JSLHD 2025. Vol. 34, No. 2. 087-095
ISSN 1226-587X / eISSN 2671-7158
Journal of Speech-Language & Hearing Disorders 언어치료연구

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2025.34.2.087
http://jslhd.org/

A Comparison of Speech Characteristics in Human-Human and Human-Conversational AI Interactions: Focusing on Speech Rate, Pitch, and Formants

Da Hee Jung[1], Yeoeun Yoon[1], Ji Yeon Lee[1], Sohyun Baik[1], Youngmee Lee[2*]

[1] *Dept. of Communication Disorders, Graduate School, Ewha Womans University, Master's Student*
[2] *Dept. of Communication Disorders, Graduate School, Ewha Womans University, Professor*

Purpose : This study aimed to investigate the differences in the acoustic characteristics of human speech depending on the type of interaction partner in real-life interaction scenarios: human-to-human and human-to-conversational AI interactions. Additionally, this study aimed to examine changes in speech production that occur during interactions in terms of temporal (speech rate), prosodic (pitch), and articulatory (formant) characteristics.
Methods : A total of 23 adults (11 men and 12 women) participated in a twenty-question game on the topic of animals, first with human partners and then with a conversational AI (Chat GPT 4.0). During a three-minute real-time interaction involving questions and answers, voice samples collected in the interactions from the participants were analyzed using the Praat speech analysis program to evaluate speech rate, pitch, and formant features.
Results : The analysis revealed significant differences in speech rate and formant between human and conversational AI interactions. In other words, interactions with the conversational AI exhibited a faster speech rate and higher formant values (F1 and F2) compared to interactions with humans, whereas pitch-related parameters (F0 average, F0 range) did not exhibit significant differences.
Conclusions : This study identified distinctive speech patterns emerging during direct interactions with human partners versus conversational AI partners, showing that humans tended to produce faster speech rates and clearer articulation when interacting with conversational AI. These findings suggest foundational data that could be the potential for developing conversational AI systems capable of precisely adjusting acoustic elements, which could have promising applications in speech and language therapy.

Keywords : Human-conversational AI interactions, speech rate, pitch, formants

Correspondence : Youngmee Lee, PhD
E-mail : youngmee@ewha.ac.kr

Received : February 12, 2025
Revision revised : April 06, 2025
Accepted : April 30, 2025

ORCID
Da Hee Jung
https://orcid.org/0009-0002-6673-7745
Yeoeun Yoon
https://orcid.org/0009-0002-0294-1771
Ji Yeon Lee
https://orcid.org/0009-0009-5087-1107
Sohyun Baik
https://orcid.org/0009-0000-0249-6693
Youngmee Lee
https://orcid.org/0000-0003-1809-5944

Ⅰ. 서론

인공지능(artificial intelligence: AI)은 현대 사회에서 필수적인 기반 기술로 자리 잡았으며, 산업 전반에서 핵심적인 역할을 수행하고 있다(Kim, 2016). 과학기술정보통신부와 소프트웨어 정책연구소의 인공지능 산업 실태조사(Ministry of Science and ICT & Software Policy Institute, 2023)에 따르면, 2022년에서 2023년 사이 AI 매출액은 평균 21.5% 증가하였으며, AI는 학습용 데이터, AI 서비스, 가상 증강현실, 의료 분야(ICT Statistics Portal, 2022)뿐만 아니라 일상생활의 다양한 영역에서 활용되고 있다. AI는 인간처럼 학습할 수 있는 기계로 정의되며, AI 분야의 선구자인 John McCarthy는 이를 "지능적인 기계를 만드는 과학과 공학"으로 규정하였다(Rajaraman, 2014). 이 중에서도 대화형 AI(conversational AI)는 AI의 하위 분야로, 자연어 처리 기술을 기반으로 사용자와 플랫폼 간의 상호작용을 강화하는 데 중점을 둔다(Choi & Song, 2024). 이에 따라 인간과 대화형 AI 간의 상호작용 및 정서적 지원에 관한 연구가 활발히 진행되고 있다(Broekens et al., 2009; Skjuve et al., 2021; Song, 2022). Choi(2023)는 지능화된 대화형 AI가 인간과 유사한 대화 기술을 가지고 있으며, 장기간의 상호작용을 통해 수집된 사용자 정보를 활용해 애정과 신뢰의 관계를 형성할 수 있다고 주장하였다. Brandtzaeg 등(2022)의 연구에서도 대화형 AI와 인간의 상호작용이 높은 신뢰 관계를 형성할 수 있으며, 이를 통해 생각, 경험, 아이디어를 공유하는 등 사용자가 대화형 AI를 상호적인 관계로 인식하는 모습을 보였다고 보고하였다. 이러한 연구는 대화형 AI가 단순한 기술적 도구를 넘어 인간과의 상호작용에서 정서적 유대와 신뢰를 형성할 가능성을 지니고 있음을 시사한다.

인간은 특정 상황에서 상대방의 특성에 따라 발화 패턴을 조정하는 경향이 있다. 이러한 발화 조정은 단순히 대화의 흐름을

자연스럽게 연결하고 효율성을 높이는 것에 그치지 않고, 상대방의 특성을 고려한 의사소통 전략으로 나타난다. 예를 들면, 부모가 어린 자녀에게 사용하는 아동지향어(infant-directed speech)는 성인지향어(adult directed speech)에 비해 높은 음도, 넓은 음도 범위, 느린 말속도, 반복적 어휘 사용 등의 특징을 지니며, 이러한 아동지향어는 아동의 주의집중과 어휘 습득을 촉진하고 정서적 반응성을 향상하는 데 기여하는 것으로 알려져 있다(Bergeson et al., 2006; Park & Lee, 2023). 나아가 인간은 특정 참조물(referents)을 의인화할 때, 독특한 음성학적 특성이나 발화 유형을 보이는 경향이 있다. 선행 연구에 따르면, 인간은 의인화된 대상과 상호작용을 할 때 말속도, 억양, 음도, 명료도 등의 다양한 음향학적 요소를 조정한다. 예를 들면, 강아지를 대상으로 하는 발화에서는 과장된 억양과 높은 음도가 두드러지며(Ben-Aderet et al., 2017), 로봇과의 상호작용에서도 유사한 패턴이 나타난다. 특히, 로봇이 인간과 유사한 외형을 가질 경우, 인간은 더욱 명료하고 의식적인 발화 특성을 보이는 것으로 보고되었다(Kalashnikova et al., 2023). 이러한 발화 패턴은 의사소통의 효율성을 넘어, 동물이나 사물과의 상호작용에서도 인간이 상대방을 사회적 대상으로 인식하고 있음을 시사한다. 이와 같은 의인화 발화의 특성은 대화형 AI와의 상호작용에서도 음향학적 요소를 포함한 여러 요인에 따라 인간의 발화가 조정되는 유사한 양상으로 나타날 가능성이 있다. Cohn 등(2022)의 연구에 따르면, 인간은 사전 녹음 및 스크립트를 기반으로 한 인간과 인간, 인간과 대화형 AI와의 상호작용에서 모두 발화 조정을 하는 것이 확인되었다. 예를 들어, 대화형 AI의 음성 인식 오류율이 높을 경우, 인간의 발화에서 평균 F0와 F0 범위가 증가하고 모음 과장이 나타났고, 음성 인식 오류율이 낮을 때는 평균 F0와 F0 범위가 감소하였다. 또한, 대화형 AI의 말속도는 인간의 말속도에도 영향을 미쳤으며(Cohn et al., 2021), 느린 말속도뿐만 아니라 평균 F0와 F0 변이성 증가도 확인되었다(Cohn & Zellou, 2021). 이는 대화형 AI가 단순한 기계가 아닌 인간과 유사한 존재로 인식되어 '지향어(directed speech)'와 같은 특수한 발화 형태를 유도할 수 있음을 시사한다.

이처럼 인간 음성과 대화형 AI 음성 간 음향학적 요소를 간접적으로 측정하거나 대화형 AI에 대한 친밀감과 의인화에 대한 인식을 조사한 연구들은 다수 존재한다. 또한, 대화형 AI와 유사한 기술적 혹은 기계적 시스템을 대상으로 발화 차이를 분석한 연구도 보고되고 있다(Burnham et al., 2010; Mayo et al., 2012; Siegert et al., 2019). 그러나 대화형 AI와의 직접적인 상호작용에서 나타나는 인간 발화의 음향학적 특성을 체계적으로 분석한 연구는 제한적이며, 대화 상대의 유형에 따른 발화 특성 변화를 심층적으로 탐구한 사례는 드물다. 앞서 제시한 여러 선행연구로 미루어 볼 때, 인간-인간, 인간-대화형 AI 간 상호작용 비교는 대화형 AI와의 상호작용 시 나타나는 발화 조정 유무와 그 양상까지도 함께 파악할 수 있는 근거 자료로 활용될 수 있을 것이다. 따라서 본 연구에서는 인간-인간 상호작용과 인간-대화형 AI 상호작용에서 나타나는 발화의 음향학적 특성을 비교하고, 대화 상대 유형에 따라 시간적, 운율적, 조음적 차원에서 발화가 어떻게 변화하는지 살펴보고자 하였다. 이에 따른 본 연구의 구체적인 연구 질문은 다음과 같다.

첫째, 양방향 소통 게임 과제에서 대화 상대 유형(대화형 AI, 인간)에 따른 대화자 발화의 시간 차원의 음성 특성(말속도)에 유의한 차이가 있는가?

둘째, 양방향 소통 게임 과제에서 대화 상대 유형(대화형 AI, 인간)에 따른 대화자 발화의 운율 차원의 음성 특성(F0 평균 및 범위)에 유의한 차이가 있는가?

셋째, 양방향 소통 게임 과제에서 대화 상대 유형(대화형 AI, 인간)에 따른 대화자 발화의 조음 차원의 음성 특성(F1값, F2값)에 유의한 차이가 있는가?

II. 연구 방법

1. 연구 대상

본 연구는 20~30대 청년 23명(남 11명, 여 12명)을 대상으로 하였으며, 선별 기준은 다음과 같다. 모든 대상자는 (1)한국어를 모국어로 사용하고, (2)자가 보고(self-report)로 청력에 이상이 없으며, (3)음성 분석 시 방언 특성이 변수로 작용할 수 있으므로 서울말 사용자로 한정되었다. 연구 대상자는 지역 커뮤니티 등을 통하여 비대면 방식으로 모집되었다. 연구 대상자의 평균 연령은 28.57세(SD=4.85)였으며, 연령 범위는 22~38세였다. 연구 대상자는 모두 대화형 AI를 사용한 경험이 있었으며, 주된 사용 목적은 정보 습득 및 학습, 생산성, 유희와 휴식으로 나타났다.

2. 연구 과제

본 연구에서는 인간-대화형 AI 상호작용의 발화 특성을 비교하기 위한 과제로 질문과 대답을 주고받을 수 있는 양방향 소통게임을 선정하였다. 일반적으로 자발화 과제로는 경험 말하기, 그림 설명하기, 이야기 다시 말하기, 대화 과제 등이 다양하게 사용되고 있다(Smith et al., 2003). 그러나 대화형 AI는 인간과 달리 실제 경험을 공유하며 소통하는 것이 불가능하기 때문에 경험 유무에 상관없이 질문과 대답으로 상호작용이 가능한 스무고개 게임으로 대화 주제를 통제하였다. 이는 반구조화된 자발화 과제에 해당하며, 그중에서도 '스무고개 게임'을 선택한 구체적인 이유는 다음과 같다. 첫째, 스무고개 게임은 하나의 주제에 대해 연구 대상자들이 질문과 대답을 주고받으며 단어를 맞히는 방식으로 진행된다(Kim et al., 2024). 둘째, 스무고개 게임은 연구 대상자들이 실시간으로 정보를 주고받고 협력적으로 문제를 해결하는 특징을 지닌다(Stocco et al., 2015). 이러한 이유로, 본 연구에서는 스무고개 게임을 인간-대화형 AI 상호작용 상황에 적용하여 대화 상대의 유형에 따라 나타나는 발화 특성의 차이를 확인하고자 하였다.

본 연구에서는 스무고개 게임을 다음과 같이 구성하였다. 연구자들은 Chat GPT 4.0과 스무고개 게임을 실시한 후, 질문과 대

답이 15회 이상 도출될 수 있는 동물 단어를 선정하여 단어 목록을 구성하였다('청설모', '악어', '물소', '공작새', '비버'). 이후, 연구 대상자는 다섯 개의 단어 목록 중 하나의 단어를 선택하여 인간과 대화형 AI 순서로 스무고개 게임에 참여하였다. 이때, 연구 대상자는 대화 상대 유형에게 스무고개 게임을 제안하고, 동물과 관련된 질문을 할 것을 요청하였다. 대화 상대는 질문자, 연구 대상자가 응답자의 역할을 맡았으며, 대화 상대가 정답을 맞히거나 제한 시간이 경과하면 실험이 종료되었다.

3. 실험 절차

발화 수집은 소음이 없는 공간에서 진행되었으며, 모든 발화는 iPhone 13 스마트 레코더 앱(Shenzhen DSQN Investment Co., Ltd, China)과 소형 핀 마이크(HC-220UC, DoowonTrade, Daegu, Korea)를 사용하여 녹음되었다. 녹음 과정에서 표본 추출률은 44.1kHz, 양자화 24bit로 설정하였고, 입과 마이크 사이의 거리는 10cm를 유지하였다. 대화형 AI로는 대용량 데이터를 학습하여 인간과 자연스러운 상호작용이 가능한 Chat GPT의 최신 모델인 Chat GPT 4.0을 사용하였다. 보다 구체적으로 설명하자면, Multi-Task 데이터 기반의 훈련된 자동 음성 인식(automatic speech recognition: ASR) 시스템인 Whisper (Open AI, 2022), 텍스트 처리 변환이 가능하고 인간과 유사한 속도의 응답을 지원하는 Autoregressive Omni Model, 인간과 대화형 AI 간의 자연스러운 상호작용을 지원하는 음성 합성 시스템(text to speech: TTS) 등을 활용하여(Open AI, 2024), 휴대폰에 내장된 음성을 이용하였다. 또한, 남성 음성보다 여성 음성이 사용자에게 높은 신뢰도와 매력도를 준다는 연구 결과(Beak & Jung, 2022)를 바탕으로, 대화형 AI 음성은 여성 음성으로 설정하였으며, 인간과의 상호작용 시에도 여성 음성으로 동일하게 적용하였다.

연구 대상자에게 연구 절차를 사전에 안내하고 자발적 동의를 획득한 후, 과제에 익숙해질 수 있도록 충분한 연습 기회를 제공하였다. 본 실험에서 연구 대상자는 먼저 3분간 동물을 주제로 인간과 스무고개 게임을 진행하였다. 이때, 비언어적 요소(예, 표정, 제스처 등)가 발화에 미치는 영향을 통제하기 위하여 연구 대상자는 연구자와 분리된 공간에서 음성 통화를 하는 방식으로 상호작용하였다. 이후, 연구 대상자는 연구자가 있는 공간으로 이동하여 동일한 시간 조건 하에서 Chat GPT 4.0 기반의 대화형 AI와 스무고개 게임을 진행하였다. 이때, 연구자는 연구 대상자의 발화만을 관찰하고 실험에 직접적으로 개입하지 않았다.

4. 음성 분석

녹음된 발화는 Praat(Ver. 6.4.23)을 사용하여 시간적, 운율적, 조음적 특성을 분석하였다. 분석에 사용된 데이터는 포먼트 주파수 변화를 가장 잘 관찰할 수 있는 연속된 7개의 발화를 전사한 후 선별하여 처리하였다(Kang et al., 2024; Oh et al., 2014). 이때, 발화 단위는 "일정한 언어 행위를 수행하는 의사소통의 최소 단위로, 문맥과 상황성이 포함된 단위"로 정의하였다 (Ahn, 2007; Zifonun, 1987, as cited in Lee, 2002). 발화에 대한 시간적, 운율적, 조음적 특성은 다음과 같은 방법으로 분석하였다.

1) 시간 차원의 음성 특성

전체 말속도(overall speech rate)는 초당 음절 수를 기준으로 측정하였다(Shin, 2018; Yu, 2019). 선별된 연속 7개의 발화의 전체 발화 시간을 음절 수로 나누고, 이 과정에서 휴지 시간을 포함하여 계산하였다.

2) 운율 차원의 음성 특성

F0의 평균은 Praat의 voice report 기능을 통해 측정하였으며(Kwon et al., 2022), F0의 범위는 최고 F0와 최저 F0의 차이로 계산하여 산출하였다.

3) 조음 차원의 음성 특성

발화 전사 후, 가장 빈도수가 높은 모음 /ㅏ/, /ㅣ/, /ㅗ/가 포함된 발화를 선정하였다(Kang et al., 2024). 포먼트 분석을 위해 자음이 포함되지 않은 음절의 모음을 우선 선정하였으며(Lee et al., 2016), 대상자의 F1값과 F2값은 스펙트로그램 상의 에너지가 집중된 안정 구간에서 Praat의 formant listing을 통해 측정하였다(Kang et al., 2024). 조음 차원의 음성 분석 시 사용한 문장의 예시는 "아니요, 포유류가 아닙니다"이며, 이는 다음과 같다(Figure 1).

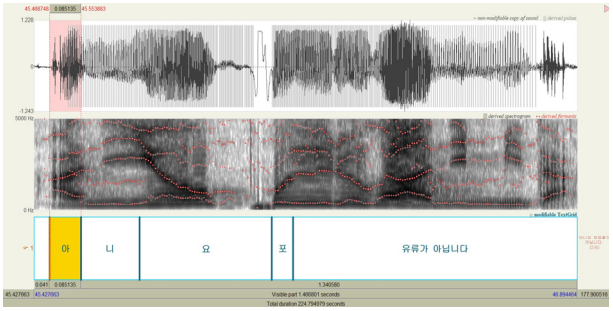

Figure 1. Example of formant analysis using Praat

5. 신뢰도

평가자 내 신뢰도(intra-rater reliability) 및 평가자 간 신뢰도(inter-rater reliability)를 산출하기 위하여, 전체 발화 자료 중 20%에 해당하는 발화를 무작위로 선택하여 분석하였다. 평가자 내 신뢰도는 제 4저자가 전체 자료의 20%에 해당하는 자료를 무작위로 선정하여 첫 번째와 두 번째 분석을 비교하였다. 피어슨 상관계수(Pearson correlation coefficients)를 사용하여 신뢰도를 측정한 결과, 말속도(r=.919, p=.001), F0 평균(r=.994, p=.000), F0 범위(r=.988, p=.000), F1/ㅏ/(r=.939, p=.001), F1/ㅣ/ (r=.959, p=.000), F1/ㅗ/(r=.863, p=.006), F2/ㅏ/(r=.903, p=.002), F2/ㅣ/(r=.935, p=.001), F2/ㅗ/(r=.995, p=.000)였다. 평가자 간 신뢰도를 측정하기 위해,

제 2저자와 제 4저자가 독립적으로 변수를 분석하였다. 피어슨 상관계수로 신뢰도를 측정한 결과, 말속도(r=.839, p=.009), F0 평균(r=.868, p=.005), F0 범위(r=.752, p=.031), F1/ㅏ/(r=.929, p=.001), F1/ㅣ/(r=.804, p=.016), F1/ㅗ/(r=.936, p=.001), F2/ㅏ/(r=.779, p=.023), F2/ㅣ/(r=.852, p=.007), F2/ㅗ/(r=.948, p=.000)였다.

6. 자료의 통계 처리

본 연구에서는 IBM SPSS statistics version 30.0(IBM, Armonk, NY, USA)을 이용하여 통계분석을 실시하였다. 양방향 소통 게임 상황에서 대화 상대 유형(대화형 AI, 인간)에 따른 대화자 발화의 말속도, F0 평균, F0 범위, F1값, F2값에서 유의한 차이가 있는지를 확인하기 위해 대응표본 t-검정(paired samples t-test)을 실시하였다.

Ⅲ. 연구 결과

1. 대화 상대 유형에 따른 시간 차원의 음성 특성

대화 상대 유형에 따른 대화자 발화의 말속도에 대한 통계 결과는 Table 1에 제시하였다. 대응표본 t-검정을 실시한 결과, 대화 상대 유형에 따른 대화자 발화의 말속도에 유의한 차이가 있었다(t=-3.121, p=.005).

Table 1. Descriptive statistics of temporal acoustic features of speaker utterances

	Human	Conversational AI
Speech rate (sps)	5.27 (.87)	5.73 (.75)

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

2. 대화 상대 유형에 따른 운율 차원의 음성 특성

대화 상대 유형에 따른 대화자 발화의 F0 평균, F0 범위에 대한 통계 결과는 Table 2에 제시하였다. 대응표본 t-검정을 실시한 결과, 대화 상대 유형에 따른 대화자 발화의 F0 평균(t=-1.083, p=.290), F0 범위(t=.536, p=.597)에 유의한 차이가 없었다.

Table 2. Descriptive statistics of the prosodic features of speaker utterances

	Human	Conversational AI
F0 average (Hz)	183.65 (49.60)	186.15 (51.70)
F0 range (Hz)	224.67 (80.52)	217.83 (79.19)

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

3. 대화 상대 유형에 따른 조음 차원의 음성 특성

대화 상대 유형에 따른 대화자 발화의 F1값, F2값에 대한 통계 결과는 Table 3에 제시하였다. 대응표본 t-검정을 실시한 결과, 대화 상대 유형에 따른 대화자 발화의 F1/ㅣ/값에 유의한 차이가 있었으나(t=-3.024, p=.006), F1/ㅏ/값과 F1/ㅗ/값에서는 유의한 차이가 없었다(t=-.484, p=.633; t=-1.242, p=.227). 그리고 대화 상대 유형에 따른 대화자 발화의 F2/ㅣ/값과 F2/ㅗ/값에 유의한 차이가 있었으나(t=-2.138, p=.044; t=-3.053, p=.006), F2/ㅏ/값에서는 유의한 차이가 없었다(t=1.007, p=.325).

Table 3. Descriptive statistics of the articulatory features of speaker utterances

	Human	Conversational AI
F1/ㅏ/ (Hz)	853.59 (190.13)	865.66 (180.88)
F1/ㅣ/ (Hz)	362.61(80.30)	473.39 (176.62)
F1/ㅗ/ (Hz)	485.29 (93.79)	515.04 (113.93)
F2/ㅏ/ (Hz)	1419.61 (212.35)	1378.03 (194.20)
F2/ㅣ/ (Hz)	1656.58 (660.88)	1994.07 (710.13)
F2/ㅗ/ (Hz)	968.84 (194.56)	1329.30 (514.87)

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

Ⅳ. 논의 및 결론

본 연구에서는 대화 상대 유형(대화형 AI, 인간)에 따라 시간, 운율, 조음 차원에서 대화자 발화의 음향학적 특성이 어떻게 변화하는지 살펴보았다. 그 결과, 시간 차원에서는 대화자가 인간과 상호작용할 때보다 대화형 AI와 상호작용할 때 말속도가 유의하게 빠른 것으로 나타났다. 운율 차원에서는 대화 상대 유형에 따라 대화자의 F0 평균과 F0 범위에서 유의한 차이가 나타나지 않았다. 조음 차원에서는 대화자가 인간과 상호작용할 때보다 대화형 AI와 상호작용할 때, F1/ㅣ/값, F2/ㅣ/값, F2/ㅗ/값이 유의하게 높았다. 이러한 연구 결과에 대한 논의는 다음과 같다.

첫째, 대화 상대 유형에 따른 대화자의 말속도를 비교한 결과, 대화자는 인간보다는 대화형 AI를 대상으로 발화할 때 말속도가 유의하게 빠른 것으로 나타났다. Cohn 등(2021)의 연구에서는 실험실 내 관찰자의 존재가 참여자들에게 심리적으로 영향을 미쳐 말속도가 느려졌을 가능성을 제시하였다. 그러나 본 연구에서는 관찰자가 실험실 내에 있었음에도 불구하고, 최신 기술을 탑재한 Chat GPT 4.0의 자연스러운 대화 능력으로 관찰자의 존재가 대화 속도에 영향을 미치지 않았을 가능성이 있다. 이러한 결과는 선행 연구(Cohn et al., 2021)에서 대화자가 소셜 봇(social bot)과 대화 시에 말속도가 감소했다고 보고

한 연구 결과와 상반된다. 선행 연구에서 사용된 Amazon Alexa는 날씨 알려주기, 알람 설정하기, 일반적인 질문에 대답하기와 같은 간단한 명령 처리만이 가능하다(Lopatovska et al., 2019). 반면 본 연구에서 사용된 Chat GPT 4.0은 대규모 언어 모델(large language model: LLM)을 기반으로 방대한 자연어를 학습하기 때문에 자연스러운 대화와 정교한 응답이 가능하다(Han, 2023). 따라서 Chat GPT 4.0이 대화 상대의 반응 속도와 명료성을 강화하여 대화자가 AI를 인간과 유사한 존재로 인식하는 의인화 효과를 유발했으며, 이로 인해 말속도가 증가했을 가능성이 있다. 이와 유사하게, Nass와 Moon(2000)은 인간이 컴퓨터와 상호작용할 때 컴퓨터가 기계임을 인지하고 있음에도 불구하고, 무의식적으로 인간과 상호작용하는 것처럼 반응한다고 보았다. 또한, 인간은 특정한 사회적 상호작용 방식 하에서 상황적 단서를 탐색하지 않고 해당 방식에 따라 자동적으로 반응한다고 하였다. 이는 대화자가 대화 상대의 상호작용 특성에 따라 말하기 방식을 자동적으로 조절할 수 있다는 것을 시사하며, 대화형 AI의 빠른 반응 속도가 인간의 발화에 영향을 주는 방식으로 작용하여 말속도 증가를 자연스럽게 유발했을 가능성이 있다.

둘째, 대화 상대 유형에 따른 대화자의 F0 평균과 F0 범위를 비교한 결과, 유의한 차이가 나타나지 않았다. 이러한 결과는 대화자가 대화형 AI와 상호작용할 때 인간과 상호작용할 때와 유사한 음도를 유지했음을 의미한다. 이는 선행 연구(Cohn & Zellou, 2021; Cohn et al., 2022)에서 보고한 결과와 일치하지 않는다. Cohn과 Zellou(2021)는 대화자가 인간과 상호작용할 때보다 대화형 AI와 상호작용할 때 더 높은 평균 F0와 더 넓은 F0 범위를 보였다고 하였으며, Cohn 등(2022)의 연구에서는 대화형 AI의 오류율에 따라 대화자의 F0 평균과 범위가 변화한다고 보고하였다. 본 연구에서 대화자가 대화형 AI와 상호작용할 때 인간과 상호작용할 때와 유사한 음도를 산출한 것은 대화형 AI가 대화 흐름을 자연스럽게 조절하고, 인간 대화 상대와 유사한 상호작용 경험을 제공했기 때문으로 해석된다(Kwon & Jeong, 2024). 특히, 초기 AI 모델과 달리 Chat GPT와 같은 최신 대화형 AI는 자연어 처리 기능이 향상되어 대화 중 음도 조정의 필요성을 감소시킨 것으로 보인다.

셋째, 대화 상대 유형에 따른 대화자 발화에서의 F1값과 F2값을 비교한 결과, F1/ㅣ/값, F2/ㅣ/값, F2/ㅗ/값에서 유의한 차이가 나타났다. Cohn과 Zellou(2021)는 대화형 AI가 발화를 잘못 인식했을 경우, 대화자가 '마스크(mask)-모스크(mosque)'와 같은 최소쌍(minimal pair)을 활용하여 오류를 수정하며, 후설모음의 모음공간을 확장하는 방식으로 F2값을 조정한다고 보고하였다. 이와 같은 조정은 모음과 모음을 더 뚜렷하게 변별하기 위한 시도로 해석할 수 있다(Cohn & Zellou, 2021). 본 연구에서도 대화형 AI를 대상으로 한 발화에서 모음을 과도하게 명료화하는 모음 과조음(vowel hyperarticulation) 현상이 나타난 것으로 보인다(Cohn et al., 2022). 이는 대화형 AI와의 상호작용에서 인간이 대화형 AI의 대화 능력을 인간과 유사하게 인식하면서도, 기계적 한계를 고려하여 말 명료도를 높이려는 경향을 보였음을 시사한다.

본 연구의 의의는 다음과 같다. 첫째, 본 연구는 반구조화된 자발화를 기반으로 한 직접적인 양방향 소통 과제를 통해, 인간이 대화형 AI와 상호작용할 때 빠르게 말속도를 조정하고, 더 정확하게 조음하려는 경향이 있다는 것을 확인하였다. 이는 대화자의 발화 특성에 따라 음향학적 요소를 정교하게 조정할 수 있는 대화형 AI가 개발될 수 있는 가능성을 제시한다. 둘째, 본 연구는 인간이 대화형 AI와 상호작용할 때 인간과 상호작용할 때와 마찬가지로 음도 측면에서 자신의 기본적인 발화 특성을 안정적으로 유지할 수 있다는 것을 확인하였다. 이는 대화자의 특성에 맞게 유연하게 발화를 조절할 수 있는 시간적, 일부 조음적 차원의 음성 특성과 달리, 발성기관의 특징에 영향을 받는 운율적 차원의 음성 특성에서는 일관된 발화 산출을 보인다는 것을 의미한다. 이러한 점을 바탕으로, AI 스피커가 말소리 장애 아동의 조음 훈련(Park et al., 2020), 언어장애 아동과 학습장애 아동의 읽기 유창성(Hwang et al., 2020), 자폐 스펙트럼 장애 아동의 부르기, 대답하기 기술 향상(Jung, 2020)에 긍정적인 영향을 준 선행 연구를 미루어 볼 때, 대화형 AI도 마찬가지로 향후 언어재활에 있어 활용 가능성이 있다는 것을 시사한다. 즉, 대화형 AI는 대화자 발화의 유연성과 안정성을 모두 고려한 중재 도구로써 임상 현장에서 개인화된 맞춤형 언어재활 서비스를 제공하는 데 기여할 수 있을 것이다.

이러한 본 연구의 의의에도 불구하고, 본 연구의 한계는 다음과 같다. 첫째, 본 연구는 청년층 23명만을 대상으로 하였기 때문에, 연구 결과를 다양한 연령대와 문화적 배경을 가진 집단으로 일반화하기 어렵다. 후속 연구에서는 다양한 연령대와 경험을 가진 연구 대상자를 모집하여, 연령별, 경험별 차이를 확인할 필요가 있을 것이다. 둘째, 본 연구는 반구조화된 자발화 과제를 사용하여 설계되었기 때문에, 연구 대상자의 이중모음 산출이 비일관적으로 나타나는 한계가 있었다. 이에 따라, 조음의 동적 특성이나 조음 차원의 다양한 음향학적 요소를 충분히 분석하지 못하였다. 이중모음은 단모음에 비해 복잡한 조음 운동 능력을 요구하므로(Lee et al., 2023), 후속 연구에서는 문단 또는 문장 읽기와 같은 통제된 발화 과제를 활용하여 이중모음의 음향학적 특징과 동시 조음 상황에서의 발화 특성을 심층적으로 분석할 필요가 있다.

Reference

Ahn, B. S. (2007). A study on utterance as prosodic unit for utterance phonology. *The Journal of Korean Studies, 26,* 233-259. doi:10.17790/kors.2007..26.233

Beak, 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

Ben-Aderet, T., Gallego-Abenza, M., Reby, D., & Mathevon, N. (2017). Dog-directed speech: Why do we use it and do dogs pay attention to it? *Proceedings of the Royal Society B: Biological Sciences, 284*(1846), 20162429. doi:10.1098/

rspb.2016.2429

Bergeson, T. R., Miller, R. J., & McCune, K. (2006). Mothers' speech to hearing-impaired infants and children with cochlear implants. *Infancy, 10*(3), 221-240. doi:10.1207/s15327078in1003_2

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human-AI friendship. *Human Communication Research, 48*(3), 404-429. doi:10.1093/hcr/hqac008

Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive social robots in elderly care: A review. *Gerontechnology, 8*(2), 94-103. doi:10.4017/gt.2009.08.02.002.00

Burnham, D., Joeffry, S., & Rice, L. (2010). Computer- and human-directed speech before and after correction. *Proceedings of 13th Australasian International Conference on Speech Science and Technology,* 13-17.

Choi, S. (2023). A study on building a university e-learning center through cloud-based conversational AI chatbots (machine learning). *Proceedings of the Korean Policy Studies Association Summer Conference,* 1-4.

Choi, S. A., & Song, Y. H. (2024). A study on technology convergence trends and promising technologies for promising technology-based entrepreneurship: Focus on conversational AI. *Journal of the Korea Entrepreneurship Society, 19*(2), 151-171. doi:10.24878/tkes.2024.19.2.151

Cohn, M., Liang, K.-H., Sarian, M., Zellou, G., & Yu, Z. (2021). Speech rate adjustments in conversations with an Amazon Alexa socialbot. *Frontiers in Communication, 6.* doi:10.3389/fcomm.2021.671429

Cohn, M., Segedin, B. F., & Zellou, G. (2022). Acoustic-phonetic properties of Siri- and human-directed speech. *Journal of Phonetics, 90,* 101123. doi:10.1016/j.wocn.2021.101123

Cohn, M., & Zellou, G. (2021). Prosodic differences in human- and Alexa-directed speech, but similar local intelligibility adjustments. *Frontiers in Communication, 6.* doi:10.3389/fcomm.2021.675704

Han, S. H. (2023). Korean speaking study using conversational generative AI (artificial intelligence) ChatGPT: Based on role-playing, from using Talk-to-ChatGPT to utilizing AIPRM-for-ChatGPT. *Journal of Learner-Centered Curriculum and Instruction, 23*(18), 651-674. doi:10.22251/jlcci.2023.23.18.651

Hwang, D. J., Pyo, S. M., & Kim, B. A. (2020). The effects of repetitive reading intervention through artificial intelligence on reading fluency of children with language learning disabilities. *Proceedings of the 21st Korean Speech-Language & Hearing Association,* 292-297.

ICT Statistics Portal. (2022). 2022 4th Industrial revolution indicators [Infographic]. Retrieved from https://url.kr/ebfgyo

Jung, S. I. (2020). Case study on intervention of conversational skills in children with autism spectrum disorder ssing AI speaker. *Proceedings of 21st Korean Speech-Language & Hearing Association,* 309-314.

Kalashnikova, N., Hutin, M., Vasilescu, I., & Devillers, L. (2023). Do we speak to robots looking like humans as we speak to humans? A study of pitch in French human-machine and human-human interactions. *Proceedings of Companion Publication of the 25th International Conference on Multimodal Interaction,* 141-145.

Kang, S. M., Kang, E. H., & Lee, J. Y. (2024). The correlations between vowel acoustic variables and speech intelligibility and acceptability in school-aged children with intellectual disabilities. *Journal of Speech-Language & Hearing Disorders, 33*(1), 91-100. doi:10.15724/jslhd.2024.33.1.091

Kim, B. W. (2016). Trend analysis and national policy for artificial intelligence. *Informatization Policy, 23*(1), 74-93. doi:10.22693/NIAIP.2016.23.1.074

Kim, Y. S., Kim, Y. S., Lee, H. J., & Chae, M. (2024). A study on a platform for treating autistic children with interactive AI. *Proceedings of 2024 Annual Conference of KIPS, 31*(1), 480-481.

Kwon, H. N., & Jeong, J. J. (2024). A multidimensional approach to the anthropomorphism of conversational generative AI: Centered on Haslam's dehumanization model. *Broadcasting & Communication, 25*(3), 44-99. doi:10.22876/bnc.2024.25.3.002

Kwon, S. H., Jeon, Y. S., Yoo, S. A., Oh, Y. R., & Lee, Y. M. (2022). A comparison of acoustic and perceptual measures of a female synthesized voice versus a real voice: A multi-dimensional analysis. *Journal of Speech-Language & Hearing Disorders, 31*(4), 81-90. doi:10.15724/jslhd.2022.31.4.081

Lee, H., Lee, J., & Kim, Y. (2016). The Lombard effect on the speech of children with intellectual disability. *Phonetics and Speech Sciences, 8*(4), 115-122. doi:10.13064/KSSS.2016.8.4.115

Lee, H. J. (2002). 'Utterance' as communicative minimal units and 'Sentence'. *Text Linguistics, 13,* 343-366. uci:G704-000916.2002.13..001

Lee, R., Han, J., & Lee, E. (2023). Development and applied research of Korean-Index of Phonetic Complexity-Revision (K-IPC-R). *Communication Sciences & Disorders*, *28*(3), 595-607. doi:10.12963/csd.23973

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., . . . Martinez, A. (2019). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science, 51*(4), 984-997. doi:10.1177/0961000618759414

Mayo, C., Aubanel, V., & Cooke, M. (2012). Effect of prosodic changes on speech intelligibility. *Proceedings of Interspeech 2012 ISCA's 13th Annual Conference,* 1708-1711.

Ministry of Science and ICT & Software Policy Institute. (2023). 2023 artificial intelligence survey: Statistical tables. Retrieved from https://spri.kr/download/23473

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103. doi:10.1111/0022-4537.00153

Oh, K. A., Nam, K. W., & Kim, S. J. (2014). The development of verbs through semantic categories in the spontaneous speech of preschoolers. *Journal of Speech-Language & Hearing Disorders, 23*(4), 63-72. doi:10.15724/jslhd.2014.23.4.006

Open AI. (2022). Introducing whisper. Retrieved from https://openai.com/index/whisper

Open AI. (2024). GPT-4o system card. Retrieved from

https://openai.com/index/gpt-4o-system-card

Park, H., & Lee, Y. (2023). Korean mothers' speech to young children with cochlear implants in parent-child interaction. *Communication Sciences & Disorders, 28*(4), 862-872. doi:10.12963/csd.231019

Park, H. J., Park, B. S., Song, B. D., Kwon, S. B., & Shin, B. J. (2020). The effects of home training for children with speech sound disorders using artificial intelligence speakers. *Proceedings of 21st Korean Speech-Language & Hearing Association,* 319-322.

Rajaraman, V. (2014). JohnMcCarthy—Father of artificial intelligence. *Resonance, 19,* 198-207. doi:10.1007/s12045-014-0027-9

Shin, J. Y. (2018). Breath and memory in speech based on quantitative analysis of breath groups and pause units in Korean. *Korean Linguistics, 79,* 91-116. doi:10.20405/kl.2018.05.79.91

Siegert, I., Nietzold, J., Heinemann, R., & Wendemuth, A. (2019). The restaurant booking corpus: Content-identical comparative human-human and human-computer simulated telephone conversations. *Proceedings of Konferenz Elektronische Sprachsignalverarbeitung,* 126-133.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion: A study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149,* 102601. doi:10.1016/j.ijhcs.2021.102601

Smith, R., Heuerman, M., Wilson, B. M., & Proctor, A. (2003). Analysis of normal discourse patterns. *Brain and Cognition, 53*(2), 368-371. doi:10.1016/S0278-2626(03)00145-3

Song, M.-S. (2022). A study on interactive talking companion doll robot system using big data for the elderly living alone. *The Journal of the Korea Contents Association, 22*(5), 305-318. doi:10.5392/JKCA.2022.22.05.305

Stocco, A., Prat, C. S., Losey, D. M., Cronin, J. A., Wu, J., Abernethy, J. A., & Rao, R. P. N. (2015). Playing 20 questions with the mind: Collaborative problem solving by humans using a brain-to-brain interface. *PloS One, 10*(9), e0137303. doi:10.1371/journal.pone.0137303

Yu, J. C. (2019). *Comparison of TTS (text-to-speech) devices' and human announcers' paralinguistic characteristics and audience perceptions about paralinguistic characteristics* (Doctoral dissertation). Sungkyunkwan University, Seoul.

언어치료연구 2025년 제34권 제2호 087-095
ISSN 1226-587X / eISSN 2671-7158
JSLHD (Journal of Speech-Language & Hearing Disorders)

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2025.34.2.087
http://jslhd.org/

인간-인간, 인간-대화형 AI 상호작용에서의 발화 특성 비교: 말속도, 음도, 포먼트를 중심으로

정다희[1], 윤여은[1], 이지연[1], 백소현[1], 이영미[2*]

[1] 이화여자대학교 일반대학원 언어병리학과 석사과정
[2] 이화여자대학교 일반대학원 언어병리학과 교수

목적: 본 연구는 인간-인간, 인간-대화형 AI 간 실제 상호작용 상황에서, 상호작용 대상에 따른 인간 발화의 음향학적 특성 차이를 비교하고자 하였다. 그리고 시간(말속도), 운율(피치), 조음(포먼트) 차원에서, 상호작용 시 나타나는 발화의 변화도 함께 살펴보고자 하였다.

방법: 20~30대 성인 총 23명(남성 11명, 여성 12명)이 인간과 대화형 AI(Chat GPT 4.0) 순으로, 동물을 주제로 한 스무고개 게임에 참여하였다. 실시간으로 질문과 대답을 주고받을 수 있는 3분간의 상호작용 동안 연구 참여자의 음성 샘플을 수집하였고 Praat 음성 분석 프로그램을 사용하여 말속도, 피치, 포먼트 특징을 분석하였다.

결과: 인간-인간, 인간-대화형 AI 간 상호작용에서 말속도와 포먼트에 있어 유의미한 차이가 나타났다. 즉, 대화형 AI와의 상호작용에서 인간과의 상호작용에 비해 더 빠른 말속도와 더 높은 포먼트값(F1값, F2값)을 보였다. 반면, 운율 관련 매개변수(피치의 평균과 범위)에서는 상호작용 간 유의미한 차이가 관찰되지 않았다.

결론: 본 연구는 인간과 대화형 AI와의 직접적인 상호작용에서 나타나는 특정 음성 패턴을 확인하였으며, 인간이 대화형 AI와 상호작용할 때 더 빠른 말속도를 보이는 것과 더 명확한 발음으로 말하는 경향이 있다는 점을 시사하였다. 이를 기반으로 본 연구는 음향학적 요소를 정교하게 조정할 수 있는 대화형 AI의 개발 가능성을 제시하는 기초 자료로 활용될 수 있으며, 향후 언어재활 분야에서의 활용 가능성을 제시하였다는 점에서 의의가 있다.

검색어: 인간-대화형 AI 상호작용, 말속도, 음도, 포먼트

교신저자 : 이영미(이화여자대학교)
전자메일 : youngmee@ewha.ac.kr

게재신청일 : 2025. 02. 12
수정제출일 : 2025. 04. 06
게재확정일 : 2025. 04. 30

ORCID
정다희
https://orcid.org/0009-0002-6673-7745
윤여은
https://orcid.org/0009-0002-0294-1771
이지연
https://orcid.org/0009-0009-5087-1107
백소현
https://orcid.org/0009-0000-0249-6693
이영미
https://orcid.org/0000-0003-1809-5944

참고문헌

ICT 통계포털 (2022). 2022 4차 산업혁명 지표 [인포그래픽]. https://url.kr/ebfgyo

강수미, 강은희, 이지윤 (2024). 학령기 지적장애 아동의 모음 음향학적 변수와 말 명료도 및 말 용인도 간의 상관관계. **언어치료연구, 33**(1), 91-100.

과학기술정보통신부, 소프트웨어정책연구소 (2023). 2023년 인공지능산업 실태조사. https://spri.kr/download/23473

권순하, 전예솔, 유성아, 오유림, 이영미 (2022). 다차원적 음성 프로파일 분석을 통한 20-30대 한국 여성의 합성음성과 실제음성의 특성 비교. **언어치료연구, 31**(4), 81-90.

권하나, 정정주 (2024). 대화형 생성 AI 의인화에 대한 다차원적 접근: 하슬람의 탈인간화 모델을 중심으로. **방송과 커뮤니케이션, 25**(3), 44-99.

김병운 (2016). 인공지능 동향분석과 국가차원 정책제언. **정보화정책, 23**(1), 74-93.

김유선, 김윤선, 이현진, 채민아 (2024). 대화형 AI 를 통한 자폐 아동 치료 플랫폼 개발. **한국정보처리학회 2024 학술대회논문집, 31**(1), 480-481.

박희선, 이영미 (2023). 인공와우이식 영유아 어머니의 아동지향어 특성. **Communication Sciences & Disorders, 28**(4), 862-872.

박희준, 박병석, 송복덕, 권순복, 신범주 (2020). 인공지능 스피커를 활용한 말소리장애 아동의 가정훈련 효과. **한국언어치료학회 2020년도 제21회 학술발표대회 논문집,** 319-322.

백승주, 정윤혁 (2022). AI 음성 에이전트의 음성 특성에 대한 사용자 반응 연구. **정보시스템연구, 31**(2), 137-158.

송문선 (2022). 빅데이터를 이용한 독거노인 돌봄 AI 대화형 말동무 아가야(AGAYA) 로봇 시스템에 관한 연구. **한국콘텐츠학회 논문지, 22**(5), 305-318.

신지영 (2018). 언어 수행에서의 호흡과 기억: 호흡 단위와 휴지 단위의 양적 분석 결과를 바탕으로. **한국어학, 79,** 91-116.

안병섭 (2007). 언어 분석 단위로서의 '발화' 설정 방법론 연구. **한국학연구, 26,** 233-259.

오경아, 남경완, 김수진 (2014). 학령전기 아동의 자발화에 나타난 동사의 의미 분류. **언어치료연구, 23**(4), 63-72.

유지철 (2019). **뉴스 낭독에 나타난 TTS(음성합성기)와 아나운서의 준언어 특성 비교 및 준언어에 대한 수용자 인식에 관한 연구.** 성균관대학교 대학원 박사학위 논문.

이란, 한진순, 이은주 (2023). 한국어 조음복잡성 지표-수정판(K-IPC-R)의 개발과 적용 연구. **Communication Sciences & Disorders, 28**(3), 595-607.

이현주, 이지윤, 김유경 (2016). 지적장애 아동의 롬바드 효과에 따른 말산출

특성. **말소리와 음성과학, 8**(4), 115-122.

이희자 (2002). '의사소통의 최소단위'로서의 '발화문'과 '문장'. **텍스트언어학, 13,** 343-366.

정상임 (2020). 인공지능 스피커를 사용한 자폐스펙트럼장애 아동의 대화 기술 중재 사례 연구. **한국언어치료학회 2020년도 제21회 학술발표대회 논문집,** 309-314.

최선아, 송영화 (2024). 기술기반 창업 활성화를 위한 기술융합 트렌드와 유망 기술 분석에 대한 연구: 대화형 AI를 중심으로. **한국창업학회지, 19**(2), 151-171.

최성 (2023). 클라우드 기반 대화형 AI 챗봇(기계학습: machine learning)을 통한 대학 이러닝학습센터 구축 연구. **한국정책학회 하계학술발표논문집,** 1-4.

한송희 (2023). 대화형 생성 AI(인공지능) Chat GPT를 활용한 한국어 말하기 연구. **학습자중심교과교육연구, 23**(18), 651-674.

황동준, 표승민, 김보애 (2020). 인공지능 기기를 통한 반복 읽기 중재가 언어학습장애 아동의 읽기 유창성에 미치는 영향. **한국언어치료학회 2020년도 제21회 학술발표대회 논문집,** 292-297.