

## Effects of Speech Rate, Message Type, and Listener age in TTS on Voice Attractiveness

Nagyeong Ha<sup>1</sup>, Yuna Jin<sup>1</sup>, Yoonji Park<sup>1</sup>, Eunjin Sim<sup>1</sup>, Youngmee Lee<sup>2\*</sup>

<sup>1</sup> Dept. of Communication Disorders, Graduate School, Ewha Womans University, Master's Student

<sup>2</sup> Dept. of Communication Disorders, Graduate School, Ewha Womans University, Professor

**Purpose:** This study investigated the effects of TTS (text-to-speech) speech rate, message type, and listener age on the perception of voice attractiveness.

**Methods:** Two TTS-generated female voices (speakers in their 20~30s) were manipulated to produce speech at three different rates (slow, moderate, fast) and deliver two distinctive message types (emotional, rational). A total of 40 participants, comprising younger adults ( $n=20$ ) and older adults ( $n=20$ ), evaluated the attractiveness of these voices using a 5-point Likert scale.

**Results:** TTS speech rate had a statistically significant effect on voice attractiveness, with slower or faster rates showing higher likability than moderate speeds. Emotional messages were rated more favorably than rational ones, and older adults provided higher overall ratings compared to younger adults. In terms of interaction effects, emotional messages were rated higher at slower rates, while rational messages were rated higher at faster rates.

**Conclusions:** TTS speech rate was found to be a critical factor in voice attractiveness. Differences emerged between slow and normal speeds, as well as between normal and fast speeds. Emotional messages were most appealing at a slow rate, whereas rational messages were most favorably received at a fast rate. Older adults, overall, provided higher ratings than younger adults. These findings suggest that tailoring TTS speech rate and message type to specific contexts can enhance user experience, offering practical insights for AAC systems and related applications.

**Correspondence:** Youngmee Lee, PhD

**E-mail:** youngmee@ewha.ac.kr

**Received:** February 12, 2025

**Revision revised:** April 02, 2025

**Accepted:** April 30, 2025

**ORCID**

Nagyeong Ha

<https://orcid.org/0009-0009-5502-1902>

Yuna Jin

<https://orcid.org/0009-0005-2205-7336>

Yoonji Park

<https://orcid.org/0009-0003-6976-8480>

Eunjin Sim

<https://orcid.org/0009-0007-7058-9294>

Youngmee Lee

<https://orcid.org/0000-0003-1809-5944>

**Keywords:** TTS, speech rate, message type, listener age, voice attractiveness

### 1. 서론

현대 사회에서 TTS(text-to-speech) 기술은 일상생활의 다양한 분야에서 점점 널리 활용되고 있다. TTS는 입력된 텍스트를 음성으로 변환하여 출력하는 음성 합성 기술로, 주로 음성신호(speech signal)를 자동 생성하는 시스템으로 정의된다(Choi et al., 2018; Jeong, 1994). 초기 TTS 기술은 구어 의사소통에 어려움을 겪는 사람들을 위한 보조 수단으로 개발되었으나, 현재는 뉴스, 날씨 정보 제공, 내비게이션 안내, 전자책 낭독 등 다양한 용도로 활용되고 있다(Yoo, 2018). 특히, ARS 전화 음성 서비스, AI 스피커, 음성 번역기와 같은 상호작용 기술의 핵심 요소로 자리 잡고 있다(Voiceware, n.d.). TTS 기술은 모든 연령대에서 친숙한 기술로 인식되고 있으며, 특히 노년층에서도 그 중요성이 강조되고 있다. Jeon 등(2020)의 연구에 따르면, 노년층이 AI 스피커를 사용하는 비율이 꾸준히 증가하고 있으며, 이를 통해 날씨 확인, 음악 감상, 뉴스 청취 등 다양한 활동을 수행하고 있다(Song et al., 2021). 보 건복지부와 SK하이닉스가 독거노인에게 AI 스피커를 지원한 사례는 이러한 기술이 사회적 돌봄 서비스의 핵심 도구로 자리 잡고 있음을 시사한다(Park, 2020). 따라서 TTS 기술은 단순한 정보 전달을 넘어 사용자 경험(user experience: UX)을 개선하고 정서적 유대감을 형성하는 역할로 발전하고 있다.

이처럼 TTS 기술은 다양한 분야에서 활용되며 사용자와 상호작용하는 중요한 도구로 자리 잡고 있다. 그러나 TTS 기술의 성공적인 활용을 위해서는 단순히 음성을 생성하는 기술적인 완성도를 넘어, TTS 음성이 사용자에게 친근하고 매력적으로 들릴 수 있도록 설계하는 것은 사용자 경험을 향상시키는 핵심 요인으로 주목받고 있다(Jang & Lee, 2019). 음성의 자연스러움, 말속도, 억양, 그리고 메시지의 전달 방식은 모두 사용자와의 정서적 유대감을 형성하고, TTS 기술에 대한 신뢰와 만족도를 높이는 데 기여할 수 있다(Moore, 2012). 이러한 맥락에서, 청자가 TTS 음성을 얼마나 호감 있게 평가하는지에 대한 연구는 TTS 기술의 상호작용 디자인(interaction design)과 실제 활용 가능성을 높이는 데 필수적이라 할 수 있다.

청자의 음성 호감도에 영향을 미치는 주요 요소로는 억양, 말속도, 음의 고저, 쉼, 음조, 목소리의 강도 등이 있다(Kwon, 2015). 특히 말속도는 음성 호감도의 중요한 요소로 주목받고 있지만, 기존 연구 결과는 상반된 결과를 보인다. 예를 들어, Chun 등(2024)은 보통 속도의 발화에서 가장 높은 호감도를 보였으나, Song 등(2022)은 느린 말속도와 풍부한 억양이 높은 호감도를 이끌어낸다고 보고하였다. 국외 연구에서는 빠른 말속도가 화자를 신뢰롭고 능력 있는 사람으로 인식하게 한다는 결과가 보고되었다(Smith et al., 1975; Street et al., 1984). 이러한 연구들은 말속도가 청자에게 미치는 영향이 상황과 맥락에 따라 달라질 수 있음을 시사한다. 또한, 청자의 연령과 성별에 따라 TTS 음성에 대한 평가가 달라질 수 있다. 예를 들어, Wingfield와 Ducharme (1999)의 연구에서는 노년층이 청년층보다 느린 말속도를 선호하는 경향을 보였으며, Chun 등(2024)은 장년층 남성만이 빠른 말속도에서 유의미하게 낮은 호감도를 보인다고 보고하였다. 여성 청자는 남성 청자에 비해 더 높은 호감도를 보였으며, 남성은 느린 말속도를, 여성은 빠른 말속도를 더 선호하는 경향이 나타났다(Kwon, 2016; Song et al., 2022).

말속도와 메시지 유형 간의 상호작용에 대한 연구도 중요한 시사점을 제공한다. 청자들은 콘텐츠의 목적에 따라 선호하는 재생 속도가 다르다(Kim, 2015). 예를 들어, 학생들은 정보전달이 목적인 인터넷 강의를 배속으로 시청하는 경향을 보인다. 이는 정보전달 상황에서 빠른 말속도가 선호되며 학습 효율 향상에도 도움이 될 수 있음을 시사한다(Murphy et al., 2022). 정보 전달 상황에서는 빠른 말속도가 더 높은 호감도를 유도하지만(Miller et al., 1976), 감성적 메시지의 경우 보통 또는 느린 속도가 더 효과적이었다(Lee, 2020). 특히 메시지의 특성이 청자의 정서적 반응을 유도하는 데 중요한 역할을 한다. Kwon과 Na(2011)는 정보 전달 상황에서 이성적 메시지가 감성적 메시지보다 더 호의적인 반응을 유도할 수 있다고 보고하였다. 따라서 이성적 TTS 음성의 말속도, 메시지 유형, 청자의 연령대는 각각 음성 호감도에 영향을 미치는 독립적인 요소로 밝혀져 있지만, 이 세 가지 변수를 종합적으로 고려하는 연구는 매우 제한적이다. 이 세 변수를 통합적으로 분석하는 것은 단일 요소 연구만으로는 확인할 수 없는 상호작용 효과를 이해하는 데 필수적이다. 예를 들어, 말속도가 청자의 연령대에 따라 다르게 인식되거나, 특정 메시지 유형이 특정 속도에서 더 효과적으로 전달될 가능성이 있다(Kwon, 2015; Street et al., 1984). 또한, 노년층과 청년층이 TTS 음성을 평가하는 기준이 다를 수 있으며, 이러한 차이는 사용자 경험 설계에 중요한 시사점을 제공할 수 있다(Walker et al., 2004). TTS 음성은 다양한 맥락에서 사용되기 때문에, 특정 변수가 단독으로 작용하는 경우보다 여러 변수가 상호작용할 때 발생하는 복합적인 사용자 경험을 이해하는 것이 중요하다. 예컨대, 내비게이션과 같은 정보 전달 목적의 TTS에서는 빠른 말속도가 선호될 수 있지만, 감성적 메시지를 전달하는 AI 스피커의 경우 느린 속도가 더 효과적일 수 있다(Lee, 2020; Miller et al., 1976). 따라서 말속도, 메시지 유형, 청자의 연령대는 독립적으로 음성 호감도에

영향을 미치는 요소로 밝혀졌으나, 이 세 변수를 통합적으로 분석하는 연구는 제한적이다. 세 변수를 종합적으로 고려하면, 특정 말속도가 특정 메시지 유형에서 더 효과적이거나, 청자의 연령에 따라 선호가 달라질 수 있는 복합적인 상호작용 효과를 이해할 수 있다(Kwon, 2015; Walker et al., 2004). 이에 따라, 본 연구에서는 TTS 음성의 말속도(느림, 중간, 빠름), 메시지 유형(이성적 메시지, 감성적 메시지), 청자의 연령대(청년, 노년)에 따른 음성 호감도를 비교 분석하고자 하였다. 이를 통해 TTS 음성이 사용되는 다양한 환경에서 사용자의 경험을 개선하고 실질적 활용 가능성을 제고하기 위한 정보를 제공하고자 한다.

## II. 연구 방법

### 1. 연구 대상

본 연구에 참여한 대상자는 만 20~39세의 청년층 여성 20명( $M=26.25$ ,  $SD=4.47$ )과 만 60세 이상의 노년층 여성 20명( $M=63.20$ ,  $SD=3.10$ )으로 구성된 청자 집단이다(Table 1). 대상자의 청력 및 정신 상태가 연구 결과에 미칠 가능성을 최소화하기 위해, 본 연구의 참여자는 (1)한국 노인성난청의청각장애지수-선별(The Korean Hearing Handicap Inventory for Elderly-Screening: KHHIE-S, Hong et al., 2004)에서 8점 이하이고, (2)축약형 노인성 우울증 척도(Short form of Geriatric Depression Scale-Korean Version: SGDS, Jo et al., 1999)에서 8점 미만인 자로 선정하였다.

Table 1. Participants' information

	Young adults ( $n=20$ )	Older adults ( $n=20$ )
Mean ( $SD$ )	26.25 (4.47)	63.20 (3.10)

Note. Values are presented as mean ( $SD$ ).

### 2. 연구 도구

#### 1) 음성 데이터

본 연구에서 사용된 음성 데이터는 네이버 클로바 더빙의 TTS 기술을 활용하여 제작된 합성 음성으로, 청년 여성 화자의 목소리를 기반으로 하였다. 이는 Ko(2019)에 근거하여 선정된 것으로, 해당 연구에서는 AI 여성 음성이 청자들에게 남성 음성보다 높은 호감도를 유발한다는 점이 보고되었다. AI 음성의 성별 선호도를 반영하여, 본 연구에서는 여성 음성을 채택하여 실험을 설계하였다. 음성 선정 과정은 '자연스러움'과 '명료도'를 주요 평가 기준으로 설정하여 진행되었다. 네이버 클로바 더빙의 다양한 여성 음성을 대상으로 4명의 연구자가 '자연스러움'과 '명료도'를 주요 평가 기준으로 하여 5점 척도로 평가를 실시하였으며, 음성 선정의 기준은 다음과 같다: (1)각 평가 항목에서 모든 평가자가 4점 이상을 부여한 음성을 후보군으로 선정하고, (2)총점이 가장 높은 상위 4개의 음성을 추출하였다. 이후 Praat 6.4.25(Boersma &

Weenink, 2023) 프로그램을 사용하여 말속도가 가장 유사한 두 개의 음성을 선정하였으며, 이 과정을 통해 '지윤'과 '나래'라는 이름의 음성이 최종 선정되었다. 두 음성의 메시지 유형별 말속도와 평균 주파수는 Table 2에 제시되어 있다.

**Table 2.** Average frequency of Jiyeon and Narae TTS voices by message type and speaking rates

Message type	Jiyeon		Narae	
	Speech rate	Frequency	Speech rate	Frequency
Emotional message	Original	235.07Hz	Original	237.21Hz
	Slower	235.71Hz	Slower	237.03Hz
	Normal	235.78Hz	Normal	237.41Hz
	Faster	236.31Hz	Faster	237.39Hz
Rational message	Original	243.69Hz	Original	246.89Hz
	Slower	244.64Hz	Slower	247.09Hz
	Normal	244.63Hz	Normal	247.89Hz
	Faster	245.43Hz	Faster	248.11Hz

## 2) 청자 반응 자료 제작

본 연구에서 청자에게 제시된 음성 파일의 수는 총 12개로, 이는 2개의 TTS 음성 화자, 3개의 말속도(느림, 중간, 빠름), 2개의 메시지 유형(이성적, 감성적)의 조합으로 구성되었다. 학습 효과와 순서 효과를 배제하기 위해 음성 파일은 무선휘(randomization) 과정을 통해 A와 B 두 개의 세트로 나누어졌다. 연구 참여자는 총 40명으로, 청년층(만 20~39세)과 노년층(만 60세 이상) 각 20명으로 구성되었으며, 각 연령층은 다시 10명씩 두 그룹으로 나누었다. 각 그룹은 세트 A 또는 세트 B 중 하나를 청취하고, 음성 호감도를 측정하였다.

## 3) TTS 음성 말속도 작업

최종 선정된 TTS 음성은 말속도에 따른 실험 조건을 충족하기 위해 Praat 6.4.25(Boersma & Weenink, 2023) 프로그램을 사용하여 각 속도에 맞는 배속 비율을 적용하였다. 말속도는 느림, 중간, 빠름의 세 가지로 조정되었으며, 중간 말속도는 Shin과 Lee(2017)의 연구를 참고하여 20대 여성의 평균 말속도인 5.42SPS(syllable per second,  $SD=0.46$ )를 기준으로 설정하였다. 느린 속도와 빠른 속도는 Jeong(2017)의 연구에서 제시된 '평균  $\pm 2SD$ '의 기준을 적용하여, 각각 4.50SPS와 6.34SPS로 설정하였다. 선정된 TTS 음성은 말속도만을 조정하여 빠름, 중간, 느린 속도로 설정하였고, 말속도 외의 변수(예, 음도, 강도 등)는 조정하지 않고 일정하게 유지하였다.

## 4) 메시지 유형별 문장 선정

실험 문장은 Kwon과 Na(2011)의 연구에서 제시한 두 가지 메시지 유형인 이성적 메시지와 감성적 메시지를 기반으로 구성하였

다. 이성적 메시지는 Song(2023)의 연구를 참고하여 공공데이터포털(Public Data Portal, 2024)의 '대전도시철도 1호선 열차 안내방송'을 활용하였으며, 감성적 메시지는 Lee(2020)가 제안한 감동적인 발화와 핵심 메시지 예시를 바탕으로 선정되었다. 두 메시지 유형은 모두 51음절로 통일되도록 수정되었다. 최종적으로 선정된 문구는 다음과 같다. 이성적 메시지는 "에스컬레이터를 이용하실 때는 사고 예방을 위해 노란선 안쪽에 서 주시고, 손잡이를 꼭 잡아 안전하게 이용해 주시기 바랍니다."이고, 감성적 메시지는 "여러분도 언젠가 여유가 생기면 여기까지 힘들게 걸어온 당신의 모습을 따뜻하게 돌아보고 스스로를 인정해주시길 바랍니다."이다.

## 5) 음성 호감도 측정을 위한 설문 문항 제작

본 연구에서는 청자가 제시된 TTS 음성을 듣고 음성 호감도를 평가하기 위해 Paek과 Jung(2022)의 설문 항목을 참고하여 보완된 설문지를 사용하였다. TTS 음성의 특성을 반영하기 위해, 각 문항에 TTS라는 용어를 추가하여 연구 목적에 적합한 설문지를 구성하였다. 설문지는 Paek과 Jung(2022)의 AI 음성 호감도에 대한 선행 연구를 기반으로 수정되었으며, 매력도(attractiveness), 신뢰도(trustworthiness), 의인화(personification)라는 세 가지 주요 항목으로 구성되었다. 각 주요 항목은 5개의 하위 문항으로 통일하였다. TTS 음성에 대한 청자의 감성적 반응(affectivereaction)을 확인하기 위해 매력도 항목을 선정하였다. 음성에 대한 인지된 반응을 확인하기 위해 신뢰도 항목을 선정하였다. TTS 음성은 실제 사람의 음성과 다르기 때문에, AI와 관련된 선행 연구에서 사용하는 '인간 유사성'과 '의인화'라는 측정 기준을 활용하여 TTS 음성에 대한 호감도를 평가하고자 의인화 항목을 선정하였다. 각 설문 문항의 구체적인 내용은 Appendix 1에 제시되어 있다. 실험 중, 참여자는 12개의 TTS 음성을 청취한 후 제시된 설문 문항을 읽고 주관적으로 평가하였다. 평가 방식은 5점 척도를 사용하였으며, 1점은 '전혀 그렇지 않다', 5점은 '매우 그렇다'를 의미하도록 지시하였다. 참여자는 각 문항에 대해 자신이 느끼는 정도를 점수로 표현하였으며, 이를 통해 TTS 음성에 대한 청자의 주관적인 평가 데이터를 수집하였다.

## 6) 연구 절차

본 연구는 조용한 공간에서 일대일 대면 방식으로 검사자와 연구 참여자가 직접 만나 진행되었다. 실험은 구글 설문지 양식(google forms)을 활용하여 진행되었으며, 실험 환경과 절차는 다음과 같다. 연구 참여자는 실험 참여 전에 선별검사로 KHHIE-S와 SGDS를 실시하였다. 이후, 참여자는 제공된 노트북(NT950XGQ, Samsung, Suwon, Korea)과 헤드폰(IT-HDS02W, KC Korea, Goyang, Korea) 기종을 사용하여 구글 설문지를 확인하며 실험에 참여하였다. 검사자는 실험 시작 전에 연구 절차와 주의사항을 직접 설명하고 참여자가 이를 충분히 숙지할 수 있도록 하였다. 본 실험에 앞서 참여자는 연습 음성 청취를 통해, 적절한 음량을 스스로 조절하였다. 음량 설정이 완료된 후, 연구 참여자는 본 실험 음성을 하나씩 청취하며, 각 음성의 내용과 목소리에 집중하여 평가하였다. 음성

파일은 반복 청취가 가능하였으나, 이전 문항으로 돌아가 응답을 수정할 수는 없도록 설계되었다.

### 3. 자료 분석

본 연구에서 수집한 호감도 설문 자료는 모두 IBM SPSS Statistics version 29(IBM, Armonk, NY, USA) 프로그램을 사용하여 분석하였다. TTS 음성의 말속도(느림, 보통, 빠름), 메시지 유형(이성적 메시지, 감성적 메시지), 청자의 연령대(청년층, 노년층)에 따른 음성 호감도에 유의한 차이가 있는지를 확인하기 위하여 삼원혼합분산분석(three-way mixed ANOVA)을 실시하였다. 이때, 구형성 검정을 통해 구형성 가정을 충족하지 않을 경우, Greenhouse-Geisser로 수정된 자유도와 *F*값을 사용하여 결과를 해석하였다.

### III. 연구 결과

TTS 음성의 말속도(느림, 보통, 빠름), 메시지 유형(이성적, 감성적), 청자의 연령대(청년, 노년)에 따른 음성 호감도에 대한 기술통계 결과는 Table 3에 제시하였다.

**Table 3.** Descriptive results of voice attractiveness ratings by TTS speaking rate, message type, and listener age group

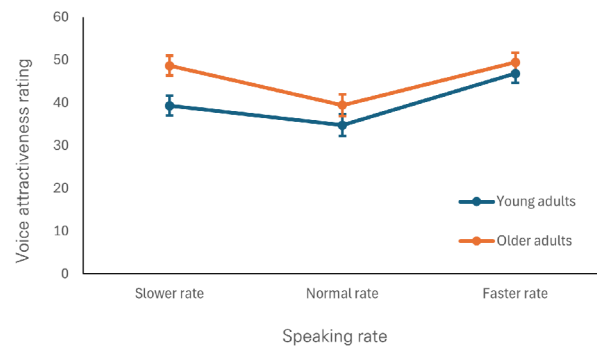
TTS speaking rate	Young adults ( <i>n</i> =20)		Older adults ( <i>n</i> =20)	
	Rational	Emotional	Rational	Emotional
Slower rate	39.275 (11.05)	44.325 ( 8.66)	78.650 ( 9.53)	48.675 (11.24)
Normal rate	34.725 (11.43)	41.050 (11.29)	39.400 ( 9.95)	49.325 ( 8.56)
Faster rate	46.800 (11.12)	43.675 ( 9.98)	49.400 (11.50)	46.100 (10.79)

Note. Values are presented as mean (*SD*).

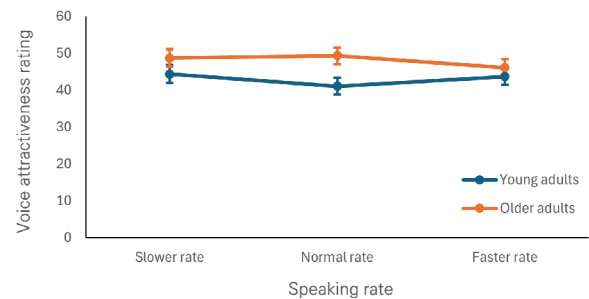
삼원혼합분산분석 결과, TTS 말속도에 대한 주효과가 통계적으로 유의하였다( $F_{(2,76)}=12.164, p<.001$ ). 이에 대한 Bonferroni 사후검정을 실시한 결과, TTS 말속도가 느림과 보통( $p<.01$ ), 보통과 빠름( $p<.001$ )에서 유의한 차이가 있었으나, 느림과 빠름 간에는 유의한 차이가 없었다(Figure 1). 즉, TTS의 말속도가 보통일 때보다 느리거나 빠를 때, 청자의 음성에 대한 호감도가 유의하게 높은 것으로 나타났다. 메시지 유형에 대한 주효과도 통계적으로 유의한 것으로 나타났다( $F_{(1,38)}=8.130, p<.01$ ). 청자는 이성적 메시지보다 감성적 메시지를 청취할 때 TTS 음성에 대해 유의하게 더 높은 호감도를 보였다. 또한, 집단에 대한 주효과도 유의하게 나타나( $F_{(1,38)}=6.082, p<.05$ ), 노년층이 청년층에 비해 TTS 음성에 대해 더 높은 호감도를 보였다.

본 연구에서 TTS 말속도와 청자의 연령, 메시지 유형과 청자의

연령 간의 이차 상호작용 효과는 통계적으로 유의하지 않았다. 그러나 TTS 말속도와 메시지 유형 간의 이차 상호작용 효과는 유의하게 나타났다( $F_{(1,685, 64.025)}=5.497, p<.01$ , Figure 2). 이를 검증하기 위해, 메시지 유형별로 TTS 말속도에 대한 일원배치분산분석을 실시하였다. 분석 결과, 느린 속도와 빠른 말속도에서는 이성적 메시지와 감성적 메시지 간에 호감도 차이가 유의미하지 않았지만, 보통 말속도에서는 이성적 메시지와 감성적 메시지 간에 호감도 차이가 유의하였다( $p<.001$ ). 이러한 결과는 보통 말속도에서 감성적 메시지가 이성적 메시지에 비해 청자의 더 높은 호감도를 유도했기 때문에 이차 상호작용 효과가 나타났음을 시사한다. 마지막으로, TTS 말속도, 메시지 유형, 집단 간의 삼차 상호작용 효과는 통계적으로 유의하지 않았다.

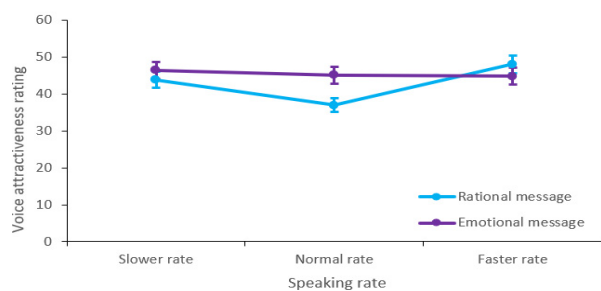


(a) Rational message



(b) Emotional message

**Figure 1.** Voice attractive rating by the listener's age and TTS speaking rate



**Figure 2.** Voice attractiveness rating by speaking rate and message type

#### IV. 논의 및 결론

본 연구는 TTS 음성의 말속도, 메시지 유형, 청자의 연령대가 음성 호감도에 미치는 영향을 살펴보았다. 연구 결과, TTS 말속도는 음성 호감도에 유의미한 영향을 미친 주요 요인으로 나타났다. 구체적으로, 느린 속도와 보통 속도, 보통 속도와 빠른 속도 간에는 유의미한 차이가 관찰되었으나, 느린 속도와 빠른 속도 간에는 유의미한 차이가 나타나지 않았다. 또한, 빠른 속도가 가장 높은 호감도를 유도하였고, 느린 속도가 그 뒤를 이었으며, 보통 속도는 가장 낮은 평가를 받았다. 이는 보통 속도에서 가장 긍정적으로 평가되었다고 보고한 기존 연구(Chun et al., 2024)와 상반되는 결과로, 본 연구의 실험에 포함된 메시지 유형 변수가 말속도 효과에 영향을 미쳤을 가능성을 시사한다. 본 연구는 메시지를 감성적 메시지와 이성적 메시지로 구분하여 제시하였으며, 이러한 메시지 유형이 말속도와 상호작용하여 청자의 호감도 평가에 중요한 영향을 미친 것으로 보인다. 특히, 보통 속도는 특정 메시지 유형과 결합할 때 효과적으로 작용하지 못했을 가능성이 제기된다. 이와 관련하여 Street 등(1984)은 발화 속도의 효과가 발화가 전달되는 상황적 맥락에 따라 달라질 수 있다고 제시하였다. 발화 속도, 멈춤, 발화 길이와 같은 말 행동이 특정 맥락에서 청자의 평가와 인식 형성에 영향을 미칠 수 있다는 관점은 본 연구 결과를 뒷받침한다. 느린 속도와 빠른 속도는 각각 감성적 메시지와 이성적 메시지의 전달력을 강화하며, 청자가 메시지를 더 긍정적으로 평가하도록 유도한 것으로 보인다.

메시지 유형은 음성 호감도에 유의미한 영향을 미치는 중요한 요인으로 나타났으며, 감성적 메시지가 이성적 메시지보다 더 높은 호감도를 유도하였다. 이는 메시지 유형이 청자의 인식에 미치는 영향을 설명하는 기존 연구와 일치하며, 감성적 메시지가 청자의 감정과 느낌을 자극하여 메시지에 대한 긍정적인 반응을 유도하는데 효과적이라는 점(Park & Park, 2007)을 뒷받침한다. 감성적 메시지는 청자의 정서적 유대감을 형성하고 메시지를 매력적으로 전달하여 호감도를 높이는 효과를 발휘한 것으로 보인다. 이러한 결과는 감성적 메시지의 효과를 다룬 기존 연구와도 부합한다. 예를 들어, Ahn 등(2017)은 메시지 소구유형(감성적 소구, 이성적 소구)과 조절 초점 유형이 청자 반응에 미치는 영향을 분석하면서, 감성적 소구가 이성적 소구보다 청자에게 더 긍정적인 반응과 높은 평가를 유도한다고 보고하였다. 본 연구의 TTS 음성 실험에서도 감성적 메시지가 청자의 정서적 반응을 효과적으로 자극하여 메시지를 더 매력적으로 인식하도록 만든 주요 요인으로 작용했을 가능성이 있다. 이는 TTS 음성에서 메시지 유형과 맥락을 고려한 설계가 청자의 호감도를 높이는 데 중요한 역할을 할 수 있음을 시사하며, 감성적 메시지와 조합이 다양한 응용 분야에서 유의미한 사용자 경험을 제공할 수 있음을 보여준다.

본 연구는 말속도와 메시지 유형 간 상호작용이 음성 호감도에 유의미한 영향을 미쳤음을 확인하였다. 감성적 메시지는 느린 속도에서 가장 높은 호감도를 기록하였으며, 이는 느린 속

도가 감성적 내용을 정서적으로 전달하는 데 효과적이라는 기존 연구(Lee, 2020)와 일치한다. 반면, 이성적 메시지에서는 빠른 속도가 가장 높은 호감도를 유도하였고, 이는 빠른 속도가 메시지의 전문성과 신뢰성을 강화한다는 선행 연구(Miller et al., 1976)와 부합한다. 이처럼 메시지 유형에 따라 말속도를 전략적으로 조정함으로써, TTS 음성이 다양한 맥락에서 청자의 호감도를 높이고 긍정적인 사용자 경험을 제공할 수 있음을 시사한다. 예를 들어, 정서적 연결을 강조해야 하는 맥락에서는 느린 속도가, 정보 전달과 같은 효율성이 요구되는 맥락에서는 빠른 속도가 각각 효과적으로 작용할 수 있다.

청자의 연령대(청년층과 노년층)와 말속도의 상호작용은 메시지 유형에 대한 호감도에 유의미한 영향을 미치지 않은 것으로 나타났다. 이는 TTS 음성이 연령대에 관계없이 일관된 사용자 경험을 제공할 가능성을 시사하는 결과로 해석된다. 반면, 연령대의 주효과는 유의미하게 나타났으며, 노년층이 청년층보다 TTS 음성에 대해 더 높은 호감도를 보인 것으로 확인되었다. 이러한 결과는 TTS 기술의 사용이 노년층에서 점차 증가함에 따라(Jeon et al., 2020; Song et al., 2021), 노년층이 TTS 음성에 익숙해지고 이를 보다 긍정적으로 수용하게 되었을 가능성을 시사한다. 이와 같은 결과는 TTS 기술이 모든 연령대에서 안정적이고 일관된 사용자 경험을 제공할 수 있는 잠재력을 가지는 동시에, 노년층 사용자들 사이에서 특히 효과적으로 활용될 수 있는 가능성을 뒷받침한다.

본 연구에서는 TTS의 말속도와 메시지 유형의 조합이 TTS 음성의 호감도에 영향을 미치는 주요 요인임을 확인하였다. 감성적 메시지는 느린 말속도에서 정서적 전달력을 강화하고, 이성적 메시지는 보통 또는 빠른 말속도에서 명료성과 신뢰성을 증대시키는 경향을 보였다. 이러한 결과는 메시지의 특성과 맥락에 따라 말속도를 조정함으로써 TTS 기술의 사용자 경험을 최적화할 수 있음을 시사한다. 더불어, 본 연구는 TTS 음성이 연령대와 관계없이 유사한 호감도를 유도했다는 결과를 통해, TTS 기술이 다양한 사용자층에서 안정적이고 일관된 경험을 제공할 수 있는 잠재력을 보여준다. 특히, 노년층의 AI 스피커 활용 증가(Jeon et al., 2020)와 결합해 볼 때, TTS 기술이 디지털 접근성을 확대하고, 세대 간 기술 격차를 줄이며, 다양한 사용자층에서 사회적 연결성을 강화하는 데 기여할 가능성을 시사한다. 이러한 결과는 TTS 기술이 정보 전달과 정서적 유대 형성을 모두 지원하며, 전 세대를 아우르는 보편적 설계로 발전할 수 있는 기반을 제공한다.

그러나 본 연구에는 몇 가지 한계가 존재한다. 첫째, 말속도와 메시지 유형 외에 억양, 강세, 리듬 등의 초분절적 요소는 고려되지 않았다. 이러한 요소들이 호감도 형성에 미치는 영향을 분석한다면, TTS 음성의 효과를 더 구체적으로 이해할 수 있을 것이다. 둘째, 연구에 참여한 대상자의 표본 크기가 제한적이었다. 총 40명의 참여자로 구성된 본 연구는 샘플의 대표성이 부족하여 결과의 일반화 가능성에 한계가 있을 수 있다. 셋째, 실험은 통제된 환경에서 진행되었기 때문에 실제 의사소통 상황에서 TTS 음성의 효과를 충분히 반영하지 못했을 가능성이 있다. 넷째, 본 연구에서는 실험 참여 전 선별검사로 본

연구에서는 실험 참여 전 선별검사로 KHIE-S(한국어판 노인 난청 지수 검사)와 SGDS(단축형 노인 우울 척도)를 실시하여 참여자의 청력과 정신적 상태를 평가하였다. 그러나 60세 이상의 참여자에 대해서는 인지 상태 검사를 별도로 수행하지 않았다. 이는 고령층 참여자의 잠재적 인지 기능 저하 여부가 충분히 반영되지 못했을 가능성을 내포한다.

후속 연구에서는 초분절적 요소와의 상호작용을 분석하고, 더 다양한 연령대와 배경을 가진 대규모 표본을 대상으로 연구를 진행해야 할 것이다. 또한, TTS 음성이 실제 사용 환경에서 어떤 영향을 미치는지 검증함으로써, 음성 인터페이스 설계와 응용에서 더 구체적이고 실질적인 지침을 제공할 수 있을 것이다. 이를 통해 TTS 기술은 정보 전달뿐 아니라 정서적 유대 형성을 포함한 사용자 경험의 전반적인 향상을 도모할 수 있을 것이다.

## Reference

- Ahn, D., Wang, J., Lee, I., & Jeon, M.-A. (2017). A study on the effects of CSR messages of an airline company on consumer attitude: Focusing on the interactive effects of types of airlines and messages. *Professional Management Studies*, 20(2), 147-165.
- Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer (Version 6.4.25) [Computer software]. <http://www.praat.org>
- Choi, Y., Jung, Y., Kim, Y., Suh, Y., & Kim, H. (2018). An end-to-end synthesis method for Korean text-to-speech systems. *Phonetics and Speech Sciences*, 10(1), 39-48. doi:10.13064/KSSS.2018.10.1.039
- Chun, E., Jeong, Y., Kim, H., Kim, N., Kim, S., & Lee, Y. (2024). Perception of voice attractiveness: Effects of speaking rate, gender, and age. *Communication Sciences & Disorders*, 29(2), 462-472. doi:10.12963/csd.240029
- Hong, B. N., Hong, H., & Lee, J. H. (2004). Development of a screening tool for presbycusis. *Journal of Audiology and Speech Research*, 8(1), 49-57.
- Jang, S., & Lee, J. (2019). User experience research on intimacy and usability when an AI voice-activated personal assistant uses a dialect. *Archives of Design Research*, 32(4), 71-83. doi:10.15187/adr.2019.11.32.4.71
- Jeon, C., Lee, H., & Kim, K. (2020). Trends and business opportunities in the voice AI market. Samjong KPMG Economic Research Institute. Retrieved from <https://assets.kpmg.com/content/dam/kpmg/kr/pdf/2020/kr-im-126-voice-ai-20200414.pdf>
- Jeong, J. H. (1994). *Development of signal processing algorithms to improve the naturalness of synthetic speech*. Korean Institute of Communications and Information Sciences, Electrical Communications Research Project.
- Jeong, S. H. (2017). *Characteristics of speech rate perception according to speech rate in adults* (Master's thesis). Myongji University, Seoul.
- Jo, M. J., Bae, J. N., Seo, G. H., Ham, B. J., Kim, J. K., Lee, D. W., & Kang, M. H. (1999). Validation of geriatric depression scale, Korean version (GDS) in the assessment of DSM-III-R major depression. *Journal of Korean Neuropsychiatric Association*, 38(1), 48-63.
- Kim, I.-G. (2015). A study about the users' preferred playing speeds on categorized video content using WSOLA method. *Journal of Digital Contents Society*, 16(2), 291-298. doi:10.9728/dcs.2015.16.2.291
- Ko, Y. T. (2019). Do you know? Gender-neutral voices... Why AI voices are female. KBS News. Retrieved from <https://news.kbs.co.kr/news/view.do?ncd=4191793>
- Kwon, S. (2015). An experimental study of favorable voice analysis and good impressions using paralinguistic construction elements. *Journal of Speech-Language & Hearing Disorders*, 24(1), 157-167. doi:10.15724/jslhd.2015.24.1.013
- Kwon, S. (2016). Characteristics of the auditory evaluation of good impression using speech manipulation scripts. *Phonetics and Speech Sciences*, 8(4), 131-138. doi:10.13064/KSSS.2016.8.4.131
- Kwon, Y., & Na, E.-Y. (2011). A Study on the effects of the credibility of message source and rational/emotional message types on social support, attitudes, and the intention of health-related Behavior: Focused on the answers of doctors and general public in the Internet health search engine. *Korean Journal of Journalism & Communication Studies*, 55(5), 128-157. uci:G704-000203.2011.55.5.008
- Lee, S. H. (2020). A study on the paralanguage through the examination of main message in persuasive speech. *The Journal of Linguistics Science*, 93, 147-181. doi:10.21296/jls.2020.6.93.147
- Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology*, 34(4), 615-624. doi:10.1037/0022-3514.34.4.615
- Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (6th ed.). Brill Academic Publishers.
- Murphy, D. H., Hoover, K. M., Agadzhanyan, K., Kuehn, J. C., & Castel, A. D. (2022). Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. *Applied Cognitive Psychology*, 36(1), 69-82. doi:10.1002/acp.3899
- Paek, S. J., & Jung, Y. H. (2022). AI voice agent and users' response. *The Journal of Information Systems*, 31(2), 137-158. doi:10.5859/KAIS.2022.31.2.137
- Park, J. W., & Park, H. S. (2007). The effect of involvement, message appeals, and self-efficacy on nonprofit organization's fund-raising campaign. *PR Studies*, 11(1), 107-140. doi:10.15814/jpr.2007.11.1.107
- Park, M. (2020, July 17). Activating elderly care with ICT in collaboration with companies. Boan News. Retrieved from <https://www.boannews.com/media/view.asp?idx=89864>
- Public Data Portal. (2024). Daejeon Metropolitan rapid transit corporation train announcement. Retrieved from <https://www.data.go.kr/data/15104411/fileData.do?recomm>

ndDataYn=Y6-2

- Shin, M., & Lee, K. (2017). Disfluency, speech rate, and communication attitude differences according to gender and age in adults who do not stutter. *Communication Sciences and Disorders, 22*(4), 794-805. doi:10.12963/csd.17431
- Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech, 18*(2), 145-152. doi:10.1177/002383097501800203
- Song, E. S. (2023). *An investigation of factors affecting listening difficulty in announcement for the acoustical design of subway platform* (Master's thesis). Chonnam National University, Gwangju.
- Song, H., Kim, Y., Lee, S., Choi, S., Noh, S., & Lee, Y. (2022). Effects of the acoustic features of a speaker's voice on listener's attractiveness evaluation. *Journal of Speech-Language & Hearing Disorders, 31*(2), 11-21. doi:10.15724/jslhd.2022.31.2.011
- Song, Y. J., Choi, S. J., Kim, J., & Sung, Y. (2021). A study on the factors influencing elderly users' satisfaction with the continued use intention of AI speakers. *Korean Journal of Broadcasting & Telecommunications Research, 114*, 9-37. doi:10.22876/kjbtr.2021.114.001
- Street, R. L., Jr., Brady, R. M., & Lee, R. (1984). Evaluative responses to communicators: The effects of speech rate, sex, and interaction context. *Western Journal of Speech Communication, 48*(1), 14-27. doi:10.1080/10570318409374138
- Voiceware. (n.d.). TTS solution overview. Voiceware. Retrieved from <http://www.voiceware.co.kr/kor/product/product1.ph>
- Walker, M. A., Whittaker, S. J., & Stent, A., Maloor, P., Moore, J., Johnston, M., & Vasireddy, G. (2004). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science, 28*(5), 811-840. doi:10.1016/j.cogsci.2004.06.002
- Wingfield, A., & Ducharme, J. L. (1999). Effects of age and passage difficulty on listening-rate preferences for time-altered speech. *The Journals of Gerontology: Series B, 54*(3), P199-P202. doi:10.1093/geronb/54B.3.P199
- Yoo, J. (2018). Comparisons of paralanguage features between a human announcer and text-to-speech (TTS) devices during the out-loud reading of news sentences. *Journal of Speech Communication, 41*, 49-83. doi:10.18625/jsc.2018.41.49

Appendix 1. TTS voice attractiveness survey questions

분항	
매력도	1. 본 TTS 목소리를 듣고 좋은 감정이 든다. 2. 본 TTS 목소리는 호감 있게 느껴진다. 3. 본 TTS 목소리는 매력적으로 느껴진다. 4. 본 TTS 목소리가 마음에 든다. 5. 본 TTS 목소리는 안정감 있게 느껴진다.
신뢰도	1. 본 TTS 목소리는 설명이 쉽게 들린다. 2. 본 TTS 목소리는 믿음직스럽게 느껴진다. 3. 본 TTS 목소리는 논리적으로 들린다. 4. 본 TTS 목소리는 신뢰가 있게 들린다. 5. 본 TTS 목소리는 신경쓰이게 들린다.
의인화	1. 본 TTS 목소리는 의식을 가지고 있는 것처럼 느껴진다. 2. 본 TTS 목소리는 실제 목소리 같다. 3. 본 TTS 목소리는 사람처럼 느껴진다. 4. 본 TTS 목소리와 대화하고 싶다. 5. 본 TTS 목소리와 감정적 유대가 느껴진다.



## TTS 음성에서 말속도, 메시지 유형 및 청취자 연령이 음성 호감도에 미치는 영향

하나경<sup>1</sup>, 진유나<sup>1</sup>, 박윤지<sup>1</sup>, 심은진<sup>1</sup>, 이영미<sup>2\*</sup>

<sup>1</sup> 이화여자대학교 일반대학원 언어병리학과 석사과정

<sup>2</sup> 이화여자대학교 일반대학원 언어병리학과 교수

**목적:** 본 연구는 TTS 말속도, 메시지 유형, 청자 연령이 음성 호감도에 미치는 영향을 분석하였다.

**방법:** TTS 음성 목록 중 두 개의 여성 화자(20~30대) 음성 선정한 뒤, 세 가지 말속도(느림, 보통, 빠름)와 두 가지 메시지 유형(이성적, 감성적)을 조합하여 총 12개의 음성을 생성하였다. 총 40명의 참가자가 연구에 참여하였으며, 청년층( $n=20$ ), 노년층( $n=20$ )으로 구분되었다. 각 음성에 대한 호감도 평가 후, 연령대별 음성 호감도 차이를 분석하였다.

**결과:** TTS 말속도는 음성 호감도에 유의미한 영향을 미쳐 보통 속도보다 느리거나 빠를 때 더 높은 호감도를 보였다. 감성적 메시지는 이성적 메시지보다, 노년층은 청년층보다 호감도가 높았다. 말속도와 메시지 유형 간 상호작용에서 감성적 메시지는 느린 속도에서, 이성적 메시지는 빠른 속도에서 더 높은 호감도를 보였으며, 삼차 상호작용은 유의하지 않았다.

**결론:** 본 연구는 TTS 음성의 말속도가 음성 호감도에 미치는 영향을 분석하고, 메시지 유형 및 청자의 연령대와의 상호작용 효과를 검증하였다. 연구 결과, TTS 말속도는 음성 호감도에 유의미한 영향을 미쳤으며, 느린 속도와 보통 속도, 보통 속도와 빠른 속도 간 유의미한 차이가 관찰되었다. 감성적 메시지는 느린 속도에서 가장 높은 호감도를 보였고, 이성적 메시지는 빠른 속도에서 가장 긍정적인 평가를 받았다. 또한, 노년층은 청년층보다 전반적으로 더 높은 호감도를 나타냈다. 이러한 결과는 TTS 음성 설계에서 메시지 유형과 말속도를 맥락에 맞게 조정함으로써 사용자 경험을 최적화할 수 있는 가능성을 제시하며, 보완대체의사소통(AAC) 시스템과 같은 응용 분야에서의 활용을 시사한다.

**검색어:** 음성 합성, 말속도, 메시지 유형, 청자 연령, 음성 호감도

**교신저자:** 이영미(이화여자대학교)

**전자메일:** youngmee@ewha.ac.kr

**게재신청일:** 2025. 02. 12

**수정제출일:** 2025. 04. 02

**게재확정일:** 2025. 04. 30

**ORCID**

하나경

<https://orcid.org/0009-0009-5502-1902>

진유나

<https://orcid.org/0009-0005-2205-7336>

박윤지

<https://orcid.org/0009-0003-6976-8480>

심은진

<https://orcid.org/0009-0007-7058-9294>

이영미

<https://orcid.org/0000-0003-1809-5944>

### 참고 문헌

- 권순복 (2015). 준언어적 구성 요소를 통한 매력적인 목소리 분석과 호감도에 관한 실험 연구. **언어치료연구**, 24(1), 157-167.
- 권순복 (2016). 말소리 변조 스크립트를 이용한 호감도 청취평가 특징. **말소리와 음성과학**, 8(4), 131-138.
- 권예지, 나은영 (2011). 정보원의 공신력과 이성적·감성적 메시지 유형이 사회적 지지, 태도, 건강행동의도에 미치는 영향: 건강 관련 검색 서비스의 답변을 중심으로. **한국언론학보**, 55(5), 128-157.
- 김이길 (2015). WSOOLA를 이용한 동영상 미세배속 재생 서비스에 대한 콘텐츠별 배속 선호도 분석 연구. **디지털콘텐츠학회논문지**, 16(2), 291-298.
- 박장원, 박현순 (2007). 기부 관여도, 메시지 소구방식, 자아효능감의 기부 의사 제고 효과에 관한 연구. **PR 연구**, 11(1), 107-140.
- 백승주, 정운혁 (2022). AI 음성 에이전트의 음성 특성에 대한 사용자 반응 연구. **정보시스템연구**, 31(2), 137-158.
- 송유진, 최세정, 김정원, 성용준 (2021). 고령자의 인공지능 스피커 만족도와 지속사용의도에 미치는 영향 요인. **방송통신연구**, 114, 9-37.
- 송은성 (2023). **지하철 승강장 음향설계를 위한 음성내음 청취 어려움의 영향요인 조사**. 전남대학교 대학원 석사학위 논문.
- 송혜선, 김예지, 이승민, 최소라, 노신희, 이영미 (2022). 발화자 목소리의 음성·음향학적 특징이 청자의 호감도 평가에 미치는 영향. **언어**

**치료연구**, 31(2), 11-21.

- 신문자, 이재경 (2017). 일반 성인의 연령과 성에 따른 비유창성 빈도, 말속도와 의사소통 태도. **언어치료연구**, 22(4), 794-805.
- 안대천, 왕진, 이인구, 전민아. (2017). 항공서비스 기업의 CSR 메시지가 소비자 태도에 미치는 영향에 관한 연구: 메시지 특성과 항공사 유형 간의 상호작용효과를 중심으로. **전문경영인연구**, 20(2), 147-165.
- 이소현 (2020). 설득적 말하기의 핵심 메시지 발화에 나타나는 준언어 연구: 속도와 휴지를 중심으로. **언어과학연구**, 93, 147-181.
- 유지철 (2018). 아나운서와 음성합성기의 뉴스 낭독 음성 비교 연구: 유사언어에서의 차이를 중심으로. **화법연구**, 41, 49-83.
- 장순규, 이지훈 (2019). AI 음성비서의 사투리 구사가 사용자의 친밀감과 사용성에 미치는 영향. **디자인학연구**, 32(4), 71-83.
- 전창희, 이효정, 김기범 (2020). **음성 AI 시장의 동향과 비즈니스 기회**. 삼정 KPMG 경제연구원.
- 정상희 (2017). **일반 성인의 말속도에 따른 말속도 지각의 특성**. 명지대학교 대학원 석사학위 논문.
- 정재호 (1994). 합성음의 자연성 향상을 위한 신호처리 알고리즘의 개발. **한국통신학회 전기통신학술연구과제**, 1-108.
- 조맹제, 배재남, 서국희, 함봉진, 김장규, 이동우, 강민희 (1999). DSM-III-R 주요우울증에 대한 한국어판 Geriatric Depression

Scale(GDS)의 진단적 타당성 연구. **신경정신의학**, 38(1), 48-63.  
천은빈, 정유라, 김희연, 김나현, 김수진, 이영미 (2024). 화자의 말속도, 청자의 성별, 연령과 말속도가 청자의 호감도 평가에 미치는 영향. **Communication Sciences & Disorders**, 29(2), 462-472.  
최연주, 정영문, 김영관, 서영주, 김희린 (2018). 한국어 text-to-speech

(TTS) 시스템을 위한 엔드투엔드 합성 방식 연구. **말소리와 음성 과학**, 10(1), 39-48.  
홍빛나, 홍하나, 이정학 (2004). 노인성난청 선별을 위한 검사도구 개발. **대한청각학회지**, 8(1), 49-57.