**AI\_speech HW3**

***2016707072 전자통신공학과 신민정***

*계수들은 text파일로 저장하여 python으로 plot하였습니다. code에는 최종 spectral envelop만 export하였습니다.*

*보고서 양이 많아 코드는 전체code란에 한번에 첨부하였습니다.*

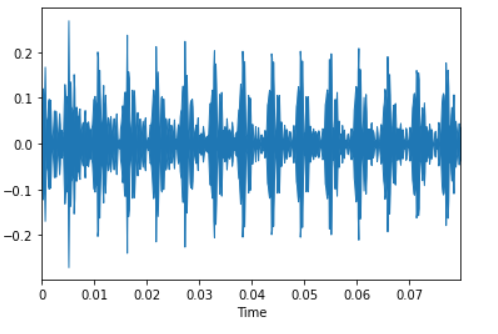
**과제목표**

LP와 Cepstrum 기반으로 음성 신호의 spectral envelope을 구하시오

**Input Data**

‘Male.raw’

음성 전체의 wave form을 plot해 보았습니다.



**Windowing**

Sample rate = 8,000. 처음 320개의 sample만 추출하여 Hanning window를 적용하였습니다.

텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

순서대로 Hanning window의 그래프, “Male.raw”의 waveform, Hanning window를 취한 waveform의 그래프 입니다.

Windowing은 HW1에서 구현하였던 함수를 사용하였습니다.

Hanning winodw뿐만 아니라 Hamming,Blackman을 audio waveform에 적용해본 결과입니다.

텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

텍스트, 지도이(가) 표시된 사진

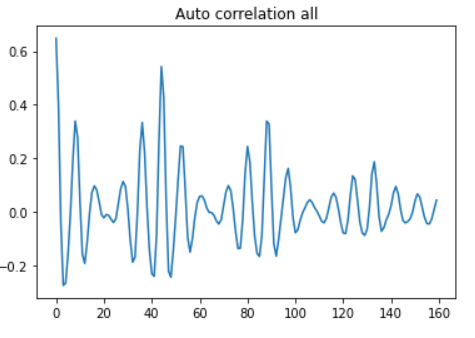
자동 생성된 설명

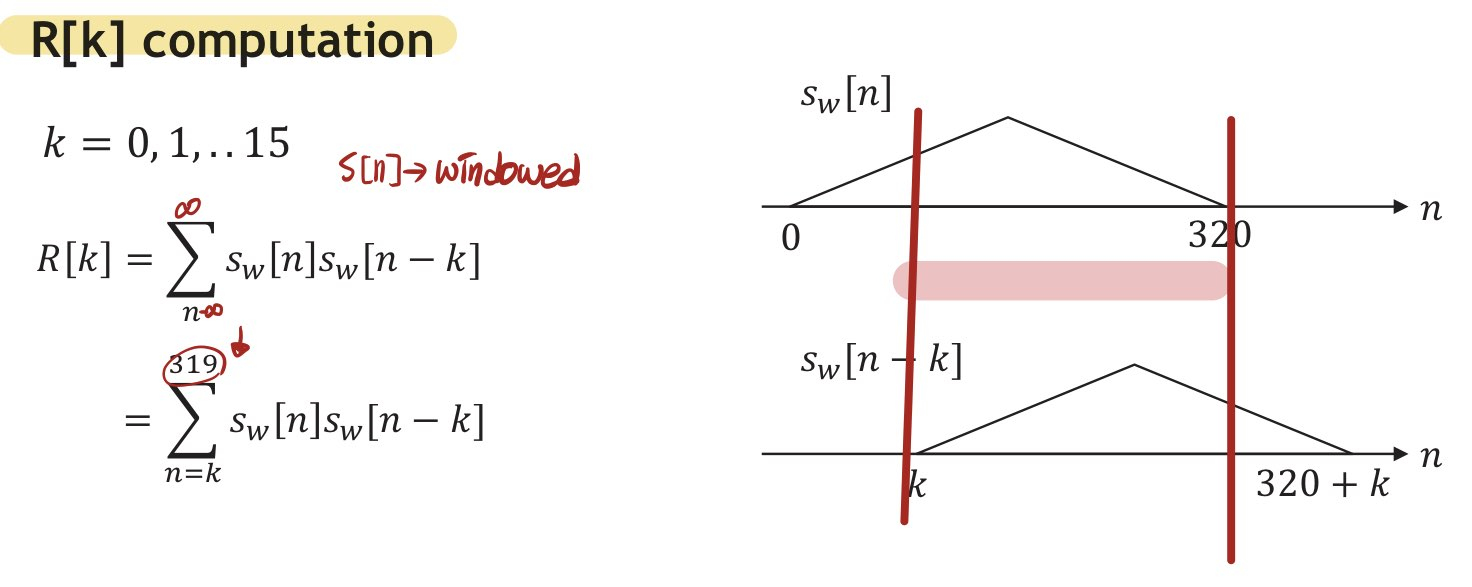
**LP model**

**Auto Correlation**

다음으로 R-coefficiant를 얻기 위해 auto correlation을 진행하였습니다.

전체의 auto correlation을 진행한 결과입니다.





R을 구하는 식은 다음과 같으며, autocorrelation은 -∞ ~ ∞의 범위이지만 이 경우 320 sample만 있으니 범위가 위와같이 줄어들게 됩니다.

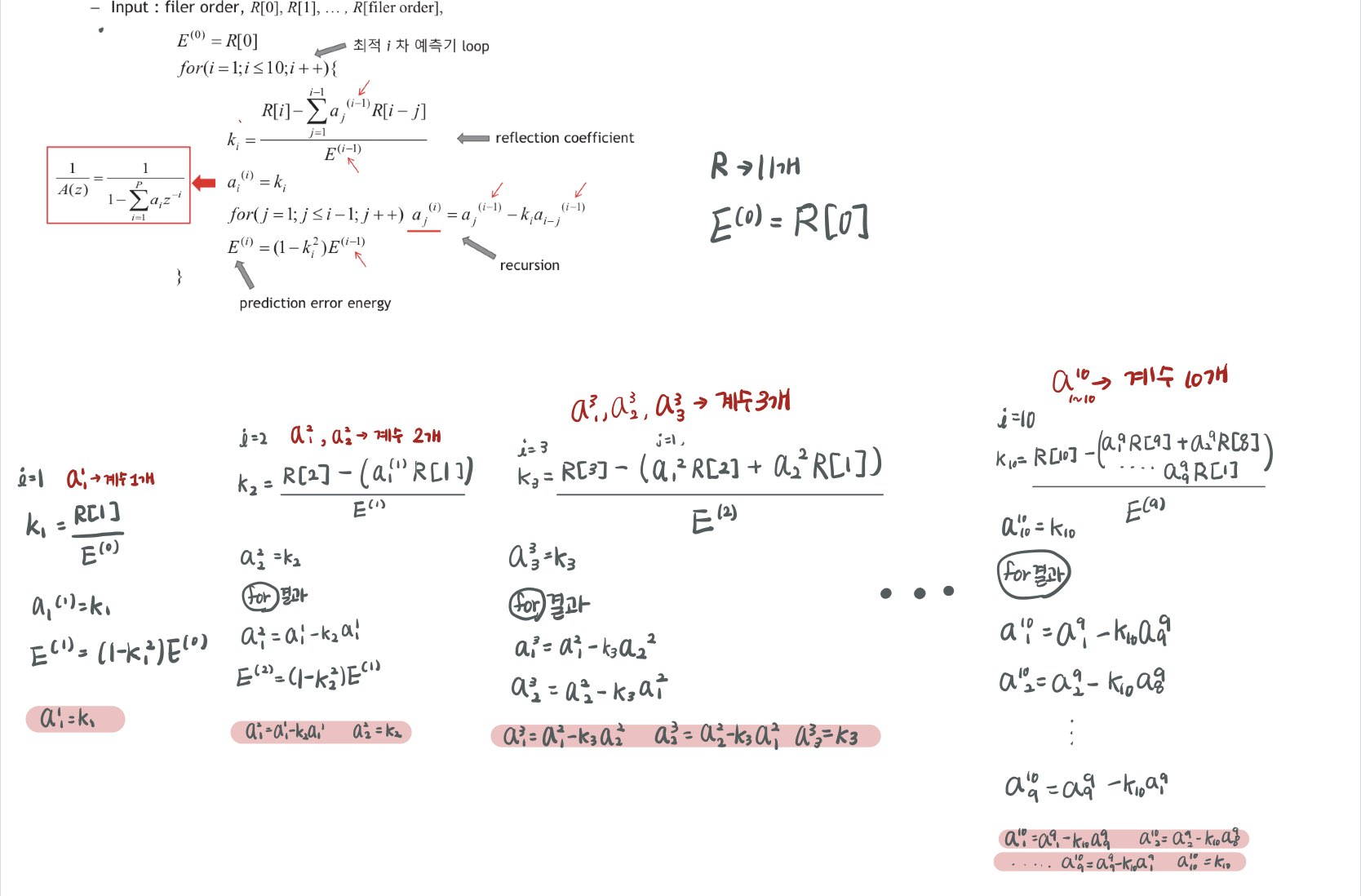
10차 필터를 설계할 것이기 때문에 k = 10이 되고, 총 0~10 (11개)의 coefficient가 나옵니다.

텍스트, 지도이(가) 표시된 사진

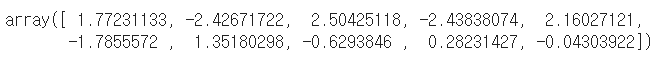
자동 생성된 설명

**Durbin Algorithm & Get an**

더빈알고리즘은



다음 배열은 더빈알고리즘을 통과한 10차 필터의 계수입니다(a10 )

****

an[0]=1,

an[1~10] = 더빈알고리즘을 통과한 a10계수

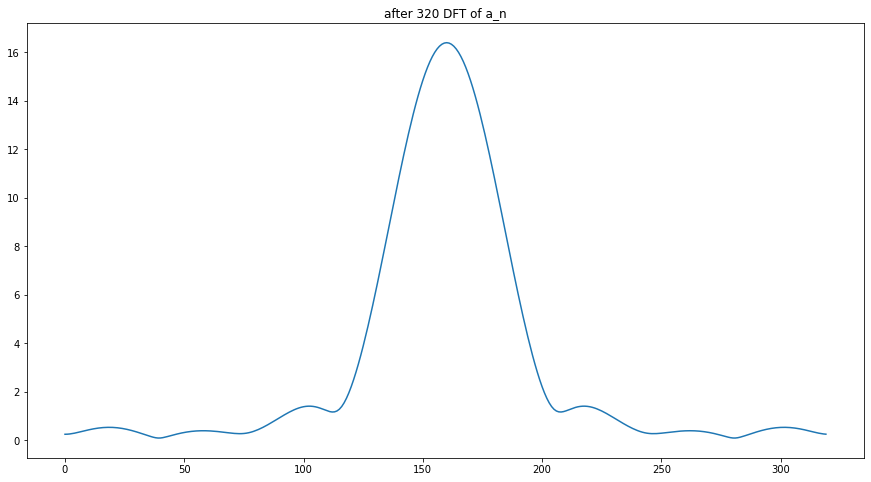
an[11~320] = 0 (zero padding)

을 한 an을 plot해 보았습니다.

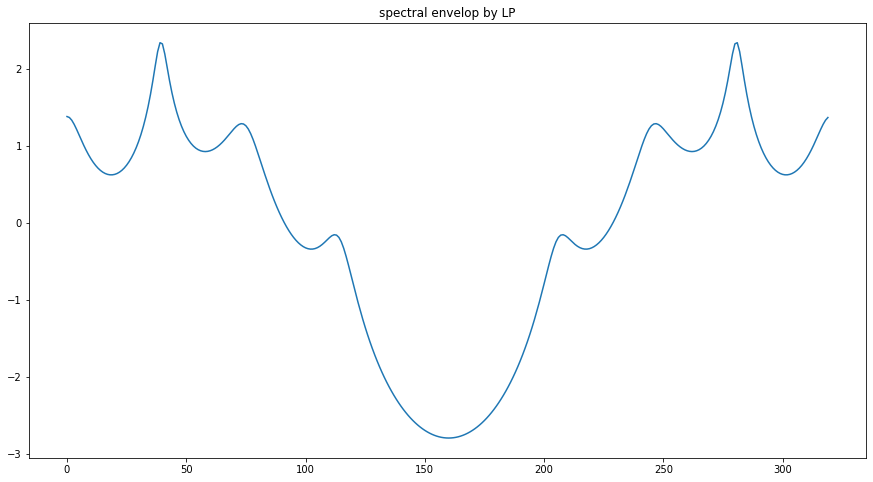
**스크린샷이(가) 표시된 사진

자동 생성된 설명**

**320-DFT**

****

**Get spectral Envelop**

****

**Spectral envelop from 15-order LP model**

**텍스트, 지도이(가) 표시된 사진

자동 생성된 설명** R15

**지도이(가) 표시된 사진

자동 생성된 설명**

**Spectral envelop from 30-order LP model**

**지도, 텍스트이(가) 표시된 사진

자동 생성된 설명**R30

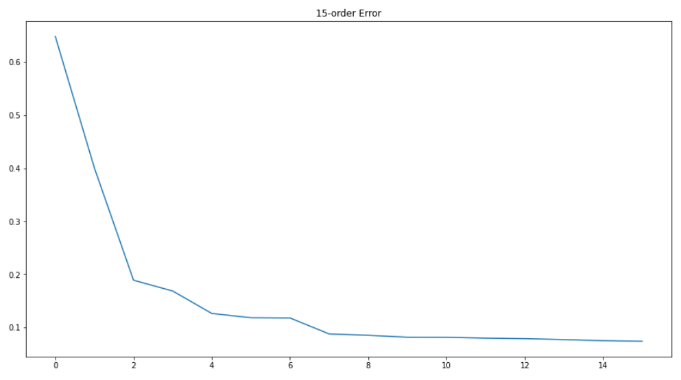
**텍스트, 지도이(가) 표시된 사진

자동 생성된 설명**

10차,15차,30차 LP model로 얻은 spectral envelop을 각각 비교해 보았습니다. 그래프에서 바로 알 수 있듯 대락적인 형태는 비슷하나, 차수가 높을수록 더 세밀한 spectral envelop을 얻을 수 있었습니다.

**Error 비교**

지도, 텍스트이(가) 표시된 사진

자동 생성된 설명지도이(가) 표시된 사진

자동 생성된 설명

각 필터의 Error를 plot해 보았습니다.

차수는 그래프에 함께 표기하였습니다.

이 경우 약 9~10차에 error가 수렴하기 때문에 10차 LP model로 spectral envelop을 얻는 것이 적절하다고 판단됩니다.

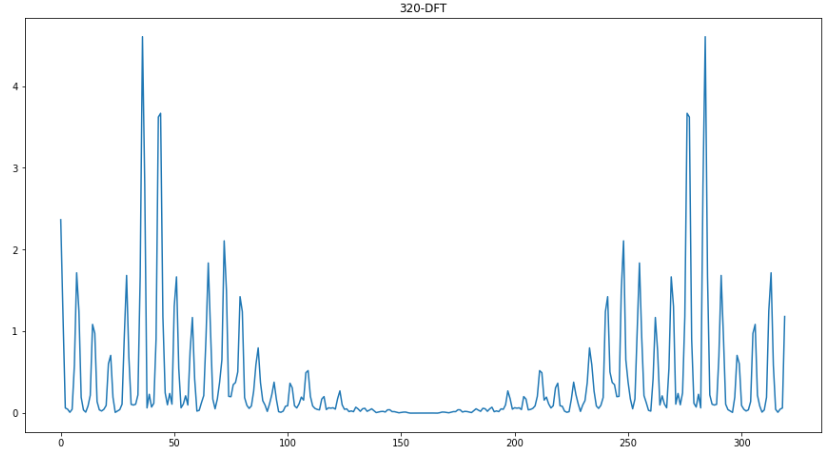
**Cepstrum**

LP와 마찬가지로 “Male.raw”의 320-sapmle에 Hann window취하였습니다.

텍스트, 지도이(가) 표시된 사진

자동 생성된 설명

**320-DFT**

****

Window를 적용한 음성신호를 320-DFT를 하였습니다.

**안테나, 개체, 캐비닛이(가) 표시된 사진

자동 생성된 설명**

우리의 귀는 Linear가 아닌 Log scale의 신호를 더 잘 듣는 구조입니다.

Window를 적용한 신호에 Log를 씌워 Log scale로 변환하였습니다.

이 신호에는 주파수가 높은 excitation signal과 주파수가 낮은 formant filter가 혼합되어있습니다.

**Inverse Fourier Transform**

**스크린샷이(가) 표시된 사진

자동 생성된 설명**

절댓값을 취한 후 log scaling한 신호를 inverse fourier transform을 적용하여 time domain으로 변환합니다.

원신호를 ifft한 것이 아니기 때문에, 처음의 domian과는 다른 제3의 domain으로 가게됩니다.

**Low time leftering**

**스크린샷이(가) 표시된 사진

자동 생성된 설명**

우리가 원하는 것은 low-time성분의 formant filter 입니다. 따라서 High time성분을 제거해 주어야 합니다. 중요한 사실은, real 신호이기 때문에 160을 기준으로 대칭되어야 한다는 점입니다.

이 경우 앞에서 15개의 sample만을 추출할것이기 때문에 앞(0부터)의 15개의 sample과, 뒤(319부터)의 15개의 sample을 추출하여 대칭을 맞추어 줍니다.

**Spectral envelop 30-point low-time liftering**

Time domain에서 (제3의 domain) filtering을 하였으므로 다시 주파수 영역으로 변환합니다.

**지도, 텍스트이(가) 표시된 사진

자동 생성된 설명**

**Spectral envelop 30-point low-time liftering**

기존 10차가 아닌 다른 차수를 filtering하여 얻은 spectral envelope을 비교해 보겠습니다.

해당 filter의 차수는 그래프에 함께 나타내었습니다.

**지도이(가) 표시된 사진

자동 생성된 설명**

**지도, 텍스트이(가) 표시된 사진

자동 생성된 설명**

**지도이(가) 표시된 사진

자동 생성된 설명**

**지도이(가) 표시된 사진

자동 생성된 설명**

기본적인 spectral envelope의 형태에는 변화가 없지만, 차수가 높을수록 세부적으로 표현하는 것을 알 수 있습니다.

**LP VS Cepstrum**

**지도, 텍스트이(가) 표시된 사진

자동 생성된 설명**

10차 LP로 얻은 spectral envelop과 15-point low-time liftering으로 얻은 spectral envelop을 각각 standardization을 한 후 plot해 보았습니다.

160을 향해 내려오는 형태의 그래프임은 동일하나, 세부적인 모양이 약간 다른 것을 알 수 있습니다.

Spectral envelop의 차이는 발음의 차이를 만들기 때문에, 소리로 변환하여 비교한다면 약간의 발음의 차이가 날것이라고 예상할 수 있습니다.

**최종Code**

**스크린샷이(가) 표시된 사진

자동 생성된 설명**

**스크린샷이(가) 표시된 사진

자동 생성된 설명스크린샷, 화면, 전화이(가) 표시된 사진

자동 생성된 설명화면, 앉아있는, 모니터, 검은색이(가) 표시된 사진

자동 생성된 설명스크린샷이(가) 표시된 사진

자동 생성된 설명스크린샷, 노트북이(가) 표시된 사진

자동 생성된 설명**