

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 진동빠면 ch

year vs ear

목의 진동 유무 유성음 voiced

무성음 voiceless sound

모든 모음은 유성음

자음은 유성 아니면 무성

lwrj-voiced

ptkh - voiceless

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

upper 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 포먼트라고 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 번 진동하는지 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 진동수 적은 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 늘어놓은것

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 칭 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  $f_0 \times 2, 3, 4, \dots$

여성의 경우  $f_0$ 가 더 높아서 음성 음성 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 똑같이 나타나  
첫 번째 산맥에 해당하는(밑에서부터) 주파수- 첫 번째 포만트 그다음에 두 번째 포만트 F1 F2

spectrum에서 첫 번째 harmonics F0

스펙트로그램 source에서 formant가 만들어지는 것이 F1

굵는 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 인지청각적으로 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