영어음성학
한국어 --조사와 명사가 붙어있고 사전에서 조사 따로 명기하기 때문에 분석어려워

음성데이터가 가장 처리하기 어려워

모음은 중심 자음은 초성이나 종성으로
철자와 소리는 구분할 필요 gap에서 g는 'ㄱ'소리 자음
vision mission sh 소리 는 목에 진동없어 혀 위에 안닿지만 vision 소리는 진동
G는 혀 닿아야해 여기서 진동빼면 ch

year vs ear
목의 진동 유무 유성음 voiced
무성음 voiceless sound
모든 모음은 유성음
자음은 유성 아니면 무성

lwrj-voiced
ptkh - voicelss

코로 나오는 소리 mn n
입막고도 나오는 소리

pbm 양순음
fv 아랫입술 윗니
혀랑 윗니- thigh thee

단모음 monothongs
복합모음 diphthongs
phonology 음운론- 소리 그룹, 시스템에 대한 이론 머릿속에서 추상적으로
phonetic음성-- 더 물리학적으로 인지적인 게 아닌 물리적인 측면
speech는 기본적으로 사람의 말을 지칭

1)articulatory phonetics
공기를 보내 고기압에서 저기압으로
성대의 기문이 완전 열린 상태 '하'

아 - 진동이 생기는 것은 기문이 막혀있고 바람에 의해서 떨리기 때문
남자 1초에 100번 여자 1초에 200번 정도
우리는 아에이오우로 생각을 하지만 phonetics 상에서는 연속적으로 변화
'싱크'

아에이오우의 차이는 성대는 아님(소리 높이)
이 차이는 결국 입모양 (혀 위치와 턱 등)
볼펜-- 턱의 위치가 변하지 않아 이 상태에서 혀로
턱의 높낮이가 주된 결정요인은 아니야
한국어는 음절이 반복 리듬 시간 같게
영어는 accent 'stress'가 반복
한국어는 턱을 많이 쓰는 언어
그와 반대는 혀로 혀가 잘 움직이는게 중요

2)공기를 타고 가는 과정 acoustic phonetics 공기가 어떻게 공명하는지 그런 일반적 원리 사람이 개입되지 않는 분야
acoustic physics

3)auditory phonetics
귓바퀴 미세한 진동을 증폭시켜 더 잘 듣기 위함
고막 ear drum 여기도 물리

귀 코 인강(목젖부터 후두까지의 긴 관) 후두(larynx)
경구개 연구개 hard soft(velum) palate
윗니 위에 alveolar -> d ㄷ은 윗니
영어의 많은 소리들이 여기서
uvula 목젖

uppper structure(고정), lower structure(유연)
lower - epiglottis - epi-뚜껑??후두개
tract라는 관 식도와 기도

lower structure - tongue

nasal tract -m n때 열려 oral tract은 닫혀
'아'-oral만 사용 nasal은 사용하지 않아

velum이 올라가면 nasal tract 막혀
모든 모음과 비음제외 모든 자음은 nasal 막혔을 때
비음 자음만 nasal tract --velum이 lower인 상태
숨쉴때도 velum이 lower

oro nasal process
larynx - 후두 진동 voice or not
모든 모음과 유성자음

articulatory process
phonation (voiced voiceless)

voiced일때는 닫혀서 진동발생

velum lower- 숨쉴 때, m,n, ng
대부분은 nasal tract 닫혀

articulatory process
lips -bp /tongue tip-dt /tongue body -g, ng
파 / 타 / 카
이 세 기관을 constrictor 협착을 만드는 주체
constriction location CL
constriction degree CD
lips 2개
yearn g 둘다 body를 쓰지만 후자는 뒤쪽 2개
th-- tip 윗니를 쳐 뒤로 alveolar d t n 4개
CD
at t --stop 완전 막혀 ptkbdg m n ng

s- fricative s z f v th sh dg
영어 approximants 4개 r l w j(y)
vowel- 막힘이 없는 것이 정의 그 자체
CD에서 자음은 세종류
국어 폐쇄음 마찰음

velum raised larynx의 틈인 glottis open
tongue tip location alveolar
stop -> t
모든 모음은 constrictor로써 tongue body만 사용
자음-- k
여기서 nasal tract - velum lower되면 ng
(여기서 glottis는 closed)

phoneme

pitch intensity spectrogram
스펙트로그램의 띠들을 포만트라고 지칭 f1, f2 ->모음을 결정하는 요소

pitch setting

vowel acoustics

Signal processing
DSP

0924

입, 후두, 연구개
1)constrictor CL CD
lips t t tb
2) velum
3) larynx

p
-cl bilabial cd stop
velum raised

larynx open (voiceless)

z -fricative tt -> alveolar??

frequency 주파수 사인곡선 1초에 몇 번진동하는지 주기 + 진폭
진동수-- 성대의 떨림 횟수와 동일
vibration of vocal folds
목에 대고 성대의 진동만 녹음하면 어떤 모음이 발화되는지 파악하기 어려워

신호
사인 파동 가장 기본적인 형태고 결정짓는 것은 frequency와 magnitude(amplitude)
결론 -세상에 존재하는 모든 소리를 포함한 신호들은 여러 사인 파동의 결합으로 표현된다
복잡한 신호들을 쪼갤 수 있다는 것
Fourier
simplex tone complex tone
??- complex tone -> 반복주기 가장 진동수적은 사인파와 같은가 ??

사인파에서의 x 축 -t
y 축-- value

x축 frequency y 축 magnitude-- spectrum --- 이퀄라이저

simplex -> complex (synthesis)
<- (analysis)

spectrum은 시간개념없이 특정 time point
그러나 spectrogram은 spectrum을 시간축으로 늘여놓은것

spectral analysis에서 가장 왼쪽에 있는 simplex tone이 내 목소리의 pitch
그리고 그 진동수의 정수배들의 simplex tone들을 합치는 것

목소리의 음의 높낮이는 pitch이고 이는 가장 왼쪽 스펙트럼의 진동수 ?

성대의 소리를 바로 뽑으면 source라고 지칭 그리고 tube에 따라 어떻게 달라지는지 -> filter

source는 점진적으로 줄어드는 모양
처음의 simplex tone -F0 fundamental frequency pitch, number of vocal folds in a second -- 같은 대상 지칭

배음 harmonics f0 *2 3 4....
여성의 경우 f0가 더 높아서 듬성 듬성
남자가 10000hz이런 기준선까지 갖는 배음의 숫자가 더 많아

filter보면 -- 배음의 구조는 그대로 유지되지만 amplitude의 패턴이 깨져
또 스펙트로그램에서 보면

wave와 스펙트로그램 모두 x축은 시간
스펙트로그램은 y축 -- frequency

스펙트로그램에서 까만게 강한 것 - low frequency에서 강해

스펙트럼이라는 한순간의 그림을 시간을 축적시켜서 스펙트로그램으로 가면 amplitude가 z축으로 간다고 생각

source filter 읽어오기

0926
simplex sound pure tone

spectrum x frequency y amplitude (at specific time point)
wave form x time y value

spectrogram x time y frequency

목소리 source--f0의 배음의 합으로 이루어짐
F0 pitch hz
filter도 source가 그렇기 때문에 고주파로 갈수록 약해지는 경향은 있지만

source -harmonics-사인파의 배음의 합으로 이루어져
filter-
EGG-voice source에서 녹음한 것
audio-실제 목소리
-> vocal tract에서 filtered 됨

peak-
valleys

누가하든 '아' 소리의 패턴은 똑같게 나타나
첫 번째 산맥에 해당하는(밑에서부터) 주파수- 첫 번째 포만트 그다음이 두 번째 포만트 F1 F2

spectrum에서 첫 번째 harmonics F0
스펙트로그램 source에서 formant가 만들어지는 것이 F1

긁는 소리등은 배음이 안나와
기타소리는 목소리처럼 배음 complex tone
인지하는 음의 높이는 같아
praat으로 voice source 만들기

10개 만들기
stereo 하나의 object 10개의 채널을 가진 스테레오
combine stereo
수학적으로 합하지는 않은 상태
독립적으로 stereo 로 존재하는 상태
stereo 합쳐 반대는 mono
convert to mono -complex tone

배음을 무한대로 합친다면

반복주기 - f0랑 같고 그리고 인지청각학적으로 100hz랑 높이 같다고 인식 (1000이 들어간 소리이긴 하지만 인지적으로 들리지 않아)

무한대개수로 합치면 피크 하나 0000 피크 하나 000 이런식
pulse train

source spectrum x vocal tract
output spectrum --
처음나오는게 f0 산맥이 F(ormant) 1 (peak) 그에 해당하는 frequency를 읽으면 돼
F3 F4는 무시해도 되고 F1 F2로 웬만한 모음은 커버 가능
vowel space
F1 F2 입의 위치와 일치
F1- 높낮이를 결정 F2-전후를 결정
F2 x축 F1 y 축

한국어 ㅏ와 영어 a의 차이
영어가 더 back and low

drag release 이중모음

------모음

coding -

1001
자동화의 반복을 위한 코딩
컴퓨터 언어의 단어란 '변수'
1) 변수에 정보를 할당하기
2) 조건절 문법
3) 반복 for 문
4) 함수 def

오픈북 (노트북 전자기기 제외)

단순암기는 지양
여러 가지 개념의 결합도 가능

변수에는 문자와 숫자 할당 가능
데이터 형에는 int(정수) float(실수), str(문자형) 존재
a=3, b='English'
type(a)-> int
type(b)-> str

데이터구조
리스트 a= [1,2,3]
a=[[1,2],[3,4]] 이렇게 리스트 안의 리스트도 가능

b=(a,b,c) -튜플 수정이 불가하기 때문에 보안측면에서 유리한 면 존재

c={'Thor':1500, 'Cap':100, 'Tony':50}
이런 형식을 딕셔너리라고 한다

인덱싱을 할 때는 대괄호를 쓴다
a[0]일 때 첫요소 불러온다
그러나 딕셔너리는 key와 value가 존재
c['Cap]->100이런 방식이다

문자형 string 다루기
s='abcd
s.upper() -대문자로 ABCD
s.lower() 소문자로
s.find('b')=>1 b의 위치 파악
s.rindex()=>오른쪽에서부터의 위치 파악
s.replace('a','g') ->gbcd
s.split('b')=> 'a', 'cd' ->b기준으로 나눠
token=s.split('b')
st='b'.join(token)-> 다시 합쳐

정리

5speech organ
major articulation 3가지 (혀입술)
+ velum , larynx 총 5가지

larynx vocal cord
close-> vibration, voiced vzlmai
open-> voiceless->f s k p h

velem oro nasal process
nasal -m n ng
velum= soft palatte
velum lower-> nasal tract open
코로 숨을 쉬니 같은 메커니즘 lower

lips tongue tip tongue body 이 세가지에 의해 정확한 소리 정의돼
lips p
tip t
body k

control of constrictors
좀더 미세하게
Constriction location & degree
 x y
아래가 유연 위가 고정
lips - 두가지로 bilabial 아랫입술과윗니 labiodental

body -palatal velar
tip이 가장자세하게 분류
dental-th alveolar -대부분
retroflex혀말아서
sh-> palato alvelolar

constriction degree
-hit- stop
-hit-turbulence -fricative
-모음과 구분안되는 approx r l w j
-모음

larynx
velum
뭘 쓰는지
CD CL- >영어 모든 소리 define

모든 가능성
모든 조합이 있다는 보장 x
-> 영어에 없을 가능성

이런소리가 영어에 존재하는지 없는지 얘기해야
lips가 CL로 velar가질수 있는가 -아닌 듯?
영어에서의 gap인가 사람이 못하는가-사람이 못하는 것
lips가 alveolar CL인 언어가 이론상 가능한가 -> 생리적으로 가능한가 그러함
velar쪽은 신체적으로 불가능함
가능해도 accidental gap으로 없을수잇고
그냥 신체적으로 불가능할 수도 있어

intensity pitch formant

source filter theory 그림 설명 가능해야
vocal tract없을 때 나는 소리 source
source에서 pitch 조절가능 -이 pitch는 첫 pure tone F0의 크기에 의해 결정돼

같은 pitch여도 아와 이가 다른 것은 vocal tract 때문
source의 spectrum 120 240 360....
점차 줄어드는 방식으로
입모양이 filter 역할로 ->source의 스펙트럼을 재구성해
peak와 valley 결정
peak가 어디서 이루어지는지->formant
F1 F2 F3 포만트도 무한으로 있지만 실질적으로 F1 F2
f0의 경우는 source spectrum의 첫 tone의 frequency가 pitch

F1 F2에 의해 모음이 결정돼
F1 혀의 높낮이 혀가 높으면 F1이 더 밑에 있어
F2 -전후 - I전 ae후
f2가 더 클수록 혀가 앞쪽

wave
spectrum

합쳐진 wave에서 각 성분알기 힘든데-> spectrum을 보면 분석이 돼서 나와
-equalizer

변수할당
변수종류-수와 문자
수-int float
하나 이상의 정보 저장
list tuple 괄호의 차이
[] ()
dictionary

(numpy xxx)

string가 list의 유사점
정보에 접근하는 방식 -index이용
딕셔너리 index에서는 key입력
string도 []로 index

-1맨뒤 range [1:2] 1위치부터 두 번째이전ㄲ지 -**<u>뒤에 범위는 자기자신포함하지 않아</u>**

strip-
split 과 join

syntax for과 if
함수들 -type, len, print, range
for I in a:
    ... indent해줘야
for안에 for 넣기
range로 명시적으로

enumerate
zip
format
if a==0 :
else: : 필요

뭔가쓸수도 있어

harmonics - f0부터 모든 것을 포함
배음

모음 무조건 tongue body
모음 CL -정의를 따로 하지 않음
praat specific한 사용법은 나오지 않아
amplitude의 결정요인은 pressure 및 소리가 나는 크기 강도
wave form에서 강하게 나오나 약하게 나오나

int( ) -> 이런것도 함수
class의 method - join split
파찰음은 stop과 fricative가 결합된 것
혀의 앞뒤
실제 발화
heed가 더 front(hid)보다

spectrum상에서 f0가 들리는가 - pitch
harmonics들의 위치는 그대로고 filter되면서 shaping만 달라져
source 상에서 f0와 그 정수곱들
filter 지나고 위치는 안바뀜
개개 pure tone들의 amplitude만 바뀜
peak가 중요 그것이 f1 f2
여자가 f0가 높아
그러면 듬성듬성있는게 여성
주어진 frequency range속에 여성의 pure tone 더 적어
0-10000 hz 120hz인 사람이 pure tone 몇개 존재하는지 -그 숫자가 filter 된 이후에 변하는가 안바뀜

1029

이미지
픽셀별로 숫자 부과한 것들-> 행렬로 볼 수 있어
크기와 방향을 가진 벡터-> 열벡터나 행벡터로 표현
흑백의 경우 한 장 컬러의 경우 세장 RGB

영상의 경우 계속
차원--
이미지 2차원(흑백) 3차원 (컬러 RGB)
영상 (4차원)

소리의 벡터화
waveform을 이산적인 값들로 나눠

텍스트의 벡터화
50000 표제어 사전의 경우 0 50000개 그중 하나만 1로 표현해서 순서 나타내기

1031

패키지 안의 패키지
numpy.A.D.(함수명) 이런식으로 불러올 수 있어

from numpy import A 이런식으로도 가능
numpy 와 list 유사성 -모든 데이터 numpy

np
arange [시작값, 끝값, 간격, dtype='float64']

np.array -> np 행렬로 만들어

astype
data= np.random.normal(0,1,100)
->표준정규분포에서 100개의 임의가치 뽑기

np 행렬에서 대괄호의 개수에 따라 차원을 계산할 수 있어

np.savez->npz파일로 행렬 저장

연산기능
sum min max mean median std
axis=0-> 행끼리 합치기 1 열끼리 합치기

sampling rate ==10000 -> 1초에 만 개의 숫자를 담아
44000 cd 음질 그 이상은 구분하기 어려워
sine 곡선-- 192000 1초에
Hz -pitch

1105

pure tone(sin cos 곡선)들의 합으로 복잡한 소리 만들어내기
sinusoidal function -> 이를 만들어내는 것 phasor
sin ->rad 값으로
rad: 호의 길이와 반지름의 길이가 같아지는 각 1 rad
180도는 파이 rad
 + Euler's theorem

e^(theta* I) = cos theta + sin theta I

I-> imaginary <-> real
e I 는 모두 상수로 주어져
결국위의 식은 theta에 대한 상수

complex number 복소수가 모두 포함
a+bi

theta= 0, pi/2 pi 3pi/2 2pi
f(theta)= 1, I, -1, -i, 1
주기를 가져

복소수 plot하는 방법
at complex plain 복소평명
x축에 a y축에 b
a+bi
(1,0), (0,1) (-1,0) (0,-1)

원에서의 각도를 theta로
theta 값의 증가에 따라서 원위를 이동

projection 투사 정사형
x축에 투사하면 x축내에서 이동
y축도 마찬가지
실수의 관점에서만 보겠다고 할 때
sin 0부터 cos 1부터

실수관점 cos과 같아
허수부분- sin과 같아

pure tone 의 frequency
sin theta->시간의 개념이 안들어있어
그저 각도일 뿐
초당 정의 안되어있으면 소리를 만들어 내기 어려워

time 과 theta의 벡터 크기는 커 5000

1107
sincosine phasor
Euler phasor --

오일러공식에는 각도값인 radian 만을 넣어야함

1112
phasor 두 종류
sampling rate와 frequency간의 연결점

sr 이 100헤르츠라면 표현하는 숫자가 1초에 100개
freq 1헤르츠 표현할 수 있는가

->표현할 수 있어
한번의 사인파 주기 있으면 돼
2hz freq도 가능

10000hz는 가능한가
sr 100이니 1초 주어진 숫자 100개
10000번의 진동을 표현하지 못해
sr 이 충분히 있어야 그만큼의 진동수 표현 가능

sr=10hz면 최대 5번의 주기만을 표현 가능
절반이 최대 5최대
-> Nyquist Frequency sr/2 hz
CD 음질 - sr 44100hz
n f 는 22050hz 아주 높은 소리까지 표현 가능

왜 cd 음질을 이렇게 잡았는가 -- 인간의 가청주파수가 20000
유선전화 sr 8000
말소리 웬만하면 4000이내에서 가능
그러나 누구인지 구별하는 것은 좀 더 높은 수준에서
핸드폰 16000sr nf 8000
초음파 인간의 가청주파수인 16kHz넘어서는 범위

1114

spectrum 으로 봤을 때
freq 100 200 5000
각각 비슷한 amplitude

Formant 산맥
그 위치에 따라 아도 되고 이도 되고

코딩
1)flat 하기보다는 gradually decreasing
2)여기에 산맥만들기

우리의 귀는 sin cos phasor의 이동은 못느끼지만 주파수에는 가능

500
1500 2000 3500 주파수에 산맥 할당

1119

Fourier Transform

matrix and vector
AI which transforms data
 and data needs to have a form of vector
데이터의 형식에 따라 그 인공지능의 유형이 결정돼
그 중간의 것이 행렬의 형식을 가져

기계학습 -- 그 행렬의 원소값들을많은 데이터를 통해 얻어내는 과정

column vector

벡터공간과 선형결합
c d scalar v w vector
cv + dw
벡터공간은 무수한 벡터들이 선형결합을 통해 만들어내는 공간
1사분면은 벡터공간이라고 할 수 없어
R^n 공간-> n개의 기저벡터 basis
벡터들이 n개의 원소가진 열벡터들로 구성

column space
null space
row space
left null space

A= [[2,1],[-1,3]] 2 1 -1 3 두 개의 열벡터를 통해서 하나의 평면을 만들어 (선형결합)-> 열벡터에의한 column space
열벡터의 차원은 2차원 -> 그 Column space는 2차원 넘지 못해

n차원의 공간은 n개의 기저벡터들에 의해 spanned

한 선상에 있지 않으면 독립 -
한 선에 있을 경우-> dependent
2,1 -1,-0.5
whole space = R2
column space=>R1

whole-dim ->nrows
column-dim-> n of indep column

R3 -세 벡터가 같은 평면에 있지 않는다면 모두 독립
3x2 행렬 whole r3 columnspace plane

transpose
2X3 행렬 열들 독립이면
whole r2 column r2

4개의 space
columnvector 관점에서의 whole space column
row vector 관점의 whole, row

3x2 column 3차원
row 2차원

여기서 두 열벡터가 종속이라서 비어있는 공간 나머지 두차원-- null space
null space의 기하적 개념은 whole space에서 row space 제외한 나머지
Ax=0 만족하는 벡터 x 들의 집합

row null
column left null 이렇게 연결

mxn m-r left null r column whole-m
n-r null r row whole -n

linear transformation
Ax=b 입력벡터와 출력벡터의 차원은 달라질 수 있어
mn n1 -> m1
A: transformation matrix

grid 좌표계적인 설명
1 1 일 경우
1 0 0 1 이렇게 기저벡터를 만들어 그리고 이 좌표계를 행렬따라 좌표계기울어
transformation matrix의 두 열벡터가 dependent 하다면
inverse가 안돼
Ax=b x= A-1 b
dependent하다면 변환행렬의 좌표계가 일직선처럼되기 때문에 펴서 어떻게 될 지 몰라찾아갈수없게돼

determinant의 값이 좌표계기울어진 그 사각형이나 다이아몬드의 면적과 같아
0이라면 역행렬 불가능

eigenvector
행렬을 곱했을 때 크기는 달라지지만 방향은 같은 벡터를 지칭
그 크기의 스칼라를 eigenvalue라고 지칭