتی بارگذاری شده مربوط به یک خانم است و فرکانس پیچ خانم ها بین 150 تا 450 است، مقدار تقریبی 380.95 برای فرکانس پیچ فریم مورد نظر (129) بدست آمده که خود تاییدی بر درستی الگوریتم است.

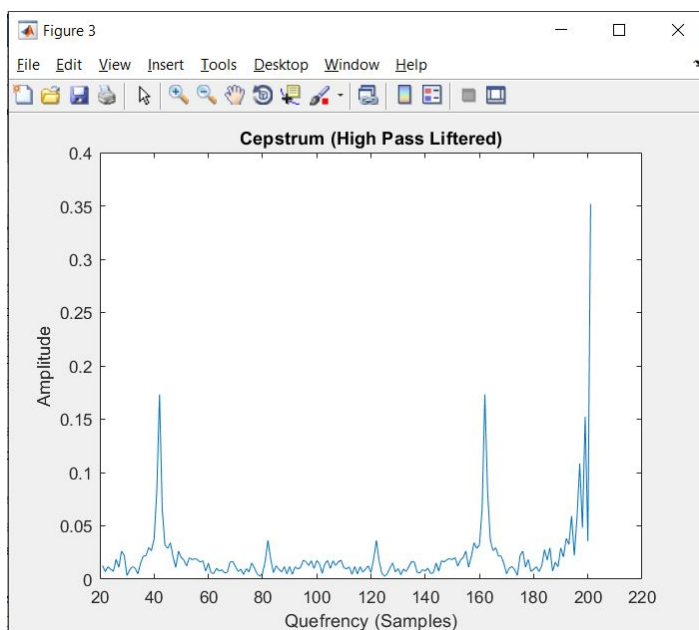
تصاویر خروجی :
● فریم در حوزه زمان :



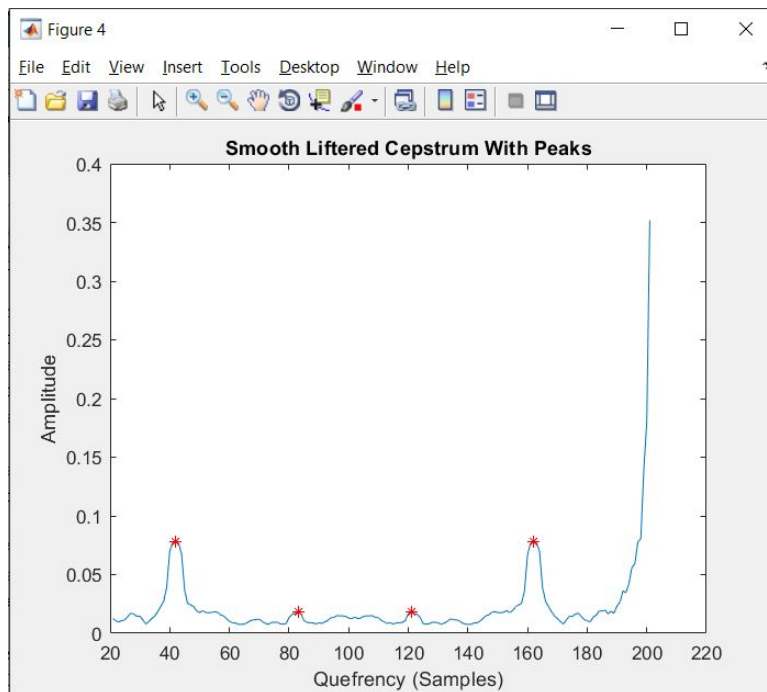
● فریم در حوزه فرکانس :



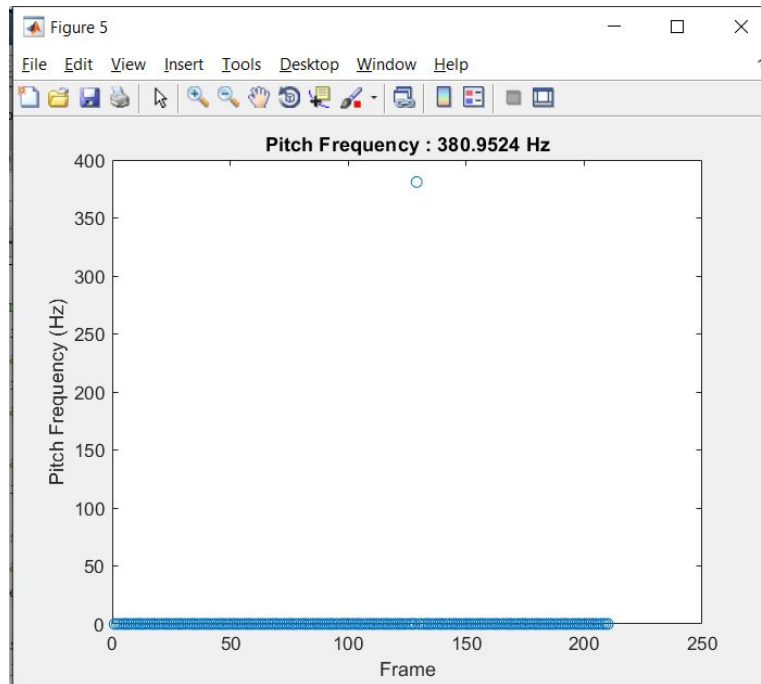
- کپستروم بعد از لیفتر بالاگذر :



- کپستروم بعد از یکنواخت شدن به همراه رسم قله :



- فرکانس پیچ فریم انتخاب شده :



کد مطلب :

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Mohammad Jalili Torkamani %
% Student No : 9523783 %
% Speech Processing %
% Kh.N.Toosi University %
% Dr.R.Doost %
% JUN 2020 %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%clear screen
clc
%clear local variables
clear all
%declare TO-PROCESS frame number (should be less than numberOfSamples later)
FRAME_NUMBER=129
%declare file name
fileName='a.wav'
%read audio and store samples and sampling frequency
[samples,FS] = audioread(fileName);
```

```

%calculate number of samples
numberOfSamples = length(samples);
%display number of Samples
disp(numberOfSamples)
%declare N in high pass liftering
liftering_N=20
%declare desired frame length
fl=400;
%declare frame shifting length
fs=fl*0.4;
%calcualte number of frames with shifting in mind
FN=(numberOfSamples-fl)/fs+1;
FN=round(FN)-1;
%declare an empty array to store pitch frequency(ies)
pitches=zeros(1,FN);
%define hamming windowing
w=hamming(fl);
%for """"JUST THE SELECTED FRAME_NUMBER"""" do processing
for i=FRAME_NUMBER:FRAME_NUMBER
    %do slicing to get i'th frame
    s0=samples((i-1)*fs+1:(i-1)*fs+fl);
    %hamming window declaration
    s=w.*s0;
    %calculate time with dividing by FS (NOTE : time is relative to zero (and starts from 0))
    t = (1/FS)*(1:fl)
    %%%%%%%%%%% plot frame in (time domain) %%%%%%%%%%%
    %prepare for plot
    figure
    %plot signal in time domain
    plot(t,s)
    %label x-axis as "Time (s)"
    xlabel("Time (s)")
    %label y-axis as "Amplitude"
    ylabel("Amplitude")
    %choose title for plot
    title("Frame (Time Domain)")

    %%%%%%%%%%%
    %%%%%%%%%%%

    %move signal to frequency domain and calculate it's absolute value
    s_fft = abs(fft(s))
    %discard half of it (because it's periodic nad repetetive)

```

```
s_fft = s_fft (1:f/2);  
%calculate frequency by multiplying to FS/fl  
f = FS*(0:f/2-1)/fl;
```

```
%%%%%%%%%%%%%%% plot frame (frequency domain)
%%%%%%%%%%%%%%
%prepare for plot
figure
%plot signal in frequency domain
plot(f,s_fft )
%label x-axis as "Frequency (Hz)"
xlabel("Frequency (Hz)")
%label y-axis as "Amplitude"
ylabel("Amplitude")
%choose plot title as "Frequency Domain"
title("Frame (Frequency Domain)")
```

%%%%%%%%%%%%%%
 %%%%%%%%%%

```
%calculate logarithm with base 10 of absolute values of signal
s_fft_log10=log10(s_fft)
%inverse FFT (move signal to time domain) and store absolute values
s_fft_log10_inverse=abs(ifft(s_fft_log10))
%High pass liftering
s_fft_log10_inverse=s_fft_log10_inverse(liftering_N:length(s_fft_log10_inverse))
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% plot Cepstrum
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%prepare for plot
figure
%plot logarithm diagram
plot(1+liftering_N:length(s_fft_log10_inverse)+liftering_N,s_fft_log10_inverse)
%label x-axis as "Quefreny (Samples)"
xlabel("Quefreny (Samples)")
%label y-axis as "Amplitude"
ylabel("Amplitude")
%showing plot title as "Cepstrum (High Pass Liftered)"
title("Cepstrum (High Pass Liftered)")
```

%%%%%%%%%%%%%%
 %%%%%%%%%%

%smooth the diagram for easy extremum finding(note : maximums will be remained as maximum !)

```
s_fft_log10_inverse = smooth(s_fft_log10_inverse,6);  
%find all local and global extremum booleans with at least 0.01 as the  
%minimum prominence distance  
TF = islocalmax(s_fft_log10_inverse,'MinProminence',0.01);  
%declare all possible indices  
x=(1:fl)  
%get all local maxima indices  
xtf=x(TF)  
%declare Quefrequencies  
s=1+liftering_N:length(s_fft_log10_inverse)+liftering_N
```

%%%%%%%%%%%%%% plot Cepstru with peaks
%%%%%%%%%%%%%%

```
%prepare for plot  
figure  
%plot smooth cepstrum liftered with peaks on it  
plot(s,s_fft_log10_inverse,s(xtf),s_fft_log10_inverse(xtf),"r*")  
%label x-axis as ""  
xlabel("Quefreny (Samples)")  
%label y-axis as  
ylabel("Amplitude")  
%title plot  
title("Smooth Liftered Cepstrum With Peaks")
```

%%%%%%%%%%%%%%
%%%%%%%%%%%%%%

```
%if found indices  
if length(xtf) >0  
    %get the max one  
    iPos=liftering_N+min(xtf)  
    %calculate pitch frequency  
    pitches(i)=FS/iPos;  
else %othersiwe , put zero as pitch frequency  
    pitches(i)=0;  
end  
end  
%%%%%%%%%%%%%% plot frame pitch frequency  
%%%%%%%%%%%%%%  
%prepare for plotting  
figure
```

```

%plot "Frame" as x-axis and "pitch values" as y-axis
%by 'o' drawing on it.
plot((1:FN),pitches,'o');
%label x-axis as "Frame"
xlabel('Frame')
%label y-axis as "pitch frequency (Hz)"
ylabel('Pitch Frequency (Hz)')
%title whole the plot as the average value(avera pitch in all frames)
title("Pitch Frequency : "+sum(pitches)+" Hz")

```

پاسخ سوال 2 :

میدانیم که سیگنال گفتار از دو بخش e و h تشکیل شده است که در یکدیگر تنیده (کانوالو) شده اند. یعنی

$$x[n] = e[n] * h[n]$$

و ما به یک تبدیلی نیاز داریم که این کانولوشن را به جمع تبدیل کند تا بتوانیم با جدا سازی آنها، هر بخش را مجدداً آنالیز کنیم. این تبدیل همان تبدیل کپستروم است که با گرفتن لگاریتم (که عملیات ضرب را به جمع تبدیل میکند)، این عملیات را برای ما میسر میسازد. بعد انجام لگاریتم، سیگنال به صورت حاصل جمع سیگنال e (منبع تحریک) و h (فیلتر مجرای گفتار) در می آید. کافی است برای یافتن فرکانس پیچ، حاصل را از لیفتر بالا گذر (HPL) عبور بدهیم تا نمونه های N (معمولاً 20) به بعد کیو فرکانسی بدست بیایند. این بخش همان بخش e (منبع تحریک) است، پس با یافتن مکان های قله های نمودار آن و تقسیم فرکانس نمونه برداری به آنها، میتوان به فرکانس پیچ دست یافت.

پایان