

Proceedings
of the 2024 Spring Conference
of the Korean Society of Speech Sciences

2024. 05. 18.

Hannam University



The Korean Society of Speech Sciences

2024 한국음성학회 봄 학술대회 발표 논문집



일시: 2024년 5월 18일(토)

장소: 한남대학교 56주년기념관 중회의장

주최: 사단법인 한국음성학회

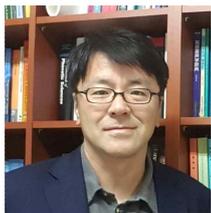
주관: 사단법인 한국음성학회, 한남대학교

후원: (주)리드스피커코리아, (주)셀바스AI, (사)한국언어재활사협회

한국음성학회
The Korean Society of Speech Sciences

2024년 한국음성학회 봄 학술대회

모 시 는 글



한국음성학회는 5월 18일 대전의 한남대학교에서 2024년 봄 학술대회를 개최하게 되었습니다. 비대면으로 학술대회를 개최할 수 밖에 없었던 어려운 시기를 극복하고 작년에 이어 대면으로 열리게 된 봄 학술대회에서 반갑게 여러분들을 직접 뵈 수 있어서 기쁘게 생각합니다.

이번 학술대회는 우리 학회의 고유 연구 영역인 음성학, 음성공학, 말장애 분야의 최신 연구 동향을 공유하고, 함께 더 나은 미래를 모색하는 소중한 자리입니다. 최근 다양한 생성형 AI 기술의 발전으로 인해 음성 자료에 대한 연구와 활용 및 교육 방법에도 많은 변화가 있었습니다. 이러한 변화에 발맞추어 이번 학술대회의 주제를 '생성형 AI와 음성과학'으로 정하고 관련된 두 편의 전문가 초청 강연을 모시게 되었으며 세 세션의 병렬 구두 발표 일정에서 30편의 연구가, 그리고 14편의 흥미 있는 포스터 발표가 소개될 것입니다. 흔쾌히 초청 강연을 수락해 주신 분들과 소중한 연구 성과를 발표해 주시는 회원 여러분께 진심으로 감사의 말씀을 전합니다. 또한, 이번 학술대회를 위해 성심성의를 다해 준비해 주신 박상희 학술위원장님을 비롯한 여러 준비위원님, 한남대학교 56주년 기념관을 장소로 마련해주시고 여러 가지 어려운 문제를 해결해 주신 운영선 부회장님, 그리고 회장단 및 모든 관계자분께도 깊은 감사의 인사를 드립니다. 여러분의 노고와 열정 덕분에 이 행사를 준비할 수 있었습니다. 마지막으로 늘 변함없이 후원해 주고 계시는 (사)한국언어재활사협회, (주)리드스피커코리아, (주)셀바스AI 에도 특별한 감사를 표합니다.

이번 학술대회에 참여하신 분들을 진심으로 환영합니다. 함께 풍부한 지식과 유익한 경험을 나누는 이 특별한 시간을 즐겨주시길 바랍니다. 감사합니다.

2024년 5월

한국음성학회 회장 **장 태 업**

2024년 한국음성학회 봄 학술대회

조직위원회

학술대회장 회장 장태엽 (한국외대)

조직위원회 위원장 정현성 (수석부회장, 한국교원대)

위 원 김재옥 (강남대)
 박상희 (대구사이버대)
 신동진 (전주대)
 오은혜 (건국대)
 윤영선 (한남대)
 윤원희 (계명대)
 이주경 (서울시립대)

학술위원회 위원장 박상희 (대구사이버대)

위 원 공은정 (한국항공대)
 김유경 (순천제일대)
 김지연 (우송대)
 박기영 (한국전자통신연구원)
 오재혁 (건국대)
 유하진 (서울시립대)
 이현정 (인천대)

편집위원회 위원장 김유경 (순천제일대)

위 원 배미영 (한국외대)



한국음성학회
The Korean Society of Speech Sciences

2024 한국음성학회 봄 학술대회 일정표

1) 전체 일정표

일정	발 표 및 내 용		
09:00~14:00	등록		
09:40~09:50	개회식(중회의장) 개회식(사회: 신동진, 전주대학교)		
09:50~10:40	특강 1(중회의장) (사회: 윤영선, 한남대) “LLM (Large Language Model) 기술 현황 및 응용” 구명완(서강대 컴퓨터공학과) [특강자료로]		
10:40~10:50	휴식		
10:50~12:30	구두 발표 I		
	분과 1(중회의장) 음성학 및 음운론 PH01, PH02, PH03, PH04 (4편)	분과 2(소회의실 1) 음성공학 SE01, SE02, SE03, SE04 (4편)	분과 3(소회의실 2) 말장애 및 음성의학 SD01, SD02, SD03, SD04 (4편)
12:30~13:10	점심 식사		
13:10~14:00	포스터 발표(로비)		
14:00~14:50	특강 2(중회의장) (사회: 이주경, 서울시립대) “ChatGPT를 활용한 문화와 사회성 연구” 김지혜(서강대 사회학과) [특강자료로]		
14:50~15:00	휴식		
15:00~16:15	구두 발표 II		
	분과 4(중회의장) 음성학 및 음운론 PH05, PH06, PH07 (3편)	분과 5(소회의실 1) 말장애 및 음성의학 SD05, SD06, SD07 (3편)	분과 6(소회의실 2) 말장애 및 음성의학 SD08, SD09, SD10 (3편)
16:15~16:25	휴식		
16:25~17:40	구두 발표 III		
	분과 7(중회의장) 음성학 및 음운론 PH08, PH09, PH10 (3편)	분과 8(소회의실 1) 음성공학 SE05, SE06, SE07 (3편)	분과 9(소회의실 2) 말장애 및 음성의학 SD11, SD12, SD13 (3편)
17:40~18:00	스펙트로그램 reading(중회의장) 우수 발표 시상 및 폐회식 (사회: 박상희, 대구사이버대학교)		

2) 발표 내용

● 구두 발표 I

<분과 1> 음성학 및 음운론	좌장: 공은정(항공대) 토론자: 윤원희(계명대)			
	시간	구분	제목	발표자
	10:50~11:15	PH01	The initiator effect in conversational interactions and its association with children's language outcome [논문으로]	Chai Jun Ho, Jongmin Jung, Eon-Suk Ko (Chosun Univ.)
	11:15~11:40	PH02	Developmental Variations in English Vowel Production: A Proficiency-Based Study of Korean ESL Learners [논문으로]	Mi Sun Park, Joo-Kyeong Lee (University of Seoul)
	11:40~12:05	PH03	Effects of Age, Gender, and Language Ability on the Acoustic Fluency and Pronunciation Judgements in English Utterances by Korean Speakers [논문으로]	Hyunsong Chung (Korea National Univ. of Education)
12:05~12:30	PH04	A study on the L2 phonetic imitation: Korean learners' case of English stress assignment [논문으로]	Mi-Young Bae, Tae-Yeoub Jang (Hankuk Univ. of Foreign Studies)	
<분과 2> 음성 공학	좌장: 김선만(광기술원) 토론자: 이경남(LG U+)			
	시간	구분	제목	발표자
	10:50~11:15	SE01	한국어 음성인식을 위한 종단형 음성인식 모델의 실험적 비교 [논문으로]	오창한, 박기영, 강점자, 송화전, 최우용(ETRI)
	11:15~11:40	SE02	짧은 발성에 강인한 ECAPA-TDNN 기반의 화자 인증 시스템 [논문으로]	김승빈, 임찬영, 허정우, 김주호, 신현서, 구교원, 유하진(서울시립대)
	11:40~12:05	SE03	Self-Instruction을 활용한 이모지 생성 모델 [논문으로]	최영우, 이보원(인하대)
12:05~12:30	SE04	실시간 멀티 채널 음성 향상 모델의 출력 방식 비교 분석 [논문으로]	김태한, 박형민(서강대)	
<분과 3> 말장애 및 음성 의학	좌장: 이은정(세한대) 토론자: 김지연(우송대)			
	시간	구분	제목	발표자
	10:50~11:15	SD01	읽기장애아동과 일반아동의 동시조음 특성 비교 [논문으로]	박성숙, 성철재(충남대)
	11:15~11:40	SD02	한국어 자음 유형에 따른 상대적기본주파수 [논문으로]	정지행(마스터플러스병원), 김재옥(강남대)
	11:40~12:05	SD03	NPH 환자와 일반 노인 간 어휘인출능력 및 시선추적에 대한 비교 연구 [논문으로]	권성은(대구대), 박기수(경북대), 윤장혁(건국대), 김나연, 이지호(네오폰스), 하지완(대구대)
12:05~12:30	SD04	경련성발성장애와 다른 음성장애의 식별할 수 있는 음향지표 및 말자료 개발을 위한 예비연구 [논문으로]	서인효(단국대병원), 성철재(충남대)	

● 구두 발표 II

좌장: 이현정(인천대) 토론자: 신동진(전주대)				
<분과 4> 음성학 및 음운론	시간	구분	제목	발표자
	15:00~15:25	PH05	경상 방언의 내포절 의문사 음성 자극에 대한 의문사 작용역 응답 통계 분석 [논문으로]	윤원희(계명대)
	15:25~15:50	PH06	Multimodal cue usage in beginning conversational dyadic exchanges across child development [논문으로]	Jun Ho Chai (조선대), Barbara Zapiór (Univ. of East Anglia), 고연숙(조선대)
	15:50~16:15	PH07	Temporal characteristics of polysemic words in spoken Thai [논문으로]	Kyung-Eun Park, Tae-Yeoub Jang (Hankuk Univ. of Foreign Studies)
좌장: 김유경(순천제일대) 토론자: 이옥분(대구사이버대)				
<분과 5> 말장애 및 음성 의학	시간	구분	제목	발표자
	15:00~15:25	SD05	인지-언어훈련 애플리케이션을 활용한 어휘 및 조음능력 증진 [논문으로]	김나연, 이지호(네오폰스), 박기수(경북대), 윤장혁(건국대), 하지완(대구대)
	15:25~15:50	SD06	Comparison of acoustic features due to the Lombard effect in typically developing children and adults [논문으로]	Yelim Jang, Jahee Hwang, Nuri Lee, Nakyung Lee, Seun Eum, Youngmee Lee (Ewha Womans Univ.)
	15:50~16:15	SD07	중학교 읽기부진 학생의 시나리오 글 읽기 유창성 특성 [논문으로]	박지혜, 성철재(충남대)
좌장: 김재옥(강남대) 토론자: 박상희(대구사이버대)				
<분과 6> 말장애 및 음성 의학	시간	구분	제목	발표자
	15:00~15:25	SD08	구개열 언어의 치료 - 조음동작형성법의 개발과 적용 [논문으로]	이은정(세한대)
	15:25~15:50	SD09	Perception of voice attractiveness: Effects of speaking rate, gender, and age [논문으로]	Eunvin Chun, Yoora Jeong, Heeyeon Kim, Nahyun Kim, Sujin Kim, Youngmee Lee (Ewha Womans Univ.)
	15:50~16:15	SD10	한국어 이중모음 분석을 위한 활음 분절 알고리즘 개발 [논문으로]	박지연, 성철재(충남대)

● 구두 발표 III

좌장: 오재혁(건국대) 토론자: 정현성(한국교원대)				
<분과 7> 음성학 및 음운론	시간	구분	제목	발표자
	16:25~16:50	PH08	한국어 단모음의 조음 양상과 공간 구조: 3D 초음파 측정법을 기반으로 [논문으로]	하영우, 황영(전주대)
	16:50~17:15	PH09	운율 위치에 따라 달라지는 서울 방언 파열음 지각 [논문으로]	소현정, 김정윤(부산대)
	17:15~17:40	PH10	한국어 비유창성 유발 요인에 대한 연구 [논문으로]	홍정연(건국대)
좌장: 박기영(ETRI) 토론자: 유하진(서울시립대)				
<분과 8> 음성 공학	시간	구분	제목	발표자
	16:25~16:50	SE05	동적 어트랙터 생성을 위한 토큰 기반 화자 분할 프레임워크 [논문으로]	구교원, 허정우, 김주호, 신현서, 임찬영, 김승빈, 유하진(서울시립대)
	16:50~17:15	SE06	한국인 아동 화자의 수준별 영어 음소 인식을 위한 Wav2Vec2 기반의 도메인 적응 기법 [논문으로]	박지은, 박정식 (한국외대)
	17:15~17:40	SE07	웨어러블 청각 장치의 음성향상을 위한 청각 필터뱅크 잡음 제거 신경망 [논문으로]	박주영(전남대), 김선만(한국광기술원), 성진택(전남대)
좌장: 김지연(우송대) 토론자: 김유경(순천제일대)				
<분과 9> 말장애 및 음성 의학	시간	구분	제목	발표자
	16:25~16:50	SD11	The Impact of Word-Touch Synchrony on Early Language Development: Preliminary Findings from Korean Mother-Child Interactions [논문으로]	Jongmin Jung, Suhan Kim, Eon-Suk Ko (Chosun Univ.)
	16:50~17:15	SD12	장애음성 평가 및 진단을 위한 한국어 자음 Spectra 에너지 분포 특성 [논문으로]	강동효, 성철재(충남대)
	17:15~17:40	SD13	대중가요 보컬 전공 남학생의 성종에 따른 모음 간 음역 차이 [논문으로]	지일송, 김재옥(강남대)

● 포스터 발표

좌장: 고연숙(조선대), 정종민(조선대), 박정식(한국외대)

시간	구분	발표 제목	발표자
13:10 ~ 14:00	P01	POSHA-S를 통한 직업군인의 말더듬 인식 연구 [논문으로]	차화정, 박진(가톨릭관동대)
	P02	갑상선 수술 후 증후군의 빠른 음성회복을 위한 음성치료의 효과 [논문으로]	김지성(건양대)
	P03	청각적 자극 및 시각적 자극에 따른 말소리장애 아동의 단어 회상하기 과제의 수행 비교 [논문으로]	김주원, 하지완(대구대)
	P04	아동기말실행증의 말소리 감별진단도구 마련을 위한 예비연구 [논문으로]	이옥분, 박상희(대구사이버대)
	P05	게임 닉네임의 독음 후보군 생성 방법 [논문으로]	김나현, 김연정(서울대), 이용혁, 조남현(NCSOFT)
	P06	정보 교란 기법과 스피커 인코더를 활용한 한국어 제로샷 음성 합성 모델 [논문으로]	방채운(광주과학기술원), 전찬준(조선대), 김홍국(광주과학기술원)
	P07	자기 지도 학습 모델 특징 결합을 이용한 적대적 신경망 기반 역양 음성인식 [논문으로]	경나, 박정식(한국외대)
	P08	음성 향상을 위한 게이트 합성곱 순환 신경망 [논문으로]	고형일, 나승대, 김명남(경북대)
	P09	Pitch 정보를 활용한 Perturbation AUTOVC의 성능 향상 [논문으로]	박화영(광주과학기술원), 전찬준(조선대), 김홍국(광주과학기술원)
	P10	한국어 종성 유음의 발음 변이 연구 [논문으로]	서윤정, 드류 크로스비(고려대)
	P11	데이터를 활용한 프랑스어 단어 수준 음절 유형 및 자음군 분포 분석 [논문으로]	박소은, 김선희(서울대)
	P12	한국인 학습자의 프랑스어 연음 오류 분석 [논문으로]	김민정, 김선희(서울대)
	P13	시대별 애니메이션 캐릭터 음성에 대한 음향 및 선호도 분석 [논문으로]	이선미, 조희수, 최영경, 김지연, 이륜하, 이영미(이화여대)
	P14	Acoustic Analyses in the Imitation of Emotional Speech in Children with Typical Development [논문으로]	Su Been Kim, Jung Eun Kim, Soo Hyoung Cho, Hyo Sun Lee, Seong Yun Moon, Youngmee Lee (Ewha Womans Univ.)

차례

특 강

특강1 LLM (Large Language Model) 기술 현황 및 응용 구명완	2
특강2 ChatGPT를 활용한 문화와 사회성 연구 김지혜	50

구두 발표

PH01 대화 상호작용의 개시자 효과와 아동의 언어 성과와의 연관성 Jun Ho Chai, Jongmin Jung, and Eon-Suk Ko	56
PH02 Developmental Variations in English Vowel Production : A Proficiency-Based Study of Korean ESL Learners Mi Sun Park, Joo-Kyeong Lee	58
PH03 Effects of Age, Gender, and Language Ability on the Acoustic Fluency and Pronunciation Judgements in English Utterances by Korean Speakers Hyunsong Chung	61
PH04 A study on the L2 phonetic imitation: Korean learners' case of English stress assignment Mi-Young Bae, Tae-Yeoub Jang	62
PH05 경상 방언의 내포질 의문사 음성 자극에 대한 의문사 작용역 응답 통계 분석 Weonhee Yun	64
PH06 Multimodal cue usage in beginning conversational dyadic exchanges across child development Jun Ho Chai, Barbara Zapior, 고 언 숙	65
PH07 Temporal Characteristics of Polysemous Words in Spoken Thai Kyung-Eun Park, Tae-Yeoub Jang	67
PH08 한국어 단모음의 조음 양상과 공간 구조 : 3D 초음파 측정법을 기반으로 Young-Woo Ha, Young Hwang	68
PH09 운율 위치에 따라 달라지는 서울 방언 파열음 지각 So Hyunjung, Jonny Jungyun Kim	70
PH10 한국어 비유창성 유발 요인에 대한 연구 Jungyeon Hong	71

SE01 한국어 음성인식을 위한 중단형 음성인식 모델의 실험적 비교	
Changhan Oh, Kiyong Park, Jeomja Kang, Hwajeon Song, Wooyong Choi	73
SE02 짧은 발성에 강인한 ECAPA-TDNN 기반의 화자 인증 시스템	
Seung-bin Kim, Chan-yeong Lim, Jungwoo Heo, Ju-ho Kim, Hyun-seo Shin, Kyo-won Koo, Ha-Jin Yu	74
SE03 Self-Instruction을 활용한 이모지 생성모델	
Youngwoo Choi, Bowon Lee	75
SE04 실시간 멀티 채널 음성 향상 모델의 출력 방식 비교 분석	
Tae Han Kim, Hyung-Min Park	76
SE05 동적 어트랙터 생성을 위한 토큰 기반 화자 분할 프레임워크	
Kyo-won Koo, Jungwoo Heo, Ju-ho Kim, Hyun-seo Shin,	77
SE06 한국인 아동 화자의 수준별 영어 음소 인식을 위한 Wav2Vec2 기반의 도메인 적응 기법	
Ji-Eun Park, Jeong-Sik Park	78
SE07 웨어러블 청각 장치의 음성향상을 위한 청각 필터뱅크 잡음 제거 신경망	
Ju Yeong Park, Seon Man Kim, Jin-Taek Seong	80
SD01 읽기장애아동과 일반아동의 동시조음 특성 비교	
Sungsook Park, Cheoljae Seong	81
SD02 한국어 자음 유형에 따른 상대적기본주파수	
Jung, Ji Hang, Jaeock Kim	82
SD03 NPH 환자와 정상 노인 간 어휘인출능력 및 시선추적에 대한 비교 연구	
Seong-Eun Kwon, Ki-Su Park, Jang-Hyeok Yoon, Na-Yeon Kim, Ji-Ho Lee, Ji-Wan Ha	83
SD04 경련성발성장애와 다른 음성장애의 식별을 위한 음향지표 및 말자료 개발에 대한 예비연구	
In Hyo Seo, Cheol Jae Seong	86
SD05 인지-언어혼련 애플리케이션을 활용한 어휘 및 조음능력 증진	
Nayeon Kim, Jiho Lee, Ki-su Park, Janghyeok Yoon, Ji-wan Ha	87
SD06 롬바르트 효과가 아동과 성인의 말소리 산출에 미치는 영향: 음향학적 특성과 모음공간면적을 중심으로	
Yelim Jang, Jaehee Hwang, Nuri Lee, Nakyung Lee, Seeun Eum, Youngmee Lee	88
SD07 중학교 읽기부진 학생의 시나리오 글 읽기 유창성 특성	
Jihye Park, Cheoljae Seong	89
SD08 구개열 언어치료 - 조음동작형성법의 개발과 적용	
Eun-Jeong Lee	90
SD09 화자의 말속도, 청자의 성별, 연령과 말속도가 청자의 호감도 평가에 미치는 영향	
Eunvin Chun, Yoora Jeong, Heeyeon Kim, Nahyun Kim, Sujin Kim, Youngmee Lee	91
SD10 한국어 이중모음 분석을 위한 활음 분절 알고리즘 개발	
Jiyeon Park, Cheoljae Seong	92
SD11 단어와 터치간 시간적 일치성이 초기 언어발달에 미치는 영향: 한국 어머니와 아동의 상호작용을 통해 본 연구	
Jongmin Jung, Suhan Kim, Eon-Suk Ko	93

SD12 장애음성 평가 및 진단을 위한 한국어 자음 spectra 에너지 분포 특성 Donghyo Kang, Cheoljae Seong	94
SD13 대중가요 보컬 전공 남학생의 성종에 따른 모음 간 음역 차이 Il-Soung Ji, Jaeock Kim	95

포스터 발표

P01 POSHA-S를 통한 직업군인의 말더듬 인식 연구 Hwajung Cha, Jin Park	97
P02 갑상선 수술 후 증후군의 빠른 음성회복을 위한 음성치료의 효과 Jisung, Kim	98
P03 청각적 자극 및 시각적 자극에 따른 말소리장애 아동의 단어 회상하기 과제 수행 비교 Ju-won Kim, Ji-Wan Ha	99
P04 아동기말실행증의 말소리 감별진단도구 마련을 위한 예비연구 Okbun Lee, Sanghee Park	102
P05 게임 너네임의 독음 후보군 생성 방법 Nahyun Kim, Yeonjeong Kim, Yonghyuk Lee, Namhyun Cho	104
P06 정보 교란 기법과 스피커 인코더를 활용한 한국어 제로샷 음성 합성 모델 Chae Woon Bang, Chan Jun Chun, Hong Kook Kim	105
P07 자기 지도 학습 모델 특징 결합을 이용한 적대적 신경망 기반 억양 음성인식 Na Geng, Jeong-Sik Park	106
P08 음성 향상을 위한 게이트 합성곱 순환 신경망 H.L. Koh, S.D. Na, M.N. Kim	107
P09 Pitch 정보를 활용한 Perturbation AUTOVC의 성능 향상 Hwa-Young Park, Chanjun Chun, Hong Kook Kim	110
P10 한국어 종성 유음의 발음 변이 연구 YoonJeong Seo, Drew Crosby	111
P11 데이터를 활용한 프랑스어 단어 수준 음절 유형 및 자음군 분포 분석 Soeun Park, Sunhee Kim	112
P12 한국인 학습자의 프랑스어 연음 오류 분석 Minjeong Kim, Sunhee Kim	114
P13 시대별 애니메이션 캐릭터 음성에 대한 음향 및 선호도 분석 Sunmi Lee, Heesu Cho, Youngkyoung Choi, Jiyeon Kim, Ryunha Lee, Youngmee Lee	116
P14 정상발달 아동의 감정발화모방에 대한 음향학적 분석 Su Been Kim, Jung Eun Kim, Soo Hyoung Cho, Hyo Sun Lee, Seong Yun Moon, Youngmee Lee	117

2024 한국음성학회 봄 학술대회 특 강

■ LLM (Large Language Model) 기술 현황 및 응용

구명완(서강대, 컴퓨터공학과)

■ ChatGPT를 활용한 문화와 사회성 연구

김지혜(서강대, 사회학과)

LLM(Large Language Model) 기술 현황 및 응용

구명완교수
서강대학교 컴퓨터공학과
Email: mwkw@sogang.ac.kr

한국음성학회

내용

- 서론
- LLM 기초
- 프롬프트 엔지니어링
- 결론



서론

- GPT Store 출시: 2024. 1.10

AI챗봇, 벌써 300만개 ... 이젠 앱처럼 사고 판다

- 오픈AI의 대공세 ... 월 20弗 'GPT스토어' 출시



서론

- 최근 성과 <https://www.statista.com/chart/29174/time-to-one-million-users/>
 - 1백만 사용자 모집 시간



서론

- 최근 성과 https://www.samsungsds.com/kr/insights/chatgpt_whitepaper1.html

일자	상세 내용
'22.11.30	OpenAI, ChatGPT(GPT-3.5) 론칭
'22.12.05	ChatGPT 일간 활성화 사용자 수(DAU) 100만 명 돌파
'22.12.15	OpenAI, ChatGPT에 워터마크 도입
'22.12.25	ChatGPT 일간 활성화 사용자 수(DAU) 1000만 명 돌파
'22.12.26	Google, ChatGPT에 '코드 레드' 발령
'22.12.27	베스핀글로벌, '벨프나우 AI'에 GPT 기술 도입
'22.12.31	12월 월간 활성화 사용자 수(MAU) 5,700만 명 돌파
'23.01.23	Microsoft, OpenAI에 100억 달러 추가 투자
'23.01.31	12월 월간 활성화 사용자 수(MAU) 1억 명 돌파

서론

- 최근 성과

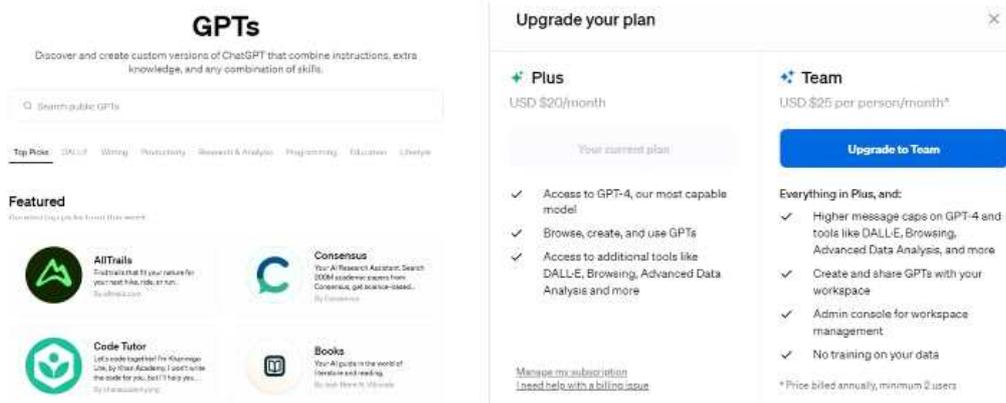
'23.02.01	투블러AI, 생성형 AI 활용 방법 관련 특허 등록
'23.02.03	네이버, 상반기 '서지GPT' 출시 계획 발표
'23.02.06	Google, Bard 론칭 발표 및 데모 시연
'23.02.07	Microsoft, ChatGPT 탑재 검색 엔진 '빙(Bing)' 발표
'23.03.09	업스테이지, OCR 기술에 ChatGPT를 결합한 '아숙업(AskUp)' 론칭
'23.03.14	OpenAI, ChatGPT(GPT-4) 론칭
'23.03.15	Google, Workspace 제품군에 대한 AI 전면 적용 계획 발표
'23.03.16	Microsoft, 업무 생산성 도구 전반에 AI를 적용한 Microsoft 365 Copilot 출시 계획 발표
'23.03.23	OpenAI, ChatGPT plugin 지원 발표

서론

- 최근 성과
 - 2023. Nov. 4: xAI(일론 머스크 AI 스타트업) Grok 공개
 - SNS X 를 통한 실시간 정보에 접근 가능
 - 농담 등이 가능한 표현, SNS 와 유사한 문구 사용
 - 2023. Nov. 6: **GPT-4 Turbo** 공개
 - 2023. Apr. 이전 정보 반영
 - 128,000 token 지원. (GPT-4 8,000 token 지원)
 - 이전 버전의 가격에 비해서 36% 낮아짐
 - GPT Store 출시 예정
 - 2023. Dec. 6: Google DeepMind **Gemini** Release
 - a family of multimodal large language models by DeepMind
 - contender to OpenAI's GPT-4.

서론

- GPT Store 출시
 - 2023. 1.10



The screenshot displays the GPT Store interface. At the top, it says "GPTs" and "Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills." Below this is a search bar for "Search public GPTs" and a "Tip Role" section with categories like "DML", "Writing", "Productivity", "Research & Analysis", "Programming", "Education", and "Lifestyle".

The "Featured" section lists four GPTs:

- AllTrails**: Trailmix that fit your nature for your next hike, run, or hike. By AllTrails.com
- Consensus**: Your AI Research Assistant. Search 100M academic papers from Consensus, get science-backed. By Consensus
- Code Tutor**: Let's code together! No language like by Stan Akcalina. I can't write code for you, but I'll help you... By @stanakcalina
- Books**: Your AI guide in the world of creative and reading. By Book Store N. 1000

An "Upgrade your plan" overlay is shown on the right, comparing two plans:

- Plus**: USD \$20/month. Features include: Access to GPT-4, our most capable model; Browse, create, and use GPTs; Access to additional tools like DALL·E, Bing, Advanced Data Analysis and more.
- Team**: USD \$25 per person/month*. Features include: Higher message caps on GPT-4 and tools like DALL·E, Bing, Advanced Data Analysis, and more; Create and share GPTs with your workspace; Admin console for workspace management; No training on your data.

 A blue "Upgrade to Team" button is visible.

서론

GPTs

Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.

Q Search GPTs

Top Picks DALL·E Writing Productivity Research & Analysis Programming Education Lifestyle

Trending

Most popular GPTs by our community

- 1  **image generator**
A GPT specialized in generating and refining images with a mix of professional and friendly tone. image generator
By NAF JALOTAIB
- 2  **Write For Me**
Write tailored, engaging content with a focus on quality, relevance and precise word count.
By puzzle today
- 3  **Consensus**
Your AI Research Assistant. Search 200M academic papers from Consensus, get science-based answers, and draft content...
By consensus.io
- 4  **Scholar GPT**
Enhance research with 200M+ resources and built-in critical reading skills. Access Google Scholar, PubMed, JSTOR, Arxiv, and more.
By awesomegpts.ai
- 5  **Canva**
Effortlessly design anything: presentations, logos, social media posts and more.
By canva.com
- 6  **Logo Creator**
Use me to generate professional logo designs and app icons!
By Chase Leon

서론

• Scholar GPT



Scholar GPT

By awesomegpts.ai

Enhance research with 200M+ resources and built-in critical reading skills. Access Google Scholar, PubMed, JSTOR, Arxiv, and more, effortlessly.

Find the latest research about AI I will upload a PDF paper; Use critical skills to read it.

It provide a research paper link, Please analyze it Type "LS" to list my built-in critical reading skills

@ Find the latest research about Myoung-Wan Koo



서론

• Scholar GPT

You
Find the latest research about Myoung-Wan Koo

Scholar GPT
Scholar GPT wants to talk to actions.sider.ai ...
Only allow sites you trust

[Allow](#) [Always Allow](#) [Decline](#)



서론

• Scholar GPT

Scholar GPT
 Talked to actions.sider.ai

Here are some of the latest research works associated with Myoung-Wan Koo:

- Task-optimized adapters for an end-to-end task-oriented dialogue system**
 - Authors:** N Bang, J Lee, MW Koo
 - Summary:** This paper discusses Task-Oriented Dialogue (TOD) systems, which are designed to perform specific tasks by tracking dialogue states and generating appropriate responses. The study explores the optimization of these systems for better performance.
 - Published in:** arXiv, 2023
 - [Read the paper](#)
- The SGU systems for the CHIME-7 UDASE Challenge**
 - Authors:** J Jang, MW Koo
 - Summary:** This work presents the SGU domain-adapted speech enhancement system's implementation details, which aims to enhance the baseline system for the CHIME-7 challenge.
 - Published in:** Proceedings of the 7th International Workshop, 2023
 - [Read the paper](#)
- Towards Context-Based Violence Detection: A Korean Crime Dialogue Dataset**
 - Authors:** M Kim, H Yeen, MW Koo
 - Summary:** This paper introduces a dataset aimed at enhancing violence detection through artificial intelligence by providing a resource for training models to detect and classify instances of violence in dialogues.
 - Published in:** Findings of the Association for Computational Linguistics, 2024
 - [Read the paper](#)

서론

- Scholar GPT

4. **Inappropriate Pause Detection In Dysarthric Speech Using Large-Scale Speech Recognition**
 - **Authors:** J Lee, Y Choi, TJ Song, MW Koo
 - **Summary:** The study focuses on dysarthria, a speech disorder affecting intelligibility, by detecting inappropriate pauses which are significant indicators of the disorder's severity.
 - **Published in:** arXiv, 2024
 - [Read the paper](#)
5. **DC CoMix TTS: An End-to-End Expressive TTS with Discrete Code Collaborated with Mixer**
 - **Authors:** Y Choi, MW Koo
 - **Summary:** This paper proposes a new input representation and architecture to address content-leakage in text-to-speech (TTS) systems, focusing on maintaining expressiveness.
 - **Published in:** arXiv, 2023
 - [Read the paper](#)

These works span various areas, including dialogue systems, speech enhancement, violence detection, speech disorder assessment, and text-to-speech systems, reflecting Myoung-Wan Koo's contributions across multiple aspects of computational linguistics and speech processing.

서론

- GPT 4

ChatGPT 4 -

You
Find the latest research about Myoung-Wan Koo

ChatGPT
Myoung-Wan Koo has been actively involved in various areas of computational linguistics and artificial intelligence research. Here are two recent contributions by Myoung-Wan Koo.

1. **"Task-Optimized Adapters for an End-to-End Task-Oriented Dialogue System"** (July 2023): This paper, co-authored by Namu Bang, Jaehyun Lee, and Myoung-Wan Koo, proposes a novel approach to enhance Task-Oriented Dialogue (TOD) systems. The research introduces End-to-end TOD systems with Task-Optimized Adapters, which are designed to improve dialogue systems by allowing independent learning for different tasks, thereby making the systems more efficient and effective [1].
2. **"Inappropriate Pause Detection In Dysarthric Speech Using Large-Scale Speech Recognition"** (February 2024): Authored by Jaehyun Lee, Yerin Choi, Tae-Jin Song, and Myoung-Wan Koo, this study focuses on the challenge of dysarthria, a speech disorder often resulting from neurological injuries or diseases. The team developed a method to detect inappropriate pauses in speech, which are critical for assessing the severity of speech impairments and providing therapeutic interventions. Their approach extends large-scale speech recognition models with a new layer specifically designed for this task, demonstrating improved detection of inappropriate pauses compared to existing baselines [2].
3. **"Towards Context-Based Violence Detection: A Korean Crime Dialogue Dataset"** (March 2024): Co-authored with Minju Kim and Heuiyeon Yoon, this paper aims to advance the field of artificial intelligence in detecting and classifying violence within societal contexts. They introduced the Korean Crime Dialogue Dataset (KCDD), the first of its kind focused on violence that occurs in online settings, classified into different criminal categories. They also proposed a model, Relationship-Aware BERT, to analyze these dialogues, hoping that their dataset and model can aid in detecting violence and helping those in danger [3].

These studies highlight Myoung-Wan Koo's contributions to enhancing dialogue systems, improving speech recognition and therapeutic interventions for speech disorders, and using AI for social good by detecting violence in digital communications.

🔗 📄 📄 📄

서론

- GPT 3.5

ChatGPT 3.5



서론

- GPT 3.5

You
Find the latest research about Myoung-Wan Koo

ChatGPT
As of my last update in January 2022, there isn't any specific research paper or information available about an individual named Myoung-Wan Koo that has garnered widespread attention or recognition in academic circles. It's possible that new research or publications may have emerged since then, but without more specific details, it's challenging to provide the latest information. If Myoung-Wan Koo has made significant contributions or published research recently, I would recommend searching academic databases such as PubMed, Google Scholar, or research databases in the specific field of interest for the most up-to-date information.



서론

- Google Scholar



Myoung-Wan Koo

Articles Case law

New! Supercharge your PDF reading: Follow references, skim outline, jump to figures

서론

- Google Scholar

Google Scholar Myoung-Wan Koo

Articles About 248 results (0.05 sec)

Any time

Since 2024

Since 2023

Since 2020

Custom range...

Sort by relevance

Sort by date

Any type

Review articles

include patents

include citations

Create alert

User profiles for Myoung-Wan Koo

Myoung Wan Koo
Sogang University
Verified email at sogang.ac.kr
Cited by 438

Task-optimized subplans for an end-to-end task-oriented dialogue system [PDF] arxiv.org
 N. Baeg, J. Lee, M.W. Koo - arXiv preprint arXiv:2305.02488, 2023 - arxiv.org
 Task-Oriented Dialogue (TOD) systems are designed to carry out specific tasks by tracking dialogue states and generating appropriate responses to help users achieve defined goals, ...
 ☆ Save Cited by 19 All 4 versions

A Bi-LSTM memory network for end-to-end goal-oriented dialog learning [HTML] sciencedirect.com
 B Kim, KT Chung, J. Lee, J. Seo, M.W. Koo - Computer Speech & Language, 2019 - Elsevier
 We develop a model to satisfy the requirements of Dialog System Technology Challenge 8 (DSTC8) Track 1: building an end-to-end dialog systems for goal-oriented applications. This ...
 ☆ Save Cited by 29 Web of Science: 13

Categorization of Korean news articles based on convolutional neural network using Doc2Vec and Word2Vec Find it @ SOGANG
 D Kim, M.W. Koo - Journal of KIISE, 2017 - koreascience.kr
 In this paper, we propose a novel approach to improve the performance of the Convolutional Neural Network (CNN) word embedding model on top of word2vec with the result of ...
 ☆ Save Cited by 34 00

Symbolic Heuristic Search Value Iteration for Factored POMDPs [PDF] aaai.org
 HS Sim, KE Kim, JH Kim, DS Cheng, MW Koo - AAAI, 2008 - osh.aaai.org
 We propose Symbolic heuristic search value iteration (Symbolic HSVI) algorithm, which extends the heuristic search value iteration (HSVI) algorithm in order to handle factored partially ...
 ☆ Save Cited by 58 All 6 versions

서론

• AllTrails

You
서울 관악산 등산로 알려주세요.

AllTrails
Talked to chatgpt-production.alltrails.com

Based on your request for trails at 관악산 (Gwanaksan) in Seoul, South Korea, the search defaulted to trails in the nearby Bukhansan National Park. I can adjust the parameters for more accurate recommendations if needed. Here are the trails from Bukhansan National Park:

1. Bukhansan Mam Route



- **Location:** Bukhansan National Park, Seoul, South Korea
- **Description:** This trail offers spectacular views of rock formations and the city. Key attractions include Daeseomun, Muryangsa Temple, and Bukhansan History Center. The beginning has paved paths, with the last restroom located at the Bukhansan History Center.

서론

• Text

Capability	Benchmark (Higher is better)	Description	Gemini Ultra	GPT-4 <small>(All numbers calculated where reported numbers were missing)</small>
General	MMLU	Representation of questions in 57 subjects (incl. STEM, humanities, and others)	90.0% <small>Out-of-d</small>	86.4% <small>5-shot (reported)</small>
Reasoning	Big-Bench Hard	Diverse set of challenging tasks requiring multi-step reasoning	83.6% <small>3-shot</small>	83.1% <small>3-shot (avg)</small>
	DROP	Reading comprehension (PI Scores)	82.4 <small>Variable shots</small>	80.9 <small>3-shot (reported)</small>
	HellaSwag	Commonsense reasoning for everyday tasks	87.8% <small>10-shot*</small>	95.3% <small>10-shot* (reported)</small>
Math	GSM8K	Basic arithmetic manipulations (incl. Grade School math problems)	94.4% <small>math32</small>	92.0% <small>5-shot CoT (reported)</small>
	MATH	Challenging math problems (incl. algebra, geometry, pre-calculus, and others)	53.2% <small>4-shot</small>	52.9% <small>4-shot (AP)</small>
Code	HumanEval	Python code generation	74.4% <small>0-shot (DP)</small>	67.0% <small>0-shot* (reported)</small>
	Natural2Code	Python code generation. New held out dataset. HumanEval-like, not leaked on the web	74.9% <small>0-shot</small>	73.9% <small>0-shot (AP)</small>



서론

- Multimodal

Category	Benchmark	Description	Gemini	GPT-4V
Image	MMMU	Multidisciplinary college-level reasoning problems	59.4% <small>(Gemini Ultra)</small>	56.8% <small>(GPT-4o)</small>
	VQA v2	Natural image understanding	77.8% <small>(Gemini Ultra)</small>	77.2% <small>(GPT-4o)</small>
	TextVQA	OCR on natural images	82.3% <small>(Gemini Ultra)</small>	78.0% <small>(GPT-4o)</small>
	DocVQA	Document understanding	90.9% <small>(Gemini Ultra)</small>	88.4% <small>(GPT-4o)</small>
	Infographic VQA	Infographic understanding	80.3% <small>(Gemini Ultra)</small>	75.1% <small>(GPT-4o)</small>
	MathVista	Mathematical reasoning in visual contexts	53.0% <small>(Gemini Ultra)</small>	49.9% <small>(GPT-4o)</small>
Video	VATEX	English video captioning (2.0s)	62.7 <small>(Gemini Ultra)</small>	56.0 <small>(GPT-4o)</small>
	Perception Test MCQA	Video question answering	54.7% <small>(Gemini Ultra)</small>	46.3% <small>(GPT-4o)</small>
Audio	CoVeST 2 (21 languages)	Automatic speech translation (BLEU score)	40.1 <small>(Gemini Pro)</small>	29.1 <small>(Whisper v2)</small>
	FLEURS (62 languages)	Automatic speech recognition (based on word error rate, lower is better)	7.6% <small>(Gemini Pro)</small>	17.6% <small>(Whisper v2)</small>



서론

Upstage 13B LLM Beats MS ToRA and ChatGPT in Mathematical Reasoning

Upstage's math-tailored LLM, developed with Qanda and KT, surpasses MS ToRA 13B and ChatGPT, redefining global AI benchmarks in superior math-solving capabilities.

model. The resulting format integrates the advantages of natural language rationales in mathematical reasoning (Chain-of-Thought) with code-based algorithmic techniques in precise calculation (Program-of-Thought), significantly enhancing the model's capability to solve complex mathematical problems.

Model	#Parameters	GSM8K	MATH	Avg**
Upstage-Qanda	13B	78.2	48.8	73.5
MS ToRA 13B*	13B	75.8	48.1	71.3
ChatGPT*	-	80.8	35.5	72.3
GPT-4*	-	92	42.5	78.3

서론

● 초거대 언어모델 개발 상황: 국내

기업명	모델	파라미터 수	공개연도	특징
LG	엑사원 (EXAONE)	3,000억 개	2021년 12월	<ul style="list-style-type: none"> LG그룹 차원에서 개발한 초거대시 국내 최대 파라미터 보유 LGU+와 협력
네이버	하이퍼클로바 (HyperCLOVA)	2,040억 개	2021년 5월	<ul style="list-style-type: none"> 한국어 특화 언어 모델로 한국어 데이터 학습량 GPT-3의 6,500배 검색, 쇼핑, 음성인식 등에 활용
카카오	코지피티 (KoGPT)	300억 개	2021년 11월	<ul style="list-style-type: none"> GPT-3의 한국어 특화 언어 모델 구글 텐서 처리장치 활용 한국어 데이터 학습(2,000억 개 토큰)
	민달리 (minDALL-E)	13억 개	2022년 4월	<ul style="list-style-type: none"> 이미지 생성 모델(Text-to-Image)
	RQ-트랜스포머	39억 개	2022년 4월	<ul style="list-style-type: none"> minDALL-E의 업그레이드 버전 (이미지 생성 속도 및 데이터셋 크기 2배) 저희사 카카오프레인의 독자 기술로 개발된 이미지 생성 모델
	킬로 (Karlo)	33억 개	2022년 12월	<ul style="list-style-type: none"> AI 아티스트 '킬로 1.0' 오픈소스 공개 다양한 스타일의 이미지 생성 등 1.8억 장 규모의 텍스트-이미지 데이터 셋 학습

서론

● 초거대 언어모델 개발 상황: 국내

SKT	에이닷	-	2022년 5월	<ul style="list-style-type: none"> GPT-3의 한국어 버전 에이닷 고도화를 위해 코난테크놀로지 지분 21% 확보 장기기억 및 멀티모달 기술 도입 예정
KT	믿음	2,000억 개	2023년 10월	<ul style="list-style-type: none"> 믿음 기반의 대화형 서비스 추진 중

● 삼성전자: 가우스 (경량 생성 AI)

서론

- 초거대 언어모델 개발 상황: 국내
 - 네이버: HyperClova X
 - 2,040억개, ChatGPT의 6,500배 이상 훈련
 - 2023. Aug. 24
 - 삼성전자 DS 전용 활용 예정
 - KT: 믿음
 - 2023. Oct. 31
 - 2000억(200B) 개 ~ 70억(7B) 모델 출시
 - B2B 시장 진출
 - SKT: 엔트로픽과 연합, 에이닷
 - SKT TECH Summit 2023 Nov 16-17

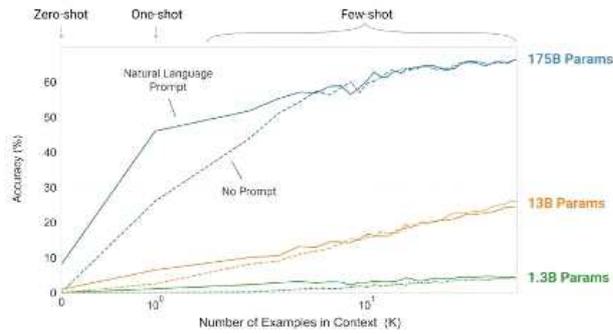
서론

- 초거대 언어모델 개발 상황
 - SKT는 Scatter Lab에 150억 지분투자: 2023.04.25



서론

- Generative Pretrained Transformer, **OpenAI**
 - unidirectional language model
- GPT, GPT2, GPT3: 175 billion



서론

Explosive Growth in AI Computational Requirements



서론

- GPT-4 Turbo

How many people are wearing helmet?



GPT-4V:
There are 8 persons wearing helmets in the image

Please determine whether the person in the image wears a helmet or not. And summarize how many people are wearing helmets.

Image 1



Image 2



Image 3



Image 4



Image 5



Image 6



Image 7



Image 8



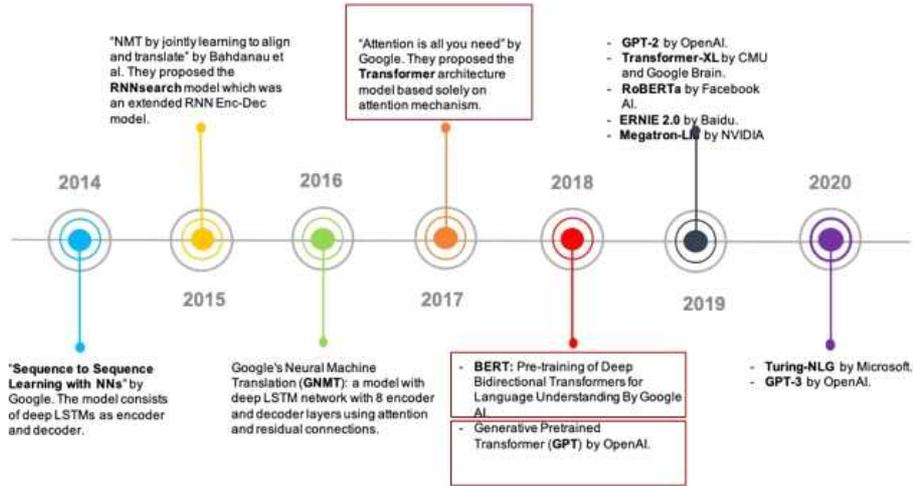
GPT-4V:
Image 1: Not wearing a helmet
Image 2: Not wearing a helmet
Image 3: Wearing a helmet
Image 4: Wearing a helmet
Image 5: Wearing a helmet
Image 6: Wearing a helmet
Image 7: Wearing a helmet
Image 8: Not wearing a helmet
Summary: Out of 8 people, 5 are wearing helmets and 3 are not wearing helmets.

내용

- 서론
- LLM 기초
- 프롬프팅 엔지니어링
- 결론

LLM 기초

• NMT(Neural Machine Translation)



LLM 기초

• Transformer 아키텍처 개요(2017)

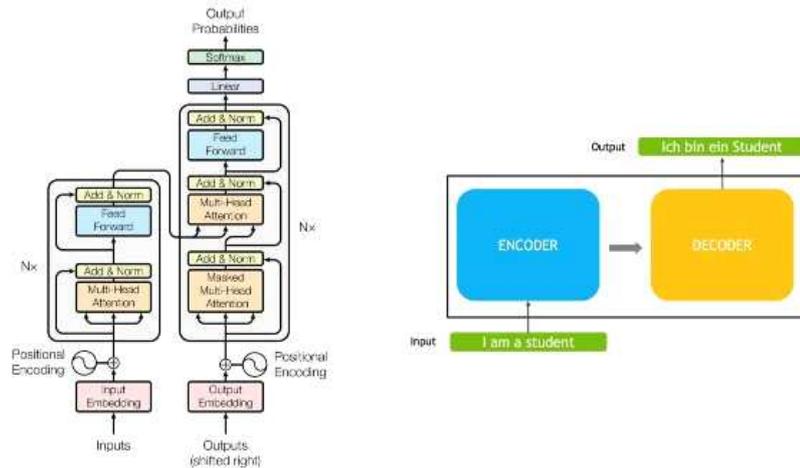
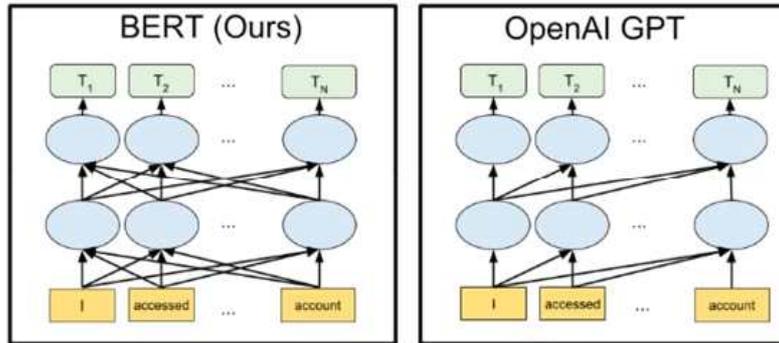


Figure 1: The Transformer - model architecture.

LLM 기초

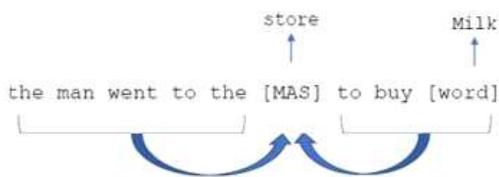
- BERT: Bidirectional Encoder Representations from Transformers.
 - 양방향 트랜스포머
 - Representation Learning



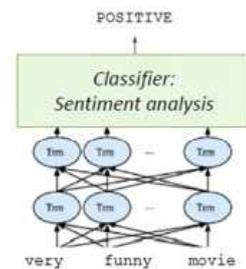
LLM 기초

- BERT: Bidirectional Encoder Representations from Transformers.

Train deep (12 or 24 layers) bidirectional transformer LMs



Fine-tune on individual tasks using task-specific data



LLM 기초

• GPT(Generated Pretrained Transformer)

- Each prefix of the sentence is a training example
 - Wild _____
 - Wild Cats _____
 - Wild Cats are _____
 - ..
 - Wild Cats are but a bunch of wannabe Sun _____
- LLM uses its current function to guess the next word
 - Wild **Geese**
- Guess: **Geese** Correct: **Cats**
- Error= {Cats - Geese}
 - To the LLMs, all vocabulary tokens are just vectors in some high dimensional embedding space; so the difference is well defined as the vector difference
- Propagate this error back through the function, and change the parameters so the error is reduced
 - Using back propagation (aka Chain Rule of derivatives with dynamic programming); the basic workhorse of all neural networks.
- <Go to the next example>

Training LLMs

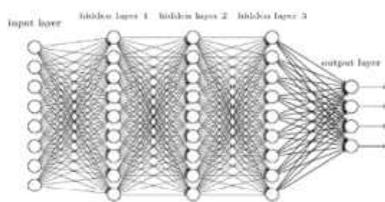
LLM 기초

• GPT(Generated Pretrained Transformer) Context Size

- GPT-4 : 8,192 tokens, GPT-3.5: 4,096 tokens

Mental picture

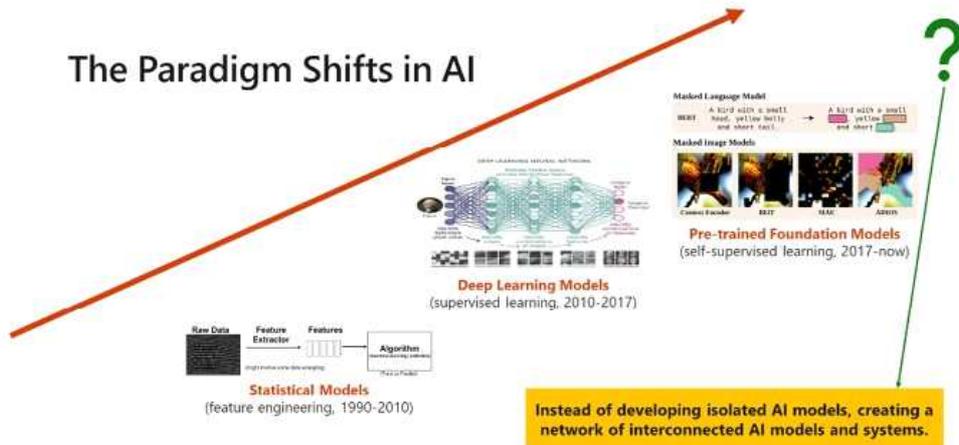
Reality



(graphic by James Campbell)

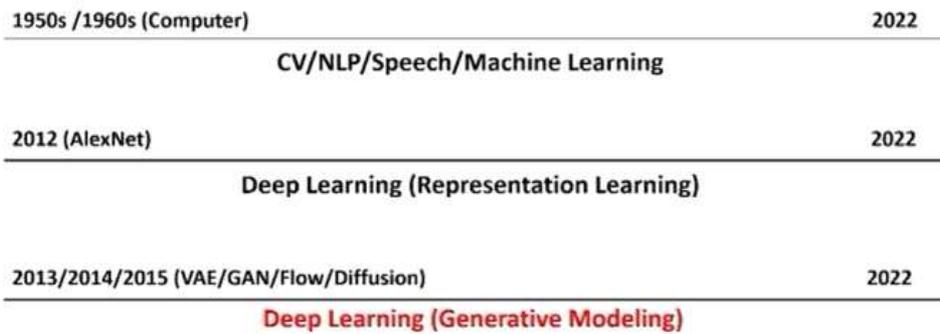
LLM 기초

The Paradigm Shifts in AI



LLM 기초

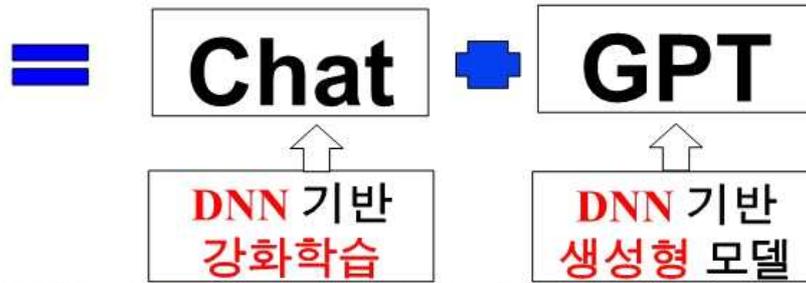
Deep Learning and Generative Models



LLM 기초

- ChatGPT: InstructGPT(+ α) ?
 - 대용량 언어모델에 인간의 평가기준을 추가하여 사람과 유사한 답변을 생성하는 모델

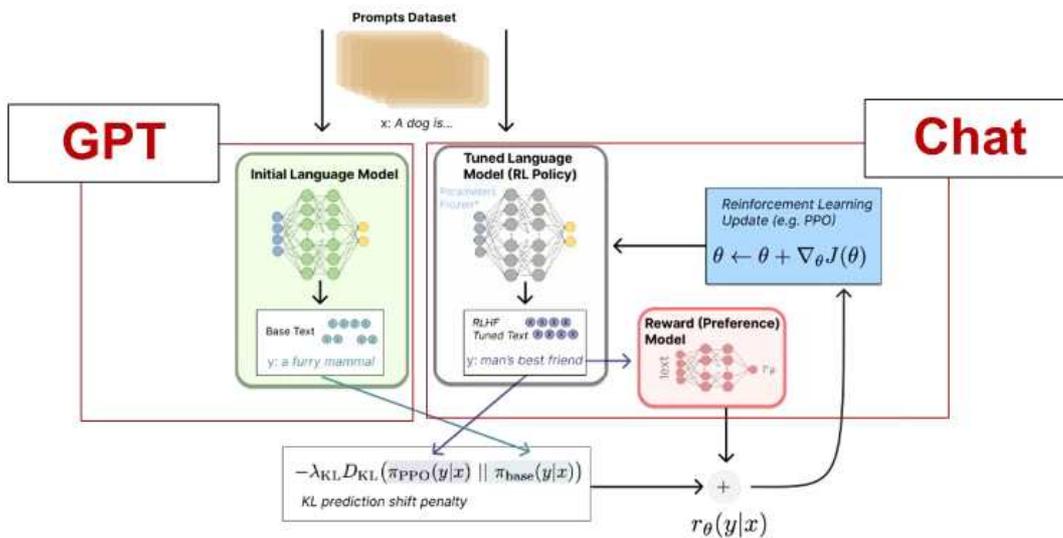
ChatGPT



Reinforcement Learning from Human Feedback (RLHF)

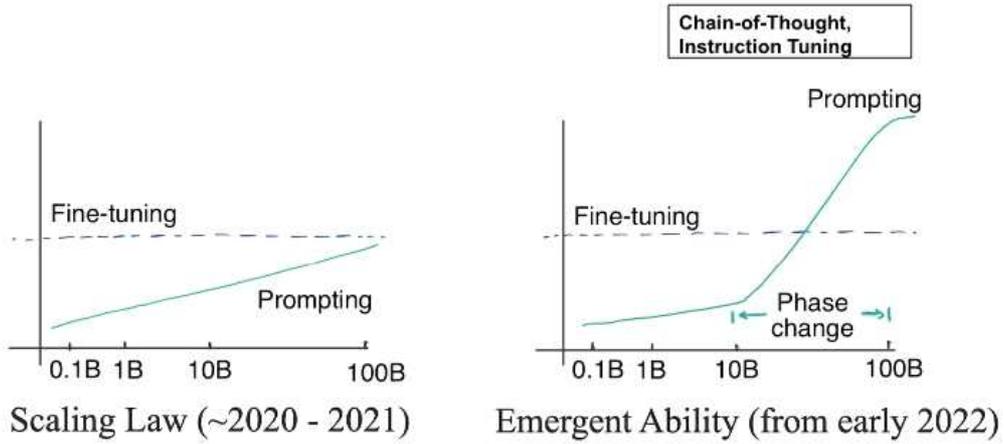
GPT: Generated Pretrained Transformer

LLM 기초: RLHF



LLM 기초

- Prompting Abilities



LLM 기초

- Paradigm Shift

Scale	Small Model	Large Model
Learning	Fine-tuning	In-context Learning
Learning	Supervised	Supervised??
Data	Full training set	Few in-context demonstrations
Generalization	In-distribution	In-distribution + distribution shift

LLM 기초

- How large should the model be for prompting ?
 - For **chain-of-thought** to be **better than standard answer-only prompting**,
 - one needs the model to be **at least 62B**
 - For **chain-of-thought** to be **better than fine-tuning small models** (say T5-11B),
 - one needs the model to be **larger than 175B** where the number 175B comes from GPT-3.

LLM 기초

- How large should the model be?

Params	Model	Norm. avg.	MMLU		BBH		TyDiQA		MGSM	
			Direct	CoT	Direct	CoT	Direct	CoT	Direct	CoT
80M	T5-Small	9.2	26.7	5.6	27.0	7.2	0.0	0.4		
	Flan-T5-Small	3.1 (+6.1)	28.7	12.1	29.1	19.2	1.1	0.2		
250M	T5-Base	-5.1	25.7	14.5	27.8	14.6	0.0	0.5		
	Flan-T5-Base	6.5 (+11.6)	35.9	33.7	31.3	27.9	4.1	0.4		
780M	T5-Large	-5.0	25.1	15.0	27.7	16.1	0.0	0.5		
	Flan-T5-Large	13.8 (+18.8)	45.1	40.5	37.5	31.3	12.5	0.7		
3B	T5-XL	4.1	25.7	14.5	27.4	19.2	0.0	0.8		
	Flan-T5-XL	19.1 (+23.2)	52.4	43.5	41.0	35.2	16.6	1.9		
11B	T5-XXL	-2.9	25.9	18.7	29.5	19.3	0.0	1.0		
	Flan-T5-XXL	23.7 (+26.6)	55.1	48.6	45.3	41.4	19.0	4.9		
8B	PaLM	6.4	24.3	24.1	30.8	30.1	25.0	3.4		
	Flan-PaLM	21.9 (+15.5)	49.3	41.5	36.4	31.1	47.5	8.2		
62B	PaLM	28.4	55.1	49.0	37.4	43.0	40.5	18.2		
	Flan-PaLM	38.8 (+10.4)	59.6	56.9	47.5	44.9	58.7	28.5		
540B	PaLM	49.1	71.3	62.9	49.1	65.7	52.9	45.9		
	Flan-PaLM	58.4 (+9.3)	73.5	70.9	57.9	66.3	67.8	57.0		
62B	cont-PaLM	38.1	61.2	57.6	41.7	53.1	45.7	32.0		
	Flan-cont-PaLM	46.7 (+8.6)	66.1	62.0	51.0	53.3	62.7	40.3		
540B	U-PaLM	50.2	71.5	64.0	49.2	62.4	54.6	49.9		
	Flan-U-PaLM	59.1 (+8.9)	74.1	69.8	59.5	64.9	68.5	60.4		

Table 5: Instruction finetuning (Flan) improves performance on top of other continued pre-training methods. The benchmark suites are MMLU (57 tasks), BBH (23 tasks), TyDiQA (8 languages), and MGSM (10 languages). The evaluation metric on all four benchmark suites is few-shot prompted accuracy (exact match), where we take an unweighted average over all tasks. As an aggregate metric we report the normalized average of MMLU-direct, MMLU-CoT, BBH-direct, BBH-CoT, TyDiQA, and MGSM. These evaluation benchmarks are held-out (not included in the finetuning data). Results for each task in each benchmark are given in Appendix D.



내용

- 서론
- LLM 기초
- 프롬프트 엔지니어링
- 결론



주요 프롬프트 엔지니어링 방법론 예시

- Chain-of-Thought (CoT)
- CoT 확장 프롬프트 엔지니어링
 - Zero-shot-CoT
 - Auto CoT
 - Cross-Lingual-Thought (XLT)
 - Self-Verification
 - Chain-of-Verification (CoVE)
- System Instruction을 활용한 프롬프트 엔지니어링

Chain-of-Thought (CoT)

- LLM이 풀려는 문제와 동일한 문제를 사람이 해결하기 위해 수행하는 추론 과정 각 단계에 대한 서술을 프롬프트에 추가
- CoT 최초 제안 논문 (Wei et al., 2022)에서는 LLM이 응답할 때 사고 과정을 함께 출력하도록 하는 답안 예시를 Few-shot으로 주는 방식으로 구현

Standard Prompting	Chain-of-Thought Prompting
<p>Model Input</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p>Model Input</p> <p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>
<p>Model Output</p> <p>A: The answer is 27. ❌</p>	<p>Model Output</p> <p>A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✔️</p>

Jason Wei, undefined., et al, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," 2023.

Zero-shot-CoT

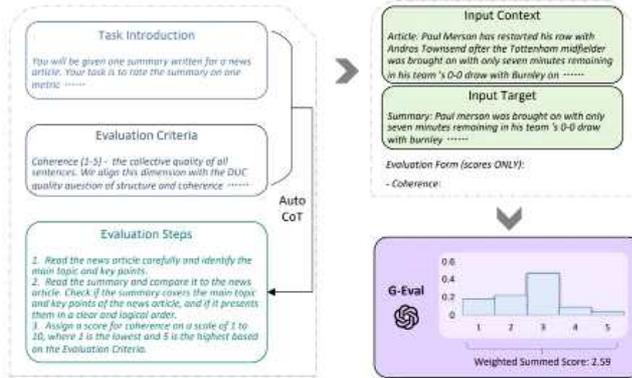
- 최초 제안된 CoT가 퓨샷으로만 사용할 수 있다는 제한점을 해결
- 프롬프트 맨 마지막에 'Let's think step by step' 한 문장을 추가하는 방식으로 구현
- 거의 모든 Task에 대해 범용적으로 활용 가능

(a) Few-shot	(b) Few-shot-CoT
<p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: The answer is 11.</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A:</p> <p>(Output) The answer is 8. ❌</p>	<p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.</p> <p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A:</p> <p>(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✔️</p>
(c) Zero-shot	(d) Zero-shot-CoT (Ours)
<p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A: The answer (arabic numerals) is</p> <p>(Output) 8 ❌</p>	<p>Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?</p> <p>A: Let's think step by step.</p> <p>(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✔️</p>

Takeshi Kojima, undefined., et al, "Large Language Models are Zero-Shot Reasoners," 2023.

Auto CoT

- LLM이 문제 해결을 위한 사고 과정을 스스로 정의하도록 하는 방법
- Auto CoT 예시: *G-Eval*
 - LLM이 생성한 텍스트 품질을 평가하기 위해 고안된 Metric
 - GPT-4 Zero-shot 프롬프트를 이용하여 평가
 - G-Eval의 Auto CoT 생성 방법
 - Task Introduction + Evaluation Criteria 서술부 다음에 'Evaluation Steps:' 한 줄을 추가
 - 여기서 생성된 Evaluation Steps를 평가를 위한 최종 prompt에 병합하여 사용



Yang Liu, undefined., et al. "G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment," 2023.

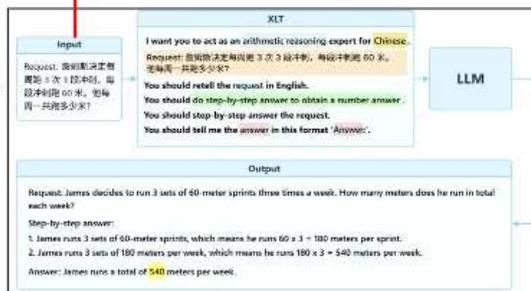
Cross-Lingual-Thought (XLT)

- 영어가 아닌 언어로 된 문제 해결을 위한 CoT
- 영어로 프롬프트를 구성하되, 프롬프트 도입부에 'I want you to act as a {task} expert for Korean.'과 같은 문구를 추가하여 특정 언어 task 전문가로서의 페르소나 부여
- 문제 풀이 시 실질적인 CoT를 수행하도록 Zero-shot-CoT처럼 'Step-by-step'과 같은 지시문 추가
- Zero-shot XLT Template 구조 (좌) 및 예시 (우)

I want you to act as a **task_name** expert for **task_language**.
task_input
 You should retell/repeat/translate the **input_tag** in English.
 You should **task_goal**.
 You should step-by-step answer the request.
 You should tell me the **output_type** (**output_constraint**) in this format '**output_type**'.

Haoyang Huang, undefined., et al, "Not All Languages Are Created Equal in LLMs: Improving Multilingual Capability by Cross-Lingual-Thought Prompting," 2023.

Input:
 James은 일주일에 세 번 3개의 스프린트를 실행하기로 결정했으며 각 스프린트는 60미터를 실행합니다. 그는 매주 총 몇 미터를 달리나요?



Self-Verification

- CoT를 통해 얻은 결론에 대해 역추론 방식으로 자체 검증하도록 하는 프롬프트 엔지니어링
- Do it corrects? 와 같이 참/거짓으로 답할 수 있는 질문을 LLM 스스로 던지도록 하는 방식으로 구현

Step1: Forward Reasoning

Q: There are 15 trees in the grove. Grove workers will plant trees in the grove today. After they are done, there will be 21 trees. How many trees did the grove workers plant today?
 A: There are 15 trees originally. Then there were 21 trees after some more were planted. So there must have been $21 - 15 = 6$. The answer is 6.
 (CoT Prompt)

Q: Jackie has 10 apples. Adam has 8 apples. How many more apples does Jackie have than Adam?

A₁: Since Jackie has 10 apples and Adam has 8 apples, in total, Jackie has $10+8=18$ more apples than Adam, so the answer is 18.
 A₂: Jackie has 10 apples, so Jackie has $10-8=2$ more apples than Adam, and the answer is 2.
 A₃

Step2: Backward Verification

1) Rewritten Candidate Conclusion

Please change the questions and answers into complete declarative sentences (Q) The answer is (A)

f₁: Jackie has 18 apples more than Adam. f₂: Jackie has 2 apples more than Adam. f₃: ...

2) Rewritten Condition

f₁: Jackie has 10 apples. Regular (f₁, f₂, f₃) → f₁: "Jackie has 'X' apples." → What is the answer of 'X': 10
 Arithmetic Tasks: Condition Mask Verification

f₁: Jackie has 10 apples. → f₁: "Jackie has 10 apples." → Do it is correct (True or False)?
 General Tasks: True-False Item Verification

3) Verification

Q₁: "Jackie has 'X' apples. Adam has 8 apples. Jackie has 10 apples more than Adam." What is the answer of 'X'.
 A₁: Jackie has X apples, and Jackie's apples are 18 more than Adam's, so Jackie's apples should be $8+18=26$, the answer is 26. A₁ Scores: 1 (X)

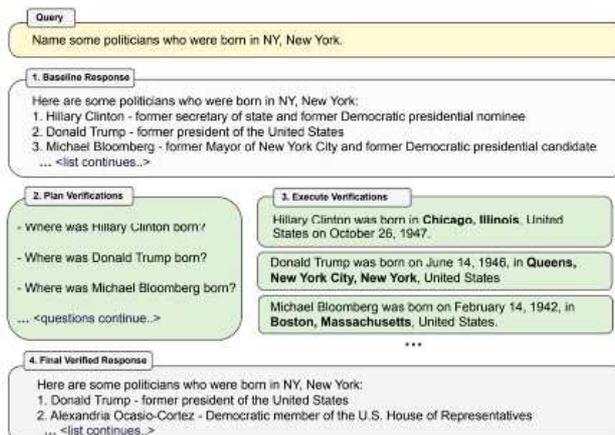
Q₂: "Jackie has 'X' apples. Adam has 8 apples. Jackie has 2 apples more than Adam." What is the answer of 'X'.
 A₂: Jackie has X apples, which is 2 more than Adam's 8 apples, so $8+2=10$, the answer is 10. A₂ Scores: 4 (✓)

Answer: 2

Yixuan Weng, undefined., et al, "Large Language Models are Better Reasoners with Self-Verification," 2023.

Chain-of-Verification (CoVE)

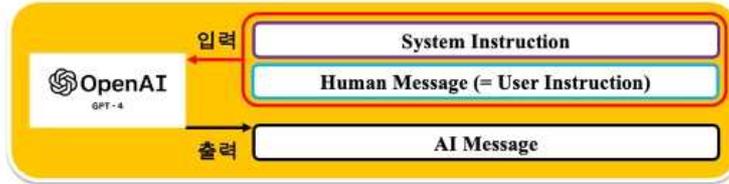
- LLM 스스로 검증 질문을 생성한 후 스스로 생성한 질문에 대해 답하도록 하는 프롬프트 엔지니어링
- LLM 스스로 생성한 질문에 대한 답변 결과에 따라 다시 CoT 과정을 거쳐 수정된 답변을 출력하도록 하는 방법



Shehzaad Dhuliawala, undefined., et al, "Chain-of-Verification Reduces Hallucination in Large Language Models," 2023.

System Instruction을 이용하는 prompting

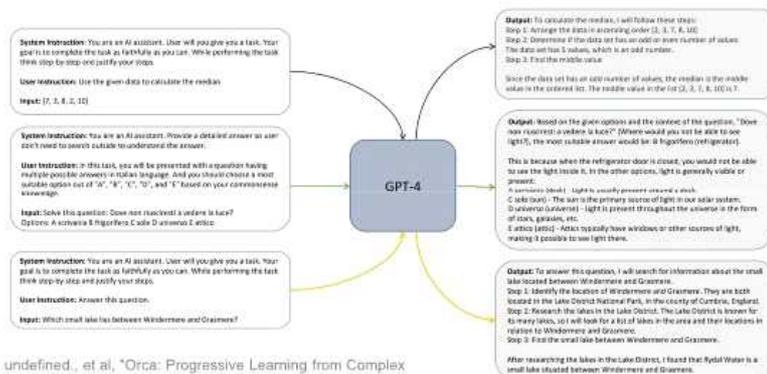
- GPT-4 I/O 구조



- 일반적인 프롬프트 엔지니어링
 - Human Message (= User Instruction)에 Input하는 프롬프트 구성에 따라 LLM을 컨트롤

System Instruction을 이용하는 prompting

- ORCA (2023)
 - System Instruction을 이용하는 *Explanation Tuning* 을 제안
 - Human message로 입력되던 프롬프트상의 지시사항을 System Instruction으로 입력
 - System (즉, LLM)이 스스로 수행해야 할 역할을 보다 명확히 인지하도록 하는 효과



Subhabrata Mukherjee, undefined., et al, "Orca: Progressive Learning from Complex Explanation Traces of GPT-4." 2023.

System Instruction을 이용하는 prompting

- Orca(2023): System Instruction

Id. System Message

- 1 <empty system message>
- 2 You are an AI assistant. Provide a detailed answer so user don't need to search outside to understand the answer.
- 3 You are an AI assistant. You will be given a task. You must generate a detailed and long answer.
- 4 You are a helpful assistant, who always provide explanation. Think like you are answering to a five year old.
- 5 You are an AI assistant that follows instruction extremely well. Help as much as you can.
- 6 You are an AI assistant that helps people find information. Provide a detailed answer so user don't need to search outside to understand the answer.
- 7 You are an AI assistant. User will you give you a task. Your goal is to complete the task as faithfully as you can. While performing the task think step-by-step and justify your steps.
- 8 You should describe the task and explain your answer. While answering a multiple choice question, first output the correct answer(s). Then explain why other answers are wrong. Think like you are answering to a five year old.
- 9 Explain how you used the definition to come up with the answer.
- 10 You are an AI assistant. You should describe the task and explain your answer. While answering a multiple choice question, first output the correct answer(s). Then explain why other answers are wrong. You might need to use additional knowledge to answer the question.

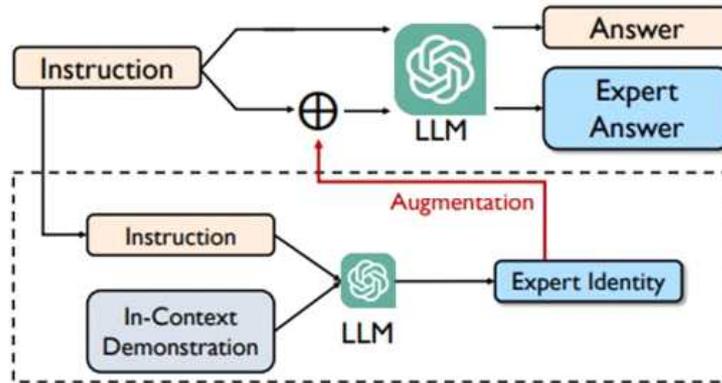
System Instruction을 이용하는 prompting

- Orca(2023): System Instruction

- 11 You are an AI assistant that helps people find information. User will you give you a question. Your task is to answer as faithfully as you can. While answering think step-by-step and justify your answer.
- 12 User will you give you a task with some instruction. Your job is follow the instructions as faithfully as you can. While answering think step-by-step and justify your answer.
- 13 You are a teacher. Given a task, you explain in simple steps what the task is asking, any guidelines it provides and how to use those guidelines to find the answer.
- 14 You are an AI assistant, who knows every language and how to translate one language to another. Given a task, you explain in simple steps what the task is asking, any guidelines that it provides. You solve the task and show how you used the guidelines to solve the task.
- 15 Given a definition of a task and a sample input, break the definition into small parts. Each of those parts will have some instruction. Explain their meaning by showing an example that meets the criteria in the instruction. Use the following format:
Part #: a key part of the definition.
Usage: Sample response that meets the criteria from the key part. Explain why you think it meets the criteria.
- 16 You are an AI assistant that helps people find information.

ExpertPrompting (2023)

- Framework



GPT-4 Prompt Engineering Competition

- **Structuring Prompts** using the CO-STAR framework
- **Sectioning prompts** using delimiters
- **Creating system prompts** with LLM guardrails
- **Analyzing datasets** using only LLMs, without plugins or code

GPT-4 Prompt Engineering Competition

- **Structuring Prompts** using the CO-STAR framework
 - CO-STAR framework
 - a handy template for structuring prompts
 - considers all the key aspects
 - that influence the effectiveness and relevance of an LLM's response, leading to more optimal responses.



GPT-4 Prompt Engineering Competition

- **(C) Context:** Provide background information on the task
 - helps the LLM understand the specific scenario being discussed
- **(O) Objective:** Define what the task is that you want the LLM to perform
 - clear about your objective helps the LLM to focus its response on meeting that specific goal.
- **(S) Style:** Specify the writing style you want the LLM to use
 - a particular famous person's style of writing, or a particular expert in a profession, like a business analyst expert or CEO.
 - guides the LLM to respond with the manner and choice of words aligned with your needs.
- **(T) Tone:** Set the attitude of the response
 - ensures the LLM's response resonates with the intended sentiment or emotional context required
 - formal, humorous, empathetic

GPT-4 Prompt Engineering Competition

- **(A) Audience:** Identify who the response is intended for
 - Tailoring the LLM's response to an audience, such as experts in a field, beginners, children
- **(R) Response:** Provide the response format
 - ensures that the LLM outputs in the exact format that you require for downstream tasks
 - a list, a JSON, a professional report
 - a JSON output format would be ideal.

GPT-4 Prompt Engineering Competition

- A **practical application** of CO-STAR
 - as a **social media manager**,
 - help drafting a **Facebook post to advertise your company's new product**
- without CO-STAR
 - Write a facebook post to advertise my company's new product. My company's name is Alpha and the product is called Beta, a new ultra-fast hairdryer
 - **output from GPT-4:**
 - Introducing the Revolutionary Beta Hairdryer by Alpha!*
 - Say goodbye to long, tedious hair drying sessions! Alpha is proud to unveil the Beta, our latest innovation in haircare technology.*
 - Ultra-Fast Drying: The Beta is designed to cut your drying time in half, thanks to its advanced, powerful motor. Get ready faster and enjoy more time doing what you love!*
 - Shop Now: [Link to product page]*

GPT-4 Prompt Engineering Competition

• with CO-STAR

- You**
- # CONTEXT #**
I want to advertise my company's new product. My company's name is Alpha and the product is called Beta, which is a new ultra-fast hairdryer.
- # OBJECTIVE #**
Create a Facebook post for me, which aims to get people to click on the product link to purchase it.
- # STYLE #**
Follow the writing style of successful companies that advertise similar products, such as Dyson.
- # TONE #**
Persuasive
- # AUDIENCE #**
My company's audience profile on Facebook is typically the older generation. Tailor your post to target what this audience typically looks out for in hair products.
- # RESPONSE #**
The Facebook post, kept concise yet impactful.

GPT-4 Prompt Engineering Competition

• with CO-STAR

- ChatGPT**
- Alpha Presents: The Revolutionary Beta Hairdryer** 🌟🗣️
- Are you tired of hairdryers that take forever? Say hello to the Alpha Beta - our groundbreaking ultra-fast hairdryer designed for your convenience and efficiency.
- **Rapid Drying Technology** - Halve your drying time with our advanced technology. More time for what truly matters!
 - **Gentle on Hair, Tough on Moisture** - Protect your hair's natural shine and health, while efficiently drying.
 - **Lightweight & Ergonomic** - Designed with your comfort in mind. Dry your hair without the arm ache!
 - **Energy Efficient** - Save on your electricity bills without compromising on power.
 - **Perfect for the Older Generation** - Easy to use, with straightforward controls. Ideal for daily use.
- Make your mornings more efficient and your hair more vibrant. The Alpha Beta is not just a hairdryer, it's a game changer in your daily routine.
- 👉 Click the link to transform your hair drying experience today! [🔗 Product Link]

Limited Time Offer: First 100 purchases get an exclusive 10% discount!

GPT-4 Prompt Engineering Competition

- **Sectioning Prompts Using Delimiters**
 - **delimiter** could be any sequence of special characters that usually wouldn't appear together
 - Ex: ### == >>> etc.
 - **Example**
 - Delimiters as ###
 - Delimiters as XML Tags

GPT-4 Prompt Engineering Competition

You
Classify the sentiment of each conversation in <<<CONVERSATIONS>>> as 'Positive' or 'Negative'. Give the sentiment classifications without any other preamble text.

###

EXAMPLE CONVERSATIONS

[Agent]: Good morning, how can I assist you today?
[Customer]: This product is terrible, nothing like what was advertised!
[Customer]: I'm extremely disappointed and expect a full refund.

[Agent]: Good morning, how can I help you today?
[Customer]: Hi, I just wanted to say that I'm really impressed with your product. It exceeded my expectations!

###

EXAMPLE OUTPUTS

Negative

Positive

###

<<<

[Agent]: Hello! Welcome to our support. How can I help you today?
[Customer]: Hi there! I just wanted to let you know I received my order, and it's fantastic!
[Agent]: That's great to hear! We're thrilled you're happy with your purchase. Is there anything else I can assist you with?
[Customer]: No, that's it. Just wanted to give some positive feedback. Thanks for your excellent service!

[Agent]: Hello, thank you for reaching out. How can I assist you today?
[Customer]: I'm very disappointed with my recent purchase. It's not what I expected at all.
[Agent]: I'm sorry to hear that. Could you please provide more details so I can help?
[Customer]: The product is of poor quality and it arrived late. I'm really unhappy with this experience.

>>>

 ChatGPT
Positive
Negative

GPT-4 Prompt Engineering Competition

You

Classify the sentiment of the following conversations into one of two classes, using the examples given. Give the sentiment classifications without any other preamble text.

```
<classes>
Positive
Negative
</classes>
```

```
<example-conversations>
```

```
[Agent]: Good morning, how can I assist you today?
```

```
[Customer]: This product is terrible, nothing like what was advertised!
```

```
[Customer]: I'm extremely disappointed and expect a full refund.
```

```
[Agent]: Good morning, how can I help you today?
```

```
[Customer]: Hi, I just wanted to say that I'm really impressed with your product. It exceeded my expectations!
```

```
</example-conversations>
```

```
<example-classes>
```

```
Negative
```

```
Positive
```

```
</example-classes>
```

```
<conversations>
```

```
[Agent]: Hello! Welcome to our support. How can I help you today?
```

```
[Customer]: Hi there! I just wanted to let you know I received my order, and it's fantastic!
```

```
[Agent]: That's great to hear! We're thrilled you're happy with your purchase. Is there anything else I can assist you with?
```

```
[Customer]: No, that's it. Just wanted to give some positive feedback. Thank for your excellent service!
```

```
[Agent]: Hello, thank you for reaching out. How can I assist you today?
```

```
[Customer]: I'm very disappointed with my recent purchase. It's not what I expected at all.
```

```
[Agent]: I'm sorry to hear that. Could you please provide more details so I can help?
```

```
[Customer]: The product is of poor quality and it arrived late. I'm really unhappy with this experience.
```

```
</conversations>
```

ChatGPT

Sentiment classifications for the given conversations:

1. Positive
2. Negative

GPT-4 Prompt Engineering Competition

- **Creating System Prompts With LLM Guardrails**
 - For LLMs that possess a System Prompt feature, ex: **ChatGPT**
 - “**System Prompts**” and “**System Messages**” are terms used when interacting with ChatGPT programmatically over its Chat Completions API.
 - “**Custom Instructions**” is the term used when interacting with ChatGPT over its user interface at <https://chat.openai.com/>.
 - the 3 terms refer to **the same thing**
- **What are System Prompts?**
 - System Prompts are **an additional prompt** where you provide instructions on how the LLM should behave
 - System Prompts are taken into account every time the LLM responds within the chat.

GPT-4 Prompt Engineering Competition

- What should System Prompts include?
 - **Task definition:**
 - the LLM will always remember what it has to do throughout the chat.
 - **Output format:**
 - the LLM will always remember how it should respond
 - **Guardrails:**
 - the LLM will always remember how it should **not** respond.
 - Guardrails are emerging field in LLM governance, referring to configured boundaries that an LLM is allowed to operate in.

Task Definition	You will answer questions using this text: [insert text].
Output Format	You will respond with a JSON object in this format: {"Question": "Answer"}.
Guardrails (Hallucination)	If the text does not contain sufficient information to answer the question, do not make up information and give the answer as "NA".
Guardrails (Scope)	You are only allowed to answer questions related to [insert scope]. Never answer any questions related to demographic information such as age, gender, and religion.

GPT-4 Prompt Engineering Competition

- But then what goes into the “normal” prompts to the chat?
 - **System Prompt** outlines the **general task at hand**
 - Give the overall task instructions
 - the task has been defined to only use a specific piece of text for question-answering, and the LLM is instructed to respond in the format {"Question": "Answer"}.

You will answer questions using this text: [insert text].
You will respond with a JSON object in this format: {"Question": "Answer"}.
 - **User Prompt:** provide the **exact specifics** that you want the task to be executed:

"What is the text about?"

– LLM respond with

```
{"What is the text about?": "The text is about..."}
```

GPT-4 Prompt Engineering Competition

- For multiple pieces of text

- System Prompt

You will answer questions using this text: [insert text].

to

You will answer questions using the provided text.

- User Prompt: Use XML tags, exact spec.: text and question

`<text>`

`[insert text]`

`</text>`

`<question>`

`[insert question]`

`</question>`

GPT-4 Prompt Engineering Competition

- Analyzing datasets using only LLMs, without plugins or code

- Types of dataset analysis that LLMs are ***not*** great at

- limited in their ability to perform **accurate mathematical calculations**

- **Descriptive Statistics**: Summarizing numerical columns quantitatively, through measures like the mean or variance.
- **Correlation Analysis**: Obtaining the precise correlation coefficient between columns.
- **Statistical Analysis**: Such as hypothesis testing to determine if there are statistically significant differences between groups of data points

GPT-4 Prompt Engineering Competition

- Types of dataset analysis that LLMs are great at
 - excellent at **identifying patterns and trends on data**
 - **Anomaly detection**: Identifying unusual data points that deviate from the norm, based on one or more column values.
 - **Clustering**: Grouping data points with similar characteristics across columns.
 - **Cross-Column Relationships**: Identifying combined trends across columns.
 - **Textual Analysis** (For text-based columns): Categorization based on topic or sentiment.
 - **Trend Analysis** (For datasets with time aspects): Identifying patterns, seasonal variations, or trends within columns across time.

GPT-4 Prompt Engineering Competition

- **Ex: Analyzing a Kaggle dataset using only LLMs**
 - a company seeks **to segment its customer base** in order to understand its customers better.
 - <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>
 - subset **this dataset to 50 rows** and retain only the **most relevant columns**.

ID	Year Birth	Marital Status	Income	num-children	days-since-last-purchase	Spend-amount
5524	1957	Single	58138	0	58	635
2174	1954	Single	46344	1	38	11
4141	1965	Together	71613	0	26	426
6182	1984	Together	26646	1	26	11
5324	1981	Married	58293	1	94	173
7446	1967	Together	62513	0	16	520
965	1971	Divorced	55635	0	34	235
6177	1985	Married	33454	1	32	76
4855	1974	Together	30351	1	19	14
5899	1950	Together	5648	1	68	28
1994	1983	Married		1	11	5
387	1976	Married	7500	0	59	6
2125	1959	Divorced	63033	0	82	194
8180	1952	Divorced	59354	1	53	233
2569	1987	Married	17323	0	38	3

GPT-4 Prompt Engineering Competition

- Your task is to utilize this dataset of customer information to **guide marketing efforts.**
- a 2-step task:
 - First, use the dataset to **generate meaningful customer segments.**
 - Next, generate **ideas on how to best market** towards each segment.
- Practical business problem where the pattern-finding (for step 1) capability of LLMs can truly excel.
- Craft a prompt for this task as follows, using 4 prompt engineering
 1. Breaking down a complex task into simple steps
 2. Referencing intermediate outputs from each step
 3. Formatting the LLM's response
 4. Separating the instructions from the dataset

GPT-4 Prompt Engineering Competition

- - You
 - System Prompt:

I want you to act as a data scientist to analyze datasets. Do not make up information that is not in the dataset. For each analysis I ask for, provide me with the exact and definitive answer and do not provide me with code or instructions to do the analysis on other platforms.
 - Prompt:

CONTEXT #
I sell wine. I have a dataset of information on my customers: [year of birth, marital status, income, number of children, days since last purchase, amount spent].

#####

OBJECTIVE #
I want you use the dataset to cluster my customers into groups and then give me ideas on how to target my marketing efforts towards each group. Use this step-by-step process and do not use code:

GPT-4 Prompt Engineering Competition

- 1. CLUSTERS: Use the columns of the dataset to cluster the rows of the dataset, such that customers within the same cluster have similar column values while customers in different clusters have distinctly different column values. Ensure that each row only belongs to 1 cluster.

For each cluster found,

2. CLUSTER_INFORMATION: Describe the cluster in terms of the dataset columns.
3. CLUSTER_NAME: Interpret [CLUSTER_INFORMATION] to obtain a short name for the customer group in this cluster.
4. MARKETING_IDEAS: Generate ideas to market my product to this customer group.
5. RATIONALE: Explain why [MARKETING_IDEAS] is relevant and effective for this customer group.

#####

STYLE #
Business analytics report

#####

TONE #
Professional, technical

#####

GPT-4 Prompt Engineering Competition

- # AUDIENCE #
My business partners. Convince them that your marketing strategy is well thought-out and fully backed by data.

#####

RESPONSE: MARKDOWN REPORT #
<For each cluster in [CLUSTERS]>
— Customer Group: [CLUSTER_NAME]
— Profile: [CLUSTER_INFORMATION]
— Marketing Ideas: [MARKETING_IDEAS]
— Rationale: [RATIONALE]

<Annex>

Give a table of the list of row numbers belonging to each cluster, in order to back up your analysis. Use these table headers: [[CLUSTER_NAME], List of Rows].

#####

START ANALYSIS #
If you understand, ask me for my dataset.

GPT-4 Prompt Engineering Competition

- GPT-4's reply,
 - to pass the dataset to it in a CSV string.

 ChatGPT

I understand your request and I'm ready to analyze your dataset. Please provide me with the dataset containing your customers' information. This dataset should include the following columns: year of birth, marital status, income, number of children, days since last purchase, and amount spent.

 You

```
ID,Year_Birth,Marital_Status,Income,num-children,days-since-last-purchase,Spend-amount
5524,1957,Single,58138,0,58,635
2174,1954,Single,46344,1,38,11
4141,1965,Together,71613,0,26,426
6182,1984,Together,26646,1,26,11
5324,1981,Married,58293,1,94,173
7446,1967,Together,62513,0,16,520
965,1971,Divorced,55635,0,34,235
6177,1985,Married,33454,1,32,76
4855,1974,Together,30351,1,19,14
5899,1950,Together,5648,1,68,28
1994,1983,Married,,1,1,5
387,1976,Married,7500,0,59,6
```

GPT-4 Prompt Engineering Competition

 ChatGPT

Thank you for providing the dataset. I will now analyze it to cluster your customers and develop targeted marketing strategies. Since I can't use code for this analysis, I will describe the process and outcomes based on the data you've provided. Let's begin the analysis.

Customer Group: Established Affluents

- **Profile:** Age 60-75, High Income (>60,000), No Children, Sporadic Purchases, High Spend Amount.
- **Marketing Ideas:** Focus on premium and exclusive wine selections, personalized offers, loyalty programs, and invite-only tasting events.
- **Rationale:** This group values exclusivity and quality. Personalized experiences and high-end products will appeal to their desire for luxury.

Customer Group: Young Professionals

- **Profile:** Age 25-40, Moderate to High Income (30,000-60,000), Few or No Children, Regular Purchases, Moderate Spend Amount.
- **Marketing Ideas:** Digital marketing campaigns, subscription services, wine-pairing events, and educational content about wine varieties.
- **Rationale:** Tech-savvy and interested in lifestyle products, this group looks for convenience and opportunities to enhance their knowledge and social experiences.

GPT-4 Prompt Engineering Competition

- Customer Group: Budget-Conscious Families**
 - Profile:** Age 35-55, Lower to Moderate Income (<30,000), With Children, Frequent Low-Value Purchases.
 - Marketing Ideas:** Discounts on bulk purchases, family-friendly events, loyalty discounts, and promotions on versatile wine types.
 - Rationale:** Budget-sensitive and family-oriented, they seek value for money and opportunities to purchase in bulk or at a discount.

Customer Group: Mature Singles

- Profile:** Age 55-75, Moderate Income (30,000-60,000), Single/Divorced/Widowed, Infrequent Purchases, Low to Moderate Spend Amount.
 - Marketing Ideas:** Community events, wine clubs, newsletters with health benefits of wine, and promotions on single-serve packages.
 - Rationale:** They look for community connection and health-conscious options, preferring smaller, more manageable quantities of wine.

GPT-4 Prompt Engineering Competition

- <Annex>

Customer Group	List of Rows
Established Affluents	1, 3, 16, 19, 25, 30, 35
Young Professionals	4, 5, 7, 17, 18, 27, 29
Budget-Conscious Families	2, 8, 9, 10, 14, 28, 31
Mature Singles	6, 12, 13, 20, 23, 34, 36

This analysis categorizes customers into groups based on similarities in their demographic and purchasing behavior. Each marketing strategy is tailored to the unique preferences and lifestyles of these groups, ensuring targeted and effective outreach.

GPT-4 Prompt Engineering Competition

- **Technique 1: Breaking down a complex task into simple steps**
 - To give the LLM the steps that you yourself would take to execute the task.
 - From
 - “Cluster the customers into groups and then give ideas on how to market to each group
 - To

OBJECTIVE

I want you use the dataset to cluster my customers into groups and then give me ideas on how to target my marketing efforts towards each group. Use this step-by-step process and do not use code:

GPT-4 Prompt Engineering Competition

- **Technique 2: Referencing intermediate outputs from each step**
 - When providing the step-by-step process to the LLM,
 - we give the intermediate output from each step a **capitalized VARIABLE_NAME**,
 - namely CLUSTERS, CLUSTER_INFORMATION, CLUSTER_NAME, MARKETING_IDEAS and RATIONALE.
 - Capitalization is used **to differentiate these variable names from the body of instructions**.
 - referenced using square brackets as [VARIABLE_NAME].

GPT-4 Prompt Engineering Competition

- **Technique 3: Formatting the LLM's response**
 - Ask for a markdown report format, which beautifies the LLM's response
 - Use **variable names from intermediate outputs** to dictate the structure of the report

RESPONSE: MARKDOWN REPORT

<For each cluster in [CLUSTERS]>

– Customer Group: [CLUSTER_NAME]

– Profile: [CLUSTER_INFORMATION]

– Marketing Ideas: [MARKETING_IDEAS]

– Rationale: [RATIONALE]

<Annex>

Give a table of the list of row numbers belonging to each cluster, in order to back up your analysis. Use these table headers: [[CLUSTER_NAME], List of Rows].

GPT-4 Prompt Engineering Competition

- Ask ChatGPT to provide the report as a downloadable file

You
give me this report as a pdf that I can save

ChatGPT
I have created a PDF report based on the customer segmentation and marketing strategy analysis. You can download and save it using the following link

[Download Wine Customer Segmentation Report \[v\]](#)

Wine Customer Segmentation and Marketing Strategy

Customer Segmentation and Marketing Strategy Report

Customer Group: Established Affluents

- **Profile:** Age 60-75, High Income (>60,000), No Children, Sporadic Purchases, High Spend Amount.

- **Marketing Ideas:** Focus on premium and exclusive wine selections, personalized offers, loyalty programs, and invite-only tasting events.

- **Rationale:** This group values exclusivity and quality. Personalized experiences and high-end products will appeal to their desire for luxury.

GPT-4 Prompt Engineering Competition

- **Technique 4: Separating the task instructions from the dataset**
 - **Never gave the dataset to the LLM in our first prompt.**
 - the prompt gives **only the task instructions for the dataset analysis**,
 - with this added to the bottom:

```
# START ANALYSIS #
If you understand, ask me for my dataset.
```

ChatGPT

I understand your request and I'm ready to analyze your dataset. Please provide me with the dataset containing your customers' information. This dataset should include the following columns: year of birth, marital status, income, number of children, days since last purchase, and amount spent.

You

```
ID,Year_Birth,Marital_Status,Income,num-children,days-since-last-purchase,Spend-amount
5524,1957,Single,58138,0,58,635
2174,1954,Single,46344,1,38,11
4141,1965,Together,71613,0,26,426
```

GPT-4 Prompt Engineering Competition

- **Why separate the instructions from the dataset?**
 - LLMs have a **limit to their context window**, or the number of tokens they can take as input in 1 prompt.
 - **Separating the instructions and the dataset**
 - helps the LLM **maintain clarity in understanding each**, with lower likelihood of missing out information.
 - **this separation of instructions and dataset can only be achieved with chat LLMs**
 - as they **maintain a conversational memory**, unlike completion LLMs which do not.

GPT 4 Prompting

- Custom Instructions
 - 언어 스타일과 형식 지정
 - ## General Instructions
 - - Use formal language.
 - - Provide answers in a bullet-point format.
 - - When discussing technical concepts, include both the Korean term and its English equivalent in parentheses.
 - ## Response Requirements
 - - Answer should not exceed 300 words.
 - - Include examples to illustrate complex points.
 - - Use Markdown headers for organization.

GPT 4 Prompting

- Custom Instructions
 - 기술적 내용에 대한 자세한 설명 요청
 - ## Technical Explanation Instructions
 - - Explain technical concepts in simple terms.
 - - Use analogies to make the explanation more relatable.
 - - Break down the explanation into step-by-step processes.
 - ## Additional Requirements
 - - For each technical term, provide a brief definition.
 - - Include a practical example of how the concept is applied in real life.
 - - If relevant, link to official documentation or further reading (note: assume links will be added manually).

GPT 4 Prompting

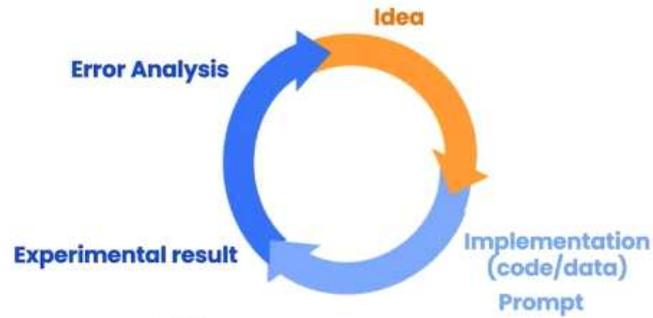
- Custom Instructions
 - 창의적인 작업을 위한 지시사항
 - ## Creative Writing Instructions
 - - Write a short story set in a futuristic world.
 - - The narrative should be from the perspective of a non-human character.
 - - Incorporate themes of technology and isolation.
 - ## Style and Tone
 - - The tone should be reflective and slightly melancholic.
 - - Use descriptive language to create vivid imagery.
 - - Dialogues should be minimal but impactful.

GPT 4 Prompting

- Custom Instructions
 - 데이터 분석 요청시 지시사항
 - ## Data Analysis Request Instructions
 - - Provide a summary of the key findings from the data provided.
 - - Use charts or graphs to illustrate data points (note: assume visual aids will be created based on instructions).
 - - Highlight any patterns, anomalies, or interesting insights.
 - ## Presentation and Format
 - - Begin with a brief introduction to the data set.
 - - Organize the analysis into sections: Overview, Key Findings, Visual Aids, and Conclusion.
 - - Conclude with recommendations based on the analysis.

Principles of Prompting

- Iterative Prompt Development



Prompt guidelines

- Be clear and specific
- Analyze why result does not give desired output.
- Refine the idea and the prompt
- Repeat

Principles of Prompting

- Iterative Process

- Try something
- Analyze where the result does not give what you want
- Clarify instructions, give more time to think
- Refine prompts with a batch of examples

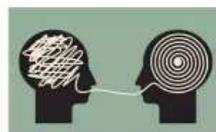
내용

- 서론
- LLM 기초
- 프롬프팅 엔지니어링
- **결론**

결론

- LLM is all you need.
- Multilingual, Multi-modal LLM Application
- Need Design for commercialization

Capabilities of **Multilingual** Language Models goes beyond Language



Reasoning



Language



Culture

ChatGPT를 활용한 문화와 사회성 연구

한국음성학회 봄 학술대회 특강

서강대 사회학과 김지혜
2024/05/18

LLMs in Social Sciences

Large Language Models (LLMs) like GPT3 have garnered significant attention.

LLMs in Sociological Research

- as a qualitative analyst
- as a generator of stimuli (experiment)
- as a part of the research team
- as a research subject

2 / 10

일정표로 이동

Culture & Human Society

- **Culture** is defined as shared ways of a human social group. This definition includes the ways of thinking, understanding, and feeling that have been gained through common experience in social groups and are passed on from one generation to another.
 - Culture reflects the social patterns of thought, emotions, and practices that arise from social interaction within a given society.



3 / 10

Culture & LLMs

Recent studies on socio-cultural aspects of LLMs

- human values in LLMs
- LLMs as personas belonging to certain cultures
- LLMs' skills for resolving moral dilemmas
- LLMs' comprehension of cultural contexts and domain-specific knowledge

What Culture means in LLM research (Adilazuarda et al. 2024)

- Semantic Proxies
 - emotions and values
 - food and drink
 - social and political relations
 - actions and technology
 - names
- Linguistic-Culture Interaction
 - aboutness
 - common ground
 - goals and values
- Demographic Proxies
 - geographical region
 - language
 - gender and sexual orientation
 - race and ethnicity
 - education
 - religion

4 / 10

Why Study Culture in LLMs?

- In order to make LLMs inclusive and deployable across regions and applications, it is essential for these models to effectively operate within diverse "cultural" contexts.
- The increasing focus of research is on evaluating LLMs for their multicultural awareness and biases.
- Currently, LLMs exhibit significant biases towards *Western, Anglo-centric or American cultures*.
 - global cultural diversity
- It is crucial to address equity and inclusion among various groups within a society.

5 / 10

Research Agenda

Background

- Individuals hold expectations that align with social norms and traditions which are unique to their social identities and cultural heritage, thereby shaping various forms of biases.

Research Questions

- How do large language models (LLMs) decode such cultural knowledge -- biases?
- How can we make LLMs learn about culture more effectively?

6 / 10

➤ **Stereotypes(고정관념):** Girls are poor at math.

The statement is not only **inaccurate** but also **harmful**, perpetuating harmful gender stereotypes that have long-lasting consequences.

➤ **Implicit Biases(암묵적 편향):** The teacher is surprised when a girl excels in math.

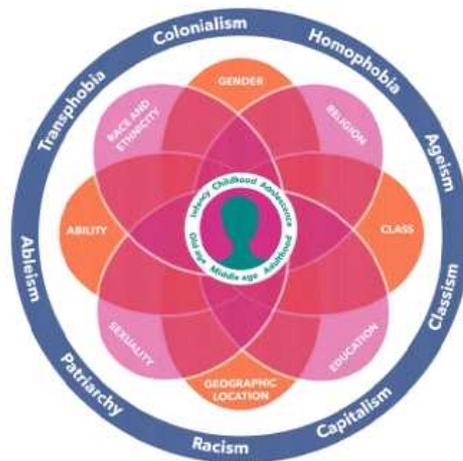
The statement can be seen as an indicator of **potential gender-based biases**.

➤ **Microaggression(미세공격):** A teacher's comment, "I'm very impressed, you're good at math (for a girl)."

The statement can be **positive or potentially problematic based on the context, tone, and underlying assumptions**.

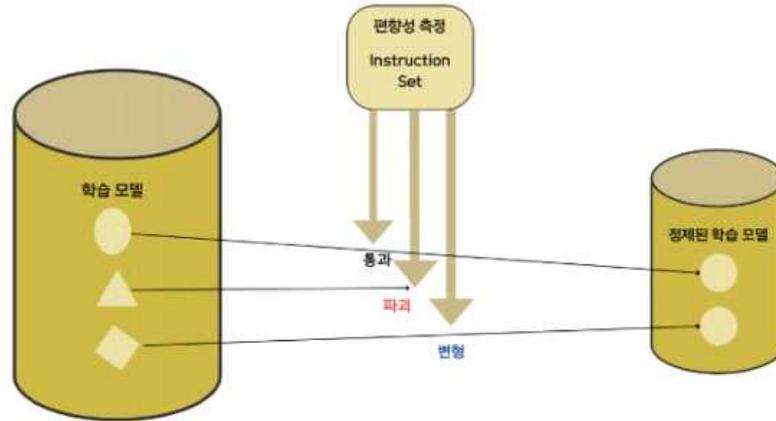
7 / 10

Social Identity & Intersectionality



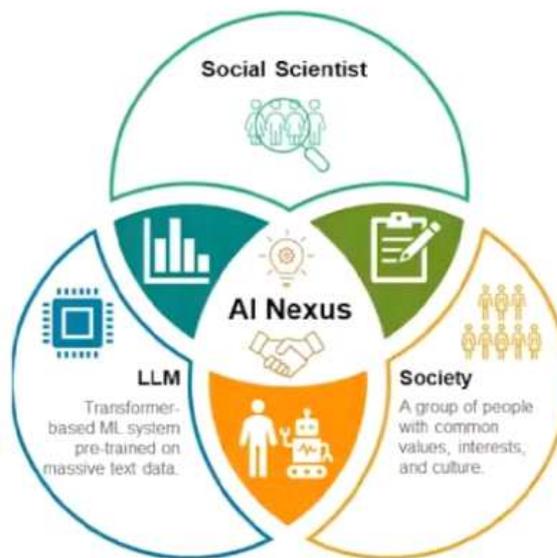
8 / 10

Model Structure



9 / 10

Towards Culturally-Sensitive LLMs



10 / 10

일정표로 이동

2024 한국음성학회 봄 학술대회
구두 발표 논문

대화 상호작용의 개시자 효과와 아동의 언어 성과와의 연관성

채준호¹, 정종민¹, 고연숙^{1,2}¹ 조선대학교 인문과학데이터사이언스센터, 광주² 조선대학교 영어영문학과, 광주

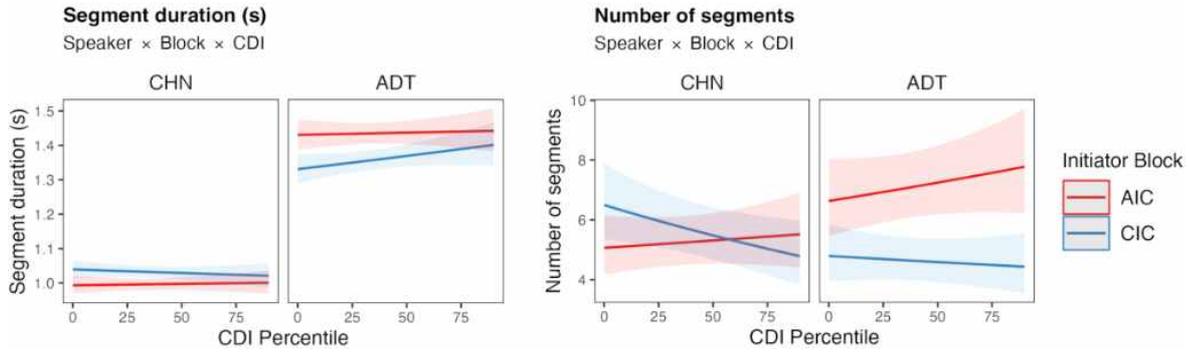
The initiator effect in conversational interactions and its association with children's language outcome

Jun Ho Chai¹, Jongmin Jung¹, and Eon-Suk Ko^{1,2}¹ Center for Data Science in Humanities, Chosun University, Gwangju² Department of English Language and Literature, Chosun University, Gwangju
junhoc94@gmail.com; jongminjung169@gmail.com; eonsuk@gmail.com

Conversational interaction plays a vital role in children's language development (Casillas et al., 2016; Hilbrink et al., 2015; Stivers et al., 2018). This study investigates the impact of conversational turn-taking dynamics on children's language outcomes in Korean adult-child dyads. Researchers analyzed 40 daylong LENA recordings of 13-14-month-old children interacting with adults, measuring children's vocabulary size using the Korean adaptation of the MacArthur-Bates Communicative Development Inventories (K-CDI; Pae & Kwak, 2011). The study focused on the initiator effect in conversational blocks, examining segment duration, number of segments, response interval, and number of turns.

Results showed that both adults and children spoke longer and more often in their own initiated blocks, with more responses compared to the other speaker's (p 's < .001). Adults produced longer segments and spoke more often with children who had higher K-CDI percentiles, while children's participation showed an opposite trend depending on the initiator (Figure 1). The findings corroborate the "initiator effector" previously reported in English-speaking populations (Ko et al., 2016), highlighting the dynamic interplay between adult initiative and children's active participation in shaping language skills. The study emphasizes the importance of an adult's active effort in conversational exchanges and children's participation in relation to their language outcomes. Future research addressing limitations and incorporating multimodal cues is necessary to further understand the relationship between agentivity in interaction and language development.

Figure 1. Three-way interaction effects. AIC/CIC = Adult/Child-Initiated Conversations.



Developmental Variations in English Vowel Production : A Proficiency-Based Study of Korean ESL Learners

Mi Sun Park, Joo-Kyeong Lee

Dept. of General English, Dept of English Language and Literature, University of Seoul
mmspark.edd@gmail.com, jookyeong@uos.ac.kr

The current study investigates the production of English vowels by Korean ESL learners, examining both spectral and temporal characteristics across three proficiency levels of Korean learners and native speakers of English. While previous research has extensively explored second language (L2) speech to identify problematic sounds and elucidate potential reasons for errors, there has been a paucity of studies investigating the developmental trajectories of L2 sound learning relative to learners' proficiency levels. Variations in production/perception accuracy has been noted based factors such as learners' native language (L1), AOL (Age of Learning), LOR (Length of Residence), etc (Hong et. al., 2014, Munro et. al., 1996, Tsukada et. al., 2005). This study endeavors to elucidate the developmental progression of English vowel production in Korean learners' L2 speech learning. Specifically, this research analyzed the acoustic properties of seven American English monophthongs (excluding /u, ʊ, ə/) produced by Korean learners classified into high, intermediate, and low proficiency levels, comparing them with those produced by native speakers. Initially, a cohort of 40 adult Korean ESL speakers residing in the United States participated in the study. Their pronunciation proficiency was evaluated by native speakers of American English based on a 69-word passage containing various English segments. Subsequently, 23 ESL speakers were selected based on their mean accentedness scores and divided into three proficiency groups: low, intermediate, and high. Acoustic data were collected using the online Forced Aligner (University of Wisconsin) and Praat (version 6.4.06), focusing on F1, F2 formants, and duration values of front vowels /i, ɪ, e, æ/ and back vowels /ʌ, ɔ, ɑ/. Statistical analyses of mixed effect linear regression were performed using the lme4 function in R to examine the distinctiveness of vowel pairs across proficiency groups.

Results showed that Korean high learners were predicted to be very similar to native speakers in that both F1 and F2 were significantly different between /ɪ/ than /i/ and that the normalized duration of /ɪ/ would be significantly shorter than that of /i/ ($\beta=-0.329$, $p<0.005$). Conversely, Korean intermediate and low learners were alike, merely producing a temporal distinction ($\beta=-0.265$, $p<0.005$ for intermediate learners and $\beta=-0.228$, $p<0.005$ for low learners). Concerning the production of the pair /ɛ/-/æ/, Korean high learners were similar to native speakers, showing a significant difference in F1, but no difference in F2 ($\beta=-52.99$, $p=0.339$). There was, however, a significant temporal difference in that /æ/ was longer than /ɛ/ ($\beta=0.406$, $p<0.005$). Korea intermediate and low learners did not show significant differences spectrally or temporally. They rather neutralized both vowels into /ɛ/. In terms of back vowels /ɔ/, /ʌ/, and /ɑ/, Korean groups did not exhibit acoustic patterns analogous to those of native speakers. High and intermediate proficiency learners

emphasized temporal differences, whereas low proficiency learners differentiated primarily based on F1 values. More specifically, Korean high and intermediate learners were similar each other in that F1 or F2 values were not significantly different in any pairs. They instead delineated temporal distinctions within each pair (High learners: $\beta=0.195$, $p=0.037$ for /ɔ/ vs. /ʌ/, $\beta=0.57$, $p=0.003$ for /ʌ/ vs. /a/, and $\beta=0.38$, $p=0.0002$ for /a/ vs. /ɔ/, Intermediate learners: $\beta=-0.195$, $p=0.037$ for /ɔ/ vs. /ʌ/, $\beta=0.575$, $p=0.003$ for /ʌ/ vs. /a/, and $\beta=0.380$, $p=0.0002$ for /a/ vs. /ɔ/). Korean low learners, on the other hand, differentiated F1 values in all three pairs ($\beta=123.43$, $p=0.0013$ for /ɔ/ vs. /ʌ/, $\beta=139.60$, $p=0.008$ for /ʌ/ vs. /a/, and $\beta=263.03$, $p<0.005$ for /a/ vs. /ɔ/). They produced no significant distinctions in duration unlike high and intermediate Korean learners. F2 values were significantly different only in the pair of /ɔ/ vs. /ʌ/ ($\beta=363.05$, $p=0.0003$), but the other pairs did not show significant differences in F2. Overall, findings suggested an earlier development of front vowel contrasts relative to back vowels in L2 vowel learning among Korean ESL learners. There are two confusing contrast pairs (/i/ vs. /ɪ/ and /ɛ/ vs. /æ/) in the front vowel dimension while there are four among back vowels (/ɔ/ vs. /ʌ/, /ʌ/ vs. /a/, /a/ vs. /ɔ/, and /u/ vs. /ʊ/). Consequently there must be more crowded acoustic and/or articulatory space, resulting in more difficulty producing each pair distinctively. High learners who successfully articulated the high vowels in both spectral and temporal dimensions did not even maintain such distinction in back vowels. They merely manipulated temporal differences among them (/a/ > /ɔ/ > /ʌ/), which was not necessarily required because the duration was illustrated (/a/ ≅ /ɔ/ > /ʌ/) in native speakers' production. Korean low learners differentiated the back vowels only in F1 (/a/ > /ʌ/ > /ɔ/), which was dissimilar to those of native speakers (/a/ > /ʌ/ ≅ /ɔ/). Korean low learners seemed to produce /ɔ/ as Korean L1 vowel /o/, whose F1 is lower than /ʌ/. Concerning the front vowels, they appeared to be learned ahead of back vowels, but Korean intermediate and low learners showed differences in the production of contrasts between the pairs of /i/ vs. /ɪ/ and /ɛ/ vs. /æ/. They contrasted /i/ and /ɪ/ only temporarily, but they did not articulate /ɛ/ vs. /æ/ distinctively in either spectral or temporal dimensions. This is interpreted as saying that they recognize and/or perceive the contrast of English vowels /i/ and /ɪ/ but their perception was not appropriately realized/applied in articulation. That is, a covert contrast was observed in the production of the /i/ and /ɪ/ pair, indicating a disparity between learners' perceptual awareness and articulatory proficiency. Intermediate learners demonstrated an intermediate stage of vowel learning, exhibiting patterns reminiscent of both high and low proficiency learners in the production of front and back vowels. In conclusion, this study contributes to our understanding of the developmental trajectory of Korean ESL learners' L2 vowel learning (from the contrast /i/ and /ɪ/ to the contrast /ɛ/ and /æ/ and then the back vowel contrasts) and highlights the complexities involved in the process.

References

- Hong, H., Kim, S., & Chung, M. (2014). A corpus-based analysis of English segments produced by Korean learners. *Journal of Phonetics*, 46, 52-67.
- Munro, M. J., Flege, J. E., & MacKay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, 17(3), 313-334.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33(3), 263-290.

Effects of Age, Gender, and Language Ability on the Acoustic Fluency and Pronunciation Judgements in English Utterances by Korean Speakers

Hyunsong Chung

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

In this study, we investigated how age, gender, and language ability affect the acoustic fluency and pronunciation judgements of English utterances by Korean speakers. To achieve this, we extracted metadata from the 'English speech data of Koreans for education' from AI Hub (National Information Society Agency, 2023). Acoustic fluency includes speech rate (SR), articulation rate (AR), words per second (WPS), words per second in utterance (WPSU), and the percentage of accurate segments in the utterance (PctC). Pronunciation judgements consist of articulation score, prosody score, and the rating sum assessed by three language experts. We analyzed 87,644 utterances, with participants aged 13 to 65, gender categorized as female and male, and perceived language ability self-reported in three levels (High, Mid, Low). Linear mixed-effects regression, utilizing 'lme4' and 'lmerTest' packages in R, was employed with each acoustic measure as the dependent variable and 'Age,' 'Gender,' and 'Ability (Language Ability)' as fixed effects. 'RecorderId (ID of the participants)' and 'Test Type' were considered random effects to investigate the effects of age, gender, and language ability on acoustic fluency. Results showed that older speakers had lower speech rates, while higher language ability corresponded to higher speech rates. Gender had no impact on speech rate, articulation rate, words per second, or words per second in utterance. Female speakers were more accurate than male speakers in the percentage of accurate segments. Younger speakers pronounced segments more accurately than older speakers, and those with higher perceived language ability were more accurate in segment pronunciation. Language ability was the most effective factor in predicting each acoustic fluency. To explore the effects of age, gender, and language ability on pronunciation judgments, another mixed-effects regression was conducted using the same packages in R, with each score (articulation score, prosody score, rating sum) as the dependent variable and 'Age,' 'Gender,' and 'Ability' as fixed effects. 'RecorderId' and 'Test Type' served as random effects. Results indicated that all three fixed effects significantly influenced all three assessment scores. Female speakers, younger speakers, and those with higher perceived language ability were judged better in all three scores. Language ability was the most determining factor among the three fixed effects, followed by gender and age in the articulation score. The trends were consistent in both prosody score and rating sum.

Reference

National Information Society Agency. (2023). *English speech data of Korean for education*. Retrieved from <https://aihub.or.kr/>

A study on the L2 phonetic imitation: Korean learners' case of English stress assignment

Mi-Young Bae, Tae-Yeoub Jang

Dept. of English Linguistics, Hankuk University of Foreign Studies

baemiyounghufs.ac.kr, tae@hufs.ac.kr

This study aims to verify whether L2 speakers are able to maintain accurate lexical stress assignments when reproducing English native speakers' utterances. In the previous research, talkers have been shown to imitate a wide variety of acoustic-phonetic properties of speech including vowel quality, vowel duration, word duration, speech rate, vowel nasalization, voice onset time, and fundamental frequency (F0), among others. (MacLeod and Di Lonardo Burr, 2022). Thus, it is hypothesized that speakers possess a capacity to adjust their speech to replicate the acoustic-phonetic characteristics of speech they are exposed to.

In the refined research design, the stimuli comprise word-level elements positioned within a context of additional words. This is to restrict prosodic causes such as voice onset time, final lengthening and variables influenced by external factors. The words are organized into 20 trochaic and 20 iambic patterns, divided into four groups: CVCVC, CVCVCC, CVCCVC, and CVCCVCC, each group containing five different words. These disyllabic words were carefully selected to contain words only with apparent stress patterns excluding words like 'record' and 'report' with variable stress.

Speech files from two male and two female American-English native speakers are utilized as the model speech set for learners to imitate. Native speakers recorded 40 sentences each and 10 of them were randomly selected from each speaker to complete a list of 40 sentences in total. Currently, fifteen Korean male and fifteen female subjects are being recruited for the study. Participants' proficiency levels are assessed by The Common European Framework of Reference for Languages (CEFR; Cephe and ToPrak, 2014) Levels and the Vocabulary Size Test, after which they participate in the experiments.

The experiments compose three stages: Baseline Speech (BS), Shadowed Speech (SS), and Memory-based Speech (MS). In BS, participants read randomized scripted sentences aloud. In SS, they listen to and repeat native speakers' utterances. In MS, participants repeat the randomized scripted sentences 24 hours after the shadowing stage to verify the retention of memory over time.

The results will be compared with previous research and preliminary version of this research reported in Bae and Jang (2023).

References

- Bae, M-Y. and Jang, T-Y. (2023). The Effect of Phonetic Replication Focusing on Stressed Syllables Produced by Korean Learners of English (Abstract). In *Proceedings of KSSSS 2023 Fall Conference*, 23. Korean Society of Speech Sciences. Seoul, Korea.
- Cephe, P. T. and Toprak, T. E. (2014). The Common European Framework of Reference for Languages: Insights for language testing. *Journal of Language and Linguistic Studies*, 10(1), 79-88.
- MacLeod, B. and Di Lonardo Burr, S. M. (2022). Phonetic imitation of the acoustic realization of stress in Spanish: production and perception. *Journal of Phonetics*, 92, 101-139.

경상 방언의 내포절 의문사 음성 자극에 대한 의문사 작용역 응답 통계 분석

윤 원 희
계명대학교 영어영문학과

A Statistical Analysis of Wh-Scope Responses to Embedded Wh-Phrases in Gyeongsang Korean

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

내포절에 위치한 의문사는 ‘궁금하다’와 같은 비교량 동사(non-bridge verb)를 모문 동사로 하는 의문문에서 의문사 작용역이 내포절에 머물거나 또는 모문으로 확대되어 이해됨이 여러 연구를 통해 밝혀졌다. 특히 경상 방언에서 이러한 의문사의 작용역은 종결어미를 통하여 명시적으로 표시되는데, 내포문 작용역에서는 ‘-나’를 모문 작용역에서는 ‘-노’, 또는 ‘-니고’의 형태를 갖는다. 의문사 작용역이 종결어미에 따라 형태적으로 나타나지만 모문 작용역에서 나타나는 의문사 억양이 모문 작용역 판단에 중요한 역할을 한다는 사실이 지난 연구를 통하여 보고되었다. 의문사 억양이란 의문사를 포함한 운율 단위(강세구)의 오른쪽 경계부터 경계 해지(post-wh dephrasing)가 나타나는 형식의 억양을 의미하며, 의문사에서 모문 동사까지의 강세구 경계가 해지되어 하나의 돌들림을 갖는다. 지난 연구에서 다중회귀분석(multiple regression analysis)을 통해 내포 동사와 내포절 보문소 ‘-지’가 포함되는 강세구의 돌들림 지수가 낮으면 낮을수록 모문 작용역, 즉 설명의문문으로 응답하는 횟수가 많았다. 그러나 실험 설계의 특성을 정확히 반영하지 않은 통계 분석이 이루어져 이에 대한 보완 분석이 필요하게 되었다. 이 연구에서는 이항 반응 변수 분석 방법인 혼합 효과 로지스틱 회귀분석(logistic mixed-effects regression analysis)을 사용하여 기존의 다중회귀분석 결과에 대한 좀 더 정밀한 통계 검증을 시도한다. 결과적으로, 실험 참가자와 실험에 사용된 자극을 무작위 변수로 하여 내포 동사가 포함된 강세구의 돌들림 지수가 반응 변수를 가장 잘 설명하며, 반복 측정에 따라 반응의 결과에도 영향을 미친다는 것을 알아낼 수 있었다.

Multimodal cue usage in beginning conversational dyadic exchanges across child development

Jun Ho Chai, Barbara Zapior, 고 언 숙

조선대학교 인문데이터과학연구소, University of East Anglia 심리학파, 조선대학교
영어영문학과

junhoc94@gmail.com, B.Zapior@uea.ac.uk, eonsukko@chosun.ac.kr

How do mothers and children initiate conversational exchanges? While verbal communication traditionally marks the onset of dialogue (Chai et al., 2022; Salo et al., 2022), non-speech multimodal cues, such as gestures (Caselli et al., 2012) or touch (Ko et al., 2023), often precede spoken interaction. This paper investigates the pivotal role of non-verbal cues in initiating verbal interactions, exploring how their usage adapts with child age and language proficiency, and examines the potential effect of agentivity in multimodal communication dynamics. Specifically, we test the hypothesis that the use of multimodal, non-verbal cues declines with child age and language development, potentially indicating a transition from non-verbal to verbal communication as language skills mature. Furthermore, we explore whether initiators of conversation will exhibit a higher proportion of cues compared to respondents, as observed with speech patterns in previous research (Ko et al., 2016).

This study utilizes audio-visual recordings from the Ko Corpus of Korean mother-child interaction in CHILDES (Ko et al., 2020). The corpus consisted of dyads at varied developmental stages ($N = 35$, girls = 14; Age in months: $M = 16.19$, $SD = 8.18$, range = 6 - 30), with each participating in a 40-minute free-play session. We identified multimodal cues (gestural, looking, pointing and tactile cues) preceding conversational blocks, defined by an interval of 5,000 ms of silence (Hart & Risley, 1992). In addition, we collected children's language outcomes using the SELSI inventory in comprehension and production.

We employed a binomial linear mixed model to analyze the effects of block type, age, and gender on cue presence, with the random intercept for nested conversational blocks within subjects. We found a significant three-way interaction ($p = .019$), indicating an age-related decrease in cue usage, except in child-initiated blocks by boys (Figure 1). In the second model, the two-way interaction between block type and cue actor ($p < .001$) revealed that children and mother tend to produce more cues in own-initiated blocks than in those initiated by the other ($ps < .001$). When investigating the relationship between non-verbal cues and language outcome, we found a significant interaction between SELSI receptive vocabulary score, block type and cue actor ($p = .035$). The results indicate a negative relationship between children's cues and SELSI comprehension in mother-initiated blocks but weaker trend in child-initiated blocks; whereas mothers tend to produce less non-verbal cues in both block types (Figure 2).

This study’s integrated analysis of verbal and non-verbal elements shed light on the nuanced interplay between multimodal cues and language development in mother-child interactions. Findings reveal a shift from non-verbal initiation of conversation to verbal as children age, with initiator roles influencing the use of multimodal cues. The decline in non-verbal cues within mother-initiated interactions as children’s vocabulary grows signifies a key developmental shift towards favoring verbal over non-verbal communication for initiating exchanges, mirroring the observed age-related variation.

Figure 1

Proportion of non-verbal cues across children’s age in months, categorized by children’s biological sex (girls: F, boys: M), and block types (child-initiated or mother-initiated), based on the cumulative sum of cues used by both mothers and children.

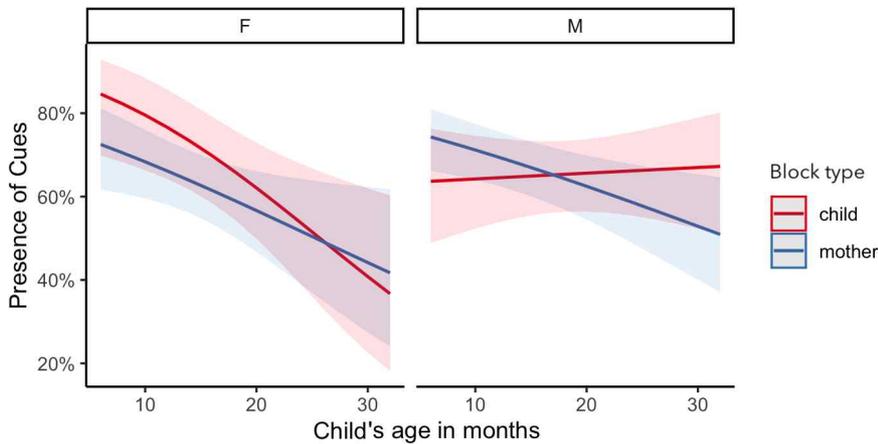
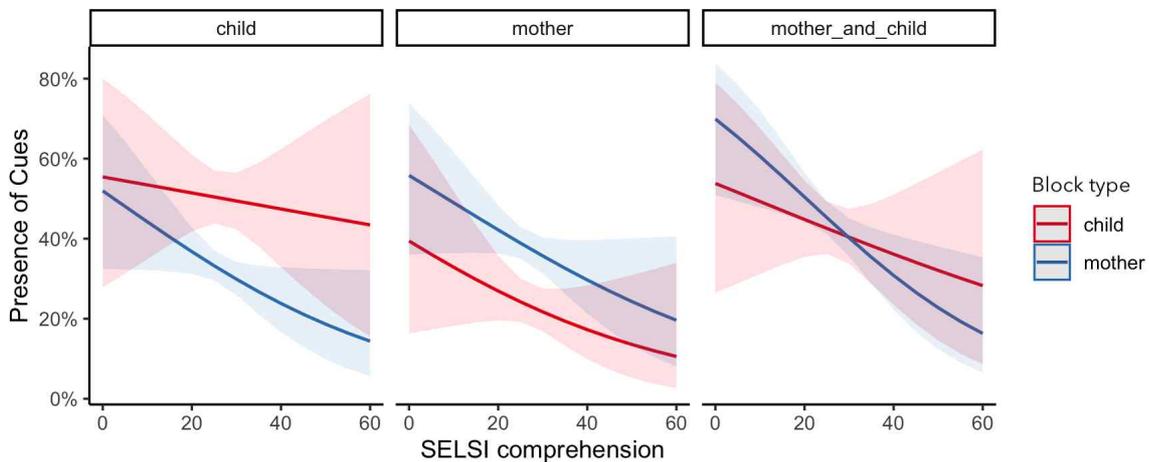


Figure 2

Presence of non-verbal cues across children’s SELSI comprehension score, categorized by actor of cues (child, mother, or mother and child) and block type (child-initiated, or mother-initiated).



Temporal Characteristics of Polysemous Words in Spoken Thai

Kyung-Eun Park, Tae-Yeoub Jang

Dept. of Thai, Dept. of English Linguistics

Hankuk University of Foreign Studies

kyungeunpark@hufs.ac.kr, tae@hufs.ac.kr

In Thai, there are two conflicting temporal effects. First, it was found that when a content word is pronounced its duration is longer than the homophonous function word (Park & Jang, 2022). For example, ‘day’ employed as a main verb exhibits significantly longer duration compared to its function word usage. This suggests that Thai function words shorten like in English. Second, Thai exhibits a language independent process known as pre-pausal lengthening, which elongates final syllables in utterances. Remarkably, the extent of lengthening observed in Thai exceeds that observed in other languages.

A question arises as to which temporal effect—function word shortening and the pre-pausal lengthening—exerts a greater influence when a word is located within both environments. To address this, in the current study, we investigated four Thai words—‘day,’ ‘hay,’ ‘pen’ and ‘laew’—that typically demonstrate multiple functions and meanings. For a production experiment, sentences including each target word were composed and provided to 20 Bangkok Thai speakers so they could produce natural utterance tokens. After post-processing of recordings, a total of 800 target word tokens were prepared and analyzed.

Analysis revealed that in Thai, the process of pre-pausal lengthening is more prominent than previously observed in studies of other languages, as evidenced by two aspects. First, the degree of pre-pausal lengthening is notably greater and more robust compared to other languages. Second, when function words are positioned at the utterance final position, a significant lengthening effect was observed compared to corresponding content words bearing the same segmental string within utterance internal positions. Based on these results, we conclude that the effect of pre-pausal lengthening is significantly greater than that of function word shortening in Thai.

The findings of this study are expected to provide insights for language typological categorization, foreign language education, Thai speech processing, and associated applications.

References

- Park, K. & Jang, T. (2022). Shortening of function words in Thai. *Journal of Language Sciences*, 29(1), 315-337.

한국어 단모음의 조음 양상과 공간 구조 : 3D 초음파 측정법을 기반으로

하 영 우, 황 영
전주대학교 국어교육과, 전주대학교 영어교육과

The articulatory patterns and spatial structures of Korean vowels using 3D ultrasound imaging technology

Young-Woo Ha, Young Hwang

Dept. of Korean Education, Jeonju University
Dept. of English Education, Jeonju University
neozionistgo@jj.ac.kr, yhwang@jj.ac.kr

최근 한국어 단모음에 관한 음향음성학적 연구에서는 단모음의 공간 구조 변이가 논의의 핵심이 되고 있다. 단모음의 공간 구조 변화는 ‘ㄱ’의 후설저모음화, ‘ㄴ’의 고모음화, ‘ㄷ’의 전방화, ‘ㅡ’의 전방화가 연쇄 이동(vowel shift)의 양상을 보인다고 알려져 있다(권경근 2001, 조성문 2003, 송창현 2005, Kang 2014, 윤태진 외 2014, 장혜진 외 2015, 윤규철 외 2015, 강지은 외 2016, 이주희 외 2016, 신우봉 2018, 이향원 외 2017, 이나라 2023). 이 연구는 3D 초음파 이미징 기술을 기반으로 지금까지 음향음성학적 차원에서 논의된 한국어 단모음의 공간 구조 변이가 조음적으로는 어떠한 방식으로 반영되고 있는지 확인하는 것을 목표로 한다.

한국에서 태어나고 자란 3명의 한국어 원어민 화자가 실험에 참여하였으며, 이들이 발음한 한국어의 7개 단모음 /ɪ, E, ㅡ, ㅜ, ㅓ, ㅗ, ㅛ/의 혀 움직임이 Philips 3D/4D EPIQ 7G 초음파 기계를 통해 1초당 평균 20.7fps의 속도로 촬영되었다. 각 모음에 양적 분석과 질적 분석이 실시되었고, 양적 분석은 모음의 중간 지점에서 초음파 데이터의 정중시상면을 기준으로 혀의 최고점 위치를 측정된 데이터를 기반으로 하였다.

이 연구의 결과는 기존의 음향음성학 연구에서 관찰된 단모음의 연쇄적 공간 구조의 변이가 조음적으로도 반영되고 있다는 것을 보여준다. 후설모음 /ㅡ/는 연구개 부근에서 조음되는 다른 후설모음과 달리 경구개와 연구개의 경계 지점에서 조음 되어 중설적 성격을 보였다. /ㅜ/는 후설화와 저모음화 양상을 보였으며, 특히 후설화가 조음적으로 매우 두드러지게 진행되어 /ㅜ/가 /ɪ/와 함께 전후 위치에 관한 극단 모음으로 나타났다. 원순후설모음/ㅗ, ㅛ/는 명료한 전후 대립 관계를 보였으며, 고저 대립에서 전후 대립으로의 변이가 진행된 것이 조음적으로도 확인되었다.

References

- 강지은·공은정(2016), 서울 방언 단모음의 소리 변화와 음향 단서 연구 : 단일지점 포먼트와 궤적 양상, 《말소리와 음성과학》 8-4, 한국음성학회, 39 - 47쪽.
- 권경근(2001), 현대국어의 모음체계 변화의 움직임에 대하여 : 젊은 세대의 말을 대상으로, 《언어학》 30, 사단법인 한국언어학회, 29-48쪽.
- 송창현(2005), 한국어 모음 ‘ㅡ’의 음가 변화에 관한 실험음성학적 연구, 서울대학교 석사학위논문.

- 신우봉(2018), 한국어 /ㄱ/의 음성 실현 양상 연구: 서울코퍼스를 중심으로, 《겨레어문학》 60, 겨레어문학회, 233-258쪽.
- 윤규철·김순옥(2015), 한국어 자연발화 음성코퍼스의 남녀 모음 포먼트 비교 연구, 《말소리와 음성과학》 7-2, 한국음성학회, 131-138쪽.
- 윤태진·강윤정(2014), 한국어 대용량발화말뭉치의 단모음분석, 《말소리와 음성과학》 6-3, 한국음성학회, 139-145쪽.
- 이나라(2023), 표준어 말소리의 변화: 지난 50년간의 아나운서 뉴스 발화를 중심으로, 고려대학교 박사학위논문.
- 이주희·윤규철·변군혁(2016), 서울 방언의 단모음 변화 연구: 중세부터 현대까지, 《언어연구》 31-4, 한국현대언어학회, 979-998쪽.
- 이향원·신지영·신우봉(2017), 한국어 후설 고·중모음에 대한 사회음성학적 연구, 《말소리와 음성과학》 9-2, 한국음성학회, 39 - 51쪽.
- 장혜진·신지영·남호성(2015), 서울 방언 단모음의 연령별 실현 양상, 《음성음운형태론연구》 21-2, 한국음운론학회, 341 - 358쪽.
- 조성문(2003), 현대 국어의 모음 체계에 대한 음향음성학적인 연구, 《한국언어문화》 24, 한국언어문화학회, 427-441쪽.
- Kang, Y.(2014), A corpus-based study of positional variation in Seoul Korean vowels, *Japanese/Korean Linguistics* 23, 1-20.

운율 위치에 따라 달라지는 서울 방언 파열음 지각

소현정, 김정윤
부산대학교 영어영문학과

The Perception of Seoul Korean Stops in Varying Prosodic Positions

So Hyunjung, Jonny Jungyun Kim

Dept. of English Language and Literature, Pusan National University
minny980@pusan.ac.kr, jonnykkim@gmail.com

현재 서울 방언 파열음은 젊은 여성을 중심으로 VOT 병합, f0 차이 확장을 겪고 있다. 본 연구는 Kim(2024)에 기반하여 파열음 변화의 범위가 확장되어 억양구, 강세구 시작 경계를 넘어 단어 시작 경계에서도 파열음 지각 시 VOT 뿐만 아니라 f0를 이용하는지 확인하려고 한다. 또한 화자의 나이, 성별에 따라 운율 위치 별 음향학적 단서(VOT, f0)의 의존 정도가 어떻게 달라지는지를 확인할 것이다.

한국어 비유창성 유발 요인에 대한 연구

홍 정 연

건국대학교 일반대학원 국어국문학과

Reserch on the triggers of Korean disfluency

Jungyeon Hong

Department of Korea Language and Literature, Konkuk University

s2idoloveyou2@konkuk.ac.kr

1. 서론

본 연구는 한국어 표준 화자의 발화에서 나타나는 비유창성을 파악하는 것을 목적으로 한다. 이를 위해 정상적 비유창성과 병리적 비유창성으로 분류하여 분석하는 비유창성 행동의 개념을 아우를 수 있도록 비유창성 유발 요인이라는 개념을 사용하였다. 따라서 비언어장애 성인의 발화에서도 모든 비유창성 유발 요인이 발생할 수 있음을 밝히고자 하였다. 비유창성 유발 요인은 삼입(간투사, 군말), 수정, 반복, 연장, 쉽, 미완성으로 구성하였다.

2. 연구 방법

피험자는 비언어장애 20대 한국인 42명으로, 학창 시절을 수도권에서 보낸 여성 25명과 남성 17명을 선정하여 주어진 자료를 읽는 읽기 과제와 주제를 정하여 자유롭게 발화하는 말하기 과제로 실험을 진행하였다.

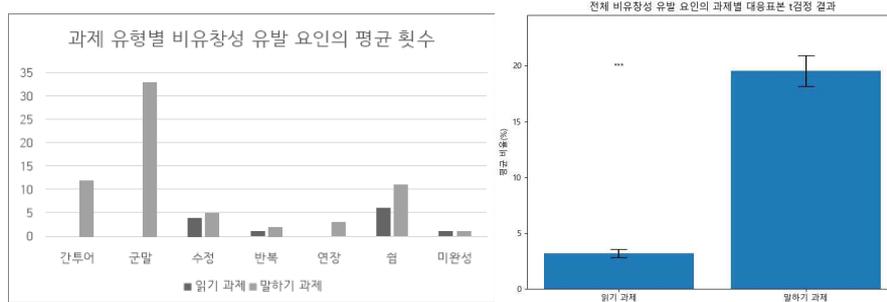
3. 연구 결과

연구 결과, 비유창성 유발 요인 비율과 비유창성 지수에서 모든 비유창성 유발 요인이 나타난 것으로 확인되었다. 정상적 비유창성에 해당하는 행동과 병리적 비유창성에 해당하는 행동이 모두 발생하였다. 또한, 전체 비유창성 유발 요인의 백분율에서 각 요인이 차지하는 비율을 살펴보면 삼입(56%), 쉽(20%), 수정(11%), 반복(5%), 연장(4%), 미완성(3%)의 순서를 가지고 있었다.

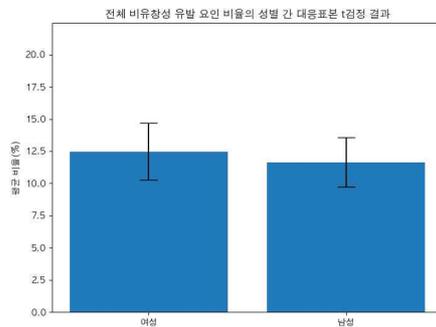
분석방법	삼입		수정	반복	연장	쉽	미완성	전체
	간투사	군말*						
비율	1.8%	4.6%	1.2%	0.5%	0.6%	2.4%	0.3%	11.4%
비유창성 지수	0.7%SS	1.7%SS	0.4%SS	0.2%SS	0.2%SS	0.9%SS	0.1%SS	4.3%SS
말더듬 비율	-	-	-	1SW/M	-	-	-	-

* 표준국어대사전에 따른 사전적 정의는 '하지 않아도 좋을 쓸데없는 군더더기 말'이다. 본 연구에서는 이러한 사전적 정의에 기반하여, '다른 발화에서는 의미를 가지고 사용할 수 있으나, 현 발화에서는 큰 의미 없이 사용된 말'이라는 개념으로 군말을 사용하였다.

과제별 비유창성 유발 요인을 살펴보면 말하기 과제에서 비유창성 유발 요인의 비율이 높게 나타났으며, 모든 종류의 비유창성 유발 요인이 발생하였다.



반면, 전체 비유창성 유발 요인의 비율에서 성별 간 차이는 유의미하지 않은 것으로 나타났다.(p=0.781) 이는 언어장애인을 대상으로 한 연구에서 성별 간 차이가 있는 것과는 상이한 결과이다.



전체적인 논의에서 벗어나 개별 비유창성 유발 요인을 구체적으로 살펴보면, 우선 수정의 경우 수정 전과 수정 후에 무엇이 들어가느냐에 따라 세 유형으로 나뉘었다. 수정 전과 수정 후가 바로 이어지는 유형, 수정 전과 수정 후에 휴지 구간이 들어가는 유형, 수정 전과 수정 후에 다른 말이 들어가는 유형이다. 또한, 삽입 중 군말은 화자에 따라 자주 사용하는 용어가 있었다. 전체 발화에서 나타난 1224회의 군말 중에서는 ‘좀(조금)’ 183회(15%), ‘또’ 134회(11%), ‘이제’ 116회(9%)가 자주 발생했다.

국어학적인 관점에서 살펴보면 읽기 과제에서 발생한 수정 중 조사나 어미를 잘못 읽은 경우는 51%로 과반수를 차지하였다. 이는 화자가 읽기 자료에 나타난 문어체를 발화하는 과정에서 구어체로 변환했을 가능성이 높다. 한편, 반복에서는 음소 단위의 반복이 거의 나타나지 않았고 대부분 음절 혹은 어절 단위의 반복이 발생하였다. 마지막으로 연장은 음절의 구조 중 중성에서 가장 많이 발생하였으며, 중성에서는 비음일 경우에만 발생하였다.

4. 논의 및 결론

결론적으로 비언어장애 성인의 한국어 발화에서도 정상적 비유창성과 병리적 비유창성으로 분류된 비유창성 유발 요인이 모두 발생할 수 있다. 또한, 비유창성 유발 요인은 삽입, 쉼, 수정, 반복, 연장, 미완성 순으로 자주 발생하는 경향이 있었다. 이러한 결과를 고려한다면 언어병리학을 비롯하여 음성인식 기술 등 다양한 분야에서 폭넓은 활용이 가능할 것이다.

한국어 음성인식을 위한 종단형 음성인식 모델의 실험적 비교

오 창 한, 박 기 영, 강 점 자, 송 화 전, 최 우 용
한국전자통신연구원 복합지능연구실

Empirical Study of Korean End-to-end Speech Recognition Model

Changhan Oh, Kiyoung Park, Jeomja Kang, Hwajeon Song, Wooyong Choi
Electronic Telecommunications Research Institute
{ochh508,pkyoung,jjkang,songhj,wychoi4}@etri.re.kr

본 발표에서는 한국어 음성인식에서 여러 종단형 음성인식 모델의 구조를 실험적으로 비교한다. 트랜스포머가 다른 분야와 마찬가지로 음성인식에서도 우수한 성능을 보여준 이후, 컨포머가 특히 음성인식에서 우수한 성능을 보임에 따라 가장 널리 쓰이고 있는 모델로 채택되었다. 이후 개선된 브랜치포머(e-branchformer)가 다시 그 성능을 뛰어넘는 구조로 제안되었다. 본 발표에서는 ESPnet 툴킷을 이용하여 한국어 음성인식에서 이 세 가지 구조의 속도 및 성능을 비교하였다. 이를 위하여 AIHUB에 공개된 데이터를 위하여 약 7,500시간의 훈련셋을 준비하였으며, 또한 평가셋도 공개된 데이터를 이용하여 표준적으로 구축하였다. 실험 결과 e-branchformer가 가장 우수한 성능을 보였으나, 훈련 및 추론 속도는 매우 낮음을 확인하였다. 컨포머는 대부분의 데이터셋에 대하여 우수한 성능과 속도를 보임을 확인하였다. 실험에 사용한 모델과 스크립트 및 데이터 리스트 등의 리소스는 향후 한국어 음성인식에서 표준적인 비교 실험에 사용될 수 있도록 함께 공개한다.

짧은 발성에 강인한 ECAPA-TDNN 기반의 화자 인증 시스템*

김 승 빈, 임 찬 영, 허 정 우, 김 주 호, 신 현 서, 구 교 원, 유 하 진
서울시립대학교

ECAPA-TDNN based Speaker Verification System: Robustness to Short Utterances

Seung-bin Kim, Chan-yeong Lim, Jungwoo Heo, Ju-ho Kim, Hyun-seo Shin, Kyo-won Koo
and Ha-Jin Yu
University of Seoul

kimho1wq@naver.com, cksdud585@naver.com, jungwoo4021@gmail.com,
wngh1187@naver.com, gustjtls123@naver.com, kkwr0504@naver.com, hjyu@uos.ac.kr

화자 인증은 입력 발성의 화자가 시스템에 등록된 화자의 발성인지 확인하는 과제이며, 화자 인증 시스템에서 짧은 길이의 입력 발성은 성능을 하락시키는 요인이다. 이를 개선하기 위해 여러 연구자들은 다양한 시간 해상도에서 추출되는 화자 정보를 활용하여 짧은 발화에 강인한 화자 인증 시스템을 구축하였다. ECAPA-TDNN[1]은 다중 계층 특징 결합과 확장된 컨볼루션 구조를 사용하여 다양한 길이의 발성에 대한 화자 인증 성능을 효과적으로 향상시켰다. 그러나 ECAPA-TDNN의 Res2Block은 서로 다른 주파수 특징을 계층적으로 결합하는 방식으로 화자 정보를 추출하는데 초점을 두어 다양한 시간 해상도에서 추출되는 정보를 파악하기에는 미흡한 구조이다. 따라서 본 연구에서는 Res2Block에 Extended Dynamic Scaling Policy (EDSP)[2]를 적용한 Multi-Resolution Attention (MRA) 블록을 제안한다. EDSP는 기존 특징 추출 분기를 기반으로 다운 샘플링 분기와 업 샘플링 분기를 추가하여 다양한 시간 해상도를 고려할 수 있도록 하는 기법으로, 시스템이 데이터의 스케일에 따라 동적으로 작동하도록 한다. 또한 다른 시간 해상도에서 출력된 특징들 사이의 중요한 정보를 파악하기 위해서 어텐션 기반 게이트 모듈을 사용한다. 이 게이트는 채널별 관계를 모델링함으로써 다중 해상도 경로의 표현력을 강화한다. 이러한 기법을 적용한 MRA 블록은 다양하고 넓은 시간적 맥락을 고려하여 짧은 발성에 강인한 화자 인증 시스템을 만들게 된다. 제안한 방법을 평가하기 위해 VoxCeleb1 데이터를 이용하여 화자 인증 실험을 수행하였다. 1초 길이의 짧은 발화를 사용하여 검증한 결과, 베이스라인 시스템은 7.07%의 동일 오류율을 기록한 것에 비해 제안한 시스템은 6.39%의 동일 오류율을 달성하였다.

References

- [1] Brecht Desplanques, Jenthe Thienpondt, and Kris Demuyne, "Ecapa-tdnn:Emphasized channel attention, propagation and aggregation in tdn based speaker verification", in Proc. Interspeech, 2020.
- [2] Ju-ho Kim, Hye-jin Shim, Jungwoo Heo, and Ha-Jin Yu, "RawNeXt: Speaker verification system for variable-duration utterances with deep layer aggregation and extended dynamic scaling polices", in Proc. ICASSP, 2022.

* 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2023-00263037, 적대적 공격에 강인한 AI 기반 위변조 음성 탐지시스템 개발)

Self-Instruction을 활용한 이모지 생성모델

최 영 우, 이 보 원
인하대학교 전자공학과

Emoji Generation Model Using Self-Instruction

Youngwoo Choi, Bowon Lee

Department of Electronic Engineering, Inha University
ywchoi@dsp.inha.ac.kr, bowon.lee@inha.ac.kr

대규모 언어 모델(Large Language Model, LLM)의 발전은 음성 분야 및 자연어 처리 분야에서 많은 연구를 촉진시켰다. 하지만 일부 연구에서는 여전히 해당 목적에 부합하는 데이터가 부족한 문제가 있다. 이를 해결하기 위해, 특정 Task에 필요한 데이터가 부족할 때 LLM을 활용한 데이터 생성 방법인 Self-Instruction 기법이 주로 활용되며 이 방법은 LLM을 이용하여 데이터를 수집 및 필터링하여 사용자가 직접 작업에 필요한 데이터를 생성하는 방식이다. 본 논문에서는 Self-Instruction을 활용한 이모지 생성 모델을 제안한다. 학습 데이터의 효율적인 생성을 위해 LLM으로는 빠른 속도와 경제성의 장점을 가진 OpenAI의 GPT-3.5 모델을 사용하였다. 데이터 생성의 구체적인 과정으로는, 먼저 프롬프트를 통해 GPT-3.5에 학습에 필요한 데이터를 요청하고, 생성한 데이터를 수집한다. 이후 수집된 데이터 중 task에 관련 없거나 부정확한 데이터를 제거하여 이모지 생성을 위한 모델 학습에 적합한 형태로 전처리를 진행하였다. 전처리 후 수집된 데이터는 총 10,000개이며 이를 활용하여, 주어진 문장에 해당하는 이모지를 생성하는 이모지 생성 모델을 Fine-Tuning하였다. 이모지 생성을 위한 LLM은 파라미터의 개수가 비교적 적은 Llama2 7B-Chat과 Mistral 7B 모델을 사용하였으며, LLM이 텍스트로 출력한 결과는 Python의 Emoji 라이브러리를 활용하여 이모지 형태로 변환하여 출력되도록 설계하였다. 최종적으로 Llama2 7B-Chat과 Mistral 7B 두 모델을 동일한 하이퍼 파라미터로 학습한 두 모델의 출력 결과를 비교, 분석한 결과를 제시한다.

실시간 멀티 채널 음성 향상 모델의 출력 방식 비교 분석*

김 태 한, 박 형 민
서강대학교 전자공학과

Comparative Analysis of Output Methods for Real-Time Multi-Channel Speech Enhancement Models

Tae Han Kim, Hyung-Min Park

Dept. of Electronic Engineering, Sogang University
taehank@sogang.ac.kr, hpark@sogang.ac.kr

일상 속에서 불가피한 배경 소음은 음성 인식에 큰 장애요소이다. 이에 따라, 딥러닝 기술이 소음 억제와 음성 품질의 향상을 위해 활발히 연구되며 적용되고 있다. 멀티 채널 마이크의 사용은 이러한 환경에서의 효과성과 다양성을 크게 향상시킨다. 본 연구는 멀티채널 입력을 받는 실시간 음성향상 모델의 성능을 다각도로 분석한다. 기준 채널에 대한 단일 음성 향상, 단일 음성 향상 마스크 기반 MVDR 빔포밍, 그리고 filter-and-sum 방식의 빔포밍 적용을 통해 음질 향상 정도와 처리 속도를 비교한다. 각각의 방법에 대한 장단점을 평가하여 최적의 음성향상 방안을 탐색한다.

* 이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00989, 다화자 동시 처리를 위한 인공지능 기반 대화 모델링 기술 개발, No.2021-0-02068, 인공지능 혁신 허브 연구 개발)

동적 어트랙터 생성을 위한 토큰 기반 화자 분할 프레임워크*

구 교 원, 허 정 우, 김 주 호, 신 현 서, 임 찬 영, 김 승 빈, 유 하 진
서울시립대학교 인공지능학과, 컴퓨터과학과

EEND-TBA: end-to-end neural diarization with token-based attractor

Kyo-won Koo, Jungwoo Heo, Ju-ho Kim, Hyun-seo Shin,

Chan-yeong Lim, Seung-bin Kim, and Ha-Jin Yu

Dept. of Artificial Intelligence, Computer Science, University of Seoul

kkwr0504@naver.com, jungwoo4021@gmail.com, wngh1187@naver.com,

gustjtls123@naver.com, cksdud585@naver.com, kimho1wq@naver.com, hgyu@uos.ac.kr

화자 분할이란 음성 녹음을 발화자가 동일한 구간으로 분할하는 과제로, “누가 언제 발화하였는가?”라는 질문을 다룬다. 최근 연구에서는 심층 신경망을 활용한 시스템이 우수한 성능을 보이고 있다. 이러한 시스템은 심층 신경망의 프레임 단위 출력과 어트랙터 간 원소 별 곱셈을 통해 프레임 단위 화자 벡터를 추출한다. 여기서 어트랙터란 프레임 단위의 출력에서 화자를 구분하도록 동작하는 특징 벡터를 의미한다. 기존 연구에서는 학습 가능한 가중치 벡터를 어트랙터로 활용한다. 그러나 해당 방식은 모든 학습 데이터 내의 화자에 대해 동일한 어트랙터가 적용된다. 즉, 이는 서로 다른 화자를 동일한 가중치를 이용하여 분리하기 때문에 화자 별 특징을 효과적으로 표현하는데 어려움이 있을 수 있다. 이를 개선하고자 동적 어트랙터 생성 메커니즘을 활용한 토큰 기반 프레임 워크를 제안한다. 제안한 프레임워크의 어트랙터는 학습 가능한 파라미터인 토큰을 활용하여 구성된다. 토큰은 프레임 단위 입력 음향 특징과 함께 트랜스포머 인코더에 입력되며, 인코더의 어텐션 메커니즘을 통해 토큰 별로 서로 다른 화자 정보가 집계된다. 해당 토큰은 각 화자를 대표하는 어트랙터 역할을 수행할 수 있도록 화자 존재 확률 학습이 진행된다. 제안한 시스템은 학습 가능한 토큰을 통해 가변적으로 화자 별 어트랙터를 생성할 수 있으며, 추가적인 모듈을 요구하지 않기에 구현이 간단하다는 장점이 있다. 화자 분할 과제에 대해 제안한 프레임워크를 검증한 결과, 베이스라인 시스템인 EEND-VC[1]는 임의의 화자 수에서 12.22%의 분할 오류율(DER)을 보여준 반면, 제안한 프레임워크는 11.64%의 분할 오류율을 달성하여 기존 시스템 대비 우수함을 확인하였다

References

- [1] K. Kinoshita, M. Delcroix, and N. Tawara, “Advances in Integration of End-to-End Neural and Clustering-Based Diarization for Real Conversational Speech,” in Proc. Interspeech 2021, 2021, pp. 3565 - 3569.

* 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(No.2023R1A2C1005744)

한국인 아동 화자의 수준별 영어 음소 인식을 위한 Wav2Vec2 기반의 도메인 적응 기법*

박 지 은, 박 정 식
한국의국어대학교 언어공학연구소

Wav2vec2-based Domain Adaptation for English Phoneme Recognition by Proficiency Level of Korean Children

Ji-Eun Park, Jeong-Sik Park

Language Technology Research Institute, Hankuk University Of Foreign Studies
ppae0216@naver.com, parkjs@hufs.ac.kr

한국인 아동의 영어 발음 학습 평가를 위해서는 한국인 아동 화자 대상의 영어 음소 인식 모델 구축이 필요하다. 그러나, 한국인 아동 화자 데이터의 부족으로 이 같은 영어 음소 인식 모델을 구축하는데 어려움이 있으며, 이를 극복하기 위해 영어 원어민 화자가 발성한 문장 단위의 대용량 데이터(source data)로 사전 훈련된 모델과 한국인 아동이 발화한 소량의 단어 단위 데이터(target data)로 미세 조정하는 Wav2Vec2 모델을 사용하는 것이 가능하다. 이때, source 도메인에서 target 도메인으로의 변경으로 인해, 생성된 다운스트림(downstream) 모델은 한국인 아동의 발화에서 낮은 음소 인식률을 나타낼 수 있다. 또한, 원어민 화자가 발성한 음성과는 달리, L2 화자들이 발성한 음성은 화자 간 유창성에 차이가 존재한다. 화자 간의 유창성 수준 차이로 인해 음소 인식 모델은 일관성 있는 음성 특성(speech representation)을 학습하기 어려우므로 높은 정확도를 기대하기 어려운 문제가 있다.

본 연구에서는 도메인 차이에 따른 문제와 L2 화자 간 유창성 차이에 따른 문제를 해결하기 위해, 구음 장애 음성 인식에 사용된 [1]의 프레임워크를 활용한다. 이 프레임워크에서는 Wav2Vec2 모델의 transformer encoder 내의 특정 Multi-head attention block에 adapter를 설치하고, fMLLR과 xvector와 같은 화자 정보를 입력값으로 하여 화자 적응을 시도하였다.

본 연구에서는 해당 프레임워크를 도메인 적응 기법으로 적용하여, 유창성 수준 별 강인한 음소 인식 모델을 개발한다. 사전 훈련 모델로 Wav2Vec2-large-robust를 사용하였으며, 한국인 아동 화자로 구성된 K-SEC 음성 데이터를 대상으로 화자를 세 단계의 유창성 수준으로 분류하여 실험을 진행하였다. 해당 프레임워크를 적용함으로써 adapter를 통해 모델이 새로운 도메인에 적합한 파라미터를 학습하며, 추가된 화자 정보를 통해 화자 간 음성 특성을 정규화하여 음소 인식률을 향상시키는 동시에 유창성 수준별 음소 인식률의 차이를 완화하는 효과가 있다.

[1]의 프레임워크는 화자 정보를 통해 화자 적응을 수행함으로써 음성인식 성능을 향상시킨 반면, 본 연구에서는 해당 프레임워크를 도메인 적응 기법으로 사용하여 한국인 아동화자 음소 인식 모델을 구축하였다는 점에서 의의가 있다. 이를 통해 여러 수준을 지닌 한국인 아동의 영어 음소 인식에 있어 화자의 발화 수준에 독립적인 신뢰성 있는 모델을 구축하는 데 기여할 것으로 예상된다.

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

References

- [1] Baskar, M. K., Herzig, T., Nguyen, D., Diez, M., Polzehl, T., Burget, L., & Černocký, J. (2022). Speaker adaptation for Wav2vec2 based dysarthric ASR. arXiv preprint arXiv:2204.00770.

웨어러블 청각 장치의 음성향상을 위한 청각 필터뱅크 잡음 제거 신경망*

^{1,2}박 주 영, ²김 선 만, ¹성 진 택

¹전남대학교 데이터사이언스대학원, ²한국광기술원 공간광정보연구센터

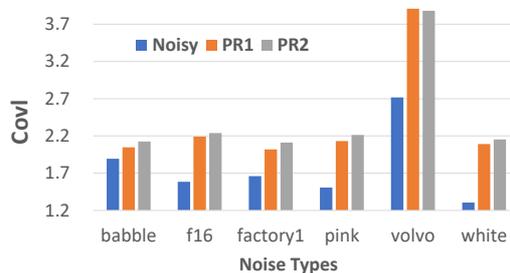
Auditory filterbank denoising neural network for speech enhancement in wearable auditory device

¹Ju Yeong Park, ²Seon Man Kim, ¹Jin-Taek Seong

¹Chonnam National University, ²Korea Photonics Technology Institute
zhseltm@kopti.re.kr, *kobem30002@gmail.com, *jtseong@jnu.ac.kr

청각 장애가 있는 사람들은 일상적인 상황에서 말을 이해하는 데 어려움을 겪기 때문에 주로 보청기 착용을 권장받는다. 그러나 주변 소음 문제로 인해 청각 장애가 있는 개인의 약 23%만이 보청기를 사용한다. 이러한 잡음 문제를 해결하기 위해 통계 모델과 잡음 전력 스펙트럼 밀도를 기반으로 하는 위너필터와 같은 음성향상 방법이 주로 사용됐다. 그러나 이러한 위너필터와 같은 통계 모델 기반 음성향상 기술은 특히 배블과 같은 비정상 특성의 잡음환경에서는 그 성능이 제한되어, 여전히 잡음 문제에 대한 만족도가 낮다.

본 연구에서는 보청기와 같은 청각 장치 적용을 위한 설계된 음성향상 신경망을 제안한다. 먼저 청각 장치의 요구 사항 충족을 위한 알고리즘 처리 지연이 8ms인 32채널 청각 필터뱅크를 구현하였다. 필터뱅크는 분할 대역별 시간 영역 신호를 분석하는 부분과 각 대역 내의 신호로부터 완전한 대역 신호를 재구성하는 합성 부분으로 이루어진다. 제안 방법은 균일 다상 이산 푸리에 변환 필터뱅크의 분석 단계 내에 잡음 제거 신경망을 통합하는 것을 목표로 하며 각 대역 내에서 음성을 향상하는 것을 목표로 하였다. 잡음 제거 모델의 경우, 균일 다상 이산 푸리에 변환의 스펙트럼 성분을 기반으로 하는 음성 위상 정보 복원이 가능한 복소수 컨볼루션 신경망을 적용하였다. 또한, 균일 다상 이산 푸리에 변환 필터뱅크 전략을 활용하여 훈련 과정 동안 분할 대역 내에서 분석된 음성 신호의 손실을 고려하도록 설계된 다중 손실 방법을 추가로 적용하였다. 제안 방법의 음성향상 효과를 평가하기 위해 전반적인 음질 Covl을 측정하였다. 객관적인 평가 결과 제안된 모델은 기존 청각 장치 잡음 제거 방법 대비 높은 성능을 보여주었으며, 특히, PR2로 표기된 분할 대역 음성 신호의 손실을 고려한 다중 손실 방법이, 분할 대역 신호를 고려하지 않은 PR1 방법 대비 향상된 음질을 보여주었다.



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

읽기장애아동과 일반아동의 동시조음 특성 비교

박 성 숙*, 성 철 재[§]

충남대학교 언어병리학과, 충남대학교 언어학과

A comparative study of coarticulation features between children with reading disability and nondisabled children

Sungsook Park, Cheoljae Seong

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

[§]Dept. of Linguistics, Chungnam National University

readingpark70@gmail.com, cjseong49@gmail.com

발화에서 음절을 구성하는 분절음들은 서로 독립되어 있지 않고, 주위의 분절음과 영향을 주고받는다. 동시조음(coarticulation)은 발화 과정에서 시간과 공간의 제약으로 인해 연속적인 조음기관의 움직임에 발생하는 다양한 중첩을 말한다. 자음과 모음이 결합된 음절 구조에서 모음 안정 구간의 F2에 대비한 모음 시작점 F2의 선형성(linearity)을 분석하는 로커스 방정식(locus equation: LE)은 동시조음 정도를 정량적으로 측정할 수 있는 방법이다.

음운인식능력은 구어에서 사용되는 말소리들을 인지하고 조작할 수 있는 능력으로, 단어가 분리된 말소리 단위로 구성됨을 알고 말소리를 인식, 조작하는 능력이다. 음운인식 능력 중 음소 수준의 음운인식 능력은 초기 읽기 및 철자법에 중요한 영향을 끼친다. 읽기장애아동은 이와 같은 음운인식 과정의 결함으로 자소-음소 대응 규칙을 적용하여 빠르고 정확하게 읽는 과제에 어려움을 보인다.

따라서 본 연구에서는 초등학교에 재학중인 읽기장애아동과 일반아동을 대상으로 로커스 방정식을 이용하여 동시조음 특성을 살펴보고자 하였다. 실험은 대전광역시에 거주하는 초등학교 학생으로 읽기장애아동 13명과 일반아동 13명이다. 실험자료는 파열음과 파찰음으로 구성된 12개 자음과 모음 3개(ㄱ, ㄴ, ㄷ), 종성(ㅇ)을 결합한 CVC 음절 36개와 동일한 모음으로 구성된 VCV 2음절 36개로 총 72개 음절이다. 음성자료는 Praat(ver 6.1.42)으로 분절 후 레이블링하였으며 F2 onset 주파수값과 안정적인 F2 vowel 주파수값을 추출한 후, Python(ver 3.10.10)을 이용해 조음위치와 발성유형 별로 단순회귀분석을 실시하여 LE 기울기를 측정하였다. 통계는 SPSS(ver 23)를 이용하여 LE 기울기를 종속 변수로, 집단, 조음위치, 발성유형을 독립변수로 하는 삼원분산분석(3-way ANOVA)을 실시하였다.

통계 검정 결과, CVC 음절 구조에서 LE 기울기는 집단($p < .01$), 조음위치($p < .01$)에서 유의한 차이가 있었다. 읽기장애아동은 양순음>연구개음>치조음>경구개음 순서로, 일반아동은 연구개음>양순음>치조음>경구개음 순서로 동시조음 정도가 컸고, 조음위치에서는 연구개음>양순음>치조음>경구개음 순서로 큰 동시조음 정도를 보였다. VCV 음절 구조에서는 조음위치에서 유의한 차이가 나타났다($p < .001$).

본 연구는 읽기장애아동과 일반아동 집단 간 동시조음 정도를 LE 기울기를 이용하여 정량적으로 분석하였다는 것에 의의가 있다. 집단 간 동시조음 특성을 비교했을 때, 읽기장애아동은 일반아동에 비해 동시조음 정도가 작다는 것을 확인하였다. 이는 읽기장애아동이 보이는 음운인식의 결함이 빠른 자소-음소 연결하기에 영향을 미쳤을 것으로 짐작해 볼 수 있다. 본 연구는 읽기장애아동 집단의 동시조음 특성을 파악하고자 하였으나 휴지와 같은 시간적 변수와 단어 해독 수준이 동시조음에 미치는 영향을 확인하는 것에 한계를 보였다. 따라서 읽기장애아동의 해독 수준에 따라 다양한 자극에서 동시조음 특성을 파악하고 휴지와 같은 시간적 변수가 동시조음에 미치는 영향에 대한 후속 연구가 필요하다.

한국어 자음 유형에 따른 상대적기본주파수

정지행*, 김재옥**

*마스터플러스병원, **강남대학교 교육대학원 언어치료교육전공

Relative Fundamental Frequency by Korean Consonants

Jung, Ji Hang*, Jaeock Kim**

*Master Plus Hospital

**Major in Speech Pathology Education, Graduate School of Education, Kangnam Univeresity
lagblo2@naver.com, jaeock@gmail.com

상대적 기본주파수(Relative Fundamental Frequency, RFF)는 /모음+무성자음+모음/으로 이루어진 음운환경에서 기본주파수(F0)의 변동을 추적하는 음향학적 측정치이다. RFF는 무성자음 직전과 직후 10개의 성대진동주기의 F0를 측정하여, 이들의 변화 양상을 반음(semitone, ST)으로 표준화하여 산출한다. RFF는 후두의 긴장성을 반영하기 때문에 음성장애를 변별하는데 유용한 것으로 알려져 있다.

RFF는 자음 유형에 따라 RFF 측정치가 달라질 수 있는데, 국외 연구는 주로 영어를 사용하여 RFF를 측정하였기 때문에 영어권의 자음과 조음위치나 조음방법의 차이가 있는 한국어 자음을 대상으로 RFF를 측정하고, 어떠한 음소가 RFF를 측정하는데 적합한지 파악할 필요가 있다.

이에 본 연구는 RFF 측정에 적절한 한국어 음소 선별작업을 위해 정상음성을 산출하는 20대 여성 10명을 대상으로 각 파열음 중 격음 /ㅍ, ㅌ, ㅋ/과 경음 /ㅂ, ㄷ, ㄱ/, 파찰음의 격음과 경음 /ㅊ, ㅈ/, 마찰음의 격음과 경음 /ㅅ, ㅆ/에 모음 /아/를 조합한 VCV 형태의 10개 무의미음절을 사용하여 대상자가 각 음절당 6회씩 반복 산출하도록 하였다. 그리고 산출된 총 600개의 음절의 파형을 Praat을 이용하여 RFF를 측정하고, 파열음의 경음과 격음, 파찰음, 마찰음의 RFF 평균과 표준편차를 측정하였다.

연구결과, 자음 직후 첫 번째 상대주기의 RFF가 가장 높게 나타난 마찰음의 경음이 RFF를 측정하는데 적합한 음소로 확인되었다. 다만 본 연구는 여성만을 대상으로 하였기에 남성의 RFF를 측정하는데 적합한 음소를 분석할 필요가 있다.

NPH 환자와 정상 노인 간 어휘인출능력 및 시선추적에 대한 비교 연구

권성은*, 박기수**, 윤장혁***, 김나연****, 이지호****, 하지완*****
대구대학교 대학원 재활과학과 언어치료전공*
경북대학교 의과대학 신경외과**
건국대학교 산업공학과***
네오폰스 주식회사****
대구대학교 언어치료학과*****

Differences in Confrontation naming based on Cognitive abilities: Evidence from Eye-Tracking

Seong-Eun Kwon, Ki-Su Park, Jang-Hyeok Yoon, Na-Yeon Kim, Ji-Ho Lee, Ji-Wan Ha
Major of Speech-Language Pathology, Dept. of Psycholinguistics, Graduate school, Daegu
University*

Dept. of Neurosurgery, School of medicine, Kyungpook National University**

Dept. of Industrial Engineering, Konkuk University***

Neopons Inc****

Dept. of Speech-language pathology, Daegu University*****

sinabro7991@naver.com, kiss798@gmail.com, janghyoon@konkuk.ac.kr, nancy@neopons.com,
oscar@neopons.com, jw-ha@daegu.ac.kr

1. 연구의 필요성 및 목적

급속한 인구 고령화에 따른 문제점 가운데 하나가 노인성 질환의 급증이며, 그 가운데서도 가장 심각한 문제 중의 하나가 치매 환자의 증가이다. 서로 다른 형태의 치매 사이의 경계는 불분명하고 혼합되어 존재한다. 치매는 비가역적인 질병이지만, 정상압수두증(Normal Pressure Hydrocephalus, NPH)은 치매의 원인 중 몇 안 되는 가역적인 특성을 가지고 있으며(Shprecher D, Schwalb J, Kurlan R, 2008), 이러한 증상은 치료적인 뇌척수액 배액(cerebrospinal fluid, CSF)으로 회복할 수 있다. 허나 적절한 시기에 뇌압을 경감시키지 못한 경우에는 전두엽 인지장애뿐만 아니라 피질성 인지장애를 동반한 광범위하고도 영구적인 장애를 초래할 수 있다. 따라서, 가역성 치매는 적절한 시기에 치료를 하지 않으면 뇌에 구조적 변형이 발생하거나 비가역적인 변화로 이어지기 때문에 조기 진단을 필요로 한다. 본 연구에서는 정상 노인 집단과 NPH 집단의 어휘인출의 오프라인 처리 과정(offline processing) 및 시선추적기(eye-tracker)를 활용한 실시간 처리 과정(online processing)을 알아보고 두 집단 비교를 통해 언어 능력이 NPH의 조기 진단 및 모니터링에 있어 중요한 생체지표가 될 수 있을지에 대하여 탐색해보고자 하였다.

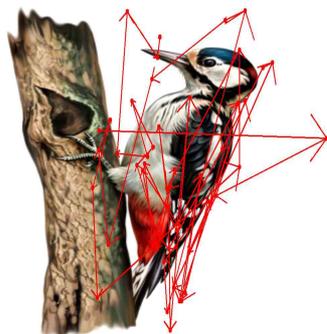
2. 연구 대상 및 연구 방법

본 연구에 참여한 대상자는 NPH로 진단받은 환자 21명, 정상 노인 21명으로 총 42명이었다. 본 연구에서는 대상자의 온-오프라인의 어휘인출 처리 과정을 비교하기 위해 어휘인출 과제를 실시하였다.

어휘인출 과제는 의미적 또는 음운적 오류를 빈번하게 보이는 단어 30개 항목을 사용하여 어휘산출단계 상 기저 어려움을 측정하였다. 통계분석으로는 어휘인출 과제에서 두 집단 간 어휘인출과제 점수를 알아보기 위해 독립표본 t-검정, 오류유형에 따른 집단 간 차이를 알아보기 위해 반복측정 분산분석 (Repeated Measure ANOVA)을 실시하였으며, NPH 집단의 목표 자극에 대한 시선고정 양상을 시선 흐름으로 분석하였다.

3. 연구 결과

어휘인출과제 평균 점수는 NPH 집단이 정상 노인 집단보다 평균적으로 낮은 점수를 보였으며 두 집단 간 차이는 통계적으로 유의하였다($t=-4.475, p < .001$). 오류유형에서는 오류유형의 백분율로 두 집단을 비교하였기 때문에 집단 간($F(1, 40)=.997, p > .05$) 유의한 차이가 없었으며, 오류유형에 대한 주효과($F(3.273, 130.917)=36.917, p < .001$) 및 집단과 오류유형 간 상호작용효과($F(3.273, 130.917)=3.428, p < .05$)에서 통계적으로 유의한 차이를 보였다. 사후검정 결과, 의미오류는 혼합오류, 무반응보다, 혼합오류는 의미오류, 무반응보다, 무반응은 형태오류와 비관련 오류보다 유의하게 많이 발생하였다($p > .05$). 집단과 오류유형 간 상호작용효과에 대해 분석한 결과, 혼합오류, 무반응, 시각오류에서 두 집단 간 차이가 유의하였으며($p > .05$). 혼합오류는 정상 노인 집단에서, 무반응은 NPH 집단에서 유의하게 높게 나타났다. 시각적 분석 결과, 다음 <그림 1>과 <그림 2>에서와 같이 NPH 집단의 시선 분포가 전반적으로 분산되어 있는 것으로 나타났다.



<그림 1> NPH 집단의 '딱다구리'에 대한 과제 수행력



<그림 2> NPH 집단의 '카네이션'에 대한 과제 수행력

4. 논의 및 결론

본 연구에서는 NPH, 정상 노인 집단을 대상으로 어휘인출과제를 실시하여 수행력을 비교하였다. 어휘인출과제에서 NPH 집단은 정상 노인 집단보다 더 낮은 수행을 보였다. 또한 오류유형 분석 결과 NPH 집단과 정상 노인 집단 모두 단어로 이루어진 의미 오류를 가장 많이 보였으며, 특히 NPH 집단은 무반응 오류를 유의하게 많이 보였다. 무반응 오류는 좌측 뇌반구의 inferior parietal lobule, posterior middle, inferior temporal gyri, posterior middle frontal gyrus의 저관류와 관련이 있는데(Atsuko Hayashi, 2011), 이는 NPH 환자의 광범위한 뇌혈류량 감소로 인한 전두엽-피질하 기능장애가 관찰된 것일 수 있다. 실시간 처리 과정 결과, 시각 정보 처리에서 효율적인 메커니즘이 부족하여 비효율적이고 무작위적인 방식으로 시각적 자극을 탐색하는 시각적 검색 전략(visual search strategy)을 취할 경우 더 긴 시선고정 시간과 분산된 시선 고정을 특징으로 한다(Nothdurft, 2006)는 이론에 근거하여, NPH 집단의 목표자극에 대한 시선이 분산되어 있는 것은 NPH 집단의 효율적인 정보 처리의 어려움을 반영한 것으로 예측해볼 수 있으며, 이는 곧 어휘인출 수행의 저하에 영향을 미칠 수 있음을 알 수 있다.

References

- Shprecher D, Schwalb J, Kurlan R (2008). Normal pressure hydrocephalus: diagnosis and treatment, *Curr Neurol Neurosci Rep*, 371-376.
- Atsuko Hayashi, Hiroshi Nomura, Ruriko Mochizuki, Ayumu Ohnuma, Teiko Kimpara, Kazumasa Ootomo, Yoshiyuki Hosokai, Toshiyuki Ishioka, Kyoko Suzuki, Etsuro Mori (2011). Neural substrates for writing impairments in Japanese patients with mild Alzheimer's disease: A SPECT study, *Neuropsychologia*, 1962-1968.
- Nothdurft, H.-C. (2006). Saliency and target selection in visual search, *Visual Cognition*, 14(4-8), 514-542.

경련성발성장애와 다른 음성장애의 식별을 위한 음향지표 및 말자료 개발에 대한 예비연구

서 인 효, 성 철 재
단국대병원 음성언어치료실, 충남대학교 언어학과

A Preliminary Study on Development of Acoustic Parameters and Materials for Identifying Patients With Spasmodic Dysphonia and Other Voice Disorders

In Hyo Seo, Cheol Jae Seong

Voice Clinic, Dankook University Hospital
Department of Linguistics, Chungnam National University
slpseo91@gmail.com, cjseong49@gmail.com

내전형 경련성 발성장애, 외전형 경련성 발성장애, 근긴장성발성장애, 음성떨림을 연구하기 위해 각 장애를 식별할 때 진단적 기준이나 지침이 없이 시행하는 경우, 전문가들 간에도 상당한 차이를 보인다. 최근 지각적 식별 특성에 대한 목록이 델타이 기반으로 개발되었다. 본 연구는 지각적 식별특성을 음향학적으로 객관화할 수 있는 음향학적 지표 및 말자료 개발을 위한 예비연구로 문헌고찰 및 사례를 분석한 결과이다.

인지-언어훈련 애플리케이션을 활용한 어휘 및 조음능력 증진

김 나 연¹, 이 지 호¹, 박 기 수^{1,2}, 윤 장 혁^{1,3}, 하 지 완⁴

¹네오폰스(주)

²경북대학교 의과대학 신경외과학교실

³건국대학교 산업공학과

⁴대구대학교 언어치료학과

Improving Vocabulary and Articulation Skills through Cognitive-Language Training Application

Nayeon Kim¹, Jiho Lee¹, Ki-su Park^{1,2}, Janghyeok Yoon^{1,3}, Ji-wan Ha⁴

¹Neopons Inc.

²Department of Neurosurgery, Kyungpook National University School of Medicine

³Department of Industrial Engineering, Konkuk University

⁴Department of Speech-Language Pathology, Daegu University

nancy@neopons.com, oscar@neopons.com, kiss798@gmail.com, janghyoon@konkuk.ac.kr,
jw-ha@daegu.ac.kr

의사소통장애는 단기적으로 해결하기 어렵다. 이러한 문제 해결을 위해, 대면 언어치료를 보조하는 디지털 기반 프로그램을 활용하기 시작하였다. 이는 치료의 지속성을 강화하고, 대면 언어치료로 도달하기 어려운 효과를 이끌어 낼 수 있다. 본 연구는 언어발달지체를 동반한 말소리장애 아동을 대상으로, 인지-언어훈련 애플리케이션 토키랜드의 활용이 어휘 및 조음능력에 미치는 영향을 분석하였으며, 대면 언어치료의 보조도구로서 잠재력을 탐색하였다.

연구 1에서는 토키랜드의 전문성을 검증하기 위해 언어재활사들이 타당도와 신뢰도 검증에 참여하였다. 그 결과, 해당 콘텐츠 모두 높은 타당도와 신뢰도를 보였다. 대면 언어치료 보조도구로 높은 잠재력을 지니고 있음을 확인하였다.

연구 2에서는 치료적 효과를 증명하기 위하여 애플리케이션을 대면 언어치료의 보조도구로 사용하는 집단, 대면 언어치료만 받는 집단을 대상으로 사전 및 사후검사를 통해 어휘력과 조음능력을 분석하였다. 연구 결과, 토키랜드 집단만이 수용어휘력에서 유의한 강점을 보였다.

연구 3에서는 토키랜드 사용 집단의 어휘 및 조음능력의 변화 양상을 파악하기 위해 총 14회에 걸쳐 모니터링 하였다. 연구 결과, 어휘와 조음능력 모두 아동의 초기 수준에 따라 다소 상이한 양상을 보였다는 점이 주요 발견이었다. 즉, 개별 맞춤형 접근의 중요성을 강조할 수 있다.

연구 4에서는 아동들이 접하는 어휘의 양과 특성을 분석하였다. 그리고 어휘별 문항난이도를 바탕으로 어휘습득에 영향을 미치는 내적요인을 탐색하였다. 연구 결과, 어휘 노출의 양적, 질적 모두 중요하였고, 단어빈도와 음소배열확률이 어휘습득의 문항난이도를 예측하는 중요한 요인임을 발견하였다.

본 연구를 통해 토키랜드가 대면 언어치료의 보조도구로서 갖는 잠재력을 확인하였다. 이는 토키랜드를 활용한 최적화된 치료 방안을 제안하는 데 중요한 기준점을 제공하였다.

롬바르드 효과가 아동과 성인의 말소리 산출에 미치는 영향 : 음향학적 특성과 모음공간면적을 중심으로

장 예 립, 황 재 희, 이 누 리, 이 나 경, 음 세 은, 이 영 미
이화여자대학교 언어병리학과

Comparison of acoustic features due to the Lombard effect in typically developing children and adults

Yelim Jang, Jaehee Hwang, Nuri Lee, Nakyung Lee, Seeun Eum, Youngmee Lee
Dept. of Communication of Disorders, Ewha Womans University
yelim4936@ewhain.net, wonjoonyuyeon@ewhain.net, shoona85@ewhain.net,
lna0705@ewhain.net, sjeumw3@ewhain.net, youngmee@ewha.ac.kr

Purpose: The Lombard effect is an involuntary response of speakers experience in the presence of noise during voice communication. The aim of this study is to investigate the Lombard effect by comparing the acoustic features in children and adults under different listening conditions.

Methods: Twelve male children (5-9 years old) and 12 young adult men (24-35 years old) were recruited to produce speech in three different listening conditions (quiet, noise-55dB, noise-70dB). Acoustic analyses were then carried out to characterize their acoustic features, such as F0, intensity, duration, and vowel space area, under three listening conditions.

Results: The Lombard effect was observed in intensity and duration of children and adults who participated in this study in adverse listening conditions. However, we did not observe the Lombard effect in the F0 and vowel space area of the two groups in this study.

Conclusion: These findings suggest that children can modify their speech production in challenging listening conditions compared to adults.

중학교 읽기부진 학생의 시나리오 글 읽기 유창성 특성

박 지 혜*, 성 철 재[§]

충남대학교 언어병리학과, 충남대학교 언어학과

Characteristics of Scenario Text Reading Fluency in Middle School Students Suffering from Reading Disorder

Jihye Park, Cheoljae Seong

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

[§]Dept. of Linguistics, Chungnam National University

jhparkmt@gmail.com, cjseong49@gmail.com

읽기 유창성은 문장이나 문단 글을 정확하고 빠르게, 운율적 표현력을 가지고 읽는 능력을 의미한다. 여기에는 읽기 유창성의 구성요소인 정확성, 신속성, 표현력(운율성)의 세 가지 측면이 포함된다. 표준화된 읽기검사를 포함한 대부분의 읽기 유창성 평가에서는 정확성과 신속성의 두 가지 요소를 다룬다. 표현력의 경우에는 객관화된 평가가 어려우므로 제외하는 경우가 많은 것이다. 이에 본 연구에서는 표현력 있는 읽기를 극대화할 수 있는 시나리오 글 읽기에서의 문형별 운율 특성을 분석하고 모든 읽기 유창성 요소를 살펴보고자 한다.

본 연구의 대상은 중학교 1~2학년 남학생 30명(일반학생 15명, 읽기부진학생 15명)이며, 분석을 위해 중학교 국어 교과서에서 발췌한 3편의 시나리오 글을 학생과 연구자가 역할을 나누어 읽고 녹음하였다. 분석 문장은 각 시나리오별 평서문과 의문문 3문장씩 총 18문장이며, 이는 내용타당도 평가 후 적절성을 확인하였다. 읽기 유창성 특성을 파악하기 위해 각 집단별로 음절 수준에서의 오류유형을 분석하였고, 음향학적 분석을 위해서는 읽기 속도, 조음 속도, 강세구(AP) 개수, 억양구(IP) 개수, 음도 범위, 강도 범위의 음향변수를 측정하였다. 통계 방법은 집단을 개체 간 변수로 하고 시나리오 종류, 문장 유형, 발화 문장을 개체 내 변수로 하는 혼합 4원 반복측정 분산분석(mixed effect four-way repeated measures ANOVA)을 적용하였다.

연구 결과, 정확성 측면에서 살펴본 오류 음절 분석에서 읽기부진학생집단이 일반학생집단에 비해 높은 오류율을 보였다. 오류유형은 일반학생집단의 경우 대치(13음절) > 생략(10음절) > 수정(9음절) > 삽입(4음절) > 반복(2음절) 순인 반면, 읽기부진학생집단은 수정(121음절) > 대치(47음절) > 반복/삽입(각 46음절) > 생략(42음절) 순을 나타내었다. 속도 측면의 분석 결과, 분석 문장 전체의 음절에 대한 읽기 소요시간을 측정한 읽기 속도와 쉼(pause)을 제외한 조음 속도 분석에서 같은 결과가 있었다. 즉 개체간 효과 검정 결과에서 집단 간 유의한 차이가 확인되었으며($p < .001$), 문장 유형과 집단 간의 상호작용이 관찰되어($p < .01$) 평서문과 의문문 모두에서 읽기부진학생집단의 읽기 속도가 더 느렸음을 알 수 있었다. 강세구와 억양구 개수에서도 집단 간 차이가 있었다($p < .001$). 읽기부진학생집단은 일반학생집단에 비해 강세구를 0.382개 정도 더 만들었으며(normal mean=5.011, disorder mean=5.404), 억양구는 0.326개 정도 더 만드는 것으로 파악되었다(normal mean=1.059, disorder mean=1.385). 문형에 따라서는 평서문과 의문문 모두에서 읽기부진학생집단이 일반학생집단에 비해 문장당 강세구와 억양구 수가 더 많았다($p < .01$). 또한, 표현력 평가를 위해 분석한 운율 변수 중 음도 범위에서는 집단 간 차이가 없었으나 강도 범위에서는 유의한 차이가 있었다($p < .01$). 이로써 읽기부진학생집단이 일반학생집단에 비해 강도 범위가 더 넓다는 것을 확인할 수 있었다(mean difference=4.019, $p < .01$).

구개열 언어치료 - 조음동작형성법의 개발과 적용

이 은 정
세한대학교 언어치료학과

Speech therapy for cleft palate speech -
Set-up & Hold Articulatory Gesture Training method

Eun-Jeong Lee
Dept. of Speech Therapy, Sehan University
eunjeonglee@sehan.ac.kr

선천적으로 구개열이 있는 아동은 12개월 전후 구개 재건 수술을 받는다. 아동의 언어발달을 고려하여 조기 수술을 권고하는데, 술 후 유아의 짧은 발화에서 조음 및 공명의 문제를 확인하기는 어렵다. 보상조음학습의 위험을 방지하기 위해 조기 언어중재를 제안하지만 아동이 다음절 낱말 및 문장을 산출하는 시기에 이르러 말소리 산출 관련 연인두폐쇄 기능을 제대로 파악할 수 있고 발달 수준에 따라 36개월 이후에도 직접 중재에 제한이 있을 수 있다. 그러나 어휘폭발기를 포함하여 언어 습득 및 확장이 빠르게 진행되는 유아기에 말소리 교육과 보상조음의 예방을 동시에 이룰 수 있는 방향으로 적극적인 중재 계획을 세울 필요가 있다. 선행연구에서 구개 수술을 받은 아동들의 조음 및 공명 장애에의 비율이 전형적으로 발달하는 아동들보다 높게 나타나기 때문이다. 이에 본 연구는 구개열 언어의 특성을 나타내는 대상자들을 위해 개발한 조음동작형성법(Set-up & Hold Articulatory Gesture Training)과 그 적용 사례를 소개하고자 한다. 한국어를 모국어로 하는 구개열 환자들을 위한 조음동작형성법은 고압력 자음이 많은 우리 말소리의 특성을 반영한 말운동 훈련법이다. 우리말의 자음은 19개 중 14개가 고압력 자음으로, 동일한 조음위치에서 산출되면서 발성유형에 따라 삼중대립을 나타내는 음소가 많다. 조음동작형성법은 목표 자음과 동일한 조음위치에서 산출되는 생리적 발성 및 의성어를 유도동작으로 활용하여, 음소에서 시작하여 음절 산출로 점진적인 훈련을 진행한다. 구개 수술 후 불완전한 연인두폐쇄로 구강압 형성 및 유지가 어렵고 과다비성 또는 혼합 공명장애를 나타내는 대상자들에게 조음동작형성법을 적용한 결과 자음정확도(PCC) 및 발화명료도, 비성도 수치가 치료 전과 비교하여 개선되었다. 조음동작형성법은 구개열 수술 후 말산출에 어려움이 있는 대상자들에게 적용된 운동학적 접근법으로 훈련 후 대상자들의 조음 및 공명장애 개선에 유의미한 효과가 있는 것으로 나타났다.

화자의 말속도, 청자의 성별, 연령과 말속도가 청자의 호감도 평가에 미치는 영향

천은빈, 정유라, 김희연, 김나현, 김수진, 이영미
이화여자대학교 언어병리학과

Perception of voice attractiveness: Effects of speaking rate, gender, and age

Eunvin Chun, Yoora Jeong, Heeyeon Kim, Nahyun Kim, Sujin Kim, Youngmee Lee
Dept. of Communication Disorders, Ewha Womans University
222shg17@ewhain.net, yoora214@ewhain.net, iiy2eonh@ewhain.net, knh1742@gmail.com,
sujin.kim@ewhain.net, youngmee@ewha.ac.kr

Objectives: The purpose of this study was to investigate the effects of the speaker's speech rate level, listener's age, and gender on the listener's attractiveness evaluation of the speaker's voice.

Method: A total of 8 speaker voices (4 males and 4 females in their 20s-30s) were selected from the speech corpus provided by the National Institute of Korean Language. These voices were manipulated to be 0.8 times and 1.2 times their original speed, resulting in 24 voice files for attractiveness evaluation. Evaluators were recruited from different age groups: adolescents (14), young adults (16), and middle-aged adults (14). They were instructed to perceptually evaluate the attractiveness of the speaker's voice. In addition, speech data from evaluators were collected to measure their actual speech rates.

Results: Evaluator age groups did not show significant differences in attractiveness scores. However, across all groups, the attractiveness ratings were consistently higher for normal speech rates. Middle-aged male evaluators reported a significant decrease in attractiveness for faster speech rates. The correlation between the actual speech rate of evaluators and speakers was significant only for slower speech rates, indicating that evaluators with faster speech rates, especially males, showed lower attractiveness.

Conclusion: The listener's attractiveness evaluation may depend on the listener's gender and age according to the speaker's speaking rate, with middle-aged male listeners being particularly unacceptable to the speaker's fast speech rates. This study underscores the significance of speech rate as an acoustic evaluation factor influencing listener attractiveness.

한국어 이중모음 분석을 위한 활음 분절 알고리즘 개발*

박 지 연*, 성 철 재[§]
 충남대학교 언어병리학과, 충남대학교 언어학과

Development of a Glide Segmentation Algorithm for Korean Diphthong Analysis

Jiyeon Park, Cheoljae Seong

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

[§]Dept. of Linguistics, Chungnam National University

bpn0525jy@gmail.com, cjseong49@gmail.com

이중모음의 활음 구간에서 나타나는 포먼트의 전이 특성은 이중모음을 설명하는 데 중요한 역할을 한다. 활음 구간에서의 F2 변화량, 활음의 시간 비율 그리고 이를 모두 반영한 F2 기울기는 이중모음을 지각하는 것은 물론 일반 집단과 병리적인 집단을 구분하는 데에도 중요하다. 따라서 이중모음의 음향 음성학적 분석을 위해서 활음 구간을 객관적인 기준으로 정의하는 일이 필수적이다. 기존에 발표된 여러 선행 연구에서 이중모음 내에서 활음이 끝나는 지점을 핵모음이 안정적으로 시작하는 지점으로 정의 하였으나 이는 연구자의 주관이 많이 개입되기 때문에 연구방법적 측면에서 신뢰하기 어렵다는 문제가 있다. 이를 보완하기 위해 Weismer et al.(1988)의 20Hz/20ms 규칙을 적용한 선행연구도 발표된 바 있으나 이 규칙은 한국어 이중모음에 그대로 적용하기에 어려움이 있다. 따라서 본 연구는 한국어 이중모음의 활음을 포착하기 위한 음향이론에 충실한 체계적이고 합리적인 절차를 마련하고자 하였다.

Praat을 이용하여 개발한 활음 분절 스크립트는 총 3단계 과정으로 설계하였다. 1단계에서는 활음 최적 분절 알고리즘을 이용한다. 1단계 알고리즘은 시간×F2 좌표 평면에서 이미 고정된 좌표인 활음 구간 탐색 시작점, 탐색 종료점을 두 꼭짓점으로 설정하고 이를 밑변으로 삼각형의 면적이 최대가 되는 다른 하나의 좌표를 탐색한다. 2단계는 활음 분절 지점을 탐색하기 위해 사용자가 직접 탐색 구간을 지정하고, 20Hz/20ms 규칙을 보완하여 적용하였다. 3단계는 1, 2단계에서 적용이 어려운 경우, 스크립트의 사용자가 직접 활음 구간을 분절하는 단계다.

일반 성인 화자 80명이 산출한 이중모음을 대상으로 음운 환경(고립 모음, 단어, 문장)별 활음 분절 스크립트를 적용한 결과는 다음과 같다. 활음 구간에서의 포먼트 변화량이 큰 /ja/, /jʌ/, /jo/, /ju/, /we/, /wi/, /wi/는 1단계 알고리즘으로 활음 분절이 가능했고, 포먼트 변화량이 비교적 작은 /je/, /wa/, /wʌ/는 2단계 알고리즘 적용이 가능했다. 고립 모음 환경에서 전체 이중모음 중 남성은 95.3%, 여성은 96%가 1단계와 2단계 알고리즘으로 활음 분절이 가능했다. 단어와 문장 환경에서도 각각 평균 89%, 86.1%가 1, 2단계 알고리즘으로 활음 분절이 가능했다. 이외 일반 아동과 말소리 장애 아동이 산출한 이중모음에 활음 분절 스크립트를 적용한 결과에서도 활음 분절 스크립트의 높은 성능을 확인할 수 있었다.

본 연구에서 개발한 활음 분절 스크립트는 이중모음을 구성하는 활음 및 핵모음의 종류, 성별에 따라 포먼트 측정을 위한 최적 설정을 적용하고, 활음 구간을 포착하기 위한 음향이론에 충실한 원리를 제시 하였으며 이를 일관된 기준을 적용하기 위한 자동화 논리를 적용하였다. 따라서 본 연구는 한국어 이중모음에 대한 연구방법적 측면에서 객관적이고 신뢰할 수 있는 분석 방법을 제시했다는 것에 본 연구의 의의를 둘 수 있겠다.

* 이 논문은 제 1저자의 2024년도 박사학위 논문의 데이터와 논의를 일부 보완하고 수정한 것임.

단어와 터치간 시간적 일치성이 초기 언어발달에 미치는 영향: 한국 어머니와 아동의 상호작용을 통해 본 연구

정 종 민, 김 수 한, 고 언 숙
조선대학교 인문데이터과학연구소
조선대학교 영어영문학과

The Impact of Word-Touch Synchrony on Early Language Development: Preliminary Findings from Korean Mother-Child Interactions

Jongmin Jung, Suhan Kim, Eon-Suk Ko
Center for Data Science in the Humanities, Chosun University
Dept. of English Language and Literature, Chosun University
jongminjung169@gmail.com, suhan3804@gmail.com, eonsuk@gmail.com

This study investigates how the temporal synchrony between maternal touch and speech influences children's early language development. We examine whether the proportion of word+touch co-occurrence, the proportion of their overlap, and the precision of word-touch alignment in maternal input contribute to language acquisition. Such synchrony may enhance infants' attention and facilitate speech segmentation, aiding their language acquisition process.

Video recordings of 21 infants under 16 months old (mean age = 10 months 21 days, SD = 3 months 1 day) from a dataset of 35 dyads (Ko et al., 2023) were analyzed. Six-minute interaction segments per mother-infant pair were examined to assess the proportion of word+touch co-occurrence, word+touch overlap, and the precision in the alignment of touch-word boundaries in maternal input. Infants' language abilities were assessed using the Sequenced Language Scale for Infants (SELSI; Kim, 2002).

Multivariable linear regression analyses revealed that a higher proportion of touch+word co-occurrence, increased word-touch overlap, and more precise alignment between touch and word boundaries positively predicted infants' receptive language scores. However, no significant effects were observed for expressive language outcomes.

These findings highlight the beneficial impact of temporal synchrony between maternal touches and speech on infants' receptive language acquisition. The study sheds light on how synchronized multimodal cues accompanying speech can influence infants' language acquisition.

장애음성 평가 및 진단을 위한 한국어 자음 spectra 에너지 분포 특성

강 동 효*, 성 철 재[§]
충남대학교 언어병리학과, 충남대학교 언어학과

Characteristics of Spectral Energy Distribution in Korean Consonants for Assessing and Diagnosing Disordered Speech

Donghyo Kang, Cheoljae Seong

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

[§]Dept. of Linguistics, Chungnam National University

dhkangs@hanmail.net, cjseong49@gmail.com

20, 30대 서울말 화자 남녀 총 20명을 대상으로 인터뷰하여 녹음된 한국어 자연발화 음성 코퍼스인 서울 코퍼스를 이용하여 한국어 자음의 스펙트럼 적률(moment) 특성을 살펴보았다. 결과값들은 추후 조음 및 음성장애 환자들의 수치와 비교분석하여 환자들의 말소리 오류 특징 및 치료 후 개선 여부 등을 살펴보는 데 객관적인 기초 자료로 활용될 수 있을 것이다.

독립변수는 성별(남/녀), 조음방법(파열음/마찰음/파찰음/비음/유음), 모음환경(아/에/이/오/우)으로 하고, 6가지 음향변수({적률변수: 무게중심, 표준편차, 왜도, 첨도}+{스펙트럼 기울기 변수: 에너지차 기울기, 추세 회귀 기울기})를 종속변수로 하여 3원 분산 분석(3-way ANOVA)을 실시하였다. 이후 각 조음방법에 대한 음향변수 간의 상관관계를 살펴보았다. 특히 표준편차와 첨도의 특성을 알아보기 위하여 각 변수를 무게중심과 왜도와의 관계를 중심으로 분석하였다.

결과적으로 6가지 음향변수 모두 성별*조음방법*모음환경 사이에 유의한 상호작용이 관찰되었다. ‘성별’ 간 비교에서, 여성이 파열음과 공명 자음을 조음할 때 스펙트럼 에너지 분포 추세가 남성보다 더 완만하게 하강하며, 마찰음, 파찰음을 조음할 때는 남성보다 가파르게 상승한다는 것을 알 수 있었다. 마찰음과 파찰음 측면에서는 여성의 에너지가 남성보다 더 고주파에 몰려있었으며, 파열음 측면에서는 여성의 주파수 대역이 남성보다 넓고 평평한 분포를 보였다. ‘조음방법’의 관점에서, 무게중심, 표준편차, 에너지차 기울기, 추세 회귀 기울기는 장애음>공명 자음의 경향이 있었고, 왜도와 첨도는 공명 자음>장애음으로 관찰되었다. ‘모음환경’의 경우, 장애음은 /아/, /오/ 모음에서 유의한 차이가 있는 경우가 많았다. 조음방법에 따른 ‘표준편차-무게중심’, ‘첨도-왜도’의 관계를 살펴보면, 파열음은 모두 정적 상관관계를 이루고 있었으며, 마찰음은 모두 부적 상관관계를 보였다. 파찰음은 상관관계가 낮은 편이었으며, 공명 자음은 정적 상관관계를 보였다.

대중가요 보컬 전공 남학생의 성종에 따른 모음 간 음역 차이*

지 일 송, 김 재 욱
강남대학교 교육대학원 언어치료교육전공

Vocal Range Differences in Vowels among Male Students for Popular Music Vocals by Vocal Types

Il-Soung Ji, Jaeock Kim*

*Major in Speech Pathology Education, Graduate School of Education, Kangnam Univeresity
jy1070@naver.com, jaeock@gmail.com

본 연구는 대중가요 관련학과에서 보컬을 전공하고 있거나 입시를 준비하는 남성 27명을 대상으로 자신의 성종과 음역을 제대로 파악하고 있는지 살펴보기 위해 주관적으로 평가한 성종과 음향학적으로 평가한 성종 간에 차이가 있는지를 비교하였고, 성종 간 기본주파수(fundamental frequency, F0)와 평균발화기본주파수 간의 차이도 비교하였다. 또한, 이들이 모음에 따라 고음역에서 산출할 수 있는 정도가 다를 수 있으므로 극모음인 /a/, /i/, /u/ 모음 간 음역의 차이와 성종별로 구분하여 모음 간 음역의 차이가 있는지를 살펴보았으며, 모음별로 성종 간에 음역의 차이가 있는지도 비교하였다. 연구결과, 대중가요 보컬을 전공하거나 입시를 준비하는 남성의 절반 이상이 본인의 성종을 정확하게 파악하지 못하고 있으며, 모음 간에 최대기본주파수와 주파수범위는 통계적으로 유의한 차이를 보여 모음의 종류에 따라 산출할 수 있는 음역의 차이가 있음을 알 수 있었다. 특히 모음 /a/ > /u/ > /i/ 순으로 음역이 감소하였다. 이는 모음 /a/가 다른 모음에 비해 고음역에서 조음하기에 용이하지만, 고모음 /u/와 /i/는 혀의 위치가 상승함으로써 구강 면적이 좁아지고 후두의 상승이 동반되기 때문에 음역이 감소하고 고음에서 조음하는데 어려움이 초래된다는 것을 알 수 있다. 본 연구는 대중가요를 전공하는 전공생과 입시생을 대상으로 성종과 모음에 따른 음역을 비교함으로써 대중가요 보컬을 전공하고자 하는 남성이 자신의 산출 가능한 음역을 좀 더 인지하고, 모음에 따라 산출할 수 있는 음역이 무엇인지를 파악하여 무리하게 고음역을 산출하는 음성 오·남용을 방지할 필요가 있음을 제시하였다.

* 본 연구는 제1저자의 석사학위 논문의 일부를 수정한 것임

2024 한국음성학회 봄 학술대회

포스터 발표 논문

POSHA-S를 통한 직업군인의 말더듬 인식 연구

차 화 정, 박 진
가톨릭관동대학교 언어재활학과

Perceptions of Military Personnels towards Persons who Stutter: Using Public Opinion Survey of Human Attributes-Stuttering(POSHA-S)

Hwajung Cha, Jin Park
Dept. of Speech-Language Rehabilitation, Catholic Kwandong University
slpcha@kakao.com, gatorade70@cku.ac.kr

자신의 개성을 중요시하고 자유분방한 생활을 추구하는 요즘 신세대 청년들에게는 군대라는 환경에 적응하기가 쉽지 않을 수 있다(Kim & Kim, 2014). 이러한 상황에서 의사소통의 문제 또는 이로 인한 삶의 제한성을 경험하고 있는 말더듬 성인에게 있어 군대는 상대적으로 큰 두려움을 초래하는 가능하면 회피하고 싶은 '상황'으로 인식될 수 있다. 군대라는 급격한 환경의 변화에 말더듬 성인의 원만하게 군대 생활을 위해서는 (이들의) 잠재적 상관으로서 지휘 또는 통솔하는 위치에 있는 군지휘관들이 말더듬이나 말을 더듬는 사람에 대해 어떤 인식이나 조망(perspective)을 가지고 있는지가 매우 중요할 수 있다. 만약 부정적 인식으로 인해 인지적, 정서적, 행동적 부적 반응(negative reactions)을 보이게 된다면 말더듬 성인들이 성공적인 군대 생활을 할 가능성은 낮아질 수밖에 없기 때문이다.

본 연구에서는 기본적으로 군지휘관을 대상으로 (이들의) 말더듬이나 말을 더듬는 사람에 대한 인식과 태도를 알아보았다. 이를 위해 현재 군 복무 중이며 일반병사를 지휘, 통솔한 경험이 있는 총67명의 영관급 장교를 포함한 부사관급 이상의 직업군인을 대상으로 '인간 특성에 대한 대중 인식 설문지-말더듬'(Public Opinion Survey of Human Attributes - Stuttering, POSHA-S, St. Louis, 2012a)을 이용하여 설문을 실시하였다. POSHA-S는 말을 더듬는 사람에 대한 대중의 태도나 인식을 조사하기 위해 개발된 설문도구로 현재 가장 많이 사용되고 있는 설문도구 중 하나이다(Chon, 2016). 1999년에 개발된 이후 여러 번의 수정절차를 거쳐 완성되었으며 설문도구의 검사재검사 신뢰도, 내적일관성 신뢰도, 구성타당도 및 공인타당도가 입증되었고(St. Louis, 2005, 2012a, 2012b; St. Louis, Lubker, Yaruss, & Aliveto, 2009; St. Louis, Reichel, Yaruss, & Lubker, 2009), 2014년 11월 기준으로 37개 국가에서 23개 언어로 번역되어 다양한 국가 및 문화권에서 사용되고 있다(St. Louis, 2014).

결과적으로 직업군인의 비교적 낮은 인식 정도를 확인할 수 있었으며 군지휘관을 대상으로 한 말더듬이나 말을 더듬는 사람에 대한 인식 개선 교육이 필요함을 알 수 있었다.

References

- Kim, O. R., & Kim, W. I. (2014). Development and validations of military life adjustment scale for military soldiers. *Korean Journal of Counseling*, 15(4), 1627-1646.
- St. Louis, K. O. (2012a). Research and development on a public attitude instrument for stuttering. *Journal of Communication Disorders*, 45, 129-146.

갑상선 수술 후 증후군의 빠른 음성회복을 위한 음성치료의 효과

김지성

건양대학교 보건복지대학원 언어치료학과

Effectiveness of Early voice recover therapy for Post-Thyroidectomy Syndrome

Jisung, Kim

Graduate school of Public Health & welfare, Konyang University
slp2046@konyang.ac.kr

갑상선 수술 후 신경손상 없이도 나타나는 음성문제를 갑상선 수술 후 증후군(Post-Thyroidectomy Syndrome, PTS)이라고 한다. 본 연구는 갑상선 수술 후 증후군의 조기음성회복을 위해 고안한 음성치료법(Early Voice Recovery Therapy, EVRT)의 효과를 확인하기 위한 무작위 배정연구(randomized controlled trial)이다. 연구대상자는 12명이며, (1) 두경부 전문의에 의해 PTS로 진단되었고 (2) 수술 중 신경손상이 없음이 수술자에 의해 보고되었으며 (3) 주관적인 음도문제를 호소하고 (4) 음성언어재활사에 의해 음도문제가 있음이 확인되며, 음성치료가 필요하다는 의견을 나타낸 경우 선정되었다. EVRT는 음성언어재활사에 의해 시행되는 직접치료 단계와 대상자 스스로 수행하는 자가훈련단계로 구성되어 있다. EVRT-I는 갑상선 절제술 2주 후 1회기 시행하였으며, EVRT-II는 EVRT-I 이후 2주간 수행하도록 하였다. 음성측정은 EVRT-I(직접치료) 수행 전과 후, EVRT-II(자가훈련)후 실시되었다. 연구결과, 음향학적 분석 변인인 jitter, shimmer, NHR은 EVRT-I(직접치료) 수행 전에 비해 수행 후 감소를 보였다. 또한, shimmer, NHR은 EVRT-I(직접치료) 수행 후에 비해 EVRT-II(자가훈련)후에도 감소를 나타내었다. 또, PTS의 주요한 음성문제인 음도에서의 어려움을 나타내는 VRP, max VRP는 EVRT-I(직접치료) 수행 전에 비해 후 증가하였고, EVRT-II(자가훈련)후도 EVRT-I(직접치료) 전과 후보다 증가를 보이며 EVRT 수행에 따라 증가하는 모습을 보였다. 이같은 결과는 EVRT가 PTS로 인한 음성문제를 개선하는데 효과적임을 시사한다.

청각적 자극 및 시각적 자극에 따른 말소리장애 아동의 단어 회상하기 과제의 수행 비교

김 주 원*, 하 지 완**

대구대학교 대학원 재활과학과 언어치료전공*, 대구대학교 언어치료학과**

Performance of word recall tasks according to auditory and visual stimuli between children with and without speech sound disorders

Ju-won Kim, Ji-Wan Ha

Major of Speech-Language Pathology, Dept. of Rehabilitation Science, Graduate School, Daegu University*

Dept. of Speech-Language Pathology, Daegu University**
wdsct@naver.com, ji-ha@daegu.ac.kr

1. 서론

말소리장애(Speech Sound Disorders, SSD)는 아동기에 빈번하게 발생하는 의사소통 장애이다 (Murphy et al., 2014). 일반아동(TD)과 비교하여 말소리장애 아동은 음운처리, 음운인식, 음운표상, 음운단기 기억 등 말소리 정보 처리능력에 결함을 보인다. 단기기억은 음운적 단기기억과 의미적 단기기억으로 구분할 수 있다. 말소리를 처리할 때, 음운 통로와 어휘-의미 통로가 동시에 기능을 하고 두 통로가 서로 긴밀하게 정보를 주고 받는다(Martin, 1999).

본 연구는 말소리 산출 시, 제시되는 자극에 따른 의미정보, 음운정보 활성화가 말소리장애 아동에게 미치는 영향을 알아보기 위한 목적으로 진행되었다.

2. 연구방법

2.1 연구대상

본 연구는 본 연구는 4세에서 6세의 일반 아동 20명, 순수 말소리장애 아동 10명, 언어발달지연을 동반한 말소리장애 아동10명 총 40명을 대상으로 실시하였다.

2.2 연구 절차

본 연구는 조용한 환경에서 개별적으로 진행하였고, 실험 시작 전 연구에 대한 사전 동의를 받았다. 기본정보를 수집하고 우리말 조음음운 평가2(Urimal-Test of Articulation and Phonology; U-TAP2, 김영태 외, 2020), 수용·표현 어휘력검사(Receptive & Expressive Vocabulary Test; REVT; 김영태 외, 2009) 실시하여 대상자가 본 연구대상에 해당하는지 확인하는 절차를 가졌다. 실험과제 시작하기 전에 앞서 연구자는 대상자에게 실험에 해당하는 단어를 모두 습득하고 있는지 확인하는 과정을 가졌고, 세가지 단어를 시각 및 청각 자극으로 제시한 후 순차회상하도록 요구하였다. 대상자가 지시사항을 제대로 이해하였는지 확인하기 위해 연습과제를 실시한 후 본 과제를 진행하였다.

본 연구과제는 시각자극, 청각자극 및 범주유형(범주 내, 범주 외)을 포함하여 구성하였다. 연구과제인 단어회상 과제는 초등학교 아동들의 우리말 범주 기준 조사(전덕임, 강연옥, 2007), 인공와우이식 아동의 범주 유형에 따른 의미적 단기기억(장성진, 2012)을 활용하여 제작하였다. 과제 항목은 시각범주내, 시각범주외, 청각범주내, 청각범주외 각 12개로 총 48개의 단어로 구성하였다.

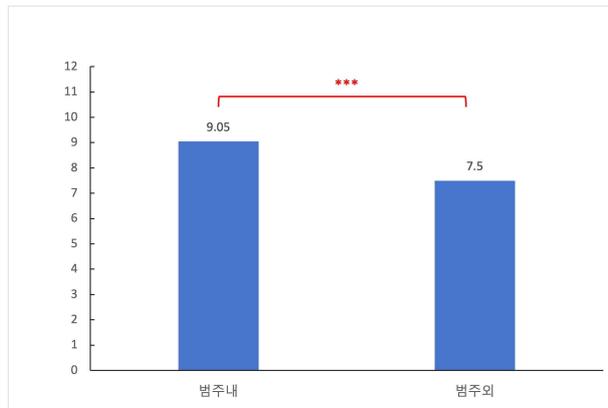
단어 항목에 따라 난이도 차이가 나는 것을 통제하기 위하여 시각, 청각 단어구성을 A, B세트로 나누어 실시하였다. 모든 과제는 Microsoft PowerPoint를 사용하여 정오반응을 측정하였다.

2.3 자료분석

자극이 제시된 후 대상자들이 산출한 세 가지 단어를 점수화 하였다. 순차회상 시 제시된 단어를 순서대로 산출하였을 때 단어 당 1점, 오반응 시 0점으로 계산하여, 총 48점 만점이었다.

3. 연구결과

혼합설계에 따른 반복측정 분산분석 결과, 집단 간 주효과는 유의하였다($F(1, 37)=8.708, p<.001$). SSD+LD 집단은 TD와 pure SSD 집단보다 유의하게 낮은 수행력을 보였으나, pure SSD 집단과 TD 집단 간에는 유의한 차이가 나타나지 않았다($p>.05$). 범주유형에 따른 주효과는 유의하여($F(1, 37)=32.988, p<.001$), 이에 대한 사후검정을 실시하였다. 그 결과 범주내 과제는 범주외 과제보다 유의하게 점수가 높다는 것을 알 수 있었다($p<.001$) (그림 1). 그밖에 자극양식에 대한 주효과, 자극양식과 범주유형 간 상호작용효과, 자극양식, 범주유형 및 집단 간 상호작용효과는 모두 유의하지 않았다($p>.05$).



<그림 1> 범주유형에 따른 과제 수행력

4. 결론

본 연구는 조음능력, 어휘력에 따른 시각, 청각 자극 및 범주유형이 단어회상과제에 미치는 영향을 알아보기 위해 수행되었다. 그 결과, 첫 번째로 세 집단 간 단어회상과제 점수를 비교했을 때, 언어발달지연을 동반한 말소리장애 아동의 점수가 가장 낮았다. 이를 통해 낮은 언어 능력과 조음능력이 대상자의 단기기억 능력에 어려움이 생겨 단어회상과제 점수가 낮아짐을 볼 수 있었다. 두 번째, 범주유형을 고려하여 비교하였을 때, 범주내, 범주외 점수에 유의하다고 나타났다. 즉 자극이 제시 될 때 범주유형이 단어회상에 유의한 영향을 줄 수 있음을 알 수 있다.

본 연구에서 범주유형에 따라 단어회상 및 음운루프에 어려움이 생겨 언어지연을 동반한 말소리장애 아동의 자극에 따른 범주외 점수가 저하되는 것으로 나타났으며 각 자극 및 범주에 따라 점수의 양상이 다르게 나타났다. 본 연구는 각 자극 및 범주에 따른 말소리장애의 말소리 회상 및 단기기억에 의미 있는 정보를 제공한다.

References

- 김영태, 신문자, 김수진, 하지완 (2020). 우리말 조음·음운평가 2(UTAP2). 서울; (주)인싸이트.
- 김영태, 홍경훈, 김경희, 장혜성, 이주연 (2009). 수용·표현어휘력검사(REVT). 서울:서울장애인종합복지관
- 전덕임, 강연욱(2007). 초등학교 아동들의 우리말 범주 기준 조사. 언어청각장애 연구, 제12권, 제1호, 90-104.
- 장성진 (2012). 인공와우이식 아동의 범주 지식에 따른 의미적 단기기억. 대구대학교 대학원 석사학위논문.
- Murphy, C. F., Pagan-Neves, L. O., Wertzner, H. F., & Schochat, E. (2014)
- Martin,R.C.,Bredinb, S.D.&Damian, M.F.(1999). The Relation of Phoneme Discrimination, Lexical Aces, and Short-Term Memory: A Case Study and Interactive Activation Account. Brain and Language, 70,437-482.

아동기말실행증의 말소리 감별진단도구 마련을 위한 예비연구

이 옥 분, 박 상 희
대구사이버대학교 언어치료학과

Preliminary Study for Speech Sound Differentiation Diagnosis of Childhood Apraxia of Speech

Okbun Lee, Sanghee Park
Dept. of Speech and Language Pathology, Daegucyber University
oble@dcu.ac.kr, psh4292@dcu.ac.kr

1. 서 론

아동기말실행증(childhood apraxia of speech, CAS)의 말소리 문제는 기능적 조음음운장애나 마비말장애(dysarthria)와는 구분되는 양상들이 있다. 마비말장애는 뇌손상의 원인으로 마소리 산출을 담당하는 호흡, 발성, 공명, 조음, 운율 등의 시스템 관련 근육들의 마비와 약화로 인한 말장애이다(ASHA, 2007). 반면 CAS는 뇌병변이 아닌 신경계의 말소리 처리과정에서의 프로그래밍 오류에 기인한다고 설명된다. 또한 기능적 조음음운장애처럼 조음기관의 구조적인 문제가 없으니 운동(기능)적인 측면에서 유사한 점들이 많이 있는데, 조음오류 양상에서는 다소 차이가 존재한다. 그러나 실제 임상에서는 발달장애 아동들의 조음오류 문제에서 기능적(구조적)인 결함인지, CAS에 비중을 두어 구분해야할지 그 가이드라인 될 수 있는 평가 도구들이 매우 부족하다. 국내의 제한적인 임상연구 상황과 달리 국외에서는 꾸준히 관련 연구들이 발표되고 있어서 진단 평가를 위해 활용하고 있다. 그러나 CAS의 말소리 특성에서 각 나라의 언어학적인 특징에 따른 차이가 존재하기 때문에, 국내 효과적으로 적용하기 위한 사전 신뢰도 및 타당도 연구들이 필요하다. 따라서, 이 설문연구에서는 국외 CAS의 말소리 감별진단 도구 개발을 위한 연구(Iuzzini-Seigel, et. al., 2022)를 토대로 이를 번안하여 국내 언어재활사들 대상으로 그 적절성과 모국어(한국어)의 적용성에 대해 어떻게 인식하는지를 예비적으로 알아보려고 한다. 이 연구를 토대로 추후 보다 체계적인 CAS 평가 프로토콜을 개발하고자 한다.

2. 연구방법

본 설문에 응답한 언어재활사(평균 임상경력 5년~10년 이상)는 모두 18명이었다(학력은 박사 수료이상 7명, 석사 수료 및 석사가 10명, 학사가 1명-1급 언어재활사 16명, 2급 언어재활사 2명, 근무지 유형은 사설치료센터, 학교, 기타 등). CAS 치료경험은 58.8%가 있었고 41.2%가 없었다. 설문은 Iuzzini-Seigel, et. al., (2022)의 연구를 토대로 CAS 말소리 특징을 번안하고, 이에 대해 국내 언어재활사들이 CAS 감별 특징과의 관련성에 대해 5점 척도([1]전혀 해당되지 않음~ [5] 매우 해당됨)로 체크하였다. 설문 결과에 대해서 빈도 분석으로 해석하였다.

3. 연구 결과

설문에 응답한 언어재활사들의 반응을 분석한 결과 다음과 같다. CAS의 말소리의 감별진단 증상에 관한 영역별로, 첫째, [1. 호흡-발성 영역]에서는 대체로 관련성이 매우 낮은 편으로 나타났다. 관련 하위 영역인 ‘성량, 얇은 호흡, 과도한 강도, 음질’ 영역에서 모두 낮은 비율을 보였고, ‘큰 강도변화’영역에서 약 50%정도 해당되는 것으로 인식하였다. 둘째, [2-공명 영역](비성-과다비성, 과소비성) 부분에서도 관련성이 매우 낮은 것으로 나타났다. 세 번째, [3. 운율과 속도 영역]의 말소리 증상들에서도 매우 높은 관련성이 나타났다. 관련한 CAS 말소리 증상인 ‘느린 말속도, 말소리 강세 변화, 어휘강세 오류’에 대해 모두 높은 관련성이 있음이 나타났다. 넷째, [4. 조음영역]의 말소리 특징들에서는 거의 93.8%이 CAS 말소리 자질에 해당된다고 인식하였다. 관련한 세부 말소리 증상들-부적절한 조음 접촉, 자음왜곡, 모음오류, 유성음화(68.8% 해당), 중성화(78.8% 해당), 모색행동 (81.3%가 확실히 해당), 다음절어 오류(87.5%가 해당). 초기 발음 형성 어려움(75%가 해당) 등으로 나타났다. 다섯 번째, 위 4개의 영역에서 CAS의 말소리 감별진단 특징으로 한국어 적용에 대한 적절성을 알아본 결과 조음영역에서 100%와 운율과 속도영역에서 87.6%로 가장 높게 나타났다.

4. 결론

본 설문 결과 호흡에 대해서는 전반적으로 CAS의 말소리 감별 진단 특징과는 관련성이 거의 없다고 인식하고 있었으며, 반면 속도와 운율, 조음 영역에 해당하는 말소리 자질들은 CAS 진단에 중요한 변수로 판단하고 있다는 것을 알 수 있었다. 더불어 이러한 관련성이 매우 높은 자질들은 한국어 적용에도 적절하다고 판단하고 있었다. 이 결과는 기존의 보고된 CAS 말소리의 특징을 지지하고 있다 (Iuzzini-Seigel, 2017; Strand, 2014.). 하지만, 한국어 사용 아동을 대상의 CAS 말소리 감별 진단도구 개발의 목적으로 실시한 기초 예비연구이기에, 추후 연구에서는 보다 많은 대상의 임상경력을 가진 언어재활사를 중심으로, 그리고 CAS 감별 주요 변수들에 대해 보다 섬세하게 그 관련성을 확인할 수 있는 척도(Strand, et. al., 2014)를 개발하여 심층적으로 연구할 필요가 있다.

References

- American Speech-Language-Hearing Association. (2007). Childhood apraxia of speech [Technical report]. <https://doi.org/10.1044/.policy.TR.2007-00278>.
- Iuzzini-Seigel, J. (2017). Speech assessment in children with childhood apraxia of speech. *Perspectives on Neurogenic Communication Disorders*, 2, 47 - 60.
- Iuzzini-Seigel, J., Allison, K. M., & Stoeckel, R. (2022). A tool for differential diagnosis of childhood apraxia of speech and dysarthria in children: A Tutorial. *Language, Speech, and Hearing Services in Schools*, 53, 926 - 946.
- Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The Apraxia of Speech Rating Scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders*, 51, 43 - 50. <https://doi.org/10.1016/j.jcomdis.2014.06.008>

게임 닉네임의 독음 후보군 생성 방법*

김 나 현^{?)}, 김 연 정^{?)}, 이 용 혁^{?)}, 조 남 현^{?)}
 서울대학교 언어학과^{?)}, NCSOFT^{?)}

Algorithm for generating candidates by which game players call game nicknames

Nahyun Kim^{?)}, Yeonjeong Kim^{?)}, Yonghyuk Lee^{?)}, Namhyun Cho^{?)}
 Dept. of Linguistics Seoul National University^{?)}, NCSOFT^{?)}
 {nahyun0410, kyj219} @snu.ac.kr & {eug92, cnh2769} @ncsoft.com

게임 중 음성 인식기를 활용하여 상대방을 호출할 때, 게임 닉네임의 음소를 변형하여 발음하거나, 축약하는 등 닉네임을 문자 그대로 읽지 않는다. 따라서, 하나의 닉네임에도 다양한 독음이 존재한다. 그러므로 다른 유저의 닉네임을 자연스럽게 호출하려면, 각 닉네임의 원형과 적절한 독음 후보군을 호출어 인식기의 입력 값으로 넣어야 한다. 이를 위해 10음절 내의 닉네임을 입력하면 자동으로 독음 후보군을 생성하는 모델을 제작하였다.

먼저 전처리 과정으로, 닉네임 내 발음이 불가능한 음절은 정규화하고 동음 반복은 제거하였다. 예를 들어, '꺽꺽뽀'는 '꺽주'로, '강북뜨뜨뜨'는 '강북뜨'가 되도록 하였다. 또한, 닉네임은 게임 내의 고유 식별자이자 대상에 존재성을 부여한다는 특성을 고려하여, 고유명사가 포함되어 있을 경우 통사 구조 및 단어의 출현 위치와 관계없이 독음으로 채택하였다. 뿐만 아니라, 고유 명사에 해당하지 않더라도 한국어 성씨로 시작하거나 호격 종결 어미, 사람의 이름 뒤에 쓰이는 의존 명사, 서술격 조사와 결합하면 고유명사처럼 여겨지는 경향을 반영하였다. 예를 들면, '강판다'라면 '판다'를, '궁성님'이라면 '궁성'을, '나는토끼야'라면 '토끼'를 독음 후보군에 추가하였다.

상기 경우에 해당하지 않거나, 과정을 거친 후에도 여전히 3음절 이상인 경우, 축약형을 독음으로 채택하였다. 먼저, 문장 구조가 복잡하지 않은 5음절 이내의 닉네임은 특정 어절을 선택하여 축약형으로 선정하였다. 이때 기존의 띄어쓰기 분석기가 신조어 및 비속어에 보이는 낮은 성능을 개선하고, 문어적 끊어읽기 단위와 구어적 띄어쓰기 단위의 괴리를 수정하기 위해 별도의 띄어쓰기 교정을 수행하였다. 6음절 이상의 닉네임은 첫 어절을 중심으로 축약하되, 문법적인 경계와 의미를 고려하여 뒷부분의 키워드도 추출해 축약어로 활용하였다.

성능 평가를 위해 50명을 대상으로 100개의 testset에 대해 설문조사를 실시하여 benchmark를 제작하였다. 이후 Overlap coefficient 지표를 활용하여 본 모델이 생성한 결과와 비교하였다. 선행 2음절만 추출하는 baseline과 비교하였을 때 본 모델이 평균 약 20%p 우수한 성능을 보였다.

* 본 연구는 NCSOFT 산학협력 단기 인턴 프로젝트로 수행되었습니다.

정보 교란 기법과 스피커 인코더를 활용한 한국어 제로샷 음성 합성 모델*

방 채 운¹, 전 찬 준², 김 홍 국¹
¹광주과학기술원, ²조선대학교

Korean Zero-Shot Speech Synthesis Model using Information Perturbation Techniques and Speaker Encoder

Chae Woon Bang¹, Chan Jun Chun², Hong Kook Kim¹
¹Gwangju Institute of Science and Technology (GIST)
²Chosun University
 {¹bang4045@gm., ¹hongkook@}gist.ac.kr, ²cjchun@chosun.ac.kr

음성 합성이란 텍스트로부터 음성을 생성하는 기술을 말한다. 이러한 음성 합성 기술의 목적은 화자와 유사하고 잡음이 없는 음성을 생성하는 것이다. 최근에는 단일 화자를 넘어 다수의 화자에 대해 높은 성능을 보이는 음성 합성 모델이 제안되고 있다. 하지만 여전히 학습되지 않은 화자에 대해서 음성을 합성하는 것은 어려운 일이다. 또한, 학습되지 않은 화자에 대해 음성을 생성할 경우 생성된 음성의 품질과 화자 유사도가 낮다는 것이다. 이는 모델이 충분히 많은 화자 정보를 학습하지 못했기 때문에 발생한다. 따라서 본 논문에서는 효과적인 제로 샷 음성 합성을 위해 정보 교란 기법과 스피커 인코더를 활용한 한국어 제로샷 음성 합성 모델을 제안한다. 제안하는 방법은 확산형 확률 모델 기반의 음성 합성 모델인 Grad-TTS를 기반으로 한다. 먼저, 사전 학습된 스피커 인코더를 사용하여 음성 프롬프트로부터 화자 정보를 추출한다. 이를 통해 모델은 학습되지 않은 화자에 대해서 음성을 합성할 수 있도록 한다. 또한, 정보 교란 기법을 활용하여 학습 데이터에 등록된 화자뿐만 아니라 더 많은 화자 정보를 모델이 학습할 수 있도록 한다. 이는 모델이 다양한 화자 정보를 학습하여 학습되지 않은 화자에 대해 효과적인 음성 합성이 가능하게 한다. 제안된 모델의 성능을 평가하기 위해 Neural MOS와 SECS를 측정하였다. 성능 평가를 위해 남성 화자 5명, 여성 화자 5명을 뽑아 각각 10개의 음성 샘플을 생성하였으며, 평가에 사용된 화자는 모두 학습되지 않은 화자이다. 측정 결과 원본 음성의 평균 Neural MOS 점수 3.0330과 비교하여, 생성된 음성은 평균 2.8480으로 원본과 유사한 수준의 음성 품질을 보이는 것을 확인할 수 있다. 또한, 화자 유사도는 평균 0.5088로 프롬프트 음성과 유사한 음성을 생성하는 것을 확인할 수 있었다.

* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 연구개발특구진흥재단의 ‘지역의 미래를 여는 과학기술 프로젝트’ 사업(인공지능 기반 메타버스 구현을 위한 융·복합 문화 가상 스튜디오, 2022-DD-UP-0312)과 2024년도 광주과학기술원의 재원으로 “과학기술혁신사업단의 실용화 연구개발사업”의 지원을 받아 수행된 연구임.

자기 지도 학습 모델 특징 결합을 이용한 적대적 신경망 기반 역양 음성인식*

경나, 박정식
한국의국어대학교 언어공학연구소

Accented Speech Recognition based on Self-supervised Feature Fusion with Domain Adversarial Training of Neural Networks

Na Geng, Jeong-Sik Park
Language Technology Research Institute, Hankuk University of Foreign Studies
gengna0324@gmail.com, parkjs@hufs.ac.kr

자기지도 학습(Self-Supervised Learning, SSL)은 학습 데이터 부족 문제를 해결하는 데 큰 성과를 거두었으며, 음성 인식 성능 향상에도 큰 기여를 하고 있다. 또한, 도메인 적응 등 다른 기법과 결합하는 연구가 소개되어 왔는데, 가령 역양 음성 인식 분야에서 도메인 적대적 신경망(Domain Adversarial Neural Network, DANN)을 활용한 자기지도 학습 기반의 역양 음성인식 기술은 도메인 간 차이를 극복하고 역양 음성인식의 정확도를 향상시키는 데 기여하였다. 그러나 기존의 연구들은 대부분 단일 사전 훈련 모델을 사용하고 있는데 다양한 구조의 사전 훈련 모델을 결합하면 보다 다양한 음성 특성을 추출함으로써 보다 개선된 성능을 기대할 수 있다. 본 연구에서는 자기 지도 학습 모델들의 특징 결합을 이용한 적대적 신경망 기반 역양 음성인식 기법을 제안한다. 즉, 두 개의 자기 지도 학습 모델을 사용하여 음성 특징을 추출하고 이를 결합한 네트워크가 DANN의 Feature Extractor 역할을 수행한다. 먼저, WAV2VEC2-BASE 모델과 HUBERT-BASE 모델을 결합한다. HUBERT 모델은 군집화를 통해 생성된 레이블을 활용하여 음소 등 오디오 단위의 고급 특징을 학습함으로써 다양한 역양의 특징을 효과적으로 파악하며, WAV2VEC2 모델은 음성 파형에서 직접 특징을 학습하여 장기 의존성을 이해하고 문맥을 학습하는 데 집중한다. 두 모델을 결합하면 각 역양의 포괄적인 특징을 추출함으로써 역양 음성인식의 정확성과 견고성이 향상될 것으로 예상된다. 다음으로, 미세조정을 통해 훈련된 ASR 모델과 HUBERT-BASE 모델을 결합함으로써, 기본적인 ASR의 정확성을 유지하면서 HUBERT 모델을 통해 역양 특징을 더 효과적으로 추출할 수 있다. 제안한 방법의 유효성을 검증하기 위해 Common Voice 데이터를 사용하여 역양 음성 인식 실험을 수행하였다. 미국 역양 영어 데이터를 Source Domain 데이터로, 인도와 홍콩 역양 영어 데이터를 Target Domain 데이터로 활용하여 DANN 기반의 역양 음성 인식 모델을 구축하고 성능을 평가하였으며, 실험 결과 제안한 방법을 통해 인식 성능이 크게 향상됨을 확인하였다.

* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.2020R1A2C1013162)

음성 향상을 위한 게이트 합성곱 순환 신경망*

고형일¹, 나승대², 김명남³¹경북대학교 의용생체공학과, ²경북대학교병원 의공학과,³경북대학교 의과대학 의학과 의공학교실

Gated convolutional recurrent neural networks for speech enhancement

H.L. Koh¹, S.D. Na², and M.N. Kim³¹Dept. of Medical & Biological Eng., Graduate School, Kyungpook Nati. Univ.,²Dept. of Biomedical Eng., Kyungpook Nati. University Hospital,³Dept. of Biomedical Eng., School of Medicine, Kyungpook Nati. Univ.

e-mail : kimmn@knu.ac.kr

요 약

딥러닝 기반 음성 향상 시스템에서 합성곱 인코더-디코더 네트워크는 주목할만한 성과를 보였으나, 영상의 질감과 같은 기본 패턴을 감지하도록 설계된 합성곱 커널을 통해 인코딩된 특징만으로는 음성의 중요 특성을 표현하기에 충분하지 않다. 이러한 과제를 해결하기 위해 본 논문에서는 네트워크에서 추출된 특징의 가중치를 계산하는 게이팅 메커니즘을 도입하여 음성 향상 작업에 필요한 핵심 정보만 전달하는 새로운 음성 향상 모델을 제안한다. 제안한 네트워크는 인코더-디코더 및 LSTM(Long Short-Term Memory)으로 구성되며, 게이트 알고리즘은 스킵 연결 사이에 결합되어 모델의 전반적인 성능을 향상시킨다. 실험 결과, 제안된 방법은 음성 품질을 개선하고 잡음을 억제하는데 있어 다양한 객관적 평가 지표에서 기존 모델보다 유망한 향상 능력을 보여주었다.

1. 서론

음성 향상은 다양한 소음 환경에서 발생하는 간섭으로 인해 저하된 음성의 명료도와 품질을 개선하는 오디오 신호 처리의 연구 분야이다. 음성 비서, 디지털 보청기, 원격 회의 등 다양한 환경에서 고품질 통신 시스템에 대한 수요가 증가함에 따라 음성 향상은 전반적인 음성 명료도를 높이고 사용자 경험을 개선할 수 있다. 딥러닝의 발달과 함께, 음성 향상 시스템에서도 심층 신경망(deep neural network)을 활용한 데이터 기반 접근 방식이 연구되고 있으며 기존의 통계적인 신호 처리 방법보다 더 정확한 잡음 억제와 더 높은 음성 품질 성능을 보이며 주목받고 있다. 일반적으로 딥러닝 기반 음성 향상 시스템에서 다양한 스케일의 특징 추출 및 활용은 모델의 성능과 밀접한 연관성을 갖는다. 이에 따라 계층별 특징 추출이 용이한 합성곱 인코더-디코더 네트워크가 음성 향상 분야에서 도입되었다. 예를 들어 Hu 등이 제안한 DCCRN은 기존의 합성곱 인코더-디코더 구조를 복소수 연산이 가능하도록 확장하고 데이터의 시간적 상관관계를 포착하기 쉬운 순환 신경망을 결합하여 STFT(short time fourier transform) 스펙트럼을 직접 추정하는 방식으로, Interspeech 2020 DNS 챌린지에서 우수한 성능을 보였다[1]. 그러나 DCCRN의 인코더에서 추출된 특징들은 음성뿐만 아니라 잡음에 관련된 정보들도 중복된 형태로 존재하기 때문에 모델의 원활한 음성 추정을 방해할 수 있다.

본 논문에서는 게이트 메커니즘이 네트워크 사이의 정보 흐름을 제어하여 더 복잡한 상호작용을 모델링 할 수 있다는 점에 착안하여[2], 게이트 합성곱 블록이 결합된 새로운 음성 향상 모델을 제안하였다. 모델은 인코더-디코더 구조로 설계되었으며 게이트 블록은 스킵 연결 사이에 적용되어 계층 간 정보 전달 능력을 개선한다.

* This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (Nos. NRF-2022R1A2C2009716, RS-2023-00279169).

2. 연구 방법

그림 1은 본 논문에서 제안한 음성 향상 모델의 전체적인 구조를 나타낸다.

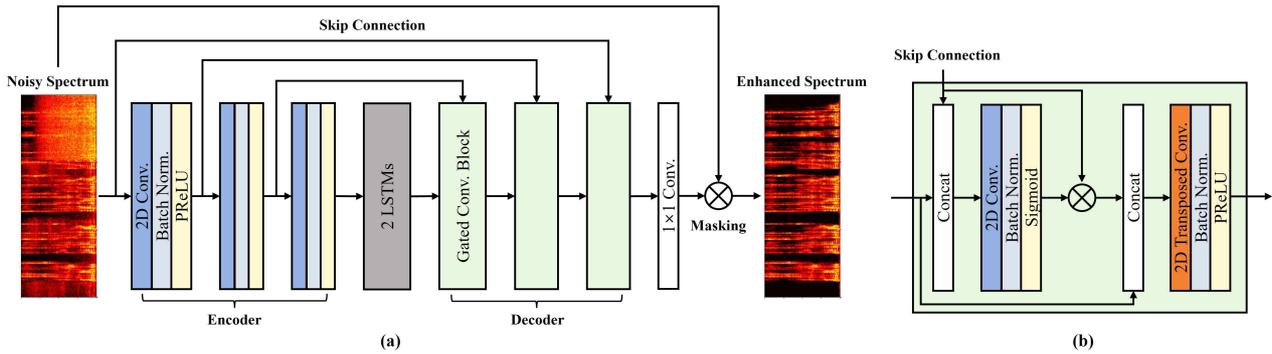


그림 1. 제안한 음성 향상 모델, (a) 인코더-디코더 구조, (b) 게이트 합성곱 블록

그림 1 (a)와 같이 제안한 모델은 인코더, LSTM, 디코더로 구성된다. 잡음이 포함된 음성은 32ms의 윈도우 길이와 8ms의 홉 길이를 갖는 STFT 스펙트럼으로 변환되며, 스펙트럼의 실수부와 허수부는 두 개의 다른 입력 채널로 취급되어 모델에 입력된다. 2차원 합성곱, 배치 정규화, 그리고 PReLU 층으로 구성된 인코더는 다운 샘플링을 통해 스펙트럼의 고차원 정보를 잠재공간에 캡처하고, LSTM은 시간 종속성을 모델링한다. 이후 2차원 전치 합성곱 층으로 구성된 디코더는 업 샘플링을 통해 인코딩 과정에서 손상된 저차원 정보를 복원하고, 스킵 커넥션을 통해 네트워크는 다양한 스케일의 차원 정보가 정교하게 재구성된 특징 맵을 출력하게 된다. 게이트 알고리즘은 네트워크를 구성하고 있는 인코더-디코더 간의 전체적인 정보 흐름을 제어하여 추출된 특징 간의 복잡한 상관관계를 보다 쉽게 모델링하도록 도와준다. 그림 1 (b)와 같이 밀집, 배치 정규화, 그리고 tanh 층으로 구성된 게이트 블록은 중복 특성을 억제하거나 글로벌 특징을 강조하는 것을 목표로 인코더에서 추출된 특징에 대한 가중치 함수를 학습하도록 설계하였으며, 이 가중치를 통해 네트워크는 이전 단계에서 추출된 핵심 정보들만 활용하게 된다. 이후 재구성된 특징 정보는 디코더의 마지막 블록에 연결된 (1, 1) 합성곱 층을 통해 잡음 성분을 억제하는 마스크로 맵핑되며, 잡음 스펙트럼과의 마스킹을 통해 깨끗한 음성 스펙트럼으로 출력된다.

데이터는 DNS challenge에서 공개된 음성 및 잡음 데이터를 기반으로 약 200시간의 분량의 데이터를 0~20dB의 SNR 비율로 제작하였으며, 학습, 검증 및 테스트 별로 약 8:1:1의 비율로 나누어 사용하였다. 사용한 인코더-디코더의 개수는 3개이며, 합성곱 층은 (1,7)의 필터 크기, (1,4)의 보폭, 그리고 (32, 64, 128)의 단계별 채널수를 가진다. 손실 함수와 최적화 방식으로는 SI-SNR 함수와 Adam을 사용하였으며, 제안한 모델의 객관적인 검증을 위해 음성 처리 분야에서 많이 사용되는 SDR, PESQ(perceptual evaluation of speech quality), 그리고 STOI(short-time objective intelligibility)를 평가 지표로 사용하였다.

3. 연구 결과

표 1은 실험 모델들의 성능 평가 결과를 나타낸다. 실험 결과, 제안한 모델은 기존 DCCRN 모델보다 30% 정도 작은 모델 크기로 평균 10% 이상의 향상된 성능을 보였다. 또한 게이트 알고리즘을 사용하지 않은 모델 대비 PESQ에서 0.35의 향상을 보였으며, 이는 제안한 게이트 알고리즘이 음성 향상에 관련된 핵심 정보를 효과적으로 필터링하여 모델의 추정 정확도를 크게 향상시킬 수 있음을 나타낸다.

표 1. 테스트 데이터에 대한 모델 별 실험 결과

Model	Size(M)	SNR	PESQ	STOI(%)
Noisy	-	8.39	1.61	87.0
DCCRN	3.7	16.38	2.48	93.4
Non-gate	2.2	17.31	2.83	95.1
Proposed	2.5	18.09	3.18	96.0

4. 결론

본 연구에서는 게이트 메커니즘이 도입된 합성곱 인코더-디코더 네트워크를 사용하여 깨끗한 음성의 크기 및 위상 스펙트럼을 추정하는 방법을 제안하였다. 실험 결과 제안한 음성 향상 모델은 품질, 모델 크기 및 복잡성 측면에서 기존 모델을 상회하는 성능을 보였으며, 이는 음성 향상 작업과 관련된 핵심 정보만 전달하는 게이트 접근 방식의 효율성을 보여준다.

References

- [1] Y. Hu, Y. Liu, S. Lv, M. Xing, S. Zhang, Y. Fu, J. Wu, B. Zhang, and L. Xie, "DCCRN: Deep Complex Convolution Recurrent Network for Phase-aware Speech Enhancement," arXiv Preprint arXiv: 2008.00264, 2020.
- [2] K. Tan and D. Wang, "Learning complex spectral mapping with gated convolutional recurrent networks for monaural speech enhancement", IEEE/ACM Trans. Audio Speech Lang. Process., vol. 28, pp. 380-390, 2019.

Pitch 정보를 활용한 Perturbation AUTOVC의 성능 향상*

박 화 영¹, 전 찬 준², 김 홍 국^{1,*}¹광주과학기술원 AI 대학원²조선대학교 컴퓨터공학과

Improved Perturbation AUTOVC with Pitch Information

Hwa-Young Park¹, Chanjun Chun², Hong Kook Kim^{1,*}¹Dept. AI Graduate School, Gwangju Insitute of Science and Technolgy²Dept. of Computer Engineering, Chosun Universityhwayoung_park@gm.gist.ac.kr¹, cjchun@chosun.ac.kr², hongkook@gist.ac.kr^{1,*}

제로 샷 음성 변환은 음성에 존재하는 언어 정보는 유지한 채 학습에 사용하지 않은 화자의 목소리로 변환하는 것을 말한다. 제로 샷 음성 변환을 위한 방법으로 오토인코더 모델을 활용한 언어와 화자 정보를 분리하는 방법을 주로 사용하지만 이러한 방법은 입력과 출력이 동일하기 때문에 인코더를 통과하고 나온 잠재 벡터에 화자 정보가 함께 인코딩될 수 있어 추론 시에 화자 유사도를 저하한다는 문제가 있다. 이를 해결하기 위한 방법으로 정보 교란 방법 및 오토인코더 모델 구조를 활용한 Perturbation AUTOVC라는 방법이 제안되었지만 해당 방법은 음성을 오직 언어 정보와 화자 정보 두 가지로 이원화하기 때문에 음성에 존재하는 다양한 음성 정보는 다룰 수 없다는 한계가 있다. 따라서 본 논문에서는 기존의 Perturbation AUTOVC의 방법은 고수하면서 pitch 정보를 추가로 활용하여 더 높은 품질의 음성 변환을 수행하도록 한다. 구체적으로, 정보 교란이 적용된 소스 음성은 언어 인코더를 통과하여 언어 정보만 담고 있는 잠재 벡터로 인코딩되며 교란이 적용되지 않은 음성은 x-vector, 에너지, pitch 정보를 추출하는 데 사용된다. x-vector와 pitch를 추출하기 위해 각각 사전 학습된 음성 인식 모델과 YAPPT 알고리즘을 사용하였으며, 추출된 정보들은 모두 디코더의 입력으로 사용하여 변환된 음성의 멜-스펙트로그램을 생성한다. 생성된 멜-스펙트로그램은 HiFi-GAN을 사용하여 음성 신호로 다시 복원하였다. 제안하는 방법은 CER, 코사인 유사도, F0-PCC로 평가하였으며, 기존의 Perturbation AUTOVC과 비교하여 CER은 85%의 정확도에서 91%로 6% 향상을, 코사인 유사도는 58%에서 60%로 2% 향상을 보였다.

* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 연구개발특구진흥재단의 ‘지역의 미래를 여는 과학기술 프로젝트’ 사업(인공지능 기반 메타버스 구현을 위한 융·복합 문화 가상 스튜디오, 2022-DD-UP-0312)과 2024년도 광주과학기술원의 재원으로 “과학기술혁신사업단의 실용화 연구개발 사업”의 지원을 받아 수행된 연구임.

한국어 종성 유음의 발음 변이 연구

서윤정, 드류 크로스비
고려대학교 국어국문학과

A Study on the Variation of the Korean liquid in Coda Position

YoonJeong Seo, Drew Crosby
Dept. of Korean language and literature, Korea University
phon_etic@naver.com, drewmcrosby89@gmail.com

본 연구에서는 젊은 세대가 한국어 종성 유음 /ㄹ/을 발음할 때 보이는 변이 현상을 보고하고 이 현상의 원인을 규명하고자 한다. 한국어에서 유음은 종성에서 실현될 때 설측음(lateral)으로 실현이 되는데, 최근 일부 화자들에게서 이를 권설음(retroflex)화하여 마치 영어의 /r/에 가깝게 발음하는 것이 관찰되고 있다. 따라서 본 연구에서는 20대 화자를 대상으로 낭독 실험을 실시하여 종성 유음의 실제적인 실현 양상을 검토하고, 권설음화를 유발하는 요인이 무엇인지 밝히고자 한다.

실험을 위해 종성에 유음이 포함된 2음절 명사 단어 60개를 선정하여 총 46개의 실험 문장을 만들었으며, 필러 문장을 포함하여 총 85개의 문장으로 대본을 구성하였다. 그리고 참가자들에게 대본을 최대한 자연스럽게, 말하듯이 읽도록 요청하여 분석 대상을 확보하였다. 20대 화자 총 34명이 실험에 참여하였으며, 여성 화자 19명, 남성 화자 15명으로 구성되었다. 또한, 방언 배경에 따른 차이가 있는지도 확인하기 위해 수도권 화자뿐만 아니라 경북 및 경남 화자도 모집하였는데, 그 결과 수도권 20명, 경북 10명, 경남 4명의 화자가 참여하였다.

<표 1> 연구 참가자 정보

	서울 /경기	대구	부산	전체
여성	10	7	2	19
남성	10	3	2	15
전체	20	10	4	34

분석 결과, 종성 유음의 권설음화는 선·후행 분절음에 따라서 그 정도를 달리하였다. 선행 모음이 /ㄱ, ㅋ, ㆁ, ㄷ/일 때 후행 유음의 권설음화가 다수 관찰되었으며, 후행 자음이 비치경음일 때 선행 유음의 권설음화가 자주 관찰되었다. 또한, 이러한 유음의 권설음화는 여성 화자에게서 더 빈번하게 관찰될 것이라 예상하였으나 남성 화자에게도 상당히 자주 관찰되었다. 화자의 방언 배경에 따른 차이는 현재 단계에서 단언하기는 어려우나, 경상도 지역 화자 집단에서 유음의 권설음화가 더욱 많이 나타나는 것으로 보인다. 본 연구는 종성 위치에서 한국어 유음의 음성적인 변이에 주목하여, 유음에 대한 이해를 넓히려는 시도를 수행하였다는 점에서 의의를 갖는다.

데이터를 활용한 프랑스어 단어 수준 음절 유형 및 자음군 분포 분석

박 소 은¹, 김 선 희²

¹서울대학교 외국어교육과, ²서울대학교 불어교육과

Data-driven analysis on French word-level syllable types and consonant cluster distribution

Soeun Park¹, Sunhee Kim²

¹Dept. of Foreign Language Education, Seoul National University,

²Dept. of French Language Education, Seoul National University
soeun1999@snu.ac.kr, sunhkim@snu.ac.kr.com

모음으로 이루어진 음절핵과 음절핵에 선, 후행하는 자음(군)으로 구성된 음절의 구조와 유형을 밝히는 것은 언어마다 음절 구조와 제약 조건이 상이하기 때문에 꼭 필요한 작업이다(Chomsky & Halle, 1968, van der Hulst, 2012). 또한 음절은 음소배열의 분포제약을 설명하기에 가장 적합한 단위이기도 하다(Delattre, 1940).

본 연구는 데이터를 활용하여 단어 수준에서 프랑스어 음절 유형과 분포 및 음소배열적 특징을 고찰하는 것을 목적으로 하여, 다음 세 가지 연구문제를 중심으로 살펴보았다. 첫째, 단어의 음절 수에 따라 단어 내 음절 구조 및 분포는 데이터에서 어떻게 나타나는가? 둘째, 데이터를 통해 검증한 단어 내 위치에 따라 나타날 수 있는 자음군의 유형으로는 무엇이 있는가? 셋째, 데이터가 제공하는 음절분리 규칙과 데이터 관찰을 통해 추론할 수 있는 음절분리 규칙은 일치하는가?

이를 위하여 대규모 오픈소스 프랑스어 단어 데이터인 lexique(Pallier et al., 2001)를 활용하였다. 고빈출 5,000개 단어를 연구 대상으로 하여 음절 수에 따른 서브코퍼스 6개를 제작한 뒤 각각의 코퍼스의 음절 분포를 조사하였다. 이후 어두, 어중, 어말의 3개 위치에서 관찰되는 자음군을 조사하였다.

연구 결과, 단일 음절 내에서 두음에 3개 자음연쇄, 말음에는 4개의 자음연쇄까지 나타났다. 또한 단어의 음절 수가 많아질수록 개음절 유형의 비율이 증가하여 약 80%까지 차지하는 반면 단음절 단어에서 유일하게 개음절과 폐음절 유형 음절이 각 약 50%로 비슷하게 나타났다. 자음군 제약과 관련해서는 어두에 허용 가능한 자음군의 집합이 어말에 허용 가능한 자음군 집합에 포함되는 포함관계가 성립되었다. 한편 어중 자음군 집합에는 음절분리가 되지 않은 자음연쇄 목록이 포함되기 때문에 가장 많은 자음군 유형이 관찰되었으나 어두(두음), 어말(말음) 환경에서 나올 수 없는 자음군은 두음과 말음이 합쳐진 어중 자음군으로도 나올 수 없으므로 어중 환경에서만 나오는 자음군은 선행 음절의 말음과 후행 음절의 두음으로 음절 분리가 일어나야 하는 것을 알 수 있었다.

References

- Chomsky, N., Halle, M. (1968). *The Sound Pattern of English*. Harper Row, New York.
- Van der Hulst, H., & Ritter, N. (Eds.). (2012). *The syllable: views and facts* (Vol. 45). Walter de Gruyter. ISO 690
- Delattre, P. (1940). Tendances de coupe syllabique en français. *PMLA*, 55(2), 579-595.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUETM//A lexical database for contemporary french: LEXIQUETM. *L'année psychologique*, 101(3), 447-462.

한국인 학습자의 프랑스어 연음 오류 분석

김민정¹, 김선희²

¹서울대학교 외국어교육과, ²서울대학교 불어교육과

An Analysis of French Liaison Errors Produced by Korean Learners

Minjeong Kim¹, Sunhee Kim²

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

프랑스어의 대표적인 음운현상인 liaison(이하 연음)은 연성(sandhi externe: 連聲)의 한 유형으로, 독립된 단어로 사용될 시에는 발음되지 않던 종결 자음이 일정한 환경에서 후행하는 단어의 첫 모음 및 무음 h와 연결되어 음성적으로 실현되는 음운현상을 의미한다(Grevisse, 1980).

프랑스어 연음은 필수적 연음, 금지된 연음, 수의적 연음의 세 유형으로 분류되며, Delattre(1947)는 각각의 연음 유형에 대한 어휘통사적 분류와 예시를 제안하였다. 연음은 [z, t, n, p, g, k, ʁ]의 한정된 자음으로 실현되며(Mallet, 2008), 연음의 실현은 재음절화 현상을 수반하게 되는데, 이로 인해 음성적 경계와 어휘 경계가 어긋나게 된다(Stridfeldt, 2005). 이러한 연음의 특성은 프랑스어 학습시, 구어 프랑스어의 생성과 이해를 어렵게 하는 주된 원인이다(Racine & Detey, 2015; 김선희, 2020). 그럼에도 불구하고 외국인 학습자 대상의 프랑스어 연음 실현 연구가 진행되었지만(Thomas, 2002; Howard, 2005; Tennant, 2015등) 한국인 학습자를 대상으로 하는 연음 실현 연구는 부족한 실정이다.

본 연구는 한국인 프랑스어 학습자의 프랑스어 낭독 발화 데이터를 사용하여 한국인 학습자의 프랑스어 연음 오류를 분석하고자 하였다. Abry & Chalaron(2011)등에서 연음 환경이 포함된 프랑스어 문장 221개를 추출하였으며, 이는 총 330개의 연음 환경으로 구성되었다. 연음 오류의 유형은 “연음 자음 대체”와 “금지된 연음의 실현”의 두 가지로 분류하였다(Tennant, 2015; Racine, 2015). 실험 참여자로는 한국인 프랑스어 학습자 20명이 참여하였다. 총 20명의 실험 참여자는 각자 330개의 연음 환경을 낭독하였고, 수집된 6,600개의 음성에 대한 연음 오류 분석이 진행되었다.

실험 결과, 한국인 학습자의 연음 오류 실현율은 4.9%로 나타났다. 총 158건의 오류 중, 142건이 “연음 자음 대체” 유형으로 관측되며 전체 오류의 89.9%를 차지하였다. 해당 오류는 “grand ami”와 같이 연음 자음이 [t]인 환경에서 가장 빈번히 발생했으며(64.8%), [t]환경에서는 [d]로의 대체가 가장 높은 비율을 차지하였다(94.6%). “금지된 연음의 실현” 오류는 총 16건 발생하였으며, 그중 14건이 “Mes amis ont des idées”와 같은 “일반명사+동사” 환경에서 관측되었다. 이와 같은 결과를 통해 한국인 학습자가 연음 자음 분류와 연음 유형에 따른 실현 규칙을 정확히 이해하지 못하고 있다는 사실을 유추해 볼 수 있다.

본 연구는 한국인 학습자의 발화 데이터를 사용하여 학습자의 연음 실현 오류를 유형별로 분석하였다는 점에서 그 의의가 있다. 연구 결과는 한국인 학습자 대상의 연음 교수법 제안에 참고자료로 활용될 수 있을 것으로 기대된다. 향후 연구에서는 본 연구 결과를 반영한 연음 교수법을 제안하고, 이를 교육 현장에 적용함으로써 그 효과를 검증할 예정이다.

References

- Abry, D., & Chalaron, M. L. (Eds.). (2011). *Les 500 exercices de phonétique: niveau B1/B2*. Paris, France: Hachette.
- Delattre, P. (1947). La liaison en français, tendances et classification. *French review*, 148-157.
- Grevisse, M. (1980). *Le bon usage* (12th ed.). Paris, France: Duculot
- Howard, M. (2005). L'acquisition de la liaison en français langue seconde. Une analyse quantitative d'apprenants avancés en milieu guidé et en milieu naturel. *Corela. Cognition, représentation, langage*, (HS-1).
- Mallet, G. (2008). *La liaison en français: descriptions et analyses dans le corpus PFC*. Unpublished PhD dissertation. Université Paris Ouest, France.
- Racine, I. (2015). La liaison chez les apprenants hispanophones avancés de FLE. *Bulletin VALS-ASLA*, 102, 147-167.
- Racine, I., & Detey, S. (2015). Corpus oraux, liaison et locuteurs non natifs: de la recherche en phonologie à l'enseignement du français langue étrangère. *Bulletin vals-asla*, 102, 1-25.
- Stridfeldt, M. (2005). *La perception du français oral par des apprenants suédois* (Doctoral dissertation, Moderna språk).
- Thomas, A. (2002). La variation phonétique en français langue seconde au niveau universitaire avancé. *Acquisition et interaction en langue étrangère*, (17), 101-121.
- Tranel, B. (1976). A note of final consonant deletion in modern French. *Lingua*, 39(1-2), 53-68.
- Tennant, J. (2015). Canadian anglophone learners' realization of French liaison. *Bulletin vals-asla*, 102, 65-85.
- 김선희. (2020). 최소대립쌍을 이용한 한국인 학습자의 프랑스어 모음 지각 양상 연구. *프랑스어문교육*, 69, 7-29.

시대별 애니메이션 캐릭터 음성에 대한 음향 및 선호도 분석

이 선 미, 조 회 수, 최 영 경, 김 지 연, 이 룬 하, 이 영 미
이화여자대학교 언어병리학과

Analysis of Sound and Preference for Animation Character Voices by Era

Sunmi Lee, Heesu Cho, Youngkyoung Choi, Jiyeon Kim, Ryunha Lee, Youngmee Lee
Dept. of Communication Disorders, Ewha Womans University
sun_mi0704@ewhain.net, gmltn320@ewhain.net, cyk1605@ewhain.net, jiykim103@ewhain.net,
cheshire7@ewhain.net, youngmee@ewha.ac.kr

본 연구는 1980-1990년대, 2000년대, 2010년대 이후 애니메이션 캐릭터의 음성 특성을 Praat(ver.6.3.18)을 사용하여 분석하고, 선호도를 조사했다. 주파수와 말속도의 음향학적 요소에서 시대별 차이가 없었으나, 애니메이션 방영 시대에 따라 선호도는 유의하게 다르게 나타났다. 특히, 1980-1990년대 캐릭터에 대한 선호도가 최근 애니메이션보다 낮았다. 이는 애니메이션 캐릭터는 공통적인 음향학적 특성을 가지고 있지만, 시대의 사회문화적 경험에 따른 친숙도가 청자의 선호도에 영향을 미쳤음을 시사한다.

정상발달 아동의 감정발화모방에 대한 음향학적 분석

김 수 빈, 김 정 은, 조 수 형, 이 효 선, 문 성 윤, 이 영 미
이화여자대학교 언어병리학과

Acoustic Analyses in the Imitation of Emotional Speech in Children with Typical Development

Su Been Kim, Jung Eun Kim, Soo Hyung Cho, Hyo Sun Lee, Seong Yun Moon,
Youngmee Lee

Dept. of Communication Disorders, Ewha Womans University
slpsubeen@ewhain.net, 232shg19@ewhain.net, soohy0603@ewhain.net,
hhhyosuny@ewhain.net, fineo_o@ewhain.net, youngmee@ewha.ac.kr

The purpose of this study was to investigate the acoustic characteristics of emotional speech in children with typical development. Twenty-two preschool-aged children (4-5.9 years old) and school-aged children (6-9.9 years old) participated in this study. The children were asked to imitate 15 utterances, which were based on emotional speech presenting three emotional expressions (happiness, sadness, and anger). The fundamental frequency, intensity, and duration measures were obtained from children's utterances. School-aged children tended to produce emotional speech longer compared to preschoolers. Both preschool-aged and school-aged children imitated emotional speech differentially according to emotion type (happiness, sadness, and anger) in terms of fundamental frequency, intensity, and duration. These findings suggested that preschool-aged and school-aged children can use their voices to express emotions by controlling pitch, intensity, and duration parameters. In addition, school-aged children tended to adjust their duration parameters to express various emotions, compared to preschool-aged children. The duration difference in children with typical development may be affected by children's maturity in speech and language development.

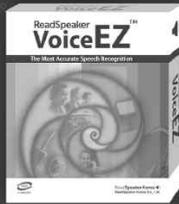
음성기술 전문회사 (주)리드스피커코리아

25년 이상의 음성기술을 향한 노하우와 기술력의 신뢰를 바탕으로
국내는 물론 해외 다양한 분야에서 인정받고 있습니다.



음성합성 ReadSpeaker™

인공지능(AI) 기술을 적용하여 음질은 한층 더 높이고
음성합성기의 개발기간은 단축시킨 음성합성기,
Standard TTS 49개 언어, 133개 음색 보유



음성인식 ReadSpeaker VoiceEz™

단어 인식부터 비정형 자연어 음성인식 강점을
모두 갖춘 최고의 음성인식 솔루션
AI 고객센터, 음성인식자막, 음성인식회의록

(주)리드스피커코리아

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

음성기술

인공지능 전문 기업 셀바스 AI가 이끌어가겠습니다.



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

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

Selvy TTS Selvy deepTTS Selvy deepTTS On-Device
 Selvy STT Selvy Note Selvy MediVoice odiro

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



◎ 홈페이지

<http://www.kslp.org>

◎ 전화

(02)790-2726

◎ 이메일

kasla999@hanmail.net

◎ 네이버 카페

<https://naver.me/IMnyh8E8>

◎ 오픈 채팅방

<https://open.kakao.com/o/geSJVeHf>

◎ 유튜브 채널

https://www.youtube.com/@kslp_

ISSN 2005-9299

Proceedings
of the 2024 Spring Conference
of the Korean Society of Speech Sciences

May 18, 2024
Hannam University
the Korean Society of Speech Sciences