2013 한국문화연구소 제1차 콜로키움

4월 17일

강제음성정렬시스템을 이용한 대용량음성자료연구

영어영문학과 윤태진



들어가는 말...

- 음성은 눈에 보이지 않 는 실체이기 때문에 쉽 게 습득하기는 어렵다.
- 하지만, 언어의 주요기 능인 의사소통을 원활 하기 위해서는, 즉 원활 한 말하기와 듣기를 위 해서는 음성이 가지고 있는 언어학적인 정보 (linguistic information) 을 파악해야 한다.





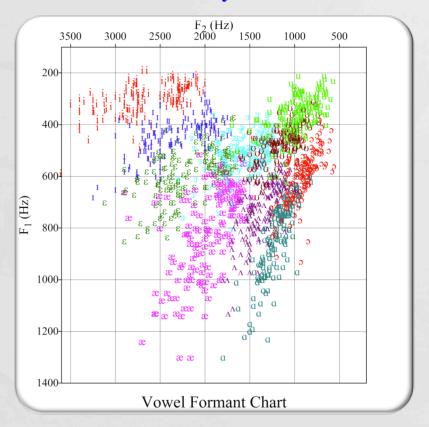
말뭉치(corpus)

- 말뭉치(corpus)는 언어 자료를 전자 자료로 만들어 데이터베이스화한 것을 주로 일컫는다.
- 언어 연구를 위해서 대용량의 말뭉치(large-scale corpus)를 구성하고, 이에 대한 연구를 시도하는 것은 비교적 최근의 일이다.
- '구어 말뭉치'(spoken corpus)는 구어(spoken language)를 대상으로 구축한 자료를 일컫는다.
- 비록 구어 말뭉치는 연구의 역사가 짧고 학계에서 도 풍부한 자료 구축이나 그 활용에 대한 깊이 있는 논의는 이루어지지 않은 상태이지만, 지속적으로 그 중요성이 강조되고 있다.

Peterson & Barney (1952)

불과 60여년 전만하더라도, 가장 대용량의 음성학 연구는모두합쳐 고작 30분도되지 않는 음성자료를 가지고 분석을 한 것이다
 (Peterson & Barney, 1952)

Peterson & Barney의 모음도



TIMIT (Texas Instrument & MIT)

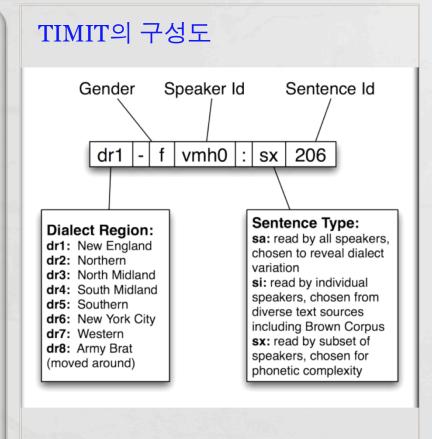
 1990년대 초반에는 약 300분(약 5시간) 정도 분량의 음성 말뭉치를 다루기 시 작했다

(TIMIT, 1991; Byrd, 1994)









SA1: She had your dark suit and greasy wash water all year

Status Quo

- 오늘날은 인터넷, 컴퓨팅 파워(Computational Power), 그리고 대용량 저장 장치의 발달로 인해 음성과 언어 연구에 새로운 paradigm을 가져오고 있다.
- 소규모의 인위적으로 개발된 자료를 연구하 던 관행에서 점차 벗어 나, 대용량의 자연 발화 말뭉치를 분석하려는 시도가 생기고 있다.

- 전화 대화 말뭉치
 (Conversational telephone speech)
 - LDC (Linguistic Data Consortium)
 - 300,000 분석 (>5000 시간)
- 미 연방대법원 구두 논쟁 (U.S. Supreme Court Oral arguments;

http://www.oyez.org)

- 540,000 minutes (> 9,000 시간)
- 뉴스, 인터뷰, pod cast, oral histories, 정치 토론, audio books 등등

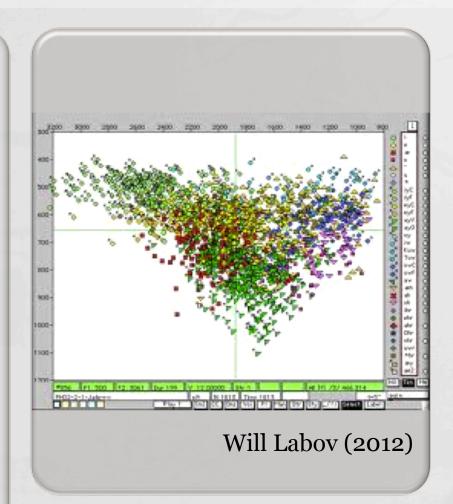
구어말뭉치 언어 현상 분석의 의의

- 연구자의 직관에 의존하는 연구가 아니라, 언어 사실에 대한 관찰을 바탕으로 결과를 도출하는 객관적인 연구이다.
- 언어가 고정되어 있는 것이 아니라 잠정적으로 변화하는 것이라는 점에서 잠정적 문법화(grammaticalization) 현상을 보여 줄수 있다.

(Cole & Hasegawa-Johnson 2012)

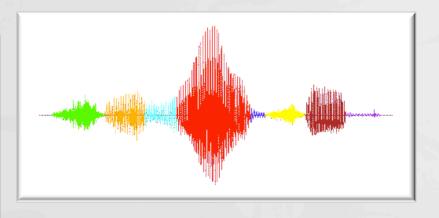
구어말뭉치 분석도구의 필요성

- 대용량의 구어말뭉치는 언어의 변이를 연구하는 데 있어서 매우 강력한 자료가 될 수있다.
- 대용량의 말뭉치를 기반으로 하는 분석은 다양한 언어연 구를 할 수 있는 가능성을 열 어줄 수 있다.
 조건:
 - ① 말뭉치를 공개하여 연구자들 이 이용할 수 있다면, 그리고
 - 2 연구자들이 그러한 대용량의 말뭉치를 보다 수월하게 연구 할 수 있도록 만들어진 도구 들이 있다면...



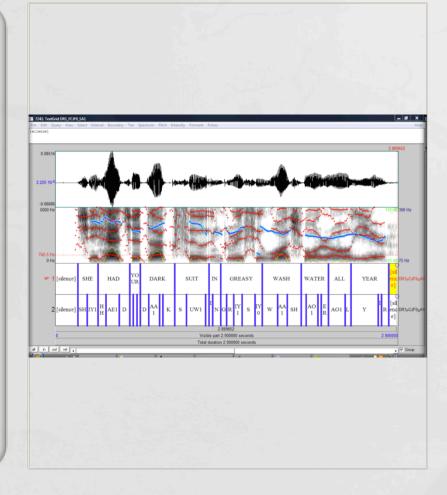
기술적 난관

- 대용량의 음성자료를 가지고 언어연구를 하고자하는 연구자들은 비록 녹음된 구어자료를 가지고 있더라도 기술적인 난관과 제약에 직면하게된다.
- 음성자료가 분절되어 전 사(transcription)되어 있는 음성자료들을 구하 기가 어렵다는 것이다.



transcription

- 보통은 음성자료를 음성분석용 컴 퓨터 프로그램의 도움을 받아 수 작업으로 전사하는 경우가 많은 데, 이는 시간이 매우 많이 들고 인 력이 많이 든다는 단점이 있다.
- 예를 들면, 음성언어를 전사하는 데 보통 실제 음성언어를 말하는 데 걸리는 시간의 400배 이상의 시간이 든다.
- 즉, 1분짜리 음성 언어를 전사하는 데 약 400분이 걸릴 수 있고, 이는 음소 하나를 전사하는 데 약 30초 의 시간이 든다는 것과 같다 (Leung & Zue 1984)



영어강제음성정렬장치 Upenn Aligner

- 펜실베니아 대학교 음성학 실 험실에서 개발한 일명 Upenn Aligner는 HTK (Hidden Mar kov Toolkit, Young et al. 200 6)에 기반을 둔 자동음성정렬 장치이다.
- 미국영어를 위해 개발
- 음성파일(wave file)과 전사파일(transcription file)이 있으면, 전사파일에 있는 단어와 음소들의 시작지점과 종결지점 등을 음성파일에서 찾아정렬시켜주는 기능을 한다.

| Upenn Aligner |
|--|
| Online Processing System |
| |
| (*) Transcription File |
| From File Choose File |
| ○ Type Here |
| (*) WAV File Choose File |
| (*) Email Again |
| First Name Last Name Organization |
| Language English ‡ |
| Progress Ver. 1.10.15.1 Submit Reset |
| Information Please choose your files and click submit. |
| The results, in the format of Praat TextGrids, will be sent to you in email. |
| |

연구목표

- 한국어를 처리할 수 있는 자동음성정렬장치를 개발하여 풍부한 음성자료를 실험음성학적인 방법으로 이용하고 분석할 수 있도록 한다.
- 대용량의 음성자료를 분석하는 데 용이하게 사용될 수 있는 도구들이 널리 사용되고 있지 않은 여건으로 인해, 아직까지 대용량의 구어 말뭉치를 이용하여 현대 국어 말소리의 변화 를 전면적이고 체계적으로 한 연구들을 찾아 보기는 힘들다.

연구 자료: 서울말 낭독체 발화말뭉치

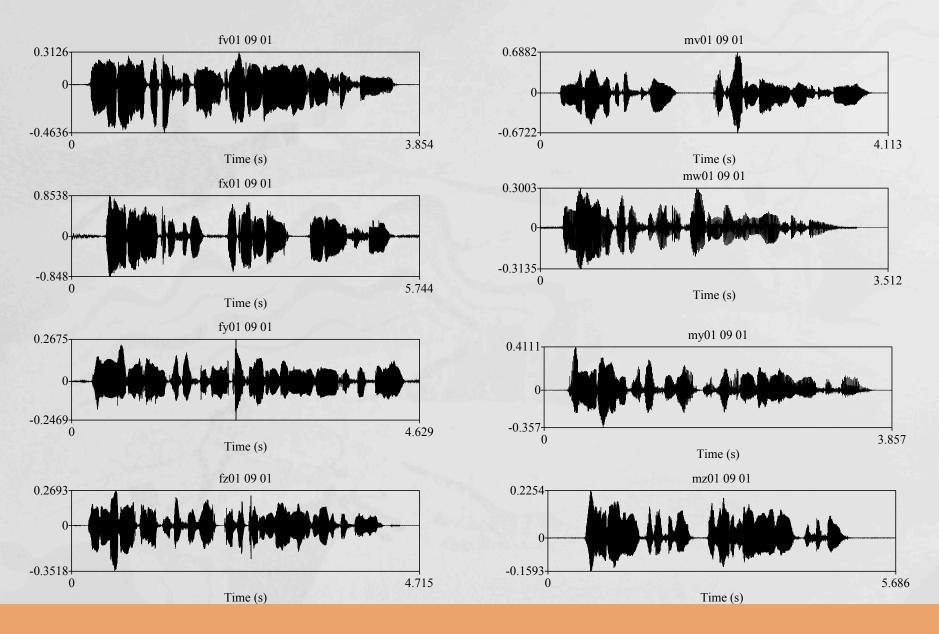
- 국립국어원에서 2005년도에 분배한 말뭉
- 2대 이상 서울 경기 지역에 거주하고 서울 말을 사용하는 20대에서 60대 이상에 이르 는 화자 120명 (남자 60명, 여자 60명)의 낭 독체 말소리를 녹음하여 만들었다.
- 비록 수년 동안 국립국어원에서 수집해 만들고 대중에게 공개한 음성자료이긴 하지만, 이 음성자료를 활용한 연구가 빈약하다.

서울말 낭독체 발화 말뭉치

- 120명의 남녀 화자들이 각자 초 등학교에서부터 고등학교에 이 르기까지 배운 소설, 수필, 논설 등 다양한 형태의 글들을 방음이 처리된 음성실험실에서 편안하 게 읽은 음성자료이다.
- 약간의 변이는 있지만, 한 명의 화자가 약 1시간 정도 낭독한 셈 이다.
- 따라서 이 말뭉치는 개인의 음성 특성 뿐 만 아니라 세대 및 성별 에 따른 변이를 연구하는 데 유 용하게 사용될 수 있는 중요한 기초연구자료이다.

- 요람기
- 소나기
- 아름다움에 대하여
- 방망이 깎던 노인
- 문학의 세계와 삶의 세계
- 독서와 인생
- 수난 이대
- 메밀꽃 필 무렵
- 토끼와 자라
- 선녀와 나무꾼
- 호랑이와 곶감
- 해님 달님
- •

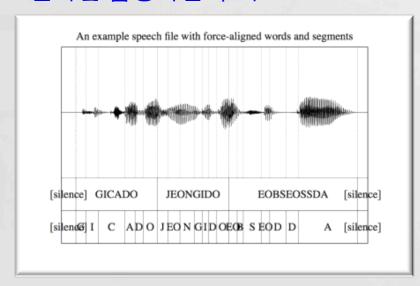
토끼와 자라



대용량 음성자료의 분석을 위한 자동음 성강제정렬장치의 필요성:

- 일반적으로 단어나 어절보다 더 작은 단위에서의 음성 및 음운 변이를 조사하고자 하면, 그러한 단위들을 음성자료에 직접 전사할 필요가 있다.
- 음소단위 전사(phone-level transcription)이음성학 연구에서 가장일반적으로 사용된다.

전사된 음성파일의 예

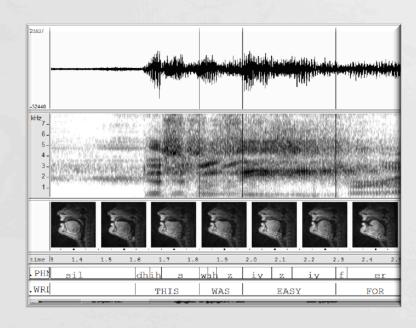


"기차도 전기도 없었다." (요람기)

대용량 음성자료의 분석을 위한 자동음성강제정렬장치의 필요성:

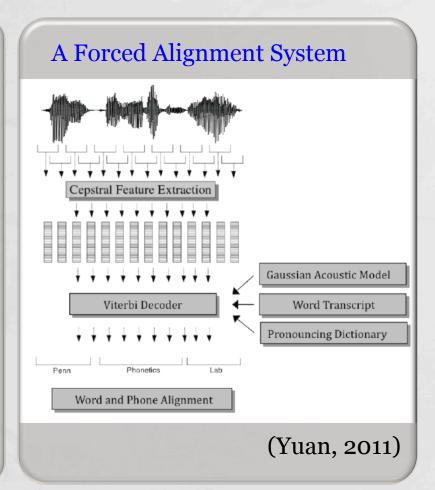
- 단어의 토큰수가 1,000개가 넘는 대용량의 음성말뭉치에 음소단위 전사(phone-level transcription) 가 되어 있는 경우는 매우 드물다.
- 음소단위의 전사는 waveform과 spectrogram을 이용해서 한다.
- 이러한 waveform과 spectrogram 을 해석할 수 있는 능력을 기르기 위해서는 많은 훈련과 연습이 필 요하다.
- 훈련을 받았다하더라도 음소단위 전사는 보통 수작업으로 하는데 이 수작업에는 너무 많은 시간이 든다.

(Cole & Hasegawa-Johnson, 2012).



강제음성정렬시스템 (Forced alignment system)

- 이러한 음소단위전사를 자동화하여 일차적으로 전사한 다음, 그 전 사된 내용 중에 사소한 오류가 있다면, 그 전사 내용을 연구목적에 맞 추어 수정하는 것이다.
- 단어나 어절에 해당하는 말뭉치자료를 자동으로 음소단위로 전사해주는 것이 강제음성 정렬장치(forced alignment system)이다.



한국어 강제음성정렬시스템 MFCC 벡터 Hidden Markov Models Acoustic Signal a 화자 8o 명 전사된 발화 말뭉치 [a] =(~100 시간) n ... 단어 ㄷ ㅏ ㄴ ㅓ DANEO DANEO 발음사전 n CEONGJU C EO NG J U g 강제음성정렬 입 력 청주 대학교 한글 2 유니코드 음운규칙 적용

CEONGJU

DAEHAGGYO

Unknown word

CEONGJU DAEHAKGGYO

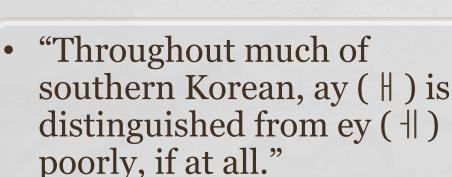
사례연구 1: 모음의 변이

- "서울말 낭독체 발화 말뭉치"에서 나타나는 8개 단모음의 전체적 윤곽을 세대별 및 성 별, 그리고 음운 및 통사 환경 속에서 체계 적으로 연구할 수 있다.
- 한 화자가 발화한 단모음의 갯수

| 단모음 | F | Н | -11 | 4 | - <u></u> | I | | Т |
|-------|-------|------|------|-------|-----------|-------|-------|------|
| Token | 2064 | 337 | 325 | 862 | 681 | 867 | 859 | 373 |
| % | 32.41 | 5.29 | 5.10 | 13.53 | 10.69 | 13.61 | 13.48 | 5.85 |



VS.



Martin 1992: 25

 "the mid front vowel merger displays a significant age variation but no significant variations in sex, class or style."

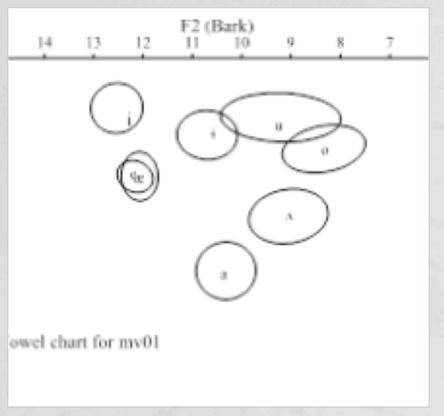
Hong 1988

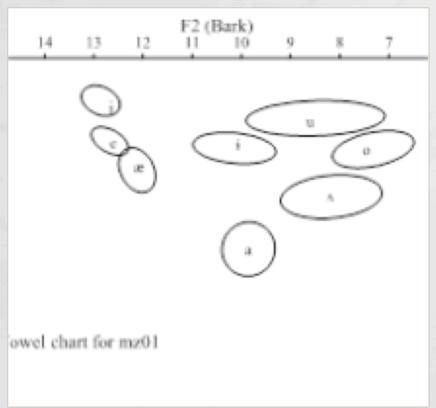


- *게가 "멍멍"하지?
- *언재 시작할까?

Silva & Jin, 2008

20대와 60대 남성화자의 모음비교

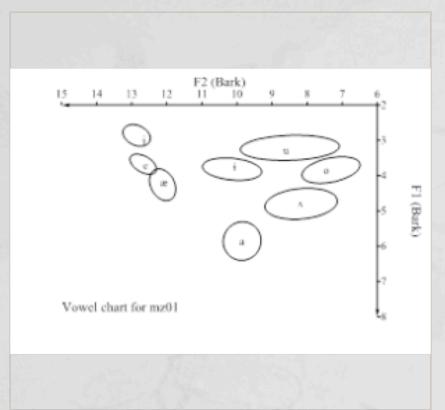




강세구 (accentual phrase) 첫 음절에서 추출 한 단모음으로 만들어진 20대 남성 화자의 모음도(vowel chart)

강세구 첫음절에서 추출한 단모음으로 만들어진 60대 남성 화자의 모음

60대 남성와 여성 화자의 모음비교



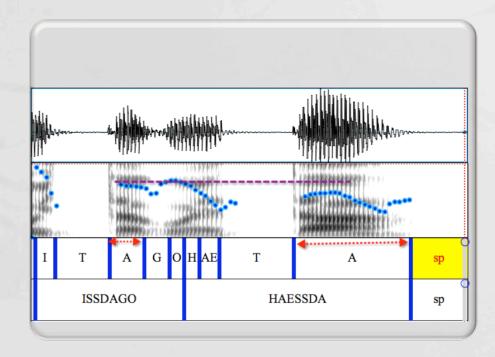
12 11 10 13 Vowel chart for fz06

강세구 (accentual phrase) 첫 음절에서 추출 한 단모음으로 만들어진 60대 남성 화자의 모음도(vowel chart)

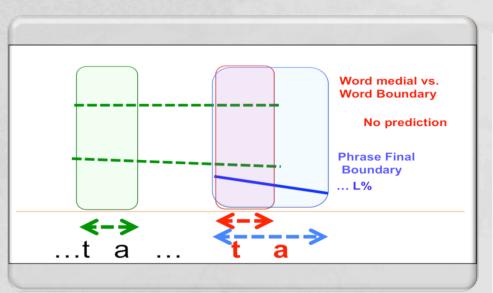
강세구 첫음절에서 추출한 단모음으로 만들어진 60대 여성 화자의 모음

사례 연구 2: 문장 위치에 따른 '다' 'ㅏ' 모음의 길이와 Fo

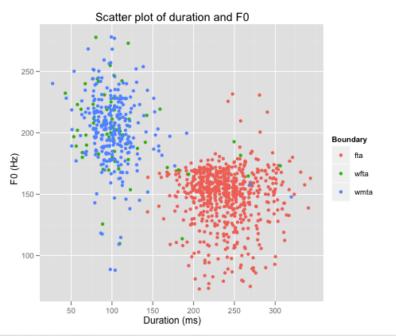
| Category | Context | 예 |
|----------------|-------------------|------|
| phrase initial | pause | 다니고 |
| word initial | no pause | 다친 |
| word medial | word- internal | 도다나기 |
| word final | no pause | 바다 |
| phrase final | pause | 볶는다 |



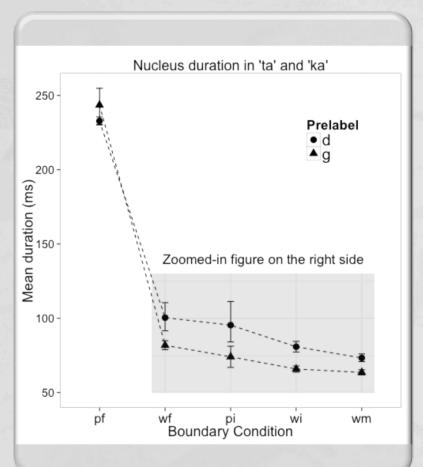
Boundary tone and final lengthening

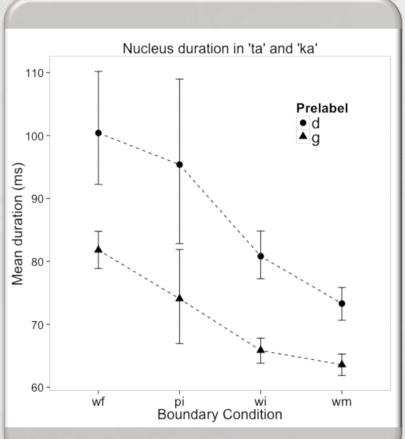


구말 -'다' vs.
어중 –'다' vs.
어말 –'다'



문맥에 따른 '다'와 '가'에서의 ' h'의 모음 길이





끝맺는 말...

- 한국어의 정보처리와 관련된 연구를 하는 사람들에게 한국어 처리는 넘어야만 하는 거대한 장벽이다.
- 영어와 달리, 한국어는 그 자료나 활용할 수 있는 도구들이 많지 않다.
- 구어 국어 연구를 위해 개발되는 자동정렬 장치(forced alignment system)는 음성언어 를 처리하는 유용한 도구로 사용될 수 있다.

끝맺는 말...

- 이러한 도구를 이용해서 변화의 과정에 있는 여러 음운 및 음성 현상들이 역사적 혹은 통시 적으로 어떤 변천을 겪어 현재에 이르게 되었 는지를 살펴 볼 수 있다.
 - 이는 우리 말소리의 변화를 문헌을 통해서 간접적으로 유추하는 방식이 아니라, 실제 음성 자료를 통하여 직접적으로 고찰할 수 있는 방법이다.
- 이러한 연구는 현대 국어 음운 및 음성의 본질을 보다 명확하게 규명할 수 있을 뿐만 아니라, 앞으로의 변화의 방향을 예측할 수 있게 해 준다는 점에서 의의가 있다.

끝맺는 말...

• 이러한 도구를 응용해서 사용하면, 한국어를 모국어로 쓰는 화자뿐만 아니라, 다문화가족 및 유학생들의 언어 학습의 연구에도 사용될 수 있다.

감사합니다!





Natural Sciences and Engineering Research Council of Canada Conseil de recherches en sciences naturelles et en génie du Canada



Social Sciences and Humanities Research Council of Canada Conseil de recherches en sciences humaines du Canada

