

# 2022 한국음성학회 가을 학술대회 발표 논문집



주제: **Speech Sciences with Big Data**

날짜: 2022년 11월 25일(금)

장소: 온라인 (주 송출: 서울대학교 신양학술관 세미나실)

주최: 사단법인 한국음성학회

주관: 사단법인 한국음성학회, 서울대학교 인문정보연구소

후원: 한국연구재단, (사)한국언어재활사협회,

(주)리드스피커코리아, (주)셀바스AI

**한국음성학회**  
**The Korean Society of Speech Sciences**

이 발표논문집은 2022년도 대한민국 교육부와 한국연구재단의  
학술대회지원사업의 지원을 받아 발간되었음(NRF-2022-2022S1A8A4A01077044)



## 모시는 글



한국음성학회는 11월 25일 온라인 웹컨퍼런스의 형식으로 가을학술대회를 개최합니다. 가을 학술대회는 우리 학회의 1년을 마무리 짓는 가장 큰 행사로서 그동안의 학술적 노력과 성과를 발표하고 의견을 나누는 자리입니다. 올해는 Speech Sciences with Big Data를 학술대회 주제로 하여 신진학자님의 발표를 포함하는 4편의 전문가 특강과 27편의 구두 발표, 그리고 7편의 포스터 발표를 준비했습니다. 초청 특강을 해 주실 네 분의 저명한 강연자들과 논문 발표를 신청해 주신 회원 여러분 모두에게 고마움을 표합니다. 어려운 가운데도 가을 학술대회 준비에 성심성의를 다하신 장태엽 학술위원장님을 비롯한 학술대회 준비위원회 여러분들과 학회 회장단 및 관계자 여러분들에게 감사의 말씀을 드립니다. 또한 서울대 신양학술관 세미나실을 섭외해주시고 학술대회의 준비에 많은 도움을 주신 정민화 수석부회장님과 서울대학교 언어학과 및 인문정보연구소에도 감사를 드립니다. 매년 도움을 아끼지 않는 한국연구재단, (사)한국언어재활사협회, (주)리드스피커코리아, (주)셀바스AI 등 후원해 주신 여러 기관에 감사의 말씀을 드립니다.

올해의 논문 발표 편수는 예년에 비해 적은 편입니다. 큰 학술대회와 겹친 이유도 있지만 코로나로 인해 제한되었던 활동반경과 데이터 수집의 어려움, 쌓여왔던 그간의 피로감 때문이 아닌가 생각하고 있습니다. 그러나 1년을 돌아보면, 코로나로 인해 이루어지지 못했던 작년까지와 달리, 여러 학술 사업과 워크숍은 성황리에 잘 진행된 한해였습니다. 학술대회와 총회를 마지막으로 학회의 1년 행사는 마무리됩니다. 코로나의 재감염과 겨울 독감이 확산될 조짐을 보이고 있습니다. 회원 여러분 건강에 유의하시고 남은 올해 잘 마무리하시기 바랍니다.

2022년 11월

한국음성학회 회장 성 철 재

## 2022 한국음성학회 가을 학술대회 조직위원회

학술대회장	회장	성철재(충남대)
조직위원회	위원장	정민화(수석부회장, 서울대)
	위원	홍기형(성신여대), 김재옥(강남대), 정현성(한국교원대), 박상희(대구사이버대), 윤태진(성신여대), 공은정(항공대), 장태엽(한국외대)
학술위원회	위원장	장태엽(한국외대)
	위원	김지연(우송대), 남호성(고려대), 박기영(ETRI), 박형민(서강대), 송윤경(동명대), 안현기(서울대), 오재혁(건국대)

## 2022 한국음성학회 가을 학술대회 일정표

11월 25일(금) 온라인 (zoom webinar, 주 송출: 서울대학교 신양학술관 세미나실)

시 간	발 표 및 내 용		
09:00~09:20	<b>개회식</b> (사회: 윤태진, 성신여대)		
09:20~10:00	<b>특강 I</b> (김형순, 부산대)		
10:00~10:40	<b>특강 II</b> (이주경, 서울시립대)		
10:40~10:50	휴식		
10:50~11:50	<b>구두 발표 I</b> (zoom 미팅)		
	음성학1 (3편)	음성공학1 (3편)	말장애1 (3편)
11:50~13:00	점심		
13:00~14:00	<b>포스터</b> (홈페이지에 포스터 게시 및 zoom 미팅)		
14:00~14:40	<b>특강 III</b> (변해원, 인제대)		
14:40~15:20	<b>특강 IV. 신진학자 초청강연</b> (송지은, KAIST)		
15:20~15:30	휴식		
15:30~16:30	<b>구두 발표 II</b> (zoom 미팅)		
	음성학2 (3편)	음성공학2 (3편)	음성학3 (3편)
16:30~16:40	휴식		
16:40~17:40	<b>구두 발표 III</b> (zoom 미팅)		
	음성학4 (3편)	음성공학3 (3편)	말장애2 (3편)
17:40~18:10	<b>총회</b> (사회: 윤태진, 성신여대)		
18:10~18:30	<b>퀴즈, 시상식 및 폐회식</b> (사회: 장태엽, 한국외대)		

## 2022 한국음성학회 가을 학술대회 세부 일정표

### 특강 I, II

좌장: 윤원희 (계명대)

시간	내용
09:20~10:00	"Voice privacy in the Big Data era" 김형순(부산대학교)
10:00~10:40	"Introducing an ultrasound analysis of sound data: case of ambisyllabicity in English" 이주경(서울시립대학교)

### 구두 발표 I

	좌장: 안현기(서울대)			
	시간	구분	제목	저자
음성학 및 음운론 1	10:50~ 11:10	PH01	한국어 자유 발화와 낭독 발화의 시간적 실현	이동은, 신지영(고려대학교)
	11:10~ 11:30	PH02	경상방언 내포문 의문사의 작용역 지각 단서	윤원희(계명대학교)
	11:30~ 11:50	PH03	Research on closure durations of Korean stops	Yungdo Yun (Dongguk University)
음성공학 1	좌장: 박정식(한국외대)			
	시간	구분	제목	저자
	10:50~ 11:10	SE01	효율적인 키워드 검출을 위한 시간 합성곱 신경망의 재매개변수화	박은익, 안대현, 김형준((주)스퀴즈비츠)
	11:10~ 11:30	SE02	종단간 음성향상을 위한 멀티 레이트 샘플링 연동 심층신경망	김선만, 이광훈(한국광기술원)
11:30~ 11:50	SE03	강인음성인식을 위해 음성/비음성 잡음으로 학습된 시청각 음성향상 딥러닝 모델 성능에 대한 교차 평가 분석	정원, 이승현, 박형민(서강대학교)	
말장애 및 음성의학 1	좌장: 송윤경(동명대)			
	시간	구분	제목	저자
	10:50~ 11:10	SD01	연령, 성별, 의문문 유형에 따른 의문사 포함 예-아니오 의문문과 의문사 의문문의 운율 특성 비교	박미진, 성철재(충남대학교)
	11:10~ 11:30	SD02	급성 COVID-19 감염의 다면적 음성 특성 추적 사례연구	이승진(한림대학교)
11:30~ 11:50	SD03	나무젓가락을 이용한 발음 연습의 음성학적 효과	강보란, 성철재(충남대학교)	

**포스터(홈페이지 내 '학술대회'에 포스터 게시 & zoom을 이용한 실시간 미팅)**  
**좌장: 이옥분(대구사이버대), 박정식(한국외대)**

시간	구분	제목	저자
13:00 ~ 14:00	P1	딥러닝 모델 전이학습 기반 한국어 방언 음성합성(Text-to-Speech)	강진아(한국전자통신연구원), 오유리(쥬스피치랩스)
	P2	매체(media) 차이가 학령기 아동의 운율과장(prosody exaggeration)에 미치는 영향	박은영, 성철재(충남대학교)
	P3	ADHD증상 및 불안, 우울을 보이는 성인의 실행기능 훈련 효과	김보선, 김성태(동신대학교)
	P4	말실행증 성인 대상 반복훈련과 말속도-리듬통제훈련 효과에 대한 사례 연구	염현자, 한지용, 박진(가톨릭관동대학교)
	P5	마스크 종류에 따른 아동의 말지각력 비교	장은영, 김지연(우송대학교)
	P6	언어발달지연 영유아에게 미치는 Floortime 중재 효과	양영진, 김성태(동신대학교)
	P7	딥러닝 기반 입술 움직임 분석 연구	강영애(충남대학교병원), 김민혁, 한연규(충남대학교), 장재원, 구본석(충남대학교병원), 이동헌(충남대학교)

**특강 III**

**좌장: 박상희(대구사이버대)**

시간	내용
14:00~14:40	"Voice pathology detection using deep learning, machine learning, and hybrid modeling" 변해원(인제대학교)

**특강 IV. 신진학자 초청강연**

**좌장: 정현성(한국교원대)**

시간	내용
14:40~15:20	"Using neurophysiological data to understand speech-recognition difficulties" 송지은(한국과학기술원)

## 구두발표 II

<b>좌장: 김지은(가톨릭관동대)</b>				
음성학 및 음운론 2	시간	구분	제목	저자
	15:30~ 15:50	PH04	A study on the predictability of acoustic power distribution of English speech for the English academic achievement in a science academy	Soon Park, Hyunkee Ahn (Seoul National University)
	15:50~ 16:10	PH05	발화 속도 통제 상태에서의 영어 능숙도 평가: 한국인 아동 영어 학습자 발화를 중심으로	최나래, 장태엽(한국외국어대학교)
	16:10~ 16:30	PH06	유아 파닉스 프로그램의 효과성 연구	정현성(한국교원대학교)
<b>좌장: 유하진(서울시립대)</b>				
음성공학 2	시간	구분	제목	저자
	15:30~ 15:50	SE04	Speech self-supervised learning을 활용한 단일 화자 한국어 합성 음성 탐지	엄영식, 김회린(한국과학기술원)
	15:50~ 16:10	SE05	한국어 음성 전사를 위한 wav2vec2.0 훈련 단위	이선우, 김선희, 정민화(서울대학교)
	16:10~ 16:30	SE06	화자 확인을 위한 음성 세그먼트 길이 임베딩	배재현, 박순찬(부산대학교), 김경화(대검찰청), 김형순(부산대학교)
<b>좌장: 이주경(서울시립대)</b>				
음성학 및 음운론 3	시간	구분	제목	저자
	15:30~ 15:50	PH07	Temporal variation in intergestural timing of CVC gestures in Korean homophones	Suyeon Im (Hanyang University), Sahyang Kim (Hongik University), Taehong Cho (Hanyang University)
	15:50~ 16:10	PH08	Prosodic marking of second occurrence focus in Korean	Seong Eun Park (Chungbuk National University), Gayeon Son (KwangWoon University)
	16:10~ 16:30	PH09	Corpus-based study of children's accentual phrases	Tae-Jin Yoon (Sungshin Women's University)



### 구두발표 III

<b>좌장: 성은경(사이버한국외대)</b>				
음성학 및 음운론 4	시간	구분	제목	저자
	16:40~ 17:00	PH10	오노마토펬이아의 음상징	권연주(고베대학교)
	17:00~ 17:20	PH11	한국인 학습자의 프랑스어 연음 실현 양상	김민정, 김선희(서울대학교)
	17:20~ 17:40	PH12	프랑스 비강모음의 포먼트 산출을 위한 안정 구간 연구	박혜숙, 김선희(서울대학교)
<b>좌장: 박형민(서강대)</b>				
음성공학 3	시간	구분	제목	저자
	16:40~ 17:00	SE07	HuBERT 기반 한국인 아동 발화 영어 음소 인식 연구	경나, 박지은, 김규진, 박정식(한국외국어대학교), 강병옥(한국전자통신연구원)
	17:00~ 17:20	SE08	외국인의 한국어 발음 오류 패턴 태깅을 위한 강제정렬 모델 활용	이주영, 황효성, 류형신, 이호영, 김선희, 정민화(서울대학교)
	17:20~ 17:40	SE09	종단간 음성인식 모델의 다층 특징을 활용한 화자인식	박순찬, 김형순(부산대학교)
<b>좌장: 김지연(우송대)</b>				
말장애 및 음성의학 2	시간	구분	제목	저자
	16:40~ 17:00	SD04	이란성 쌍둥이 구순구개열 아동과 일반 아동 파열음의 음향음성학적 특성 비교	전선영, 박지연, 성철재(충남대학교)
	17:00~ 17:20	SD05	말더듬 성인의 군대 적응 정도에 대한 군 지휘관의 인식 양상 및 관련 요인 분석	박혜린(충남대학교), 고복순, 박진(가톨릭관동대학교)
	17:20~ 17:40	SD06	물 저항 발생 시 깊이에 따른 성대접촉률과 설골하근 수축량의 변화	차준서, 김채현, 최성희(대구가톨릭대학교)

# 차 례

## 특강 I, II

Voice privacy in the Big Data era

김형순(부산대학교) / 3

Introducing an ultrasound analysis of sound data: case of ambisyllabicity in English

이주경(서울시립대학교) / 7

## 구두발표 I

[음성학 및 음운론 1]

PH01 한국어 자유 발화와 낭독 발화의 시간적 실현

이동은, 신지영(고려대학교) / 11

PH02 경상방언 내포문 의문사의 작용역 지각 단서

윤원희(계명대학교) / 12

PH03 Research on closure durations of Korean stops

Yungdo Yun (Dongguk University) / 13

[음성공학 1]

SE01 효율적인 키워드 검출을 위한 시간 합성곱 신경망의 재매개변수화

박은익, 안대현, 김형준((주)스퀴즈비츠) / 17

SE02 종단간 음성향상을 위한 멀티 레이트 샘플링 연동 심층신경망

김선만, 이광훈(한국광기술원) / 18

SE03 강인음성인식을 위해 음성/비음성 잡음으로 학습된 시청각 음성향상 딥러닝 모델 성능에 대한 교차평가 분석

정원, 이승현, 박형민(서강대학교) / 19

[말장애 및 음성의학 1]

SD01 연령, 성별, 의문문 유형에 따른 의문사 포함 예-아니오 의문문과 의문사 의문문의 운율 특성 비교

박미진, 성철재(충남대학교) / 23

SD02 급성 COVID-19 감염의 다면적 음성 특성 추적 사례연구

이승진(한림대학교) / 24

SD03 나무젓가락을 이용한 발음 연습의 음성학적 효과

강보란, 성철재(충남대학교) / 25

## 포스터발표

- P1 딥러닝 모델 전이학습 기반 한국어 방언 음성합성(Text-to-Speech)  
강진아(한국전자통신연구원), 오유리(썬스피치랩스) / 29
- P2 매체(media) 차이가 학령기 아동의 운율과장(prosody exaggeration)에 미치는 영향  
박은영, 성철재(충남대학교) / 30
- P3 ADHD증상 및 불안, 우울을 보이는 성인의 실행기능 훈련 효과  
김보선, 김성태(동신대학교) / 31
- P4 말실행증 성인 대상 반복훈련과 말속도-리듬 통제 훈련 효과에 대한 사례연구  
염현자, 한지용, 박진(가톨릭관동대학교) / 32
- P5 마스크 종류에 따른 아동의 말지각력 비교  
장은영, 김지연(우송대학교) / 34
- P6 언어발달지연 영유아에게 미치는 Floortime 중재 효과  
양영진, 김성태(동신대학교) / 35
- P7 딥러닝 기반 입술 움직임 분석 연구  
강영애(충남대학교병원), 김민혁, 한연규(충남대학교), 장재원, 구분석(충남대학교병원), 이동현(충남대학교) / 36

## 특강 III

- Voice pathology detection using deep learning, machine learning, and hybrid modeling  
변해원(인제대학교) / 39

## 특강 IV. 신진학자 초청강연

- Using neurophysiological data to understand speech-recognition difficulties  
송지은(한국과학기술원) / 63

## 구두발표 II

### [음성학 및 음운론 2]

- PH04 A study on the predictability of acoustic power distribution of English speech for the English academic achievement in a science academy  
Soon Park, Hyunkee Ahn (Seoul National University) / 67
- PH05 발화 속도 통제 상태에서의 영어 능숙도 평가: 한국인 아동 영어 학습자 발화를 중심으로  
최나래, 장태엽(한국외국어대학교) / 68

PH06 유아 파닉스 프로그램의 효과성 연구

정현성(한국교원대학교) / 70

### [음성공학 2]

SE04 Speech self-supervised learning을 활용한 단일 화자 한국어 합성음성 탐지

엄영식, 김회린(한국과학기술원) / 73

SE05 한국어 음성 전사를 위한 wav2vec2.0 훈련 단위

이선우, 김선희, 정민화(서울대학교) / 74

SE06 화자확인을 위한 음성 세그먼트 길이 임베딩

배재현, 박순찬(부산대학교), 김경화(대검찰청), 김형순(부산대학교) / 75

### [음성학 및 음운론 3]

PH07 Temporal variation in intergestural timing of CVC gestures in Korean homophones

Suyeon Im (Hanyang University), Sahyang Kim (Hongik University), Taehong Cho (Hanyang University) / 79

PH08 Prosodic marking of second occurrence focus in Korean

Seong Eun Park (Chungbuk National University), Gayeon Son (KwangWoon University) / 80

PH09 Corpus-based study of children's accentual phrases

Tae-Jin Yoon (Sungshin Women's University) / 81

## 구두발표 III

### [음성학 및 음운론 4]

PH10 오노마토펬이아의 음상징

권연주(고베대학교) / 85

PH11 한국인 학습자의 프랑스어 연음 실현 양상

김민정, 김선희(서울대학교) / 86

PH12 프랑스 비강모음의 포먼트 산출을 위한 안정 구간 연구

박혜숙, 김선희(서울대학교) / 88

### [음성공학 3]

SE07 HuBERT 기반 한국인 아동 발화 영어 음소 인식 연구

경나, 박지은, 김규진, 박정식(한국외국어대학교), 강병욱(한국전자통신연구원) / 91

SE08 외국인의 한국어 발음 오류 패턴 태깅을 위한 강제정렬 모델 활용

이주영, 황효성, 류형신, 이호영, 김선희, 정민화(서울대학교) / 92

SE09 종단간 음성인식 모델의 다층 특징을 활용한 화자인식

박순찬, 김형순(부산대학교) / 94

[말장애 및 음성의학 2]

SD04 이란성 쌍둥이 구순구개열 아동과 일반 아동 파열음의 음향음성학적 특성 비교

전선영, 박지연, 성철재(충남대학교) / 97

SD05 말더듬 성인의 군대 적응 정도에 대한 군지휘관의 인식 양상 및 관련 요인 분석

박혜린(충남대학교), 고복순, 박진(가톨릭관동대학교) / 99

SD06 물 저항 발성 시 깊이에 따른 성대접촉률과 설골하근 수축량의 변화

차준서, 김채현, 최성희(대구가톨릭대학교) / 100

# 특강 I

좌장: 윤원희(계명대)

Voice privacy in the Big Data era  
(김형순, 부산대학교)



## 빅데이터 시대의 음성 프라이버시

김형순  
부산대학교 전자공학과

### Voice Privacy in the Big Data Era

Hyung Soon Kim  
Department of Electronics Engineering, Pusan National University  
kimhs@pusan.ac.kr

음성인식, 음성합성, 화자인식을 비롯한 음성공학 기술들은 50여 년 이상 지속적으로 연구되어 왔으나, 최근 10여 년 동안 딥러닝 기술의 도입을 통해 비약적인 성능 향상을 이루었다. 그 배경에는 딥러닝 학습 연산을 가능케 한 GPU 기술도 있지만, 무엇보다도 음성 빅데이터의 지대한 역할이 자리 잡고 있다. 스마트폰과 AI 스피커 등 모바일 단말기를 통한 음성인식 응용 서비스의 보급 증대로 인해 음성 빅데이터의 수집 규모는 더욱 커지고 있으며, 자기도 학습(self-supervised learning) 기술의 발전과 더불어 음성기술 성능 개선의 지속적인 선순환에 활용되고 있다.

음성 데이터에는 화자가 말하고자 하는 메시지 내용과 함께 화자 정보, 감정 상태, 건강 상태 등 민감한 개인 정보들이 포함되어 있다. (물론 메시지 내에도 중요한 개인 정보들이 많이 포함되어 있다.) 음성 빅데이터에 의해 강력해진 음성공학 기술은 음성 데이터로부터 이들 개인 정보들을 불법적으로 탈취하는데 악용될 수 있다. 따라서 음성 빅데이터를 이용한 딥러닝을 통해 음성공학 기술의 성능 향상은 계속 추진하되, 음성 빅데이터로부터 개인 정보의 탈취를 막는 방안의 마련이 시급하다.

본 특강에서는 음성 빅데이터로부터 개인 정보를 최대한 보호하면서 음성공학 분야의 활용성은 유지하는 방안에 대한 최근의 기술개발 동향을 소개한다. 그중에서도 음성 익명화(anonymization) 기술에 초점을 맞추고, 이와 관련한 최근 행사인 VoicePrivacy 2022 Challenge 결과도 간단히 소개한다. 성능이 우수한 음성 익명화 기술은 대부분 딥러닝 기반이지만, 비공학 음성 전공자들도 쉽게 이해할 수 있는 음성 익명화 및 가명화(pseudonymization) 방법론에 대한 소개도 함께 다루도록 한다.





## 특강 II

좌장: 윤원희(계명대)

Introducing an ultrasound analysis of sound data:  
case of ambisyllabicity in English  
(이주경, 서울시립대학교)



## Introducing an ultrasound analysis of sound data: case of ambisyllabicity in English

이 주 경  
서울시립대학교 영어영문학과

Joo-Kyeong Lee  
Department of English Language and Literature, University of Seoul  
jookyeong@uos.ac.kr

This work introduces an ultrasound analysis of speech data, focusing on the phonetic, more specifically, articulatory substance of ambisyllabicity of English /l/. It presents an overall sketch of an ultrasound experiment from tongue video recordings to interpreting the ultrasound images and ultimately to conducting a quantitative/statistical analysis of ultrasound data. Articulatory studies on speech sounds, for example, EMA, EPG, and MRI, have been consistently carried on, but they are not sufficiently accessible due to subjects' susceptibility problem, a limited amount of data or high cost. Ultrasound studies, on the other hand, are more likely available with relatively little difficulty.

An intervocalic consonant is ambisyllabic in English, where it is aligned to both the onset of the following 'unstressed' syllable and the coda of the preceding 'stressed' syllable (e.g., *color*). That is, the consonant, which is originally syllabified to the onset of the following syllable according to the Onset Maximization Rule, is simultaneously connected to the coda of the preceding syllable to make it heavy and satisfy the requirement that a stressed syllable should be heavy. In the previous ultrasound studies of intervocalic /l/, its degree of darkening was found to be affected by prosodic boundaries within each category of [l] in various dialects of British English (Turton, 2017) and in American English (Lee-Kim, Davidson and Hwang, 2013). Some questions are raised up at this point: Is the intervocalic and ambisyllabic consonant phonetically substantial? Is it phonetically different from an intervocalic consonant with no ambisyllabicity like /l/ in *collar* or *collect*? Does it contain both onset and coda characteristics (clear [l] and dark-[ɫ]) as it is aligned to both onset and coda?

In this study, ultrasound tongue imaging data were used to see if an ambisyllabic /l/ is articulatorily different from non-ambisyllabic /l/s in intervocalic position (non-ambisyllabic /l/ with the preceding syllable stressed, e.g. *collar*, and non-ambisyllabic /l/ with the following syllable stressed, e.g., *collect*). As reported in Sproat & Fujimura(1993), not only dark-[ɫ] but also light-[l], to some extent, can be articulated with darkening (tongue body lowering and tongue dorsum retraction). Both vertical and horizontal degrees of the tongue body and the tongue dorsum were compared among three intervocalic /l/s with two references of a word-initial onset and a word-final coda. The comparisons were also carried out at three

temporal positions: initial, midpoint and final positions of /l/. Six native speakers of American English were recorded for the /l/ sounds of the 5 different categories. Five actual words were collected for each category and 25 words were repeated three times in the sentence frame 'This is \_\_\_\_\_ for me.' In total, 225 (25 words \* 3 repetitions \* 3 positions) ultrasound images of /l/ were analyzed for each participant and SS-ANOVA was used for statistical testing. Results showed that the tongue contours of three intervocalic /l/s commonly change from a dark-[ɫ] in initial position to a clear-[l] in final position. That is, the tongue body lowered and the dorsum was retracted like the dark-[ɫ] immediately after the articulatory gestures of the preceding vowel is over, but the tongue body was raised and the dorsum was fronted like the light-[l] immediately before the following vowel as they are all onset. What should be noticed was that the tongue movement toward the light-[l] was observed as early as in the midpoint for the ambisyllabic /l/ and the non-ambisyllabic /l/ with the preceding syllable stressed. On the other hand, such articulatory gestures were delayed for the non-ambisyllabic /l/ with the following syllable stressed as either tongue body lowering or tongue dorsum retraction or both of the gestures were observed in the midpoint. The early anticipation of the onset-like gestures might be attributed to the fact that the preceding syllable was stressed. What should be noticed from the ultrasound imaging results was that (1) the ambisyllabic /l/ did not show a significant difference in tongue body and dorsum gestures from the other two intervocalic /l/s temporally at initial and final positions, and (2) the ambisyllabic /l/ was similar to the intervocalic /l/ with the preceding syllable stressed in such a way that the tongue body and/or the tongue dorsum already anticipated in the postures of the onset variant at midpoint though the degrees were varied. This suggests that the ambisyllabicity of /l/ was articulatorily substantial as its tongue body and dorsum gestures were coda-like at the postvocalic boundary and onset-like at the prevocalic boundary, but that this was not confined to the ambisyllabic /l/ alone. The other two intervocalic /l/s also showed the same gestural patterning even though they are phonologically an onset; therefore, intervocalic consonants, at least /l/, may have ambivalent characteristics of both coda and onset regardless of their phonological structure associated with ambisyllabicity. Therefore, the theoretical/phonological proposition of ambisyllabicity did not seem to be empirically supported; it may be a peripheral solution for the syllable weight requirement concerning stress.

### References

- Lee-Kim, SangIm, Davidson, Lisa and Sangjin Hwang (2013) Morphological effects of the darkness of English intervocalic /l/. *Laboratory Phonology*, 4(2), 475-511.
- Sproat, Richard & Fujimura, Osamu (1993) Allophonic variation in English /l/ and its implications for phonetic implementation, *Journal of Phonetics*, 21, 291-311.
- Turton, Danielle (2017) Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English, *Laboratory Phonology*, 8(1): 1-31.

[Acknowledgment: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2021S1A5A2A01070152)].

구두 발표 I  
음성학 및 음운론 1

좌장: 안현기(서울대)



## 한국어 자유 발화와 낭독 발화의 시간적 실현

이동은, 신지영  
고려대학교 국어국문학과

### The temporal realization of spontaneous and read speech in Korean

Dongeun Lee, Jiyoung Shin  
Department of Korean Language and Literature, Korea University  
ap02101@korea.ac.kr, shinjy@korea.ac.kr

이 연구는 자유 발화와 낭독 발화의 휴지 구간과 조음 구간의 시간적 실현 양상을 살핌으로써 대본의 여부에 따라 발화의 시간적 구조가 어떤 차이를 갖는지를 탐구하는 데 목적이 있다. 화자는 한 발화 안에서 휴지와 조음 구간의 길이, 조음 속도 등을 발화 내 위치에 따라 달리 실현한다. 본 연구는 시간의 관점에서 한국어 화자가 발화 과제에 따라 조음 구간과 휴지 구간을 발화 내 위치에 따라 어떻게 구성하는지 살펴본다.

이를 위해 본 연구는 30명의 화자를 대상으로 두 가지 과제를 수행하게 하였다. 첫 번째 과제는 일상 주제에 대한 독백 발화 과제였고, 두 번째 과제는 다른 사람의 독백 발화 전사본을 낭독하는 과제였다. 수집한 발화를 대상으로 과제별 휴지 구간과 조음 구간의 발화 내 위치에 따른 길이와 각 조음 구간의 조음 속도를 측정하였다.

그 결과, 자유 발화에서는 발화 말 위치로 갈수록 휴지 구간이 짧아지고 조음 구간이 길어지는 경향이 관찰되었지만, 낭독 발화에서는 발화 내 위치에 따른 길이의 변화가 관찰되지 않았다. 한편, 조음 속도는 두 과제 모두 발화 말에서 증가하는 경향이 관찰되었다. 하지만 그 증가 정도가 자유 발화에서 더 큰 것으로 확인되었다.

과제에 따라 휴지 구간과 조음 구간의 시간적 구조가 달리 나타난 이유를 발화 계획의 차이로 설명해 볼 수 있다. 후행할 발화를 실시간 직접 계획해 가며 수행해야 하는 자유 발화의 경우는 발화 초에서 더 많은 시간이 필요한 만큼, 발화의 시작에서는 긴 휴지와 짧은 조음 구간을, 발화 말에서는 짧은 휴지와 긴 조음 구간을 가지며 조음의 수행 속도가 빨라진다. 반면에 직접 발화 계획을 할 필요가 없는 낭독 발화는 제시된 문장을 적절한 위치에서 끊고 휴지를 넣으며 문장을 읽어나가는 만큼, 발화 내 위치에 따른 시간적 차이가 두드러지지 않는다. 다만, 문장의 끝이 구두점에 의해 표시되는 만큼, 문장의 끝에서 발화 속도는 조금 빨라지는 것이 아닌가 해석해 볼 수 있다.



## 경상 방언 내포문 의문사의 작용역 지각 단서

윤 원 희  
계명대학교 영어영문학과

### Perceptual Cues of Wh-Scopes in Gyeongsang Korean

Weonhee Yun  
Dept. of English Language and Literature, Keimyung University  
whyun@kmu.ac.kr

의문사가 내포문에 위치할 경우 의문사의 작용역은 내포문에 머물러 판정의문문으로 해석될 수 있고, 또한 내포절의 보문소를 넘어 모문 작용역을 갖게 되면 설명의문문으로 해석될 수 있다. 경상 방언에서 모문 작용역을 갖게 될 경우 종결어미는 ‘-노(고)’로 나타나며, 내포문 작용역일 경우 ‘-나’로 나타나 형태적으로 구분되는 특징이 있다. 그러나 이러한 경우에도 모문 작용역을 갖기 위해서는 의문사 억양(wh-intonation)의 형태로 발화되어야 한다. 서울말의 경우 의문문 종결어미는 ‘-니’의 형태로 의문사 억양만이 의문사의 작용역을 알려주는 유일한 단서인 반면, 경상 방언에서는 형태적으로 다른 종결어미로 인해 의문사 억양 외에도 작용역을 알려주는 단서가 존재한다고 할 수 있다. 그렇다면 ‘-노(고)’ 종결어미를 갖는 내포문 의문사 의문문을 경상 방언 화자의 발화 자극으로 청취 실험할 경우 모문 작용역 판단이 우세하게 나타날 수 있다는 가설을 세울 수 있다. 이를 검증하기 위해 40명의 경상 방언 화자가 두 번 발화한 80개의 모문 작용역 발화를 대상으로 모문 작용역으로 이해되는지를 묻는 청취실험을 하고 각 자극에 대해 모문 작용역 응답이 얼마나 많이 나타났는지 살펴보았다. 이와 함께 다중 회귀 분석을 통하여 모문 작용역 응답 수를 설명할 수 있는 자극 문장의 운율 음향 단서를 추정하였다. 실험 결과 종결어미가 모두 ‘-노(고)’로 나타남에도 불구하고 여전히 내포문 작용역으로 이해되는 발화 자극들이 있었으며 모문 작용역 지각에 가장 큰 기여를 한 음향 단서는 내포 동사의 돌들림 지수로 돌들림 지수가 크면 내포문 작용역으로, 돌들림 지수가 작으면 모문 작용역으로 지각하는 응답이 증가함을 알 수 있었다.

## Research on Closure Durations of Korean Stops

Yungdo Yun

Dharma College, Dongguk University  
yungdoyun@dongguk.edu

This study investigates closure durations of Korean stops. Twenty two Seoul Korean speakers produced nonsense Korean words containing Korean stops found at the word-initial position, word-final position, and between vowels. Based on laryngeal contrasts, closure durations were measured in the order of fortis > aspirated > lenis for both word-initial stops and stops between vowels, but fortis > aspirated = lenis for word-final stops. Based on places of articulation, closure durations were in the order of bilabial > alveolar = velar stops for both word-initial stops and stops between vowels, but bilabial = alveolar = velar for word-final stops. It seems that word-final stops' closure durations were a little neutralized. I assume that this may be related to so-called coda neutralization in Korean phonology.



# 구두 발표 I

## 음성공학 1

좌장: 박정식(한국외대)



## 효율적인 키워드 검출을 위한 시간 합성곱 신경망의 재매개변수화

박 은 익, 안 대 현, 김 형 준  
(주) 스퀴즈비츠

### Reparameterization Temporal Convolution Network for Efficient Keyword Spotting

Eunik Park, Daehyun Ahn, Hyungjun Kim  
SqueezeBits Inc.

eunik.park@squeezebits.com, daehyun.ahn@squeezebits.com,  
hyungjun.kim@squeezebits.com

키워드 검출(Keyword spotting, KWS)은 연속적인 음성 데이터에서 특정 키워드를 검출하는 작업으로 스마트 기기를 깨우거나 사용자와 스마트기기 간 상호작용하는 방법으로 널리 사용되고 있다. 키워드 검출 기능의 특성상 리소스가 제한된 기기에서 빠르게 구동할 수 있어야 하므로 최근 시간 합성곱 신경망(Temporal Convolution Network) 기반의 모델이 연구 및 사용되고 있다. 시간 합성곱 신경망은 시간 축 방향으로 1차원 합성곱 연산을 수행하므로 다양한 주파수 영역의 정보를 동시에 활용할 수 있어, 보다 높은 인식률을 달성할 수 있다. 하지만 기존의 시간 합성곱 신경망은 9x1 크기로 고정된 합성곱 가중치를 사용하여 불필요하게 많은 가중치를 사용하는 부분이 있다. 본 연구에서는 합성곱 신경망의 재매개변수화 기법을 통해 최적의 응답속도(latency)-인식률(accuracy) trade-off를 갖는 키워드 검출 모델, RepTCNet을 제안한다. RepTCNet 모델은 학습 시에는 복잡한 multi-branch 구조를 갖지만, 추론 시에는 single-branch 구조를 가질 수 있어 효율적인 연산이 가능하다. 또한, 학습 단계에서 9x1, 7x1, 3x1 등의 다양한 크기의 합성곱 가중치를 사용할 수 있어 최적의 합성곱 가중치 크기를 찾는 데에도 유용하다. 12개의 class를 갖는 키워드 검출 데이터셋에 학습한 결과 기존 시간 합성곱 신경망보다 2.1배 빠르면서 0.2%만큼 인식률이 높은 모델을 개발하였다.

\* 본 연구는 2022년도 과학기술정보통신부의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 연구임 (A1504-22-1010)

# 종단간 음성향상을 위한 멀티 레이트 샘플링 연동 심층신경망

김 선 만, 이 광 훈  
한국광기술원 공간광정보연구센터

## Deep neural networks combined with multi-rate sampling for end-to-end speech enhancement

Seon Man Kim, Kwanghoon Lee  
Spatial Optical Information Research Center, Korea Photonics Technology Institute  
kobem30002@gmail.com, geniuspb@kopti.re.kr

지난 수십 년에 걸쳐 음성향상 시스템은 통신, 자동 음성 인식 및 보청기 장치 등에 사용될 목적으로 꾸준히 많은 연구가 진행되었다. 음성향상의 목표는 음성의 품질이나 명료도를 유지하면서 부가적인 배경 잡음 성분을 억제하는 것이다. 마이크로폰을 하나만 사용하는 단일 채널 음성향상 방법에서는 워너필터와 같은 통계적 모델 혹은 비음수 행렬 인수분해와 같은 빅데이터를 활용하는 기계학습 방식이 주로 사용되었다. 특히, 지난 10년 동안에는 심층신경망을 기반으로 한 음성향상 방법이 주로 연구되어왔으며, 이 심층신경망은 기존 방법들 대비 음성향상 성능을 크게 향상시킬수 있음이 보편적으로 확인되고 있다. 음성향상 심층신경망 접근 방식 중 Wave-U-Net과 같은 순수 웨이브 신호 데이터를 기반으로 한 종단간 전략이 음성 인식 및 음성향상과 같은 음성 처리 분야에서 최근 주목받고 있다.

본 연구에서는 Wave-U-Net 종단간 심층신경망 구조에서 웨이브 신호의 멀티 레이트 다운-업 샘플링이 극심한 잡음 조건에서 음성 파형 신호를 복원해 내기 위한 주요 인자임을 보인다. 이를 기반으로 한 제안 방법의 음성향상 효과를 평가하기 위해 단기 객관적 어음인지도 STOI를 측정하였다. 객관적인 평가 결과 제안된 Wave-U-Net을 사용하는 음성향상 방법의 성능은 -10dB~15dB 신호-대-잡음 비 잡음 조건에서 기존 심층신경망 방법들 대비 높은 성능을 보여주었다.

**Table 1.** Comparison of average improved STOI scores (%) for SNRs of -10, -5, 0, 5, 10, and 15 dB for the different methods under various noise conditions

Methods	Noise					
	babble	F16	factory	pink	car	white
U-Net [4]	3.80	10.48	6.60	8.37	3.10	6.53
SEGAN [5]	3.57	5.27	4.22	4.73	3.35	0.25
Wave U-Net [2]	12.57	13.47	13.45	13.40	5.32	11.32
Proposed Method	<b>14.48</b>	<b>15.08</b>	<b>14.53</b>	<b>13.95</b>	<b>6.45</b>	<b>12.27</b>

### 감사의 글

본 연구는 과학기술정보통신부의 재원으로 한국연구재단 (NRF-2022R1A2C2010614)의 지원을 받아 수행되었음

## 강인음성인식을 위해 음성/비음성 잡음으로 학습된 시청각 음성향상 딥러닝 모델 성능에 대한 교차 평가 분석

정 원, 이 승 현, 박 형 민  
서강대학교 전자공학과

### A cross-evaluation of a deep-learning-based audio-visual speech enhancement model trained by speech/non-speech noise for robust speech recognition

Won Jeong, Seung-Hyeon Lee and Hyung-Min Park  
Dept. of Electronic Engineering, Sogang University  
wjeong@sogang.ac.kr, ee\_seung@u.sogang.ac.kr, hpark@sogang.ac.kr

본 연구는 특정 잡음에 대해서 학습된 시청각 음성향상 딥러닝 모델을 통해 얻어진 데이터의 음성인식 결과를 비교한다. 최근 딥러닝 기술의 발전에 힘입어 연속음성인식 성능이 크게 개선되어 왔으며, 실세계 잡음 환경에서의 강인성 확보에 대한 연구도 활발히 이루어지고 있다. 그러나 음성 잡음과 비음성 잡음의 정상성 차이 등으로 인해 음성 잡음에서는 상대적으로 낮은 성능을 보인다. 근래의 몇몇 연구는 기존의 인공 신경망 모델의 입력에 영상 데이터를 추가로 사용하여 이러한 한계를 극복하기 시작했다. 본 연구에서는 시청각 음성향상 모델이 강인음성인식을 위한 전처리 기법으로 활용 가능하다는 점에 주목하여, 특정 잡음에 대해 학습한 후 이를 음성 잡음과 다른 비음성 잡음들에 의해 왜곡된 음성신호를 향상하여 음성인식 성능을 평가해본다. 본 연구의 결과를 통해 특정 상황에서의 음성 향상이 필요한 다양한 환경에서 인공 신경망 모델을 선택할 때 참고할 수 있도록 한다.

사사: 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00621, 대화 기반 설명가능성을 멀티모달로 제공하는 인공지능 기술 개발 및 No.1711152445, 인명 구조용 드론을 위한 영상/음성 인지 기술 고도화)





구두 발표 I  
말장애 및 음성의학 1

좌장: 송윤경(동명대)



## 연령, 성별, 의문문 유형에 따른 의문사 포함 예-아니오 의문문과 의문사 의문문의 운율 특성 비교

박 미 진, 성 철 재  
충남대학교 언어병리학과, 충남대학교 언어학과

### Comparison of Prosodic Characteristics of Yes-No Questions with Interrogatives and Wh-questions according to Age, Gender, and Type.

Mijin Park, Cheoljae Seong  
Dept. of Speech-Language Pathology, Chungnam National University  
Dept. of Linguistics, Chungnam National University  
twoline0112@daum.net, cjseong49@gmail.com

한국어 의문문 운율에 관한 기존 연구들의 결과는 모집군의 특징과 실험 과제에 따라 상이하다. 이에 본 연구는 연령, 성별, 의문문 유형에 따라 의문문의 운율 특성을 음향학적으로 비교하여 한국어 운율 교육의 기초 자료를 마련하고자 했다.

대상은 서울, 경기, 충남 천안에 거주하는 표준어 화자 21명으로 연령(10대, 20대, 40대, 60대)과 성별의 비율을 맞춰 모집했다. 분석에 사용한 의문문은 '의문사 포함 예-아니오 의문문'과 '의문사 의문문'이다. 이를 대화 형식의 이야기 과제에 포함시켜 참여자들이 '직접 화법형'으로 읽도록 했다. 녹음은 당시 사회적 거리두기 권고에 따라 대면 혹은 비대면으로 진행하였고 상황과 초점을 반영한 그림을 함께 제공했다. 이후 음향학적 분석을 통해 어절 및 강세구의 위치에 따른 음도, 강도, 길이, 기울기를 살피고 두 위치 간에 운율 특성과 기울기 특성을 비교하였다. 분석에는 음도, 강도, 길이를 분석하기 위해서 '정규화된 음도 중앙값', '음도 범위', '정규화된 강도 중앙값', '상대적 발화시간 비율'과 같은 운율변수를 이용했다. 기울기 분석을 위해서는 기본주파수 기울기 변수인 '음도 변화량 기울기', '음도 회귀 기울기'를 사용했다.

{어디}, {누구}, {언제}가 포함된 예-아니오 의문문과 의문사 의문문을 비교한 결과, 의문문 유형에 따라 초점 실현 양상이 다르게 나타났다. 의문사의 초점 실현에 영향을 주는 변수는 의문사 의문문의 경우, 선행 어절에 위치한 의문사의 음도 범위와 기울기, 강도가 종합적으로 영향을 미쳤다. 이와 반대로 예-아니오 의문문에는 문미 어절의 음도와 음도 범위, 길이가 영향을 주었다. 문장 전체 음도는 의문사 포함 예-아니오 의문문이 의문사 의문문에 비해 높았다. 성별에 따라서 여성은 남성보다 음도 범위가 넓었고 연령에 따라서는 20대와 40대가 이러한 경향을 공유하였다. 즉, 의문사 의문문과 의문사가 포함된 예-아니오 의문문의 중의성은 초점이 다르게 실현됨으로써 해소 되었다. 그러나 문장 어절 수와 문장 내 의문사가 포함된 어절의 위치, 종결 어미의 유형에 따라 두 문장의 운율 특성은 다양하게 나타났다. 그 외에 성별에 따라서 선호 운율이 다를 수 있으며, 연령이 운율에 미치는 일반적인 영향 외에 사회적인 특성도 변화의 요인이 될 수 있으므로 해석에 주의가 필요하다. 해당 연구의 의의는 운율 특성 차이를 살핀 연구가 부족한 상황에서 한국어 의문문의 운율 특성을 분석하고 정리했다는 것에 있다.

## 급성 COVID-19 감염의 다면적 음성 특성 추적 사례연구

이 승 진

한림대학교 자연과학대학 언어청각학부 및 청각언어연구소

### A follow-up case of multi-dimensional voice features in acute COVID-19 infection

Seung Jin Lee

Division of Speech Pathology and Audiology, College of Natural Sciences, Research Institute of Audiology and Speech Pathology, Hallym University  
sjl@hallym.ac.kr

연구배경 및 목적: 발성장애(dysphonia)는 COVID-19 감염의 이비인후과 증상 중 하나로 잘 알려져 있으며, COVID-19 환자의 음성 특성은 다양한 음성 평가의 측면에서 보고되어 왔다. 그러나 급성 COVID-19 감염 환자의 음성 변화에 대한 추적 연구는 미흡하다. 따라서 이 사례보고 연구에서는 COVID-19 확진 후 일일 음성 추적검사를 통해 급성기의 음성변화 양상을 확인하고자 하였다.

방법: 연구대상은 만 41세 남성 1명으로, 1주일 간의 자가 격리 기간 동안 매일 모음 발성 및 표준 문단 읽기 과제를 수행하여 음성을 녹음하였다. 음향학적 분석을 통해 주기성을 기반으로 한 변수와 켈스트럼 변수를 측정하였고, GRBAS 척도를 이용하여 청지각적 평가를 하였다. K-VAPP를 이용하여 자가 중증도 평가를 계산하였다.

결과: 음질과 관련된 대부분의 측정치들에서 진단 2일차에 중증도가 정점을 기록하고 7일차에 회복되는 경향이 확인되었다. 전문가와 환자 스스로가 평정한 중증도가 일치하는 경향을 확인하였다.

결론: 전반적인 측정치들이 진단 후 2일차에 음성의 질이 가장 나빠지는 경향을 시사하였으며, 7일차에 대부분 회복되었다. 이러한 경향은 추후 급성기 COVID-19 환자의 음성 변화와 예후를 예측하는 데 도움이 될 수 있을 것으로 여겨진다.

## 나무 젓가락을 이용한 발음 연습의 음성학적 효과

강 보 란, 성 철 재  
충남대학교 언어병리학과, 충남대학교 언어학과

### Phonetic Effects of Pronunciation Practice using Wooden Chopsticks

Boran Kang, Cheoljae Seong  
Dept. of Speech-Language Pathology, Chungnam National University  
Dept. of Linguistics, Chungnam National University  
2844431@naver.com, cjseong49@gmail.com

목소리는 의사소통의 준언어적인 요소들을 함축적으로 포함하고 있으며, 이에 따라 의사소통에 큰 영향을 준다. 특히, 메시지를 전달하고자 할 때 다른 요소들보다 높은 비율을 차지한다. 또한, 말명료도에도 영향을 미치는데, 이는 목소리를 개선한다면 말명료도가 향상될 수 있음을 시사한다. 본 연구에서는 ‘나무젓가락 발음 연습’이 음성 변화와 말명료도에 영향을 미치는지를 알아보고자 하였다. 이러한 훈련은 대중매체에서 많이 소개되어 알려져 있고 혼자서도 쉽게 할 수 있다는 장점이 있으나, 여전히 효과에 대해 대립하는 의견들이 많고 과학적 입증과 연구가 부족하다고 생각한다. 그러므로 본 연구가 기초자료로서 의미가 있을 것으로 판단된다.

실험은 20대 성인(남자 32명, 여자 35명)을 대상으로 발음 연습 전(사전)/발음 연습 후(사후)/10분경과 후(유지)의 상태에서 읽기 과제(모음, 자음, 문단)를 실시하였다. 실험 종료 후 전문가를 대상으로 사전, 사후, 유지 간 음성 차이가 있는지 청지각 평가를 실시하였다.

연구 결과, ‘나무젓가락 발음 연습’은 모음 포먼트 에너지를 상승시키며, F1, F2는 유지(세션)에서도 지속되었다. 모음 삼각형 면적은 사전보다 사후에 더 넓어졌으며, 유지에서는 다시 좁아졌다. 자음은 강도(intensity), 스펙트럴 에너지, 무게중심, 왜도 값을 바탕으로 사후에 말명료도를 높이는 데 영향을 주었으며, 스펙트럴 에너지와 왜도는 유지에서도 지속되었다. 따라서, 훈련이 모음, 자음 명료도 향상에 일시적으로 도움을 줄 수 있음을 의미한다. 문단은 기본주파수, 말속도 변수를 보았을 때 사후에 말명료도의 향상에 도움을 주었으나, 유지 단계에서는 지속되지 않았다. 강도, 왜도, 무게중심은 사전과 사후에 차이가 없었다. 자음, 문단 과제에서 강도는 사전, 사후보다 유지에서 현저히 낮아졌는데, 이는 음성피로로 인해 음량·음성 조절력이 감소한 것으로 추측할 수 있다. 또한, 상대적으로 모음, 자음 과제보다 문단 과제에서 훈련 효과가 나타나지 않았는데, 이는 과제의 수준이 짧을수록 효과가 있음을 시사한다. 청지각 평가 결과, 사후 집단 점수가 가장 컸지만, 집단 간 점수 비율의 차이는 크지 않았다. 이는 문단 과제에서 음향학적 차이가 크지 않아서 이러한 결과가 나타난 것으로 볼 수 있다.

객관적인 분석을 통해 훈련 직후에는 말명료도 향상에 효과가 있음을 알 수 있었으며, 의사소통 상황에서 자신의 의도를 더 명확하게 전달하기 위해서 활용할 수 있을 것으로 보인다. 그러나 1회의 한시적인 훈련만으로 유지 효과는 보기 어렵다고 판단하였다. 그러므로 훈련의 기간을 두고 음성의 변화를 살펴본다면, 훈련 효과에 대해 규명할 수 있을 것이다.



# 포스터 발표

좌장: 이옥분(대구사이버대), 박정식(한국외대)





# 딥러닝 모델 전이학습 기반 한국어 방언 음성합성(Text-to-Speech)

강진아<sup>1,\*</sup>, 오유리<sup>2</sup>

<sup>1</sup>한국전자통신연구원 기업성장지원전략실, <sup>2</sup>(주)스피치랩스

## Korean Dialectal Text-to-Speech based on Deep Learning with Transfer Learning

Jin Ah Kang<sup>1</sup>, Yoo Rhee Oh<sup>2</sup>

<sup>1</sup>SMEs and Commercialization Division, ETRI, <sup>2</sup>SpeechLabs  
jakang@etri.re.kr\*, yroh@speechlabs.co.kr

최근 비약적인 발전을 이룬 딥러닝 기술은 광범위한 산업분야에 적용되어 많은 문제해결과 다양한 응용 서비스 개발을 이루는데 기여하고 있다. 음성인식 기술이 딥러닝 기술을 토대로 또다시 성능향상을 이루고 있는 것처럼, 음성합성(Text-to-Speech, TTS) 기술 또한 다양한 딥러닝 기술을 적용함으로써 기존의 한계점을 극복하며 사람의 목소리를 필요로 하는 다양한 서비스에 부합하게 기술발전을 이루고 있다. 이와 관련하여 본 논문에서는 한국어 방언 음성합성 기능을 효율적으로 제공하기 위한 방법으로, 딥러닝 모델의 전이학습(transfer learning) 기술을 활용한다. 즉, 사전 학습된 한국 표준어 음성합성 딥러닝 모델을 소용량의 방언(사투리) 음성 데이터셋으로 전이학습하여 방언에 적응된 최종 음성합성 모델을 얻는다. 이를 위해, 본 논문에서는 먼저 Tacotron2[1,2] 모델과 약 12시간 분량의 한국어 단일화자 음성 데이터셋(KSS)[3]을 이용하여 표준어 음향모델(Text2Mel)을 학습하였다. 이후 이 모델을 약 20분 분량의 경상도 방언 단일화자 음성 데이터셋[4]으로 미세조정하여 최종 음향모델을 학습하였다. 뉴럴 보코더(Mel2Wav)로는 MelGAN[5,6]을 KSS 데이터셋으로 학습하여 사용하였다. 실험결과, 학습된 음향모델과 뉴럴 보코더에 의해 경상도 방언 음성합성이 이루어짐을 확인하였다.

### 감사의 글

본 연구는 한국전자통신연구원의 'ETRI R&D성과의 사업화 촉진사업(22ZV1100)'의 일환으로 수행되었고, 산업통상자원부의 산업기술혁신사업으로부터 지원을 받아 수행된 연구(No. 10080667, 음원 다양화를 통하여 로봇의 감정 및 개성을 표현할 수 있는 대화음성합성 원천기술 개발)의 결과물인 공개 음성 데이터베이스 일부를 사용함

### 참고문헌

- [1] J. Shen, et al., "Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions," *arXiv preprint arXiv:1712.05884*, 2018.
- [2] T. Hayashi, et al., "ESPnet2-TTS: Extending the Edge of TTS Research," *arXiv preprint arXiv:2110.07840*, 2021.
- [3] K. Park, *Korean Single Speaker Speech Dataset*, 2020, (<https://www.kaggle.com/bryanpark/korean-single-speaker-speech-dataset>)
- [4] (주)셀바스에이아이, *대화형 음성코퍼스 DB*, 2020, ([https://github.com/emotiontts/emotiontts\\_open\\_db/tree/master/Dataset/SpeechCorpus/Personality/pfo](https://github.com/emotiontts/emotiontts_open_db/tree/master/Dataset/SpeechCorpus/Personality/pfo))
- [5] K. Kumar, et al., "MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis," *arXiv preprint arXiv:1910.06711*, 2019.
- [6] T. Hayashi, *Parallel WaveGAN implementation with Pytorch*, 2022, (<https://github.com/kan-bayashi/ParallelWaveGAN>)

## 매체(media) 차이가 학령기 아동의 운율과장(prosody exaggeration)에 미치는 영향

박 은 영 , 성 철 재  
충남대학교 언어병리학과, 충남대학교 언어학과

### The effects of media differences on prosodic exaggeration in school-age children

Eun-young Park, Cheol-Jae Seong  
Dept. of Speech-Language Pathology, Chungnam National University  
Dept. of Linguistics, Chungnam National University  
chaos7928@naver.com, cjseong49@gmail.com

운율은 음높이, 크기, 길이 등과 운율 자질로 이루어져 강세, 억양 등을 구성하며 의사소통에 중요한 역할을 한다. 운율에 대한 국내외 연구 동향을 살펴보면 자폐 범주성 장애, 우뇌 손상, 청각장애, 마비말 장애, 읽기 장애 등 언어치료에서 의사소통의 주요 내용으로 다루어지고 있다. 본 연구는 초등 1학년 아동을 대상으로 소리만 들려준 집단 14명, 동영상(소리와 영상을 함께 접한) 집단 14명으로 나누어 영상효과가 운율 과장에 미치는 영향을 살펴보았다. 매체(media) 간 비교 연구에서 뮤지컬 가사만 들려준 집단과 드라마 대사만 들려준 집단을 비교한 결과 소리만 또는 영상만을 접하고 발화했을 때 실험 전/후 통계적으로 유의한 운율 변화가 관찰되지 않았으나, (소리가 결합된) 동영상을 접한 후의 말속도, 강도, 음도 측정 결과에서는 소리만 접한 집단보다 운율 과장 효과가 좀 더 큰 것으로 나타났다. 결론적으로 영상에 대한 긍정적인 효과를 기대해 볼 수 있었으며 읽기 유창성뿐만 아니라 화용적 측면에도 중요하여 언어치료 분야 중 시각적 자극에 민감한 자폐 범주성 장애 군과 구어장애 등에 응용해 볼 수 있다.

## ADHD증상 및 불안, 우울을 보이는 성인의 실행기능 훈련 효과

김 보 선, 김 성 태  
동신대학교 언어치료학과

### Effect of Execution Function Training on Adults with ADHD Tendency, Anxiety, and Depression

Boseon Kim, Seong-Tae Kim  
Dept. of Speech-Language, Dongshin University  
kbs2812@hanmail.net, voicekim@dsu.ac.kr

본 연구의 목적은 ADHD증상 및 불안, 우울을 보이는 성인의 특성을 알아보고, 실행기능 훈련을 통해 ADHD증상, 불안, 우울성이 감소하는지 확인하고자 하는 것이다. 대상은 광주광역시 소재의 브레인 센터에 다니는 성인들이며, 지능에 문제가 없고, 시각, 청각 등 기타 감각기관에 문제가 없는 7명의 성인을 대상으로 하였다. 브레인메모리 검사(종합 신경 심리 검사)를 통하여 주의력, 인지기능 및 실행기능을 평가하고, 임상 인터뷰를 통해 현재 삶에 영향을 주는 불편함을 조사하였다. 주로 주의력 결핍, 감정조절, 자기판단에 어려움을 호소하였으며, 이로 인해 불안감과 우울감을 느끼고 스트레스를 많이 받고 있는 상태였다. 실행기능 훈련 도구는 첨단인지브레인 연구소에서 개발된 심리안정, 불안감소 프로그램 'Sound brain wave'와 트라우마 전이 완화 프로그램 'Blockcog', 전환주의력 향상 프로그램 'TMT(trail making test)' 인지유연성 향상 프로그램 'WCST(wisconsin card sort test)'를 사용하여 실행기능 훈련을 진행하였다. 주 1회, 3개월 동안 실행기능 훈련을 진행한 후 사후검사를 실시하였다. 검사결과 대상자 모두 ADHD증상, 불안, 우울 수치가 감소하였다. 특히, 대상자 A는 대학성적과 동기 친구들과의 관계에도 긍정적인 영향을 주었고, 대상자 C는 자기조절력이 향상되어 자존감이 높아졌으며 취업에 성공하였다. 성인의 ADHD 증상 및 불안, 우울은 자기통제, 정서조절 결함, 의사결정 장애와 관련이 높으며 스트레스 대처능력, 문제 해결 능력과도 높은 상관관계가 있는 것으로 확인되었다. 연구 결과를 통해 실행기능 훈련으로 성인의 주의력, 감정조절, 자기판단의 어려움을 개선할 수 있음을 시사하였다.

# 말실행증 성인대상 반복훈련과 말속도-리듬 통제 훈련 효과에 대한 사례연구

염현자, 한지용, 박진  
가톨릭관동대학교 언어재활과

## Effects of Repetition Practice and Speech-Rate Control Training on the Percentage of Correct Consonants and Speech-Related Measures: A Case Study

Yom, Hyunja, Han, Jiyong, Park, Jin  
Dept. of Speech-Language Rehabilitation, Catholic Kwandong University  
knamyhj11@cku.ac.kr, slp4u@cku.ac.kr, gatorade70@cku.ac.kr

### I. 서론

반복연습은 주로 말실행증 치료 초기 단계에서 활용되는 기본적인 치료법이며, 말속도/리듬통제치료는 말실행증 성인들의 말속도를 감소시켜 말산출 능력을 향상시키는 치료법으로 말속도를 통제하기 위해 손가락 두드리기, 메트로놈 등과 같이 다양한 외부통제방법들을 사용할 수 있다(Dworkin & Abkarian, 1996). 본 연구는 말실행증 성인을 대상으로 반복연습과 말속도-리듬 통제훈련을 위계적으로 실시하였을 때 자음정확도 및 말산출 능력에 미치는 영향을 살펴보았다.

### II. 연구방법

#### 1. 연구대상자

본 연구의 대상자는 성인환자 1명으로 생활연령은 57세, 실어증과 말실행증 동반으로 발병 전 우세손은 오른손, 발병 후 경과 기간은 15년 이상, 교육년수는 16년으로 발병 전 활발한 사회활동을 하였으며, 발병 후 시력과 청력 장애는 동반하지 않았지만 우측사지 편마비 증세가 있다.

한국판 간이정신상태검사(K-MMSE) 결과 27점, 한국판 몬트리올 인지평가(K-MOCA)는 21점, P-K-WAB AQ 검사 결과 총 36.1점, 72.2%ile로 명칭실어증으로 판정되었다. 말실행증 체크리스트(Huskin, 1986) 검사 결과는 2.7로 중도(moderate) 수준의 말실행증을 나타냈다.

#### 2. 측정도구 및 방법

반복연습을 위한 훈련날말은 수용.표현 어휘력 검사(REVT: 김영태 외, 2009)에서 2음절 단어 20개, 3음절 단어 20개, 4음절 단어 10개, 총 단어 50개를 사용하였고, 노트북(LG PC DESKTOP-7L2MCBE)을 사용하여 모니터에 한 단어씩 보여주었다. 말속도통제훈련을 위해서는 수용.표현 어휘력 검사(REVT)에서 5~10음절 낱말을 무작위로 선정하여 40개 구문으로 제작하였다.

낱말 반복훈련은 매회기 낱말 10개 문항을 선택하여 5회씩 반복하기를 3회 실시, 총 15회 소리내어 (read-aloud) 반복 읽기를 실시하였다. 말속도-리듬통제훈련은 1초 간격으로 외부통제방식(손가락 두드리기)으로 40개 구문으로 실시하였고, 반복훈련과 마찬가지로 10개 문항을 5번씩 반복, 3회씩 실시하였다.

### 3. 연구절차 및 자료 분석

치료의 전 과정은 Samsung Galaxy S7 edge를 사용하여 동영상 녹화하였다. 자음정확도 평가를 위해 녹화된 피험자의 발화 중 왜곡, 대치, 첨가, 생략, 알아들을 수 없는 반응, 보속 현상의 발화, 착어로 인한 계속된 자가수정 등 피험자의 발화를 모두 전사하였다. 한 낱말에서 착어나 모색행동, 자가수정이 빈번하게 반복되는 경우 전사 후 마지막 반응을 자음정확도 평정자료로 사용하였다.

결과 산출은 첫 회기에 1번씩 조음한 것과 훈련 후 최종 반응결과를 대조하였고, 말실행증 대상자의 반복, 자음정확도, 모색행동, 착어 등의 조음오류를 평가하여 말산출 능력이 향상되었는지를 평가하였다.

## Ⅲ. 연구결과 및 논의

연구결과, 낱말반복훈련은 전반적으로 자음정확도가 향상되었고, 조음 오류 및 모색행동 등의 감소가 보였다. 말속도 리듬통제훈련에서는 외부통제없이 구문을 바로 읽은 경우에 음소, 음절, 어절에서 반복, 대치, 왜곡 등의 현상을 많이 보였으나 외부통제(손가락 두드리기)를 통한 과제를 수행한 경우 이러한 현상이 현저히 감소하였다. 다만 대상자의 자가호흡조절, 바른자세, 복식호흡 등의 의도적인 노력이 말산출에 긍정적인 영향이 있었음과 추운 날씨로 인한 급격한 온도차이 등의 환경적 요인과 긴장, 스트레스 등 대상자의 심리적 요인, 대상자의 건강 악화가 말산출시 음색이나 집중도 등의 부정적인 영향을 미침을 알 수 있었다. 하지만 전반적인 훈련 결과를 볼 때 말실행증 환자에게 낱말반복훈련과 말속도-리듬속도통제훈련을 체계적이고 목표로 지속적으로 시행하게 된다면 언어산출능력에 매우 효과적일 것이라 본다.

## 참고문헌

- 김영태, 홍경훈, 김경희, 장혜성, 이주연. (2009). *수용·표현어휘력검사(REVT)*. 서울: 서울장애인종합복지관.
- 김향희, 나덕렬. (2001). *파라다이스 한국판-웨스턴 실어증검사(P-K-WAB)*. 서울: 파라다이스 복지재단
- Dworkin, J. P. & Abkarian, G. G. (1996). Treatment of phonation in a patient with apraxia and dysarthria secondary to severe closed head injury. *Journal of Medical Speech Language Pathology*, 4, 105-115.
- Huskins, S. (1986). *Working with apraxic clients: a practical guide to therapy for apraxia*. Tucson, AZ: Communication Skill Builders.
- La, E. Y. & Sim, H. S. (2014). Repetition Practice and Speech-Rate Control Training's Effect on the Percentage of Correct Consonants and Ratio of Delayed Words in Apraxia of Speech Patients with Aphasia(반복연습 및 리듬-속도조절 훈련이 자음정확도와 반응지연문항 비율에 미치는 영향). *Communication Sciences & Disorders*, 19(3), 342-351.

## 마스크 종류에 따른 아동의 말지각력 비교

장은영, 김지연  
우송대학교 언어치료청각재활과

### Comparison of children's speech perception by mask type

Eunyeong Chang, Jiyouon Kim  
Dept. of Speech Therapy & Aural Rehabilitation, Woosong University  
alskdj3525@gmail.com, bluejiyouon@hotmail.com

코로나19 감염과 확산을 방지하기 위한 의무적인 마스크 착용은 언어의 음향학적, 시각적 정보 접근을 어렵게 한다. 영유아 시기에 감각을 통한 말소리 지각이 언어발달에 영향을 미치는데 마스크 종류마다 마스크에 사용된 재질에 따라 주파수 왜곡이 발생하여 말소리 지각에도 영향을 준다고 하였다. 따라서 본 연구에서는 정상 청력을 가진 아동을 대상으로 마스크 착용 유무 및 마스크 종류(KF-94, Lip-View)에 따른 말지각 능력에 차이가 있는지, 1음절 단어와 2음절 단어를 제시하여 음절수에 따른 말지각 능력에 유의한 차이가 있는지 살펴보고자 한다.

3-4세 건청 아동 20명을 대상으로 보기가 있는 조건(closed set)의 단어인지 항목을 사용하여 아동이 검사자가 말한 단어를 듣고 제시된 3개의 그림 중에 선택하도록 하여 검사를 실시하였다.

건청 아동들은 단어인지 정반응률 평균은 95~96%로 마스크 착용 유무 및 마스크 종류에 따른 말지각 능력에 통계적으로 유의미한 차이를 보이지 않았다. 그러나 음절수에 따른 말지각에서는 차이를 보이며 1음절 낱말에 비해 2음절 낱말의 말지각 수행이 더 좋은 것으로 나타났다.

청력에 문제가 없는 건청 아동의 경우 마스크 착용으로 인해 시각적인 정보가 없고, 음향학적인 왜곡이 있더라도 조용한 일대일 대화 환경에서는 청각적인 자극만으로도 단어 수준의 말지각 수행에는 어려움이 크게 없으나 단음절 단어에 대한 말지각 수행은 단어 내에서 문맥적인 단서가 매우 적어서 2음절 단어에 비해 수행에 어려움이 있는 것으로 해석할 수 있다. 본 연구는 언어발달 시기에 있는 건청 아동들에게 단어 수준에서 마스크 착용이 말소리 지각에 미치는 영향에 대해 알아보았는데 의의가 있다. 그러나 건청 아동과 달리 청각장애를 가진 아동들은 청각적인 정보의 제한으로 마스크 착용 유무나 시각적 정보의 활용에 따라 말지각 수행에 차이를 보일 수 있으므로 이에 대한 연구가 이루어질 필요가 있다.

#### 참고문헌

- [1] M. Y. Lee, J. C. Shin, H. H. Kim, & L. S. Kim. (2009). Open-Set Monosyllabic Speech Perception Test for Preschool Children. *Korean J Otorhinolaryngol-Head Neck Surg*, 52, 312-21.
- [2] S. R. Atcherson, L. L. Mendel, W. J. Baltimor, C. Patro, S. Lee, M. Pousson, M. J. Span. (2017). The Effect of Conventional and Transparent Surgical Masks on Speech Understanding in Individuals with and without Hearing Loss. *Journal of the American Academy of Audiology*, 28, 58-67.

## 언어발달지연 영유아에게 미치는 Floortime 중재 효과

양 영 진, 김 성 태  
동신대학교 언어치료학과

### The effect of floortime intervention on infants and toddlers with speech development delay

Young jin Yang, Kim Seong-Tae  
Dept. of Speech-Language Pathology, Dongshin University  
dudwslvn867@naver.com, voicekim@dsu.ac.kr

최근 코로나바이러스감염증-19(COVID-19)이 크게 확산되고 장기화 되면서 전 세계적으로 많은 변화를 겪게 되었다. 많은 변화들 가운데 마스크를 착용하게 됨으로써 언어발달 지연이 우려되어지고 있다. 영유아들의 언어발달은 사회적 상호작용 속에서 이루어지며, 아동이 성장하는 중요한 시기에 적절한 상호작용발달이 이루어지지 않는다면 이후 언어발달에 문제가 생길 수 있다. Floortime은 아동의 개인적인 발달 수준에 맞추어 상호관계적인 발달을 향상시킴으로써 사회적 기술을 개발하고 사용되는 사회적 실용주의적 접근방법이다. Floortime단계는 Regulation&Shared Attention, Engaging & Relating, Intentional-Reciprocal Communication, Shared Social Problem Solving & Developing a Sense of Self, Creating Symbols and Ideas, Logical Thinking, Multi-causal Triangular Thinking이다.

본 연구의 목적은 대표적인 관계기반 접근법인 Floortime 중재가 언어발달지연 영유아에게 어떠한 중재 효과가 있는지 확인해 보고자 실시하였다. 대상은 2022년 3월~ 2022년 9월까지 초기 상담 시 언어발달에 어려움을 겪는 것을 보고한 48개월 미만의 남아 8명을 대상으로 하였다. 대상자들은 모두 영유아 언어발달 검사인 SELSI와 아동의 발달을 체크할 수 있는 K-CDI를 사용하여 사전검사와 사후검사를 실시하였으며, 주 2회씩 6개월 동안 Floortime 중재 방법으로 진행하였다. 아동들의 개월 수가 적고 분리가 어려운 것을 감안하여, 주 4회까지 부모와 함께 입실하여 진행하였다.

Floortime중재 접근법으로 진행한 결과, 대상자 모두 SELSI에서 수용과 표현언어 능력이 향상되었음을 확인하였다. K-CDI에서도 사회성, 자조행동, 표현언어, 언어이해 부분에서 유의미한 향상을 보였다. 특히 일상생활에서 소극적이고 회피했던 태도들에서 상대방의 행동을 모방하거나 말소리에 집중하는 태도가 나타났고 무발화 이었던 아동들이 익숙한 놀이를 먼저 시작하고 일관성 있는 낱말을 사용하는 것을 관찰할 수 있었다. 거부하는 모습인 소리를 지르거나 물건을 던지는 행위가 감소되어 상대방에게 도움을 청하고 관습적인 몸짓언어(주세요 등)로 바뀐 것을 확인 하였다. 이는 Floortime 중재 방법으로 시행하였던 영유아의 함께하기 능력이 향상되면서 개별적인 상호작용 뿐 만 아니라 친밀한 어른, 아동 등의 상호작용 발달이 이루어진 것으로 보이며, 주고받는 의사소통의 순환고리가 유지되면서 이로 인해 언어발달을 포함한 전반적인 발달이 증가 된 것으로 사료 되어진다. 향후 지속적인 치료효과에 대한 후속 연구가 필요할 것으로 사료되어 진다.



## 딥러닝 기반 입술 움직임 분석 연구

강영애\*, 김민혁\*, 한연규\*, 장재원\*, 구분석\*, 이동현\*

\*충남대학교병원 이비인후과, \*충남대학교 의공학과

### Deep-Learning based Lip Movement Analysis Research

Youngae Kang\*, Minhyuk Kim\*, Yeongyu Han\*, Jaewon Jang\*,  
Bonseok Koo\*, Dongheon Lee\*

\*Dept. of Otorhinolaryngology, Chungnam National University Hospital

\*Dept. of Medical Engineering, Chungnam National University  
kangyoungae1@naver.com, dhlee.jubilee@gmail.com

**연구목적** : 말소리 산출은 호흡, 발성 공명과 조음기관을 협응이 필수적이다. 모든 협응이 잘 이루어져도 최종 단계인 조음기관에서 개구(입 벌림) 및 조음점이 틀려지면 의사소통에 어려움을 겪는다. 음성장애 환자들은 입술을 작게 벌려 구강 공명을 적게 하는 경향이 있다. 전통적으로 구강 면적을 알아보는 방식으로 제1과 제2 포먼트를 기반으로 모음삼각도를 구하는 접근법이 있으나 이는 음질이 안 좋은 화자(speaker)에 적용할 경우 포먼트를 정확하게 구하는데 음성분석의 한계가 있고 이는 시상면(sagittal view) 관점이다. 본 연구팀은 관상면(coronal view)적 접근으로 구강개방(입술 움직임)에 대한 연구를 딥러닝 기술을 이용한 객관적인 측정 방법을 제안하고자 한다.

**연구방법** : 딥러닝 기술로 화자(speaker)가 발화를 할 경우 입술을 벌리므로 얼굴에서 입술을 인식시키는 작업을 해야 하고 각 발성별 어떤 모음을 발화하였는지 정확하게 맞추어야 한다. 이에 모음별 발성에 대한 구강 개구(입술 벌림)를 자동으로 분할(segmentation) 하는 딥러닝 모델(DL model)을 개발하였다. segmentation 모델은 U-net(efficientNet-b3)을 사용하였다. 대규모 얼굴 이미지 오픈 데이터셋의 입술 영역을 훈련용(training set)으로 사용하고, 비디오로 모음발성을 촬영한 음성장애 환자 영상을 실험군(test set)으로 하였다. 음성장애 환자(patient)를 대상으로 모음(/아,이,우,에,오/) 발성을 관상면에서 녹화하였고 사전에 연구에 대해서 설명하고 동의를 구하였다. 촬영된 환자 영상에서 각 모음별 발성시간의 앞 뒤 구간의 값을 구하였다. 분석 변수는 입술 면적, 가로 길이, 세로 길이이며 단위는 픽셀(pixel)로 계산하였다. 환자마다 카메라와 얼굴 간의 거리가 달라 촬영화면에서 입술 크기가 달라짐을 보정하기 위하여 개인별 전체 입술외곽 면적대 안쪽 입술 면적의 비율을 설정하여 계산한 정규화(normalizing) 처리를 하였다.

**연구결과** : 음성장애 환자 3명의 모음별 입술 면적을 전체총면적, 가로 길이, 세로 길이를 분석하였고 평균 처리하였다. 모음별 원자료 값과 예측값 간의 픽셀차이(면적-가로-세로길이 순서, 발성시작점 기준, 절대값)는 모음/아/는 1472, 25, 187 픽셀이고, 모음/이/는 1350, 26, 15 픽셀, 모음/우/는 11661, 24, 14 픽셀, 모음/에/는 1445, 26, 16 픽셀, 모음/오/는 1052, 23, 16 픽셀 차이를 보였다.

**연구결론** : 딥러닝 기술을 사용하여 음질(voice quality) 여부와 상관없이 얼굴 정면에서 보는 관상면에서의 객관적 입술 면적 및 길이를 측정하는 프로그램을 개발하였다. 이는 여러 환자군(우울증 여부, 청각장애, 마비말장애 등)에게 적용하여 입술 면적 및 길이를 바탕으로 입술 움직임을 비교하는 객관적 측정 방법이 될 것으로 기대된다.

## 특강 III

좌장: 박상희(대구사이버대)

Voice pathology detection using deep learning,  
machine learning, and hybrid modeling  
(변해원, 인제대학교)



## Voice pathology detection using deep learning, machine learning, and hybrid modeling

Haewon Byeon  
Department of Digital Anti-aging Healthcare (BK-21),  
Inje University  
byeon@inje.ac.kr


Dysphonia includes components of speech characteristics that relate to your voice. For example, these include rate, intensity, pitch, and quality. Therefore, It is important to evaluate the voice from various aspects. Machine learning models, such as support vector machine (SVM), have traditionally been used to identify speech features.

Following a review of recent studies, I discovered that Convolutional Neural Networks (CNN) outperforms methods with a relatively high overall accuracy. Shallow and deep neural networks are both capable of approximating any function. When compared to shallow neural networks, deep neural networks have the advantage of being able to do feature extraction while learning on large datasets. This is primarily due to the deep models' ability to extract.

As far as I know, in order to predict laryngeal disease using negative indicators, it is more efficient to use a hybrid model rather than develop an individual (single) deep learning model. My interest raises the question of whether hybrid models outperform single models regardless of the dataset's characteristics. As a result, my research investigates how particular hybrid models respond to various kinds of datasets from various fields.

My experiments show that hybrid models beat all other investigated models in terms of dependability for voice pathology analysis. When performing voice pathology analysis, combining deep learning models with the SVM technique (or machine learning technique) produces better results than using a single model. In majority of the investigated datasets, hybrid models with machine learning outperform those without; nevertheless, the computational time for the hybrid models with machine learning is significantly longer. I also discovered that the effectiveness of the algorithms is strongly influenced by the qualities and quality of the datasets.

[감사의 글] 이 논문은 2022년도 과기정통부 동남권 개방형 혁신연구실 육성(오픈랩) 지원 사업(2022openlab(RnD)\_0361582062860202)과 연구재단 이공학개인지초연구과제(2018R1D1A1B07041091), 인문사회기초연구사업(2021S1A5A8062526)의 지원을 받아 수행된 연구임.



2022 한국음성학회 특강

## Voice pathology detection using deep learning, machine learning, and hybrid modeling

- 1)인제대학교 대학원(BK21) 디지털항노화헬스케어학과,
- 2) 과학기술정보통신부 인증 인제메디컬빅데이터 연구실 (INJEM lab: <https://sites.google.com/view/injem>)

Prof. 변해원 (Haewon Byeon)

### 강사 소개




성명	변해원
소속	인제대학교 (BK21)대학원 디지털항노화헬스케어학과
분야	메디컬 빅데이터, 노인병 고위험군 예측
홈페이지	INJEM ( <a href="https://sites.google.com/view/injem">https://sites.google.com/view/injem</a> )
유튜브	 <a href="#">빅날괴</a>





# CONTENTS

- 1 정확도 높은 예측 모델링을 위한 도전 (하이브리드 모델링)
- 2 설명 가능한 예측을 위한 도전 (X-AI)
- 3
- 4
- 5
- 6
- 7
- 8



CSD (2010)

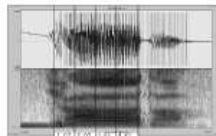
## Comparing the Acoustic Character of Diphthong Production between Flaccid Dysarthria and Spastic Dysarthria

Hae Won Byeon<sup>§</sup>

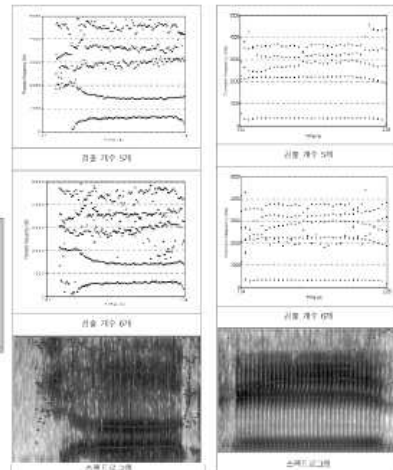
Dept. of Preventive Medicine & Public Health, Aju University, Suwon, Korea

<sup>§</sup>Correspondence to

Hae Won Byeon, MEd,  
Department of Preventive  
Medicine & Public Health,  
Aju University School of  
Medicine, San 3, Wonchon-  
dong, Teonjong-gu, Suwon,  
Korea  
e-mail: byun@ajou.ac.kr  
tel: +82 31 219 5302



<그림 1> -> 이음매 시간상에 따른 구강형 표현도



<그림 1> -> 10개 강모계구의 강모계구 F1과 F2의 구강 표현

<그림 2> -> 10개 강모계구에 따라 F1과 F2의 구강 표현의 차이

**Variability and Duration of Voice Onset Time in Childhood Apraxia of Speech in Korea: Results of the Spectrum Study**

Junbeom Park<sup>1</sup> Haewon Byeon<sup>2,3\*</sup>  
 1 JUN Rehabilitation Medicine, Seoul, South Korea

**2.2.2. VOT**

VOT of collected utterance data was analyzed by using wave form of Praat (version 5.3.17) and spectrograph. Procedure of analysis was as follows; first, in the oscillogram of spectrograph, the point where vowel section was confirmed was selected. Then, referring to a preceding study [11], sections for VOT analysis were selected. At this time, the starting point of VOT was defined as the point where irregular wave began while the end of VOT was defined as the point where the wave of vowel appears when the first and second formant (F1, F2) began. Measured section of VOT is presented in Fig. 1.

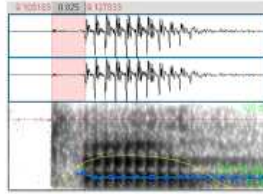


Figure 1. Measured Section of VOT in Praat

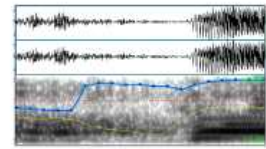


Figure 3. The Case that Magnifying Spectrogram and Oscillogram

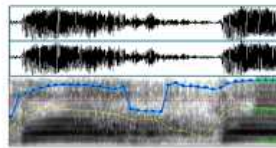


Figure 2. The Case that Difficult to Identify the VOT Section

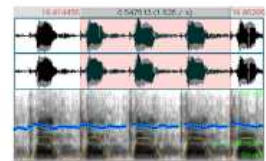


Figure 4. Calculation of Variability of VOT

**Could Neuromuscular Electrical Stimulation Really Lead to Changes in Voice of Swallowing Disorder Caused by a Stroke?**

Haewon Byeon<sup>1</sup> and Sunghyoun Cho<sup>2\*</sup>

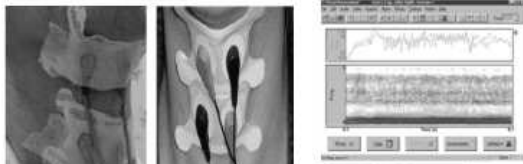


Figure 4. Electrode Placement of VitalStim

Table 2. Results of Acoustic Evaluations after Neuromuscular Electrical Stimulation

Variables	NMES		Compensatory strategies		F	p
	Pre-treatment	Post-treatment	Pre-treatment	Post-treatment		
F <sub>0</sub> (Hz)	130.8±9.0	131.8±10.8	133.9±8.7	134.2±8.0	0.001	0.955
Intensity (dB)	41.8±3.6	44.5±3.3	42.2±3.1	42.5±3.0	9.313	0.005
f <sub>0</sub> er (%)	1.1±0.3	0.5±0.2	1.1±0.2	0.8±0.2	14.661	0.001
Stimulus (%)	3.4±1.0	2.5±0.6	3.3±0.6	2.8±0.7	4.358	0.046
NNE (%)	-5.7±3.3	-6.7±2.3	-5.3±2.6	-6.8±2.3	1.203	0.266

**Exploring Potential Risk Factors for Benign Vocal Fold Mucosal Disorders Using Weighted Logistic Regression**

Haewon Byeon<sup>1,2</sup>

80% (n=8,057) [11] This study selected as subjects 1,335 persons who had completed both the questionnaire survey on health and an endoscopic laryngeal examination, of whom 208 persons were excluded because the results of their endoscopic laryngeal examinations were illegible, leaving us with a total of 1,127 persons (468 males and 659 females) as the final subjects of study.

**Table 3. Odds Ratios (OR) and Confidence Intervals (CI) of Factors Related to Benign Vocal Fold Mucosal Disorders: Weighted Logistic Regression Model**

	Crude model			Adjusted model		
	OR	p	95% CI	OR	p	95% CI
Age		0.775		0.589		
(19-39)						
40-59	1.18		0.41 3.44	1.58		0.43 6.79
≥60	9.82		0.35 2.75	1.89		0.51 7.79
Sex		0.003		0.283		
(Male)						
Female	1.45		0.64 3.77	2.06		0.52 8.15
Education level		0.041		0.025		
( <sup>1</sup> Elementary school)						
Middle school	1.38		0.24 8.00	3.49		0.73 16.44
High school	2.79		1.28 6.17	8.84		1.85 25.23
College	1.71		0.39 7.36	5.99		1.11 32.17



**Figure 1. Type of Benign Vocal Fold Mucosal Disorders**



**전통적인 회귀모형의 한계점**

정규성, 등분산성 위배  
 독립적인 개별 요인만  
 사람은 건강행위는 **복합** (다중건강행위)



다중건강행위의 파악이 필요





## 정확도 높은 예측 모델링을 위한 도전 (1)

### 1. 누적 위험요인

- 그런데, 모든 위험요인은 동일한 크기(가중치)를 가질까?

## Multiple Socioeconomic Risks and Cognitive Impairment in Older Adults

Back J.H., Kim J., Byeon H., et al.

Author affiliations

Keywords: Cognitive impairment, Socioeconomic factors, Education, Income, Wealth, Occupation

Dement Geriatr Cogn Disord 2019;29:523-529

https://doi.org/10.1159/000515587

ABSTRACT GET ARTICLE LOGIN / REGISTER



### Abstract

**Background/Aim:** To investigate the multiple effects of socioeconomic factors on cognitive impairment in older persons. **Methods:** Subjects were a nationally representative sample (n = 4,155) aged ≥65 years of the Korean Longitudinal Study of Aging. Cognitive status was assessed by the Mini-Mental State Examination. Education, income, wealth and occupation were measured, with low levels on each factor combined to produce an aggregate risk score. Multiple logistic regression was used to analyze the combined effects of socioeconomic risks on cognitive impairment. **Results:** A socioeconomic risk gradient was evident, with the combined effect being stronger than any one of the risk factors acting alone. Compared with those with no risks, the odds ratios were 2.4 (for 2 socioeconomic risks), 3.4 (3 risks) and 7.7 (4 risks) for men, and 2.8 (2 risks), 3.5 (3 risks) and 5.4 (4 risks) for women. For individual socioeconomic risks, income and occupation were significant predictors of cognitive impairment in men, whereas in women, all 4 factors were independently associated. **Conclusion:** Multiple socioeconomic risks have a combined effect on cognitive impairment in older adults.



## 정확도 높은 예측 모델링을 위한 도전 (2)

### 2. 더미변수로 변환

- 변수가 늘어나면 경우의 수도 기하급수적으로 증가
- 실제 인구 집단에서 드문 조합의 추정도 발생

Table 2. Combinations of healthy lifestyle factors in adults aged ≥65 years from the 2006 Korean Longitudinal Study of Aging: observed and expected frequency %

	S	A	P	B	Men (n = 1688) n (weighted %)	O	E	O/E (95% CI)	Women (n = 2251) n (weighted %)	O	E	O/E (95% CI)
All four	+	+	+	+	189 (10.9)	11.2	10.3	1.09 (0.94, 1.26)	287 (12.2)	12.7	12.4	1.03 (0.91, 1.15)
Three	+	+	+	-	68 (3.8)	4.0	3.2	1.26 (0.98, 1.60)**	118 (5.3)	5.2	5.2	1.01 (0.83, 1.20)
	+	+	-	-	415 (24.7)	24.6	23.1	1.07 (0.97, 1.17)	1207 (54.0)	53.6	53.5	1.00 (0.95, 1.06)
	+	-	+	-	93 (5.3)	5.5	6.1	0.91 (0.73, 1.11)	4 (0.3)	0.2	0.4	0.49 (0.13, 1.25)
	-	+	+	-	59 (3.3)	3.5	4.5	0.77 (0.55, 1.00)**	7 (0.3)	0.3	0.5	0.65 (0.26, 1.35)
Two	+	+	-	-	136 (8.0)	8.1	7.2	1.12 (0.94, 1.33)	501 (22.4)	22.3	22.5	0.99 (0.91, 1.08)
	+	-	+	-	39 (2.4)	2.3	1.9	1.23 (0.87, 1.69)	2 (0.1)	0.1	0.2	0.58 (0.07, 2.10)
	+	-	-	-	182 (10.8)	10.8	13.4	0.79 (0.68, 0.92)***	32 (1.7)	1.4	1.4	0.91 (0.62, 1.28)
	-	+	+	-	15 (0.9)	0.9	1.4	0.63 (0.35, 1.03)**	5 (0.3)	0.2	0.2	1.11 (0.36, 2.60)
	-	+	-	-	139 (8.4)	8.2	10.1	0.81 (0.68, 0.96)**	45 (2.0)	2.0	2.0	0.98 (0.71, 1.31)
One	-	-	+	-	45 (2.5)	2.7	2.7	1.00 (0.73, 1.34)	0 (0.0)	0.0	0.0	-
	+	-	-	-	51 (3.0)	3.0	4.2	0.72 (0.53, 0.94)**	17 (0.7)	0.8	0.7	1.15 (0.67, 1.84)
	-	-	-	-	41 (2.5)	2.4	3.1	0.77 (0.55, 1.03)**	17 (0.7)	0.8	0.9	0.88 (0.51, 1.41)
None	-	-	-	+	12 (0.7)	0.7	0.8	0.86 (0.44, 1.58)	1 (0.05)	0.04	0.01	7.60 (0.10, 42.15)
	-	-	-	-	166 (10.5)	9.8	6.0	1.65 (1.41, 1.92)****	3 (0.1)	0.1	0.1	2.23 (0.45, 6.50)
None	-	-	-	-	38 (2.3)	2.3	1.9	1.21 (0.86, 1.67)	5 (0.2)	0.2	0.03	8.82 (2.84, 20.56)****

Healthy lifestyle factor: present (+) or absent (-). \*\*P < 0.01, \*\*\*P < 0.05, \*\*\*\*P < 0.001 in  $\chi^2$ -test. A, no heavy alcohol consumption; B, normal body mass index; CI, confidence interval; E, expected; O, observed; P, regular physical activity; S, non-smoking.



## 정확도 높은 예측 모델링을 위한 도전 (2)

### 2. 더미변수로 변환

- 변수가 늘어나면 경우의 수도 기하급수적으로 증가
- 실제 인구 집단에서 드문 조합의 추정도 발생

Table 2. Combinations of healthy lifestyle factors in adults aged ≥65 years from the 2006 Korean Longitudinal Study of Aging: observed and expected frequency %

	S	A	P	B	Men (n = 1688) n (weighted %)				Women (n = 2251) n (weighted %)			
					O	E	O/E (95% CI)	O	E	O/E (95% CI)		
All four	+	+	+	+	189 (10.9)	11.2	10.3	1.09 (0.94, 1.26)	287 (12.2)	12.7	12.4	1.03 (0.91, 1.15)
Three	+	+	+	-	68 (3.8)	4.0	3.2	1.26 (0.98, 1.60)**	118 (5.1)	5.2	5.2	1.01 (0.83, 1.20)
	+	+	-	+	415 (24.7)	24.6	23.1	1.07 (0.97, 1.17)	1207 (54.0)	53.5	53.5	1.00 (0.95, 1.06)
	+	-	+	+	93 (5.3)	5.5	6.1	0.91 (0.73, 1.11)	4 (0.2)	0.2	0.4	0.49 (0.13, 1.25)
Two	-	+	+	+	59 (3.3)	3.5	4.5	0.77 (0.59, 1.00)**	7 (0.3)	0.3	0.3	0.63 (0.26, 1.35)
	+	+	-	+	136 (8.0)	8.1	7.2	1.12 (0.94, 1.33)	501 (22.4)	22.3	22.3	0.99 (0.91, 1.08)
	+	-	-	+	39 (2.4)	2.3	1.8	1.23 (0.87, 1.80)	2 (0.1)	0.1	0.2	0.58 (0.07, 2.10)
One	+	-	+	-	182 (10.6)	10.8	13.6	0.79 (0.68, 0.92)***	32 (1.7)	1.4	1.6	0.91 (0.62, 1.28)
	-	+	+	-	15 (0.9)	0.9	1.4	0.63 (0.33, 1.05)**	5 (0.2)	0.2	0.2	1.11 (0.36, 2.60)
	-	-	+	+	139 (8.4)	8.2	10.1	0.81 (0.68, 0.96)**	45 (2.0)	2.0	2.0	0.98 (0.71, 1.31)
None	-	-	-	+	45 (2.5)	2.7	2.7	1.00 (0.73, 1.34)	0 (0.0)	0.0	0.0	-
	-	-	-	-	51 (3.0)	3.0	4.2	0.72 (0.51, 0.99)**	17 (0.7)	0.8	0.7	1.15 (0.67, 1.84)
	-	-	-	-	41 (2.5)	2.4	3.1	0.77 (0.55, 1.05)**	17 (0.7)	0.8	0.9	0.89 (0.51, 1.41)
None	-	-	-	-	12 (0.7)	0.7	0.8	0.86 (0.44, 1.50)	1 (0.05)	0.04	0.01	7.60 (0.10, 42.15)
	-	-	-	-	166 (10.5)	9.8	6.0	1.65 (1.41, 1.92)****	3 (0.1)	0.1	0.1	2.22 (0.45, 6.01)
None	-	-	-	-	38 (2.3)	2.3	1.9	1.21 (0.85, 1.67)	5 (0.2)	0.2	0.03	8.82 (2.84, 20.58)****

Healthy lifestyle factor present (+) or absent (-); \*\*P < 0.01, \*\*\*P < 0.05, \*\*\*\*P < 0.001, \*\*\*\*\*P < 0.0001 in  $\chi^2$ -test; A, no heavy alcohol consumption; B, normal body mass index; CI, confidence interval; E, expected; O, observed; P, regular physical activity; S, non-smoking.



## 정확도 높은 예측 모델링을 위한 도전 (3)

### 3. 스코어링

- 여전히 설명변수는 연구자에 의해 선택(이론 중심 모델=이미 잘 알려진 모델)
- 보다 많은 변수를 포함하면서 정확도 높은 모델을 만들 수 있을까?

#### Calculating Global CHD Risk in Men

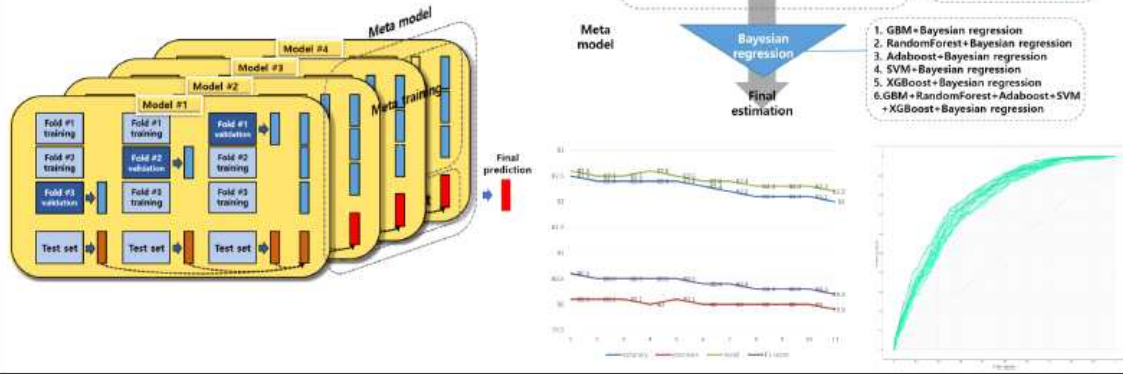
Step 1: Age				Step 5: Diabetes Mellitus		Step 8: CHD Risk			Score	Cases (n=230)	Controls (n= 230)
Years	Points	Years	Points	Present?	Points	Total points	10-year risk of CHD based on Step 2 (%)	10-year risk of CHD based on Step 8 (%)			
30 to 34	-1	55 to 59	4	No	0	-3	1	-	0-7	18	72
35 to 39	0	60 to 64	5	Yes	2	-2	2	-			
40 to 44	1	65 to 69	6			-1	2	2			
45 to 49	2	70 to 74	7			0	3	3			
50 to 54	3					1	4	4			
						2	5	5			
						3	6	6			
						4	7	7			
						5	8	8			
						6	11	10			
						7	14	13			
						8	18	16			
						9	22	20			
						10	27	25			
						11	33	31			
						12	40	37			
						13	47	45			
						14	56	53			
						15	66	63			
						16	77	74			
						17	89	85			
						18	102	97			
						19	117	111			
						20	134	127			
						21	153	145			
						22	174	165			
						23	197	187			
						24	232	221			
						25	270	258			
						26	312	299			
						27	358	345			
						28	408	395			
						29	462	449			
						30	520	507			
						31	582	569			
						32	648	635			
						33	718	705			
						34	792	779			
						35	870	857			
						36	952	939			
						37	1038	1025			
						38	1128	1115			
						39	1222	1209			
						40	1320	1307			
						41	1422	1409			
						42	1528	1515			
						43	1638	1625			
						44	1752	1739			
						45	1870	1857			
						46	1992	1979			
						47	2118	2105			
						48	2248	2235			
						49	2382	2369			
						50	2520	2507			
						51	2662	2649			
						52	2808	2795			
						53	2958	2945			
						54	3112	3099			
						55	3270	3257			
						56	3432	3419			
						57	3598	3585			
						58	3768	3755			
						59	3942	3929			
						60	4120	4107			
						61	4302	4289			
						62	4488	4475			
						63	4678	4665			
						64	4872	4859			
						65	5070	5057			
						66	5272	5259			
						67	5478	5465			
						68	5688	5675			
						69	5902	5889			
						70	6120	6107			
						71	6342	6329			
						72	6568	6555			
						73	6808	6795			
						74	7062	7049			
						75	7320	7307			
						76	7582	7569			
						77	7848	7835			
						78	8118	8105			
						79	8392	8379			
						80	8670	8657			
						81	8952	8939			
						82	9248	9235			
						83	9548	9535			
						84	9852	9839			
						85	10160	10147			
						86	10472	10459			
						87	10788	10775			
						88	11108	11095			
						89	11432	11419			
						90	11760	11747			
						91	12092	12079			
						92	12428	12415			
						93	12768	12755			
						94	13112	13099			
						95	13460	13447			
						96	13812	13799			
						97	14168	14155			
						98	14528	14515			
						99	14892	14879			
						100	15260	15247			



## 정확도 높은 예측 모델링을 위한 도전 (4)

### 4. 머신러닝

- 보다 더 정확하면서 해석까지 가능한 머신러닝 모델이 있을까?

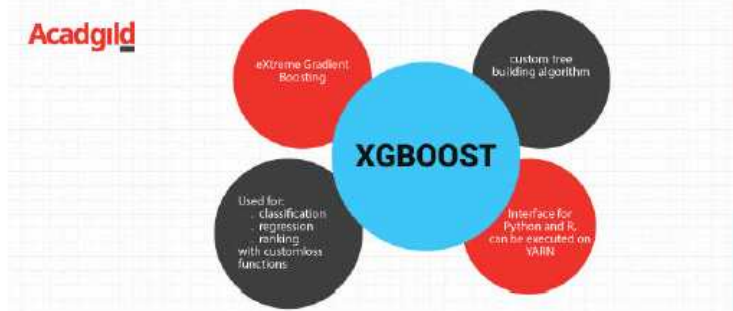


## 어떤 알고리즘이 가장 우수한가?

- Data Robot을 통한 정확도 상위 12위의 알고리즘

(2020)

- Light GBM 3.40
- XG Boost 3.50
- RuleFit 3.68
- RandomForest 5.06
- GLM 5.11
- SVM 5.44
- ExtraTrees 6.39
- GBM 6.40
- KNN 8.87
- Vowpal Wabbit 8.93
- Tensorflow 9.71
- Decision Tree 10.15



출처:

<https://blog.naver.com/nextdoor>



## 고위험군 예측을 위한 인공지능 알고리즘

- 인공지능? 머신러닝? 딥러닝?

Artificial Intelligence

### 인공지능

사고나 학습 등 인간이 가진 지적 능력을 컴퓨터를 통해 구현하는 기술



Machine Learning

### 머신러

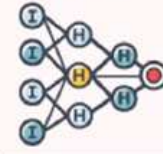
컴퓨터가 스스로 학습하여 인공지능의 성능을 향상시키는 기술 방법



Machine Learning

### 딥러닝

인간의 뉴런과 비슷한 인공신경망 방식으로 정보를 처리



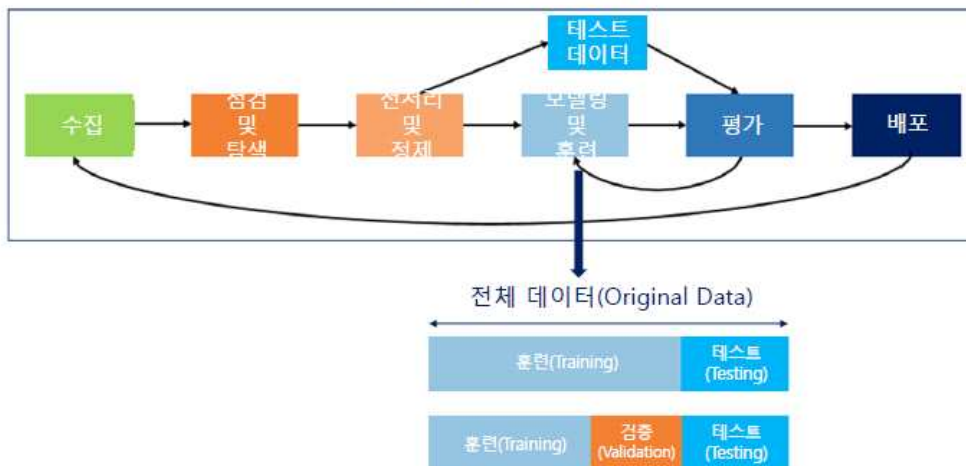
출처:

(인사이드) <https://m.blog.naver.com/jw6971/221614497987>



## 인공지능을 이용한 질병 예측의 절차

### 머신 러닝 워크플로우



출처:

(인사이드) <https://wikidocs.net/31947>

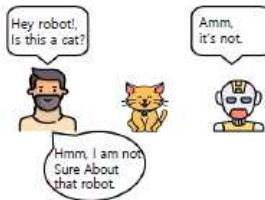


## 하이브리드 머신(모델)의 등장

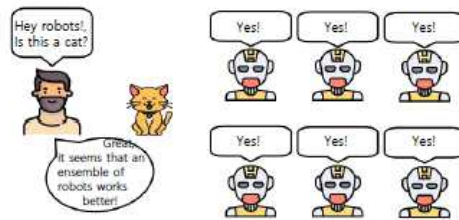
- 왜 하나의 AI로만 질병을 예측할까?
- 축구심판이 1명이 아니라 10명이라면?



## 전통적인 머신러닝 학습과 하이브리드 학습의 차이



a) Traditional Learning



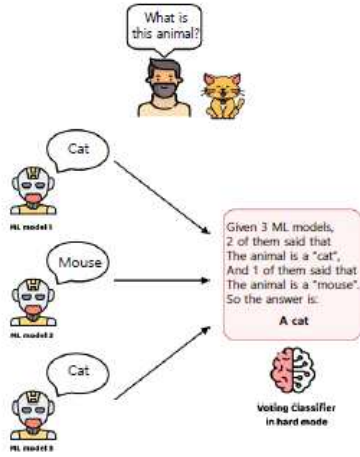
b) Ensemble Learning



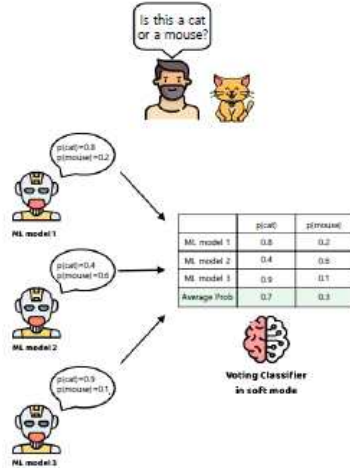


## 전통적인 머신러닝 학습과 하이브리드 학습의 차이

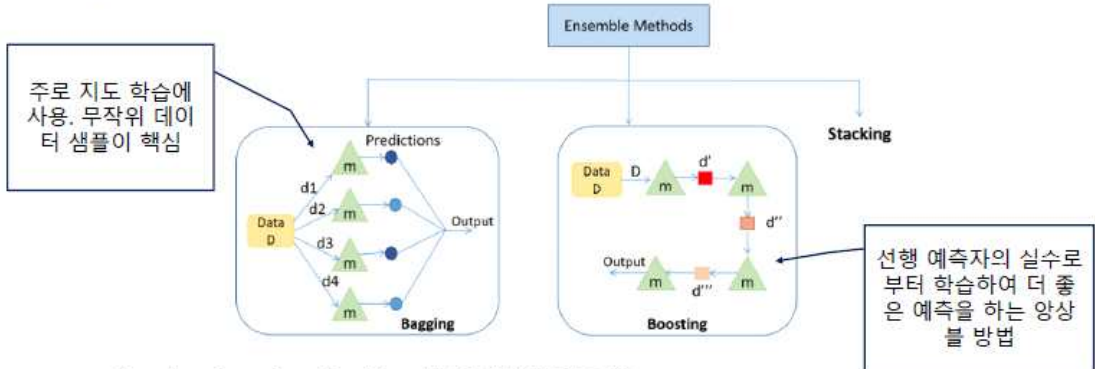
### Voting의 예



### Soft mode의 예



## 하이브리드 학습(Ensemble learning)

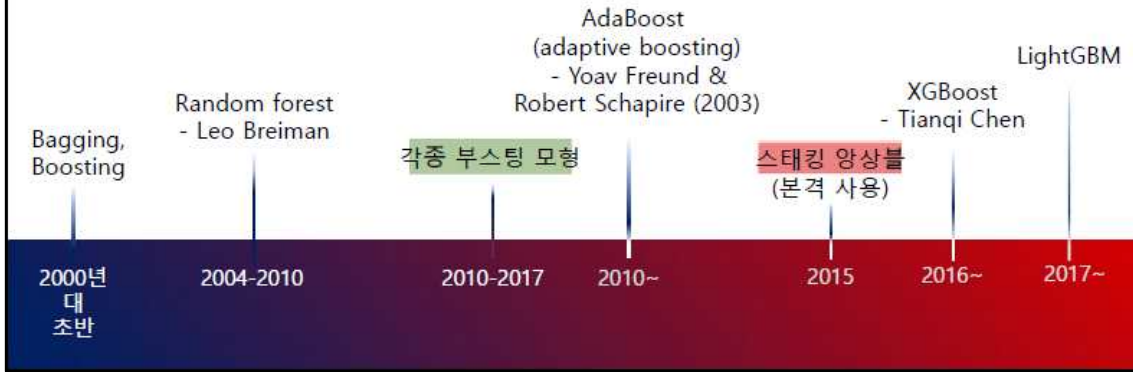


- Bagging, Boosting, Stacking 세 가지 방법이 있음
- $m$  = 약한 학습자
- $d_1, d_2, d_3 = d$ 의 무작위 샘플
- $d', d'', d''' =$  이전의 약한 학습기의 결과를 기반으로 업데이트된 훈련 데이터



## 앙상블 모델

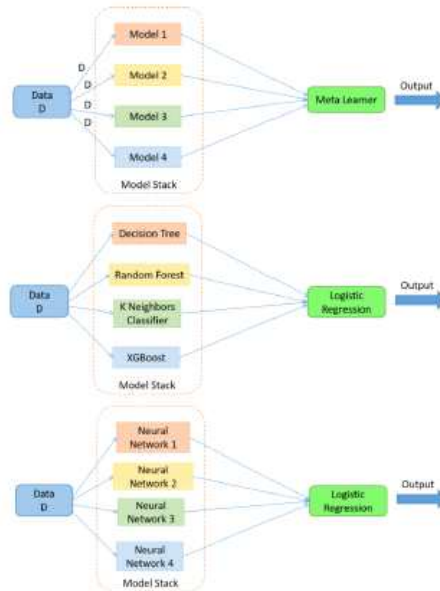
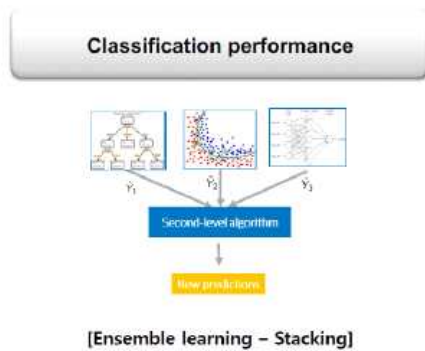
- 앙상블 모델의 유행(경향)



## 앙상블 모델

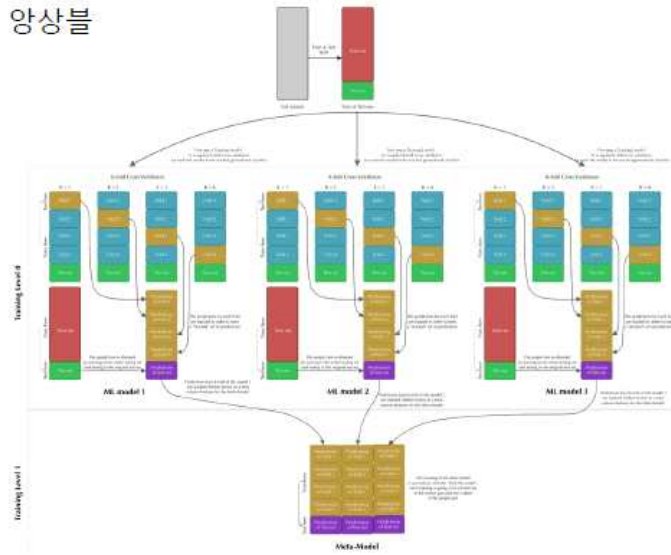
- 적층 앙상블

- Bagging, Boosting 외에 적층(Stacking)이라는 방법이 있음



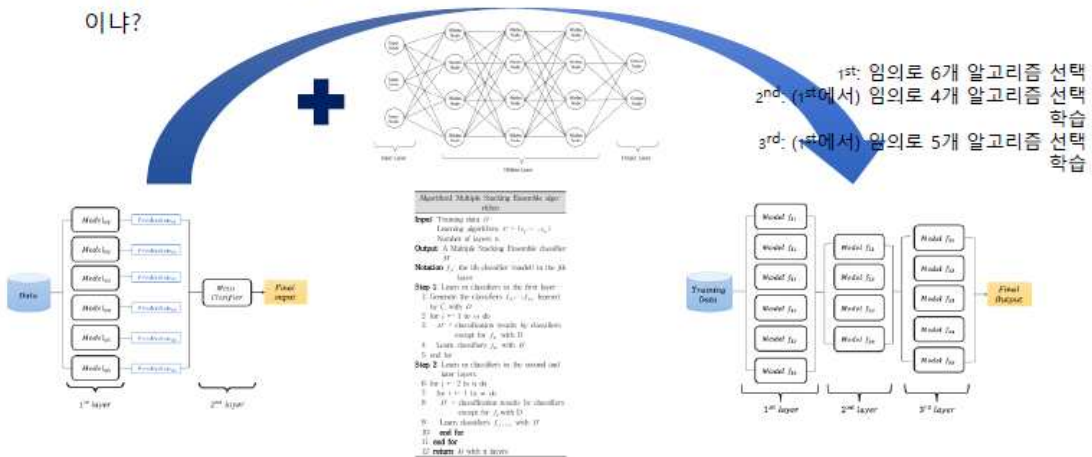


## 앙상블 모델 - 적층 앙상블



## 비교적 최근의 적층 앙상블 (하이브리드 앙상블) 방법

- 딥러닝의 구조를 모방함
- 핵심은 적층구조 - 병렬이나, 직렬 (10) 이냐?







optimal model selection: How will you choose?



Goals:  
Minimal acceleration time  
Maximal range



Size



Power



Capacity



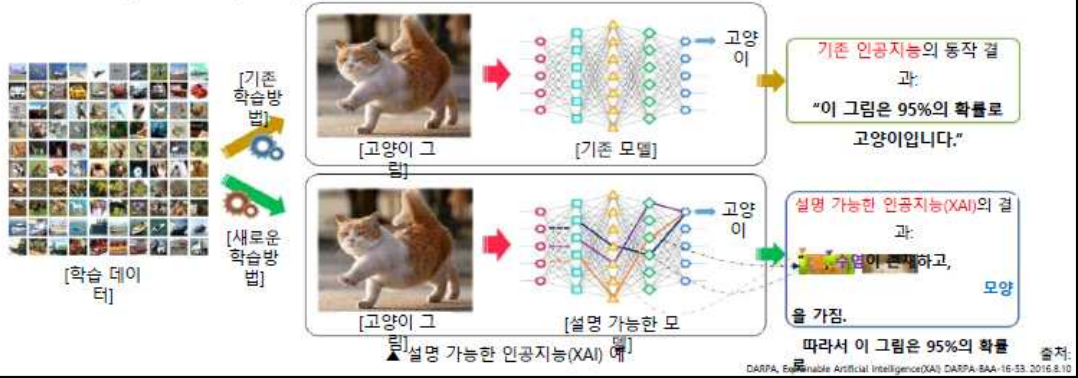
We chose the best model. next? : Explanation





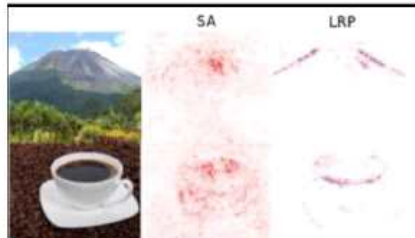
## 전통적 딥러닝에서의 해석력 문제

- 이유: 오류 역전파(back propagation)로 인한 블랙박스(Black-Box) 구조이기 때문  
-> 학습 과정에서의 네트워크를 알 수 없음
- 해석가능 모델 (Interpretable model)
- 설명가능한 모델(X-AI)

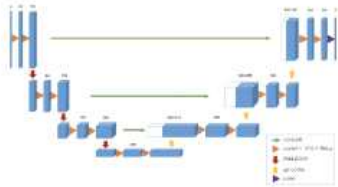


## X-AI의 예: 이미지연구

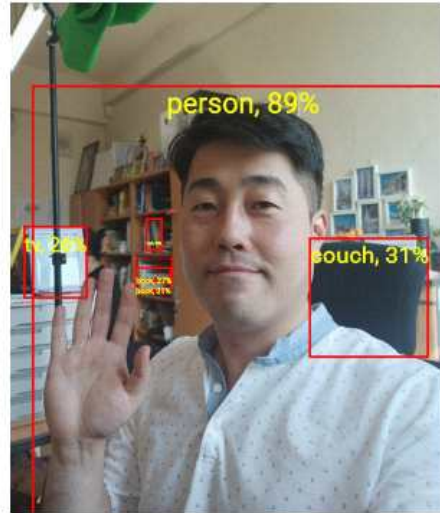
- 보편적 방법의 합의
  - 히트맵(heat map) 기반의 설명가능한 AI 방법: 딥러닝 블랙박스 모델은 그대로 유지하면서 예측 결과를 토대로 거꾸로 입력 이미지의 요소를 강조하도록 시각화하여 이미지의 어떤 부분이 결과 도출에 영향을 미쳤는지를 알아내는 방법
  - 히트맵 방법의 사례: 민감도 분석(Sensitivity analysis, SA), 계층별 관련성 전파(Layer-wise relevance propagation, LRP)

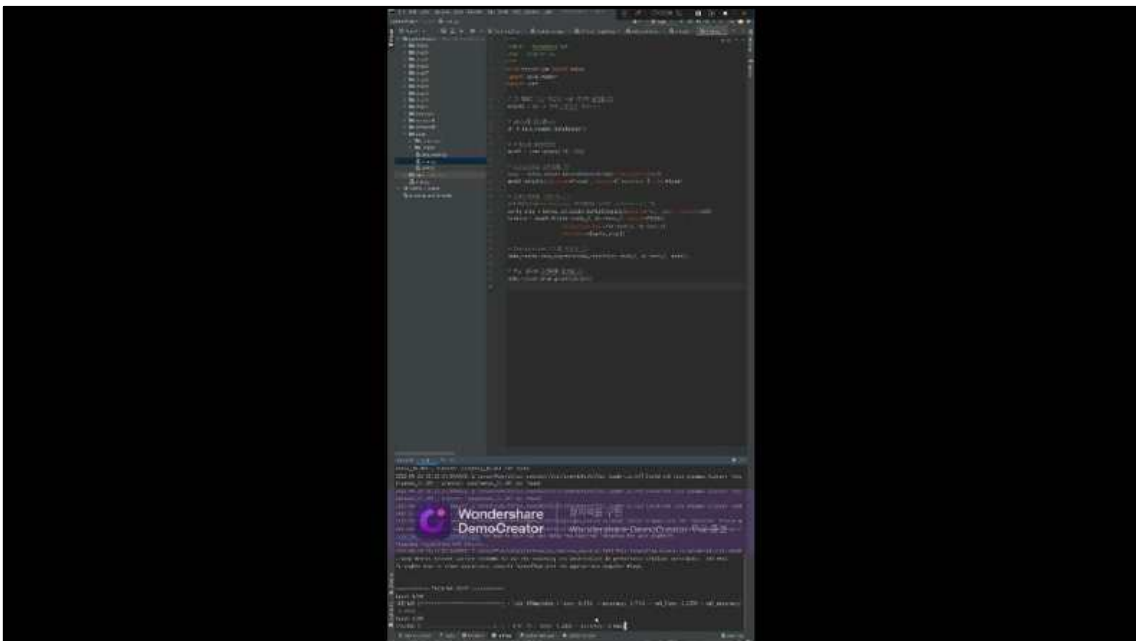
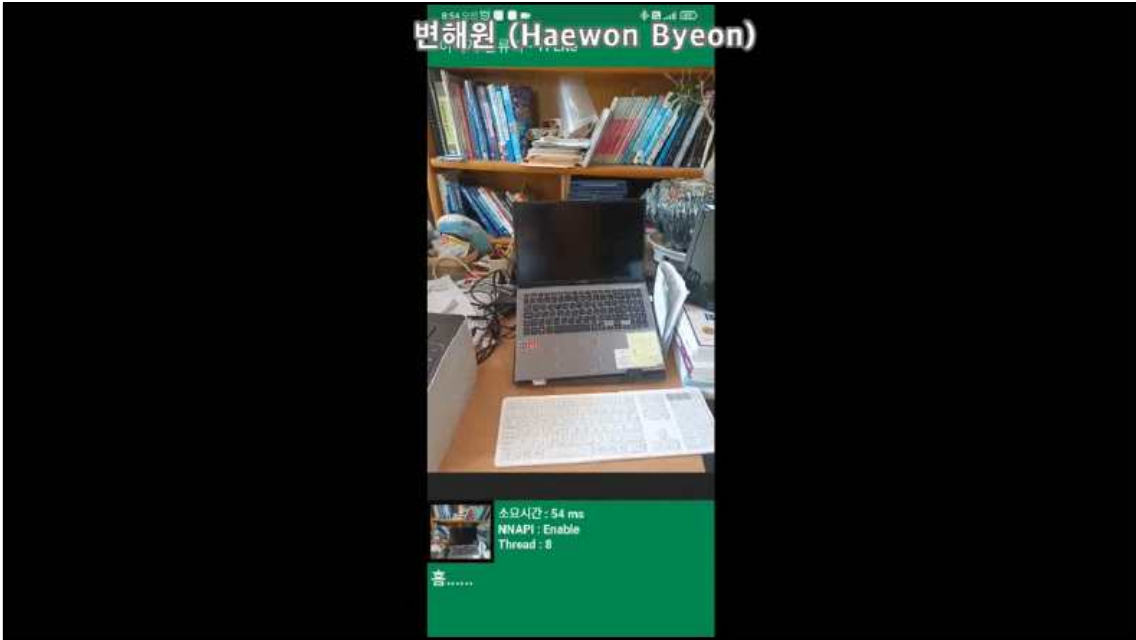


- 가장 최신 영상(이미지) 인식 기법=semantic segmentation

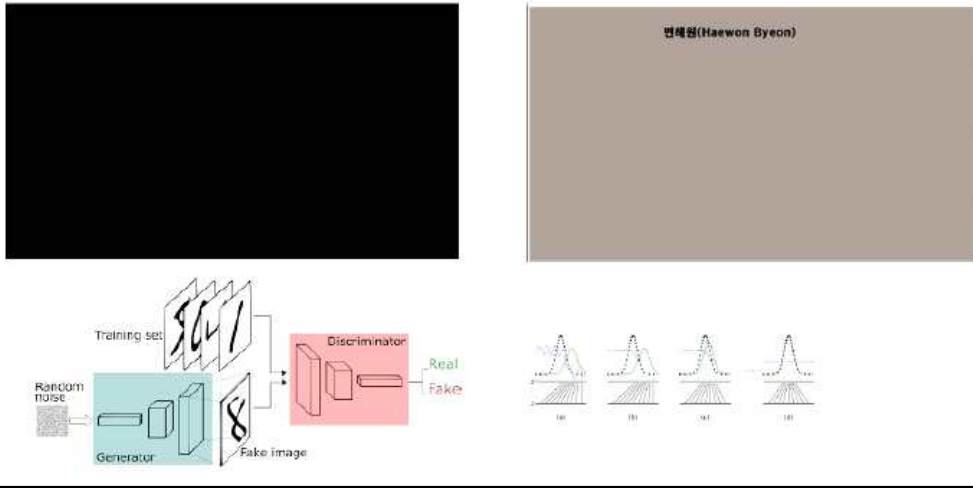


- Unet 응용 semantic segmentation  
-> Fast R-CNN (YOLO 기반)
- GAN 응용

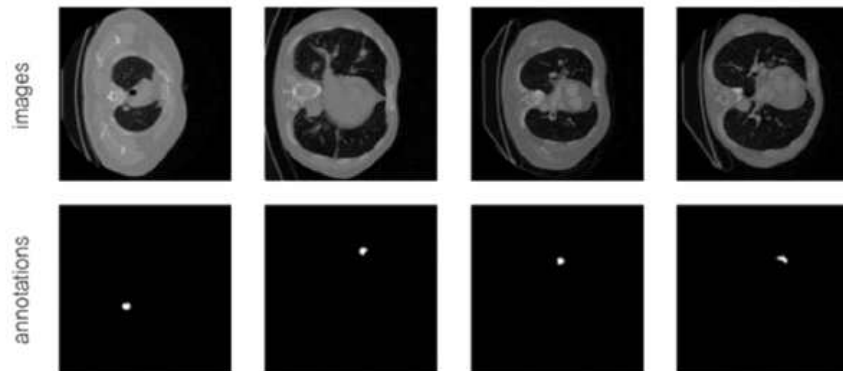




- GAN 응용



- 복부 CT 분석을 통한 암 진단
- 시맨틱 세그멘테이션 응용한 방식









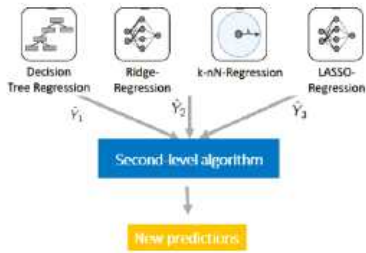
## X-AI의 예: 정형데이터

- DARPA(미국 국방성 산하 국방위고등연구계획국): 2018년부터 진행하고 있는 XAI 학습 모델 개발 및 테스트에 관한 연구에서 다음의 개념이 파생됨
  - 심층설명학습 (learning deep explanation)
  - 해석 가능한 모델(more interpretable model)
  - 귀납적 모델(Model Induction)
- 대표적인 예시: 베이지안 규칙 리스트(Bayesian Rule Lists, BRL)
  - > 고차원, 다변량 특징 공간을 단순한 if-then 문(statement)의 연속인 결정 리스트(Decision lists)로 변화하여
    - 결정문들을 쉽게 해석 가능하도록 만드는 방법



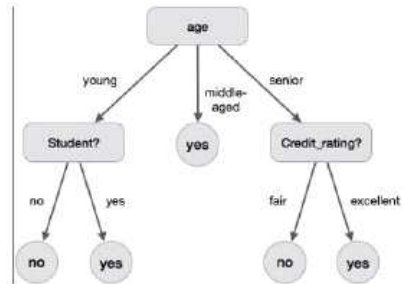
## X-AI의 예: 정형데이터

### Classification performance



[Ensemble learning – Stacking]

### Extract effective diagnostic rules



[Rule extraction method]

과학기술정보통신부 인증 **인제메디컬빅데이터 연구실**  
(**INJEM** lab)

<https://sites.google.com/view/injem>)

감사합니다





# 특강 IV

## 신진학자 초청강연

좌장: 정현성(한국교원대)

Using neurophysiological data to understand  
speech-recognition difficulties  
(송지은, 한국과학기술원)



## Using neurophysiological data to understand speech-recognition difficulties

Jieun Song

School of Digital Humanities & Computational Social Sciences,  
Korea Advanced Institute of Science and Technology  
jsong@kaist.ac.kr

Electroencephalography (EEG) is a technique that measures electrical activity of the brain, and one of the biggest advantages of using EEG for speech research is that it can allow us to examine different processes of speech recognition as they unfold in real time. Using both EEG and behavioural methods, I have investigated a variety of speech recognition problems in realistic listening situations - those arising when understanding speech in different types of noise, second-language speech, or synthetic speech. I'll talk about how I have examined these issues in terms of auditory tracking of speech, lexical-semantic processing and attention-related processes.

One of the challenges when investigating these issues using EEG is that these studies need a large number of sentence stimuli. I'll introduce a project on developing a large-scale set of Korean sentences with varying degrees of semantic predictability. A large stimuli set like this will be valuable for speech research, especially for studies with several experimental manipulations. For validating final-word probabilities of the sentences, I used a new method of combining results of cloze tests by (human) subjects with those of a state-of-the-art language model. I hope my research will help understand some of the unanswered questions on how listeners cope with speech recognition difficulties and open up new investigations in speech research.



구두 발표Ⅱ  
음성학 및 음운론 2

좌장: 김지은(가톨릭관동대)



## A study on the predictability of acoustic power distribution of English speech for the English academic achievement in a science academy

Soon Park, Hyunkee Ahn

Dept. of English Language Education, Seoul National University  
drspark@snu.ac.kr, ahnhk@snu.ac.kr

The average acoustic distribution of American English speakers was statistically compared with the English-speaking patterns of gifted students in a Science Academy in Korea. By analyzing speech recordings, the duration time of which is much longer than in previous studies, this research identified the degree of acoustic proximity between the two parties and the predictability of English academic achievement of gifted high school students. Long-term spectral acoustic power distribution vectors were obtained for 2,048 center frequencies in the range of 20 Hz to 20,000 Hz by applying an long-term average speech spectrum (LTASS) MATLAB code. Three more variables were statistically compared to discover additional indices that can predict future English academic achievement: the receptive vocabulary size test, the cumulative vocabulary scores of English formative assessment, and the English Speaking Proficiency Test scores. Linear regression and correlational analyses between the four variables showed that the receptive vocabulary size test and the low-frequency vocabulary formative assessments which require both lexical and domain-specific science background knowledge are relatively more significant variables than a basic suprasegmental level English fluency in the predictability of gifted students' academic achievement.



## 발화 속도 통제 상태에서의 영어 능숙도 평가: 한국인 아동 영어 학습자 발화를 중심으로

최 나 래, 장 태 업  
한국외국어대학교 영어학과

### Proficiency Evaluation under Controlled Speech Rate Condition: Focusing on Korean Children Learners' English Utterances

Narah Choi, Tae-Yeoub Jang  
Dept. of English Linguistics, Hankuk University of Foreign Studies  
nrchoi993@hufs.ac.kr, tae@hufs.ac.kr

제2 언어 학습자가 산출한 음성언어는 원어민보다 느리게 발화되며, 학습자의 능숙도가 낮을수록 발화 속도도 느려지는 상관관계가 있음이 여러 연구를 통해 밝혀졌다(Derwing & Munro, 1997; Huang and Gráf, 2020). 이를 바탕으로, L2 말하기 교육에서 발화 속도를 증가시키는 훈련을 통해 능숙도를 향상시킬 수 있음도 밝혀졌다(Arevart & Nation, 1991). 또한, 발화 속도는 비교적 측정이 쉬우므로 많은 제2 언어 말하기 또는 발음정확성 측정을 위한 자동평가시스템에서 가장 중요한 특징(feature) 중 하나로 사용된다(Cucchiari et al. 2000, 2002, Wet et al. 2009). 그러나 능숙도의 판정을 발화 속도에 지나치게 의존하면 분절음 및 운율 특징과 같은 언어의 본질적 특성을 반영하지 못하는 근본적인 한계에 직면하게 된다. 이에 본 연구에서는 발화 속도라는 특성이 통제된 L2 음성을 인간이 평가하여도 능숙도를 구분해낼 수 있는지에 대한 여부를 알아보기 위한 인지 실험을 진행하였다.

실험에 사용된 음성 자료는 KSEC 말뭉치(이석재 외, 2003)에서 제공하는 한국인 아동 학습자 발화이며, 사전 실험인 1차 실험은 각 발화 토큰에 대해 발화 속도가 통제되지 않은 한국인 아동 학습자의 발화를 듣고 영어 능숙도를 평가하였다. 이 평가를 위해 영어학 또는 영어교육 전공 대학원생 5인이 참여하였고, 이들의 평가 결과에 따라 능숙도가 높은 아동 4인과 능숙도가 낮은 아동 5인을 각각 선정하였다. 두 집단 간 유의미한 발화 속도 차이가 나타났으며, 리듬 수치 중 %V와 VarcoV만이 기존 연구 결과를 뒷받침하였다.

본 실험인 2차 인지 실험에서는 1차 실험과 같은 음성 자료를 사용하였지만 각 음성토큰의 발화속도를 동일하게 조작하여 제시하였다. 이 실험에서는 1차 인지 실험 참여자 5인 및 현직 중·고등학교 교사 5인이 추가로 참여하였다. 실험 참가자들은 1차 실험과 마찬가지로, 동일한 빠르기로 조정된 한국인 아동 학습자의 영어 발화를 듣고 각 화자의 능숙도를 평가하였다. 능숙도 평가 이후, 평가자들은 표본으로 선정한 일부 영어 문장 발화를 듣고 음원의 속도가 인위적으로 조작되었는지 여부를 판단하는 추가 실험도 진행하였다. 2차 실험의 결과, 평가자는 화자의 능숙도 차이를 1차 실험과 마찬가지로 구분하였으나, 발화 속도가 조정된 발화와 그렇지 않은 발화를 구분하지는 못한 것으로 나타났다.

본 연구에서 제시한 결과는 추후 발음 평가시스템의 교육적 활용에 유의미한 방향성을 제시하는 한편, 실제 영어 교육 현장에서도 학습자의 능숙도를 분석하고 피드백을 제공하는 데 도움이 될 것으로 기대된다.

## 참고문헌

- Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal*, 22(1), 84-94.
- Cucchiari, C., Strik, H., & Boves, L., (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 107(2), 989-999.
- Cucchiari, C., Strik, H., & Boves, L., (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, 111(6), 2862-2873.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 19, 1-16.
- Huang, L. & Gráf, T. (2020). Speech Rate and Pausing in English: Comparing learners at different levels of proficiency with native speakers. *Taiwan Journal of TESOL*, 17(1), 57-86.
- Wet, F., Walt, C., & Niesler, T. R. (2009). Automatic assessment of oral language proficiency and listening comprehension. *Speech Communication*, 51, 864-874.
- 이석재, 이숙향, 강석근, 이용주. (2003). 한국인의 영어 음성 코퍼스 설계 및 구축. *말소리*, 159-174.

## 유아 파닉스 프로그램의 효과성 연구

정 현 성  
한국교원대학교 영어교육과

### A Study on the Effectiveness of A Phonics Program for Children

Hyunsong Chung  
Dept. of English Education, Korea National University of Education  
hchung@knue.ac.kr

본 논문은 한 민간 어린이집 영어 교육 프로그램의 일환으로, 체계적 파닉스에 기반해 개발된 파닉스 프로그램의 효과성을 검증하는 연구이다. 파닉스 프로그램을 적용한 후 만 5세 유아들을 대상으로 음운 및 음소 인식과 음소 발화, 철자 인식 및 발화, 어휘 인식 및 발화, 조음 능력에 미치는 효과성과 정의적인 측면의 효과성을 알아보았다.

검사 도구는 끝소리 대응 과업 검사, 첫소리 대응 과업 검사, 첫소리 발화 검사, 철자 인식 검사, 철자 이름 및 음가 말하기 검사, 어휘 인식 검사, 어휘 발화 검사, 조음 능력 검사, 정의적 영역 검사를 시행하였다. 수업 관찰도 병행하였다.

효과성을 검증한 결과 유아들의 음운 및 음소 인식, 음소 발화에 통계적으로 유의미한 긍정적 효과를 보였지만, 음소 발화의 경우 낮은 수준의 향상을 보였다. 유아들의 철자 인식 및 발화에는 부분적으로 긍정적인 효과를 보여주었지만, 철자와 철자 이름, 철자와 해당 음가를 좀 더 잘 연결해 인식할 수 있도록 프로그램의 개선이 필요한 것으로 드러났다. 유아들의 어휘 인식 및 발화에도 본 프로그램이 긍정적인 효과를 보이는 것으로 나타났지만, 어휘 발화의 경우 평균 점수가 매우 낮아 이를 개선할 수 있는 프로그램을 추가할 필요가 있었다.

유아들의 조음 능력 향상에도 긍정적인 효과를 보여주었지만, 우리말의 음운 구조로 인한 모국어 간섭 현상의 영향을 받는 유아들이 적지 않아, 모국어의 간섭 현상을 줄일 수 있도록 단어 차원에서 정형적인 발음에 더 노출 시켜 줄 필요가 있었다. 정의적인 측면에서는 긍정적인 효과를 보여주지는 못했는데, 특히 자신감과 참여도에서 점수가 높지 않은 만큼 유아들이 좀 더 적극적으로 참여할 수 있는 프로그램의 개선이 필요한 것으로 나타났다.

수업 관찰 결과, 유아들이 철자 모양에 민감하게 반응하고, 손가락으로 글자 모양을 표현하는 등 정의적 영역 검사 결과와는 달리 적극적으로 참여하는 모습을 보였다. 하지만, 몇몇 유아에게 한정되어 표현되고 있기 때문에, 전체 유아가 적극적으로 파닉스 활동에 참여할 수 있는 방법을 모색해야 할 것으로 보인다.

# 구두 발표 II

## 음성공학 2

좌장: 유하진(서울시립대)



## Speech self-supervised learning을 활용한 단일 화자 한국어 합성음성 탐지

엄 영 식, 김 회 린  
카이스트 전기및전자공학부

### Single-Speaker Korean Synthetic Speech Detection Using Speech Self-Supervised Learning

Youngsik Eom, Hoirin Kim  
School of Electrical Engineering, KAIST  
iron5605@kaist.ac.kr, hoirkim@kaist.ac.kr

최근 발전된 형태의 text-to-speech(TTS)와 voice conversion(VC) 알고리즘으로부터 생성된 합성음성들은 정교하여 실제 음성과 그 차이를 느끼기 어렵다. 따라서 정교한 음성합성을 이용한 보이스피싱 등 적대적 활용에 대비하기 위한 합성음성 탐지 시스템 개발의 중요성이 대두되고 있다. 이러한 합성음성들은 매우 다양한 종류의 알고리즘으로 생성되기 때문에, 모든 가능한 합성 시스템에 대하여 훈련하는 것은 현실적이지 않다. 따라서, 합성음성 탐지 시스템을 구현함에 있어서 제한된 훈련 데이터로부터 보지 못한 합성 시스템의 음성에 잘 일반화하는 능력이 중요하다.

본 연구에서는 speech self-supervised learning으로 기학습된 Wav2vec 2.0 모델의 representation을 활용한 한국어 합성음성 탐지 시스템을 제안한다. 레이블이 없는 데이터셋을 활용하여 self-supervised learning으로 기학습된 모델을 task-specific한 projection head와 함께 합성음성 탐지 데이터셋으로 fine-tuning 하였다.

실험은 한국어 TTS, 보코더의 5가지 조합들로 생성한 2,500 문장의 단일 여성 화자 합성음과 동일한 화자의 진짜 음성 2,500 문장으로 구성된 합성음성 탐지 데이터셋을 이용하여 진행하였다. 총 5가지의 조합 중 4가지 조합의 데이터로 fine-tuning을 진행하였고, 훈련 데이터를 제외한 나머지 1가지 조합의 데이터에 대하여 테스트를 진행한 후 시스템의 성능을 평가하는 5-fold cross validation을 진행하였다. 본 연구에서는 speech self-supervised learning으로 기학습된 representation의 효과를 확인하기 위해, 서로 다른 데이터셋으로 기학습된 Wav2vec 2.0 모델들을 fine-tuning 하여 성능 변화를 비교하였다. 실험 결과, self-supervised learning을 진행한 뒤 fine-tuning 했을 경우 self-supervised learning을 진행하지 않았을 때보다 동일 오류율과 정확도 측면에서 우수한 성능을 보임을 확인하였다. 특히, self-supervised learning을 진행하지 않았을 경우 Tacotron+Griffin-Lim 조합의 합성 시스템에 대하여 일반화 성능이 저조함을 확인할 수 있었다. 이는 Griffin-Lim 보코더가 다른 보코더와는 달리 심층 신경망 기반이 아니기 때문에 추론 시 일반화에 어려움을 겪는 것으로 분석할 수 있다. 하지만, self-supervised learning을 진행한 모델은 위 경우에 대해서도 일반화 성능이 우수함을 확인하였다. 또한, Librispeech-960h과 Librispeech-960h + Ksponspeech-965h로 각각 기학습된 모델의 결과가 거의 유사하게 우수함을 확인하였다. 이를 통해, self-supervised learning으로 기학습된 representation이 한국어 합성음성 탐지에 있어 효과적이며, 영어 데이터로만 기학습된 representation 또한 한국어 합성음성 탐지에 효과가 있음을 확인하였다.<sup>1)</sup>

<감사의 글> 본 연구는 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00575, 음성·텍스트 딥러닝 기반 보이스피싱 예방 기술 개발).

## 한국어 음성 전사를 위한 wav2vec2.0 훈련 단위

이 선 우<sup>1</sup>, 김 선 희<sup>2</sup>, 정 민 화<sup>1</sup>  
<sup>1</sup>서울대학교 언어학과, <sup>2</sup>서울대학교 불어교육과

### Linguistic Unit to train wav2vec2.0 for phonetic transcription of Korean

Seonwoo Lee<sup>1</sup>, Sunhee Kim<sup>2</sup>, Minhwa Chung<sup>1</sup>  
<sup>1</sup>Dept. of Linguistics, Seoul National University  
<sup>2</sup>Dept. of French Language Education, Seoul National University  
lsw5220@snu.ac.kr, sunhkim@snu.ac.kr, mchung@snu.ac.kr

Self-supervised Learning (SSL) 방법론의 등장으로 자동음성인식 기술 기반의 음성전사 (phonetic transcription) 및 발음오류 감지 연구가 활발하게 진행되고 있다. 하지만, 음향 모델을 훈련에 사용하는 언어 단위에 대한 연구는 아직 이루어진 바가 없다. 따라서 본 연구에서는 한국어 사용 아동의 음성 데이터로 다국어 음성으로 프리트레이닝 된 wav2vec2.0-xlsr-53 모델을 파인튜닝하여 아동의 한국어 발화를 음성전사하기 위한 최적의 단위를 찾고자 하였다.

AI-hub에서 공개한 아동의 자유발화 코퍼스에서 3세에서 5세 아동의 음성 약 240시간을 이용하여 모델을 훈련하였으며, 훈련에 사용한 훈련 단위는 음절, 음소, 조음위치 및 조음방법의 조음자질 수준이었다. 조음위치는 각 음소를 자음의 조음위치와 모음 조음에서의 혀의 높이로 매핑하였고 조음방법은 각 음소를 자음의 조음방법고 모음 조음에서의 혀의 앞뒤 위치로 매핑하여 변환하였다.

모델 훈련 결과, 음절(3.84%), 조음위치(7.34%), 조음방법(7.53%), 음소(9.50%) 순으로 단위인식오류율 (UER)이 낮았다. 향후 연구에서는 조음자질 수준을 다양하게 변형하여 자음과 모음의 어떠한 측면이 함께 학습되어 성능을 향상시킬 수 있는지 파악하여 발음오류 감지 모델에 적용하고, 해당 방법론을 성인의 음성 데이터에도 적용하여 아동 발화의 음성 전사 정확도를 증진할 예정이다.

## 화자확인을 위한 음성 세그먼트 길이 임베딩

배재현<sup>1</sup>, 박순찬<sup>1</sup>, 김경화<sup>2</sup>, 김형순<sup>1</sup>  
<sup>1</sup>부산대학교, <sup>2</sup>대검찰청

### Speech Segment Duration Embedding for Speaker Verification

Jaehyoun Bae<sup>1</sup>, Sunchan Park<sup>1</sup>, Kyung-Wha Kim<sup>2</sup>, Hyung Soon Kim<sup>1</sup>

<sup>1</sup>Dept. of Electronics Engineering, Pusan National University,

<sup>2</sup>Supreme Prosecutors' Office

juliansalt@pusan.ac.kr, sunchanpark@pusan.ac.kr, savoix@spo.go.kr, kimhs@pusan.ac.kr

화자인식 모델에서는 일반적으로 입력 발화를 짧은 길이의 동일한 세그먼트 단위로 분할해서 학습함으로써 효율적으로 GPU를 사용하도록 한다. 이 방식은 짧은 길이의 평가 발화에 대해서 성능향상을 꾀한다는 이점이 있지만, 실제 발화는 일반적으로 이 길이보다는 길다. 이 문제를 보완하기 위해 입력 발화를 상대적으로 더 긴 길이의 세그먼트로 분할하여 학습했을 때 19%의 성능향상이 됨을 확인하였다.

본 연구는 이에 더 나아가서 실제 발화와 비슷한 길이의 평가 발화에 더 잘 대응할 수 있도록 ECAPA-TDNN 방식[1]에 음성 세그먼트 길이에 대한 임베딩 층을 추가한 화자확인 시스템을 제안한다. 제안한 시스템은 입력 발화를 특정 범위 내에서 랜덤한 길이를 가지는 세그먼트로 분할하여 학습하며, 모델 내 임베딩은 학습가능한 두 임베딩의 음성 세그먼트 길이에 따른 가중합으로 생성되어 기존 화자인식 모델의 특정 계층에 더해진다. 성능평가를 위해 한국어 표준음성 DB[2]를 데이터셋으로 사용하여 일부는 미세조정 추가학습용으로, 다른 일부는 화자확인 평가용으로 나누었고, VoxCeleb DB[3]로 사전훈련한 ECAPA-TDNN 모델을 SpeechBrain 오픈소스 음성 툴킷에서 가져와 화자 인식기의 사전학습모델로 사용하였다. 실험 결과, 제안한 시스템은 기존의 짧은 세그먼트 단위로 분할하여 학습했을 때보다 38%의 성능이 향상되고, 긴 길이의 세그먼트 단위로 분할하여 학습했을 때보다는 24%의 성능이 향상됨을 확인하였다.

이 논문은 2022년 대검찰청 연구용역의 일부로 수행되었음.

#### 참고문헌

- [1] B. Desplanques, J. Thienpondt and K. Demuynck. (2020). "ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification," in *Proc. Interspeech 2020*, 3830-3834.
- [2] 신지영, 장혜진, 강연민, 김경화. (2015). "한국인 표준 음성 DB 구축," *말소리와 음성과학*. 139-150.
- [3] A. Nagrani, J. S. Chung, and A. Zisserman, (2017). "VoxCeleb: A large-scale speaker identification dataset," in *Proc. Interspeech 2017*, 2616-2620.





구두 발표Ⅱ  
음성학 및 음운론 3

좌장: 이주경(서울시립대)



## Temporal Variation in Intergestural Timing of CVC Gestures in Korean Homophones

Suyeon Im<sup>1</sup>, Sahyang Kim<sup>2</sup>, Taehong Cho<sup>1,3</sup>

Hanyang Institute for Phonetics and Cognitive Sciences of Language, Hanyang University<sup>1</sup>,

Dept. of English Education, Hongik University<sup>2</sup>,

Dept. of English Language & Literature, Hanyang University<sup>3</sup>

suyeonim@hanyang.ac.kr, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Articulatory gestures are realized with some degree of invariance to encode phonological/lexical information, but they should also accommodate some degree of variability arising with other factors that affect speech production. This study investigates how gestures in homophonous words in Korean (CVL words with a liquid coda; CVN words with a nasal coda) are temporally realized and how their temporal variability is manifested in interaction with factors such as gestural composite of coda, lexical frequency, and focus. Results showed that these factors induced no difference in intergestural timing itself, but that they affected temporal variability. First, a liquid coda showed more variability compared to a nasal coda, indicating that a coda with more complex gestures induces temporal instability. Second, homophones with high lexical frequency showed more stable intergestural timing than their low frequency counterparts, indicating a possible motor-practice effect. These two factors further interacted with focus, showing reinforced effects under focus. Finally, intergestural timing was more variable under focus, possibly due to a larger degree of freedom in temporal stretching associated with focus-induced hyperarticulation. The results imply that temporal variability encodes the complex interplay between various factors that all contribute to forming a linguistic message to be available to the listener.

## Prosodic Marking of Second Occurrence Focus in Korean

Seong Eun Park, Gayeon Son  
Dept. of Eng. Lang. & Lit, Chungbuk National University,  
Dept. of Eng. Lang. & Industry, KwangWoon University  
seongpark@chungbuk.ac.kr, gson@kw.ac.kr

This study investigates the intonational expression of second occurrence (SO) focus in Korean and its implications on the semantics-phonology interface. SO focus is a repeated focused item within the scope of a focus sensitive operator such as *only*, which is preceded by another newly focused item. While focused items are normally marked with higher pitch, longer duration and more energy, SO focus is known to be marked with less or little prosodic prominence: previous studies show that SO focus in English is marked with longer duration and more intensity and SO focus in German with longer duration and optional higher pitch (Beaver et al. 2007; Howell 2009; Ishihara and Féry 2006). However, phonological marking of SO focus in Korean, whose prosodic structure is significantly different from that of English or German, has not been measured yet. In this study, therefore, two types of discourse data containing SO focus items produced by 10 Korean speakers were analyzed: one with a focus-sensitive operator *-man* and the other without it (question-answer pairs). The statistical results showed that duration, intensity, and F0 at focus, SO focus, and nonfocus positions did not have significant differences. However, depending on the existence of *-man*, F0 differences among the different focus positions were reported: when *-man* was not used, F0 values significantly varied according to the focus position. Phonetic realization of SO focus is susceptible to multifarious linguistic features within sentences and a substantial source of evidence for the interfaces between grammatical domains.

## Corpus-based study of Children's Accentual Phrases

Tae-Jin Yoon

Dept. of English Language and Literature, Sungshin Women's University  
tyoon@sungshin.ac.kr

The Accentual Phrase (AP) in Korean is said to be conditioned by the glottal status in the AP-initial segments, with the canonical shape of LHLH, when the initial segment of AP bears [-stiff vocal cords]. While the tonal pattern of adult speakers has been examined extensively, relatively few studies exist that examine the tonal patterns realized by children's speech. The paper examines the developmental trajectory of the canonical tonal shape of AP in Korean by analyzing speech samples of children aged 3 to 10 years old in a publicly available large-scale corpus. The VOT of the initial plain stop was also examined to see whether the tonal shape of the AP in children's speech was affected by the phonetic properties of the initial segment. The results showed that the tonal contours of the Accentual Phrases were best understood to be phonologically encoded and produced by children across most age groups.



구두발표 III  
음성학 및 음운론 4

좌장: 성은경(사이버한국외대)





## 오노마토편어의 음상징

권 연 주  
고베대학 국제인간과학부

### Sound Symbolism in Onomatopoeia

Yeonjoo Kwon  
Dept. of Human Development, Kobe University  
reonisa@gmail.com

하마노(2017)에서는 일본어 [스쿠스쿠]와 [쿠스쿠스]등을 예로 같은 자음이라도 단어내의 위치에 따라 의태어에서의 역할이 다르다고 주장하였다. 2음절 C1VC2V 구조의 의태어의 제 1 자음 (C1)은 대상의 표면의 성질을 나타내고, 제 2 자음 (C2)는 대상의 움직임을 묘사한다고 언급하고 있다.

본 연구에서는 한국어 의태어에서도 위치에 따른 음상징의 차이가 존재하는지를 검증한다. 구체적으로는 한국어의 1음절어기 의태어[초성+중성+종성]에서 음상징이 초성과 종성에서 어떻게 나타나는지 확인한다.

## 한국인 학습자의 프랑스어 연음 실현 양상

김민정<sup>1</sup>, 김선희<sup>2</sup>

<sup>1</sup>서울대학교 외국어교육과, <sup>2</sup>서울대학교 불어교육과

### A Study on the Realization of French Liaison Produced by Korean Learners

Minjeong Kim<sup>1</sup>, Sunhee Kim<sup>2</sup>

<sup>1</sup>Dept. Foreign Language Education, <sup>2</sup>Dept. French Language Education,  
Seoul National University  
monavocado@snu.ac.kr, sunhkim@snu.ac.kr

프랑스어의 대표적 음운현상인 liaison(이하 '연음')은 일정한 환경에서 병치되는 두 개의 단어에서 선행하는 단어의 어말에 위치한 철자상의 자음이 후행하는 단어의 첫 모음과 연결되어 음성적으로 실현되는 음운현상을 의미한다. 연음 실현에는 두 단어의 길이와 빈도, 통사적 범주 등이 요인으로 작용하는 것으로 알려져 있다(Fougeron et al. 2020 ; Bybee, 2005). 본 연구는 연음 실현의 대표적 요인인 두 단어의 길이와 통사적 범주를 기준으로 하여, 한국인 학습자의 연음 실현 양상을 살펴보는 것을 목적으로 한다.

연음 실현의 양상을 살펴보기 위하여 프랑스인 2인과 한국인 학습자 3인이 연음이 포함된 프랑스어 문장 180문장을 낭독하는 음성 데이터를 수집하였다. 발화 문장의 연음 환경에 대한 품사 분류는 Abry & Chalaron (2010)의 분류를 채용하였고, 녹음에 사용된 180개 문장은 Moisset (2005)와 Abry & Chalaron (2010)의 예문에서 추출하였다. 녹음된 음성 데이터에 대하여 전체 문장에 대한 두 발화자 그룹의 연음 실현 비율을 비교 분석하였고, 각 그룹의 필수적 연음과 수의적 연음의 실현 비율을 품사적 분류에 따라 정리하였다. 또한, 180개 문장의 단어 길이를 기준으로 한 연음 실현율을 학습자 그룹과 원어민 그룹으로 나누어 비교 분석하였다.

실험 결과, 원어민 그룹의 필수적 연음 실현율은 93.64%, 학습자 그룹의 필수적 연음 실현율은 70.90%로서 학습자 그룹은 원어민 그룹에 비해 낮은 연음 실현율을 보였다. 특히 수의적 연음의 경우 원어민 그룹은 연음의 51.41%를 실현한 것에 반하여 학습자 그룹은 전체의 17.47%만을 실현하였다. 필수적 연음의 경우, 학습자 그룹의 연음 실현율은 "1음절 부사+형용사" 유형에서 16.6%이었다. 또한, 수의적 연음 환경에서 학습자 그룹은 총 7개의 유형 중, "부정성분", "동사+동사" 유형에서 각각 14.8%, 16.6%의 낮은 실현율을 나타내며, 학습자-원어민 간 연음 실현율에서 큰 차이를 보였다. 단어의 길이를 기준으로 분석한 결과, 학습자의 연음 실현율은 "monosyllabique + polysyllabique" 유형에서 원어민과 가장 큰 차이를 보였다.

#### 감사의 글

이 연구는 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행되었습니다 (NRF-2021S1A5A2A03067523)

#### 참고문헌

Abry, D., & Chalaron, M. L. (2010). Les 500 exercices de phonétique [ensemble multi-support]:

niveau A1/A2.

Bybee, J. (2005). La liaison: effets de fréquence et constructions. *Langages*, 24-37.

Fougeron, C., Goldman, J. P., Dart, A., Guélat, L., & Jeager, C. (2020). Influence de facteurs stylistiques, syntaxiques et lexicaux sur la réalisation de la liaison en français. In *8ème Conférence Traitement Automatique des Langues Naturelles*.

Moisset, Ch. (2000). *Variable liaison in parisian French*. University of Pennsylvania.

## 프랑스어 비강모음의 포먼트 산출을 위한 안정 구간 연구

박혜숙<sup>1</sup>, 김선희<sup>2</sup>

<sup>1</sup>서울대학교 불어교육과, <sup>2</sup>서울대학교 불어교육과

### A Study on the Stable Section for Extracting the Formant Values of French Nasal Vowels

Hye-Sook Park<sup>1</sup>, Sunhee Kim<sup>2</sup>

<sup>1</sup>Dept. French Language Education, <sup>2</sup>Dept. French Language Education,  
Seoul National University  
cielcine@snu.ac.kr, sunhkim@snu.ac.kr

모음에 대한 포먼트 값의 산출은 시간 축을 따라 중간값을 취하는 것이 일반적이다(Agwuele, Sussman, & Lindblom 2008 ; Stevens & House 1963 등). 비강모음 스펙트로그램을 보면, 모음 초반부의 주요 포먼트는 검은 띠 모양으로 평행을 이루다가, 발성 에너지가 구강과 동시에 비강으로 유출되는 지점부터 갑자기 흐려져 반전 공명(antiformant)이 나타난다. 반전 공명 구간의 포먼트는 비강으로 유출되는 에너지로 인해 그 값에 대한 신뢰도가 떨어지는데, 비강모음의 중간값이 이 반전 공명 위에 종종 나타난다. 본 연구는 비강모음의 포먼트에서 객관적 정보를 얻을 수 있는 안정 구간(stable section)의 도출 방법을 상응 구강모음과 함께 고찰하고자 한다. Mermelstein (1978)은 청지각 실험을 통해 모음의 F1의 차이가 60Hz보다 크고, F2가 200Hz 이상을 상회할 때 서로 다른 소리로 지각된다는 차이식역(difference thresholds)을 보고한 바 있다. 본 연구에서는 이러한 Mermelstein의 제안을 채용하여 각 지점들(points)의 증감값이 차이식역 안에 연속하여 한 개 이상 들어가는 구간들(intervals)을 안정된 구간으로 제안한다. 프랑스어의 비강모음과 이에 상응하는 구강모음에서의 안정 구간을 도출하기 위하여 각 모음 구간을 10개 지점으로 분할하고, 두 지점 사이의 F1과 F2의 증감값을 구하였다. 프랑스어 데이터로는 CV.CV 구조의 2음절 단어 오디오를 네이버 프랑스어 사전에서 추출하였다. 첫음절에 파열음이 선행하는 비강모음 232개, 그리고 이에 상응하는 구강모음 249개가 사용되었다. 실험 결과, 안정 구간은 비강모음의 경우 모음 전체의 20~50%의 구간, 그리고 상응하는 구강모음의 경우 20~80%의 구간으로 나타났다.

#### 감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음(IITP-2022-2018-0-01833)

#### 참고문헌

- Agwuele, A., Sussman, H. M., & Lindblom, B. (2008). The Effect of Speaking Rate on Consonant Vowel Coarticulation. *Phonetica*, 65(4), 194-209.
- Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *The Journal of the Acoustical Society of America*, 63(2), 572-580.
- Stevens, K. N., & House, A. S. (1963). Perturbation of Vowel Articulations By Consonantal Context: An Acoustical Study. *Journal of Speech and Hearing Research*, 6(2), 111-128.

구두발표 III  
음성공학 3

좌장: 박형민(서강대)



## HuBERT 기반 한국인 아동 발화 영어 음소 인식 연구

경 나<sup>1</sup>, 박 지 은<sup>1</sup>, 김 규 진<sup>1</sup>, 박 정 식<sup>1</sup>, 강 병 옥<sup>2</sup>  
<sup>1</sup>한국외국어대학교 언어공학연구소, <sup>2</sup>한국전자통신연구원 복합지능연구실

### Study on Recognition of English Phonemes in Korean Children's Speech using HuBERT

Na Geng<sup>1</sup>, Ji-Eun Park<sup>1</sup>, Kyu-Jin Kim<sup>1</sup>, Jeong-Sik Park<sup>1</sup>, Byung Ok Kang<sup>2</sup>  
<sup>1</sup>Language Technology Research Institute, Hankuk University of Foreign Studies  
<sup>2</sup>Integrated Intelligence Research Section, Electronics and Telecommunications Research  
Institute  
gengna0324@gmail.com, ppae0216@naver.com, rndlwjs@gmail.com, parkjs@hufs.ac.kr,  
bokang@etri.re.kr

딥러닝 기술의 발전으로 음소 단위의 음성 인식 모델은 높은 인식률을 보이고 있다. 하지만 대용량의 훈련데이터가 모델에 반영된 원어민(L1) 화자의 발화와 달리 비원어민(L2) 화자가 발화한 음성에 대해서는 인식률이 저하되는 경향이 있다. 본 연구에서는 한국인 아동 영어 학습자의 발화를 잘 인식할 수 있는 음소 모델 개발 내용을 소개한다.

딥러닝 모델 구축을 위해서는 다량의 훈련 데이터가 요구된다. 음소 모델 또한 실제 발화 음성에 정확히 매칭되는 음소 레이블이 포함된 데이터가 필요하지만, 상당량의 데이터를 취득하는 것은 어려우며, 특히 한국인 아동이 발화한 영어 음성에 음소 레이블이 표기된 데이터는 찾기 어렵다. 따라서 본 연구에서는 소량의 데이터로 처리하기 용이한 사전 학습 모델을 활용한다. 즉, 다량의 데이터로 사전 학습(pre-training)한 모델에 소량의 데이터로 미세 조정(fine-tuning)하는 기법을 사용한다.

기존 연구 결과에 따르면, 사전 학습 모델 구축 방법 중, HuBERT 기반의 방식이 음소 인식에 좋은 성능을 보이는 것으로 나타났다. 따라서, 본 연구에서는 HuBERT의 Base 모델, Large 모델을 통해 실험을 진행하였다. 우선 HuBERT의 Base 모델과 이의 선행연구 모델인 Wav2Vec 2.0의 Base 모델로 baseline을 구축한 후, 이 모델을 사용한 실험 결과를 반영하여 한국인 아동 발화에 적절하도록 음소 모델을 미세 조정했다. 한국인 아동이 발화한 단어 데이터를 대상으로 PER을 기준으로 성능을 평가한 결과, HuBERT 모델과 Wav2Vec 2.0 모델은 각각 38.58%, 38.68%의 성능을 보였으며, HuBERT 기반의 Fine-tuning 모델은 17.69%의 PER를 나타냈다.

본 연구에서 개발한 음소 인식 모델은 한국인 아동의 영어 학습자의 잘못된 발음을 교정하는데 활용될 수 있을 것으로 기대된다.

\*이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(2019-0-00004, 준지도학습형 언어지능 원천기술 및 이에 기반한 외국인 지원용 한국어 튜터링 서비스 개발). 또한 한국연구재단의 지원(No. 2020R1A2C1013162)을 받아 수행된 연구임.



## 외국인의 한국어 발음 오류 패턴 태깅을 위한 강제정렬 모델 활용

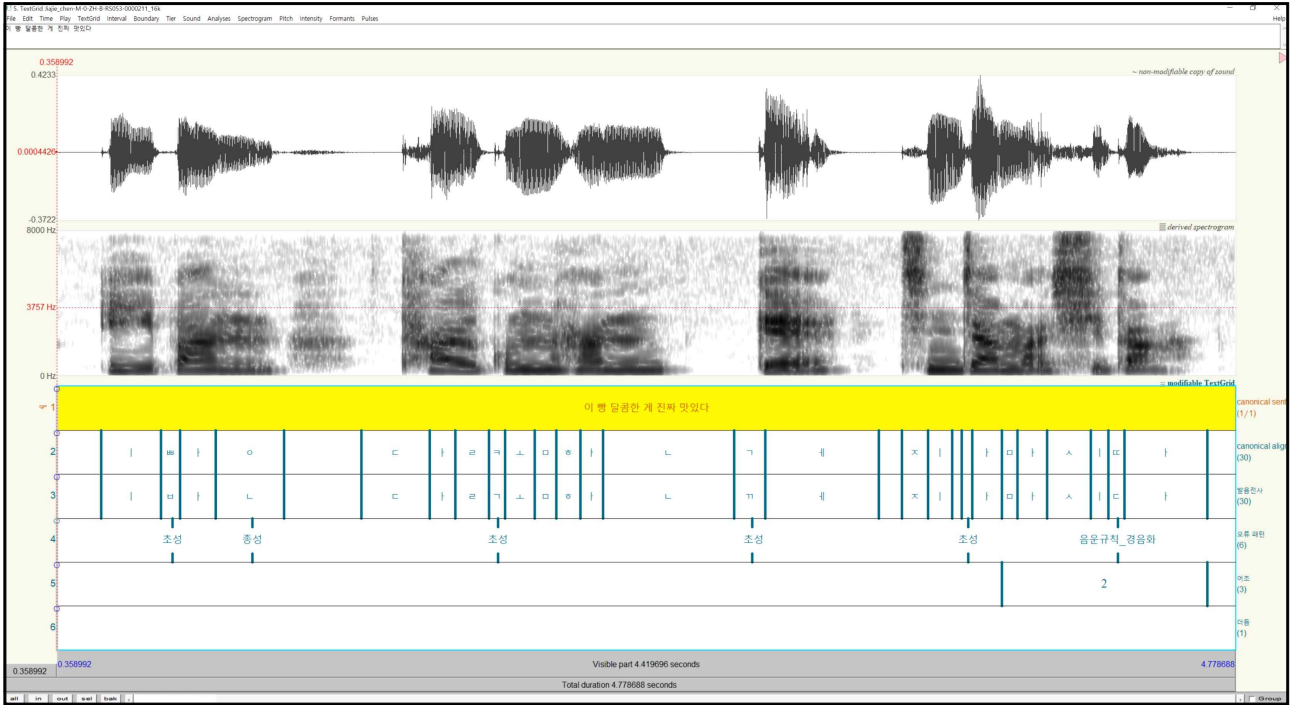
이 주 영<sup>1</sup>, 황 효 성<sup>1</sup>, 류 형 신<sup>1</sup>, 이 호 영<sup>1</sup>, 김 선 희<sup>2</sup>, 정 민 화<sup>1</sup>  
서울대학교 언어학과<sup>1</sup>, 서울대학교 불어교육과<sup>2</sup>

### A Forced-Alignment Model for Mispronunciation Tagging on Korean Speech of Non-native Speakers

Jooyoung Lee<sup>1</sup>, Hyosung Hwang<sup>1</sup>, Hyungshin Ryu<sup>1</sup>,  
Ho-Young Lee<sup>1</sup>, Sunhee Kim<sup>2</sup>, Minhwa Chung<sup>1</sup>  
Dept. of Linguistics, Seoul National University<sup>1</sup>  
Dept. of French Language Education, Seoul National University<sup>2</sup>  
excalibur12@snu.ac.kr<sup>1</sup>, pyscho77@snu.ac.kr<sup>1</sup>, rhss10@snu.ac.kr<sup>1</sup>,  
hylee@snu.ac.kr<sup>1</sup>, sunhkim@snu.ac.kr<sup>2</sup>, mchung@snu.ac.kr<sup>1</sup>

강제정렬(forced-alignment)은 발화 음성을 주어진 발음열에 맞추어 강제로 발화 구간을 음소 단위로 분절하는 기법으로, 음성을 단어 혹은 음소 단위로 세밀하게 분석할 때 많이 활용된다. 이와 같은 음소 단위 분석은 외국인의 한국어 발화에서 표준 한국어 발음에 비해 어느 음소가 어떻게 잘못 발음되었는지 살펴볼 때 활용할 수 있다. 본 연구에서는 외국인의 한국어 낭독 발화에서 생긴 발음 오류 패턴을 태깅하기 위한 작업 공간을 제안하며, 이때 강제정렬 모델이 어떻게 활용되었는지도 함께 설명한다.

발음 오류 패턴 태깅을 위한 작업 공간은 Praat[1]에서 열람할 수 있는 TextGrid 형태로 되어 있으며, 총 6개의 Tier로 구성하였으며 [그림 1]과 같이 나타내었다. 각 Tier는 (1) 대본 낭독 문장, (2) 표준 발음 열 기준 한글 자소별 강제정렬 결과, (3) (2)의 강제정렬 결과를 바탕으로 수정한 외국인의 실제 음소별 발음, (4) 발음 오류 유형, (5) 특이 억양 구간, (6) 말더듬 구간을 의미한다. 이때 (2)의 강제정렬 결과는 공개 엔진인 Montreal Forced-Aligner(MFA)[2]를 사용하여, (1) 대본 낭독 문장의 발음열에 맞추어 음소 단위로 강제정렬 하였다. 또한 발음 오류 태깅 작업의 단순화를 위해 [표 1]과 같이 복수로 표준 발음이 허용되는 경우에도 대본 문장과 1:1 대응이 되도록 하나의 고유 발음열을 수동으로 지정하여 작성하였다. MFA 엔진은 영문 발음열 형태의 단어 단위와 음소 단위 강제정렬 결과만 생성하기 때문에 본 연구에서는 태깅 작업 공간에 맞게 (1), (4), (5), (6) Tier 추가 및 MFA의 결과인 (2)와 (3) Tier의 한글 발음열로의 변환을 추가적으로 진행하여 오류 태깅 작업에 바로 사용할 수 있도록 준비한 것이다.



[그림 1] 강제정렬 모델을 활용하여 발음 오류 태깅 한 예시

[표 1] 대본 텍스트 및 표준발음열의 1:1 대응표 (문장 리스트 일부)

문장	표준발음열
가족이 몇 명이예요	가 자고 기 며 나 며 이 게 표
여기 비빔밥 한 그릇 갖다 주세요	여 기 비 비빔 밥 한 그릇 갖 다 주 세 요 가 드 따 자 서 게 표
십만 원만 빌려줄 수 있어요	사 모 마 나 거 나 마 나 비 러 러 자 서 쓰 서 쓰   쓰 서 표

### 감사의 글

이 논문은 2022년도 한국지능정보사회진흥원의 재원으로 과제 '언어교육용 서양어, 아시아어 사용자의 한국어 음성 데이터'의 지원을 수행된 연구임(PA202201662).

### 참고문헌

- [1] Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.23, retrieved 8 October 2022 from <http://www.praat.org/>
- [2] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. (2017, August). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In Interspeech (Vol. 2017, pp. 498-502).

## 종단간 음성인식 모델의 다층 특징을 활용한 화자인식

박 순 찬, 김 형 순  
부산대학교 전자공학과

### Speaker Recognition using Multi-Layer Features of End-to-End Speech Recognition Model

Sunchan Park, Hyung Soon Kim  
Dept. of Electronics Engineering, Pusan National University  
sunchanpark@pusan.ac.kr, kimhs@pusan.ac.kr

음성인식 시스템의 화자적응 기술은 동일한 음소로 구분되는 음성이라도 화자에 따른 발화 방식의 차이를 반영하여 특정 화자에 대한 음성인식 성능을 개선한다. 이와 같은 화자에 따른 음소 실현 방식의 차이는 화자인식에서도 유용하게 사용될 수 있는데, 화자인식 모델에 다중작업학습(multi-task learning)을 적용하거나 사전학습된 음성인식 모델을 결합하는 방식으로 구현된다. 본 연구에서는 다중작업학습을 기반으로 종단간 음성인식모델의 다층 특징을 화자인식 모델에 결합하여 문장 독립형 화자인식 성능을 개선하는 방법을 제안하였다. 전체 네트워크 구조는 기존 ECAPA-TDNN 화자인식 모델에 이와 유사한 구조의 종단간 음성인식 모듈을 추가로 결합하고, 음성인식 모듈의 마지막 레이어에서 문자열을 추정하도록 CTC 손실함수를 통해 훈련하였다. 기존 ECAPA-TDNN 화자인식 모델에서는 네트워크 중간의 다층 출력을 MFA(Multi-Feature Aggregation) 레이어를 통해 결합하고, ASP(Attentive Statistical Pooling) 레이어를 거쳐 화자 임베딩(embedding)을 추출한다. 제안하는 모델에서는 음성인식 모듈에서도 다층 출력을 MFA 레이어를 통해 결합하고, 화자인식 모듈의 MFA 레이어 출력과 가중합을 통해 통합하고 ASP 레이어를 거쳐 화자 임베딩을 추출하도록 구성하였다. 이때 가중합 과정에서 음성인식 모듈 출력의 가중치를 0.1로 화자인식 모듈의 가중치 0.9보다 상대적으로 낮게 조절하여 학습 데이터의 특정 화자와 음소에 대해 과적합 되는것을 방지하였다. 화자확인 성능비교를 위해 화자인식을 위한 VoxCeleb2, 음성인식을 위한 LibriSpeech 데이터셋을 통해 학습하고 VoxCeleb1 데이터셋에 대해 평가한 결과, 화자인식만으로 학습한 모델 대비 제안한 모델의 동일오류율이 0.90%에서 0.81%로 감소하여 10%의 성능 개선 효과를 확인할 수 있었다.

#### 감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00575, 음성-텍스트 딥러닝 기반 보이스피싱 예방 기술 개발)

구두발표 III  
말장애 및 음성의학 2

좌장: 김지연(우송대)



## 이란성 쌍둥이 구순구개열 아동과 일반 아동 파열음의 음향음성학적 특성 비교

전 선 영\*, 박 지 연\*, 성 철 재<sup>§</sup>  
충남대학교 언어병리학과, 충남대학교 언어학과

### Comparison of the acoustic phonetic characteristics of plosive sounds between fraternal twins with cleft lip & palate and normal children

Sunyoung Jeon, Jiyeon Park, Cheoljae Seong

\*Dept. of Speech & Language Pathology, Linguistics, Chungnam National University

<sup>§</sup>Dept. of Linguistics, Chungnam National University

shashack@hanmail.net, bpn0525jy@gmail.com, cjseong49@gmail.com

구순구개열 아동은 파열된 구개로 인해 일반 아동과 다르게 음운 발달 및 말소리 산출이 나타난다. 일반 유아들이 생후 6개월 경 양순·치조파열음이 포함된 음절성 발음이 나타나는 반면 구순구개열 유아들은 성문·연구개파열음과 같이 구강 뒤쪽 자음 위주로 산출하며 입술, 설첨 부위를 덜 사용한다. 또한 돌 전후로 구개성형술이 이루어져도 입천장 및 연인두의 선천적인 조음기제의 결함으로 구강파열음을 성문파열음으로 산출하는 보상적 오류와 일반 아동에 비해 낮은 말 명료도와 말 용인도가 부정적 요인으로 남는다. 본 연구는 이란성 쌍둥이 구순구개열 아동과 일반 아동의 구강파열음에 대해 청지각적 평가를 실시하여 정확도를 분석하고 아동 간 조음 위치와 발성 유형에 따라 어떤 음향음성학적 변수에 유의미한 차이가 있는지 살펴보고자 하였다.

연구에 참여한 이란성 쌍둥이는 2014년 8월생(7:4)으로 언어발달과 인지능력 모두 정상이다. 쌍둥이 형은 조음기관의 구조적 문제가 없으나, 동생은 Bilateral cleft lip with full palate로 2014년 11월에 Cheiloplasty, 2015년 10월에 Palatoplasty, 2017년 11월에 Palatal lengthening, 2019년 5월에 VPI로 인해 Palatal lengthening 2차 수술을 받았다. 음성 표본은 두 아동에게 구강파열음/p, t, k, p', t', k', p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ 9개와 단모음/i, ε, i, ə, u, o, α/ 7개가 결합한 CV, VCV 글자를 일정 속도로 제시하여 조음 속도를 일정하게 3회 녹음하여 수집하였다.

청지각 실험은 Praat의 ExperimentMFC를 이용하여 15%의 dummy를 포함한 파열음을 조음 위치와 발성 유형으로 나누어 1점(부정확함)에서 6점(정확함) 리커트 척도로 정확도를 평가하였다. 청지각 평가자는 총 8명으로 언어병리학을 석·박사 이상 전공하였거나 언어치료 임상 경력 9년 이상 자이다. 평가자 내 일치도는 감마계수 값이 .877, 카파계수 값이 .588, 스피어맨 로우 값이 .793으로 높은 것으로 나타났다( $p < .0001$ ). 평가자 간 신뢰도는 급간내 상관계수(ICC) Cronbach's Alpha 값의 평균측도가 .879로 매우 높게 평가되었다[AM(754, 5278)=.879,  $p < .0001$ ].

청지각 평가 결과, 정확도는 대상자, 발성유형 각각의 점수에서 유의한 차이가 있었고( $p < .001$ ), 조음 위치에서는 유의한 차이가 관찰되지 않았다( $p = .426$ ). 상호작용은 대상자\*발성유형( $p < .001$ ), 발성유형\*조음위치( $p < .001$ ), 대상자\*조음위치( $p < .01$ ), 대상자\*조음위치\*발성유형( $p < .001$ )에서 모두 유의하였다.

측정한 음향음성학적 변수는 총 13가지로 유성 개시 시간, 유성 개시 후 첫 5ms에 대한 적률변수(무게 중심, 표준편차, 왜도, 첨도)와 스펙트럼 에너지 총량, 기본주파수, 후행 모음의 첫 번째 pitch point부터 5ms의 적률변수, H1-H2, H1-A2이다. 대상자, 조음위치, 발성유형을 독립변인으로 유성 개시 시간, 유성 개시 후 스펙트럼 전체 에너지, 기본주파수는 각각 3way ANOVA를 실시하였다. 상호 밀접한 상관관계를 보이는 적률변수군이나 H1-H2, H1-A2는 1종 오류 감소를 위해 다변량 분산분석(MANOVA)을 하였고, 사후검정은 본페로니 다중 짝대응 비교를 실시하였다. 로지스틱 회귀 분석을 시행하여 대상자를 구분하는 작

업과 청지각 평가 결과에 따라 고득점과 저득점을 구분하였다.

검정 결과, 유성 개시 후 5ms의 적률변수와 전체 에너지, 기본주파수가 대상자 간 가장 유의한 차이가 있었다. 조음위치는 유성 개시 시간, 유성 개시 후 첫 5ms의 무개중심, 표준편차, 왜도와 전체 에너지가 유의미했고, 발생유형은 유성 개시 시간, 유성 개시 5ms의 왜도, 첨도와 기본주파수, H1-H2, H1-A2에 유의한 차이가 나타났다. 유성 개시 후 왜도는 모든 독립 변수 및 상호작용에서 유의한 차이가 나타났다. 로지스틱 회귀분석 결과, 구순구개열 아동과 일반 아동에 대한 분류 정확도는 74.2%, 청지각 점수 3.5점을 기준으로 고득점과 저득점 분류 시 82.5%, 청지각 점수 3점 미만은 초저득점, 4점 초과는 초고득점으로 분류 시 86.5%로 나타났다. 모든 분류 작업에서 기여도가 가장 높은 변수는 유성 개시 시간, 유성 개시 후 5ms의 왜도와 전체 에너지, 기본주파수였다. 아래 표에 독립 변수별로 음향변수들의 주효과와 상호작용 유무를 표시하였다.

음향변수		독립변수	대상자	조음위치	발성유형	대상자* 조음위치	대상자* 발성유형	조음위치* 발성유형	대상자* 조음위치* 발성유형
유성 개시 시간				***	***	***	***	*	
유성 개시 후 5ms	무개중심		***	***		***	*	*	
	표준편차		***	***		***			
	왜도		***	***	***	***	***	***	**
	첨도		***		***	**	***	**	*
전체 에너지			***	***		***			
기본주파수			***		***		***		
1st pitch point 5ms	무개중심				**		*		
	표준편차		**						
	왜도		*						
	첨도		*						
H1-H2					***				
H1-A2			*		***				

\*:  $p < .05$ , \*\*:  $p < .01$ , \*\*\*:  $p < .001$

본 연구에서 이란성 쌍둥이 구순구개열 아동과 일반 아동의 파열음에 대한 정확도를 청지각 평가를 통해 확인하였으며, 청지각 평가에 영향을 미치는 음향음성학적 요인을 살펴보았다. 청지각 판단에 유성 개시 시간과 기본주파수가 가장 주요한 단서로 인식되었으나 유성 개시 후 5ms의 왜도와 에너지 또한 파열음을 정확하게 지각하는데 중요한 음향 단서로 나타났다. 또한 구순구개열 아동과 일반 아동의 유성 개시 시간 차이는 없었으나, 유성 개시 후 5ms의 적률변수와 전체 에너지가 유의한 차이를 보여 이는 구순구개열 아동이 모여 있는 압력을 터트리는 개방 파열의 초입에서 파열과 발성의 시간 차에 의한 것이나 구개천공으로 인한 압력 저하로 여겨진다. 향후 더 많은 구개열 아동을 대상으로 조음기제와 보상조음에 대한 음향음성학적 연구가 다양하게 이루어지길 바란다.

### 참고문헌

- [1] 박선영, & 성철재. (2022). 한국어 파열음 조음자리와 발성유형 구분을 위한 음향변수 분석 및 선정. 한글, 83(3), 571-610.
- [2] 한진순, & 심현섭. (2008). 구개열 아동과 일반 아동 및 기능적 조음장애 아동의 자음정확도, 말 명료도 및 말 용인도 비교. Communication Sciences & Disorders, 13(3), 454-476.
- [3] 김호, & 하승희. (2021). 수술 후 구개열 아동과 일반 아동의 음절성 발성 및 음운 발달. Communication Sciences & Disorders, 26(3), 672-685.

## 말더듬 성인의 군대 적응 정도에 대한 군지휘관의 인식 양상 및 관련 요인 분석

박 헤 린\*, 고 복 순\*\*, 박 진\*\*

\*충남대학교 대학원 언어병리학과, \*\*가톨릭관동대학교 언어재활학과

### Perceptions of Military Officers on Adults who Stutter and their Adaptation in the Military and the Associated Factors

Hyelin Park\* · Boksun Go\*\* · Jin Park\*\*

\*Dept. Linguistic Pathology, Chungnam National University

\*\*Dept. of Speech-Language Rehabilitation, Catholic Kwandong University

linday96@hanmail.com, tk-37@hanmail.net, gatorade70@cku.ac.kr

본 연구에서는 직업군인의 말더듬 성인에 대한 군대 적응 예측 정도와 이러한 인식에 유의한 영향을 미치는 요인들이 무엇인지를 알아보았다. 이를 위해 89명의 직업군인(부사관, 장교)을 대상으로 발화 유창성 정도와 관련해 세 가지 버전(즉, 유창한 발화(23명), 경도 말더듬(34명), 심도 말더듬(32명)) 중 무작위로 배정된 하나의 음성샘플을 들려주고 발화 특성 관련 요인(유창성, 명료도, 자연스러운 정도, 말속도), 전반적인 인상 특성 관련 요인(호감 정도, 불안 정도, 지적 수준 정도, 사회성), 그리고 군대 적응 예측 정도를 평가하였다. 일변량 분산분석(one way ANOVA) 결과, 군대 적응 예측 정도에서 경도 말더듬과 심도 말더듬 간, 그리고 유창한 발화와 심도 말더듬 간에서 유의한 차이가 나타났다( $p < .05$ ). 하지만 경도 말더듬과 유창한 발화에서는 유의한 차이가 나타나지 않았다. 또한, 상관관계 분석(Pearson correlation test) 결과, 군대 적응 예측 정도에 있어 발화 특성 요인 중에서는 발화 유창성, 자연스러운 정도, 명료도, 말속도 순으로, 전반적인 인상 특성 관련 요인 중에서는 호감 정도, 사회성, 불안 정도, 지적 수준 정도의 순으로 높은 상관관계가 나타났다. 이는 발화 유창성 정도가 군대 생활 예측 정도에 상대적으로 높은 상관관계를 미치는 요인으로 해석할 수 있다. 본 연구는 말을 더듬는 사람에 대한 군대 내에서의 인식 개선은 물론 말더듬으로 인한 군대 적응의 어려움 혹은 말더듬에 대한 선입견을 해소하는 방안을 마련하는 실제적 근거로서의 의의가 있다고 하겠다.



## 물 저항 발성 시 깊이에 따른 성대접촉률과 설골하근 수축량의 변화

차 준 서, 김 채 현, 최 성 희  
대구가톨릭대학교 언어청각치료학과

### Contact Quotient and Myoelectric Activity in the Infrahyoid Muscles during Tube Phonation Submerged Underwater at Varied Depths

Junseo Cha, Chaehyun Kim, Seong Hee Choi  
Dept. of Audiology & Speech-Language Pathology  
junseo94@cu.ac.kr, ulche99@cu.ac.kr, shgrace@cu.ac.kr

**서론:** 물 저항 발성의 해부적, 공기역학적 이점은 다양한 시뮬레이션 및 후두내시경 연구로 진행이 되어왔으나, 환자가 즉각적인 바이오피드백으로 사용하여 반폐쇄성도 발성 훈련을 할 수 있는 도구는 부족하다. 이를 보완하기 위하여 표면 근전도 검사 기기를 이용해 설골하근에 표면 근전도 스티커를 부착 후 근 수축량을 통한 후두 하강 정도를 간접적으로 조사하였다.

**방법:** 음성적으로 건강한 정상인 62명 (남성 31명, 여성 31명) 대상으로 설골하근에 표면 근전도 스티커를 부착 후, 물저항 발성을 6단계의 수심(2cm, 4cm, 7cm, 10cm, 15cm, 20cm)에서 실시하였다. 성별과 수심에 따라 나타나는 성대접촉률과 설골하근의 수축량의 변화를 알아보기 위해 이원 혼합 분산분석을 실시하였다.

**결과:** 성대접촉률은 성별과 물 깊이에 따라 상호작용이 관찰되지 않았으며 ( $p = .092$ ), 주효과 분석 시, 물의 깊이에 따라 유의한 차이가 보였다 ( $p < .001$ ). 상호작용이 통계적으로 유의하게 관찰되었다 ( $p < .001$ ). 남성 집단의 경우 수심이 깊어 질수록 설골하근의 수축량이 증가하는 경향이 보인 반면, 여성 집단의 경우 수심이 증가함에 따라 설골하근의 수축량의 증가율이 비교적 낮게 관찰되었다. 성별 내 주효과를 확인 한 결과 남성의 경우 2cm와 4cm, 4cm와 7cm, 7cm와 10cm를 제외한 모든 수심에서 통계적으로 유의한 차이를 보였고, 여성의 경우 모든 수심에서 통계적으로 유의한 차이가 나타나지 않았다.

**논의 및 결론:** 물 저항 발성 시 남성 같은 경우에는 수심이 깊어짐에 따라 수축량이 증가하는 설골하근의 정도를 바이오피드백으로 이용해 훈련이 가능한 반면, 여성 같은 경우에는 효과가 없을 수 있다.

**키워드:** 물 저항 발성, 표면 근전도 검사, 설골하근

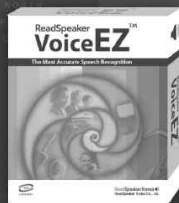
## 음성기술 전문회사 (주)리드스피커코리아

20년 이상의 음성기술을 향한 노하우와 기술력의 신뢰를 바탕으로  
국내는 물론 해외 다양한 분야에서 인정받고 있습니다.



### 음성합성 ReadSpeaker™

인공지능(AI) 기술을 적용하여 음질은 한층 더 높이고  
음성합성기의 개발기간은 단축시킨 음성합성기,  
DNN TTS(Deep Neural Network TTS) 37개 언어, 106개 음색 보유



### 음성인식 ReadSpeaker VoiceEz™

차세대 Human - Machine Interface의 핵심 음성인식

(주)리드스피커코리아

www.readspeaker.co.kr  
sales@readspeaker.co.kr  
tel 02-3016-8500

## 음성기술

# 인공지능 전문 기업 셀바스 AI가 이끌어가겠습니다.

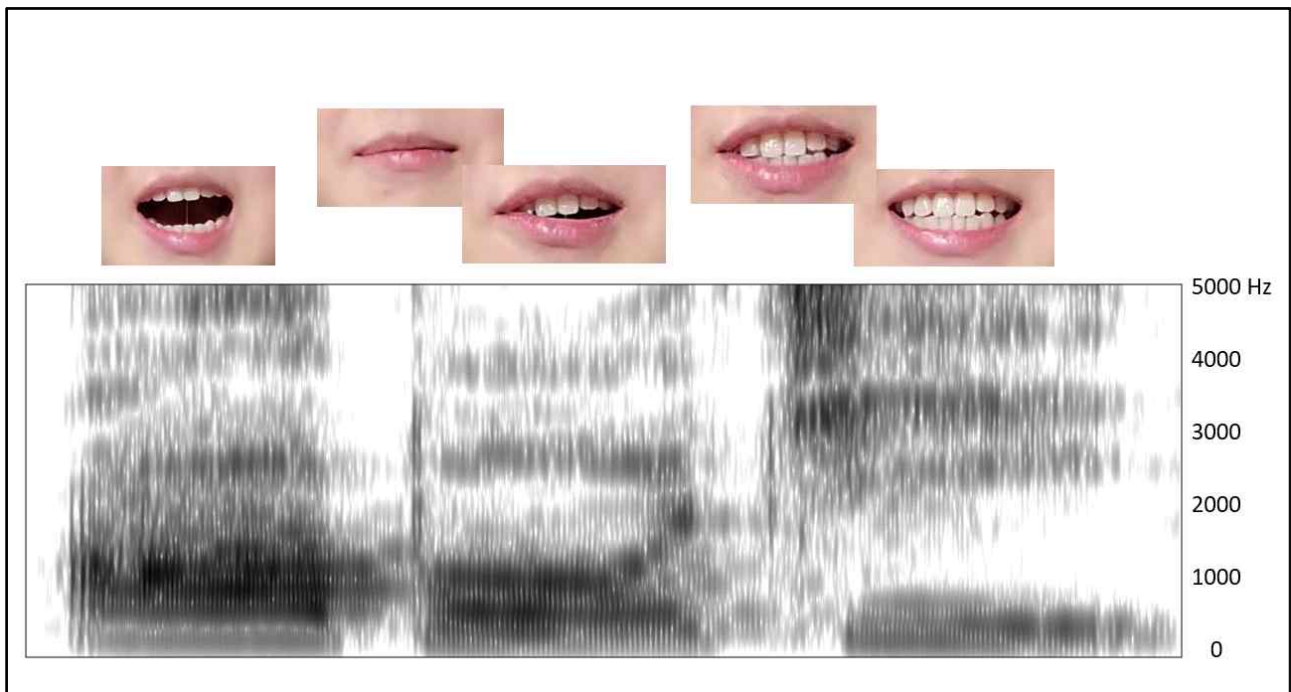
셀바스 AI는 음성인식 및 음성합성 등 핵심 음성기술 분야의  
독보적 원천 기술을 보유한 국내 대표 음성기술 기업입니다.

셀바스 AI는 꾸준한 연구개발을 통해 음성기술의 성능 향상에 힘쓰고 있으며,  
다양한 환경에 최적화 된 인공지능 기반 제품화를 통해  
고객들에게 새로운 가치를 제공하고 있습니다.



기업 selvasai.com | 채용 careers.selvasai.com | 제품 selvy.ai  
이메일 문의 support@selvasai.com | 전화 문의 02.852.7788

**Proceedings**  
**of the 2022 Fall Conference**  
**of the Korean Society of Speech Sciences**



**November 25, 2022**  
**Zoom Webinar**  
**The Korean Society of Speech Sciences**