ISSN 2005-9299

2019 한국음성학회 봄 학술대회

발표 논문집

주제: AI 시대의 음성ㆍ언어 연구

일시: 2019년 6월 1일(토)

장소: 연세대학교 새천년관

주최: 사단법인 한국음성학회

주관: 사단법인 한국음성학회

후원: 한국연구재단, ㈜엔씨소프트, ㈜코난테크놀로지, ㈜네이버
㈔한국언어재활사협회, KAIST, ㈜보이스웨어, ㈜셀바스AI
연세대학교 언어정보연구원

한국음성학회
The Korean Society of Speech Sciences

이 발표논문집은 2019년도 정부 재원(교육부)으로 한국연구재단의 지원을 받아 발간되었음

모시는 글

안녕하십니까.

2019년이 시작된 지 얼마 지나지 않은 것 같은데 벌써 한여름의 더위를 느끼게 됩니다. 올해도 어김없이 본 학회의 가장 중요한 행사 중 하나인 봄 학술대회가 개최되었습니다. 학술대회에 참석해주신 모든 음성학회 회원 여러분께 진심으로 감사를 드립니다.

이번 학술대회를 통해 한 분의 자문위원께서 은퇴하시게 되었음을 알려드리게 되었습니다. 이용주 명예회장님께서 2019년 2월에 교육 일선에서는 퇴직하셨지만, 더욱 폭넓은 연구 및 사업 활동을 위해 여러 가지 일들을 준비하고 계신 것으로 알고 있습니다. 특별히 오늘 지난 40년간의 음성 연구를 회고하는 소중한 기회도 주시기로 하셨습니다. 그동안 본 학회의 발전을 위하여 지대한 공헌을 하셨고, 다시 새로운 시작을 준비하고 계신 분께 여러분께서도 함께 응원의 큰 박수를 보내주십시오.

학술대회를 개최하기 위해서 많은 분들이 수고해주셨습니다. 이석재 조직위원장님, 박한상 학술위원장님을 비롯한 학술대회 위원님들 그리고 논문을 보내주신 모든 분들께 감사드립니다. “AI 시대의 언어지능 융합”이라는 주제로 특강을 준비해주신 ETRI 김영길 박사님께도 감사드립니다. 또한 “공연음성의 이론과 실제”라는 주제로 흥미로운 강연을 해주시는 Arkansas대학교의 박문숙 교수님께도 감사드립니다. 더욱이 한국음성학회 봄 학술대회의 성공적 개최를 위해서 많은 기관과 기업에서 후원해주셨습니다. 장소를 제공하고 후원도 해주신 연세대학교를 비롯하여 한국연구재단, 엔씨소프트, 코난테크놀로지, 네이버, (사)한국언어재활사협회, KAIST, 보이스웨어, 셀바스AI에서 후원해주셨습니다. 대단히 감사합니다.

오늘 함께 해주신 모든 분들에게 이 학술대회가 학문적 교류의 귀중한 기회가 되기를 진심으로 기원합니다.

2019년 6월

한국음성학회 회장 김 회 린

2019 한국음성학회 봄 학술대회 준비위원회

학술대회장	회장	김회린(KAIST)
조직위원회	위원장	이석재(연세대)
	위원	박한상(홍익대), 장태엽(한국외대), 성철재(충남대), 정민화(서울대), 최성희(대구가톨릭대), 정현성(한국교원대), 김지환(서강대)
학술위원회	위원장	박한상(홍익대)
	위원	남호성(고려대), 오재혁(건국대), 김성태(동신대), 박상희(대구사이버대), 박기영(ETRI), 박형민(서강대), Jeffrey Holliday(고려대)

2019 한국음성학회 봄 학술대회 일정표

6월 1일(토) 연세대학교 새천년관

<table>
<tr><th>시 간</th><th colspan="3">발 표 및 내 용</th></tr>
<tr><td>09:20~9:50</td><td colspan="3">등록(1층 로비)</td></tr>
<tr><td>9:50~10:50</td><td colspan="3">포스터 발표 Ⅰ
(1층 로비)</td></tr>
<tr><td>10:50~11:00</td><td colspan="3">개회식
(101호)
사회: 정현성(한국교원대)</td></tr>
<tr><td>11:00~11:30</td><td colspan="3">특강 Ⅰ
(101호)
음성 연구 40년의 회고
강사: 이용주(원광대)</td></tr>
<tr><td>11:30~12:20</td><td colspan="3">특강 Ⅱ
(101호)
AI 시대의 언어지능 융합
강사: 김영길(ETRI)</td></tr>
<tr><td>12:20~13:20</td><td colspan="3">점심(지하 2층 푸드코트)</td></tr>
<tr><td rowspan="2">13:20~14:20</td><td colspan="3">구두 발표 Ⅰ</td></tr>
<tr><td>음성학
(101호)</td><td>말장애 및 음성의학
(111호)</td><td>음성공학
(112호)</td></tr>
<tr><td>14:20~14:30</td><td colspan="3">휴식</td></tr>
<tr><td>14:30~15:30</td><td colspan="3">포스터 발표 Ⅱ
(1층 로비)</td></tr>
<tr><td>15:30~15:40</td><td colspan="3">휴식</td></tr>
<tr><td rowspan="2">15:40~16:40</td><td colspan="3">구두 발표 Ⅱ</td></tr>
<tr><td>음성학
(101호)</td><td>말장애 및 음성의학
(111호)</td><td>음성공학
(112호)</td></tr>
<tr><td>16:40~17:30</td><td colspan="3">특강 Ⅲ
(101호)
공연음성의 이론과 실제
강사: 박문숙(University of Arkansas)</td></tr>
<tr><td>17:30~18:00</td><td colspan="3">우수발표시상 및 폐회식
(101호)
사회: 박한상(홍익대)</td></tr>
</table>

2019 한국음성학회 봄 학술대회 세부 일정표

포스터 발표 I (1층 로비)

좌장: 윤영도(동국대), 김재옥(강남대), 유하진(서울시립대)

시간	구분	제목	저자
9:50 ~ 10:50	P01	과제, 세대, 성별에 따른 휴지 간 발화 길이의 실현 양상	유도영, 신지영(고려대)
	P02	시간 벌기 담화표지의 지연 정도와 분포	석수경, 신지영(고려대)
	P03	인공지능 스피커 음성 특성에 따른 한국어 화자의 인식 양상	김승록, 신지영(고려대)
	P04	공용 미디어 자료를 이용한 발화 호감도에 대한 음성 산출 및 청지각 연구	이현정, 성철재(충남대)
	P05	독일어의 말음 무성음화에 따른 비완전 중화현상 재고찰	이동은, 이석재(연세대)
	P06	한국인의 L2 영어 고립어 발화에서 유창성 등급 대비 어말 폐쇄음 유무성성에 따른 선행 모음 길이 비율 연구	이현전, 박우지, 이석재(연세대)
	P07	한국인의 L2 영어 [관사 - 1음절 명사]NP에서 기능어 - 내용어 사이 모음 길이 비율과 영어 능숙도 사이의 상관 관계 연구	박우지, 모란, 이석재(연세대)
	P08	문장 조건에서 3~6세 아동의 음운지표 발달	박나래, 김수진(나사렛대), 하지완(대구대), 신문자(조선대), 김영태(이화여대)
	P09	학령전 아동의 UTAP2 문장과제에 나타난 음운오류패턴	최영빈, 김수진(나사렛대)
	P10	과소비성 분류를 위한 음향음성학적 접근 - 3가지 머신러닝 기법의 비교 -	김난숙, 성철재(충남대)
	P11	연결발화에서 근긴장성발성장애 음성의 켑스트럼 분석	윤주원(충남대), 심희정(한림대), 성철재(충남대)
	P12	과제 유형에 따른 음성장애 화자의 청지각적 및 음향학적 상관성	박상희, 이옥분(대구사이버대)
	P13	Nasality Visualization System(NVS)를 이용한 청년 및 중년기 성인의 말소리문맥에 따른 비음치 연구	정재만, 하지완(대구대). 우승탁, 박영빈(경북IT융합산업기술원)
	P14	Influence of the nasometer structure on the nasalance: evaluation of acoustic clinical experimental and lumped-element model analysis	Seong Tak Woo, Young-Bin Park(GITC), Seong Dae Na(KNUH), Euisung Jung(DGMIF), Da Hee Oh, Ji-Wan Ha(Daegu Univ.)
	P15	노년층의 조음교대운동 속도와 빠른 씹기 속도 간의 최대수행력 비교	이지언, 이현정, 김향희(연세대)
	P16	The Effect of Unvoiced Segments on the Cepstral Analysis in Age-related Dysphonia	Geun-Hyo Kim, Yeon-Yoo Lee, In-Ho Bae(PNUH), Hee-June Park(Catholic Univ. of Pusan), Chang-Yoon Lee(DIRAMS), Soon-Bok Kwon(Pusan National Univ.)
	P17	다채널 마이크로폰 배열에서의 DNN 기반 빔형성에 관한 연구	김태우, 김남균, 이건우, 김홍국(광주과학기술원)
	P18	ITD와 분산도를 활용한 가우시안 혼합 모델 기반 음원 국지화 방법	김정민, 김보현(서강대), 김이경, 박강호(ETRI), 박형민(서강대)
	P19	자기부호화기 특징을 이용한 딥러닝 기반 비침입적 음성명료도 추정 방법	김윤희, 윤덕규, 최승호(서울과학기술대), 김선교(국방과학연구소)
	P20	순차적 공분산 추정을 이용한 마스크 기반 온라인 빔포밍 구현	윤성욱, 권오욱(충북대)
	P21	Conditional GAN 기반 가창 합성 시스템	최순범, 남주한(한국과학기술원)
	P22	음향적 특징과 Attention 기반 LSTM 모델을 활용한 한국어 방언 추정 연구	이주영(서울대), 김경화(대검찰청), 이규환, 정민화(서울대)

특강 Ⅰ·Ⅱ (101호)

시간	내용
11:00~11:30	**특강 Ⅰ** 좌장: 김회린(KAIST) 음성 연구 40년의 회고 강사: 이용주(원광대)
11:30~12:20	**특강 Ⅱ** 좌장: 정민화(서울대) AI 시대의 언어지능 융합 강사: 김영길(ETRI)

구두 발표 Ⅰ

	좌장: **Jeffrey Holliday(고려대)**			
	시간	구분	제목	저자
음성학 (101호)	13:20~ 13:35	PH1	The effect of base token selection for stimuli manipulation on the perception of lenis and aspirated stops in Seoul Korean	Eunjin Oh(Ewha Womans Univ.)
	13:35~ 13:50	PH2	Effect of short-term English exposure on Korean children's English perception	Margarethe McDonald(Univ. of Wisconsin-Madison) & Eon-Suk Ko(Chosun Univ.)
	13:50~ 14:05	PH3	A cross-linguistic study of sound symbolism in personal names	Jihyo Kim, Hyunji Kim, Eon-Suk Ko (Chosun Univ.)
	14:05~ 14:20	PH4	Articulatory and acoustic characteristics of Korean and English word-final laterals produced by Korean learners of English	Young Hwang(Indiana Univ.)
	좌장: **하지완(대구대)**			
	시간	구분	제목	저자
말장애 및 음성의학 (111호)	13:20~ 13:35	SD1	내전형 연축성 발성장애 감별진단을 위한 선별검사의 내용타당도	조재경(강북삼성병원), 최성희(대구가톨릭대), 이상혁, 진성민(강북삼성병원)
	13:35~ 13:50	SD2	병리적 음성떨림의 음향음성학적 특성	서인효(단국대 의대), 성철재(충남대)
	13:50~ 14:05	SD3	온라인 뉴스 발화의 말 속도 특성	김수연, 김국환, 유재연(호남대)
	14:05~ 14:20	SD4	직업적 음성사용 패턴에 따른 음성 활동 및 참여 프로파일 비교	이승진(연세대 의대), 임성은(강남세브란스병원), 최홍식(하나이비인후과병원), 임재열(연세대 의대)
	좌장: **박기영(ETRI)**			
	시간	구분	제목	저자
음성공학 (112호)	13:20~ 13:35	SE1	자기 모방 발음 교정을 위한 GAN의 스펙트로그램 학습	양승희, 정민화(서울대)
	13:35~ 13:50	SE2	Learnable Dictionary Encoding 기반의 문장 독립 화자 검증 방법	정영문, 최연주, 김회린(한국과학기술원)
	13:50~ 14:05	SE3	스푸핑 음성 검출의 성능향상을 위한 다양한 고해상도 정보의 활용	심혜진, 정지원, 허희수, 유하진(서울시립대)
	14:05~ 14:20	SE4	가창 음원의 음악적 표현 이식 시스템	용상언, 남주한(한국과학기술원)

포스터 발표 Ⅱ (1층 로비)

좌장: **손민정(한남대), 박상희(대구사이버대), 박형민(서강대)**

시간	구분	제목	저자
14:30 ~ 15:30	P23	형태소 유형에 따른 한국어 단모음의 음성적 실현	이향원, 신지영(고려대)
	P24	다문화 가정 이주여성들의 한국어 파열음 유성개시시간 비교	박선영, 성철재(충남대)
	P25	한국인 영어 학습자의 영어 숙련도에 따른 모음 사이 폐쇄음의 유성음화 경향	유지윤, 전지현, 이석재(연세대)
	P26	YELC 음성코퍼스 구축과 과제 유형별 분석적 평가의 총체적 점수에 대한 예측 분석	이석재(연세대)
	P27	한국인 EFL 학습자 영어 발화에서 능숙도에 따른 모음약화 현상 연구	강태훈, 임유경, 이석재(연세대)
	P28	한국인 화자 연령 지각 연구	서윤정, 신지영(고려대)
	P29	Second language acquisition of intonation: nuclear peak alignment in English	Hye Jeong Yu(Hanshin Univ.)
	P30	주관적 음성장애 인지와 비염의 상관성 연구	이인애, 성철재(충남대)
	P31	아동의 유음 /ㄹ/에 대한 청지각적·음향학적 연구	김인경, 성철재(충남대)
	P32	소아의 안면근육 움직임 평가도구 개발: 예비연구	조태양, 봉성록, 정해철, 윤혜수, 박소현, 김성태(동신대)
	P33	음성장애 유형에 따른 음향학적 특성과 청지각적 평가 간의 상관성	이옥분(대구사이버대), 심희정(한림대)
	P34	증강현실 기반 음성치료 소프트웨어의 개발 및 적용	박희준(부산가톨릭대), 김근효, 이연우(부산대병원), 배인호(양산부산대병원), 권순복(부산대)
	P35	아동의 말소리산출 능력에 대한 음운산출지식(PPK) 분석 연구	이손희, 하지완(대구대)
	P36	일반아동의 자극양식, 자극길이, 음운유사성 및 시연여부에 따른 음운단기기억 수행력 비교	전은혜, 하지완(대구대)
	P37	정상 여성의 음성범위 프로파일과 발화범위 프로파일	김재옥(강남대), 이승진(연세대 의대)
	P38	심층 신경망의 화자 기저를 활용한 화자인증 손실 함수 정의 기법	허희수, 정지원, 심혜진, 유하진(서울시립대)
	P39	감정 음성 합성을 위한 토큰 기반의 음향 모델링	최희진, 박진욱, 한민수(한국과학기술원)
	P40	단어의 Text Label을 활용한 Acoustic Word Embedding 추출	정명훈, 김회린(한국과학기술원)
	P41	Research on Novel Speech Intelligibility Correlation Coefficient Based on Time-Frequency Domain SNRs	Seon Man Kim(KOPTI)
	P42	딥 러닝 오토인코더 기반 보코더 파라미터 감축	이정혁, 김홍국(광주과학기술원), 김선교(국방과학연구소)
	P43	자동차 잡음 제거를 위한 잡음이 혼합된 목표신호를 사용한 심층신경망의 성능	조현호, 권오욱(충북대)
	P44	속삭임 음성인식을 위한 비병렬 코퍼스 환경에서의 유사 속삭임 음성 생성	김민식, 정용원(부산대), 이성주(ETRI), 김형순(부산대)

구두 발표 Ⅱ

	좌장: **박시균(군산대)**			
	시간	구분	제목	저자
음성학 (101호)	15:40~ 15:55	PH5	선·후행 음절에 따른 L2 학습자의 한국어 폐쇄음 지각	정환민, 홍민경, Chiin Ngaihmuan Ngaihte, Jeffrey Holliday(고려대)
	15:55~ 16:10	PH6	성인 포먼트 측정에서의 최적 세팅 구현 - Praat software와 관련하여 -	박지연, 성철재(충남대)
	16:10~ 16:25	PH7	절의 통사·의미 유형에 따른 운율적 실현 양상	전태희, 신지영(고려대)
	16:25~ 16:40	PH8	서울말 2,3,4음절어의 운율패턴	변희경(국제교양대)
	좌장: **김성태(동신대)**			
	시간	구분	제목	저자
말장애 및 음성의학 (111호)	15:40~ 15:55	SD5	2-7세 일반아동의 발성유형 별 말소리발달 연구	류은주, 하지완(대구대)
	15:55~ 16:10	SD6	말·언어유창성 평가 시스템 SDUA 개발	심현섭(이화여대), 이수복(우송대), 박혜연, 곽효정(이화여대), 신동춘, 고연정(보리㈜), 윤재민(얄리㈜)
	16:10~ 16:25	SD7	SOVTE를 이용한 기능적 실성증의 음성치료 효과	김지성, 채혜림(대구가톨릭대), 이동욱(충북대병원), 최성희(대구가톨릭대)
	좌장: **박정식(한국외대)**			
	시간	구분	제목	저자
음성공학 (112호)	15:40~ 15:55	SE5	양방향 학습 방법을 이용한 Flow 기반의 뉴럴 보코더	이경훈, 이준모, 이협우, 김영익, 조훈영(엔씨소프트)
	15:55~ 16:10	SE6	CNN에 기반한 방송 콘텐츠에서의 음악 신호 분리	허운행(충북대), 김혜미(ETRI), 권오욱(충북대)
	16:10~ 16:25	SE7	나이 정보를 활용한 화자 인증 시스템	김주호, 허희수, 정지원, 심혜진, 유하진(서울시립대)
	16:25~ 16:40	SE8	Transformer 기반 Seq2seq 모델 및 WaveGlow 보코더를 활용한 음성합성 시스템	최연주, 서영주, 정영문, 최희진, 김회린(한국과학기술원)

특강 Ⅲ (101호)

시간	내용
16:40~17:30	**특강 Ⅲ** 좌장: 최성희(대구가톨릭대) 공연음성의 이론과 실제 강사: 박문숙(University of Arkansas)

차 례

특강 I

특강 II

포스터발표 I

구두발표 I

[음성학]

[말장애 및 음성의학]

[음성공학]

포스터발표 II

구두발표 II

[음성학]

[말장애 및 음성의학]

[음성공학]

특강 III

특강 I

좌장: 김회린(KAIST)

음성 연구 40년의 회고
(이용주, 원광대)

음성 연구 40년의 회고
남기고 싶은 이야기들

이 용 주

주 제

1. 한국음성학회의 발자취
   - 통합과 발전을 중심으로

2. 우리나라의 음성공학 연구의 발자취

우리나라의 음성공학 연구의 역사

IT 분야 연구소의 음성연구

외국의 사례
미국: AT&T 벨 연구소
일본: NTT 연구소, 전자기술총합연구원(ETL)
한국: 한국전자통신연구원
(ETRI: Electronics & Telecommunications Research Institute)

* 60~70년대: 체신부 전기통신연구소
* 80년대 이후 한국전자통신연구원
(KIST 부설 한국전자통신연구소 -> 한국통신기술연구소 ->
한국전기통신연구소 -> 한국전자통신연구소 ->
한국전자통신연구원)

음성공학 분야의 선구자

최진태(? ~ 1984?)

- 체신부 전기통신연구소 음향연구계장(~1979)
- ETRI 음향연구실장(1979~1982)

음향, 음성 연구 선구자 (음향학회 창립에도 기여(초대 감사)
(한국음향학회는 1981년 차일환 초대회장(본 학회 자문위원) 께서 주도하여 설립됨)

주요 연구들 (체신부 시절)
- 통화품질 연구 (아날로그 통신 시대)
- 전화기 송수화갑 개발(전화기 국가 보급 시절)
- 음성분석 기초연구들
  - spectrogram(Sonagraph)을 통한 음성 분석 연구
  - analog filter bank를 이용한 음성인식 기초연구

ETRI 의 초기 음성연구 (1980~)

학제적 연구를 위한 준비

- 음성학/언어학 (음성학?)
- 디지털신호처리 (음성공학?)
- 청각 및 음향, 인지 심리 (음성의학?)

1983년 이후 처음 디지털 음성처리에 의한 연구 시작

1985년부터 본격적인 음성연구 시작
(자동통역 연구, 음성기술을 이용한 언어교육, 음성 네비게이션 등)

•추억1 !

음성연구의 저변 확대를 위하여
한 해에 14개의 국내외 위탁연구 시행

또 한 분의 선구자

은종관교수(1940~1997)
KAIST 전자공학과 교수 역임

speech coding (RELP vocoder,..)
음성합성, 인식 연구

음성분야 인재 양성에 크게 기여

본격적인 정부의 지원

산자부와 정통부의 경쟁적 지원(2001~)
2000년 중순 경 산자부 디지털전자산업과의 음성분야 지원 계획
2000.12 음성정보기술산업 발전안 기획 및 발표

당시 정통부 크지는 않지만 ETRI의 음성연구를 간헐적 지원
2002년부터 대규모 연구비를 ETRI에 지원

산자부, 정통부, 문광부 등의 정부 부서 간 관할 영역에 따른 다툼
예: 문화컨텐츠 (정통부<-> 문광부)
음성기술 (산자부<-> 정통부)

결론: 기초 및 통신분야 응용-> 정통부
일반산업분야 응용 -> 산자부

2001.5 ~

(산자부)
음성정보 기술산업지원센터(SiTEC) 원광대에 설립(이용주)
음성정보기술산업협회 설립(이인섭)
중기거점사업 (자동차 음성 HMI 기술개발, 책임자 정민화)
음성기술 표준화 포럼
HCI 연구조합(정도상, 이용주)

(정통부)
ETRI 를 통한 대규모 연구비지원
음성처리산업협의회 설립

기억에 남은 일들

SiTEC 아카데미

1. 명 강의들을 통한 종합적인 음성연구 입문 교육과정(OJT교육처럼)
(음성학 + 음성공학)
이제는 우리 학회에서?
(음성학 + 음성공학 + 음성의학(언어치료))

2. 음성코퍼스 음운 및 운율 레이블링 과정 교육
국내 각 대학의 실험음성학 연구의 밑거름

3. 국립 국어원 및 한국방언학회
방언(지역어) 수집 및 전사 교육 과정
한국방언학회 설립에 기여

한국음성학회의 발자취

통합과 발전을 중심으로

연혁

- 1935.04 조선음성학회 창립 (간사 정인섭, 김상용, 이하윤)
  (발기인: 김선기, 김윤경, 방종현, 양주동, 이극로, 이숭녕, 이희승, 최현배, 홍기문)
- 1948.08 한국음성학회로 명칭 변경 (회장 정인섭)
- 1983.09 정인섭 별세로 학회의 활동 마감

- 1976.12 대한음성학회 창립(회장 이현복)
  (발기인: 김방한, 김영송, 성백인, 이명섭, 김선희, 표진희, 김학주, 함기선, 김차균, 이현복 등 20여명)

- 창립 초기의 연구 분야
  1) 일반 및 한국어, 영어 등의 개별어 음성학,
  2) 실험 및 음향 음성학,
  3) 언어 치료

- 학회의 활동
  - 학회지 "말소리" 발간
  - 음성학 연수회 (자격증: "말소리 갈닦기 보람" (음성학 자격증))
  - 한국어 영어 발음 진단사업 시행
  - 음성학 관련 자료 및 음성자료의 보급
  - 서울 국제 음성학 학술대회 개최
  - 일본 영어 음성학회와 공동 학술 행사

- 학회 발전을 위한 움직임들

  - 1990년대 중반? 제1차 발전위원회 가동
    (이용주, 이숙향, 김향희,....)

  - 2002.05 제2대 회장 취임 (유만근)

  - 2002. 5월 이후? 제2차 발전위원회 가동
    (이용주, 김주원....)

- **대한음성학회 발전위원회 회의록**

◈ 시간: 2002년 6월 17일 오후 4시 ~ 7시
◈ 장소: 충남대학교 문과대 248호 세미나실
◈ 참석자: (존칭생략, 가나다순)
- 음성학: 김주원, 성철재, 신지영, 이숙향, 이호영
- 음성공학: 김형순, 이용주, 한민수
- 저녁회식에 구명완, 김회린, 오영환
- 회의록 작성: 성철재

주요 논의 내용:
- 조직 개편(평의원 제도도입, 회장 1년 임기, 사단법인화 등)
- 이후 이용주 김주원 공동위원장 주관으로 구체적인 활동 진행

이후, 음성공학분야 및 언어치료 분야 연구자 대거 합류

- 2003.03 사단법인 인가 (정보통신부) [대한음성학회]
- 2004.~ 김종교, 김기섭, 오영환, 김상준, 이용주 회장 역임

- 한국음성과학회 연혁
- 1996.11 이전 음성연구회? (김대원, 구희산, 고도흥, 김현기, ...)
- 1996.11 한국음성과학회 창립 (회장 김대원)
- 1999.~ 차일환, 신효근, 구희산, 최홍식, 고도흥, 정옥란, 이정원, 김현기 회장 역임
(최홍식, 차일환, 신효근 이사장 역임)

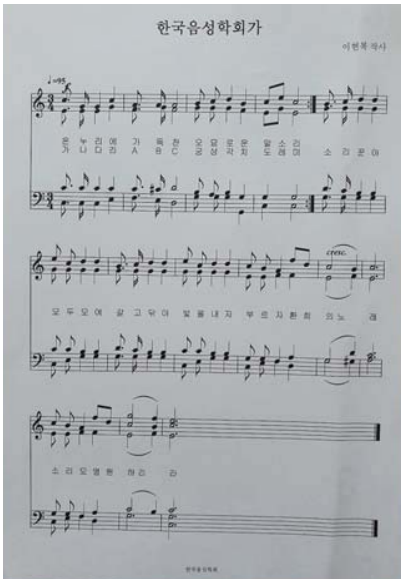
- 한국음성학회

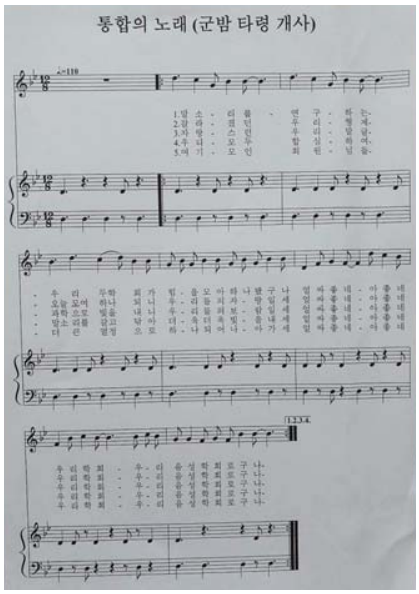
  - 2007 양 학회 통합 학술대회
  - 2008 통합추진위원회 발족
  - 2008.11 양 학회 통합 학술대회
  - 2009.~ 김기호, 김주원, 한민수, 이숙향, 구명완, 김형순, 양병곤, 조철우, 이호영, 김수진, 김회린 회장 역임.

- 통합의 추억

  학회명: 한국음성학회 (대한음성학회 + 한국음성과학회)
  학회지: 말소리와 음성과학

- 통합학회 학회가
- 통합의 노래


한국음성학회가
이현복 작사


통합의 노래 (군밤 타령 개사)

강연을 마치며

음성분야의 지속적인 발전을 위하여
- 산학연의 긴밀한 협력
- 학회의 학술활동 강화(특히 학제적 협력활동의 강화)
- 구성원 간의 협력의 활성화

- 음성학분야 : 통합 이후의 발전
- 음성의학 및 언어치료 분야: 응용 분야에 따른 경쟁 발전
- 음성공학 분야: 학회 별 경쟁 및 상생 발전
  (특히 음성학회와 음향학회)

* 협력방안(예)
일본음향학회와 전자통신학회의 음성연구회 운영 사례

최근
음성정보기술 및 산업 발전을 위한 워크샵 개최
(2019.4.25.~27, 제주)

- 음성포럼의 결성
- 소속학회에 상관없이 음성공학 분야의 산학연 협력을 위한 위한 구심체

(단톡방 가동하고 가칭 Speechnet (mailing list) 운영 준비중)

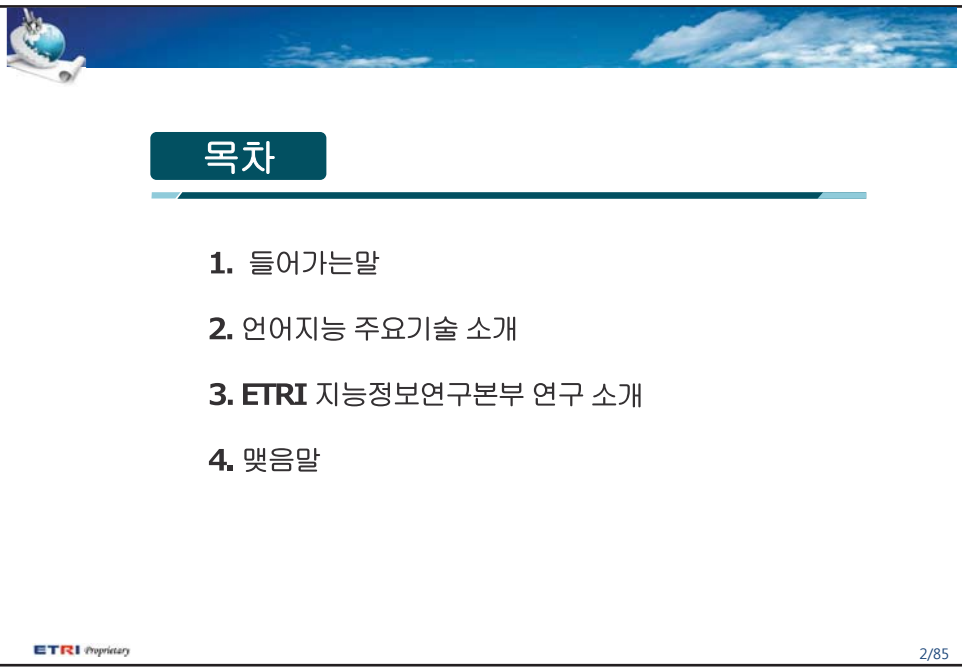
감사합니다.

특강 II

좌장: 정민화(서울대)

AI 시대의 언어지능 융합
(김영길, ETRI)


ETRI
www.etri.re.kr
AI 시대의 언어지능 융합
2019. 6. 1.
SW.콘텐츠연구소
언어지능연구그룹 김영길


목차
1. 들어가는말
2. 언어지능 주요기술 소개
3. ETRI 지능정보연구본부 연구 소개
4. 맺음말
ETRI Proprietary
2/85

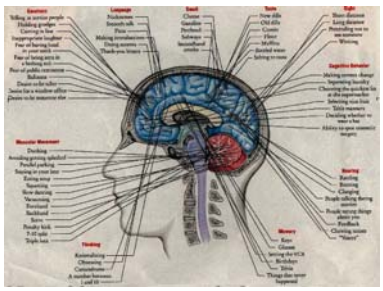
1. 들어가는말

ETRI Proprietary

3/85

들어가는말 – AI 정의

인간이 하는 지적 활동
(인식, 이해, 학습, 추론, 예측 -> 문제풀기, 창작)을
컴퓨터가 하도록 하는 모든 기술



듣기, 말하기, 걷기, 물건 들기, 길찾기, **보기**, **정보 찾기, 요약하기, 번역하기**, 게임하기, **정해진 업무 대화하기**, **퀴즈 풀기**, … …,

인간이 하는 생각, 판단, 예측 따라 하기 (전문가 시스템: 의사, 판사, 교사, …)

새로운 생각하기, 새로운 창조하기, 새로운 판단하기, 감정을 가지기

ETRI Proprietary

4/85

들어가는말 - 딥러닝

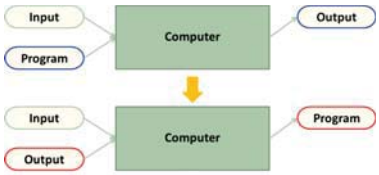
개념 및 내용

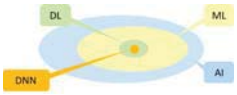
Machine Learning

- 1959 Arthur Samuel
- 컴퓨터에게 배울 수 있는 능력, 즉 코드로 정의하지 않은 동작을 실행하는 능력에 대한 연구분야(출처 : 위키피디아)

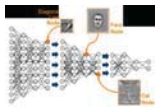
Deep Learning

- 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화를 시도하는 기계학습 알고리즘의 집합
- Deep Neural Network : 입력 계층과 출력 계층 사이에 복수개의 은닉 계층들로 이뤄진 인공신경망





※ AI : Artificial Intelligence
※ ML : Machine Learning
※ DL : Deep Learning
※ DNN : Deep Neural Network



2006 : The Deep Breakthrough

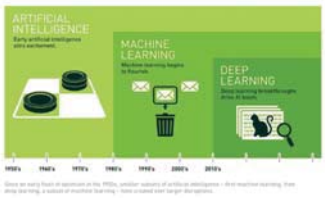
- Before 2006 : Failing to train deep architectures
- 3 Papers : A fast learning algorithm for deep belief nets, Greedy layer-wise training of deep networks, Efficient learning of sparse representations with an energy-based model)

들어가는말 - 딥러닝

현재 혁신적 인공지능 기술 발전을 주도

인간이 사전에 고정한 모델에 의존 없이 학습을 통해 스스로 진화

- 음성과 이미지 인식 분야의 획기적 성능 향상
  - 얼굴인식률: 구글 '페이스넷(FaceNet)' 99.96%
- 가능성
  - 추상적 개념 기반 작업의 해결가능성 제시
- 한계
  - 사물의 이해 측면에서 아직 초보적 수준
- 향후 연구방향
  - 추론 과정에 대한 체계적 설명 (Explainable AI)
  - 인간 지식에 의존하지 않는 학습 (Unsupervised Learning)



(출처: NVIDIA)

들어가는말 – 플랫폼 전쟁

"플랫폼을 선점하는 자가 AI 세상 지배"
(에릭 슈미트, 2016.3.24, GCP 사용자 컨퍼런스)

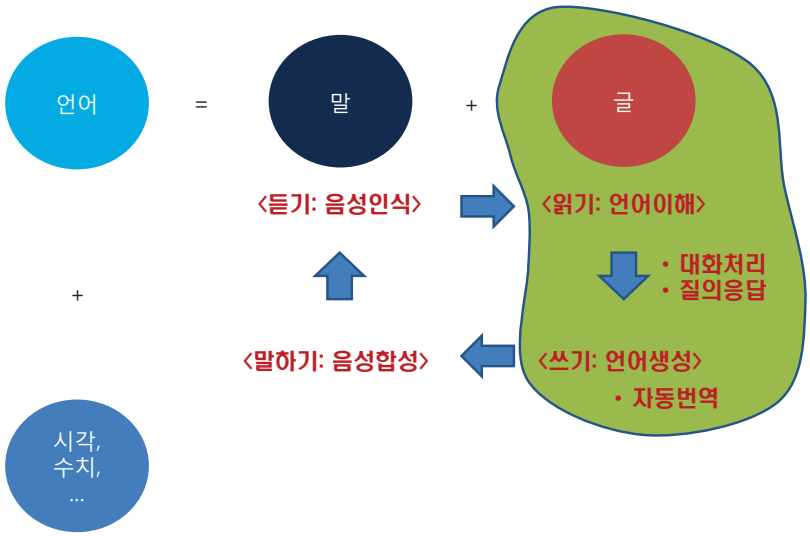
AI는 전문기업에 의해 구현되고, 플랫폼을 통한 서비스 형태로 활용되는 환경 도래

- 스마트폰 기반 또는 거치형 음성인식 개인비서 서비스가 치열한 경쟁
  - 스마트폰 기반: 애플 시리('11) 이후 구글 나우('12), MS 코타나('14) 등
  - 거치형: 아마존 에코, 구글 홈, 레노버 스마트 어시스턴트 등
- Google 기술력, 방대한 데이터 기반 범용 플랫폼 구축 (ML Platform)
- facebook 고도화된 개인맞춤형 인공지능 플랫폼
- IBM 의료, 금융 등 도메인 특화 인공지능 플랫폼 (Watson)
- amazon 알렉사 음성대화 플랫폼 (Alexa Platform)

http://blog.lgcns.com/1400 일부 인용

ETRI Proprietary 7/85

들어가는말 – 언어지능



ETRI Proprietary 8/85

들어가는말 – 융합

■ 4차산업 혁명의 핵심, AI 융합의 시대

- 기존 산업분야에서의 응용 도메인 융합(의료, 관광, 국방, 자동차, 금융 등)

인공지능 자연어처리는 IT 분야뿐만 아니라
경영, 마케팅, 디자인, 예술, 창업 등
모든 분야에 접목 가능해.

- 기술적 융합(언어, 음성, 시각, 감성, 상황, 행동, 추론 등 -> 복합 지능)
  . Knowledge based method, Data based method
  . Supervised learning, Unsupervised learning, Reinforcement learning, …

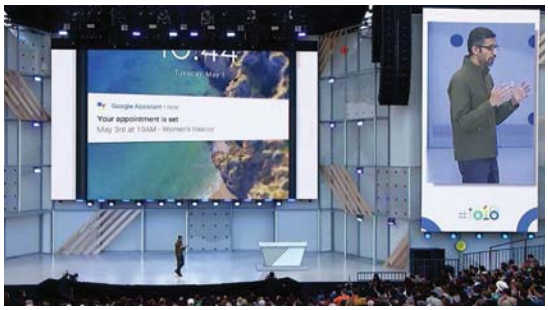
■ 기술적 융합(언어, 음성, 시각 등)의 예

- AI 스피커, 콜센터 AI
  . 음성인식/합성, 대화처리, 질의응답, (감성인식) 등
- 음성 통역, 이미지 번역
  . 음성인식/합성, 자동번역, 문자인식 등
- 대화형 외국어 교육
  . 발음평가, 음성인식/합성, 대화처리, 문법처리 등

ETRI Proprietary 9/85

들어가는말 – 융합 (음성+대화+질의응답: AI 스피커)

■ AI 스피커, 개인비서 (음성, 대화, 질의응답)



"(AI) 화요일 헤어컷을 예약하고 싶은데요~"

"(사람) 그러시군요. 잠시만요~.
몇 시를 괜찮으시겠어요?"

"오후 12시요."

"오후 12시엔 예약이 다 차서 안되겠네요.
요청하신 시각과 가장 가까운 예약 시간은
오후 1시 15분인데요~."

"그럼 혹시 오전 10시부터 오후 12시 사이에
가능한 시각은 없을까요?"

"으흠~~ 고객님께서 어떤 서비스를 받고자 하시는 지
에 따라
달라질 수 있는데요. 어떤 서비스를 받고 싶으신가요?"

"헤어트리 그냥 커트를 하려고 하는데요."

"그럼 오전 10시에 예약 가능합니다."

"오전 10시가 좋겠네요."

"좋습니다. 고객님 성함은 어떻게 되시죠?"

"리사입니다."

"네 좋습니다. 5월 3일 화요일 오전 10시에
리사 고객님을 뵙도록 할게요."

ETRI Proprietary 10/85

들어가는말 – 융합 (음성, 대화, 질의응답)

■ 차량용 음성대화 서비스 (음성, 대화)

한국일보 (풍자) (...

"차와 대화 나눠요"...신형 쏘나타에 카카오 'AI 비서'가 탄다
매일경제 - 2019. 3. 17.
현대자동차는 21일 출시하는 신형 쏘나타에 카카오의 인공지능(AI) 플랫폼 '카카오 i'를 활용한 음성인식 대화형비서 서비스를 적용한다고 18일 ...
현대차, 신형 쏘나타에 카카오와 개발한 음성인식 대화형서비스 최초 적용
한국일보 (풍자) (보도자료) (블로그) - 2019. 3. 17.
현대자동차, 신형 쏘나타에 카카오 인공지능 비서 탑재
로봇신문사 - 2019. 3. 18.
신형 쏘나타, 차량 커넥티비티 연결 '카카오 i' 인공지능 비서 최초 적용
오토데일리 (보도자료) - 2019. 3. 17.
현대차 '신형 쏘나타', 대화로 작동하는 AI 비서 서비스 제공
조이스경제 - 2019. 3. 17.

■ 스마트홈 서비스 (음성, 대화)

조선일보

카카오-호반건설, 스마트홈 서비스 적용 '맞손'
대한금융신문 - 4시간 전
특히 대화형 스마트홈 서비스를 통해 음성 명령만으로 냉난방과 조명, 출 등 다양한 기능을 컨트롤 할 수 있게 된다. 카카오미니를 ...
카카오, 호반건설과 홈 IoT MOU 체결
조선일보 - 3시간 전
거실에 에어컨 켜줘...카카오-호반건설과 홈 IoT 기술 공동 개발
비아이뉴스 - 3시간 전
카카오-호반건설, 홈 IoT 기술 '맞손'
브릿지경제 - 3시간 전
카카오-호반건설, 스마트홈 구축 맞손..."카카오i 빌트인"
뉴시스 - 4시간 전

■ 금융 AI 서비스(음성, 대화, 질의응답)

은행권 AI 어디까지 왔나..디지털 금융 총력전

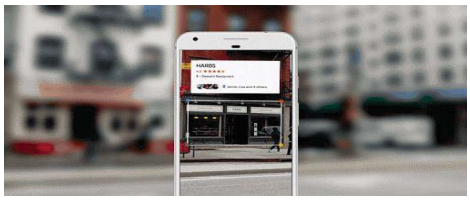
ETRI Proprietary

11/85

들어가는말 – 융합 (음성, 문자인식, 번역)

■ Wearable Speech Translator (음성+번역+웨어러블 IF: 웨어러블 통역)

■ Google Lens (문자인식, 위치인식, 검색, 번역: 이미지 번역, 정보서비스)



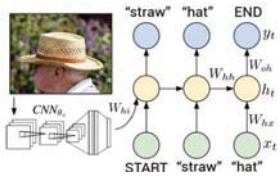
ETRI Proprietary

12/85

들어가는말 – 융합 (영상, 언어)

■ Image Captioning(CNN+LSTM)

Network Topology



<출처: https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2>

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

2. 언어지능 주요 기술 소개

(언어/의미분석, 대화처리, 자동번역)

주요 기술 소개 – 언어

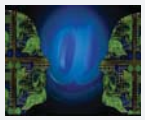
개념 및 내용

사람의 **언어에 대한 분석, 이해, 생성, 지식화 등을 다루는 기술**로서
대표적인 응용 분야는 정보검색/질의응답, 대화처리, 자동번역 등 있음

정보검색/질의응답

자동통번역

대화처리



.....

주요 기술 소개 – 언어(언어/의미분석)

소피스트란 그리스어로 지혜로운 자 또는 지혜를 만들어내는 사람이라는 뜻으로, BC 5~4세기의 그리스의 철학자들을 말한다.
이들은 아테네 사람들을 대상으로 하였고, 수사학과 웅변술을 가르쳤다.

형태소 분석

소피스트/NNG+란/JX 그리스/NNP+이/XSN+로/JKB 지혜롭/VA+ㄴ/ETM 자/NNB 또는/MAG 지혜/NNG+를/JKO 만들/VV+어/EC+내/VX+는/ETM 사람/NNG+이/VCP+라는/ETM 뜻/NNG+으로/JKB+,/SP BC/SL 5/SN+~/SO+4/SN+세기/NNP+의/JKG 그리스/NNP+의/JKG 철학/NNG+자/XSN+들/XSN+을/JKO 말/NNG+하/XSV+ㄴ다/EF+./SF 이/NP+들/XSN+은/JX 아테네/NNP 사람/NNG+들/XSN+을/JKO 대상/NGG+으로/JKB 하/VV+았/EP+고//EC+,/SP 수사/NNG+학/XSN+과/JC 웅변/NNG+술/XSN+을/JKO 가르치/VV+았/EP+다/EF+./SF

개체명 인식

<CV_OCCUPATION:소피스트/NNG>+란/JX <CV_LANGUATE:그리스/NNP+이/XSN>+로/JKB 지혜롭/VA+ㄴ/ETM 자/NNB 또는/MAG 지혜/NNG+를/JKO 만들/VV+어/EC+내/VX+는/ETM 사람/NNG+이/VCP+라는/ETM 뜻/NNG+으로/JKB+,/SP <DT_DURATION:BC/SL 5/SN+~/SO+4/SN+세기/NNP>+의/JKG <LCP_COUNTRY:그리스/NNP>+의/JKG 철학/NNG+자/XSN+들/XSN+을/JKO 말/NNG+하/XSV+ㄴ다/EF+./SF 이/NP+들/XSN+은/JX <LCP_CAPITALCITY:아테네/NNP> 사람/NNG+들/XSN+을/JKO 대상/NGG+으로/JKB 하/VV+았/EP+고//EC+,/SP <FD_ART:수사/NNG+학/XSN>+과/JC <FD_ART:웅변/NNG+술/XSN>+을/JKO 가르치/VV+았/EP+다/EF+./SF

어휘의미분석

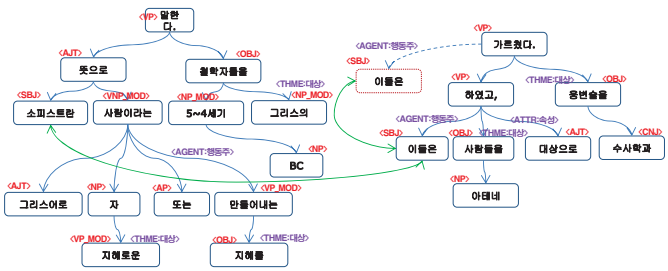
소피스트란 그리스어로 지혜로운 자__18_0000/NNB 또는 지혜__02_0001/NNG+를 만들__00_0101/VV+어내는 사람__00_0001/NNG+이라는 뜻__00_0002/NNG+으로, BC 5~4+세기__03_0002/NNG+의 그리스__02_0000/NNP+의 철학자들을 말하__00_0101/VV+ㄴ다.

의존구문분석

의미역 인식

상호참조해결

무형대용어 생략복원

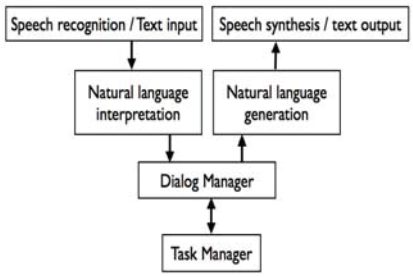


주요 기술 소개 – 언어(대화처리)

개념 및 내용

컴퓨터와 인간이 보다 자연스럽고 쉽게 인터랙션할 수 있도록 지원하는 기술

- **응용 분야**
  - 전화를 이용한 정보 시스템 (ARS 등)
  - 손을 이용한 기기 작동 등이 불가능한 상황에서의 보조 수단 (자동차 운전, 수술 등)
  - 외국어 말하기 학습
  - AI 스피커, AI 콜센터 등

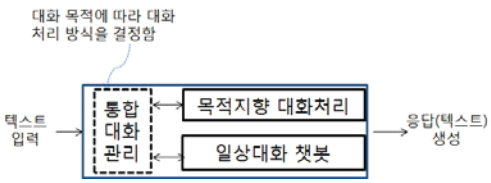


- **대화 주도(Initiative)에 따른 분류**
  - system-initiative
  - user-initiative
  - mixed-initiative

주요 기술 소개 – 언어(대화처리)

주요연구 동향

- **대화처리 기술 구성도**



- **대화처리 방법론**

중분류	소분류	방법론
대화처리	목적지향 대화처리	슬롯 채움(Slot-Filling) 방식
		정보 상태(Information State) 방식
		태스크 기반 방식
	일상대화 챗봇	패턴 매칭 기반 챗봇
		정보 검색 기반 챗봇
		생성 기반 챗봇

주요 기술 소개 – 언어(대화처리)

주요연구 동향 **Chatbot의 방법론**

- **Retrieval-based models (easier)**
  - 사전에 정의된 응답 중, 적절한 것을 선택하여 출력 (새로운 텍스트 생성 불가능)
  - 응답 선택 방식은 간단한 규칙기반 또는 머신러닝 기법을 이용

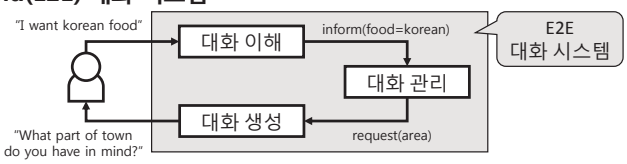
- **Generative models (harder)**
  - 사전 정의된 응답에 의존하지 않고 새로운 응답을 생성
  - 자동번역 기술과 유사한 방법론
    ※ 언어와 언어의 변환규칙 대신, 입력과 출력문장의 변환규칙 생성

Deep Learning techniques can be used for both retrieval-based or generative models, but research seems to be moving into the generative direction

ETRI Proprietary 19/85

주요 기술 소개 – 언어(대화처리)

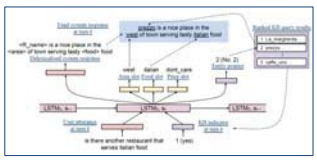
- **End-to-End(E2E) 대화 시스템**



- 기존 대화 시스템(modular 방식)
  - 대화 이해(NLU), 대화 관리(DM), 대화 생성(NLG)
- 기존 대화 시스템의 단점
  - 새로운 도메인에 적용하기 어려움
  - Error가 모듈 간에 전이되는 현상이 발생함

=> E2E 대화 시스템은 목적 함수를 직접 최적화 ->이러한 단점들을 극복

- **연구 배경 및 흐름**
  - E2E chat-oriented 대화 시스템의 성공이 배경
  - Task-oriented Trainable E2E 모델들이 다수 제안되고 있음



Bing Liu, Ian Lane, " An End-to-End Trainable Neural Network Model with Belief Tracking for Task-Oriented Dialog", 2017

ETRI Proprietary 20/85

주요 기술 소개 – 언어(대화처리)

현안 및 이슈

- 자연스러운 대화형 서비스를 위해서는 "목적지향 대화처리 기술", "챗봇 기술", "Q&A 기술"의 통합 처리가 필요함
  → 기술적 난이도가 매우 높아짐
  - 목적지향 대화처리 기술 : 상용 개인비서 서비스에서는 단순한 Slot-Filling 방식을 채택하고 있으며, 대화이해 성능 개선, 대화관리/추적에 대한 연구가 진행
  - 챗봇기술 : End-to-End 형태의 DNN 기반 챗봇 기술 연구가 활발하나 아직 좋은 성능을 보이지 않음
  - Q&A기술 : 기존 단답형 Q&A에서 서술형 Q&A 기술 및 심층 질의응답을 위한 지식화 기술 연구가 진행됨
- 음성인식 기술에 비하여 도메인 의존도가 높은 특징이 있음.
  이 문제를 해결하기 위하여 대화지식 구축 플랫폼 및 환경 지원이 필수적임.

ETRI Proprietary 21/85

주요 기술 소개 – 언어(자동통번역)

개념 및 내용

한 언어의 말 또는 텍스트를 다른 언어의 말 또는 텍스트로 자동으로 통역 또는 번역하는 기술

ETRI Proprietary 22/85

주요 기술 소개 – 언어(자동통번역)

주요연구 동향

- **Google, MS 등 글로벌 기업, 중국의 바이두, 일본의 NTT, 국내의 시스트란인터내셔널를 중심으로 다국어 자동통번역 기술 경쟁이 치열함**
  - **(Google)** 통계/신경망 기반 자동번역 기술에 기반하여 한국어를 포함한 100개 국어 이상의 다국어에 대한 자동통번역 서비스를 실시함
  - **(MS)** 화상 전화 통역 등 영어를 중심으로 연속 발화에 대한 실시간 자동통역 서비스를 시작함
  - **(바이두)** 광동어, 고대 중국어, 중국어 번체를 포함하여 총 26개 언어간 웹서비스 및 모바일 자동통번역 서비스를 실시함
  - **(NTT)** 국가적으로 지원을 받아 NICT와 함께 2020년 동경올림픽을 목표로 본격적으로 다국어 자동통번역 기술 개발에 착수함

    * 일본 총무성에서 글로벌 커뮤니케이션 계획(http://gcp.nict.go.jp)을 수립하였으며, 2015~2019년 총 100억엔 예산 투입
  - **(시스트란인터내셔널)** 한중일영 자동통번역 및 130여 언어의 다국어 문서 자동번역 솔루션을 확보함.

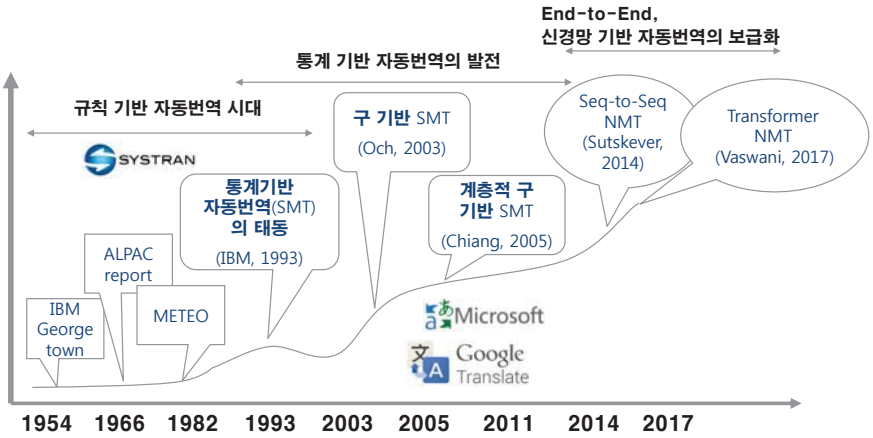
ETRI Proprietary 23/85

주요 기술 소개 – 언어(자동통번역)

- **최근에 자동통번역 기반의 다양한 응용 솔루션이 시장에 나타남**
  - **MS는 Skype를 인수하고, 2015년 영어와 스페인어, 독일어, 이탈리아어, 중국어 등 실시간 화상 자동통역 서비스를 시작함**
  - **Google은 증강현실 번역 앱 'Word Lens'의 퀘스트비주얼을 인수하고, 증강현실 자동번역 서비스를 실시함**
  - **채팅 자동번역, SNS 자동번역, OCR 자동번역 등의 다양한 스마트폰 앱 출현(대부분 Google translator API 사용)**
- **실시간 동시 통역, 멀티모달 기술 개발 등 미래형 기술 연구 진행 중**
  - **전화 대화, 강연, 연설, 구두 발표용 실시간 동시통역 기술 개발 연구 진행 중**
  - **얼굴 인식에 의한 감정 전달, 멀티모달 인터페이스에 관한 연구 등**

ETRI Proprietary 24/85

주요 기술 소개 - 자동번역 기술 진보의 흐름도



ETRI Proprietary

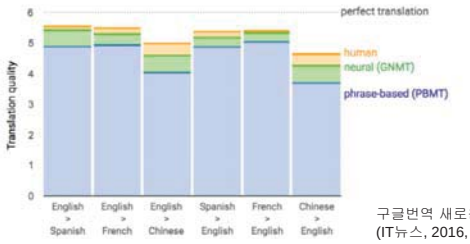
25/85

주요 기술 소개 - 언어(자동통번역)

- NMT 방법론에 의한 번역 성능 향상(번역 오류가 55%~85%가량 감소)

  "문장을 쪼개지 않고 문맥을 중시해 문장 전체로 번역한다는 점에서 지난 10년 간 업데이트를 뛰어넘는 도약을 했다" (2016.11, Google)

Input sentence:	Translation (PBMT):	Translation (GNMT):	Translation (human):
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.



구글번역 새로운 알고리즘 'NMTS'
(IT뉴스, 2016, 9, 29)

ETRI Proprietary

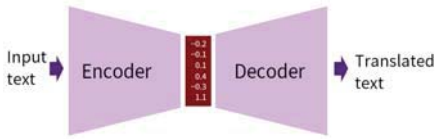
26/85

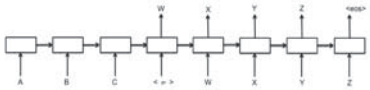
자동번역 기술 – 신경망 자동번역 기술 부상

신경망 자동번역 기술

- 'Sequence to Sequence' or 'Encoder-Decoder' Architectures

  : Combination of Many – to – One(Encoder) Neural Network and One – to – Many (Decoder) Neural Network



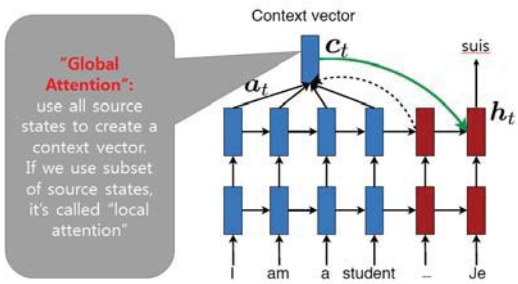


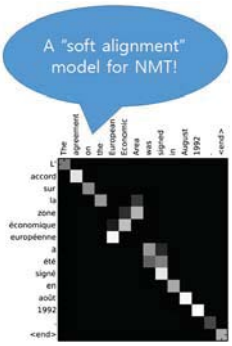
<From ACL 2016 NMT Tutorial Presentation(Luong-Cho-Manning-NMT-ACL2016-v4.pdf), and Sutskever et al., 2014 paper>

자동번역 기술 – 신경망 자동번역 기술 부상

신경망 자동번역 기술: Attention mechanism

- In Vanilla NMT, source context is used only once, to initialize decoder state
- Bahadanau et al., "Neural Machine Translation by Jointly Learning to Translate and Align", ICLR'15. (Luong et al., "Effective Approaches to Attention-based NMT", EMNLP'15 is better(=easy) for me)





<From ACL'16 NMT Tutorial Slides (Luong-Cho-Manning-NMT-ACL2016-v4.pdf)>

자동번역 기술 – 신경망 자동번역 기술 부상

신경망 자동번역 기술: Sub-word tokenization strategies

- 단어보다 더 작고, 문자(음절) 단위 보다 더 큰 단위를 사용해보는 건 어떨까?
- Taboo: 전통적인 Word-to-Word, Phrase-to-Phrase Transfer 과정을 필요로 하는 MT 시스템에서는 생각해서는 안될 것
- ASR(LM)에서 먼저 시작된 Sub-word tokenization - Wordpiece model (Schuster et al., 2012)
- Word 기반의 NMT 모델의 약점
  - 큰 Vocabulary 사용의 어려움 – "GPU의 VRAM은 12GBytes로도 충분하지 않다. 비싸서 못하는 것일 뿐."
    - 더 큰 Vocabulary를 위해 더 큰 Word Embedding Dimension을 요구
    - 대역어 선택 과정 간 Softmax()에서 발생하는 연산 부하량이 매우 높음
  - Copy mechanism으로는 OOV/Unseed words를 핸들링하기 충분하지 않다
    - Transliteration, Multi-word alignment, Rich morphology, Informal spelling 등
- NMT: Word보다 작은 단계를 Transferring 하기 위한 상황이 준비됨 – 많은 연구에서 적용
  - Byte-pair encoding scheme (aka BPE; Sennrich'16)
  - Mixed Word/Character Model (Wu et al., 2016, a Famous GNMT paper)

ETRI Proprietary

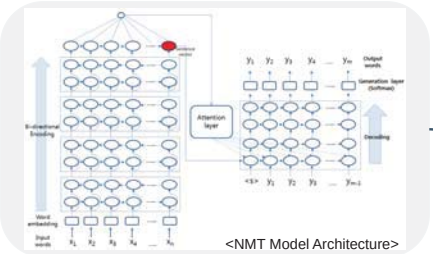
29/85

자동번역 기술 – 신경망 자동번역 기술 부상

신경망 자동번역(Neural Machine Translation; NMT) 기술 개발

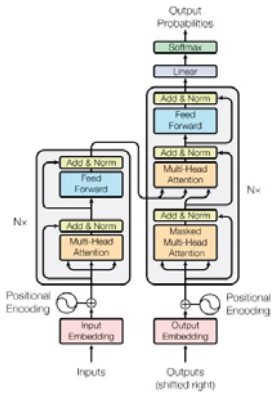
: 최근에 RNN-LSTM 모델에서 Transformer Model로 발전하고 있음

- Attention-based **RNN-LSTM** Seq2seq Encoding-decoding Model



<NMT Model Architecture>

- Multi-head Attention based **Transformer Model**



ETRI Proprietary

30/85

자동번역 기술 – Transformer Architecture(1/3)

- Standard Encoder-Decoder Structure
  - Encoder
    - Composed of a stack of N=6 layers
    - Each layer has 2 sub-layers
      - "Multi-head Self-Attention Mechanism"
      - Simple, position wise Fully-connected, Feed-forward NN
  - Decoder
    - Also composed of a stack of N = 6 identical layers
    - Each layer has 3 sub-layers
      - Third sub-layer: "Multi-head attention over the output of encoder stack"

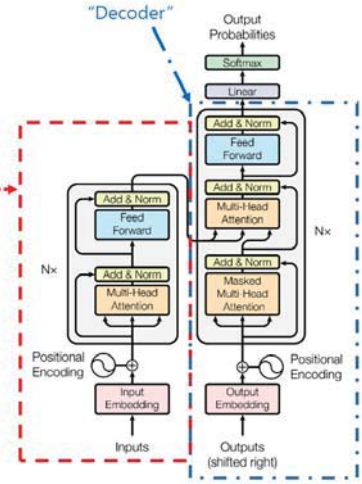


Diagram from "Attention is all you need", Vaswani et al., 2017

자동번역 기술 – Transformer Architecture(2/3)

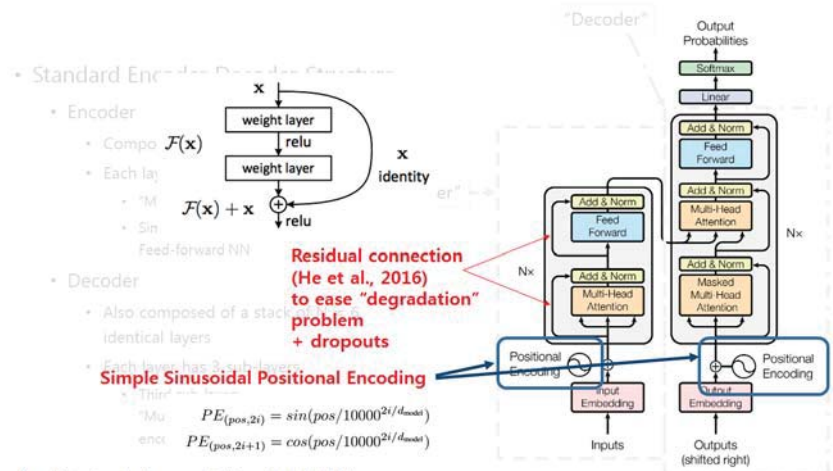


Diagram from "Attention is all you need", Vaswani et al., 2017
Residual block images from "Deep Residual Learning for Image Recognition", He et al., 2016

자동번역 기술 – Transformer Architecture(3/3)

- Standard Encoder-Decoder Structure
  - Encoder
    - Composed of a stack of N = 6 identical layers
    - Each layer has 2 sub-layers
      - "Multi-head Self-Attention Mechanism"
      - Simple, position wise Fully-connected, Feed-forward NN
  - Decoder
    - Also composed of a stack of N = 6 identical layers
    - Each layer has 3 sub-layers
      - Third sub-layer: "Multi-head attention over the output of encoder stack"

Layer Normalization (ba et al., 2016)

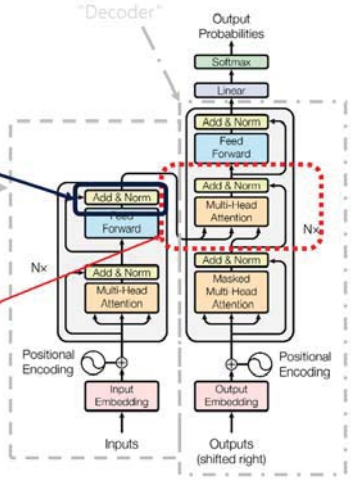
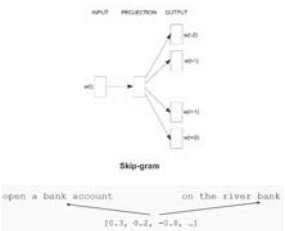


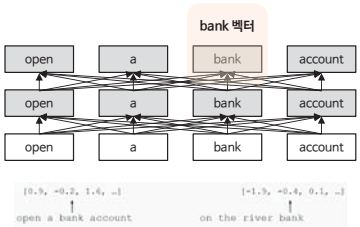
Diagram from "Attention is all you need", Vaswani et al., 2017

주요 기술 소개 - BERT

- BERT 개발 동기 (→ contextual representation)
  - 문제점
    - 딥러닝 자연어처리에서 단어를 벡터로 표현하는 워드임베딩은 필수적임
      - 심볼인 단어를 실수 벡터로 표현해야 뉴럴넷(i.e. FFNN) 적용 가능
    - 기존 워드임베딩 접근 방법은 문맥을 고려하지 못하는 한계를 지님
  - 접근 방법
    - 입력 문장(N개 단어)에 대해서, 뉴럴넷을 적용한 결과(N개 출력)를 단어의 워드임베딩 벡터로 고려



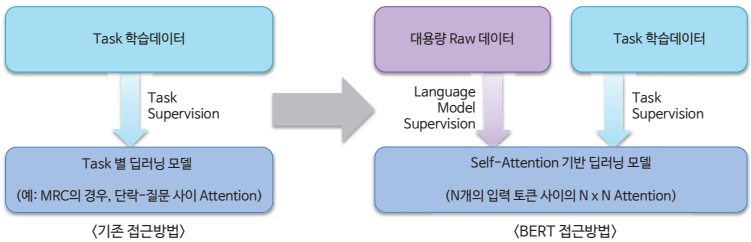
〈기존 워드임베딩 접근방법〉
(context free manner)



〈BERT 접근방법〉
(contextual representation)

주요 기술 소개 - BERT

- Contextual Representation 적용 전/후
  - 적용 전
    - 응용 태스트 별 별도의 딥러닝 모델과 학습 데이터를 이용하여 모델 개발
  - 적용 후
    - 1) 대용량 Raw 데이터로부터 공통 언어모델(Language Model) 미리 학습(pre-train)
    - 2) 학습된 공통 모델을 응용 태스크 별로 재학습(fine-tuning)하여 적용



〈기존 접근방법〉 〈BERT 접근방법〉

주요 기술 소개 - BERT

- 공통 언어모델 학습 단계
  - 1) 공백 단어 예측
    - 입력 문장 중, 15%의 단어를 masking 후 해당 단어를 맞추는 태스크

```
                    store          gallon
                      ↑              ↑
the man went to the [MASK] to buy a [MASK] of milk
```

    - BERT 특징: 양방향 정보를 이용한 단어 예측
      - 기존 연구: 단어 예측 시, 단일 방향의 정보만 고려
  - 2) 문장 선후관계 예측
    - 임의의 두 문장에 대해, 두 문장이 선/후 관계가 맞는지 맞추는 태스크

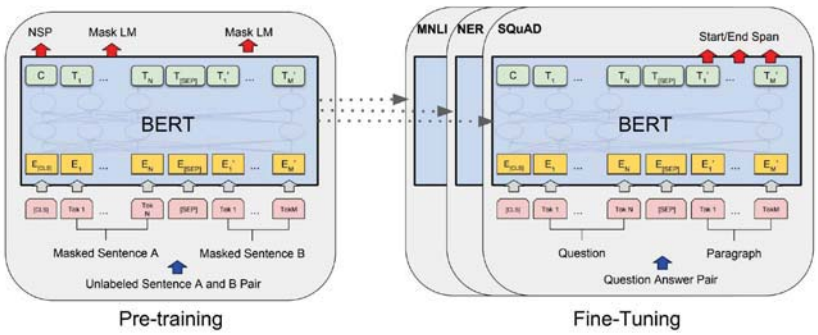
Sentence A = The man went to the store.
Sentence B = He bought a gallon of milk.
Label = IsNextSentence

Sentence A = The man went to the store.
Sentence B = Penguins are flightless.
Label = NotNextSentence

  - (비고) 두 태스크 모두 별도의 정답 말뭉치 없이 대용량 원시 말뭉치로부터 자동으로 학습 가능

주요 기술 소개 - BERT

- 응용 태스트 학습 단계
  - 언어처리, 텍스트 분류, 기계독해 등에 적용
    - 언어처리: 형태소분석, 개체명인식 등 언어처리 문제에 적용
    - 문장분류: 단일 문장 주제 분류 또는 두 문장 사이의 유사성 분석 문제에 적용
    - 기계독해: 질문과 단락을 입력 받은 후, 단락에서 정답 경계 인식 문제에 적용



주요 기술 소개 - BERT

- 자연어처리 분야 적용 현황
  - 기계독해를 비롯한 여러 태스크에서 SOTA(state-of-the-art) 기록
    - 스탠포드 MRC 챌린지(SQuAD): 상위 20여 팀에서 BERT 기반 모델 적용
    - 딥마인드 문장 분류 평가(GLUE): BERT 기반 multi-task 모델 최고 성능
    - 구글 문서 단위 질의응답 챌린지(NQ): BERT 기반 통합 학습 모델 최고 성능

유형	태스크	BERT 성능 (개선정도)
언어처리	개체명 인식 (* 평가셋: CoNLL2003)	92.8% F1 (+0.2%)
문장분류	GLUE (* 문장, 문장 쌍의 11종 분류 평가)	83.0% Score (+10.2%)
기계독해	SQuAD 2.0 (* 질문-단락 간 관계분석)	89.1% F1 (+6.2%)
질의응답	Google NQ (* 질문-단락 간 패러프레이즈 인식)	53.9% F1 (+22.5%)

3. ETRI 지능정보연구본부 소개

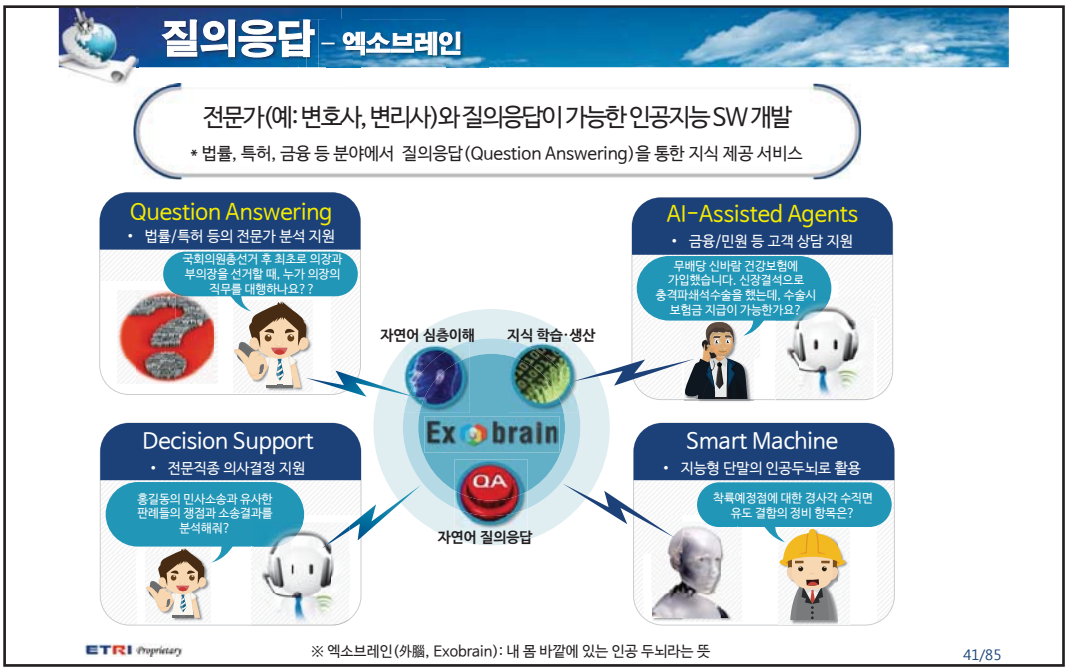
ETRI Proprietary

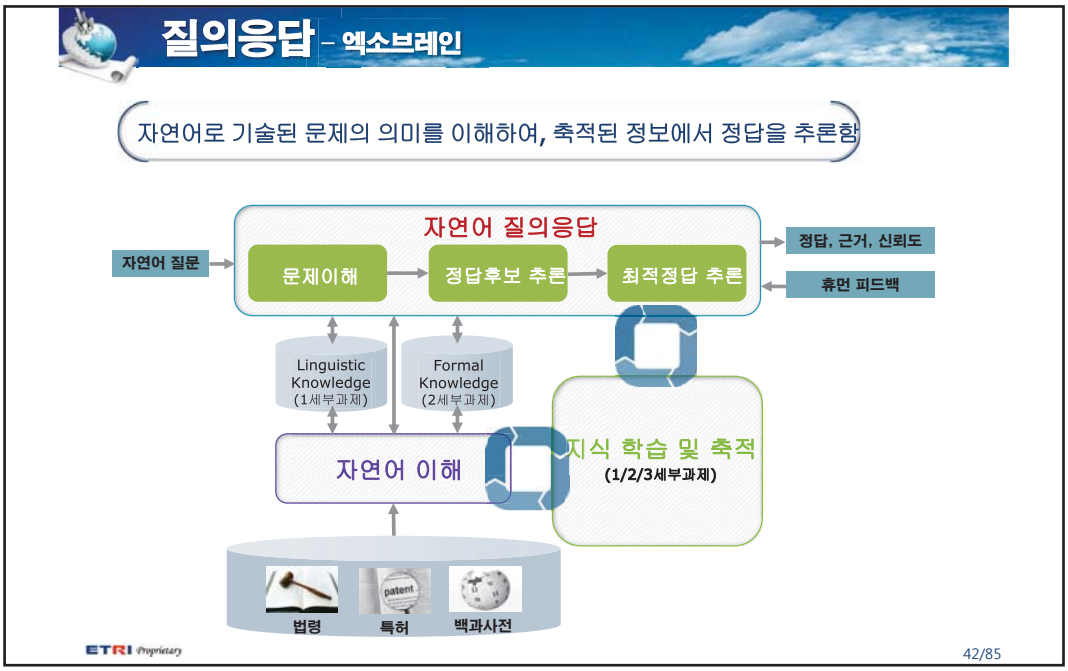
39/85

ETRI 지능정보연구본부

ETRI Proprietary

40/85


질의응답 – 엑소브레인
전문가(예: 변호사, 변리사)와 질의응답이 가능한 인공지능 SW 개발
* 법률, 특허, 금융 등 분야에서 질의응답(Question Answering)을 통한 지식 제공 서비스
Question Answering
• 법률/특허 등의 전문가 분석 지원
국회의원총선거 후 최초로 의장과 부의장을 선거할 때, 누가 의장의 직무를 대행하나요??
AI-Assisted Agents
• 금융/민원 등 고객 상담 지원
무배당 신바람 건강보험에 가입했습니다. 신장결석으로 충격파쇄석수술을 했는데, 수술시 보험금 지급이 가능한가요?
자연어 심층이해
지식 학습·생산
Exobrain
QA
자연어 질의응답
Decision Support
• 전문직종 의사결정 지원
홍길동의 민사소송과 유사한 판례들의 쟁점과 소송결과를 분석해줘?
Smart Machine
• 지능형 단말의 인공두뇌로 활용
착륙예정점에 대한 경사각 수직면 유도 결함의 정비 항목은?
ETRI Proprietary
※ 엑소브레인(外腦, Exobrain): 내 몸 바깥에 있는 인공 두뇌라는 뜻
41/85


질의응답 – 엑소브레인
자연어로 기술된 문제의 의미를 이해하여, 축적된 정보에서 정답을 추론함
자연어 질의응답
자연어 질문
문제이해
정답후보 추론
최적정답 추론
정답, 근거, 신뢰도
휴먼 피드백
Linguistic Knowledge (1세부과제)
Formal Knowledge (2세부과제)
자연어 이해
지식 학습 및 축적
(1/2/3세부과제)
법령
특허
백과사전
ETRI Proprietary
42/85

엑소브레인 오픈API서비스 ('17.10.30 시작, '18.12.11 개편)

- 중소 · 벤처 기업, 학교, 개인 개발자 등의 다양한 응용 개발 촉진과 인력 양성 지원 (http://www.etri.re.kr/aiopen)
  - 언어처리 학습데이터 및 구축 표준안 추가 : 국립국어원 및 학계에서 언어처리 표준안으로 사용
  - 언어처리 성능개선, 위키백과 QA API 추가

<언어처리 학습데이터 목록>

<언어처리 API>

<언어처리 학습데이터 구축 표준안>

<위키백과 QA API>

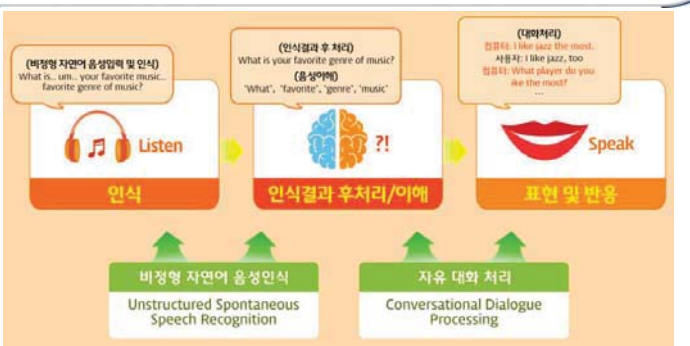
ETRI Proprietary

43/85

자유발화형 음성대화처리

개념 및 내용

자유발화형 음성 대화처리 = 비정형 자연어 음성인식 + 자유대화 처리



현재 기술 한계

- 비정형 자연어 음성인식 정확도
  - 65~70% (NIST 측정 결과 전화 대화 음성 인식률 40% 대)
- 자유대화처리 정확도 (비교 대상의 측정 결과 없음)
  - 60~75%로 예상 (SIRI 등 상용시스템의 경우 패턴기반 단순 대화처리 수준)

ETRI Proprietary

44/85

자유발화형 음성대화처리

대표적인 응용 서비스

음성대화처리 원천기술을 외국어 교육에 적용 – 지니튜터

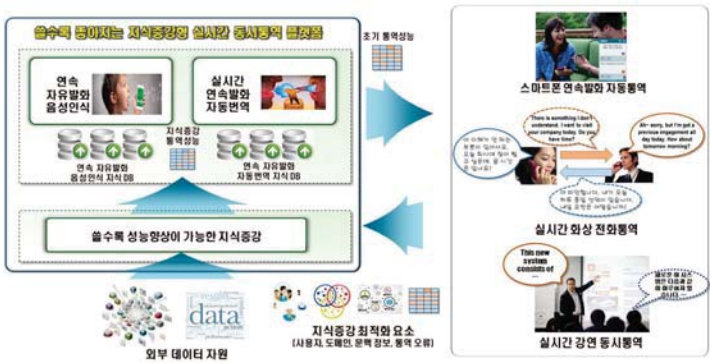
컴퓨터가 외국어 교사 역할을 대체/보완하여,
듣고(음성인식/이해), 말하고(대화처리), 가르쳐주는(평가/튜터링) 서비스

ETRI Proprietary 45/85

실시간 동시통역

개념 및 내용

관광/국제행사/전화/강연/회의/동영상 등에서 연속된 자유발화에 대해
의사소통 단위로 실시간 통역이 가능하고, 쓰면 쓸수록 성능이 개선되는
지식증강형 실시간 동시통역 원천기술 개발(2016.4~2020.12)



ETRI Proprietary 46/85

실시간 동시통역

"화상 전화, 강연 동시통역, 웨어러블 동시통역 등 **실시간 동시통역 기술 급부상**"

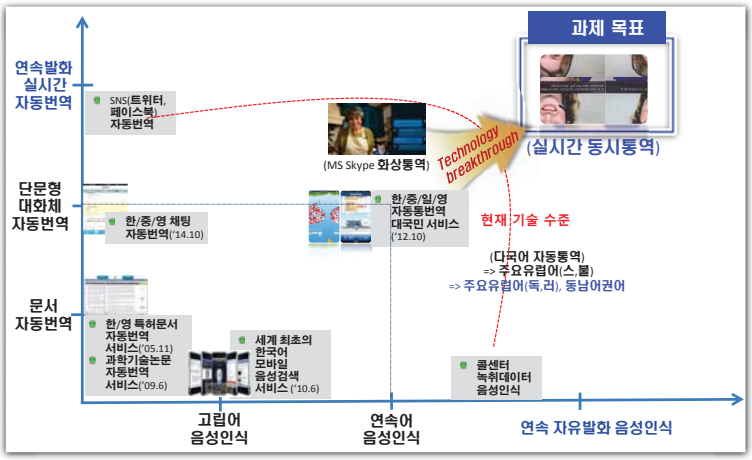
- (MS) 영어를 중심으로 Skype 화상 전화 통역 서비스
- (Google) Youtube 동영상 음성인식 및 자동번역 서비스

ETRI Proprietary 47/85

실시간 동시통역

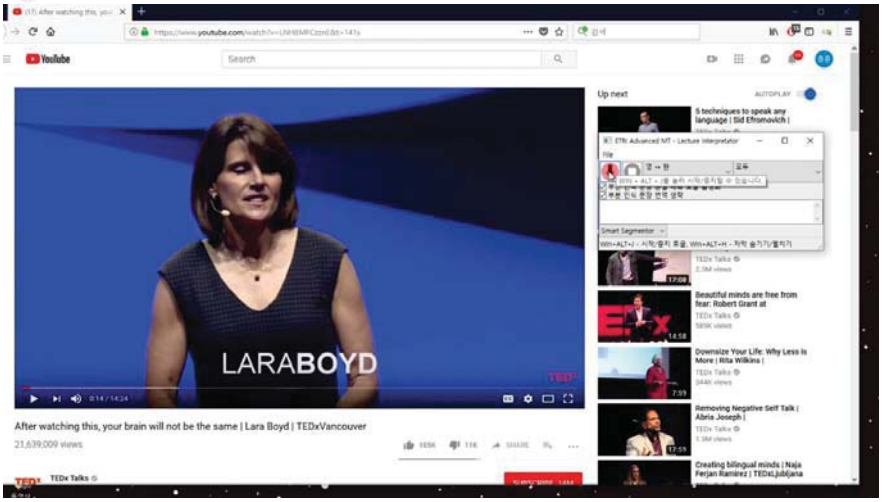
- 단문형 제한 발화 자동통역에서 **무제한 연속 자유발화 실시간 동시통역 원천 기술 확보**

※ 2015년 MS는 Skype 환경에서 영어를 중심으로 화상 자동통역 서비스를 시작함



ETRI Proprietary 48/85

실시간 동시통역 – 동영상 실시간 통역



4. 맺음말

맺음말

- 해외 글로벌 기업들은....
  - 구글 2001년부터 14년간 280억불 투자, IBM 왓슨 개발 20억불 투자,
  - 아마존 알렉사 개발 20억불 투자, MS 코타나를 위해 링크드인 262억불 인수 ....
- 국내 기업들은...
  - 삼성전자, 통신사, 네이버 등 포탈은 조단위 규모의 투자…
  - 중소/중견/벤처/스타트업은 ?
- ETRI는 ?
  - 선택과 집중 + Vertical AI → 핵심연구역량 확보
  - 핵심기술 공개 → 민간산업, 학교 생태계 활성화 지원
  - 국가 지능정보산업 발전에 기여

ETRI Proprietary 51/85

감사합니다

ETRI Proprietary 52/85

포스터 발표 I

좌장: 윤영도(동국대)
김재옥(강남대)
유하진(서울시립대)

과제, 세대, 성별에 따른 휴지 간 발화 길이의 실현 양상

유 도 영, 신 지 영
고려대학교 국어국문학과

A Realization of Utterance Duration between Pauses across Speech style, Gender and Generation

Doyoung Yoo, Jiyoung Shin
Korean Language and Literature, Korea University
doong723@korea.ac.kr, shinjy@korea.ac.kr

이 연구는 휴지 간 발화의 길이가 어떤 양상으로 나타나는지 관찰하여 이것이 인지적 차이나 사회 언어적 변수 간 차이를 어떻게 반영하고 있는지 관찰하는 것이 목적이다. 여러 연구들에서 발화의 양이 발화 계획 여부나 기억력과 같은 인지적 이유 및 발화 실수와 같은 유창성을 반영하며, 호흡군의 길이가 폐활량과 같은 신체적인 차이를 반영한다는 것이 보고되어 왔다[1][2]. 그런데 한국어를 대상으로 발화 길이에 대해 면밀하게 살핀 연구는 거의 수행되지 않았으며, 수행된 연구도 분석의 규모가 크지 않았다는 점이 아쉬움으로 남는다. 이 연구는 과제, 세대, 성별에 따른 발화 길이를 살펴 변수에 따른 발화 길이가 어떤 의미를 갖는지 관찰하였다. 이를 위하여 48명의 발화를 분석하여 휴지 간 발화 길이를 측정하였다. 발화 길이는 크게 두 가지로 나누어 살폈는데, 1) 발화에서 실현된 각 휴지 사이에 관찰되는 발화의 길이와, 2) 한 호흡으로 이어갈 수 있는 발화의 길이를 살필 수 있는 들숨 동반 휴지 간 발화 길이를 관찰하였다. 그 결과 발화 계획이 필요한 자유 발화와 휴지의 실현 빈도가 높은 장년층은 짧은 발화 길이가 나타나는 것으로 확인되었다. 한편 들숨 동반 휴지 간 발화 길이는 과제 간 차이를 보이지 않았는데, 이는 낭독 과제의 경우 호흡량에 무관하게 문장이 끝나는 곳에서 들숨을 쉬려는 경향이 나타났기 때문이다. 또, 폐활량이 가장 큰 청년층 남성이 가장 긴 들숨 동반 휴지 간 발화 길이를, 장년층 남성은 가장 짧은 들숨 동반 휴지 간 발화 길이를 보였으며 청장년층 여성의 들숨 동반 휴지 간 발화 길이는 차이를 보이지 않았다.

참고문헌

[1] Hnderson, Goldman-Eisler, & Skarbek. (1965). Temporal patterns of cognitive activity and breath control in speech, Language and Speech 8(4), 236-242.
[2] 신지영(2018). 언어 수행에서의 호흡과 기억. 한국어학 79. 한국어학회. 91-116.

시간 벌기 담화표지의 지연 정도와 분포

석 수 경, 신 지 영
고려대학교 국어국문학과

The degree of delay and distribution of delaying tactic discourse marker

Sukyung Seok, Jiyoung Shin
Department of Korean Language and Literature, Korea University
seok.sukyung89@gmail.com, shinjy@korea.ac.kr

본 연구에서는 시간 벌기 담화표지 간의 지연 정도와 분포 차이를 분석하고자 한다. 시간 벌기 기능은 담화표지의 주요한 기능 중 하나이지만, 기존 한국어 담화표지 연구들 중 해당 기능을 중심으로 담화표지 간 비교를 수행한 연구는 부족하다. 이에 시간 벌기 기능을 주로 수행하는 담화표지 '어, 음, 그, 뭐'를 대상으로 지연 정도와 분포의 차이를 알아보고자 한다. 이를 위해 20대 남녀 54명의 108개 자유 독백 발화 자료를 대상으로, 담화표지에 동반된 휴지의 비율과 길이 그리고 담화표지의 지속 시간을 측정하고, 운율 단위 내 위치와 선후행 품사에 따른 비율을 살펴보았다. 본 연구를 통해 말 산출 과정에서의 담화표지의 역할이 보다 주목받기를 기대한다.

인공지능 스피커 음성 특성에 따른 한국어 화자의 인식 양상

김 승 록, 신 지 영
고려대학교 국어국문학과

Korean Speakers' Perception by the Voice Characteristics of A.I. Speakers

Seungrok Kim, Jiyoung Shin
Department of Korean Language and Literature, Korea University
ohdde@naver.com, shinjy@korea.ac.kr

본 연구는 AI 스피커가 가진 음성적 특성이 스피커에 대한 사용자들의 인식 및 평가에 미치는 영향을 알아보는 데 그 목적이 있다. AI 스피커의 음성적 특성은 사회언어학적 변수인 성별(남성, 여성), 높임법(존댓말, 반말)의 변수로 구분되었으며, 각각의 음성을 모두 제공하는 카카오미니 C 스피커를 연구 대상으로 삼았다.

피험자로는 서울·경기 방언의 2~30대 남녀 피험자 총 20명을 모집하였다. 각 스피커 음성마다 정보 전달, 정서 표현의 두 가지 담화 상황을 상정하고 그에 따른 세부 주제들을 부여하였고, 피험자들이 이에 따라 AI 스피커와 자유롭게 대화하게 유도한 뒤 각 음성의 명료도, 자연성, 말투 선호도를 6점 척도로 평가하게 하였다.

그 결과 정보 전달, 정서 표현의 담화 상황 모두에서 스피커의 음성적 특성에 따른 말투 선호도에서의 차이가 가장 유의미했고, 공통으로 높임법에서 존댓말 음성이 더 선호되었다. 성별의 영향은 상대적으로 적었고, 여성 음성이 비교적 더 선호되었다. 자연성에서도 차이가 나타났으나 상대적으로 뚜렷하지 않았고, 명료도에서는 이렇다 할 차이가 발견되지 않았다.

한편, 피험자 성별에서도 남녀 공통으로 말투 선호도에서의 차이가 유의미했다. 단, 남성의 경우 높임법과 성별이 두루 영향을 미쳤으나 여성의 경우 높임법에서만 뚜렷한 차이가 발견되었다.

종합적으로 합성음의 명료도 및 자연성 등의 품질 평가보다는, 높임법이나 성별 등의 사회언어학적 요인에서 드러나는 말투의 차이가 인공지능 스피커 음성 각각에 대한 평가에 큰 영향을 미쳤다는 것을 확인할 수 있었다. 이러한 결과는 같거나 유사한 상황과 주제라 하더라도 AI 스피커가 가지는 성별, 높임법 등의 사회언어학적 특성에서 드러나는 말투의 차이가 사용자들의 평가에 큰 영향을 미침을 보여 준다.

참고문헌

[1] 남기춘·구민모·김종진·최양규. (2002). 합성음 평가 방법 고찰. 한국음향학회 2002년도 학술발표대회 논문집. 21(1). 297-300.

[2] 이익섭. (1994/2000). 사회언어학. 서울: 민음사.

[3] Cardoso. W.. Smith. G., & Garcia Fuentes. C. (2015). Evaluating text-to-speech synthesizers. proceedings of the 2015 EUROCALL Conference. 108-113.

[4] Reeves. B., & Nass. C. (1996). The media equation. Cambridge University press. New York.

[5] Nass. C., & Y. Moon. (2000). Machines and Mindlessness: Social Responses to Computers. Journal of Social Issues. 56(1). 81-103.

공용 미디어 자료를 이용한 발화 호감도에 대한 음성 산출 및 청지각 연구

이 현 정, 성 철 재
충남대학교 언어학과

A Study of Speech Production and Perception According to the Speech Likability about Public Media Materials

Hyunjeong Lee, Cheoljae Seong
Department of Linguistics, Chungnam National University
lhj811@gmail.com, cjseong49@gmail.com

사람 간 소통에서 목소리는 중요한 역할을 한다. 특히 말소리 운율(speech prosody)은 의사소통에서 의미와 감정 전달의 주요 변수를 만들어 낸다. 제4차 산업혁명의 도래로 인공지능(AI), 사물인터넷(IoT)등 새로운 기술이 등장하면서 다양한 콘텐츠와 플랫폼이 형성되고 있다. 음성인식 및 음성합성 서비스를 소비할 때 사람들은 자연스럽고 듣기 좋은 목소리를 선호할 것이라는 관점으로 연구에 착수하였다. 본 연구에서는 현 시대 다수의 사람들이 인정하는 호감 목소리를 찾고자 카카오 AI 서비스 관련 모바일 설문조사('음성 비서 목소리에 어울리는 연예인' 항목)에서 상위권으로 선정된 4명의 연예인 말소리를 활용하였다. 각 인물의 음성을 문장 단위로 처리한 후, 청지각 실험을 실시하였다. 내용에 대한 호감도와 말소리 운율에 대한 호감 정도를 발화자 성별 및 인물에 따라 판단하여 연구를 위해 고안된 35가지의 음향 변수로 통계를 진행하였다. 호감도의 분포를 살피고, 발화자 성별과 2지선다형 호감도에 따른 음향학적 특성을 알아내며, 호감도 분류정확도를 얻어낼 수 있었다.

독일어의 말음 무성음화에 따른 비완전 중화현상 재고찰

이 동 은, 이 석 재
연세대학교 독어독문학과, 영어영문학과

Incomplete Neutralization of Final Obstruents in German Revisited

Dong Eun Lee, Seok-Chae Rhee
Dept. of German (& English) Language and Literature, Yonsei University
ellielee@yonsei.ac.kr, scrhee@yonsei.ac.kr

본고는 독일어 음운규칙 중 잘 알려진 말음 무성화 규칙(Final Obstruent Devoicing(FOD))과 그로 말미암아 음성학적으로 관찰되는 중화 현상의 비완전성에 대한 연구이다. 독일어 말음은 무성음화를 거쳐 최소쌍의 경우 완전히 중화되어 동음 관계에 있다고 알려져 있지만, Port & O'Dell(1985), 그리고 Charles & Luce(1985) 등의 연구로부터는 무성음화에 따른 중화가 발생하더라도 기저 유성장애음과 기저 무성장애음에 선행하는 모음 길이 등 몇몇의 청각 단서들에 의해 기저 유성장애음과 기저 무성장애음을 포함한 최소 단어쌍들이 구분될 수 있음이 보고된 바 있다. 본고는 이러한 선행연구에 대한 검증 및 무성음화의 대상이 되는 장애음들의 음성적 환경의 차이에 따른 중화 관련 음성 차이 가능성을 알아보고자 한다. 분석 대상은 독일어 L1 화자 6명의 녹음물이고, 각 환경에 따른 유/무성 장애음 앞의 모음 길이 측정 및 비교에 초점을 두었다. 차이 값의 평균을 등분산 t-검정을 통해 검증한 결과 말음무성음화가 일어난 21개의 쌍 중 18개 쌍의 차이가 통계적으로 유의미한 것으로 나타났으며 ($p \leq 0.1$), 모음 길이가 비교적 짧은 비중을 갖는 3음절 어휘에서의 어중 음절말 장애음 간의 차이가 1음절과 유사하게 나타남에 따라 어중 음절말 위치에서 이들의 차이가 실재함을 유추할 수 있다. 실험의 결과를 종합하여 보았을 때, 장애음의 위치를 막론하고 기저형이 유성음인 장애음에 선행하는 모음의 길이가 무성음 기저형에 선행하는 모음의 길이보다 길다는 관찰이 이루어져, 독일어에서 소위 말음 무성음화에 따른 중화는 비완전하다는 가설을 뒷받침한다.

한국인의 L2 영어 고립어 발화에서 유창성 등급 대비 어말 폐쇄음 유무성성에 따른 선행 모음 길이 비율 연구

이 현 전, 박 우 지, 이 석 재
연세대학교 영어영문학과

Effects of Word Final Stops' Voicing on Vowel Duration and its Relation with L2 Proficiency

Hyunjun Lee, Wooji Park, Seok-Chae Rhee
Department of English Language and Literature, Yonsei University
hjne95@yonsei.ac.kr, woojipark@yonsei.ac.kr, schree@yonsei.ac.kr

본 연구는 한국어 화자의 영어 발화에서 어말 폐쇄음의 유무성성(voiced/voicelss)에 따른 선행 모음 길이와 영어 유창성(fluency)과의 관계를 밝히고자 한다. 영어에서 어말 폐쇄음이 유성음인 cab과 무성음인 cap의 경우에 cab의 모음 길이가 cap의 모음 길이보다 더 길게 발화 된다는 것은 잘 알려진 사실인 바, 이를 토대로 본 연구는 이러한 단어들에 있어 모음 길이 간 비율의 차이와 한국인 학습자의 영어 유창성 등급 간의 관계를 밝히는 것이 그 목적이며, 영어 유창성이 높을수록 어말 무성폐쇄음에 선행하는 모음의 길이 대비 유성폐쇄음에 선행하는 모음의 길이 비율(ratio: 유성음 앞 모음길이/무성음 앞 모음길이)이 더 큰 값을 보일 것이라는 가설을 검증한다. 분석 자료로는 K-SEC 음성코퍼스를 대상으로 유창성 평가 후 3개의 등급(Novice, Intermediate, Advanced)으로 구분된 32명의 서울 및 경기지역 초등학생의 고립형 단어 발화를 대상으로 하며, 각 환경에서 모음 길이 측정 후 비율값을 구하고, 이 비율값이 각 유창성 등급에 따라 어떠한 상관관계를 보이는지를 밝히게 된다. 분석 결과 연구 대상 단어들에 대하여 모든 등급에서 어말 유성 폐쇄음에 선행하는 모음이 어말 무성 폐쇄음에 선행하는 모음보다 더 길게 발화됐으며 그 비율값은 상-중-하 그룹 각각 1.2915:1, 1.1821:1, 1.3257:1로 나타났다. 그러나 본 연구의 핵심인 영어 유창성과 모음 비율 간의 상관관계는 찾지 못하였다. 즉, 한국인 영어 학습자들은 어말 유성폐쇄음에 선행하는 모음을 어말 무성폐쇄음에 선행하는 모음보다 더 길게 발화하는 경향을 보이나 이 두 모음 간의 비율은 영어 유창성과는 독립적이라는 것이 연구의 결론이다.

한국인의 L2 영어 [관사 - 1음절 명사]$_{NP}$ 에서 기능어 - 내용어 사이 모음 길이 비율과 영어 능숙도 사이의 상관 관계 연구

박 우 지, 모 란, 이 석 재
연세대학교 영어영문학과

The Correlation between the Vowel Duration in Articles and Monosyllabic Nouns within English NP and L2 Proficiency of Korean Learners

Wooji Park, Ran Mo, Seok-Chae Rhee
Department of English Language and Literature, Yonsei University
woojipark@yonsei.ac.kr, 0728mr@yonsei.ac.kr, schree@yonsei.ac.kr

본 연구는 기능어인 관사(a function word) 및 1음절의 내용어인 명사(a content word)로 구성된 명사구(NP) 내 두 단어의 모음 길이 비율과 영어 능숙도 사이의 상관관계를 밝히는데 목적이 있다. 일반적으로 영어 발화 시 내용어는 기능어 대비 상대적으로 더 두드러진(prominent) 소리를 가지게 된다. 즉, 강세를 받는 단어의 모음은 강세를 받지 않는 단어에서 보다 길게 발음되며, 더 큰 소리로 발화되며, 더 높은 소리를 가지게 된다. 이를 근거로 본 연구에서는 이들 중 모음의 길이 차이에 초점을 두고 영어 유창성에 따라 4개의 등급 (Novice, Intermediate-Low, Intermediate-High, Advanced)으로 구성된 한국인의 L2 영어 음성 코퍼스에서 [관사 - 1음절 명사]$_{NP}$가 포함 된 문장 발화 시 관사 대비 명사의 모음 길이 비율과 영어 능숙도 간에 어떠한 상관관계가 있는지 살펴보고자 한다. 등급화 된 한국인의 L2 영어 음성코퍼스(rated K-SEC)의 일부를 활용하여 진행하였다. 총 853의 음성 파일을 추출, Praat을 이용하여 문장 안에 들어 있는 명사구 내의 관사의 모음 길이와 명사의 모음 길이를 확인하고, 그 비율을 측정하였다. 분석 결과 관사의 모음 대 명사의 모음 비율은 능숙도와 양의 정적 상관관계를 갖는 것을 확인하였다 (Advanced (2.453) > Intermediate High (1.465) > Intermediate Low (1.174) > Novice (0.979), 괄호 안의 수치는 관사의 모음 길이 대비 명사의 모음길이 비율임). 또한 각 그룹 간, 내용어-기능어 간 모음 길이의 차이, 모음 비율의 차이가 모두 유의미한 것으로 나타났다 ($p<.001$). 즉, 발화자의 영어 유창성이 높아질수록, 영어 관사 내의 모음보다 명사 내의 모음이 더 길게 발화 된다.

문장 조건에서 3~6세 아동의 음운지표 발달

박나래[1], 김수진[2], 하지완[3], 신문자[4], 김영태[5]
[1]나사렛대학교 일반대학원 언어치료학과, [2]나사렛대학교 언어치료학과,
[3]대구대학교 언어치료학과, [4]조선대학교 언어치료학과, [5]이화여자대학교대학원
언어병리학과

Development of phonological whole-word measures in sentence condition in children at 3-6 years of age

Narae Park[1], Soo-Jin Kim[2], Ji-Wan Ha[3], Moon Ja Shin[4], Young Tae Kim
[1]Department of Communication Disorders, The Graduate School of Korea Nazarene University, [2]Department of Communication Disorders, Korea Nazarene University, [3]Department of Speech Pathology, Daegu University, [4]Department of Speech Language Pathology, Chosun University, [5]Department of Communication Disorders, Ewha Womans University
[1]201816002@slp.education, [2]skim@gmail.com, [3]jw-ha@daegu.ac.kr, [4]moonjashin@chosun.ac.kr, [5]youngtae@ewha.ac.kr

1. 서론

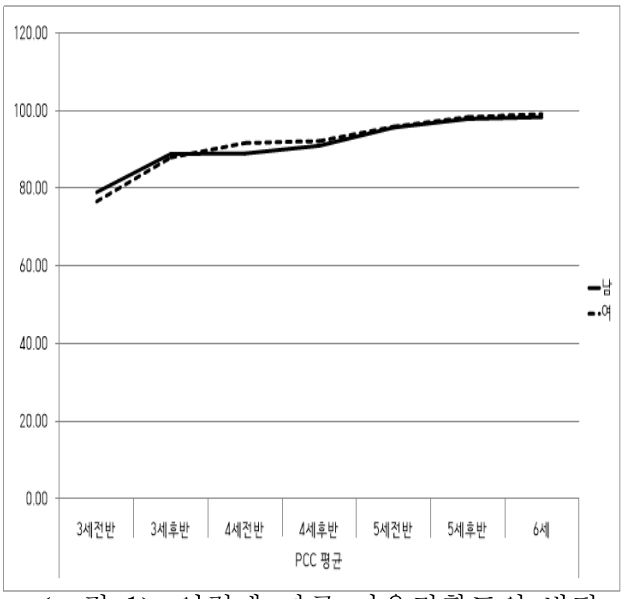
말소리장애의 중증도를 평가하는데 널리 사용되는 자음정확도(PCC)는 단어보다 자발화 혹은 유도문장에서 평가하는 것이 꼭 필요하다([1]). 그러나 임상장면에서는 표준화 검사들이 모두 단어수준에서만 규준이 제공되기 때문에 단어수준의 자음정확도만으로 중증도의 기준으로 사용되고 있다. 이에 본 연구에서는 3세부터 6세까지 일반아동을 대상으로 자발화와 보다 유사한 평가 문맥이라고 할 수 있는 따라말하기를 통한 유도문장수준에서 자음정확도와 단어단위 음운지표인 평균음운길이(PMLU)의 발달양상을 분석하여 임상장면에서 활용할 수 있는 준거 자료를 제공하고자 한다.

2. 연구방법

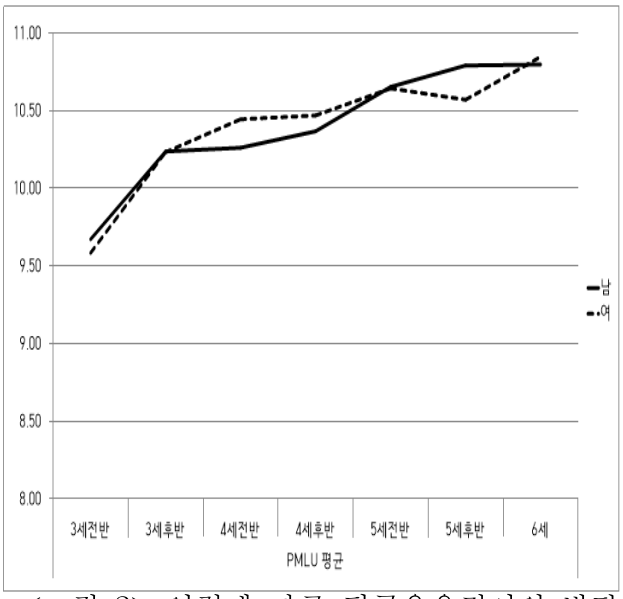
서울, 경기, 충청, 경상, 전라, 제주 지역에 거주하고 있는 3~6세까지 일반 아동 468명을 대상으로 어휘발달검사(REVT)와 U-TAP2의 문장수준검사를 실시하고, 전사한 후 평균음운길이(PMLU)와 개정자음정확도(PCC-R)를 구하고, 연령과 성별에 따른 발달양상을 분석하였다.

3. 연구결과

연령이 증가할수록 자음정확도와 평균음운길이가 증가하였으며 그 양상은 <그림 1, 2>와 같다.



<그림 1> 연령에 따른 자음정확도의 발달



<그림 2> 연령에 따른 평균음운길이의 발달

4. 결론 및 논의

모든 지표는 연령에 따라 증가하였으나 같은 대상자들에게 구한 이전의 단어수준 연구결과([2])에 비해서 낮은 수준으로 나타났다. 단어수준에서 자음정확도는 4세 후반에 천정점수에 달한 반면, 문장수준에서는 4세후반에 91.5%, 5세전반 95.8%로 5세 후반이 되어야 천정점수에 달할 수 있었다. 본 연구결과는 전국규모로 수집된 자료로, 자발화의 발달 양상과 유사한 문장 수준에서 한국어 말소리 습득의 전반적인 양상을 확인할 수 있는 동시에 임상장면에서 지역에 관계없이, 자발화와 유사한 문장 조건의 진단준거로 활용될 수 있을 것으로 기대된다.

5. 참고문헌

[1] 윤미선, 김정미, 김수진(2013). 자발화 문맥에서의 단어단위 음운 평가 2~4세 아동. *언어치료연구, 22*(4), 69-85.

[2] 하지완, 김수진, 김영태, 신문자(2019). 자음정확도와 단어단위 음운지표를 이용한 일반아동의 말소리산출 능력에 대한 발달 연구. *Communication Sciences & Disorders, 24*, 출판예정.

학령전 아동의 UTAP2 문장과제에 나타난 음운오류패턴

최 영 빈[1], 김 수 진[2]
나사렛대학교 일반대학원 언어치료학과, 나사렛대학교 언어치료학과

Phonological Error Patterns of UTAP2 Sentence Context Produced by Preschool Children

Youngbin Choi[1], Soojin Kim[2]
Department of Communication Disorders, The Graduate School of Korea Nazarene University[1]
Department of Communication Disorders, Korea Nazarene University[2]
201816007@slp.education, sjkim@kornu.ac.kr

1. 서론

말소리장애의 정확한 진단과 평가를 위해서 다양한 문맥에서의 평가가 필요하다. 특히 음운오류패턴 분석은 발달적인 전형적 오류패턴과 비전형적 오류패턴으로 나누어 분석하는 과정을 통해 장애의 정도를 나누기도 하지만 그보다는 다른 유형으로 집단을 나누고 그에 따른 중재 전략도 다르게 세워야 한다(하승희, 2015). 본 연구는 일반아동을 대상으로 대화와 유사한 문맥이라고 할 수 있는 표준화 검사의 유도 문장 문맥에 나타난 발달적 오류패턴을 알아보고자 한다. 이 분석 결과는 말소리장애 아동의 진단 장면에서 나타나는 오류패턴이 발달적이고 전형적인 것인지 판단하는데 활용될 수 있을 것이다.

2. 연구방법

만 3세부터 만 5세까지의 아동들을 대상으로 각 연령을 6개월로 나누어 6집단으로 구성하였다. UTAP2의 문장수준 검사를 수행하면서 녹음한 아동의 발화를 전사하고 김수진(2014)의 발달적 음운변동 틀에 맞추어 분석하고 그에 맞지 않는 오류는 기타변동으로 분류하였다. 각 오류패턴별로 2회 이상 오류를 보인 아동의 비율을 확인하였다.

3. 연구결과

음절구조변동, 대치변동과 기타변동으로 나누어 보았으며 요약한 결과는 그림 1과 같다. 각 변동에 대하여 연령별로 전체 아동 중 2회 이상 해당 오류패턴을 보인 아동의

수가 3% 이상인 경우는 가장 흐린 음영으로, 5% 이상인 경우는 중간 음영으로, 10% 이상인 경우는 진한 음영으로 표현하였다.

그림 1. 3~5세 아동이 문장과제에서 보인 음운오류패턴

4. 결론 및 논의

전국적인 규모에서 문장과제로 수집한 발화자료를 통해 우리말소리 습득과정에서 보이는 오류패턴을 살펴볼 수 있었다. 특히 일반아동의 어중종성생략, 어말종성생략, 유음생략, 유음활음화, 마찰음의 파열음화와 파찰음화, 파찰음의 파열음화, 연구개음의 전방화, 어중종성역행동화 등의 오류패턴이 4세부터 5세까지 지속적으로 나타나는 것을 확인할 수 있었다. 이 결과는 임상현장에서 문장 따라말하기 과제를 통하여 말소리 오류패턴을 확인하는 데 활용될 수 있을 것이다.

5. 참고문헌

[1] 김민정, 배소영, 박창일. (2007). 아동용발음평가(APAC). 서울: 휴브알앤씨.
[2] 김수진, 신지영. (2015). 『말소리장애』. 서울: 시그마프레스.
[3] 김수진. (2014). 자발화에 나타나는 발달적 음운오류패턴. *Communication Sciences and Disorders, 19*(3), 361-370.
[4] 김영태, 박희영, 강진경, 김정아, 신문자, 김수진, 하지완. (2019). 우리말 조음·음운평가-2 (U-TAP2) 개발을 위한 타당도 및 신뢰도 분석.
[5] 김영태, 신문자. (2004). 우리말 조음-음운평가(U-TAP) 수정·보완판, 서울: 학지사.
[6] 석동일, 박상희, 신혜정, 박희정. (2008). 한국어 표준 그림 조음음운검사(KS-PAPT). 서울: 학지사.

과소비성 분류를 위한 음향음성학적 접근
- 3가지 머신러닝 기법의 비교 -

김 난 숙, 성 철 재[1)]
충남대학교 대학원 언어병리학과

Classification of Hyponasality using 3 kinds of Machine Learnning Techniques

Nansook Kim, Cheoljae Seong
Department of Speech & Language Pathology,
Chungnam National University
kidsmirae@gmail.com, cjseong49@gmail.com

구·비강 통로의 생리·해부학적 문제로 인하여 과소비성을 보이는 질환들 중에 아데노이드 편도비대(ATH: AdenoTonsillar Hypertrophy)와 앨러지성 비염(AR: Allergic Rhinitis)이 있다. ATH와 AR 두 집단의 나조미터로 측정한 비성도(nasalance)는 t 검정 결과 유의한 차이를 보이지 않았다(p=.531).

22050Hz 나이퀴스트 주파수 도메인에서 5060Hz 이하 영역 에너지 최소값(normalized)과 5060-22050Hz 주파수 영역 에너지 평균값(normalized), 켑스트럼 객체의 RNR (rhamonics to energy ratio), 그리고 LH-ratio(low_energy(0-4000) to high_energy(4000-8000))는 두 집단을 유의하게 구분하였다($p<.05$). 5060Hz는 22050Hz 주파수 영역을 16 단계 멜 스케일 로그 분할하였을 때 11번째 단계에 해당한다.

유의미한 네 가지 변수들을 기계학습(machine learning) 기법 중 트리(tree)의 배깅(bagging) 알고리듬을 가장 적극적으로 사용하는 Random Forest[1]를 이용하여 상대적 중요도를 Gini 계수로 평가한 뒤 변수 개수를 3개로 줄였다[2)]. 원 데이터를 7:3으로 무선 분류한 뒤 70% 데이터를 훈련데이터로 삼아 서포트 벡터 머신, 랜덤 포레스트, 그리고 로지스틱 회귀분석을 이용하여 평가 데이터(test data)의 분류정확도를 살펴보았다[2]. 그 결과, 30% 평가 데이터의 분류정확도가 가장 높게, 가장 안정적으로 나타난 것은 SVM (Support Vector Machine)이었다(68.7%). 그림 1과 2는 이와 같은 작업을 9회 반복한 결과다.

1) 교신저자
2) 과적합(overfitting) 방지를 위하여.

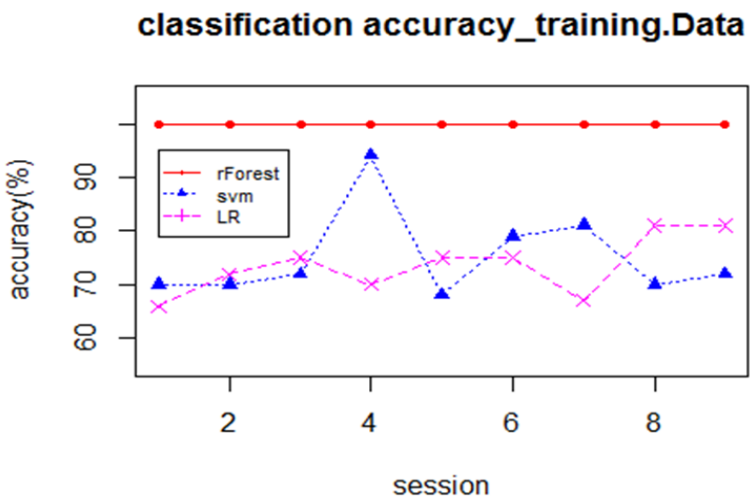


Figure 1. Classification accuracy of training data

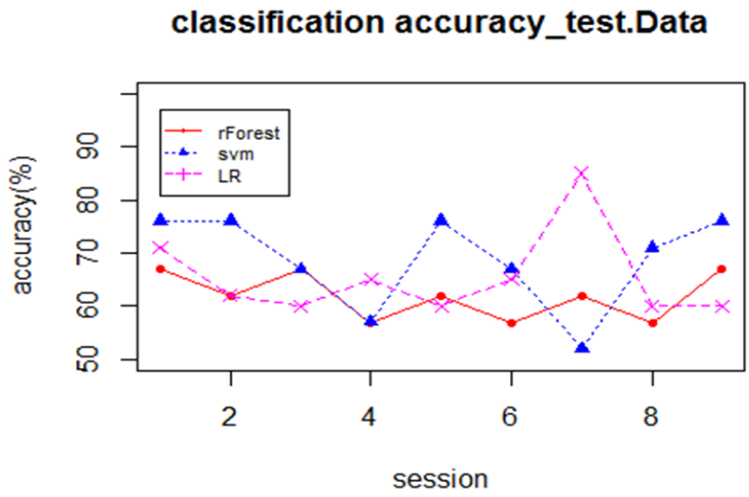


Figure 2. Classification accuracy of test data

참고문헌

[1] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.

[2] Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2), 415-425.

연결발화에서 근긴장성발성장애 음성의 켑스트럼 분석

윤 주 원* · 심 희 정** · 성 철 재*3)
*충남대학교 대학원 언어병리학과
**한림대학교 언어병리학과

Cepstral Analysis on the Connected Speech in Young Women With Muscle Tension Dysphonia

Joowon Yun*, Heejeong Shim**, Cheoljae Seong*
*Department of Speech & Language Pathology, Chungnam National University
**Division of Speech Pathology & Audiology, Hallym University
voiceslp@naver.com, amygmlwjd@gmail.com, cjseong49@gmail.com

근육 긴장성 발성 장애(MTD)환자의 모음 발성과 연결발화 과제에 대한 켑스트럼 분석결과를 비교하였다. 또한 음성장애 환자의 청지각적 특성과 음향음성학적 특성의 상관성을 살펴보고, 장애음성을 감별하는데 가장 유용한 측정치를 확인하고자 하였다. 내원 시 MTD로 진단받은 여성 36명과 정상음성을 사용하는 여성 36명이 연구에 참여했으며, 수집한 음성샘플은 ADSV™를 사용하여 분석하였다. 연구 결과, MTD의 CPP는 대조군보다 유의하게 낮았으며, MTD의 CSID는 대조군보다 높았다. 음향학적 측정치들은 모음 과제에서 대조군의 CPP, CPP Fo 값이 통계적으로 유의하게 컸으며, L/H ratio는 MTD 환자에게서 큰 양상을 보였다. 읽기 과제에서도 대조군의 CPP, CPP F0 값이 큰 반면, CSID는 MTD 환자에게서 높게 나타났다. MTD 환자의 음질 특성은 전반적인 음성중증도 (G), 조조성 (R), 기식성 (B), 노력성 (S)이 높아질수록 CPP와 L/H ratio는 감소하는 부적 상관을 보이고, CSID는 증가하는 정적 상관이 관찰되었다. 또한, 모음과 문장과제의 음향학적 분석 결과와 청지각적 평가 결과가 r=.404~.637까지 중간 정도 이상의 상관성을 보였다. 이를 통하여 연결발화에서 켑스트럼 측정치가 MTD 환자의 감별 진단에 대한 임상 활용 가능성과 CSID가 MTD 환자의 음성중증도와 중재효과를 객관적으로 살펴볼 수 있는 측정치임을 확인할 수 있었다.

참고문헌

[1] Seo, I. H., & Seong, C. J. (2013). Voice quality of dysarthric speakers in connected speech. *Journal of the Korean society of speech science*, 5(4), 33-41. (서인효·성철재 (2013).

3) 교신저자

연결발화에서 마비말화자의 음질 특성. *말소리와 음성과학*, 5(4), 33-41.)
[2] Peterson, E. A., Roy, N., Awan, S. N., Merrill, R. M., Banks, R., & Tanner, K. (2013). Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. *Journal of Voice*, 27(4), 401-410.
[3] Awan, S. N., Roy, N., Zhang, D., & Cohen, S. M. (2016). Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *Journal of Voice,* 30(2), 130-144.

과제 유형에 따른 음성장애 화자의 청지각적 및 음향학적 상관성

박 상 희, 이 옥 분
대구사이버대학교 언어치료학과

Auditory and Acoustic Correlation of Voice Disorders According to Task Type

Sanghee Park, Okbun Lee
Speech Language and Pathology, Daegu Cyber University
psh4292@dcu.ac.kr, oblee@dcu.ac.kr

1. 연구의 의의

기존의 임상현장에서는 단모음 /아/와 읽기 과제를 통해 음성장애 심한정도에 대한 평가를 실시하고 있다. 연구적인 측면에서는 모음에 대한 /아/, /이/, /우/ 3개의 모음을 선택하여 실시하고 있지만 비공식검사에서도 /아/의 연장발성을 보는 것이 보편적이다. /이/. /아/ /우/와 같이 3개의 모음을 선택하는 이유는 모음삼각도에서 극점에 있는 모음이기 때문이다. 모음이 달라지므로 성도의 길이, 구강의 모양, 혀의 위치가 다르다는 것을 의미하고 있는 공명의 차이를 가져오게 된다. 기능적 혹은 기질적인 음성문제가 있는 화자들의 목소리도 모음이 유형에 영향이 있을 수 있고, 음성 문제의 심한정도를 지각하는 데에도 영향을 미칠 수가 있다고 본다. 또한 음성문제에 대한 음향학적인 평가 뿐만 아니라 청지각적인 판단은 발성 및 발화과제에 따라서 차이가 있을 수 있고, 기류역학적인 평가 자료까지 포함된 다차원적인 결과를 도출할 필요가 있다.

2. 연구 방법

연구의 대상은 음성장애 42명으로 모음 /이/, /아/, /우/와 산책 문단산출하게 하고 음향학적 분석결과와 청지각적 평가를 실시하였다. 음향학적인 분석은 Jitter, shimmer, NHR을 실시하였고, 기류역학적인 평가는 성문하압을 측정하였다. 청지각적인 평가는 전문가의 GRBAS 분석을 실시하였으며, 각 검사 결과에 대해서는 상관분석을 실시하였다.

3. 연구 결과 및 논의

본 연구 결과 음향학적 변수와 청지각적 평가 결과, 기류역학적인 평가에 대해서 상

관이 있는 것으로 나타났다. 청지각적인 평가결과가 Jitter, shimmer, NHR, 성문하압과 상관이 높은 것으로 나타났다. 특히 성문하압의 경우는 청지각적인 평가와 상관도가 높은 것으로 나타났다. 현장에서 기류역학적인 평가기기는 보편적으로 가지고 있는 기기는 아니므로 훈련된 전문가들이 기류역학적인 평가에 대한 부분을 보완할 수 있고, 특히 GRABAS 평가의 경우는 평가요소에서 기류역학적인 측면이 반영되어 있기 때문에 이러한 결과가 도출되었다고 볼 수 있다.

참고문헌

[1] 이옥분, 권순복.(2012). 발화과제 유형에 따른 청지각적 평가와 기기적 평가의 상관성 연구. 언어치료연구. 21(2), 23-35.

[2] Maryn, Y., & Roy, N. (2012). Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. Journal of Socedade Brasileira Fonoaudiologia, 24(2), 107-112.

[3] Murry, T., Medrado, R., Hogikyan, N. D., & Aviv, J. E. (2004). The relationship between ratings of voice quality and quality of life measures. Journal of Voice. 18(2), 183-92.

Nasality Visualization System(NVS)를 이용한 청년 및 중년기 성인의 말소리문맥에 따른 비음치 연구

정 재 만*, 하 지 완**, 우 승 탁***, 박 영 빈***
*대구대학교 일반대학원 재활과학과, **대구대학교 언어치료학과
***경북IT융합산업기술원

A Study of nasalance scores according to phonetic contexts in young and middle-aged adults

Jae-Man Jung* Ji-Wan Ha**, Seong Tak Wo***, Young-Bin Park***
*Daegu Univ. Graduate School, **Daegu Univ.,
***Gyeongbuk Institute of IT Convergence Industry Technology
oenpece@naver.com, jw-ha@daegu.ac.kr, stwoo@gitc.or.kr,
ybpark@gitc.or.kr

중년기는 장년에서 노년기로 넘어가는 교량역할을 하며, 시청각 등 감각기능저하, 호르몬 불균형으로 인한 다양한 신체변화가 찾아온다. 그러나 국내 말소리장애 분야에서 중년을 대상으로 한 비음특성의 변화와 관련한 연구는 찾아보기 어렵다. 이에 본 연구에서는 국내에서 아직 소개된 바 없는 Nasality Visualization System (NVS)를 이용하여 청년층과 중년층의 비음특성을 다양한 말소리 환경에서 비교 분석하였다. 국내 비음관련 연구들은 대부분 Nasometer II (Flether, 2002)를 이용하였으나, 이는 제한된 환경 내에서만 측정이 가능하다는 단점이 있다. 반면 NVS는 상대적으로 환경에 제약을 받지 않고 사용방법 또한 용이하다. 모음조건, 인접말소리 조건, 단어/비단어 조건을 달리하여 청년기 정상성인 7명과 중년기 정상성인 7명의 비음치를 분석한 결과, 모든 조건에서 연령집단 간 유의한 차이는 관찰되지 않았으나 말소리 조건에 따른 차이는 유의하였다. 말소리 조건 간 차이를 구체적으로 살펴보면, 첫째 저모음문맥보다 고모음문맥에서 비음치가 유의하게 높았고, 둘째 모음+비음 문맥보다 비음+모음 문맥, 즉 모음 후행 문맥에서 모음의 비음치가 유의하게 높았고, 셋째 비단어 조건보다 단어 조건에서 비음치가 유의하게 높았다. 모음 종류뿐 아니라 음절 내 비음 위치와 단어여부에 따라 비음치가 유의하게 달라진다는 것을 확인하였다는 점에서 본 연구결과는 의의가 있다. 특히 비단어보다 단어에서 유의하게 높은 비음치를 보였다는 결과는 단어의 경우 내재되어 있는 운동프로그램로 인해 말소리 간 동시조음이 더욱 활발해지고, 그 결과 인접 말소리간 영향력이 더욱 커진다는 것을 시사한다.

This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Ministry of Education and Korea government (MSIP) (No. 2018 R1D1A1B07051072, No. NRF-2018R1A2B2001434) and the Daegu-Gyeongbuk Medical Innovation Foundation funded by the Gyeongsan-si.

Influence of the nasometer structure on the nasalance: evaluation of acoustic clinical experimental and lumped-element model analysis

[1]Seong Tak Woo, [1]Young-Bin Park
[1]Gyeongbuk Institue of IT Convergence Industry Technology,
stwoo@gitc.or.kr, ybpark@gitc.or.kr
[2]Seong Dae Na
[2]Dept. of Biomedical Eng., Kyungpook National Univeristy Hospital
bluepoison14@gmail.com
[3]Euisung Jung
[3]Deagu-Gyeongbuk Medical Innovation Foundation,
Medical Device Development Center,
esjung@dgmif.re.kr
[4]Da Hee Oh
[4]Dept. of Rehabilitation Science, Graduate School, Daegu University,
duhee03@daegu.ac.kr
[5]Ji-Wan Ha
[5]Dept. of Speech Pathology, Daegu University,
jw-ha@daegu.ac.kr

Due to the development of medical technology, various devices for language rehabilitation are being studied. In particular, device studies for speech disorders such as cleft palate, hypernasality and hyponasality are being studied. And, a nasometer are commonly used to diagnose these disorders. For conventional separation type nasometers, there is an effect of acoustic feedback between sounds of oral and nasal. Recently, a mask type nasometer has been developed for insensitivity to acoustic feedback, but it has not been popularized yet. In this paper, we analyzed the acoustic characteristics according to the existing nasometer structure. In particular, we evaluated the acoustic collection characteristics of the mask type structure through the lumped-element model with electro - mechanical equivalent circuit. Through the analysis, it was confirmed that the optimum area of the acoustic hole was obtained and the closed type mask structure could be designed. In addition, we were obtained clinical acoustic data from the 6 subjects and examined the significant differences in the structure of the separation type and mask type nasometer. As a results, we confirmed the significant difference of nasalance according to the acoustic collection structure of the nasometer.

노년층의 조음교대운동 속도와 빠른 씹기 속도 간의 최대수행력 비교

이 지 언[1], 이 현 정[1], 김 향 희[1,2]
[1] 연세대학교 대학원 언어병리학협동과정
[2] 연세대학교 대학원 언어병리학협동과정, 의과대학 재활의학교실

Performance of Fast Solid Chewing Test and DDK in the Elderly

Jieon Lee, Hyunjung Lee, Hyanghee Kim
Graduate Program in Speech-Language Pathology, Yonsei University
happiness10@yonsei.ac.kr, neuroslp@yonsei.ac.kr, h.kim@yonsei.ac.kr

= 초록 =

연구목적 말(speech) 산출과 씹기(chewing) 기능은 유사한 구강구조를 공유한다. 따라서 이들에 대한 최대수행력(maximum performance) 과제는 입술, 혀, 턱 등을 포함한 구강 안면 근육 운동 속도 및 규칙성에 관해 유용한 정보를 제공할 수 있다. 이에, 본 연구에서는 정상 노인의 조음교대운동(DDK) 속도와 빠른 고체 씹기 속도(fSCT) 사이의 최대 수행력 간의 상관관계를 살펴 보고자 하였다.

연구방법 정상 노인 189명(남:여=96:93, 평균 연령(±표준편차): 76.74세(±5.7), 범위: 66~95세)을 대상으로 각 개인에게 DDK 과제 및 TOMASS(The Test of Masticating and Swallowing Solids)(Huckabee et al, 2018) 과제를 모두 실시하였다. 조음교대운동속도(AMR과 SMR, 초당 최대 음절반복횟수)와 빠른 씹기 속도(초당 최대저작횟수)를 측정한 뒤 피어슨 상관분석을 실시하였다.

결과 AMR '퍼' 과제의 평균 속도는 4.29회/초, SMR 과제의 평균 속도는 1.62회/초, fSCT 평균 속도는 1.64회/초였다. AMR '퍼' 과제의 속도가 빠를 수록(4.29회/초) fSCT 속도가 빨랐고(1.64회/초)($p<.05$), SMR 과제 수행 속도가 빠를 수록(1.62회/초) fSCT 속도가 빨랐다($p<.01$).

논의 및 결론 본 연구의 결과로 객관적인 측정을 통하여 조음교대운동 속도(DDK)와 노년층의 빠른 씹기 속도(fSCT rate) 사이에 유의한 상관이 있음을 확인할 수 있었다. 이는 조음 산출과 저작 과정에서 사용되는 각각의 구강운동구조들에서 기능 간에 상관이 있음을 의미한다. 이를 통해 말 산출 수행력을 통해 씹기 기능을 유추할 수 있을 것으로 기대하며, 추후 삼킴기능 평가 및 재활과 관련된 후속 연구에 기초자료를 제공할 것으로 기대한다.

The Effect of Unvoiced Segments on the Cepstral Analysis in Age-related Dysphonia

Geun-Hyo Kim, PhD1, Yeon-Yoo Lee, PhD1, In-Ho Bae, PhD2
Hee-June Park, PhD3, Chang-Yoon Lee, MS4, Soon-Bok Kwon, PhD5
1Department of Otorhinolaryngology-Head and Neck Surgery and Biomedical Research Institute, Pusan National University Hospital, Busan, South Korea;2Department of Otorhinolaryngology-Head and Neck Surgery, Pusan National University Yangsan Hospital, Yangsan, Gyeongsangnam-do, South Korea; 3Department of Speech and Hearing Therapy, Catholic University of Pusan, Busan, Korea; 4Department of Otorhinolaryngology, Dongnam Institute of Radiological & Medical Sciences, Busan, South Korea; 5Humanities, Language and Information, Pusan National University, Busan, South Korea

kimgeunhhyo@gmail.com, slpywlee@gmail.com, voicebae@gmail.com, yuni765@naver.com, sbkwon@pusan.ac.kr

Background and Objectives: The purpose of this study is to investigate the effect of unvoiced segments on cepstral analysis in patients with presbyphonia. Materials and Methods: A total of 381 participants(263 participants with presbyphonia and 118 with normal voices participated in this study. The sustained vowel /a/ 2 seconds and one sentence from 'San-Chaek' were edited and analyzed using the Praat script. Auditory-perceptual rating was also performed by three raters. Results: All parameters except LHRatio_V and extracted continuous speech (EXT) showed a significant difference between the normal and the dysphonia groups. The normal group showed higher cepstral peak prominence (CPP) and smoothed CPP (CPPS) than the dysphonia group. Additionally, AP ratings of the normal group were lower than that of the dysphonia group. The correlation within the measured values ranged from -0.86 to -0.66. In continuous speech (CS), there was a significant difference of area under the curve (AUC) between CPP and CPPS, but there was no significant difference in EXT. Conclusion: Through this study, we confirmed the effect of unvoiced segments on cepstral analysis. By removing the unvoiced segments, the measured values of the cepstral analysis showed an increase. Through further studies, we hope to increase the accuracy of discriminating dysphonia by extracting the unvoiced segments.

참고문헌

[1] Martins RHG, Pessin ABB, Nassib DJ, Branco A, Rodrigues SA, Matheus SMM. (2015). Aging voice and the laryngeal muscle atrophy. *Laryngoscope. 125*(11): p. 2518-21.

[2] Jin SM. Voice Evaluation. (2015). *J Clinical Otolaryngol. 26*(1): p. 14-22.

[3] Ravi SK, Shabnam S, George KS, Saraswathi T. (2018). Acoustic and Aerodynamic Characteristics of Choral Singers. *J Voice.* In Press.

[4] Dejonckere PH, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier-Buchman L, Millet B. (1996). Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. Rev *Laryngol Otol Rhinol (Bord). 117*(3): p. 219-24.

[5] Hillenbrand J, Houde RA. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Lang Hear Res. 39*(2): p. 311-21.

다채널 마이크로폰 배열에서의 DNN 기반 빔형성에 관한 연구

김 태 우, 김 남 균, 이 건 우, 김 홍 국
광주과학기술원 전기전자컴퓨터공학부

A Study on DNN-Based Beamforming for Multi-Channel Microphone Array

Tae Woo Kim, Nam Kyun Kim, Geon Woo Lee, Hong Kook Kim
School of EECS, Gwangju Institute of Science and Technology (GIST)
{kimtaewoo, skarbs001, geonwoo0801, hongkook}@gist.ac.kr

근래에 다채널 마이크로폰을 활용한 원거리 음성인식, 음원 위치추정, 잔향 제거 등 많은 연구가 진행되고 있다. 이를 위한 전처리 기법의 하나인 빔형성은 다채널 마이크로폰 배열에 있어서 음성 또는 오디오 신호가 발생하는 특정 방향에 집중함으로써 그 음질을 향상시킨다. 그러나, 종래의 빔형성 기법은 설계된 특정 배열 구조에 종속적이며, 따라서 임의의 마이크로폰 배열 구조에 적용될 경우 그 성능이 저하되기 때문에 마이크로폰 배열 구조에 맞게 다시 설계해야하는 단점이 있다.

본 연구에서는 마이크로폰 배열 구조에 상관없는 심층 신경망 기반의 minimum variance distortionless response (MVDR) 빔형성 가중치 추정 기법을 제안한다. 먼저, 훈련 데이터 생성을 위해 특정 마이크로폰 배열 구조의 room impulse response (RIR)와 합성곱을 한다. 이 때, RIR이 적용된 모든 채널에 무작위로 시간 지연을 준다. 이를 통해, 다양한 마이크로폰 배열 구조의 데이터를 얻는 효과를 얻을 수 있다. 또한, REVERB Challenge에서 제공하는 훈련 잡음을 더해 훈련 데이터를 증강한다. 심층 신경망 모델의 입·출력은 각각 생성된 훈련 데이터의 음성이 존재하는 구간의 프레임당 채널 간 generalized cross correlation-phase transform (GCC-PHAT)과 MVDR 빔형성 가중치이다. 심층 신경망 모델의 입력층은 234개의 노드, 출력층은 2,056개의 노드로 구성된다. 또한, 각각 2,700개의 노드로 구성된 4개의 은닉층과 rectified linear unit (ReLU) 활성화 함수로 구성되어 있다.

성능 평가를 위해 REVERB Challenge 코퍼스의 시뮬레이션 데이터를 사용하였으며, 학습된 배열 구조와 다른 배열 구조의 테스트 데이터를 사용하여 기존의 MVDR 빔형성 기법과 제안된 기법을 적용하였다. Cepstral distance (CD) 비교 결과, 제안된 기법이 기존의 MVDR 빔형성 기법에 비해 0.71dB 더 나은 성능을 보였다.

감사의 글

본 연구는 산업통상자원부의 산업기술혁신사업으로부터 지원을 받아 수행된 연구임(No. 10063424, '실내용 음성대화 로봇을 위한 원거리 음성인식 기술 및 멀티 태스크 대화처리 기술 개발').

ITD 와 분산도를 활용한 가우시안 혼합 모델 기반 음원 국지화 방법

김 정 민[1], 김 보 현[1], 김 이 경[2], 박 강 호[2], 박 형 민[1]
[1]서강대학교 전자공학과, [2]ETRI 융복합센서연구그룹

Sound Source Localization Based on Gaussian Mixture Models Using Interaural Time Difference and Diffuseness

Jeongmin Kim[1], Bohyun Kim[1], Yi-Gyeong Kim[2], Kang-Ho Park[2], Hyung-Min Park[1]
[1]Department of Electronic Engineering, Sogang University
[2]Convergence Sensor Research Group, ETRI
kjmin0228@gmail.com, i_want_to@naver.com, kimyig@etri.re.kr, pkh@etri.re.kr, hpark@sogang.ac.kr

음원 국지화 알고리즘은 음원의 각도와 거리를 추정하는 알고리즘이다. 이를 위해 다수의 마이크로폰으로 들어온 음원에서 각도와 거리에 대해 의존도를 갖는 특징들을 추출한 뒤 각 각도와 거리에 대한 가우시안 혼합 모델(Gaussian mixture model, GMM)을 만들고, 확률 밀도를 통해 음원의 위치를 추정할 수 있다. 특정 위치에서 발화된 음원은 마이크로폰의 위치에 의해 위상차를 갖으며 이러한 위상차는 음원의 발화 위치와 마이크 쌍에 대해 달라진다. 따라서 마이크 쌍에 대한 샘플 지연 시간을 음원 국지화를 위한 특징으로 사용할 수 있다. 또한, 음원이 발화될 때 마이크로폰으로 직접 입력되는 신호 외에도 반향으로 생성되는 분산 잡음이 발생하므로 분산도를 특징으로 사용할 수 있다. 가우시안 혼합 모델 학습을 위해 마이크로폰 배열의 중심을 기준으로 일정한 간격의 각도, 거리에서 음원이 발화되도록 하였다. 마이크로폰 배열로 입력된 신호에서 앞서 설명한 두 특징을 추출한 뒤 각각의 위치에 대한 가우시안 혼합 모델을 만든다. 학습된 가우시안 혼합 모델을 이용해서 임의의 위치에서 발화된 신호의 특징 두 가지의 각 모델에 대한 likelihood를 구할 수 있다. 가장 높은 likelihood값을 갖는 가우시안 혼합 모델의 각도, 거리를 음원의 위치로 추정할 수 있다.

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00050, 신체기능의 이상이나 저하를 극복하기 위한 휴먼 청각 및 근력 증강 원천 기술 개발)

자기부호화기 특징을 이용한 딥러닝 기반 비침입적 음성명료도 추정 방법

김윤희*, 윤덕규**, 최승호*, 김선교***
* 서울과학기술대학교 전자IT미디어공학과
** 서울과학기술대학교 전자공학과
*** 국방과학연구소

A Non-intrusive Speech Quality Estimation Method Based on Deep Learning Using Autoencoder Features

Yunhee Kim, Deokgyu Yun, Seung Ho Choi, Seon-Kyo Kim
Seoul National University of Science and Technology
shchoi@seoultech.ac.kr

본 논문은 기준(reference) 음성신호가 없는 비침입적(non-intrusive) 음성명료도 추정에 관한 것으로서 자기부호화기(autoencoder)의 특징(feature)을 이용한 딥러닝 기반 방법을 제안한다. 기존의 표준 비침입적 음성명료도 추정 방법인 P.563은 다양한 잡음 환경에서의 명료도 추정성능이 미흡하다. 제안한 방법은 기준 음성신호가 있는 (intrusive) 표준 음성명료도 측정 도구인 STOI (short-time objective intelligibility) 점수를 활용하며, 음성신호의 프레임별 자기부호화기 특징벡터와 STOI 점수를 심층신경망(deep neural network)의 입력과 출력으로 하여 신경망을 훈련시킨다. 다양한 잡음 환경에서의 음성신호에 대한 기존의 P.563과 MFCC (mel frequency cepstrum coefficient) 기반 방법들과의 비교 실험을 통해 제안한 방법이 우수한 성능을 보임을 확인했다.

Acknowledgement

이 연구는 방위사업청 및 국방과학연구소의 재원에 의해 설립된 신호정보 특화연구센터 사업의 지원을 받아 수행되었음.

순차적 공분산 추정을 이용한 마스크 기반 온라인 빔포밍 구현

윤 성 욱, 권 오 욱
충북대학교

Implementation of Mask-Based Online Beamforming Using Sequential Covariance Estimation

Sung-Wook Yoon, Oh-Wook Kwon
Chungbuk National University
swyoon@cbnu.ac.kr, owkwon@cbnu.ac.kr

본 연구는 원거리 음성인식 CHiME3 Challenge에서 순차적 공분산 추정을 이용한 BLSTM 기반 마스크 추정 온라인 빔포밍의 성능을 측정한다. 일반적인 온라인 빔포밍 알고리듬에서는 음성이 존재하지 않는 환경에서 추정된 음성신호 공분산 행렬이 왜곡된 빔포밍 조향벡터를 추정함으로써 성능 열화를 보인다. 반면에 제안된 알고리듬은 학습된 공분산 모델을 순차적으로 적응함으로써 안정적인 성능을 보인다. 음성과 잡음 각각에 대한 공분산 모델을 학습한 후 batch block 단위에서 frame 단위로 추정되는 음성과 잡음의 BLSTM 기반 마스크 사후확률(posterior) 값을 추정한다. 추정된 사후확률 값을 이용하여 빔포밍 조향벡터 계산에 필요한 공분산 벡터를 순차적으로 추정한다. 제안된 온라인 빔포밍 시스템은 시시각각 변화하는 잡음 및 간섭 환경에서도 시간이 지남에 따라 안정적인 성능을 유지한다. 본 논문에서는 학습된 공분산 행렬을 적용하지 않는 기존 온라인 빔포밍 시스템과 제안한 온라인 빔포밍 시스템의 WER을 비교하고, batch block 단위에 따른 성능을 분석한다.

Conditional GAN 기반 가창 합성 시스템

최순범, 남주한
한국과학기술원 문화기술대학원

Singing Voice Synthesis based on Conditional GAN

Soonbeom Choi, Juhan Nam
Graduate School of Culture Technology, KAIST
cjb3549@kaist.ac.kr, juhannam@kaist.ac.kr

본 논문에서는 전통적인 Concatenative Synthesis의 한계를 극복하고자 Neural Network 기반의 가창 합성 기술 제안한다. Concatenative Synthesis로 가창 합성 시에 호흡 소리, 음간 연결 음고 변화 합성에서 자연스럽지 못한 한계점이 있다. 이 한계점을 극복하고자 데이터를 기반으로 가수의 특성을 고려하여 합성을 할 수 있는 모델을 연구하였다. 이미지 변환 기술에서 영감을 받아 Pix2Pix GAN을 기반으로 가창 합성 시스템을 개발하였다 [1]. 음성 합성과 가창 합성의 차이를 분석하여, Recurrent Neural Network 기반이 아닌 Frame 단위로 합성하는 Convolutional Neural Network 기반의 모델을 선택하였다. 생성 결과는 Spectrogram을 사용하였고, 오디오로 변환시키기 위해 음성 합성에 사용되는 Tacotron 모델을 참고하여 Griffin-Lim 알고리즘을 사용하였다 [2].

학습 데이터로는 자체적으로 녹음한 여성 가수 데이터를 사용하였다. 데이터는 통제된 상황에서 한국 동요를 위주로 녹음하였다. 모델 학습을 위해 녹음된 음원에 악보 가사 데이터를 추가하였다. 앞으로 해당 기술이 사용될 상황을 고려하여 음 변화, 주기성 등의 구체적인 정보를 배제하고, 악보와 가사만을 입력으로 사용하여 가창을 합성하였다.

GAN에는 Generator와 Discriminator 두 모델이 있어, 각 모델에 대하여 다른 기반 모델을 적용하였고, 실험을 통해 가창 합성에 최적화된 모델을 선택하였다. Generator의 경우 U-Net을 기반으로 한 모델이 가장 합성 음질이 높음을 확인하였다. 최종적으로 Concatenative Synthesis를 이용한 결과물과 제안한 합성 방법의 결과물을 비교 분석하였다. Conditional GAN 기반으로 합성된 음원은 Concatenative Synthesis에서 합성하기 어려운 음간 연결, 호흡 소리가 더 자연스럽게 합성됨을 확인하였다. 하지만 Phase가 연속적으로 생성되지 않는 점이 큰 한계점으로 확인되며, 후속 연구를 통해 이를 해결할 필요가 있다.

참고문헌 12p

[1] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

[2] Wang, Y., Skerry-Ryan, R. J., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., ... & Le, Q. (2017). Tacotron: Towards end-to-end speech synthesis. arXiv preprint arXiv:1703.10135.

음향적 특징과 Attention 기반 LSTM 모델을 활용한 한국어 방언 추정 연구

이 주 영[1], 김 경 화[2], 이 규 환[3], 정 민 화[1]
서울대학교 언어학과[1], 대검찰청[2], 서울대학교 인문정보연구실[3]

Korean Dialect Identification using Acoustic Features and Attention-based LSTM Model

Jooyoung Lee[1], Kyungwha Kim[2], Kyuwan Lee[3], Minhwa Chung[1]

Department of Linguistics, Seoul National University[1]
Korea Supreme Prosecutor's Office[2]
Human Information Center, Seoul National University[3]

excalibur12@snu.ac.kr[1], savoix@spo.go.kr[2],
whannylee@snu.ac.kr[3], mchung@snu.ac.kr[1]

본 연구는 보이스피싱 범죄 수사의 원활한 진행을 위한 화자 프로파일링 소프트웨어 개발의 일환으로 한국어 방언을 자동으로 식별하는 모델을 개발하는 것을 목적으로 한다. 방언 식별 모델로는 음향적 특징을 입력으로 학습을 하는 Attention 기반 Long Short Term Memory 모델을 사용하였으며, ivector 기반 DNN을 베이스라인 모델로 성능을 비교하였다.

참고문헌 12p

[1] Shon, S., Ali, A., & Glass, J. (2018). Convolutional neural networks and language embeddings for end-to-end dialect recognition. arXiv preprint arXiv:1803.04567.

[2] Dey, S., Koshinaka, T., Motlicek, P., & Madikeri, S. (2018, April). DNN based speaker embedding using content information for text-dependent speaker verification. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5344-5348). IEEE.

구두 발표 I:
음성학

좌장: Jeffrey Holliday(고려대)

The effect of base token selection for stimuli manipulation on the perception of lenis and aspirated stops in Seoul Korean

Eunjin Oh
Ewha Womans University
ejoh@ewha.ac.kr

This study investigated whether the perceptual pattern of listeners varied according to base tokens selected for stimuli manipulation in categorizing the lenis and aspirated stops of Seoul Korean. This question was based on the experimental method used in a previous study which examined the effect of listener gender on the perception of the two stops (Oh et al., 2018). In the study, a word containing an aspirated stop was used as the base token; as such, the perception stimuli may have contained other stop cues that were inclined towards the aspirated stop. Therefore, in the present study, VOT and F0 were manipulated with words containing a lenis and an aspirated stop, respectively. Other stop cues, besides VOT and F0, were left unchanged in the base tokens. Twenty native and twenty non-native listeners of Seoul Korean participated in a forced-choice identification experiment. Results found that the numbers of aspirated responses were significantly higher when the base token contained the aspirated stop for both native and non-native listeners, indicating that both groups used minor cues other than VOT and F0 in the categorization of the two stops. For native listeners, the use of the VOT and F0 cues did not differ regardless of which stop was used as the base token. This confirms that the results of the previous study concerning the effect of gender on the use of the VOT cue are tenable. The use of the VOT and F0 cues by non-native listeners was not native-like, indicating that the systematic use of the VOT and F0 cues through manipulated stimuli might be a difficult task for non-native listeners and that minor cues may have been perceived more stably in those environments.

Effect of short-term Engilsh exposure on Korean children's English perception

Margarethe McDonald [1], Eon-Suk Ko [2]
[1] Dept Communication Science & Disorders, University of Wisconsin-Madison
[2] Dept English Language and Literature, Chosun University
mlmcdonald3@wisc.edu, eonsukko@chosun.ac.kr

Children and adults are known to accommodate their perception to that of exposed speech [1][3]. In adults, this is seen in the ability to process foreign accented speech quicker with less than a minute of exposure [2]. In addition, adults have the ability to do this in their second language [4]. If children can also accommodate to exposed speech in a second language, it may be a tool to improve L2 speech perception. Korean children between ages 6 and 9 with varying levels of English ability participated in an English minimal pair forced choice recognition task. Participants matched auditorily presented words with pictures from the minimal pair *lock* and *rock* read by a native speaker. In one condition, the participants were exposed to a different native English speaker reading l- and r-initial words prior to performing the recognition task. In the other condition, they were exposed to a Korean-accented native speaker reading l- and r- initial words without a strong differentiation between the two sounds. Results indicate that short-term Korean-accented English exposure was beneficial for the English speech perception in children with low English abilities and native English exposure was more beneficial for those with high English language abilities.

참고문헌

[1] Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. Cognition, 106(2), 707-729. doi:10.1016/j.cognition.2007.04.005

[2] Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647-3658.

[3] McQueen, J. M., Tyler, M. D., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development, 8*(4), 317-339.

[4] Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 75.

A cross-linguistic study of sound symbolism in personal names

Jihyo Kim, Hyunji Kim, Eon-Suk Ko
Dept. of English Language and Literature, Chosun University
jihyo.email@gmail.com, hyunji2776@gmail.com, eonsukko@chosun.ac.kr

Sound symbolism refers to the phenomenon whereby the sound of a word is directly associated with its meaning [1]. It was reported that many languages have a large grammatically defined word class in which sound symbolism is clear [2]. In this study, we investigate the sound symbolism in personal names with regard to the gender associated with the names.

We conducted perception tests using Praat where participants judged the gender of the auditorily presented personal names in Korean, Chinese, English, German, and Igbo. Fifty names from each gender were presented twice for each language. Participants rated the gender of each name on scale of 1-(surely male) to 7(surely female). The results suggest that listeners tend to have an intuition about the gender of a name with varying degrees of accuracy, which seems to depend on the familiarity of the language.

A segmental analysis of the names suggests that there is a sound-symbolic mapping between consonant and vowel properties and maleness/femaleness. Specifically, female names tend to adopt open syllables and certain segmental features more often than male names. We will be conducting a more detailed comparison and analyzing the link between the segmental composition and the perceived gender of names.

References

[1] Pitcher BJ, Mesoudi A, McElligott AG (2013) Sex-Biased Sound Symbolism in English-Language First Names. PLoS ONE 8(6): e64825. doi:10.1371/journal.pone.0064825

[2] Sidhu DM, Pexman PM (2015) What's in a Name? Sound Symbolism and Gender in First Names. PLoS ONE 10(5): e0126809. doi:10.1371/journal.pone.0126809

Articulatory and acoustic characteristics of the Korean and English word-final laterals produced by Korean learners of English

Young Hwang
Department of Linguistics, Indiana University
hwang24@indiana.edu

The English [ɫ], a dark variant of the English lateral, is known to have different articulatory and acoustic characteristics from the Korean [l], which is identified as a light lateral, exhibiting more posterior tongue dorsum positions and low F2 values. Previous studies on the acquisition of English laterals by Korean learners, however, assert that the English [ɫ] is perceptually similar to the Korean [l], and thus it is difficult for learners to acquire a native-like production due to transfer of their articulatory strategy of the Korean [l] to the English [ɫ], without creating a new phonemic category. The aim of this study is to examine the articulatory and acoustic characteristics of the English word-final [ɫ] produced by advanced Korean learners of English, using novel 3D/4D ultrasound imaging methods. The results of this study reveal that Korean learners use completely new articulatory and acoustic strategies for their production of the English [ɫ]. All participants displayed a low tongue body and a heavily retracted tongue root and their F2 values were also greatly lower in the production of the English lateral. Although the Korean [l] is the closest L1 sound for English [ɫ], the results of this study suggest that they are perceived as two distinct phonemic categories.

구두 발표 I:
말장애 및 음성의학

좌장: 하지완(대구대)

내전형 연축성 발성장애 감별진단을 위한 선별검사의 내용타당도

조 재 경[1], 최 성 희[2], 이 상 혁[1], 진 성 민[1]
[1]성균관의대 강북삼성병원 이비인후과
[2]대구가톨릭대학교 언어청각치료학과

Content Validity of Screening Test for differential diagnosis of Adductor Spasmodic Dysphonia

Jae Kyung Cho, Seong Hee Choi, Sang Hyuk Lee, Sung Min Jin
Department of Otorhinolaryngology-Head and Neck Surgery,
Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine
Department of Audiology and Speech-Language Pathology,
Daegu Catholic University, Korea
jjk1226@naver.com, shgrace@cu.ac.kr, entlsh@hanmail.net, strobojin@hanmail.net

목적: 내전형 연축성 발성장애를 효율적으로 감별하기 위하여 선별검사 개발 및 문항의 내용타당도를 분석하고자 하였다. 방법: 개발된 문항에 대하여 20명의 언어치료전문가들이 검사항목에 대한 척도점수에 근거하여 내용타당도 지수를 산정하여 선별검사 문항을 작성한 후 내용타당도 검증을 실시하였다. 결과: 7개 평가문항들의 CVI는 모두 .60 이상으로 내전형 연축성 발성장애를 선별하는데 타당한 것으로 나타났다. 결론: 내전형 연축성 발성장애 선별검사 문항이 임상에서 감별 시 시간단축에 유용할 것으로 사료된다.

참고문헌

[1] Cannito, M. P., Doiuchi, M., Murry, T., & Woodson, G. E. (2012). Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates. *Journal of Voice*, 26(6), 818.e5-818.e13.

[2] Erickson, M. L. (2003). Effects of voicing and syntactic complexity on sign expression in adductor spasmodic dysphonia. *American Journal of Speech-Language Pathology*, 12(4), 416-424.

[3] 최성희. (2018). 음성장애 환자의 음질 평가를 위한 표준문장 개발. 한국청각언어재활학회논문집, 14(2), 128-142

병리적 음성떨림의 음향음성학적 특성 (음성떨림과 연축성발성장애, 근긴장성발성장애의 음향학적 비교)

서 인 효[1], 성 철 재4)[2]
단국대학교의과대학 이비인후과학교실[1], 충남대학교 언어학과[2]

Acoustic Characteristics of Pathological Vocal Tremor

Inhyo Seo[1], Cheol Jae Seong[2]
Department of Otorhinolaryngology-Head & Neck Surgery, College of Medicine, Dankook University[1], Department of Linguistics, Chungnam National University[2]
slpseo91@gmail.com, cjseong49@cnu.ac.kr

1. 서론

음성떨림(vocal tremor; VT)은 규칙적 저주파수 변동이 불수의적으로 나타남으로 인한 목소리 떨림으로 지각된다. 음성진전(vocal tremor; VT)과 연축성 발성장애(spasmodic dysphonia, SD), 근긴장성 발성장애(muscle tension dysphonia, MTD)는 목소리의 갑작스럽고 큰 저주파(2~12Hz) 파동(fluctuation)이 특징적이라고 하는 임상적 유사성이 있으며 합의된 진단기준이 부족하기 때문에 감별진단이 매우 어렵다. 음성병리에서 주로 분석되는 변동율변수(perturbation parameters)는 주로 청지각적 음질과 관련되어있으며 이러한 변수들로는 저주파의 큰 파동 특성을 분석할 수가 없다. 이에 본 연구자들은 Praat로 저주파 파동특성과 음질을 분석하여 병증 간 음향학적 특성을 비교하였다.

2. 연구방법

연구대상자는 음성떨림을 주소로 내원한 환자 중 후두부의 구조적 문제가 없는 79명(M:F=11:68)이었다. 목소리의 주기적 떨림, 급작스런 억압에 의한 목소리 끊김, 전반적 쥐어짬이나 걸걸거림 등의 청지각적 특성에 따라 한명의 음성언어치료사가 청지각적으로 반복측정하여 VT, SD, MTD로 분류하였다(평가자 내 신뢰도 Spearman's rho, .813 ~ .961). 대상자 중 VT는 39명(M : F = 4:35), SD는 16명(M : F = 4 : 12), MTD는 24명(M : F = 3 : 21)이었다. 평균연령은 46.41세였으며 연령범위는 21세에서 82세였다.

분석된 자료는 모음 /아/연장발성이었다. 음향학적 분석은 Praat 6.0.48로 분석사이즈를 윈도우 사이즈 30ms로 설정하여 음도와 강도 곡선 각각에서 파동을 검출하여 파동

4) 교신저자

정도(magnitude)의 변동율과 주기 불규칙성(aperiodicity)을 측정하고 켑스트럼 분석을 통해 cpp, rnr(rhamonics to noise ratio)을 측정하였다.

병증에 따른 음향학적 변수들의 차이를 알아보기 위해 정규성검정 결과에 따라 Kruskal-wallis 검정 후, 유의한 차를 보이는 변수에 대해서는 사후검정을 위해 순위변수에 대한 일원분산분석의 tukey검정을 하였다(SPSS Ver. 17.0, 25.0).

3. 결과 및 결론

3.1 음도파동의 주기 불규칙성

VT는 .47, SD는 .56, MTD는 .61이었으며 Kruskal-wallis 검정결과 병증 간 유의한 차를 보이지 않았다(p = .103)

3.2 음도파동의 정도 변동율

VT는 .26, SD는 .12, MTD는 .20이었으며 Kruskal-wallis 검정결과 병증 간 유의한 차를 보였다(p <.01). 사후검정결과 VT가 SD보다 유의하게 높았다(p <.01).

3.3 강도파동의 주기 불규칙성

VT는 .39, SD는 .53, MTD는 .51이었으며 Kruskal-wallis 검정결과 병증 간 유의한 차를 보였다(p <.01). 사후검정결과 VT가 SD와 MTD보다 유의하게 낮았다.

3.4 강도파동의 정도 변동율

VT는 .36, SD는 .015, MTD는 .018이었으며 Kruskal-wallis 검정결과 병증 간 유의한 차를 보였다(p <.01). 사후검정결과 SD가 VT와 MTD보다 유의하게 낮았다.

3.5 cpp

VT는 29.49, SD는 35.71, MTD는 28.22이었으며 Kruskal-wallis 검정결과 병증 간 유의한 차를 보였다(p <.01). 사후검정결과 SD가 VT와 MTD보다 유의하게 높았다.

3.6 rnr

VT는 1.81, SD는 5.80, MTD는 2.30이었으며 Kruskal-wallis 검정결과 병증 간 유의한 차를 보였다(p <.05). 사후검정결과 SD가 VT와 MTD보다 유의하게 높았다.

음도 파동의 주기 불규칙성을 제외한 모든 변수에서 병증 간 유의한 차를 보였다. 이 결과는 모음연장발성에서 음도 및 강도의 파동 특성과 음질에 대한 음향학적 분석이 VT, SD, MTD의 객관적 특성 규명 및 변별적 분류에 유용할 수 있음을 시사한다.

온라인 뉴스 발화의 말 속도 특성

김 수 연, 김 국 환, 유 재 연
호남대학교대학원 재활과학과 언어치료전공

The characteristics of speech rate of the online news

Suyeon Kim, Gukhwan Kim, Jaeyeon Yoo
Graduate Program in Speech-Language Pathology, Honam University, Gwangju, Korea
rarara727@naver.com, rnrghks64@naver.com, slpyoo@hanmail.net

본 연구는 온라인 뉴스 발화의 말속도 특성을 알아보고자 하였다. 화자의 말 속도는 성별, 제공되는 3가지의 말 속도(느림, 보통, 빠름), 그리고 검색사이트에 따라 어떤 차이가 있는가를 알아보기 위해 '분당 말한 음절수(SPM)'를 측정하였다. 본 연구는 '다음'과 '네이버'에 2019년 4월14일부터 4월 22일까지 탑재된 10개의 온라인 뉴스 발화(본문듣기)를 선정하였고 뉴스 주제는 일반적으로 청자들이 많이 접하는 경제, 스포츠, 생활, 세계 등으로 선택했다. 온라인 뉴스 발화의 말 속도를 분석하기 위해 녹음된 음성파일을 Praat(ver. 6.0.50) 편집 창에서 발화시간과 음절 수를 측정하였다. 자료처리는 SPSS(ver. 25)를 사용하여 기술통계를 통해 SPM의 평균값과 표준편차를 확인하였고 성별, 제공되는 3가지의 말 속도, 검색사이트에 따라 SPM에 차이가 있는가를 알아보기 위해 삼원분산분석을 실시하였다. 연구결과는 다음과 같았다. 첫째, 온라인 뉴스 발화의 전체 SPM 평균값은 341.68(±SD34.71)이었다. 둘째, 온라인 뉴스 발화의 SPM은 성별 간에 유의한 차이가 있었으며($p<.000$), 남자집단이 여자집단보다 말 속도가 유의하게 빨랐다. 셋째, 온라인 뉴스 발화에서 제공되는 3가지의 말 속도(느림, 보통, 빠름) 간에 SPM은 유의한 차이가 있었다($p<.000$). 넷째, 온라인 뉴스 발화의 SPM은 검색사이트 간에 유의한 차이가 있었으며($p<.000$), '다음'이 '네이버'보다 유의하게 말 속도가 빨랐다. 마지막으로, 말 속도와 검색사이트, 그리고 성별과 검색 사이트에서만 상호작용이 있는 것으로 나타났다($p<.000$). 본 연구 결과를 토대로 온라인 뉴스 발화의 말 속도가 일반인의 말 속도와는 어떤 관련이 있는지, 그리고 말 속도에 따라 청취자의 뉴스 이해도에 차이가 있는지 등에 대한 후속 연구가 필요할 것이다.

참고문헌

[1] Jong-Bok Ahn, Myung-Sun Shin, Do-Ha Kwon. (2002). The study of speech rate in normal-speaking adults and children, *Speech Sciences, 9*(4), 93-103.

[2] Jin-Won Park, Do-Ha Kwon. (2010). AccoAccording to the severity of stuttering adults of speed rate characteristics comparison studyrding to the severity of stuttering adults of speed rate characteristics comparison study, *The Jounal of Special Education: Theory and Practice, 11*(1), 129-146.

직업적 음성사용 패턴에 따른 음성 활동 및 참여 프로파일 비교

이 승 진[1], 임 성 은[2], 최 홍 식[3], 임 재 열[1]
1) 연세대학교 의과대학 이비인후과학교실 및 후두음성언어의학연구소
2) 강남세브란스병원 이비인후과 언어치료실
3) 하나이비인후과병원

A comparison of voice activity and participation profiles according to the patterns of professional voice use

Seung Jin Lee[1], Sung-Eun Lim[2], Hong-Shik Choi[3], Jae-Yol Lim[1]
1) Department of Otorhinolaryngology; The Institute of Logopedics & Phoniatrics, Yonsei University College of Medicine
2) Voice Clinic, Department of Otorhinolaryngology, Gangnam Severance Hospital
3) Hana ENT Hospital
slplee@yuhs.ac, selim@yuhs.ac, hschoimd@gmail.com, jylimmd@yuhs.ac

본 연구에서는 음성장애의 중증도를 통제하였을 때 양성 성대점막 질환 여부 및 직업적 음성사용 수준에 따라 음성 활동 제한 혹은 음성 활동 제약에 차이가 있는지의 여부를 알아보고자 하였다.

양성 성대점막질환으로 진단된 환자군 120명과 정상 대조군 120명이 참여하였다. 각 집단은 음성사용 수준에 따라 엘리트 음성 사용자(I), 직업적 음성 사용자(II), 직업적 비음성 사용자(III), 비직업적 비음성 사용자(IV) 각 30명으로 구성되었다. 심리측정적 평가로는 K-VAPP를, 청지각 평가로는 GRBAS를 시행하였다. K-VAPP 총점, 활동 제한 점수(ALS), 참여 제약 점수(PRS), 각 하위영역 점수(5개)를 종속변수로 하여 다중회귀분석을 시행하였다.

집단, 성별, 중증도(G)를 통제하였을 때, 기준 집단(IV)에 비해 집단 I은 총점, PRS, 중증도 인식 점수, 직업 영향 점수, 감정 영향 점수가 높았으며, 집단 II는 총점, ALS, PRS, 직업 영향 점수, 사회적 의사소통 영향 점수가 높았다. 반면, 집단 III은 기준 집단과 유의한 차이를 보이지 않았다.

직업적 음성사용 수준에 따른 이러한 차이는 엘리트 음성 사용자에 비해 직업적 음성 사용자가 겪는 음성 문제가 결코 작지 않다는 것을 시사하며, 이를 음성치료 계획 수립에 적극적으로 반영할 필요가 있을 것으로 여겨진다.

감사의 글

본 연구는 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018S1A5A2A02069988).

참고문헌

[1] Koufman, J. A. & Isaacson, G. (1991). The spectrum of vocal dysfunction. *Otolaryngologic Clinics of North America*, 24, 985-988.

[2] Lee, S. J., Choi, H. S., & Kim, H. (2018). A comparison of voice activities and participation profiles among etiological groups. *Journal of Voice*. Advanced online publication. Doi: 10.1016/j.jvoice.2018.04.016.

[3] 김재옥. (2015). 음성장애가 있는 직업적 음성사용자와 비직업적 음성사용자의 음성장애 중증도와 유형에 따른 자기보고식 음성평가 차이. 말소리와 음성과학, 7, 67-76.

[4] 이승진, 최홍식, 김향희, 변형권, 임성은, 양민교. (2016). 음성 활동 및 참여 프로파일-한국판(K-VAPP): 타당성 검증 연구. *Communication Sciences & Disorders*, 21, 695-708.

구두 발표 I:
음성공학

좌장: 박기영(ETRI)

자기 모방 발음 교정을 위한 GAN의 스펙트로그램 학습

양 승 희*, 정 민 화**
서울대학교 협동과정 인지과학*, 서울대학교 언어학과**

Self-imitating Feedback Generation Using GAN by Spectrogram Learning

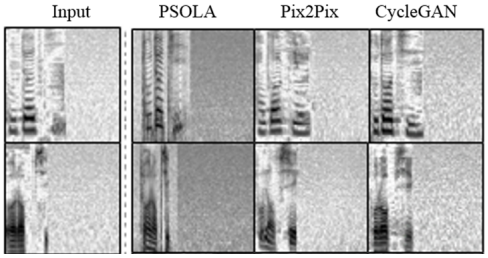
Seung Hee Yang, Interdisciplinary Program in Cognitive Science,
Minhwa Chung, Department of Linguistics, Seoul National University
sy2358@snu.ac.kr, mchung@snu.ac.kr

컴퓨터 기반 발음 교육 시스템에서 자기모방 발음 교정은 학습자 친화적이고, 효과적인 피드백 생성 방법이다. 전통적으로는 PSOLA 알고리즘을 기반으로 원어민 데이터의 음향 특징들을 추출한 후, 이를 학습자의 음성에 이식하여 들려주는 방식을 사용하였으나, 이는 초분절 단위의 교정만이 가능하다는 한계를 갖는다. 본 연구에서는 원어민과 비원어민 발화를 시각화한 스펙트로그램의 언어학적 차이점을 분석하고, 이에 대한 기계 학습을 위해 Conditional GAN과 Cycle Consistency GAN 알고리즘을 적용하여 초분절음 뿐만 아니라 분절 단위의 교정이 가능한 음성을 생성하였다.

스펙트로그램에서 확인되는 포먼트와 이미지 농도는 음소를 변별하는 정보를 나타낼 수 있으며, 단어가 주어졌을 때, 원어민 및 비원어민 발화 특성을 확인할 수 있다. 예를 들면, 외국인의 한국어에서 빈번하게 나타나는 경음의 평음화, 종성 삭제, 이중모음의 단모음화, 끊어읽기, 강세 오류, 권설음 개입 현상은 스펙트로그램에서 확인이 가능하다. 본 연구는 GAN 네트워크를 사용하여 이러한 차이점을 학습하고, 발음 오류를 교정할 수 있는지 실험하였다.

GAN의 생성모델의 목표는 학습자의 스펙트로그램을 입력으로 받아 원어민의 스펙트로그램을 출력하는 것이다. 51명 원어민과 27개 모국어 배경의 217명 학습자 발화에 대한 300단어 스펙트로그램 이미지로 학습한 결과, CycleGAN 생성모델의 교정 능력을 확인할 수 있었다. 그림 1은 생성된 스펙트로그램의 예시이다. 이들을 음성으로 변환하여 지각 평가를 시행한 결과 표 1과 같이 초분절음뿐만 아니라 분절음 단위 교정이 가능하다는 것이 관찰 되었다.

<그림 1. 스펙트로그램 생성 결과 예시>



<표 1. 지각 평가 MOS 결과 비교>

모델	교정 능력			모방 유사도	음질	평균
	총체적	분절음	초분절음			
PSOLA	3.118	3.029	3.324	4.029	2.794	3.259
Pix2Pix	1.970	2.485	2.152	2.697	1.636	2.188
Cycle-GAN	4.000	4.333	4.364	3.515	2.667	3.776

Learnable Dictionary Encoding 기반의 문장 독립 화자 검증 방법[5)]

정 영 문, 최 연 주, 김 회 린
KAIST 전기및전자공학부

Learnable Dictionary Encoding based Text-Independent Speaker Verification

Youngmoon Jung, Yeunju Choi, Hoirin Kim
School of Electrical Engineering, KAIST
{dudans, wkadldppdy, hoirkim}@kaist.ac.kr

최근 각광받고 있는 d-vector 기반 화자 검증 시스템은 다음과 같이 세 가지 모듈로 이루어진다. (1) deep neural network(DNN)을 이용하여 frame 단위의 deep feature를 추출하는 특징 추출 모듈, (2) 추출된 deep feature들을 종합하여 발화 단위의 화자 임베딩(d-vector)을 생성하는 pooling 모듈, (3) 서로 다른 화자로부터 생성된 d-vector들을 이용하여 화자를 분류하도록 훈련되는 화자 분류기이다.

본 연구에서는 learnable dictionary encoding(LDE) 기반의 두 가지 pooling 모듈을 제안한다. LDE 방식은 기존의 Gaussian mixture model(GMM) supervector 방식의 supervector 인코딩 과정을 딥러닝과 결합한 것으로, convolutional neural network (CNN) 기반의 특징 추출 모듈로부터 추출된 deep feature를 이용하여 supervector를 생성한다. 이러한 LDE 방식은 입력 feature들의 화자별 차이를 보여주는 시간 정보를 무시한다는 단점이 있는데, 이것을 극복하기 위하여 deep encoding pooling(DEP) 방식과 spatial pyramid encoding(SPE) 방식을 제안한다. DEP 방식은 LDE 방식과 temporal average pooling(TAP) 방식을 결합하여 LDE 방식을 보강한다. SPE 방식은 LDE 방식과 spatial pyramid pooling(SPP) 방식을 결합하여 LDE 방식을 보강한다.

실환경 조건에서의 화자 검증 테스트를 위하여 Voxceleb1 데이터를 이용하였다. 기존에 제안되었던 LDE pooling 모듈 및 다른 pooling 모듈들과의 비교를 통해 제안한 pooling 모듈들의 성능 향상을 보였다. 또한 제안한 SPE pooling 모듈을 최근 얼굴 인식 분야에서 도입된 ring loss 및 large margin softmax loss와 결합하여 4.03%의 equal error rate(EER)을 달성하였다.

5) 본 연구는 산업통상자원부의 산업기술혁신사업으로부터 지원을 받아 수행된 연구임 (No. 10063424, '실내용 음성대화 로봇을 위한 원거리 음성인식 기술 및 멀티 태스크 대화처리 기술 개발').

스푸핑 음성 검출의 성능향상을 위한 다양한 고해상도 정보의 활용

심혜진, 정지원, 허희수, 유하진
서울시립대학교 컴퓨터과학과

Complementary high-resolution information for improving the performance of spoofing detection

Hyejin Shim, Jeeweon Jung, Heesoo Heo, Hajin Yu
School of Computer Science, University of Seoul
shimhz6.6@gmail.com, jeeweon.leo.jung@gmail.com, zhasgone@naver.com, hjyu@uos.ac.kr

재생공격에 의한 스푸핑 음성은 사용자의 음성을 녹음한 후 다시 재생하는 과정을 통해 생성된다. 이러한 스푸핑 음성이 사용자의 음성으로 오인식 되는 경우 화자 인증 시스템의 신뢰도가 저하되기 때문에, 이를 방지하기 위해서는 스푸핑 음성을 검출하는 연구가 필요하다.

기존의 스푸핑 음성 검출 연구는 전문가들의 음향 지식에 기반한 새로운 특징을 개발하는 방향으로 진행되어왔다. 하지만 최근 음향 기기가 다양해지고 성능이 향상됨에 따라, 스푸핑 음성과 사용자의 음성을 구분하던 기존 특성들이 점차 사라지거나 줄어들고 있다. 따라서, 기존의 지식 기반 방식으로는 음향기기의 다양화 및 성능 향상을 반영한 스푸핑 음성의 검출 성능이 하락할 수 있다.

본 연구에서는 기존의 지식 기반 특징 대신 종단간 심층신경망을 이용하여 다양한 고해상도 정보를 활용한 스푸핑 검출 방법을 제안한다. 제안한 방법은 다양한 정보를 이용하여 음향 기기를 포함한 새로운 재생 공격 환경에 대응할 뿐만 아니라, 고해상도 정보를 이용해 재생 음성과 사용자 음성 간의 미세한 차이점에 집중하는 것을 목표로 하였다. 다양한 정보를 활용하기 위해, 푸리에 변환을 통해 얻을 수 있는 스펙트럼과 위상 정보, 에너지 스펙트럼 밀도 정보와 함께, 별도의 특징인 i-vector와 원신호를 사용하였다. 제안한 방법의 성능은 ASVspoof2019 데이터를 사용한 스푸핑 음성 검출 실험을 통해 확인하였다. 실험 결과, 베이스라인의 평가세트에 대한 t-DCF 와 동일 오류율이 각각 0.2454, 11.04% 인 것과 비교하여 제안한 방법의 경우 각각 0.0570, 2.45%로, 제안한 방법이 효과적인 것을 확인하였다.

이 논문은 미래창조과학부 재원으로 경찰청과 치안과학기술연구개발사업단의 지원을 받아 수행된 치안과학기술연구개발사업임. (PA-J000001-2017-101)

가창 음원의 음악적 표현 이식 시스템

용 상 언, 남 주 한
한국과학기술원 문화기술대학원

Singing Expression Transfer System from One Voice to Another

Sangeon Yong, Juhan Nam
Graduate School of Culture Technology, KAIST
koragon2@kaist.ac.kr, juhannam@kaist.ac.kr

이 논문에서는 주어진 노래에 대해 하나의 가창 음성에서 다른 가창 음성으로 음악적 표현을 자동으로 이식하는 음성 처리 알고리즘을 제시한다. 가창 음성의 경우 음색에 해당하는 부분 이외에도 가창자의 숙련도에 따라 박자와 음정, 셈여림의 측면에서 차이와 개성이 발생하게 된다. 이 논문에서 제안하는 알고리즘은 총 네 가지 단계를 거쳐서 이러한 음악적 표현들을 이식한다. 첫번째 단계는 가창 음성 추출 단계로, 기성 음원 내의 목소리로부터 음악적 표현을 추출하기 위해 음원 분리 기술을 적용하여 음원으로부터 가창 음성을 분리한다. 두번째 단계는 박자 정렬 단계로 가장 먼저 노래의 음정에 해당하는 음정정보와 가사에 해당하는 음소정보를 각각 심층학습을 활용한 음정 추출기 음소 추출기를 활용해 추출한 뒤, 이를 기반으로 유사도 행렬을 생성하여 두 음원의 박자 정보를 분석한다. 유사도 행렬로부터 시간 워핑 기술로 추출된 워핑 경로는 평활화 과정을 거친 뒤, 이를 기반으로 시간축 변환 알고리즘을 활용해 박자를 정렬한다. 두 음성의 박자가 정렬되면 박자 정렬이 이루어진 음원으로부터 음정 정보를 추출한 뒤, 이를 기반으로 시간축 변환 알고리즘을 활용해 음정을 정렬한다. 마지막으로, 두 음원으로부터 셈여림 정보를 추출한 뒤, 엔빌로프 정렬을 활용해 프레임 단위의 음정과 셈여림 정보를 정렬한다. 알고리즘을 검증하기 위한 실험 과정은 숙련된 가수의 음악적 표현을 음악적 표현이 다소 풍부하지 못한 비숙련자의 목소리에 이식하는 형태로 진행하였고, 이를 위해 총 12쌍의 가창 음원쌍에 알고리즘을 적용한 뒤 결과 비교를 진행하였다.

참고문헌

[1] Nicholas J. Bryan, Jorge Herrera, and Ge Wang (2012). User-guided variable-rate time-stretching via stiffness control. *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx)*

[2] Pedro Cano, Alex Loscos, Jordi Bonada, Maartende Boer, and Xavier Serra (2000). Voice morphing system for impersonating in karaoke applications. *Proceedings of the International Computer Music Conference (ICMC)*

[3] Tomoyasu Nakano and Masataka Goto (2009). Vocalistener:A singing-to-singing synthesis system based on iterative parameter estimation. *Proceedings of the Sound and Music Computing Conference (SMC)*

[4] Takeshi Saitou, Masataka Goto, Masashi Unoki, and Masato Akagi (2007). Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*

[5] Shimpei Aso, Takeshi Saitou, Masataka Goto, Katsutoshi Itoyama, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (2010). Speakbysinging:Converting singing voices to speaking voices while retaining voice timbre. *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx)*

[6] Jonathan Driedger and Menard Müller (2014). TSM toolbox:MATLAB implementations of time-scale modification algorithms. *Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx)*

[7] F.J. Charpentier and M.G. Stella (1986) Diphone synthesis using an overlap-add technique for speech waveform concatenation. *Acoustics, Speech, and Signal Pro-cessing, IEEE International Conference on ICASSP'86*

포스터 발표 II

좌장: 손민정(한남대)
박상희(대구사이버대)
박형민(서강대)

형태소 유형에 따른 한국어 단모음의 음성적 실현

이 향 원, 신 지 영
고려대학교 국어국문학과

The Phonetic Realization of Korean Monophthongs in Different Morphological Contexts

Hyangwon Lee, Jiyoung Shin
Department of Korean Language and Literature, Korea University
gidskan33@korea.ac.kr, shinjy@korea.ac.kr

이 연구의 목적은 한국어 단모음이 어휘 형태소에 속할 때와 문법 형태소에 속할 때 음성적 실현에서 어떠한 차이를 보이는지 밝히는 데 있다. 단어의 범주(word category)와 음성적 실현의 연관성을 탐구한 연구들에서는 기능어가 내용어에 비하여 음성적으로 약화되는 경향이 있음을 보고한 바 있다[1][2][3]. 이 연구에서는 한국어에서도 문법 형태소의 모음이 어휘 형태소에 비하여 음성적으로 약화되는 경향이 나타나는지 확인해 보고자 하였다. 이를 위해 한국어 자연발화 음성코퍼스(서울코퍼스) 20-30대 여성 10명의 발화 자료에서 관찰된 15,842개의 단모음을 대상으로 해당 모음을 포함한 형태소의 유형을 주석하였다. 또한 운율적 위치의 영향을 통제하기 위하여 발화의 중간에 위치한 어절에 속하면서 어절 말 위치에 오는 모음으로 대상으로 한정하였다. 그 결과, 문법 형태소에서 모음 공간이 더 작게 실현될 것이라는 예측과는 달리, 문법 형태소와 어휘 형태소의 모음 공간 크기는 차이가 없는 것으로 나타났다. 지속시간에서도 어휘 형태소의 모음이 더 길게 실현될 것이라는 예측과 달리 형태소 유형에 따른 차이가 나타나지 않았다. 다른 언어내적 요인의 영향을 최소화하기 위하여 품사와 어절 내 음절수를 통제하여 재분석을 수행하였을 때에도 동일한 결과를 얻었다. 이 연구에서는 한국어 단모음이 형태소 유형에 따라 음성적 실현의 차이를 보이지 않는 이유에 대해서 한국어에서는 어휘 형태소와 문법 형태소의 구분에 강세가 개입하지 않는 것과, 어두 강화 효과로 인해 어휘 형태소와 문법 형태소가 구분되므로 어말 위치에서는 상대적으로 두 형태소의 구분이 불필요한 것으로 해석하였다.

참고문헌

[1] Johnson, K. (2004). Massive reduction in conversational American English. *In Spontaneous speech: Data and analysis. Proceedings of the 1st session of the*

10th international symposium, 29-54.
[2] Meunier, C., & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, 39(3), 271-278.
[3] Nguyễn, T. A. T. (2015). Vowel reduction in conversational speech in Vietnamese. *International Journal of Asian Language Processing*, 23(2), 91-109.

다문화 가정 이주여성들의 한국어 파열음 유성개시시간 비교
- 중국, 일본, 베트남 이주민을 중심으로-

박선영, 성철재[6)]
충남대학교 언어병리학과

Comparison of VOT of Korean Plosives among Multicultural Immigrant Women

Sunyoung Park, Cheoljae Seong
Department of Speech & Language Pathology, Chungnam National University
psy6252@naver.com, cjseong49@gmail.com

본 연구는 한국인 화자와 중국인, 일본인, 베트남 다문화 가정 이주 여성의 한국어 파열음 유성개시시간(VOT)을 비교한 것이다. 실험에 참여한 대상자는 중국인 , 일본인, 베트남인, 그리고 한국인 여성 대학생 각 30명으로 총 120명이었다. 이들에게 {파열음+아}로 구성된 '바, 빠, 파, 다, 따, 타, 가, 까, 카'를 천천히 읽도록 하였다.

발성유형 3가지와 조음자리 3가지를 개체내 변수로 하고 국가를 개체간 변수로 하는 3웨이 혼합 반복검정 분산분석을 실시한 결과 개체내 변수의 경우, 조음자리 주효과[$F_{(2, 232)}$=42.513, p=.000]와 발성유형 주효과[$F_{(2, 232)}$=205.99, p=.000]는 모두 유의하였고, 조음자리*발성유형*국가의 상호작용 또한 유의하였다[$F_{(12, 464)}$=5.876, p=.000]. 개체간 변수 또한 유의하였다[$F_{(3, 116)}$=8.439, p=.000]. 상호작용의 의미를 파악하기 위하여 본페로니 교정 짝대응(pairwise comparison) 사후검정을 실시한 결과 다음과 같은 내용을 관찰할 수 있었다.

발성유형의 관점에서, 양순파열음의 경우 평음의 VOT는 한국>일본(p=.000), 한국>베트남(p<.05), 중국>일본(p<.05)으로 관찰되었고, 격음과 경음은 유의한 쌍이 없었다. 치조파열음은 평음의 경우, 한국>일본(p<.05)으로 나타났고, 격음은 중국>한국(p<.05), 중국>일본(p<.01)로, 경음은 중국>{한국, 일본, 베트남}(p=.000)으로 특이하게 관찰되었다. 연구개파열음은 평음은 유의한 쌍이 없었고, 격음은 중국>일본(p=.000), 중국>한국(p<.01), 베트남>일본(p<.01)으로 관찰되었고, 경음은 유의한 쌍이 없었다.

6) 교신저자

각 출신 국적 별 조음위치에 따른 발성유형간 VOT의 차이를 살펴보면 다음과 같다. 중국인의 경우, 양순파열음은 평음과 격음(p<.05), 경음(p=.000)사이에 유의한 차이가 있었고(평음<격음; 평음>경음), 격음과 경음도 유의한 차이가 있었다(p=.000, 격음>경음). 치조파열음은 평음과 격음(p<.05) 사이에 유의한 차이가 있었고(평음<격음), 평음과 경음(p=.27), 그리고 놀랍게도, 격음과 경음(p=.146) 간 유의한 차이가 없었다. 연구개파열음은 평음과 격음(p<.01), 평음과 경음(p=.000), 격음과 경음(p=.000) 간 모두 유의한 차이가 있었다(격음>경음>평음). 일본인의 경우, 양순파열음은 평음과 격음(p=.000), 평음과 경음(p<.05), 격음과 경음(p=.000) 간 유의한 차이가 있었고(격음>평음>경음), 치조파열음은 평음과 격음(p<.05), 격음과 경음(p=.000) 사이에 유의한 차이가 있었고(평음<격음: 격음>경음) 평음과 경음 사이에는 유의한 차이가 없었다(p=.385). 연구개파열음은 평음과 경음(p=.000), 격음과 경음(p=.000) 간 유의한 차이가 있었고, 놀랍게도, 평음과 격음 간 유의한 차이가 없었다(p=.236). 베트남인의 경우, 양순파열음은 평음과 격음(p=.000), 평음과 경음(p<.01), 격음과 경음(p=.000) 사이에 유의한 차이가 있었고(격음>평음>경음), 치조파열음도 평음과 격음(p<.01), 평음과 경음(p<.01), 격음과 경음(p=.000) 사이에 유의한 차이가 있었다(격음>평음>경음). 연구개파열음도 모든 짝대응에서 유의한 차이가 관찰되었다(p=.000, 격음>평음>경음). 한국인의 경우, 양순파열음은 평음과 격음은 유의한 차이가 없었고(p=1.000), 평음과 경음(p=.000, 평음>경음), 격음과 경음(p=.000, 격음>경음) 사이에 유의한 차가 있었다. 치조파열음도 평음과 격음 간 유의한 차이가 없었고(p=1.000), 평음과 경음(p=.000, 평음>경음), 격음과 경음(p=.000, 격음>경음) 간 유의한 차이가 있었다. 연구개 파열음도 평음과 격음 간 유의한 차이가 없었고(p=.812), 평음과 경음(p=.000), 격음과 경음 간 유의한 차이는 관찰되었다(p=.000).

선형판별분석을 이용하여 각 출신 국적 별로 발성유형을 구분하는 작업을 시도하였다. 분류에 활용된 독립변수는 유성개시시간(vot), 개방 파열 부분(5ms)의 스펙트럴 에너지(i_energy: pascal to dB conversion), 후행모음 첫번째 피치포인트의 피치값(vowelPitch)을 선정하였다(floor: 100, ceiling: 500). 결과, 중국인의 경우 구조행렬(structure matrix)의 판별적재값(load) 절대값이 vot>i_energy>vowelPitch 순으로 나타났으며 평음 정분류율은 42.2%, 격음 정분류율은 72.2%, 경음 정분류율은 61.1%였다. 일본인의 경우도 구조행렬 절대값은 중국인과 마찬가지 순서로 배열되었으며, 평음 정분류율은 43.3%, 격음 정분류율은 60%, 경음 정분류율은 72.2% 였다. 베트남인의 경우도 구조행렬 절대값 배열순서는 위와 동일하였다. 평음 정분류율은 45.6%, 격음 정분류율은 72.2%, 경음 정분류율은 71.1% 였다. 한국인의 경우 구조행렬은 vot>vowelPitch>i_energy 순으로 유성개시시간과 후행모음의 피치가 분류에 좀 더 지대한 역할을 한 것으로 관찰되었고, 평음 정분류율은 67.8%, 격음 정분류율은 71.1%, 경음 정분류율은 97.8%(88/90)로 경음의 괄목할만한 분류율이 특징적이었다.

한국인 영어 학습자의 영어 숙련도에 따른 모음 사이 폐쇄음의 유성음화 경향

유 지 윤, 전 지 현, 이 석 재
연세대학교 영어영문학과

Effects of L1 on Voicing of Intervocalic Voiceless Stops and its Relation with L2 Proficiency

Jiyun Yoo, Jihyun Jeon, Seok-Chae Rhee*
Dept. of English Language and Literature, Yonsei University
jiyun6843@naver.com, wjswlgus6623@google.com, scrhee@yonsei.ac.kr

본 발표는 한국인 영어 학습자의 (L2 영어 발화시) 모음 사이 무성폐쇄음의 발화에 있어 한국어의 유성음화 규칙(Korean intervocalic voicing rule)의 간섭(interference) 정도와 영어 발화 유창성 등급과의 관련성에 대한 연구이다. 한국어에는 무성폐쇄음이 모음 사이에 올 경우 유성음화되어 발음되는 규칙이 존재하지만, 영어의 경우는 그렇지 않다. 이러한 두 언어 간에 모음 간 유성음화 규칙의 존재 유무 차이를 통해 한국인의 L2 영어 발화에 있어 발화 평가 등급이 낮을수록 (즉, 유창성이 떨어질수록) 간섭 현상이 더 일어나 폐쇄음의 유성성(voicing)의 정도가 심해질 것이라는 예측 가설을 세울 수가 있고, 본 연구는 이러한 예측 가설을 테스트한다. 분석 대상으로는 한국인의 L2 영어 발화를 유창성에 따라 다섯 등급으로 평가한 *Genie SpeeCor* 음성코퍼스가 사용되었으며, 폐쇄음의 유성성 정도의 음향지표로는 '강도 강하 (intensity drop)' 수치로 계산하였다 ([1]. 모음 사이의 폐쇄음이 무성음일 경우 강도 강하가 큼). /t/의 경우 flapping 현상과 유성음화의 구분이 어려워 분석 대상에서 제외하였고, /p/와 /k/의 강도 강하만을 분석 대상으로 삼았다. 분석결과로는 높은 등급에 있는 영어 학습자의 모음 사이 폐쇄음 발화는 낮은 등급 영어 학습자의 발화에서보다 상대적으로 큰 강도 강하 값을 보였다. 이는 높은 등급의 영어 학습자일수록 한국어 유성음화 규칙의 간섭이 줄어 모음 사이 폐쇄음을 유성음화하여 발화하는 경향이 적음을 확인시켜준다.

참고문헌

[1] Warner, N., & Tucker, B.V. (2011). Phonetic variability of stops and flpas in spontaneous and careful speech. *The Journal of the Acoustic Society of America*, 130(3), 1606-17.

YELC 음성코퍼스 구축과 과제 유형별 분석적 평가의 총체적 점수에 대한 예측 분석

이 석 재
연세대학교 영어영문학과

Construction of YELC Speech Corpus and Prediction of Analytic Evaluation Factors to Holistic Evaluation Results

Seok-Chae Rhee
Dept. of English Language and Literature, Yonsei University
scrhee@yonsei.ac.kr

본 발표는 YELC 2011 음성코퍼스의 구축 및 평가 과정과 그 특징을 개관하고 코퍼스 내 과제 유형별 (대화형, 독백형, 낭독형) 유창성 평가 결과에서 나타나는 총체적(holistic) 평가 점수에 대비 각 분석적(analytic) 평가 항목(분절음 요인 점수, 운율적 요인 점수, 문법성 요인 점수, 어휘 선택 요인 점수)의 예측 영향력을 제시한다. YELC 2011 음성코퍼스의 두드러진 특징으로는 i) 발화 평가에 따른 등급이 매겨진(rated) 음성코퍼스이고, ii) 한 코퍼스 내에 대화형, 독백형, 낭독형 과제가 모두 포함된 다유형 음성코퍼스이며, iii) 즉흥 발화가 포함된 음성코퍼스란 점이다. 발표의 후반부는 이러한 YELC 2011 음성코퍼스의 유형별 과제에서 제반 분석적 평가 점수의 총체적 평가 점수에 대한 예측 분석 결과를 제시한다. 대화형 과제의 총체적 점수에 대한 예측력은 다중회귀분석 결과 어휘선택의 적절성이 가장 높으며, 문법성에 대한 평가가 가장 낮다고 할 수 있으며 (Word(β=.424), Segment(β=.273), Prosody(β=.212), Grammar(β=.163)의 순), 이 결과는 독백형 과제에서도 유사한 결과가 도출되었다 (Word(β=.390), Prosody(β=.329), Segment(β=.202), Grammar(β=.118)의 순). 반면 낭독형 과제에서는 (평가 루브릭에서 어휘선택과 문법성은 배제: 보고 소리내어 읽은 것이기 때문) 분절음 평가 점수, 발화속도 및 휴지 평가 점수, 억양 구현에 관한 점수, 강세와 리듬 구현에 대한 점수 등 4개의 분석적 평가 요인 중 분절음 점수 요인이 총체적 점수에 대한 예측 영향력이 가장 큰 것으로 결과가 산출되었다 (Segment(β=.452), Rate-Pause(β=.252), Intonation(β=.176), Stress-Rhythm(β=.146)의 순).

한국인 EFL 학습자 영어 발화에서 능숙도에 따른 모음약화 현상 연구

강 태 훈, 임 유 경, 이 석 재
연세대학교 영어영문학과

The Relationship Between English Vowel Reduction and Proficiency Levels in Korean EFL Learners' Speech

Tae Hoon Kang, You Kyung Lim, Seok-Chae Rhee
Dept. of English Language & Literature, Yonsei University
xogns0214@daum.net, youkyung1991@naver.com, scrhee@yonsei.ac.kr

본 연구는 한국인 영어학습자의 영어 문장 발화에서 모음약화 현상과 영어 능숙도 간의 관계성을 밝히는 데 목적이 있다. 영어에서는 한국어에서와는 달리 문장 내에서 강세가 있는 부분을 제외하고 모음이 약화되어 발화되는 경향이 강하다. 따라서 영어 능숙도가 높은 영어학습자는 'as soon as'와 같은 문장 내 어구에서 /æ/ 모음을 /ə/로 발화하는 반면, 능숙도가 낮은 학습자들은 모음을 약화시키지 않을 것으로 예상되어진다. 본 연구에서는 5가지의 레벨로 평가되어있는 지니 음성코퍼스에서 성인 한국어 발화자들이 읽은 영어 문장들을 능숙도에 따라 세 그룹(상, 중, 하)으로 나누고, 강세를 받지 않는 기능어인 'is', 'at, as', 'to' 내에서 발화되는 세 가지 모음 [ɪ], [æ], [u]의 뚜렷한 특성(F1 & F2)들이 얼마나 약화되어 /ə/의 값과 유사한지를 보고자 한다. 분석 결과 '하' 그룹에선 모든 모음을 축약 없이 특성을 뚜렷하게 발음하는 경향을 보였다. 반면 '중'과 '상' 그룹 사이에서는 모음마다 약화 현상의 차이나는 정도가 각각 달랐으며 '하' 그룹만큼 뚜렷한 차이를 보이지는 않았다. 이러한 결과로 볼 때 영어 발화에서 모음의 약화 양상은 영어 L2 화자들의 능숙도 그룹을 '하' 대비 '상'과 '중'으로 구분하는데 역할을 하는 것으로 보인다.

참고문헌

[1] Lee, B., Guion, S. G., & Harada. T. (2006) Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, *28*(3), 487-513.

[2] Flemming, E. (2009). The phonetics of schwa vowels. In Donka Minkova (ed.) *Phonological weakness in English.* London: Palgrave Mcmillan. 78-98.

한국인 화자 연령 지각 연구

서 윤 정, 신 지 영
고려대학교 국어국문학과

Age estimation from Korean adult voices

YoonJeong Seo, Jiyoing Shin
Korea University
phon_etic@naver.com, shinjy@korea.ac.kr

본고는 한국인 청자를 대상으로 지각 실험을 진행하여 화자의 실제 연령(Chronogical age)과 지각 연령(Perception age) 간의 상관관계를 살피고, 한국인 청자가 얼마나 정확하게 익명의 화자의 연령을 지각할 수 있는지를 밝히고자 한다. 또한, 이러한 연령 판단에 음성적 단서가 되는 음 높이와 발화 속도와 지각 연령 간의 영향 관계를 검토하고자 한다.

이를 위해, 한국인 성인 청자 71명을 대상으로 3가지의 지각 실험을 진행하였다. 실험 자극은 표준어 화자 40명에게서 추출되었으며, 자유 발화, 낭독 발화, 모음 연장 발성으로 구성되었다. 각 실험은 10초 내외의 음성을 듣고 연령을 구체적인 숫자로 답하는 방식으로 진행되었다. 분석 결과, 한국인 청자들은 상당히 높은 판단 정확도를 보였으며, 모음 연장 발성을 들었을 때보다 자유 발화와 낭독 발화를 들었을 때 화자의 연령을 더욱 정확하게 짐작하였다(r2 자유 발화: 0.8, 낭독 발화 0.8, 모음 연장 발성: 0.5). 이는 음성이 포함하고 있는 정보의 양의 차이에 기인한 결과로 보인다. 또한, 음성 분석을 수행한 결과 청자들은 화자의 발화 속도와 음 높이를 참고하여 화자의 연령을 추정하는 것으로 나타났다.

이러한 연령 판단 지각 실험은 지금까지 한국어 연구에서는 거의 수행되지 않았으므로, 본 연구는 연령과 음성 간의 관계에 대한 색다른 접근으로서 의의를 가질 수 있을 것이다.

Second language acquisition of intonation: nuclear peak alignment in English

Hye Jeong Yu
English language and literature, Hanshin University
hyejeong.u@gmail.com

American English has six pitch accents (H^*, L^*, H^*+L, H+L^*, L^*+H, L+H^*).[2] Standard Korean has an Accentual Phrase (AP), marked by a phrasal tone sequence, HHLH or LHLH. When the AP-initial segment is an aspirated or tense consonant, the initial segment has H; otherwise, the AP-initial segment has L.[1] This study investigated the phonetic implementation of tonal targets in English of Korean learners. The data was collected from 10 Korean English learners and 3 American English speakers. H^* and L+H^* were the two most common pitch accents observed in the English and Korean speakers. The pitch accent association to segments was found in the Korean speakers ($p<.001$). H^* occurred more frequently when the onset of the stressed syllable is voiceless and L+H^* occurred more frequently when the onset of the stressed syllable is voiced. But, regardless of the onset of the stressed syllable, the English speakers had H^* more than L+H^*. The advanced learners of English had better stress production than the beginner learners, but there was no parallel processing of the alignment of the high tone (H) with proficiency in L2.

참고문헌

[1]Jun, S. A. (2000). K-Tobi (Korean ToBI) Labeling Conventions: Version 3, *The Korean Journal of Speech Science,* 7(1), 143-169.

[2] Pierrehubert, J. (1980). *The Phonology and Phonetics of English intonation.* Ph.D. dissertation, MIT.

주관적 음성장애 인지와 비염의 상관성 연구

이 인 애, 성 철 재[7)]
충남대학교 언어병리학과

Relationship between Perception of Self-Reported Voice Problem and Rhinitis

Inae Lee, Cheol Jae Seong
Department of Speech Pathology, Chungnam National University
hadi_in@naver.com, cjseong49@gmail.com

본 연구에서는 비염의 유무가 주관적 음성장애 인지에 미치는 영향을 확인하고자 하였다. 제5기(2011년) 국민건강영양 조사 이비인후과 검진을 완료한 19세 이상의 성인을 6,010명을 대상으로 하며 주관적 음성장애 인지와 비염의 상관성을 분석하였다. 비교를 위해 한글판 SPSS 18.0(Statistics package for social science, version 18.0, USA)을 사용하였으며 chi-square, 이분형 로지스틱 회귀분석을 이용하였다.

그 결과, 주관적 음성평가를 통한 음성장애 유무와 비염의 유무는 유의한 차이를 나타내는 특성을 보였다($p<.001$). 또한 주관적 음성장애 인지에 영향을 미치는 요인으로 비염의 유무와의 관련정도를 확인한 결과, 비염이 있는 경우 주관적 음성평가를 통한 음성장애 인지도가 약 1.65배더 높았다($p<.001$). 따라서 본 연구를 통해 과소비성을 유발하는 질환 중 하나인 비염이 음성을 통한 의사소통에 어려움을 나타내며, 이로인해 주관적 음성장애 인지에 영향을 미친다는 것을 확인할 수 있었다.

서론

비음의 생성은 연인두 폐쇄와 연관이 있는데, 연구개의 기능이 정상이어도 과소비성을 일으키는 질환이 있으면 공명장애가 나타날 수 있다(김난숙 등, 2017). 그 대표적 질환 중 하나인 비염을 동반한 환자의 경우 목소리의 불편함을 호소하며, 이것이 과소비성으로 인한 비음의 산출의 불편감과 의사소통의 전달력 저하가 그 원인으로 작용함을 예측하여 볼 수 있다. 때문에 본 연구에서는 비염이라는 질환이 주관적으로 음성장애를 인식하는데 유의한 차이를 나타내는지 확인하여보고자 한다.

연구방법

7) 교신저자

본 연구에서는 제5기 2011년 국민건강영양조사 이비인후과 데이터를 바탕으로 6,010명의 남녀 성인을 대상으로 하였으며, 이비인후과 설문에서 음성장애를 평가하는 항목(변해원, 2011)과 비염경험 증상 항목을 이용하였다. 주관적 음성평가를 통한 음성장애의 유무와 비염의 유무의 명목형 변수는 Chi-square test를 이용하여 비교하였고, 비염의 유무에 따른 주관적 음성장애 위험도는 이분형 로지스틱 회귀분석을 이용하였다.

연구결과

1) 주관적 음성평가를 통한 음성장애의 유무와 비염의 유무

주관적 음성평가를 통한 음성장애 유무와 비염의 유무는 유의한 차이를 나타내는 특성을 보였다(p<.001).

비염		있음(n=1411)	없음(n=4599)	*p*-value
음성장애	있음(n=415)	136 (9.6%)	279 (6.1%)	.000**
	없음(n=5596)	1275(90.4%)	4320 (93.9%)	

***p<.001

2) 비염의 유무에 따른 주관적 음성장애 위험도

비염이 있는 경우 주관적 음성평가를 통한 음성장애 위험이 약 1.65배 더 높았다(p<.001). 또한 남자의 경우 음성장애 위험이 약 1.78배 더 높았고(p<.001), 여자의 경우 주관적 음성평가를 통한 음성장애 위험이 약 1.59배 더 높았다(p<.001).

			p-value	OR	95% CI	
전체	비염	(1= 있다)	0.000	1.652	1.333	2.046
남자		(1= 있다)	0.000	1.786	1.258	2.538
여자		(1= 있다)	0.000	1.592	1.214	2.088

***p<.001

본 연구결과를 종합하여 볼 때, 비음은 기류를 비강을 통해 내보내면서 조음하는데(성철재, 1996), 비염의 경우 비강통로의 개방을 방해하여 과소비성을 유발하고 따라서 주관적 음성장애 인지에 영향을 미친다는 것을 알 수 있다.

참고문헌

[1] 김난숙, 성철재. (2017). 과소비성(Hyponasality) 집단의 음향 특성과 분류 변인. 언어학, 78, 31-61.

[2] 변해원. (2011) 주관적 음성문제 인지와 음성장애의 연관성. 말소리와 음성과학. 3(1), 111-116.

[3] 성철재. (1996) 한국어 비음의 음향적 특성에 관한 실험음성학적 연구. 말소리, 31, 9-22.

아동의 유음 /ㄹ/에 대한 청지각적·음향학적 연구

김 인 경, 성 철 재[8)]
충남대학교 언어병리학과

A perceptual and acoustical study of Korean /ㄹ/ in children's speech

In-gyeong Kim, Cheol-jae Seong
Department of Speech Pathology, Chungnam National University
slpkyeong@naver.com, cjseong49@gmail.com

본 연구는 무의미 음절 수준에서 아동의 유음 /ㄹ/ 변이음의 조음 발달을 청지각적 및 음향학적으로 분석하고자 하였다. 본 연구에서는 친숙도에 따른 말소리 산출에 영향을 줄이기 위하여 모음 유형과 /ㄹ/의 음운 환경을 고려한 무의미 음절로 구성하였다. 모음 유형은 한국어 모음 삼각도의 각 극점에 위치한 단모음 /ㅏ/, /ㅣ/, /ㅜ/ 3가지로 제한하였다. 음운 환경은 유음 /ㄹ/의 변이음이 실현 가능한 환경 중 한국어에서 대표적으로 조음되는 환경을 선택하였다.

설측음은 단어 내 어말, 어중 초성 자음 앞의 어말, 설측음과 연결된 어중 초성(겹자음)환경을 선택하였고 탄설음은 어중초성 환경을 선택하였다. 이를 토대로 모음 /ㅏ/, /ㅣ/, /ㅜ/에 각각의 음운 환경을 결합하여 총 12개 항목을 과제로 구성하였다. 연구 대상은 만 4세 정상 발달 아동 20명이었으며 연구자의 입을 가린 후 모방 발화를 유도하여 진행하였다. 각 과제 당 5회 발음 하도록 하였고 그 중 안정적으로 녹음 된 3회분을 선택하여 분석하였다. 아동들의 조음 능력은 9명의 평가자에 의해 1점(부정확하다) ~ 7점(정확하다)의 7점 척도 방법으로 평가 되었다. 이후 유음의 음향학적 특성을 고려하여 발화 내 시간 비율, 발화 내 에너지 비율, 유음의 무게중심을 음향 변수로 선정하여 음향학적 분석을 실시하였다.

청지각 평가 결과, 자음 정확도 평정치는 어말 종성이 가장 높고 다음으로 자음 앞, 겹자음, 어중 초성으로 나타나 설측음의 정확도가 탄설음의 정확도 보다 높은 것으로 나타났다. 음향학적 분석 결과, 정조음 집단과 오조음 집단 간 발화 내 유음의 시간 비율과 발화 내 유음의 에너지 비율에서 유의한 차이가 나타났지만 유음의 무게 중심에서는 유의한 차이가 없는 것으로 나타났다.

본 연구는 유음이 낱말 내 위치에 따라 발달 상 차이가 존재한다는 사실을 확인해주었으며 아동의 유음에 대한 의미 있는 음향 변수를 측정하여 음향학적 분석에 대한 가

8) 교신저자

능성을 제시하였다. 하지만 본 연구에서 나타난 제한된 무의미 음절 환경과 표본 수를 극복한 후속 연구가 필요하다고 판단되며, 후속 연구를 통해 본 연구가 조음 중재 시 보다 실용적인 정조음 지표로 활용될 수 있다고 생각된다.

참고문헌

[1] 박애경, 이승환(2000). 모음 환경에 따른 초성 /ㅅ/, /ㅈ/ 및 /ㄹ/의 산출 연구. *언어청각장애연구*, 5(2), 74-90.

[2] 이호영(1996). *국어 음성학*. 서울: 태학사.

[3] Ray D. Kent. & Charles Read.(2002). The Acoustic Analysis of Speech. 2nd ed. New York: Thomson Learning.

소아의 안면근육 움직임 평가도구 개발: 예비연구

조태양, 봉성록, 정해철, 윤혜수, 박소현, 김성태
동신대학교 언어치료학과

Development of a Facial Muscle Motion Assessment Tool for Children: Preliminary Study

Taeyang Jo, Sungrok Bong, Haechul Jung, Hyesoo Yoon,
Sohuyn Park, Seong-Tae Kim
Department of speech-language pathology, Dongshin University
towha@naver.com, voicekim88@hnmail.net

목 적: 조음문제를 가진 아동들의 언어중재 효과를 극대화하기 위해서는 조기 감별의 필요성이 요구된다. 허나, 아직까지 조기 감별에 적합한 표준화된 선별검사는 거의 없는 실정이다. 이에, 저자들은 안면근육의 움직임과 조음문제의 연관성에 초점을 맞추어 소아를 대상으로 안면근육의 유동성과 비대칭성을 점검하고 선별하는 평가도구를 개발하고자 본 연구를 시행하였다.

방 법: 본 연구는 2018년 10월부터 12월까지 언어치료를 받기위해 본원 언어치료센터에 내원한 아동 중 조음음운장애 또는 언어발달장애로 진단되어진 14명의 아동들을 대상으로 하였다. 단, 구개열, 뇌병변 등 수술 병력이 있거나 선천적 구조결함을 가진 아동들은 실험대상에서 제외하였다. 대상자들의 성별은 남아 7명, 여아 7명이었으며, 연령은 3세 8개월에서 13세 9개월로, 평균 7세 1개월이었다. 대상자들 모두 본원에서 개발한 DSU-FACT(Dongshin university-facial activity condition test)를 시행하여 안면근육 비대칭성(7개 수행과제, 0=normalfacial activity condition, 3=severe), 안면근육 유동성(5개 수행과제, 5=normal, 1=severe) 그리고 안면근육 약화 및 마비 유무 등을 평가하였다. 총 점수는 유동성 점수에서 비대칭성 점수를 차감한 백분위점수로 환산하여 표기하였다. 또한, 실험군과 동일한 연령대의 정상대조군 13명의 아동들을 대상으로 DSU-FACT를 시행하여 실험군과 비교하였다.

결 과: DSU-FACT로 안면근육 비대칭성을 평가한 결과, 조음문제를 가진 아동들과 조음문제와 언어발달문제를 함께 가진 아동들이 정상대조군에 비해 비대칭 점수가 높게 나타났다. 또한 안면근육 유동성을 평가한 결과, 조음문제를 가진 아동들은 평균 88점으로 정상대조군(평균 95점)에 비해 유동성점수가 낮게 나타났으며, 언어발달에만 문제를 가진 아동들은 평균 84점으로 나타났다. 그러나 조음문제와 언어발달문제를 함께

가진 아동들의 경우, 평균 64점으로 정상대조군에 비해 유동성점수가 유의미하게 낮은 것으로 나타났다. 안면근육 약화 및 마비를 평가한 결과, 조음문제와 언어발달문제를 함께 가진 아동 중 80%가 약화 및 마비 증상이 있는 것으로 나타났다.

결 론: 소아들의 조음문제와 언어발달문제가 안면근육의 비대칭적 움직임과 관련이 있는 것으로 확인되었으며, 향후 의사소통에 문제를 보이는 아동들은 조기에 안면근육 움직임의 상태를 점검하고 선별해야 할 것으로 사료된다.

참고문헌

[1] Kanerva, M., Jonsson, L., Berg, T., Axelsson, S., Stjernquist-Desatnik, A., Engström, M., & Pitkäranta, A. (2011). Sunnybrook and House-Brackmann systems in 5397 facial gradings. *Otolaryngology-Head and Neck Surgery*, 144(4), 570-574.

[2] Namasivayam, A. K., Pukonen, M., Goshulak, D., Vickie, Y. Y., Kadis, D. S., Kroll, R., ... & Luc, F. (2013). Relationship between speech motor control and speech intelligibility in children with speech sound disorders. *Journal of Communication Disorders*, 46(3), 264-280.

[3] 김보선 (2018). *조음문제를 보이는 아동의 안면 특징 및 음향학적 특성*. 동신대학교 석사학위눈문.

[4] 김호정, 김병조, 고성범, 박건우, 이대희, 이승환 (2007). *Sunnybrook 척도를 이용한 특발성 안면신경마비의 acyclovir 치료 효과 및 예후 분석*. 대한신경과학회, 25(4), 455-461.

음성장애 유형에 따른 음향학적 특성과 청지각적 평가 간의 상관성

이 옥 분, 심 희 정
대구사이버대학교 언어치료학과, 한림대학교 언어청각학부

Relation Acoustic and Auditory-Perceptual Measures according to Different Dysphonic Voice

Okbun Lee, Heejeong Shim
DaeguCyber University, Hallym University
oblee@dcu.ac.kr, skim@gmail.com

마비말장애나 음성장애로 인한 구어장애 화자들의 말명료도의 청지각적 평가와 더불어 음성 음향학적 평가는 말소리 장애의 심한 정도를 진단하고 중재전략을 계획하는 데 있어서 중요한 역할을 한다. 마비말장애의 경우 불명확한 조음과 음성의 다면적인 문제가 동반되어 나타나기에 말명료도를 평가할 때 말소리 총체적 관점과 더불어 말산출시스템 영역에 대한 명료성 판단 요구된다. 왜냐면 말명료도 저하의 원인이 조음과 음성 시스템 중 어디에 비중을 두느냐에 따라 중재방향이 달라지기 때문이다. 말소리 명료성 관련 음향학적 측정치들은 모음공간면적(vowel space area, VSA), 모음 조음지수(vowel articulatory index, VAI), 모음중앙화비율(formant centralization ratio, FCR), 그리고 입술의 원순성과 평순성의 대립 정도를 지시하는(모음의 명료도와 관련) F_2 ratio($F_{2/이/}/F_{2/우/}$) 등이 있다. 이 연구에서는 마비말장애와 달리 조음문제가 없는 정상성인들의 음성문제만이 명료성 관련 음향학적 측정치들이 어떻게 차이가 있는지 그리고 음성문제의 청지각적 평가와 어떤 관련성이 있는 지 알아보았다. 본 연구에서는 음성장애(성대결절, MTD, 폴립, 마비)로 진단받은 42명의 화자를 대상으로 음성의 청지각적 평가와 함께 음향학적 분석(VSA, VAI, FCR, F_2 ratio)을 실시하였다. 연구 결과 첫째, 음성장애 유형에 따른 모음공간면적(VSA), 모음중앙화를 나타내는 FCR(formant centralization ratio)의 차이는 유의미하게 나타나지 않았다. 그러나, 조음운동성 관련 변수, VAI(vowel articulatory index)와 모음(/이/-/우/)의 원순성과 평순성 대립(모음 명료도) 관련 변수 F_2ratio에서는 통계적으로 유의한 차이가 나타났다. 이는 모음공간면적에서는 유형별 유의한 차이를 보이지 않았으나, 모음조음의 정확성에 대해서는 유의미한 차이를 보임을 알 수 있었다. 이러한 경향은 성대결절 그룹과 성대마비 환자 그룹에서 보다 두드러지게 나타났다. 이 결과는 음성장애 유형별로 모음산출의 명료성이 달라질 수 있다는 점을 시사한다. 연구결과 두번째, 피어슨 상관분석을 통해, 음성장애유형과 VAI, F2ratio의 상관성이 유의미하게 나타났다. 그리고 청지각적 평가 결과의 G와 B, A가 모음의 조음정확지수를 예표하는 F2ratio와 유의미한 상관이 있는 것으로 나타났다. 이는 음성장애 환자들의 음성산출 시 /i/와 /u/의 조음에 있어서 차이를 보일 수 있는 것으로 해석할 수 있다. 이 결과들을 토대로 임상에서는 음성장애를 동반한 다양한 말소리 장애 평가 과제를 구성할 때 모음유형에 따른 산출 정확성과 지각적 명료성을 고려해야 할 것이다.

증강현실 기반 음성치료 소프트웨어의 개발 및 적용[9)]

박 희 준*, 김 근 효**, 이 연 우**, 배 인 호***, 권 순 복****
*부산가톨릭대학교 언어청각치료학과, **부산대학교병원, ***양산부산대학교병원, ****부산대학교 언어정보학과

Development and Application of Augmented Reality Based Voice Therapy Software

Heejune Park*, Geunhyo Kim**, Yeonwoo Lee**, Inho Bae***,Soonbok Kwon****
*Dept. of Speech and Hearing Thrapy, Catholic University of Pusan, **Pusan National University Hospital, ***Pusan National University Yangsan Hospital, ****Dept. of Language and Information, Pusan National University
june@cup.ac.kr

본 연구에서는 기존 음성장애 치료에서 사용할 수 있는 실시간 음도 피드백과 강도 피드백을 이용하기 위해서 음도와 강도의 목표 역치를 6단계로 설정 할 수 있도록 구성하였고, 피험자가 마이크로폰의 음도와 강도 역치 범위 내에서 발성을 하면 역치와 대응되는 3차원 캐릭터가 모니터에 자신의 현실 모습과 동시에 출력되도록 하였다. 아동의 흥미를 유발하기 위해 애니메이션 캐릭터(에그구그, SBS 방영 애니메이션)를 치료에 맞게 자체 제작하여 사용하였다.

아동의 음성평가 결과를 바탕으로 목표 음도와 강도 범위를 설정하였으며, 목표 범위 내에서 발성하였을 때 증강현실 콘텐츠가 출력되어 목표 음도와 강도에 대한 실시간 시각적 피드백 제공이 가능하였다. 또한 기존의 2차원적인 시각적 피드백을 이용한 전통적인 치료 방법보다 흥미를 유발하는데 더 효과적이었으며 몰입도도 높은 것으로 나타났다.

치료 동기가 부족하고 집중도와 몰입도가 적은 아동 음성치료에서 증강현실 콘텐츠를 적용한 음성치료는 흥미도와 몰입도가 높았다. 향후 아동의 흥미와 몰입도를 높일 수 있는 다양한 3D 콘텐츠 개발 및 3차원 게임을 이용한 소프트웨어 개발을 통해 몰입도를 길게 유지할 수 있는 추가적인 연구가 필요할 것으로 판단된다.

9) This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5B6075173)

아동의 말소리산출 능력에 대한 음운산출지식(PPK) 분석 연구

이 손 희*, 하 지 완**
*대구대학교 일반대학원 재활과학과, **대구대학교 언어치료학과

Productive Phonological Knowledge Analysis for Children's Spontaneous Speech

Son Hee Lee* Ji-Wan Ha**
*Daegu Univ. Graduate School, **Daegu Univ.,
sonhiy9047@hanmail.net, jw-ha@daegu.ac.kr

음운산출지식(productive phonological knowledge, PPK)은 단편적인 말소리산출능력 평가방법의 한계점을 극복한 것으로, 동일한 말소리에 대한 사용능력을 다양한 조건에서 여러 차례 심층적으로 분석한다[1]. 본 연구에서는 3세 후반에서 6세 사이의 일반아동 3명과 말소리장애아동 2명을 대상으로 그림 보고 따라말하기와 스스로 말하기 과제를 이용하여 자발화를 수집한 후, 이에 대한 말소리산출 능력을 PPK를 이용하여 분석하였다. 따라말하기 과제는 모든 어절 및 음절 위치, 모든 모음문맥에서 19개의 한국어 자음이 세 번 이상 산출될 수 있는 단어들로 구성되었고, 스스로 말하기는 아동의 습관화된 발음을 유도하기 위해 실시하였다. PPK의 1단계는 목표음소가 오류 없이 모두 정확하게 산출된 경우, 2단계는 특정 문맥에서만 오류 보이는 경우, 3단계는 특정단어에서 습관화된 오류를 보이는 경우, 4단계는 초성위치에서는 정확하나 종성위치에서 오류를 보이는 경우, 5단계는 초성위치에서도 간혹 오류를 보이는 경우, 6단계는 목표음소가 전혀 산출되지 못하는 경우를 말한다. 연구결과 자음정확도가 높은 아동은 전반적으로 높은 PPK를 보인 반면, 자음정확도가 낮은 아동은 상대적으로 낮은 PPK를 보였다. 아동 내에서 완전 습득된 음소 또는 숙달된 음소의 경우 비교적 높은 PPK 단계에 해당하였다. 또한 공식검사에서 오조음으로 평가되었던 각 음소들에 대해서도 다양한 PPK 수준이 관찰되었다. 본 연구결과로 인해 PPK 분석을 통해 오류음소에 대한 음운지식정도를 보다 구체적으로 파악할 수 있음을 확인하였다. 음소 간 다양한 PPK 수준은 오류 말소리에 대한 구체적인 중재계획을 세우는 데에 기여하는 바가 있을 것이다.

참고문헌

[1] Gierut, J. A., Elbert, M., & Dinnsen, D. A. (1987). A Functional Analysis of Phonological Knowledge and Generalization Learning in Misarticulating Children. *Journal of Speech, Language, and Hearing Research, 30*(4), 462-479.

일반아동의 자극양식, 자극길이, 음운유사성 및 시연여부에 따른 음운단기기억 수행력 비교

전 은 혜*, 하 지 완**
*대구대학교 일반대학원 재활과학과, **대구대학교 언어치료학과

The effect of presentation modality, stimuli length, and phonological similarity in children's short term memory with and without vocal rehearsal

Eun Hye Jeon*, Ji-Wan Ha**
*Daegu Univ. Graduate School, **Daegu Univ.
cehye@naver.com, jw-ha@daegu.ac.kr

본 연구에서는 자극양식(시각적 자극 vs. 청각적 자극), 시연여부(외현적 시연 vs. 시연억제), 자극어 길이(숫자 네 개 vs. 숫자 여섯 개), 자극어의 음운유사성(음운유사성 높음 vs. 음운유사성 낮음)을 음운정보 저장의 효율성에 영향을 주는 요소로 보고, 각 요소들이 일반아동의 음운단기기억 수행력에 미치는 영향을 알아보고자 하였다. 음운유사성의 경우 음운유사성 정도가 높은 자극어와 낮은 자극어를 제작하기 위해 선행연구를 참고하여 다음과 같은 기준을 적용하였다. 첫째, 숫자에 동일한 음소가 많이 포함되어 있을수록 음운유사성이 높다. 둘째, 자음과 모음 중 동일한 모음이 더 많이 포함되어 있을수록 음운유사성이 높다. 셋째, 음소수준에서 유사성이 동일한 경우 음절구조가 유사할수록 음운유사성이 높다. 실험과제는 숫자 자극 제시로부터 13초 경과 후 제시된 숫자를 순서대로 회상하여 말하는 지연 숫자따라하기 과제이다. 13초 동안 자극어를 소리 내어 되뇌인 후 말하는 시연 조건과 /더/라는 단순한 소리를 반복하게 함으로써 시연활동을 방해한 후 말하게 하는 시연억제 조건에서 실험을 진행하였다. 자극어는 4개의 숫자와 6개의 숫자로 구성되었고, 각 자극어 내 숫자 간 음운유사성 정도에 따라 고-음운유사성 자극어와 저-음운유사성 자극어로 구분되었다. 또한 자극어는 시각적 글자자극와 청각적 소리자극의 형태의 두 가지 양식으로 제시되었다. 이와 같은 실험과제는 DmDx 소프트웨어 프로그램으로 제작되었고, 5세의 일반아동 7명이 본 연구에 참여하였다. 연구 결과 자극양식, 시연 여부, 자극어 길이에 대한 주효과가 유의하였으나, 음운유사성 여부에 대한 효과는 유의하지 않았다. 그리고 시연 여부와 자극어 길이의 상호작용효과 또한 유의하였다. 이와 같은 결과를 통해 시각적으로 정보를 접하고, 외현적 시연을 사용하고, 자극어의 길이가 짧을수록 아동의 음운단기 수행력이

유의하게 향상됨을 알 수 있다. 따라서 이러한 요소들을 적극 활용한다면 아동의 음운 단기기억력과 어휘 확장, 더 나아가 언어발달에 긍정적인 영향을 줄 것이라 기대한다. 한편 본 연구의 대상자인 5세 아동에게는 Baddeley (2003)의 이론[1]에서처럼 음운유사성 효과, 즉 음운적으로 유사한 말소리들로 이루어진 자극어에서 단기기억 수행력이 떨어지는 현상은 나타나지 않았다. 이와 관련하여 해당 연령대 아동들의 경우 아직 기억책략을 활발하게 사용하지 못하는 시기임을 고려할 때, 시연에 대한 부담감, 이로 인해 선행연구[1]의 성인 대상자들과 달리 음운유사성 효과가 나타나지 않았을 가능성 등을 고려해보아야 할 것이다.

참고문헌

[1] Baddeley, A (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36, 189-208.*

정상 여성의 음성범위 프로파일과 발화범위 프로파일

김 재 옥*, 이 승 진**
*강남대학교 교육대학원 언어치료교육전공
**연세대학교 의과대학 이비인후과, 음성언어의학연구소

Voice Range Profile and Speech Range Profile in Healthy Women

Jaeock Kim*, Seung Jin Lee**
*Major in Speech Pathology Education, Graduate school of Education, Kangnam University
**Department of Otorhinolaryngology, Institute of Logopedics and Phoniatrics, College of Medicine, Yonsei University
jaeock@gmail.com, metallican1024@daum.net

음성범위 프로파일(Voice Range Profile, VRP)은 주파수와 음성강도의 최저부터 최고 범위를 측정하는 것으로, 모음을 연장 발성하는 동안 평가하기 때문에 말을 산출하는 동안의 후두의 생리적인 변화로 인한 발성능력을 살펴보기 어려우며, 이를 측정하는데 상당한 시간이 소요된다[1]. 이에 기능적 말산출 수행능력을 측정할 수 있는 발화범위 프로파일(Speech Range Profile, SRP)을 살펴볼 필요가 있다.

본 연구는 정상 음성을 산출하는 24명의 20대 여성(연령 21.92±6.43)을 대상으로 VRP와 SRP를 통한 주파수와 음성강도 범위를 살펴보고자 하였다. VRP는 /a/ 모음 활창하기 과제(활창)와 축약된 VRP 과제(축약)[2]로 측정하였고, SRP는 21부터 30까지의 숫자세기(숫자세기)와 타당도 검증을 통해 새롭게 개발된 “불이야” 문장읽기 과제(문장읽기)로 측정하였다. 각 과제에서 주파수범위(최저-최대주파수, $F0_{range}$)와 음성강도범위(최저-최대음성강도, I_{range})의 평균과 표준편차를 측정하였고, 4가지 과제 간의 차이를 반복측정분산분석으로 비교하였다.

$F0_{range}$는 과제들 간에 유의한 차이를 보였으며(F=21.000, p<.001), 주효과 분석을 실시한 결과, 활창과 숫자세기 간에만 유의한 차이가 있었고, 다른 과제들 간에는 유의한 차이가 없었다. 또한 I_{range}는 과제들 간에 유의한 차이를 보였고(F=10.565, p<.001), 주효과 분석 결과, 활창과 축약 간, 활창과 문장읽기 간에 유의한 차이를 보인 반면, 활창과 숫자세기 및 축약과 문장읽기 간에는 유의한 차이가 나타나지 않았다. 이에 축약된 VRP와 문장읽기 SRP를 통해 산출된 주파수와 음성강도 범위 간에는 차이가 없어 VRP 대신 문장읽기를 통한 SRP로 발성 및 말산출 최대능력을 평가할 수 있음을 알 수 있다.

감사의 글

본 연구는 2018년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2018S1A5A2A03032902).

참고문헌

[1] D'Alatri, L. & Marchese, M. R. (2014). The speech range profile (SRP): an eary and useful tool to assess vocal limits. *Acta otorhino-laryngologica italica, 34*(4), 253-258.

[2] 정원정(2018). *음성장애 환자를 위한 축약된 음성범위프로파일(Voice Range Profile, VRP) 검사법 개발과 타당도*. 대구가톨릭대학교 석사학위논문.

심층 신경망의 화자 기저를 활용한 화자인증 손실 함수 정의 기법

허희수, 정지원, 심혜진, 유하진
서울시립대학교 컴퓨터과학과

Novel Loss Functions for Speaker Verification Based on Speaker Bases of Deep Neural Network

Hee-soo Heo, Jee-weon Jung,
Hye-jin Shim, and Ha-jin Yu
School of Computer Science, University of Seoul
zhasgone@naver.com, jeewon.leo.jung@gmail.com,
shimhyejin930615@gmail.com, hjyu@uos.ac.kr

최근 화자인증에 적합한 특징을 추출하는 심층 신경망을 학습시키기 위해 손실 함수를 설계하는 많은 연구가 수행되고 있다. 본 논문에서는 화자 기저 개념을 적용해 설계한 두 종류의 화자인증 손실 함수를 제안한다. 여기서 화자 기저는 출력층 가중치 행렬의 역할을 분석해 정의한 개념으로 학습 가능한 파라미터이다.

먼저 화자 간 분산을 증가시키는 손실 함수 ℓ_{BC}를 다음과 같이 정의하였다:

$$\ell_{BC} = \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \cos(W_i, W_j),$$

위 수식에서 N은 식별하고자 하는 화자의 수, $\cos(\cdot,\cdot)$는 두 벡터 사이의 코사인 유사도, W_i는 i번째 화자의 기저 벡터를 의미한다. 두 번째 손실 함수는 hard negative mining을 포함한 metric learning이 가능하도록 다음과 같이 정의하였다:

$$\ell_H = \sum_{W_h \in H} \log(1 + \exp(\cos(W_h, \boldsymbol{x}) - \cos(W_y, \boldsymbol{x}))),$$

위 수식에서 $\boldsymbol{x}$와 y는 각각 입력 데이터와 해당 데이터의 화자 레이블, H는 입력 $\boldsymbol{x}$와 유사한 특성을 가지는 다른 화자 집합을 의미한다. 기존의 손실 함수들의 경우 mini-batch 구성에 종속돼 한정된 수의 화자 만을 고려할 수 있는 것과 다르게, 제안한 손실 함수들은 화자 기저를 활용해 전체 화자를 한번에 고려할 수 있다.

VoxCeleb1과 VoxCeleb2 데이터셋을 활용한 실험 결과, 기존의 손실 함수를 제안한 화자 기저 기반의 손실 함수로 대체하면 화자인증 오류를 상대적으로 약 20% 감소시킬 수 있는 것을 확인하였다.

감사의 글

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초 연구사업임(2017R1A2B4011609)

감정 음성 합성을 위한 토큰 기반의 음향 모델링

최 희 진, 박 진 욱, 한 민 수
한국과학기술원 전기및전자공학부

Token based acoustic modeling for Expressive speech synthesis

Heejin Choi, Jinuk Park, Minsoo Hahn
School of Electrical Engineering, KAIST
{change, windclay, mshahn2}@kaist.ac.kr

최근 Global Style tokens(GSTs)을 이용한 비지도 스타일 모델링이 성공적으로 end-to-end 음성 합성 시스템에 적용되어 평균화된 스타일만을 표현했던 기존의 문제점을 보완하였다 [1]. GSTs는 여러 개의 임베딩 벡터들로 구성되며 음성 합성 시스템과 함께 훈련되어 데이터 셋에 의존하게 되므로 각 token이 어떤 의미를 가지는지 정의할 수 없다. 우리는 GSTs의 성공적인 사례를 차용하여 GSTs에 감정 라벨을 부여한 새로운 감정 스타일 모델을 구축하였으며 이를 음성 합성 시스템의 음향 모델링에 적용하였다. 각 감정에 대응하는 token set이 구성되어 감정 스타일 모델의 입력 감정 종류에 따라 감정 token set은 스위치처럼 작용하여 해당 감정의 데이터로만 훈련이 된다. 4가지 감정(중립, 기쁨, 슬픔, 화남)으로 발화된 데이터 셋을 이용하여 실험한 결과 기존의 GSTs을 이용한 합성음보다 제안된 감정 token set을 이용한 합성음에서 좀 더 감정이 잘 표현되었다. t-SNE를 이용한 시각화 결과 감정 스타일 모델의 출력인 감정 스타일 임베딩으로 감정 분류가 가능했으며 이를 토대로 같은 감정 스타일 임베딩을 모두 추출한 뒤 평균하여 4개의 감정 스타일 임베딩으로 통합하였다. 따라서 입력 텍스트가 주어졌을 때 정규화된 감정 스타일 임베딩을 직접 음향 모델의 입력으로 사용하여 원하는 감정 스타일로 발화하는 합성음을 얻을 수 있었다.

감사의 글

본 연구는 산업통상자원부의 산업기술혁신사업으로부터 지원을 받아 수행된 연구임 (No. 10080667, 음원 다양화를 통하여 로봇의 감정 및 개성을 표현할 수 있는 대화음성합성 원천기술 개발).

참고문헌

[1] Y. Wang, D. Stanton, Y. Zhang, R. Skerry-Ryan, E. Battenberg, J. Shor, Y. Xiao, F. Ren, Y. Jia, and R. A. Saurous, “Style tokens: Unsupervised style modeling, control and transfer in end-to-end speech synthesis,” *arXiv preprint arXiv:1803.09017*, 2018.

단어의 Text Label을 활용한 Acoustic Word Embedding 추출

정 명 훈, 김 회 린
한국과학기술원 전기및전자공학부

Extracting Acoustic Word Embedding using Text Label of Word

Myunghun Jung, Hoirin Kim
School of Electrical Engineering, KAIST
kss2517@kaist.ac.kr, hoirkim@kaist.ac.kr

단어를 고정된 길이의 벡터로 표현하는 word embedding 방식 중, 단어의 의미가 아닌 어떻게 발음되는가를 표현하는 acoustic word embedding(AWE) 방식이 keyword spotting 및 spoken term detection 등에 적극적으로 활용되고 있다. 이러한 AWE을 추출하는 모델을 학습하기 위해 triplet loss 기반의 siamese network 구조를 도입하는 많은 연구가 진행되어 왔다. 최근 단어의 character 단위 text label과 acoustic feature를 동시에 입력으로 활용하여 AWE 모델을 학습하는 multi-view learning 방식이 기존의 acoustic feature만으로 학습하는 방식보다 높은 성능을 보여주었다[1]. Character 단위의 표기법은 어떻게 발음되는가를 어느 정도 반영하지만, 같은 character여도 앞뒤에 어떤 character가 오는지에 따라 발음을 정확히 표현하는데 한계가 있다. 본 연구에서는 이러한 문제점을 해결하기 위하여 phoneme 단위의 text label 입력을 사용함으로써 발음을 보다 잘 반영할 수 있도록 했다. 또한 triplet의 negative sample을 얻는 과정에서 edit distance를 활용한 sampling 기법을 적용하여 모델의 학습을 도왔다. 실험은 WSJ DB를 이용한 word discrimination task에 대하여 진행되었으며, 평가 지표는 average precision(AP)을 사용하여 기존의 0.791보다 높은 0.814의 성능을 얻었다.

감사의 글

본 연구는 산업통상자원부의 산업기술혁신사업으로부터 지원을 받아 수행된 연구임 (No. 10063424, '실내용 음성대화 로봇을 위한 원거리 음성인식 기술 및 멀티 태스크 대화처리 기술 개발').

참고문헌

[1] W. He, W. Wang, and K. Livescu (2017). Multi-view recurrent neural acoustic word embeddings. *Proc. Int. Conf. on Learning Representations*.

Research on Novel Speech Intelligibility Correlation Coefficient Based on Time-Frequency Domain SNRs

Seon Man Kim
Korea Photonics Technology Institute
smkim@kopti.re.kr

Objective intelligibility measuring algorithms such as STOI and NCM have been widely used in the development process of speech processing systems because subjective listening tests are costly and time consuming. However, none of these methods provide cues on how intelligibility of noisy speech can be enhanced in de-noising algorithms such as spectral subtraction and Wiener filter.

We propose here an objective measure for predicting the intelligibility of noisy speech that is based on counting the number of perceived time-frequency (T-F) speech component bins. The intelligibility is determined by frame wise counting the number of T-F bins above and below threshold. This strategy is motivated by the assumption that human beings have suprathreshold speech intelligibility enhancing mechanisms, so the number of actually perceived speech components is a more important factor for speech intelligibility than the signal-to-noise ratio (SNR).

The simulation result shows that the proposed method highly correlated with the state-of-the-art conventional speech intelligibility predicting algorithms such as STOI and NCM in various noise and noise reduction situations.

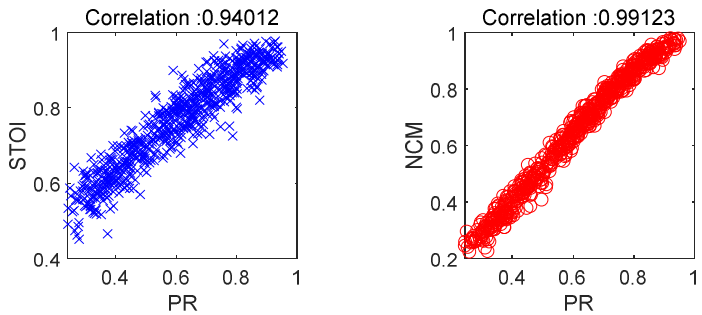


Figure 1. The relationship between conventional and proposed speech intelligibility predicting values on white noise conditions ranging -12 dB to 12 dB SNRs.

딥 러닝 오토인코더 기반 보코더 파라미터 감축

이 정 혁[1], 김 홍 국[1], 김 선 교[2]
[1]광주과학기술원 전기전자컴퓨터공학부, [2]국방과학연구소

Vocoder Parameter Reduction Based on Deep Learning Auto-Encoder

Jung Hyuk Lee[1], Hong Kook Kim[1], Seon Kyo Kim[2]
[1]School of Electrical Engineering and Computer Science, GIST
[2]Agency for Defense Development
{[1]ljh0412, [1]hongkook}@gist.ac.kr, [2]skkim@add.re.kr

보코더는 음성에서 특징 파라미터를 추출하고, 이를 비트스트림으로 양자화하여 전송하고 이를 복원한다. 이 때, 파라미터를 auto-encoder 구조의 bottleneck feature를 통하여 감축시킬 수 있다.

본 논문에서는 표준 mixed-excited linear prediction (MELP) 보코더에 대하여, 딥러닝 기반의 auto-encoder (AE) 구조를 통한 파라미터 감축 방법을 연구한다. 우선, MELP 보코더를 적용하여 각 22.5ms의 음성 프레임마다 29개의 파라미터를 추출한다. 다음으로, 이렇게 추출된 29차원의 파라미터 벡터를 입력 및 출력으로 하는 AE를 구성한다. 본 논문에서 사용된 AE는 29개 노드의 입력층은 두 개의 은닉층(여기서 각 은닉층은 64개, 8개 노드를 갖으며 전연결층 형태임)과 16개 노드를 갖는 전연결층을 병렬 구조의 skip-connection에 공유된다. 이 때, 8개 노드를 갖는 두번째 은닉층과 skip-connection 계층을 출력으로 하는 encoder와, 이로부터 획득된 24개의 파라미터를 이용하는 decoder로 나누어 생각할 수 있다. 또한, 8개 노드를 가지는 은닉층과 16개 노드를 가지는 skip-connection은 각기 29개 노드를 갖는 은닉층에 연결되며, 해당 은닉층의 각 29개의 출력을 합하여 29개의 파라미터를 회귀 출력하는 출력층으로 구성된다. 이와 같은 방법으로 기존의 29개 파라미터는 24개로 감축된다.

성능 평가를 위해, TIMIT 음성 데이터베이스 중 6,300문장을 AE 훈련 및 validation에 사용하였으며, 20개의 테스트 문장에 대해 AE로 디코딩한 후, MELP 합성한 음원에 대해 PESQ 점수를 측정하였다. 그 결과, 29개 파라미터를 갖는 표준 MELP 보코더의 경우 평균 PESQ 점수가 2.885에 대비, AE로 24개로 파라미터를 줄인 후 MELP 디코딩을 수행한 경우 2.828의 평균 PESQ 점수로 유사한 결과를 얻을 수 있었다.

감사의 글

이 연구는 방위사업청 및 국방과학연구소의 재원에 의해 설립된 신호정보 특화연구센터 사업의 지원을 받아 수행되었음.

자동차 잡음 제거를 위한 잡음이 혼합된 목표신호를 사용한 심층신경망의 성능

조 현 호, 권 오 욱
충북대학교

Performance of Deep Neural Networks Using Noise-Mixed Target Signals for Car Noise Reduction

Hyeon-Ho Jo, Oh-Wook Kwon
Chungbuk National University
jhh9601@cbnu.ac.kr, owkwon@cbnu.ac.kr

음성인식 성능이 향상됨에 따라서 다양한 환경에 음성인식이 적용되기 시작했다. 실제 환경에서 존재하는 잡음은 음성인식 성능을 떨어뜨리기 때문에 잡음을 제거하기 위한 음성강화 기술이 적용되어야 한다. 본 논문에서는 차량 내 음성인식을 위해서 잡음이 혼합된 목표신호를 사용한 DNN (Deep Neural Network) 기반 음성강화 방법을 제안한다. 기존의 DNN 기반 음성강화 방법은 깨끗한 음성을 목표신호로 사용했다. 기존 방법은 잡음 신호를 강하게 제거하여 잡음 제거 효과는 뛰어나지만 음성 신호 또한 왜곡시킨다. 본 논문에서는 잡음을 강하게 제거하는 효과를 줄이기 위해서 높은 SNR의 잡음을 목표 신호에 혼합하는 방법을 제안한다. 실험에 사용된 목표신호의 SNR은 20, 30, 40, 50, 60, 70dB이며 성능평가 지표는 PESQ 및 STOI를 사용했다. 실험 결과, 목표신호의 SNR이 50dB일 때 PESQ 2.89, STOI 0.81로 가장 높은 성능을 보였으며 제안된 방법이 기존 방법에 비해서 음성의 품질 및 명료도를 향상시키고 있음을 보여준다.

속삭임 음성인식을 위한 비병렬 코퍼스 환경에서의 유사 속삭임 음성 생성[10)]

김 민 식*, 정 용 원*, 이 성 주**, 김 형 순*
*부산대학교 전자공학과, **한국전자통신연구원

Pseudo Whisper Generation from Non-Parallel Corpora for Whispered Speech Recognition

Min Sik Kim*, Yongwon Jeong*, Sungjoo Lee**, Hyung Soon Kim*
*Department of Electronics Engineering, Pusan National University,
**Electronics and Telecommunications Research Institute
*{fire9945, jeongy, kimhs}@pusan.ac.kr, **lee1862@etri.re.kr

속삭임 음성과 같이, vocal effort 차이에 의한 발음 변이는 음성인식 성능에 많은 영향을 미친다. 다량의 속삭임 음성을 수집할 수 있다면 속삭임 음성의 발음 특성을 반영한 음향 모델을 통해 성능 개선이 가능하겠으나, 다량의 속삭임 음성을 확보하는 것에는 많은 시간과 노력이 필요하다. 이와 같이 다양한 환경의 훈련 데이터 수집이 제한적인 상황에서는 데이터 증강 방식이 자주 사용되는데, 이전 연구에서는 생성적 적대 신경망(Generative Adversarial Network, GAN) 기반의 방식으로 대규모의 유사 속삭임 음성을 생성하는 방식을 제안하였다. 하지만 그 방식은 동일한 화자와 동일한 문장에 대해 일반 음성과 속삭임 음성의 발화 쌍이 존재하는 병렬 코퍼스 환경에서만 적용이 가능하여 대다수의 비병렬 코퍼스 환경에서는 활용하지 못한다는 단점이 있다.

이에 본 연구는 생성적 적대 신경망의 생성자로 변분오토인코더(Variational Auto Encoder, VAE) 구조를 사용하여 비병렬 코퍼스 환경에서 유사 속삭임 음성을 생성하는 방식을 제안한다. 변분오토인코더는 vocal effort 종류에 독립적으로 훈련되어 vocal effort 정보를 제외한 음소 정보를 표현하는 잠재적 변수를 추정하는 인코더와, 잠재적 변수와 vocal effort 정보로부터 입력 음성을 복원하는 디코더로 구성되며, 변환 과정에서는 vocal effort 정보를 바꾸어 일반 음성을 속삭임 음성으로 변환할 수 있다. 실험에는 병렬 코퍼스인 wTIMIT corpus를 사용하였으나, 훈련과정에서 프레임 단위로 순서를 무작위로 섞어 병렬 코퍼스에 대한 어떠한 사전 정보 없이 훈련을 진행하였다.

실험 결과, 비병렬 코퍼스 환경에서 유사 속삭임 음성 생성을 통해 속삭임 음성의 인식 성능을 개선할 수 있었고, 이전 연구에서 제안한 방식의 속삭임 음성인식 성능과 비교해보니 비병렬 코퍼스 환경일지라도 본 연구에서 제안한 방식의 음성인식 성능이 더 뛰어남을 확인 할 수 있었다.

10) 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발 사업의 일환으로 수행하였음. [R0126-15-1117, 언어학습을 위한 자유발화형 음성대화처리 기술 개발]

구두 발표 II:
음성학

좌장: 박시균(군산대)

선·후행 음절에 따른 L2 학습자의 한국어 폐쇄음 지각

정 환 민, 홍 민 경, Chiin Ngaihmuan Ngaihte, Jeffrey Holliday
고려대학교 국어국문학과

The effect of stimulus length on the L2 perception of Korean stops

Hwanmin Jung, Minkyoung Hong, Chiin Ngaihmuan Ngaihte, Jeffrey Holliday
Korea University
jhm21v@gmail.com, gophang97@gmail.com, muanmok@gmail.com, holliday@korea.ac.kr

한국어 폐쇄음이 어두 위치에 있을 때 후행 모음의 음높이(f0)는 평음과 격음을 변별할 수 있는 주요한 단서가 되지만, 모국어가 성조언어인 한국어(L2) 학습자들조차도 어두 위치 폐쇄음을 잘 변별하지 못한다는 사실은 몇몇 연구(오재혁, 2013; Yang, 2017 등)을 통해 밝혀졌다. 그러나 이러한 결과를 도출한 연구들은 지각 실험을 위한 자극으로 고립(isolated) 음절 즉, CV 혹은 VCV로 구성된 자극을 사용하였는데, 이는 청자들이 주변 분절음과의 맥락을 통해 목표 분절음의 음향적 단서를 얻을 수 있다는 점을 고려하지 못한 것이다(Davidson, 2011). 따라서 선행 혹은 후행에 추가되는 분절음 정보가 L2학습자들의 목표음 지각에 어떠한 영향을 미치는지를 살펴볼 필요가 있다. 이에, 본 연구는 다음과 같이 가설을 설정 후 실험을 진행하였다. 어두 폐쇄음의 경우, 청자들은 두 번째 음절에서 제공되는 f0 정보로 인해 /Ca/보다 /Cata/에서 목표 음소를 더 정확하게 변별할 것이다. 반면 어중 폐쇄음의 경우, VCV 구성에 폐쇄음이 첫 음절 초성으로 추가되면, 이에 후행하는 모음의 f0가 오히려 청자들의 지각을 방해하므로 /taCa/ 보다 /aCa/에서 목표 음소를 더 정확하게 변별할 것이다. 본 연구의 실험을 위해, 한국어 원어민과 중국인 한국어 학습자 각 32명이 본 연구에 참여하여 /Ca/, /Cata/, /aCa/, /taCa/ 환경의 한국어 폐쇄음에 대한 음소 구별(discrimination) 및 확인(identification) 실험을 수행하였다. 통계분석(paired t-test) 결과, 두 집단 모두 /Ca/보다 /Cata/자극에서 어두 폐쇄음을 더 잘 변별하였으며 그 차이가 유의미한 반면, 어중 폐쇄음의 경우는 /taCa/보다 /aCa/에서 잘 변별하였고, 그 차이는 유의미하거나 같은 것으로 나타났다. 본 연구는 판단컨대, 한국어 L2학습자들의 폐쇄음 지각에 미치는 분절음의 맥락적 구성 즉, 목표 폐쇄음의 선·후행에 추가되는 음향 신호의 영향을 살펴본 최초의 연구라는 점에서 의의가 있다. 또한 본 연구의 결과는 추후 한국어 폐쇄음 지각 연구에 대한 한국어 학습자들의 지각 양상을 해석하는데 유용한 기준이 될 것이다.

참고문헌

[1] Davidson, L. (2011). Phonetic, Phonemic, and Phonological Factors in Cross-Language Discrimination of Phonotactic Contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 37(1), 270 - 282.

[2] Yang, Ximei. (2017). 중국인 한국어 학습자의 한국어 폐쇄음 산출과 지각에 대한 연구. 고려대학교 석사논문.

[3] 오재혁. (2013). 중국인 한국어 학습자의 발성 유형에 따른 한국어 폐쇄음의 변별 지각양상. 한국언어문화학, 10(1), 57-73.

성인 포먼트 측정에서의 최적 세팅 구현
- Praat software와 관련하여 -

박 지 연, 성 철 재[11)]
충남대학교 언어병리학과, 충남대학교 언어학과

Optimal Praat Setting in Adult Formant Measurement

Jiyeon Park, Cheoljae Seong
Speech-Language Pathology, Chungnam National University
bpn0525jy@gmail.com, cjseong49@gmail.com

성인을 대상으로 최적의 포먼트 분석이 가능하도록 자동화된 스크립트를 구현하였다. 최적의 포먼트 분석이란 Praat에서 포먼트 분석 시 설정하는 2가지 세팅 파라미터(최대 포먼트, 포먼트 개수)를 조합하여 측정된 포먼트의 편차가 최소일 때를 가리킨다(Escudero et al., 2009). LPC(Linear Predictive Coding)를 이용한 포먼트 추출 방식은 LPC 차수(order)를 포먼트 개수의 두 배로 지정하는데, 신뢰성을 높이기 위해서는 성별이나 모음의 종류에 따라 LPC 차수를 다르게 설정해야 한다(윤태진&강윤정, 2014). Praat 매뉴얼에서는 최대 포먼트 설정 값으로 남성 5,000Hz, 여성 5,500Hz, 측정개수는 5개를 권고한다. 그러나 이렇게 권고된 포먼트 세팅 설정이 한국어 모음에 대해서도 타당한지 검증이 필요하다. 본 연구에서는 박지연&성철재(2018)에서 개발된 아동 포먼트 분석 스크립트를 성인 포먼트 분석에서 최적의 포먼트 세팅을 찾도록 확장 구현하여 그 타당성을 검토하였다.

첫 번째 스크립트는 Praat의 권고 설정에 따라 최대 포먼트/포먼트 개수 설정을 남성 5,000/5, 여성 5,500/5로 모든 모음에 일괄 적용하여 포먼트를 검출하였다(fixed_5000n5_script). 두 번째 스크립트는 최대 포먼트는 고정하고, 포먼트 개수를 4부터 6까지 0.5단위로 증가하는 5가지 경우의 수를 조합하여 최적의 값을 찾도록 설계하였다(fixed_script). 세 번째 스크립트는 최대 포먼트를 100Hz씩 증가하는 20가지, 포먼트 개수 5가지 경우의 수를 조합하도록 설계하였다. 즉 하나의 모음 당 총 100가지 경우의 수에서 최적의 세팅 조건을 찾게 된다(linear_script). 네 번째 스크립트는 말소리 분석에 가장 합당한 청지각 로그스케일을 반영하기 위해 4분음 척도(quater-tone scale)을 적용하여 최대 포먼트의 경우 공비 $\sqrt[24]{2}$를 곱해가는 알고리듬으로 설계하였다(qtone_script).

스크립트 적용 결과는 각 모음별 포먼트 산점도로 확인하였는데 특히 여성의 경우

11) 교신저자

스크립트에 따라 측정된 포먼트 변이의 폭이 두드러지는 차이를 보였다. Praat의 권고 설정에 따른 fixed_5500n5_script는 [이, 에], fixed_script는 [이, 에, 어, 우]에서 넓은 편차 범위를 보이며 포먼트가 불안정하게 측정된 것을 알 수 있다(그림1, 그림2). 반면 linear_script 산점도에서는 7개 모음이 보다 뚜렷하게 구분되며 변이폭이 비교적 작은 것을 확인할 수 있다(그림3). qtone_script 산점도 역시 linear_script 산점도와 유사한 분포를 보이는 것을 확인하였다(그림4).

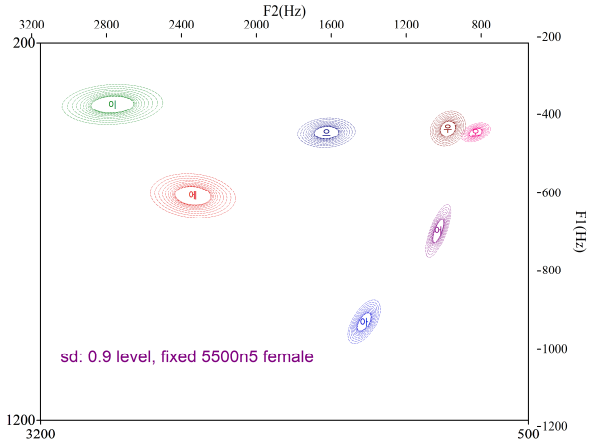


그림1. fixed_5500n5_script의 포먼트 분포(female)

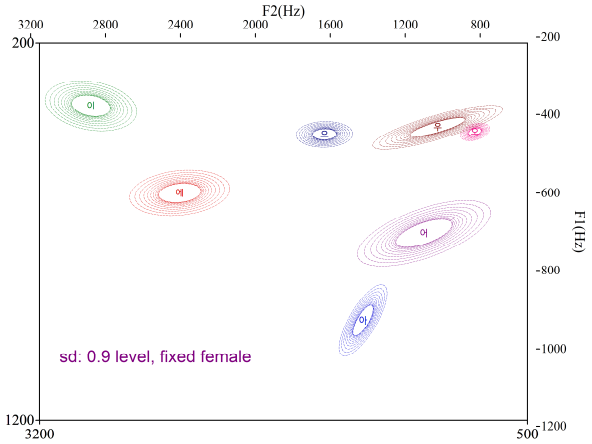


그림2. fixed_script의 포먼트 분포(female)

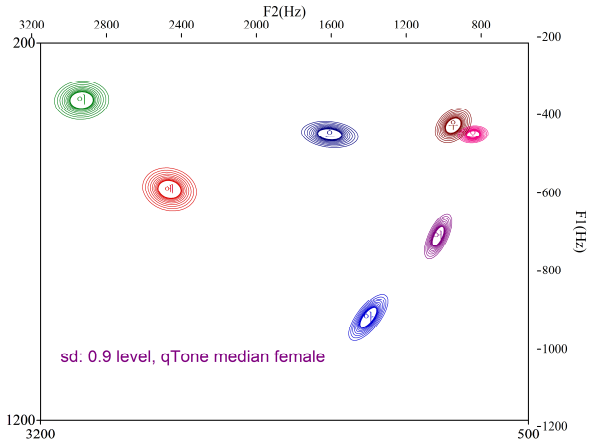


그림3. qtone_script의 포먼트 분포(female)

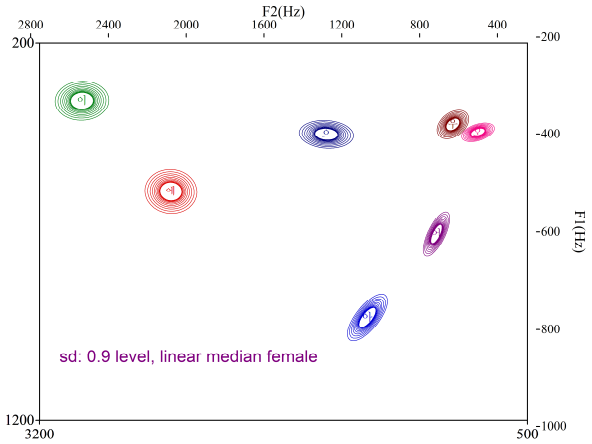


그림4. linear_script의 포먼트 분포(female)

linear_script, qtone_script에서 최적의 세팅으로 설정된 최대 포먼트와 포먼트 개수의 데이터 경향성을 살펴보면, 전설 모음 [이, 에]의 경우 권고 설정보다 최대 포먼트 값은 높게, 포먼트 개수의 값은 낮게 설정되었다. 반면 후설모음 [오, 우]의 경우, 권고 설정보다 최대 포먼트 값은 낮게, 포먼트 개수의 값은 높게 설정되는 것을 확인할 수 있었다. 최대포먼트와 포먼트 개수를 정확히 짝을 지어 언급하지는 않았으나 전설모음의 포먼트 개수를 낮게 설정한 양병곤(2019)도 비슷한 맥락임을 알 수 있다.

절의 통사·의미 유형에 따른 운율적 실현 양상

전 태 희, 신 지 영
고려대학교 국어국문학과

A Study on the Prosodic Realization of Cluases in Korean

Taehee Jeon, Jiyoung Shin
Department of Korean Language and Literature, Korea University
taeheejeon22@gmail.com, shinjy@korea.ac.kr

이 연구는 비종결 절 경계의 운율적 실현 양상을 살핌으로써 통사 단위와 운율 단위 사이의 상관관계를 밝히는 것을 목적으로 한다. 비종결 절 경계에서는 다양한 운율 단위가 실현될 수 있고, 앞뒤 절의 통사·의미 관계 또한 다양하게 나타날 수 있다. 따라서 이에 주목함으로써 절 사이의 통사·의미 거리에 대한 모국어 화자들의 인식이 운율적으로 어떻게 실현되는지를 살필 수 있을 것이다.

이 연구에서는 20대 한국어 모국어 화자 32명의 독백 발화 자료(14,120어절)를 이용하였다. 이 자료에는 억양구 단위의 운율 주석과 절의 통사·의미 유형에 대한 주석이 갖추어져 있다. 이를 토대로 절의 통사·의미 유형에 따른, 절 경계에서의 억양구 실현율, 절 경계 억양구의 후행 휴지 실현율, 절 경계 억양구의 후행 휴지 길이, 절 경계 억양구의 끝음절 길이를 살펴보았다.

절의 통사적 유형에 따라서는 통사적 독립성이 더 큰 대등 접속절, 종속 접속절, 내포절 순으로 억양구 경계 및 억양구 후행 휴지 실현율이 높았다. 억양구 후행 휴지와 끝음절 길이도 통사적 독립성이 큰 절들에서 더 길었다. 절의 의미적 유형에 따라서는 배경, 원인 등 의미적 독립성이 큰 유형의 절들이 목적, 방식/도구 등 의미적 독립성이 낮은 유형의 절들보다 더 큰 억양구 경계 및 억양구 후행 휴지 실현율을 보였다. 의미적 독립성이 큰 절은 억양구 후행 휴지와 끝음절 길이 역시 길었다. 이를 통해 통사·의미적 독립성이 큰 절일수록 그 경계에서 더 강한 운율적 경계성을 보임을 알 수 있었다.

서울말 2, 3, 4음절어의 운율패턴

변 희 경
국제교양대학

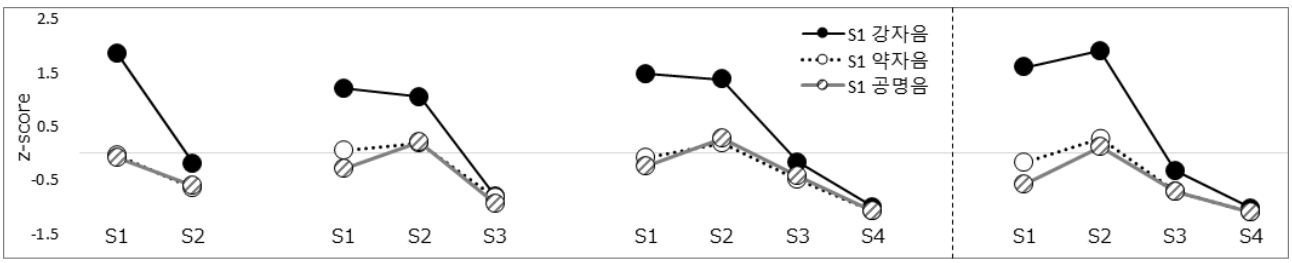
Pitch patterns in two-, three-, and four-syllable words of Seoul Korean

Hi-Gyung Byun
Akita International University
byun@aiu.ac.jp

서울말의 운율패턴은 3음절 이상에서는 많은 연구가 이루어졌으나 2음절에 관해서는 연구 자체가 적을 뿐만 아니라 선행연구 간의 결과가 일치하지 않아 전형적인 운율패턴이 무엇인지 분명하지가 않다. 한국어 교육에서 2음절어는 유기음화, 경음화 등 음운규칙을 설명하기 위해 사용되는 최소 음절수로 2음절어의 운율패턴을 밝히는 것은 학습자 교육에 있어서도 중요한 연구 과제라 할 수 있다.

분석시료는 서울말 모어 화자(대학생) 19~33명이 발화한 2음절(113단어), 3음절(16단어), 4음절(16단어)의 145단어이다. 둘째 음절의 초성 종류는 운율패턴에 관여하지 않는 것이 확인되었기 때문에 결과는 첫음절의 초성 종류에 따라 강자음(유기음, 경음, 마찰음), 약자음(연음), 공명음(모음, 비음)으로 나누어 제시한다.

그림 1. 2, 3, 4음절의 운율패턴
(왼쪽부터 2음절 단독발화, 3음절 단독발화, 4음절 단독발화, 4음절 문말(선행구 있음))



본 연구에서 확인된 서울말의 운율패턴은 그림 1 과 같다. 2음절은 첫음절의 초성 종류와 상관없이 HL, 3, 4음절은 첫음절의 초성이 강자음이면 HHL, HHLL, 약자음/공명음이면 LHL, LHLL이 기본이다. 3, 4음절의 경우, 첫음절 초성이 강자음일 때의 H는 약자음일 때의 H보다 현저하게 높다. 첫음절 초성이 약자음/공명음일 때의 L은 1음절, 3음절 째의 높이는 거의 같고, 4음절어의 4음절 째의 L은 3음절 째의 L 보다 현저하게 낮다. 그림1에서 보듯이 4음절어의 초성이 강자음일 때 선행구가 없으면 높게 시작하

고(세 번째), 선행구가 있으면 낮게 시작하는(네 번째) 경향이 있으나 3음절 이후의 음조와 비교한다면 첫음절과 둘째 음절 모두 HH로 보아야 할 것이다. 다만 첫음절과 둘째 음절의 HH, LH가 모두 실제 음성에서 존재하므로 청지각적으로는 첫음절이 강음절이어도 HHLL, LHLL 모두 자연스럽게 들릴 것이다.

위의 결과를 첫음절과 둘째 음절에 한정하여 음조의 발생 빈도를 보면 표1과 같다. S1-S2에 차이가 없으면 HH, 차이가 있으면 HL 또는 LH로 분류하였다. 1 Hz 기준은 1 Hz 라도 차이가 있으면 그에 따라 분류한 것이고 10 Hz 기준은 인지면에서 두 음이 다르다고 느끼는 참고치 10 Hz를 기준으로 분류한 것이다. 1 Hz 기준으로 분류한 음조는 HL가 압도적으로 많으나 첫음절이 강자음일 때를 제외하면 실제로는 HL의 음절간 차이가 작아서 청각적으로는 대부분 HH로 인지되기 때문에 서울말에는 높낮이가 없는 것처럼 느껴지는 것이다.

표 1. 첫음절과 둘째 음절의 음조형에 따른 발생 빈도(반올림 오차 있음)

2 음절 단독발화 (IP말)	1 Hz 기준			10 Hz 기준		
	HH	HL	LH	HH	HL	LH
강자음 n=502	1%	98%	2%	8%	91%	0%
약자음 n=868	2%	83%	15%	49%	49%	2%
공명음 n=987	3%	79%	18%	52%	44%	4%
합계 n=2357	2%	85%	13%	42%	56%	3%

3음절 단독발화 (IP말)	1 Hz 기준			10 Hz 기준		
	HH	HL	LH	HH	HL	LH
강자음 n=151	6%	60%	34%	81%	13%	5%
약자음 n=111	5%	23%	73%	76%	7%	17%
공명음 n=38	3%	3%	95%	42%	0%	58%
합계 n=300	5%	39%	56%	74%	9%	16%

4음절 단독발화 (IP말)	1 Hz 기준			10 Hz 기준		
	HH	HL	LH	HH	HL	LH
강자음 n=151	9%	58%	33%	76%	16%	8%
약자음 n=37	8%	19%	73%	68%	0%	32%
공명음 n=114	4%	5%	91%	40%	1%	59%
합계 n=302	7%	33%	60%	62%	8%	30%

4음절 문말 (IP말, 선행구 있음)	1 Hz 기준			10 Hz 기준		
	HH	HL	LH	HH	HL	LH
강자음 n=164	2%	27%	71%	55%	2%	42%
약자음 n=148	2%	15%	83%	46%	1%	53%
공명음 n=130	1%	5%	95%	26%	1%	73%
합계 n=442	2%	16%	82%	44%	2%	55%

구두 발표 II:
말장애 및 음성의학

좌장: 김성태(동신대)

2-7세 일반아동의 발성유형 별 말소리발달 연구

류 은 주, 하 지 완
대구대학교 재활과학과

Developmental study of speech production by phonation types in Korean Children

Eun-joo Ryu[a], Ji-wan Ha[b]
Department of Speech and Language Pathology, Graduate School of Rehabilitation Science, Daegu University[a]
Department of Speech Pathology, Daegu University[b]
hyej1234@hanmail.net, jw-ha@daegu.ac.kr

한국어의 발성유형은 다른 언어권과는 달리 다소 상이한 특성을 갖는다. 영어권의 경우 성대 진동의 유무에 따라 발성유형이 유성음과 무성음으로 구분되는 반면, 한국어의 경우 성대의 긴장성과 기식성의 유무에 따라 평음, 경음, 격음으로 구분된다. 발달과정을 살펴보면, 경음, 평음, 격음 순서로 세 발성유형 중 경음이 가장 우선적으로 산출된다. 또한 다른 언어권의 아동들에게서는 관찰되지 않은 경음화라는 음운오류패턴을 나타내는데, 이는 다른 언어체계와는 다른 한국어만의 특성이 반영된 것이다. 따라서 말소리 발달과정 중 나타나는 발성유형 오류에 대해 우리말 발성유형의 독특한 특성과 관련하여 알아볼 필요가 있다. 본 연구는 2세 6개월부터 7세 일반아동 537명을 대상으로 실시되었으며, 현재 개발 중인 UTAP2 단어수준 검사의 항목 중 평음, 경음, 격음이 포함된 단어들에 대한 대상자들의 정•오반응을 분석하여, 연령에 따른 발성유형별 정확도와 발달적 오류패턴의 특성을 살펴보았다. 보다 구체적인 결과를 담고자 분석은 1차, 2차로 나누어 실시하였다. 1차 분석 결과, 연령이 증가할수록 발성유형 별 정확도가 증가하였다. 반면, 발달적 오류패턴의 경우, 가장 일찍 발달하는 경음을 가장 늦게 발달하는 격음 또는 평음으로 대치하는 등 발달적 오류패턴으로 규정짓기 어려운 차이를 나타내었다. 이와 같은 결과는 한국어 마찰음의 상이한 특성과 지역적 방언 특성에 기인한 것으로 판단되었다. 따라서 이러한 고유특성에 대해 수정 및 보완하여 총 396명을 대상으로 2차 분석을 실시하였다. 2차 분석 결과, 발성유형별 정확도는 1차 분석과 마찬가지로 연령이 증가할수록 정확도 역시 향상되었다. 발달적인 오류패턴의 경우, 가장 늦게 발달하는 격음에 대해 경음 혹은 평음으로 대치하는 발달적인 오류패턴을 나타내었다. 이후 전체 아동을 대상으로 2회 이상 보인 음운오류패턴의 출현횟수 비율을 추가적으로 살펴보았고, 그 결과 전체 아동 중 10%이상의 아동이 평음의 경음화, 격음의 경음화 오류를 발달적인 오류패턴으로 나타내었다. 본 연구의 결과는 전국 규모의 아동들에 대한 발성유형별 발달 차이 및 발달적인 오류패턴에 대한 결과를 제시하였다. 또한 발달적인 오류패턴을 조사하는데 있어서 지역적 방언의 특성과 한국어 마찰음의 고유특성에 따른 차이에 대해 살펴보았다는 점에서 의의가 있다.

말·언어유창성 평가 시스템 SDUA 개발

심현섭* 이수복 박혜연* 곽효정* 신동춘** 고연정** 윤재민
이화여자대학교*, 우송대학교, 보리㈜**, 얄리㈜

Development of SDUA for speech-language fluency evaluation system

Hyunsub Sim*, Subok Lee, Hyeon Park*, Hyojung Kwak*, Dongchun Shin**, Yeonjeong Ko**, Jaemin Yun****
Ewha Womans University*, Woosong University, Bory Inc.**, Yally Inc.***
simhs@ewha.ac.kr, sblee@wsu.ac.kr, waterdrop88@hanmail.net, yd61581@naver.com, dcshin@bory.io, bory@bory.io, yally@yally.com

<요약>

본 연구는 아동 및 성인의 말·언어 유창성 평가를 위한 스마트폰, 태블릿 PC 기반 시스템인 SDUA(smart device for utterance analysis) 개발에 관한 것이다. SDUA의 주요기능은 영상녹화 및 업로드, 비유창성 분석, 문장 분석, 말속도 및 반응시간 분석, 문장완성도 분석 등의 기능을 제공한다. 본 시스템은 말더듬 아동과 부모의 자연스러운 놀이장면을 안드로이드폰이나 아이폰, 태플릿 PC 영상으로 촬영해서 서버에 업로드하면 전문가와 시스템이 협업하여 비유창성, 말·언어 특성을 분석한 후 그 결과를 확인할 수 있다. 즉, SDUA 시스템이 발화를 전사하고 말·언어 특성을 자동으로 분석할 수 있으며 언어재활전문가가 비유창성 분석과 말·언어 특성 분석내용을 검토하면 부모나 관련전문가들이 스마트 기기로 분석기록을 모니터링 할 수 있다.

본 시스템은 발화분석의 반자동화로 임상현장에서 임상가의 부담을 줄여줄 수 있고, 부모들에게 다양한 정보를 제공할 수 있어 임상적 의의가 있다.

<기술 개발 내용>

1. 시스템 구조

SDUA의 전체 시스템 구조도는 다음과 같다. 스마트폰이나 태블릿 PC 앱에서 영상을 녹화하고, 대용량 파일 전송 기능에 의해서 해당 영상을 서버로 전송하면, 전송된 영상에서 음성부분만 추출하여 발화구간을 감지하여 분석을 수행한다.

일반적으로 발화구간이란, 아동이나 부모가 자신이 말을 할 수 있는 기회에 말을 산출한 구간을 말하며, 보통 일상 대화에서는 1~2개의 문장으로 구성된다.

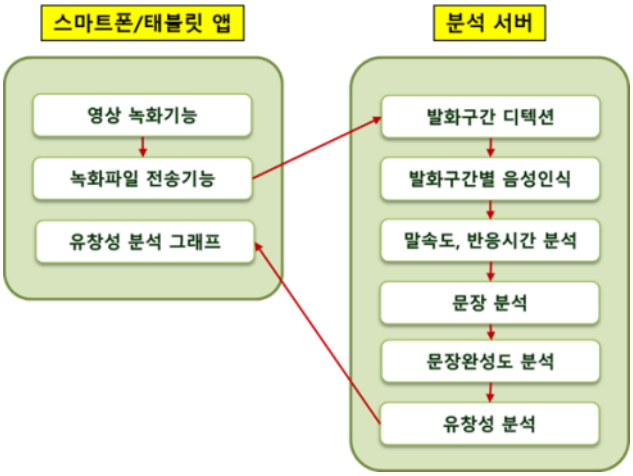


그림 1. 전체 시스템 구조도

Figure 1. Structure of Total System

1) SDUA에서 영상을 녹화하여 클라우드로 영상을 전송하면, 발화구간을 감지한 후, 발화구간별 음성인식을 수행하고, 말속도, 반응시간, 문장분석을 자동으로 실행한다.
2) 문장분석의 경우, 형태소분석을 수행해서 발화에 포함된 단어(명사, 용언 등)의 개수를 분석한다.
3) 문장완성도와 유창성 분석의 경우는 전문가가 참여하여 실제 평가를 수행한다.

그림2. 문장분석화면

Figure 2. Sentence analysis

그림3. 비유창성 분석 화면

Figure 3. Disfluency analysis

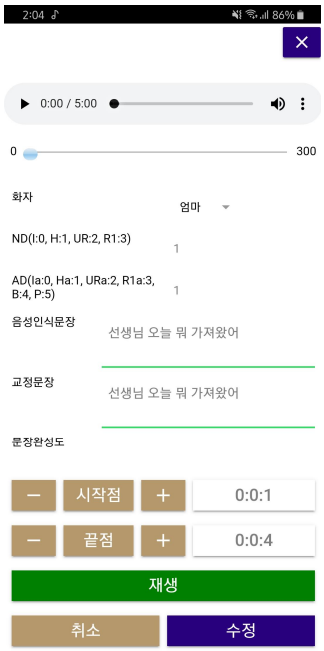


그림4. 비유창성 수정 화면

Figure 4. Disfluency modification

※ 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2-2018-1600-001-1).

SOVTE를 이용한 기능적 실성증의 음성치료 효과

김지성[1], 채혜림[1], 이동욱[2], 최성희[3]
대구가톨릭대학교 일반대학원[1], 충북대학교병원 이비인후과[2], 대구가톨릭대학교 언어청각치료학과[3]

Effect of Voice therapy Using SOVTE in patients with Functional aphonia

Graduate school of DaeguCatholic University[1], department of Otorhinolaryngology, Chungbuk national university hospital[2], Department of Audiology & Speech-Language Pathology, DaeguCatholic University[3]

slp2046@naver.com, mona@cu.ac.kr, dwlee@chungbuk.ac.kr, shgrace67@gmail.com,

기능적 실성증(functional ahponia)은 성대에 구조적인 문제는 없지만 속삭이는 음성이 산출되는 것으로, 후두내시경 검사에서는 정상적인 발성이 이루어지기에는 두 성대의 간격이 속삭이는 음성을 산출할 때와 같이 멀리 나타난다. 기능적 실성증의 음성치료는 기침하기나 흡기발성과 같이 성대접촉을 강화시키는 방법을 사용하는데 이러한 방법들은 오용되거나 남용될 경우 오히려 성대외상으로 이어질 가능성이 있어, 치료 시 주의가 요구된다.

SVOTE는 성도의 일부를 수축하는 것으로, 성도의 반폐쇄 시 나타나는 구강내압으로 인해 일반적인 음성산출보다 성대진동 시 접촉의 충격은 적지만 성대진동은 증가되는 효율적인 음성산출 방법이다(최성희, 2017; Titze, 2006). SOVTE는 근긴장성발성장애, 성대 결절과 같은 과다기능적 발성장애 뿐만 아니라 신경학적인 발성장애에서도 효과적인 치료방법으로 보고되어져 왔다(김지성 외, 2016, 채혜림, 2019). 이에 본 연구는 기능적 실성증의 새로운 치료방법으로써 SOVTE를 이용한 음성치료를 소개하고 그 효과를 알아보고자 한다. 2018년 충북대학교 이비인후과 전문의에 의해 기능적 실성증(functional dysphonia)로 진단 받은 환자 총 10명(남자 : 1명, 여자 : 9명, 평균연령 : 45.40세±12.19)을 대상으로 입술트릴 발성, 허밍, Lax Vox를 이용한 음성치료를 실시하였다. 평균 치료횟수는 1.42회이었으며 치료 효과를 측정하기 위하여 치료 전후 CSL(Computerized Speech Lab, Model 4150B, Kay-pentax)의 ADSV(Analysis of Dysphonia in Speech and VoiceTM)의 CPP를 측정하였다. 청지각적 평가는 GRBAS scale을 사용하였다. 6년 이상 경험이 있는 1급 언어재활사가 환자의 모음 /아/ 연장발성을 듣고 평

가하였다. GRBAS는 총 5점 척도로 평가하였다(0=문제 없음, 1=문제라고 보기 어려움, 2=문제 있음, 3=심한 문제, 4=아주 심한 문제). 수집된 자료의 분석은 IBM SPSS 22.0을 이용하여 Wilcoxon's signed-ranks test를 사용하였으며, 유의수준은 .05 수준에서 검정하였다. 연구결과, 음향학적 평가인 ADSV의 CPP 값이 치료 전 4.11dB에서 치료 후 13.43dB로 유의하게 증가하였으며($p<.01$), 청지각적 평가인 GRBAS 척도에서는 치료 전 Grade척도는 3.50에서 치료 후 0.58으로, Breathy척도는 치료 전 3.40에서 치료 후 0.60으로, Asthenic 척도는 치료 전 3.40에서 치료 후 0.30로 유의하게 감소하였다($p<.01$). 이러한결과는 SOVTE가 기능적 실성증의 음성치료에 효과적일뿐만 아니라 빠른 음성개선을 가져오는 효율적인 방법임을 시사한다. 음성치료에 사용된 SOVTE 중 Lax Vox는 빨대의 직경과 물의 깊이뿐만 아니라 개인적인 기류산출 능력에 따라 차이가 있을 수 있으므로 개인 맞춤 훈련을 제시하는 것이 필요하겠다.

참고문헌

[1] 김지성, 이동욱, 최철희, & 최성희. (2017). 후두마사지와 반폐쇄성도훈련이 과기능적 음성장애 환자의 음성개선에 미치는 효과. *Communication Sciences and Disorders*, *22*(4), 806-817.

[2] 채혜림. (2019). Lax Vox 음성치료가 파킨슨병 환자의 호흡 및 발성에 미치는 효과. *대구가톨릭대학교 일반대학원 석사학위논문.*

[3] 최성희(2017). 공연예술가를 위한 반폐쇄성도훈련: 악센트기법, 빨대 발성, Lax Vox. *한국공연예술발성연구회지*, *4,* 11-19.

[4] Titze, I. R. (2006). Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings. *Journal of Speech, Language, and Hearing Research, 49,* 448-459.

구두 발표 II:
음성공학

좌장: 박정식(한국외대)

양방향 학습 방법을 이용한 Flow 기반의 뉴럴 보코더

이 경 훈, 이 준 모, 이 협 우, 김 영 익, 조 훈 영
엔씨소프트

Flow based neural vocoder using bidirectional learning

Gyeong hoon Lee, Jun mo Lee, Hyeop woo Lee, Young ik Kim,
Hoon young Cho
NCSOFT
{ghlee0304, ljun4121, hwlee, youngik, hycho}@ncsoft.com

본 연구에서는 실시간 음성 합성을 위한 flow 기반의 가역(invertible) 뉴럴 보코더 모델에 대한 양방향 학습 방법을 제안한다. 기존의 학습 방법은 가우시안 확률 분포로부터 음성 데이터셋의 분포로 사상하는 가역 flow 보코더 모델을 구성하고, 입력 음성의 mel spectrogram을 조건값으로 하여 음의 로그우도(negative log-likelihood)를 최소화하도록 전방향으로 모델을 학습한다.

제안하는 방법은 기존 방향의 학습 방법과 역방향의 학습 방법을 혼합한 양방향 학습 방법이다. 기존 방법과 동일하게 모델을 학습한 뒤, 사전 분포로부터 샘플을 추출하여 학습된 모델의 역방향으로 음성을 생성하고 실제 음성과의 차이를 줄이도록 재학습함으로써, 전방향으로만 학습된 보코더보다 높은 품질의 음성을 생성할 수 있는 모델을 만들 수 있다.

실험에서는 한국어 DB에 대하여 L1, spectral convergence, log STFT magnitude와 같이 세 개의 loss를 이용하여 역방향 학습을 진행한다. 실험을 통하여 spectrogram을 기반으로 하는 각각의 loss가 합성된 음성의 품질에 미치는 영향을 분석하고, 이를 개별적으로 사용하는 것보다 가중합을 하여 모두 사용하는 loss가 가장 높은 성능을 가짐을 보인다. 실험 결과는 양방향으로의 학습 방법이 기존 학습 방법보다 고품질의 음성을 생성할 수 있음을 보여준다.

CNN에 기반한 방송 콘텐츠에서의 음악 신호 분리*

허 운 행[1], 김 혜 미[2], 권 오 욱[1]
[1]충북대학교, [2]ETRI

Music Signal Separation in Broadcast Contents Using CNN

Woon-Haeng Heo[1], Hey-Mi Kim[2], Oh-Wook Kwon[1]
[1]Chungbuk National University
[2]Electronics and Telecommunications Research Institute
{whheo, owkwon}@cbnu.ac.kr, miya0404@etri.re.kr

Abstract

We propose a CNN-based architecture to separate music in broadcast contents. We add multi-band block and dilated block to the DenseNet architecture for music source separation tasks. Experimental results show that the proposed architecture improves signal-to-distortion ratio compared with U-Net, Wave-U-Net and MDenseNet.

1. 서론

방송 콘텐츠의 음악 신호는 음성과 혼합된 부분이 많다. 이러한 혼합 신호에서 음악 신호를 분리하기 위해서 convolution neural network (CNN)에 기반한 새로운 DenseNet 구조를 제안한다.

2. 제안 방법

제안 방법의 구조는 <그림 1>과 같다. CNN에 기반한 multi-band block과 dilated dense block, down-sampling, up-sampling으로 구성되어 있다. 제안 구조는 이전의 MDenseNet 구조[1]에서 multi-band와 dilated block이 추가되었다. BN, ReLU, composite, dense block은 이전 MDenseNet 구조와 동일하게 적용하였다[1]. Composite는 BN-ReLU-Conv(3×3) 과정을 말하고[2], 출력 특징맵을 하나로 만들기 위함이다.

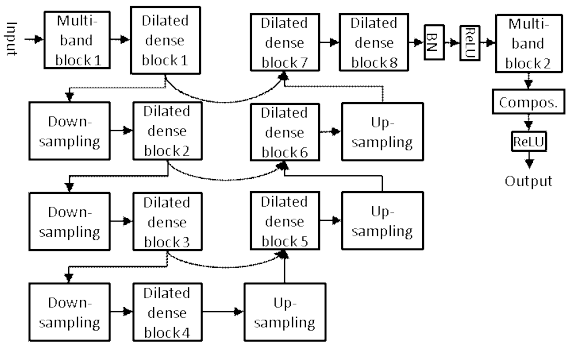


그림 1. 제안 방법 구조도

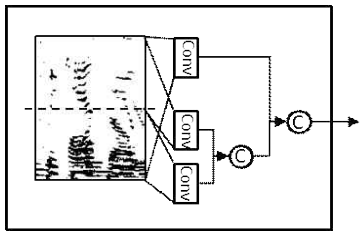


그림 2. Multi-band block

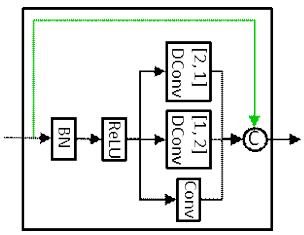


그림 3. Dilated block

* 본 연구는 문화체육관광부 및 한국저작권위원회의 2019년도 저작권기술개발사업의 연구결과로 수행되었음. [2019-micro-9500, 음악 및 동영상 모니터링을 위한 지능형 마이크로 식별 기술 개발]

음악과 음성이 혼합된 오디오 신호의 스펙트로그램이 입력으로 사용되고, 제안된 CNN 구조를 거쳐 혼합 스펙트로그램에서 음악 신호를 분리하기 위한 마스크가 최종적으로 출력된다.

2.1. Multi-band block

Multi-band block은 입력을 주파수 축으로 2등분하여 각 CNN 결과를 주파수 차원으로 연결하고, 전체에 대한 CNN 결과를 다시 특징맵 차원으로 연결하여 최종 출력 결과를 얻는다. 각 주파수 별로 서로 다른 convolution 필터를 적용하기 위해서 위와 같은 구조를 추가하였다.

2.2. Dilated block

Dilated block은 일반적인 convolution과 dilation rate를 [1, 2], [1, 2]로 주어 시간 축과 주파수 축에 따라 각각 dilated CNN을 적용하여 얻은 결과를 연결하여 최종 출력을 얻는다. 본 연구에서는 dilated block을 dense block 앞단에 추가하여 dilated dense block을 구성한다. 보통 영상에서는 dilation rate를 가로, 세로 같은 비율로 적용하지만 음성은 시간 축과 주파수 축이 서로 독립적인 특성을 가지므로 따로 적용하였다.

3. 실험 및 결과

실험 데이터베이스는 다양한 장르의 한국 방송 콘텐츠에서 얻은 음성과 배경음악으로 쓰이는 라이브러리 음악을 혼합하여 생성하였다. 학습 데이터는 0 dB에서 - 30 dB signal-to-noise ratio (SNR)로 혼합하여 약 5시간 30분 크기로 생성했고, 검증 데이터는 학습 데이터와 같은 방식으로 30분을 생성했다. 테스트 데이터는 0 dB, -10 dB SNR로 혼합하였고 각 6시간 크기로 생성했다.

스펙트로그램을 구하기 위해서 16 kHz의 샘플링 주파수, 윈도우 크기 1024, 중첩 크기 75%를 사용했다. 입력으로 patch 크기 128의 스펙트로그램을 사용했다.

실험 결과는 정답 신호와 분리된 음악신호 간의 signal-to-distortion ratio (SDR)를 측정하여 나타낸다.

<표 1>의 결과를 통해 제안된 구조의 음악 분리 성능이 가장 좋음을 알 수 있다.

4. 결론

방송 콘텐츠에서 음악 신호 분리를 위해 multi-band block과 dilated block을 추가한 dilated DenseNet을 제안하였다. 제안된 구조의 분리 성능이 기존 구조보다 더욱 성능을 개선시키는 것을 확인할 수 있었다.

표 1. 기존 및 제안 구조의 음악 분리 SDR 결과(dB)

분리 구조 \ 혼합 SNR	0 dB	-10 dB
U-Net	6.26	3.18
Wave-U-Net	6.67	3.48
MDenseNet	6.79	3.80
Dilated DenseNet (proposed)	7.43	4.32

참고문헌

[1] Takahashi, N., & Mitsufuji, Y. (2017). Multi-scale multi-band DenseNets for audio source separation. *in Proc. WASPAA*, pp. 261 - 265.

[2] Huang, G., Liu, Z., van der Maaten, L., & Weinberger.K. Q. (2017). Densely connected convolutional networks. I*n Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700 - 4708.

나이 정보를 활용한 화자 인증 시스템

김주호, 허희수, 정지원, 심혜진, 유하진
서울시립대학교 컴퓨터과학과

Speaker verification system using age information

Ju-ho Kim, Hee-soo Heo, Jee-weon Jung,
Hye-jin Shim, and Ha-jin Yu
School of Computer Science, University of Seoul
wngh1187@naver.com, zhasgone@naver.com, jeewon.leo.jung@gmail.com,
shimhz6.6@gmail.com, hjyu@uos.ac.kr

본 논문은 화자 인증에서 Deep Neural Network(DNN)의 일반화 성능을 향상시키기 위해, 화자 정보 이외의 정보를 활용하여 DNN을 학습시키는 방법을 제안한다. 구체적으로, DNN에 Multi-task learning 기법을 적용하여 화자 정보와 나이 정보를 동시에 학습시킨다. 나이 정보를 활용한 DNN 학습 과정은 다음과 같다. Mel-filterbank를 이용하여 원 음성 신호로부터 추출한 음향 특징을 DNN의 입력값으로 사용한다. 마지막 은닉층의 출력은 각각의 태스크의 출력층과 연결된다. 화자 식별의 손실함수는 categorical cross entropy(CCE)를 활용하고, 나이 추정의 손실함수는 mean abstract error(MAE)를 활용한다. 각각의 초기 손실 가중치를 1로 지정하고, 1 epoch이 끝날 때마다 CCE 가중치에 1.1을 곱하고, MAE 가중치에 1.1을 나눈다. 나이 정보를 화자 식별 학습 초기에 활용함으로써, DNN의 일반화 성능 향상을 기대하였다.

화자 인증 실험은 RSR2015 데이터를 사용하여 수행하였다. 실험에 사용할 화자는 다음과 같이 나누었다. 총 300명 화자 중 194명을 화자식별기의 학습으로 사용하였고, 53명을 학습이 끝날 때마다 화자 인증시스템의 검증으로 사용하였다. 나머지 53명 화자는 화자 인증시스템의 최종 평가를 위해 사용되었다. 등록 발성과 평가 발성을 입력하여 마지막 은닉층 출력 간의 코사인 유사도를 기준으로 성능을 평가하였다.

화자 정보만 사용하여 학습한 화자 인증 시스템은 동일 오류율 기준 6.51%의 성능을 보였다. 반면에 제안한 방법을 적용한 화자 인증 시스템은 동일 오류율 기준 4.38%의 성능을 보였다. 이 결과로부터 화자 정보와 나이 정보를 함께 활용하여 화자 인증 시스템을 구성하는 것이 DNN의 일반화 성능 향상에 효과적임을 확인하였다.

이 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(10076583)

Transformer 기반 Seq2seq 모델 및 WaveGlow 보코더를 활용한 음성합성 시스템[13)]

최 연 주, 서 영 주, 정 영 문, 최 희 진, 김 회 린
KAIST 전기및전자공학부

A Speech Synthesis System using Transformer-based Seq2seq Model and WaveGlow Vocoder

Yeunju Choi, Youngjoo Suh, Youngmoon Jung, Heejin Choi, Hoirin Kim
School of Electrical Engineering, KAIST
{wkadldppdy, yjsuh, dudans, change, hoirkim}@kaist.ac.kr

본 연구에서는 Transformer 기반 seq2seq 모델과 WaveGlow 보코더를 결합한 음성합성 시스템을 제안하였다. WaveNet 보코더와 어텐션 메커니즘을 활용한 RNN 기반 seq2seq 모델을 결합한 음성합성 시스템인 Tacotron2의 합성음 음질은 이미 사람 음성 수준에 매우 근접하였다. 하지만 음성 생성 과정에서 WaveNet 보코더가 샘플을 순차적으로 계산하기 때문에 상용화되기에는 음성 생성 속도가 너무 느리다는 한계점이 있다. 또한 RNN 기반 seq2seq 모델은 훈련과 합성 과정 모두에서 순차적인 계산을 하기 때문에 효율이 떨어진다는 한계점이 있다.

본 연구에서는 우선 전체 속도 저하의 가장 큰 요인인 autoregressive WaveNet 보코더를 병렬 계산이 가능한 flow 기반 생성 모델인 WaveGlow 보코더로 대체하여서 음성 생성 속도를 향상시켰다. 또한 기존의 Transformer TTS 연구(Li et al., 2018)에서와 마찬가지로 seq2seq 모델의 RNN 기반 구조를 어텐션 메커니즘과 feed-forward 네트워크만을 활용한 Transformer 기반의 구조로 대체하여 훈련에서의 병렬 계산이 가능하도록 하였다.

19시간 분량의 단일 여성 화자 한국어 DB를 이용하여 RNN 및 Transformer 기반의 seq2seq 모델과 WaveNet 및 WaveGlow 보코더를 각각 훈련시켰다. 그 후, seq2seq 모델과 보코더를 조합한 총 네 가지 음성합성 시스템의 음질과 훈련 및 생성 속도를 비교하였다. WaveGlow 보코더를 사용하였을 때, WaveNet 보코더를 사용하였을 때와 비교하여 음질은 유지하면서 음성 생성 속도가 빨라졌다. Seq2seq 모델 역시 Transformer 기반의 구조를 사용하였을 때, RNN 기반 구조를 사용하였을 때와 비교하여 음질은 유지하면서 훈련 속도가 빨라짐을 확인할 수 있었다.

13) 본 연구는 산업통상자원부의 산업기술혁신사업으로부터 지원을 받아 수행된 연구임 (No. 10080667, 음원 다양화를 통하여 로봇의 감정 및 개성을 표현할 수 있는 대화음성합성 원천기술 개발).

특강 III

좌장: 최성희(대구가톨릭대)

음성공연의 이론과 실제
(박문숙, University of Arkansas)

The theory and practice of the performing voice

Dr. Moon-Sook Park
Associate Professor of Music/Voice
Department of Music
J. William Fulbright College of Arts and Sciences
University of Arkansas

Classical Singing Masterclass

성악 매스터 클래스란?

"성악 매스터 클래스란?"

개인 레슨식의 성악 지도를 통해 기술적, 예술적, 가창적 측면에서 개선되어 가는 과정을 보여주는 지도방법

실기 강의 목적

- ✓ **곡 해석**
- ✓ **곡 Phrase 처리**
- ✓ **명확한 가사 전달**
- ✓ **소리의 강약 조절(dynamic abilities)**
- ✓ **예술성이 첨가된 가창**
- ✓ **노래와 반주부의 음악적 관련성 이해**

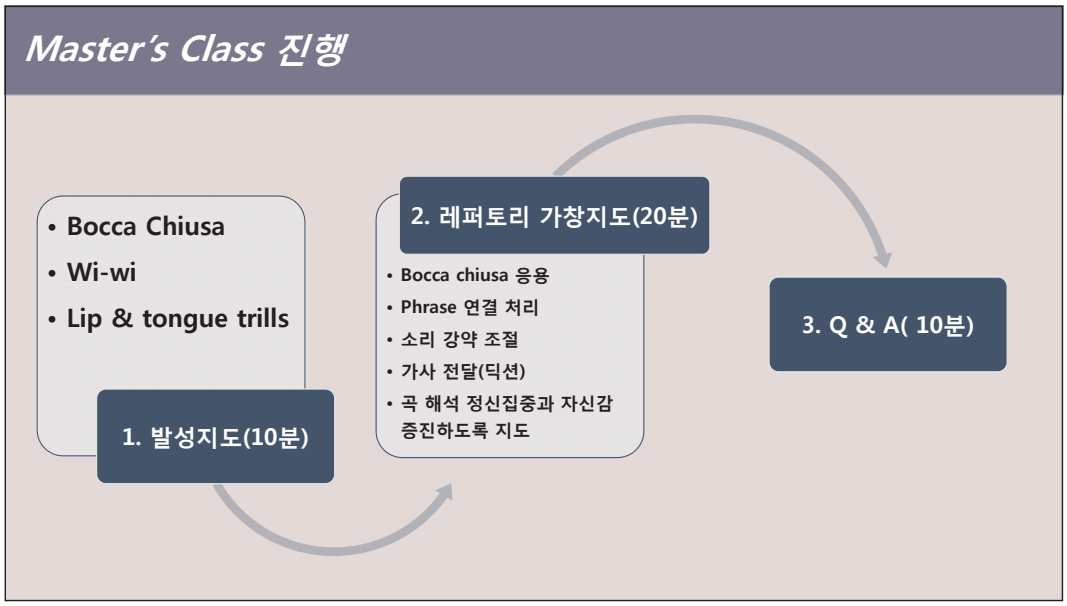
실기 내용

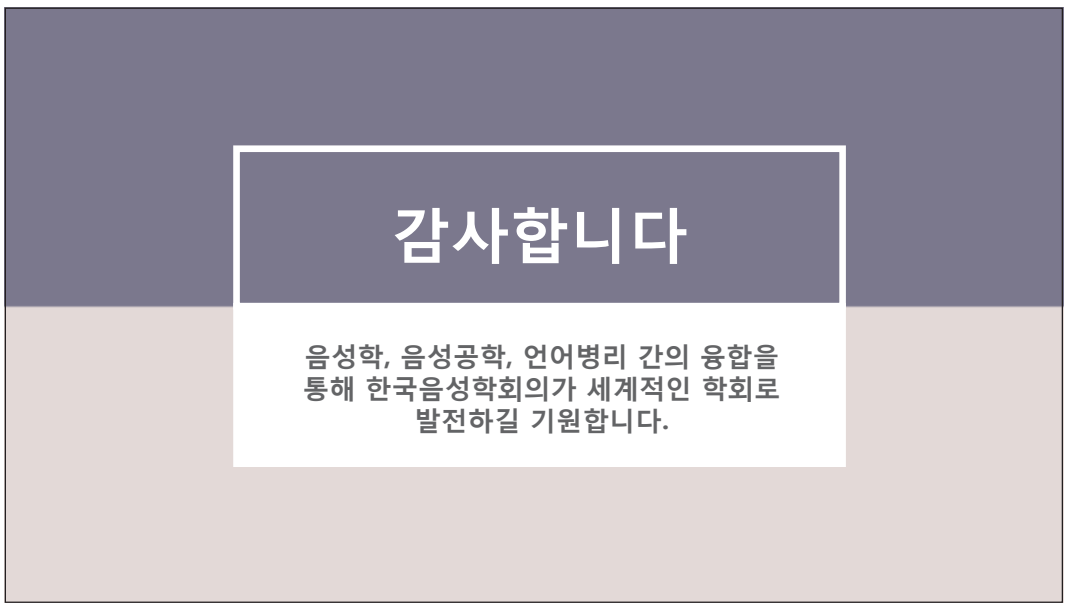
개인의 최적의 발성과 공명을 위해 적절한 호흡 조절 및 발성을 체감할 수 있도록 여러가지 발성 테크닉 및 호흡과 정신집중 훈련 방법, 균형적인 몸 자세 그리고 도움이 되는 여러 몸 동작들을 제시하면서 참가자의 최적의 가창과 자신감을 느낄 수 있도록 끌어냄

기대효과

1. 여러가지 성악적 기술과 노래 훈련 과정을 통해 공명, 발성 그리고 호흡에 대한 다른 각도의 이해도를 얻게 됨

2. 비록 짧은 시간의 매스터 클래스 강의지만 참가자 뿐만 아니라 청중들도 성악 훈련과 소리 공명, 발성과 호흡에 관련하여 다양한 성악 훈련 과정을 경험할 수 있는 긍정적인 방법이라 할 수 있는 모두에게 유익한 경험이 될 수 있다고 기대됨


Master's Class 진행
• Bocca Chiusa
• Wi-wi
• Lip & tongue trills
1. 발성지도(10분)
2. 레퍼토리 가창지도(20분)
• Bocca chiusa 응용
• Phrase 연결 처리
• 소리 강약 조절
• 가사 전달(딕션)
• 곡 해석 정신집중과 자신감 증진하도록 지도
3. Q & A(10분)


감사합니다
음성학, 음성공학, 언어병리 간의 융합을 통해 한국음성학회의가 세계적인 학회로 발전하길 기원합니다.

NCSOFT

AI R&D

AI에 다가서다

TAPPING INTO AI

채용 문의

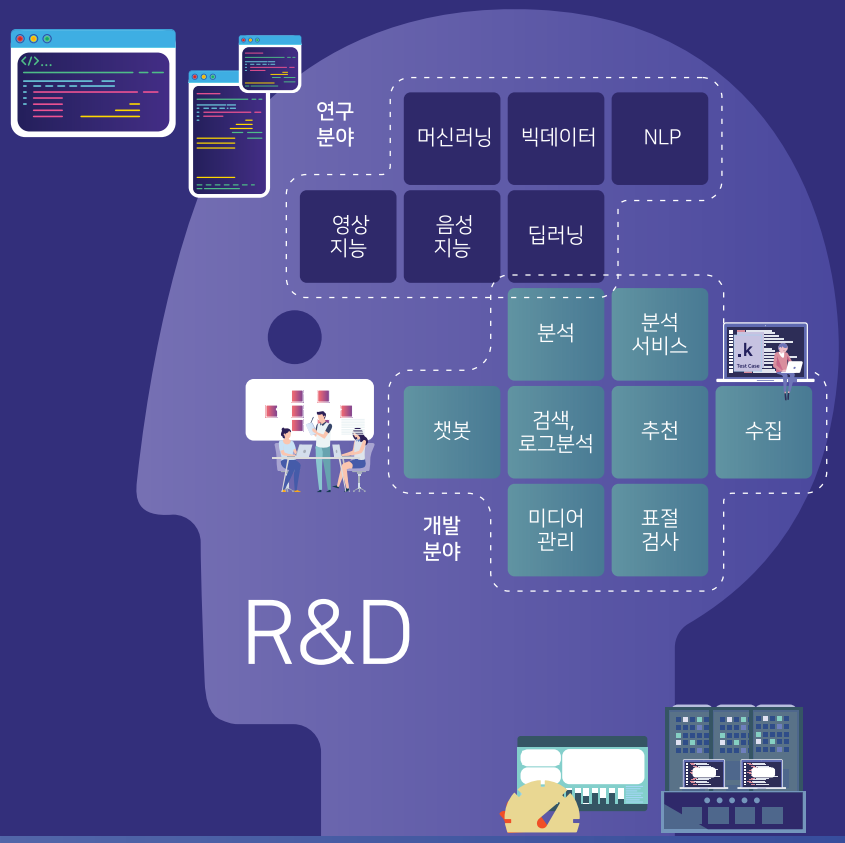
이메일 ai@ncsoft.com **전화** 02-6201-0680

연구개발 분야 및 채용에 관한 문의는 언제든지 환영합니다.

http://about.ncsoft.com

20년 토종 인공지능기업
코난테크놀로지

'KONAN'은 1999년에 설립된 기술혁신 기업으로, 자연어처리와 인공지능, 빅데이터와 비정형데이터를 포함하는 초대용량 문서처리기술을 통해 고객의 가치에 기여하는 인공지능 및 빅데이터 소프트웨어 전문기업입니다.
Korea Natural Language Analysis 의 의미를 담고 있습니다.



Why Konan Technology

원천기술에 대한 지속적 R&D
자연어처리 및 대용량 문서처리기술 기반으로 인공지능, 빅데이터 분석기술 등에 대해 지속적인 R&D 투자 시행, 지능형 검색, 빅데이터 분석, 인공지능 응용 솔루션까지 지능정보산업을 대응하는 제품 풀 라인업 확보

다양한 시스템과 유연하게 연동
여러 채널의 다양한 소스로부터 데이터 수집을 지원하고 다양한 기업 내 시스템과 상호 연동 지원(포털, CMS, ERP, 그룹웨어, CRM, VOC, BI 등)

3,000 + 레퍼런스
3,000회 이상의 프로젝트 경험이 축적된 제품과 전문인력을 보유하고 있으며 공공, 교육, 금융, ICT, 쇼핑, 기업 등 다양한 산업분야의 레퍼런스와 Domain Knowledge 보유

2,500 + 고객유지보수 체계 운영
품질관리시스템 운영을 통한 고품질 서비스를 제공하며 전담 유지보수 조직과 체계적인 이슈관리시스템 운영으로 24시간 신속하게 응대

제품/서비스

AI	Collect	Analyze	Search	Manage
Konan Bot	Konan Social Crawler	Konan Analytics	Konan Search	Konan DigitalArc
Konan Smart Match	Konan Web Crawler	Konan Log Analytics	Konan MemeChecker	Konan MediaArc
		pulsek	Konan Easyfinder	

구축 사례

- 차세대 인공지능 지향 검색 - 서울시 통합검색 분류 정비 및 시스템 통합 (I·SEOUL·U)
- HR 업무지원 인공지능 챗봇 구축 - SK하이닉스 사내 메신저 內 탑재 (SK hynix)
- 개인화 맞춤 인공지능 추천 • 검색 시스템 - 사람인 인공지능 아바타 서치 (saramin)
- 온라인 미디어 모니터링 및 분석 - 2017 제 19대 대통령 선거 MBC 개표방송 (MBC)
- 내부 직원 업무지원용 챗봇 구축 - 한국토지주택공사 LH in Helper (LH 한국토지주택공사)
- 대민용 챗봇 구축 - 서울특별시 우리마을가게 상권분석 챗봇 (서울특별시)
- 사회 • 경제 등 화제 이슈 돌출 및 분석 - 『빅데이터를 활용한 환경 스캐닝보고서』 (NIA 한국정보화진흥원)
- 표절 검사 시스템 구축 - 건국대학교 자기소개서 표절 검사 시스템 (KU KONKUK UNIVERSITY)

KONAN TECHNOLOGY

(주)코난테크놀로지 우)06627 서울특별시 서초구 강남대로 327 대룡서초타워 6,7층
TEL 02-3469 8555 / FAX 02-3469-8599,8798 www.konantech.com

제품문의 02-3469-8555
기술지원 02-3469-8588

Clova Friends

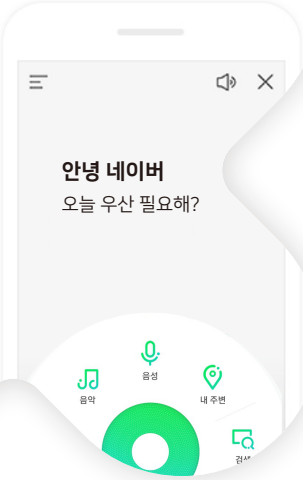
인공지능
스피커

Papago

음성통역
및 번역

네이버 그린닷

음성검색
서비스



네이버 지도

네비게이션
음성 검색
및 음성 안내

NAVER | **Clova** 클로바의 인공지능 음성기술과 함께
당신의 일상이 더욱 편리해집니다

『2019년 한국음성학회 봄 학술대회』의 개최를 축하드립니다.

(사) 한국언어재활사협회는 한국 음성학의 발전을 위해 노력해 오신 '한국음성학회'의 봄 학술대회의 개최를 진심으로 축하드립니다. 음성의 이론적 연구와 임상 현장에서의 실제적 응용과 관련된 많은 연구를 통해 발전을 거듭해 온 음성학회가 앞으로도 지속적 교류와 협력을 통해 한국의 음성학과 언어병리학의 위상을 더욱 높여주시기 바랍니다.

사단법인 **한국언어재활사협회**

문의 02) 790-2726 http://www.kslp.org

음성기술
인공지능 전문기업 셀바스 AI가 이끌어가겠습니다

셀바스 AI는 음성인식 및 음성합성 등 핵심 음성기술 분야의 독보적 원천 기술을 보유한 **국내 대표 음성기술 기업**입니다. 셀바스 AI의 인공지능 대표브랜드인 **Selvy**를 음성기술에 접목하여 꾸준한 연구개발과 음성인식 성능 향상에 힘쓰고 있습니다.

selvas AI

회사 홈페이지 www.selvasai.com **음성기술 홈페이지** speech.selvasai.com
Tel. 02 6190 7500 **E-mail.** support@selvasai.com

CALL FOR PAPERS

Seoul International Conference on Speech Sciences 2019 (SICSS 2019)
15-16 November 2019, Seoul National University, Seoul, Korea
https://sicss2019.wordpress.com/; http://www.sicss.or.kr/
"Speech Sciences Moving Forward: New Methods, Data and Populations"

Date: 15 (Fri) - 16 (Sat) November 2019
Venue: Seoul National University, Seoul, Korea
Abstract Submission Deadline: 31 August 2019

The Korean Society of Speech Sciences (KSSS) will host the Seoul International Conference on Speech Sciences 2019 (SICSS 2019) between 15 and 16 November 2019 at Seoul National University, Seoul, Korea. The abstract submission for SICSS 2019 is now open for authors. For more information, please visit https://sicss2019.wordpress.com/ or http://www.sicss.or.kr/.

Keynote Speakers

Phonetics

Allard Jongman, The University of Kansas, USA

- Keynote Speech: Enhancing intelligibility: Hyperarticulation, visual, and lexical contributions
- Tutorial Lecture: The role of training in the formation of new sound categories

Speech Disorders

Keith R. Kluender, Purdue University, USA

- Keynote Speech: Long-standing problems in speech perception dissolve within an information-theoretic perspective
- Tutorial Lecture: Virtues of (co)variance for perceptual learning

Speech Technology

Helen Meng, The Chinese University of Hong Kong, Hong Kong

- Keynote Speech: Speech AI for learning and wellbeing
- Tutorial Lecture: Research advancement in Text-to-Speech synthesis

The topics of the conference include, but are not limited to:

1. Phonetics (Theoretical & Experimental)
2. Applied Phonetics
3. Speech Perception & Production
4. Experimental Phonology
5. Speech Disorders
6. Hearing Disorders
7. Speech Synthesis & Recognition
8. Speech Coding
9. Speech Analysis & Processing
10. Pronunciation Teaching & Learning
11. Other Areas in Phonetics and Speech Sciences

Presentation Types
1. Paper presentation 20 minutes (15 minute talk + 5 minute discussion)
2. Poster presentation 60 minutes

Important Dates
1. Abstract submission opens: Wednesday, 1 May 2019
2. Abstract submission deadline: Saturday, 31 August 2019
3. Notification to be sent within two weeks of receiving the abstract.
4. Early bird registration opens: Monday, 16 September 2019
5. Author registration deadline: Monday, 30 September 2019
6. Early bird registration deadline: Thursday, 31 October 2019
7. Conference: 15-16 November 2019

Abstract Submission

Abstracts must be written in English and must not exceed two pages in length (references and graphics included). There are two abstract submission formats available: a one-page abstract (with no sections other than references) and a two-page abstract (with section numbers and headers). You may choose either option depending on how you would like to prepare your abstract. You may download the abstract templates from https://sicss2019.wordpress.com/ or http://www.sicss.or.kr/. Submissions should be made online at:

https://easychair.org/conferences/?conf=sicss2019

Please make sure that the completed file of the abstract should be converted to PDF format before submission. The call for papers will remain open until 31 August 2019. Papers will be accepted on a rolling basis, with authors of accepted abstracts notified within two weeks of submission. Full papers are not required for this conference. The authors of conference papers MUST register by 30 September 2019. If the registration has not been done by the deadline, the acceptance of the paper will be cancelled.

Contact
sicss2019@gmail.com