JSLHD 2025. Vol. 34, No. 3. 001-011
ISSN 1226-587X / eISSN 2671-7158
Journal of Speech-Language & Hearing Disorders 언어치료연구

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2025.34.3.001
http://jslhd.org/

Exploring the Speech Transcription and Analysis Capabilities of Speech Recognition and Generative AI: Focusing on the Utterances of Older Adults

Min Guk Kang[1], So Bin Lee[1], Eun Byeol Cho[2], Duk L. NA[3], Ji Hye Yoon[4*]

[1] *Dept. of Speech Pathology and Audiology, Graduate School, Hallym University, Master's Student*
[2] *Happymind Clinic, Researcher*
[3] *Dept. of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Professor*
[4] *Div. of Speech Pathology and Audiology, Hallym University, Professor*

Purpose : This study aimed to explore the clinical applicability and usefulness of automatic transcription using speech recognition AI (Clova Note) and automatic analysis using generative AI (ChatGPT-4o) for the utterances of older adults.

Methods : Expository discourse was collected through a picture description task from 60 older adults aged between 50 and 90 years (18 middle-aged and 42 older adults). The transcription results were compared between human transcribers and Clova Note based on transcription accuracy rates, and the utterance analyses were compared between human analysis and ChatGPT-4o (1st and 4th attempts) across six analytic measures (total number of utterances, total number of words, total number of syllables, disfluency frequency, maximum sentence length, number of CIU).

Results : First, Clova Note showed an average transcription accuracy rate of 80.60% for middle-aged adults and 78.92% for older adults, indicating a usable level of automatic transcription performance. However, it tended to omit or automatically correct disfluencies (e.g., repetitions, interjections) and misarticulations, making human review and revision essential for accurate transcription. Second, ChatGPT-4o showed statistically comparable results to human analysis in fluency-related indicators across both age groups. However, a detailed examination of the automatic analysis results revealed differences from the human analysis, in utterance classification and in counting words and syllables. Third, in the CIU analysis, ChatGPT-4o showed comparable results to human analysis in the middle-aged group when revision prompting (4th attempts) was provided. However, in the older adults group, no improvement was observed even after revision prompting.

Conclusions : These results show the potential clinical applicability and limitations of speech recognition and generative AI technologies in speech language pathology, and suggest the need for further research to enhance their clinical usefulness.

Keywords : Aritificial intelligence, automatic speech recognition, generative AI

Correspondence : Ji Hye Yoon, PhD
E-mail : j.yoon@hallym.ac.kr

Received : May 30, 2025
Revision revised : July 11, 2025
Accepted : July 31, 2025

This research was supported by Hallym University Research Fund, 2025 (HRF-202504-004).

ORCID
Min Guk Kang
https://orcid.org/0009-0004-3996-8104
So Bin Lee
https://orcid.org/0009-0009-9491-6899
Eun Byeol Cho
https://orcid.org/0000-0002-0327-7035
Duk L. NA
https://orcid.org/0000-0003-1572-7862
Ji Hye Yoon
https://orcid.org/0000-0003-1403-2276

Ⅰ. 서론

인공지능(artificial intelligence: AI)의 급속한 발전은 다양한 학문 분야와의 융합을 통해 기술적 혁신을 이끌고 있으며(Cho, 2023; Han, 2023; Hwang & Kim, 2019; Kwon, 2025), 언어병리학 분야에서도 인공지능 기술을 평가, 진단, 중재 등의 목적으로 활용하려는 시도가 확산되고 있다(Kang et al., 2022; Kang et al., 2022). 언어병리 임상에서의 평가 과정은 특정 장애의 유무를 판단하고 치료의 방향을 설정하는 중요한 임상적 절차이다(Sim et al., 2019). 평가 시 자연스러운 맥락에서 발화를 수집하는 것은 진단이나 중재 과정에서 대상자의 언어능력을 확인하기 위하여 빈번히 활용되지만, 수집된 발화 자료를 검사자가 수동으로 전사 및 분석하는 과정에는 많은 시간과 노력이 요구된다는 어려움이 있다(Chung, 2013; Pavelko et al., 2016; Yang et al., 2023). 국내 언어병리 분야 현직 종사자 및 학위과정생을 대상으로 생성형 AI의 활용 필요성과 인식을 조사한 선행 연구(Lee & Yoon, 2025)에 따르면, 과반수 이상의 응답자는 향후 임상 및 연구를 목적으로 생성형 AI를 활용할 의향이 있고(62.2%), 생성형 AI의 활용이 업무 또는 학업 속도 및 효율성 향상에 기여할 것이라는 인식(75.7%)을 보였다. 이는 생성형 AI를 통한 효율성 증대와 시간 절약에 대한 임상 현장의 수요가 높음을 보여준다.

AI는 응용 분야에 따라 여러 유형으로 분류되는데, 언어병리학 분야에서 발화 분석을 위해 사용할 수 있는 주요 인공지능 기술로

는 음성인식 AI와 생성형 AI가 있다. 자동음성인식(automatic speech recognition: ASR)이란 프로그램이 사람의 음성을 인식하여 문자(텍스트) 형식으로 변환하는 기술로, 흔히 STT(speech-to-text) 기술이라고도 알려져있다(Huh et al., 2023). 대표적인 음성인식 AI로는 Whisper, Google STT, Clova Note 등이 있으며, 이 중 클로바노트(Clova Note)는 한국어 음성인식에 특화된 대표적인 국내 개발 AI이다. 음성인식 기술은 언어병리 임상 현장에서 언어능력을 평가하기 위해 대상자의 발화를 전사할 때 유용하며, 현재 ASR을 전사에 활용한 국내외 연구들이 활발히 이루어지고 있다(Yang et al., 2023). 학령기 일반 아동과 언어학습장애 아동의 발화를 대상으로 클로바노트와 Google STT의 전사 오류율을 확인한 국내 선행 연구(Yang et al., 2023)에서는 두 ASR 모두 낮은 전사 오류율과 빠른 처리 속도를 보여 임상적 유용성을 확인하였다.

생성형 AI(generative AI)는 사용자의 명령어에 따라 새로운 텍스트, 이미지, 음성 등을 생성하는 데에 특화된 인공지능을 의미한다(Cho, 2023; Han, 2023). 그중에서도 입력과 출력이 모두 텍스트 형식으로 이루어지는 모델을 text-to-text 모델이라고 하며, 대표적으로 OpenAI에서 제공하는 ChatGPT(chat generative pre-trained transformer)가 이에 해당한다(Han et al., 2025). 생성형 AI는 사람의 언어, 즉 자연어(예, 한국어, 영어 등)를 이해하고 분석하거나 생성할 수 있는 자연어 처리(natural language processing) 기능을 포함하고 있다는 점에서 언어분석 과정에 유용하게 활용될 가능성이 있다. AI 모델에 제공되는 질문, 지시문, 설명 등의 텍스트를 프롬프트(prompt)라고 하며, 이러한 프롬프트를 활용하여 AI 모델과 상호작용하거나 원하는 결과를 얻어내는 일련의 과정을 프롬프팅(prompting)이라고 한다(Han & Lee, 2024). 언어병리 분야에서 생성형 AI의 응용 및 활용 현황을 보고한 연구(Lee & Yoon, 2025)에 따르면, ChatGPT는 생성형 AI 중 가장 높은 인지도를 보였으나, 인공지능 전반에 대한 개념이나 활용 방법에 대한 교육 부족으로 인해 실제 임상 및 연구 현장에서 응답자의 과반수 이상(약 70%)은 생성형 AI 활용 경험이 없는 것으로 나타났다. 그러나 다양한 유형(텍스트, 이미지, 음성, 영상 등)의 콘텐츠 생성이 가능하다는 특징으로 인해(Du & Juefei-Xu, 2023) 생성형 AI가 치료 자료 제작과 같은 보조적 수단으로서 보다 다채로운 재활 활동 구성에 기여할 수 있을 것으로 기대하고 있었다.

음성인식 AI와 생성형 AI는 웹사이트 또는 애플리케이션 프로그래밍 인터페이스(application programming interface: API)를 통해 이용할 수 있다. 웹사이트 이용은 공식 웹사이트에 접속하여 AI를 사용하는 방식을 의미한다. 클로바노트는 직접 음성 파일을 업로드하면 바로 자동전사를 실시할 수 있고, ChatGPT의 경우 질문이나 명령어를 입력하면 대화 형식으로 답변을 받을 수 있다. 이는 대량의 데이터를 처리하거나 반복 작업(예, 데이터 수동 입력 등)을 자동화하는 것이 어렵다는 단점이 있지만, 인터넷만 있다면 바로 AI 서비스를 이용 가능하기 때문에 별도의 개발 지식 없이도 사용자가 손쉽게 접근할 수 있다는 장점이 있다. API는 특정 프로그램(또는 서비스)이 다른 프로그램과 상호작용할 수 있도록 설계된 일련의 규칙이나 도구의 집합을 의미하며, 서로 다른 프로그램을 연동시키는 방식을 통해 필요한 기능이나 데이터를 요청하고 응답할 수 있도록 하는 메커니즘이다. API 이용은 앱, 시스템 또는 자체 프로그램 등에 AI 기능을 통합하여 사용하는 방식으로 이루어진다. 예를 들어, 기상청의 소프트웨어 시스템에 저장된 일일 기상 데이터를 휴대폰의 날씨 애플리케이션이 API를 통해 요청하면, 기상청 시스템은 해당 데이터를 응답하여 사용자의 휴대폰에 최신 날씨 정보를 표시할 수 있다. API 이용은 대규모 데이터 처리나 시스템 연동, 반복 작업의 자동화 등의 측면에서 장점을 가진다. 언어병리 분야에서는 ASR 프로그램과 ChatGPT를 API로 연동하면, 스마트폰 앱에서 녹음한 대상자의 음성 파일을 자동으로 전사하고, 전사된 텍스트를 기반으로 분석, 요약, 분류 등의 작업을 자동 수행하는 통합 시스템을 구현할 수 있다.

다만, API를 사용하기 위해서는 Python과 같은 프로그래밍 언어, 코드 작성 방식 등의 기본적인 프로그래밍 지식을 이해하는 것이 필요하다. 최근에는 예제 코드나 사용자 친화적 인터페이스가 함께 제공되어 초보자도 일정 수준까지 API를 활용 가능하지만, 여전히 익숙하지 않은 사용자에게는 진입 장벽이 존재한다. 따라서 단순한 소규모 작업이나 임상 현장에서의 기본적인 평가 목적이라면 웹사이트를 통한 이용 방식이 더욱 간편하고 적합하다. 기존의 인공지능을 활용한 연구들은 대부분 개발자 입장에서 프로그램 개발을 목적으로 이루어졌지만, 최근에는 클로바노트나 ChatGPT처럼 단순한 조작으로 이용 가능한 간편한 AI 기술을 연구에 활용하는 방식이 늘고 있다. 그럼에도 불구하고 언어병리 임상 및 연구 분야의 많은 사용자들이 'AI 기술을 사용하고 싶지만 방법을 몰라서' 활용하지 못하는 실태(Lee & Yoon, 2025)를 고려할 때, 현 시점에서 AI 전문가가 아니라도 쉽게 접근 가능한 음성인식 및 생성형 AI를 발화 전사 및 분석에 적용할 수 있을지 탐색하는 연구가 필요하다. 만약 이러한 단순한 활용 방식을 통해서도 정확하고 유의미한 결과를 얻을 수 있다면, 실제 언어병리 임상 및 연구 현장에서 보다 폭넓고 효율적으로 AI 기술을 도입할 수 있는 가능성을 보여줄 수 있을 것이다.

과제 및 분석 지표의 측면에서, 성인 및 노인을 대상으로 한 발화 분석 선행 연구는 대부분 그림 설명 과제를 활용하여 발화를 수집하였다(Kim et al., 1998; Kim et al., 2006; Kwon et al., 1998; Lee & Kim, 2001). 이는 그림 설명 과제가 동일한 시각적 자극에 대한 수행력을 비교적 객관적으로 측정할 수 있다는 장점을 가지기 때문이다(Lee & Kim, 2001). 분석에 사용된 지표들을 살펴보면, 정보전달의 효율성을 확인하기 위하여 '정확한 정보 단위(correct information unit: CIU)' 분석이 실시되었다(Kim et al., 2006; Kwon et al., 1998). CIU란 Nicholas와 Brookshire(1993)에 의해 도입된 개념으로, 전체 발화 중 주제에 적합하지 않거나, 무관하거나, 뜻이 명료하지 않은 낱말을 제외한 단위를 의미한다. 따라서 CIU는 내용 전달 측면에 대한 양적 또는 질적 분석을 가능하게 한다. 발화의 유창성 측면을 평가하기 위해서는 연구마다 다양한 지표들이 사용되었는데, 정상 성인을 대상으로 한 연구의 경우, 발화 속도와 머뭇거림의 정도를 확인 가능한 "분당 음절 수, 분당 낱

말 수, 분당 CIU 수, CIU 비율, 분당 머뭇거림 수(간투사, 반복, 수정)" 등을 분석하거나(Kwon et al., 1998), "발화당 음절 수, 발화당 단어 수, 간투사, 총 발화 수에 대한 문장 비율" 등을 포함한 총 21가지 변인을 통해 정상 성인의 발화 특성을 확인하고자 하였다(Lee & Kim, 2001). Kim 등(2006)에서는 총 12가지 변인(발화당 음절 수, 발화당 단어 수, 발화당 내용어 수, 초당 음절 수, 음소착어, 의미착어, 간투사, 수정, 반복, 도치, 후속발화 개시시간, CIU 비율)을 통해 알츠하이머병(Alzheimer's disease: AD) 치매 환자와 정상 노인의 발화 특성의 기본적인 양적 수치들을 비교하고자 하였다. 그러나 이러한 양적 지표들을 사람이 수기로 분석하려면 많은 시간과 노력이 필요하다는 어려움이 있다. 만약 AI 기술을 활용하여 양적인 지표에 대한 분석 및 개수 산정 작업을 자동화한다면, 분석 작업의 효율성을 크게 높일 수 있으며, 임상 및 연구 환경에서 더 많은 데이터를 보다 신속하게 처리하고 해석할 수 있는 기반을 마련할 수 있다.

연구 대상자의 측면에서는, 우리 사회가 초고령사회(super-aged society)로 진입하게 되면서(Statistics Korea, 2025) 언어병리학적 평가 및 중재가 필요한 장노년층 대상자는 앞으로 더욱 늘어날 전망이다. 이러한 사회적 변화는 노년층에서 발생하는 퇴행성 질환의 증상 중 하나인 치매의 증가로 이어질 수 있는데, 보건복지부(2025)의 보고에 따르면, 2023년 기준 65세 이상 노인의 치매 유병률은 9.25%, 경도인지장애(mild cognitive impairment: MCI) 유병률은 28.42%에 달하며, 향후 치매 환자 수는 2025년에는 97만명(치매 유병률 9.17%), 2026년에는 100만명, 2044년 이후에는 200만 명을 초과할 것으로 전망된다. 이처럼 고령층에서의 인지 및 언어기능 저하가 상당히 높은 비율로 발생하고 있으며, 기능 저하가 조기에 식별되지 않은 채 치매로 진행되는 경우도 많다. 선행 연구(Kim et al., 2019)에서 언어능력의 저하를 자각하고 있음에도 병원 내원 이력이 없고 일상생활 기능에 뚜렷한 어려움이 없는 정상 노인 27명을 대상으로 신경심리검사를 실시한 결과, 25명(92.6%)이 MCI 범주에 속하며, 그중 과반수 이상(54%)은 치매 고위험군에 해당하는 것으로 나타났다. 이는 연구 대상자가 되는 노년층에서 정상임을 가장한 인지저하 환자가 상당히 많이 포함되어 있을 가능성을 시사한다.

종합하면, 사회적 변화에 따라 정상이나 질환이 있는 장노년층의 발화를 전사하고 분석할 기회가 확대되면서 해당 집단에 대한 인공지능의 활용 가능성에 대한 탐색이 필요한 실정이다. 특히 기존의 API를 통한 전문가용 AI가 아닌 무료로도 쉽게 이용 가능한 AI를 사용한 연구를 통해 임상에서 활용 가능한 수준의 성능을 입증할 수 있다면, 언어병리 분야에서의 AI 활용도를 높이는 데에 기여할 수 있을 것으로 생각된다. 이에, 본 연구는 대중적이고 접근성이 높은 대표적인 인공지능을 활용하여, 음성인식 AI(클로바노트)를 이용한 자동전사 결과와 생성형 AI(ChatGPT-4o)를 이용한 자동분석 결과를 사람의 수기전사 및 수기분석 결과와 비교함으로써, 발화 전사 및 분석 과정에서 AI 기술의 임상적 활용 가능성과 유용성을 탐색하고자 한다.

II. 연구 방법

1. 연구 대상

본 연구는 만 50~90세 사이의 성인 총 60명을 대상으로 하였고, 보건복지법에 근거하여 만 65세를 기준으로 만 65세 미만의 장년층 18명(남 5명, 여 13명)과 만 65세 이상의 노년층 42명(남 18명, 여 24명)으로 구분하였다. 모든 대상자는 1급 언어재활사(제3저자)의 판정상 조음이나 공명 등에 영향을 미치는 말운동장애 소견이 없는 것으로 확인되었다. 또한, 한국판 단축형 노인우울척도(Korean version of the Geriatric Depression Scale-Short-form: SGDS-K, Kee, 1996) 점수는 8점 미만에 해당되어 우울감이 없는 것으로 확인되었다. 본 연구의 대상자 정보는 Table 1에 제시하였다. 두 집단 간에는 연령(p<.001)을 제외한 학력, 한국판 간이정신상태검사(Korean-Mini Mental State Examination: K-MMSE, Kang et al., 1997) 점수, 치매임상평가척도(Clinical Dementia Rating: CDR), SGDS-K 모두에서 두 집단 간 유의한 차이가 없었다. 추가적으로 시행된 서울신경심리검사(Seoul Neuropsychological Screening Battery-second edition: SNSB-II, Kang et al., 2012) 및 CDR에서 2점 이하에 해당하여 경도 또는 중등도 수준의 치매 단계에 해당되는 대상자는 장년층 집단에서 9명, 노년층 집단에서 21명으로 확인되었다.

Table 1. Participants' information

	MA (n=18)	OA (n=42)	Total (n=60)	p
Age (years)	59.50 (3.60)	75.19 (6.32)	70.72 (9.33)	MA<OA***
Education (years)	11.39 (5.36)	10.60 (5.40)	10.93 (5.30)	
K-MMSE	20.06 (7.47)	22.35 (6.08)	21.65 (6.52)	
CDR	1.00 (.68)	.74 (.39)	.81 (.50)	
SGDS-K	3.76 (1.60)	3.38 (1.21)	3.49 (1.38)	

Note. Values are presented as mean (*SD*).
MA=middle-aged adults; OA=older adults; K-MMSE=Korean-Mini Mental State Examination; CDR=Clinical Dementia Rating Scale; SGDS-K=Korean version of the Geriatric Depression Scale-Short-form.
***p<.001

2. 연구 절차

1) 발화 수집

대상자의 발화는 소음이나 타인의 개입이 없는 조용하고 독립된 공간에서, 검사자(제3저자)와 대상자 간 일대일 상호작용을 통해 그림 설명 과제를 수행하는 과정에서 수집되었다. 검사자는 흑백 선화 그림을 태블릿(Galaxy Tab S7 FE)화면에 제시한 후 대상자가 그림에 대해 설명하도록 하였다. 해당 그림에는 가족들이 함께 야외로 소풍을 나온 배경에서 자전거 타는 손자를

도와주는 할아버지, 손녀에게 책을 읽어주는 할머니, 고기 굽는 아버지, 음식을 준비하는 어머니 등 여러 인물과 사물이 포함되었다. 그림 제시와 발화 녹음은 태블릿을 사용하여 동시에 진행되었다.

2) 발화 전사

발화 전사는 사람의 수기전사와 음성전사 AI의 자동전사로 나뉘어 실시되었다. 사람은 녹음된 음성파일을 직접 듣고 전사하였으며, 동일한 음성파일 원본을 클로바노트 웹사이트에 업로드하여 자동전사를 실시하였다. 수기전사에는 총 2명의 평가자(제1, 2저자)가 전사에 참여하였으며, 각자 독립된 환경에서 전사를 시행한 후 두 저자 간 서로의 전사 자료를 검토하고, 차이가 있는 경우 최종 합의를 통해 사람의 수기전사 결과로 정리하였다. 이후 자동전사 결과와 사람의 전사 결과를 비교하여 전사일치율을 산출하였다. 전사일치율은 음절 단위를 기준으로 (클로바노트가 사람과 동일하게 전사한 음절 수)/(총 음절 수)×100의 방식으로 계산하였다.

3) 발화 분석

발화 분석은 사람의 수기분석과 ChatGPT의 자동분석으로 나뉘어 실시되었다. 분석 지표들은 관련 선행 연구(Kim et al., 1998, Kim et al., 2006; Kwon et al., 1998; Lee & Kim, 2001)를 참고하여 선정하였다. 발화 분석 지표는 총 6가지이며, 이 중 5개는 발화의 유창성 측면을 평가하기 위한 지표(총 발화 수, 총 어절 수, 총 음절 수, 유창성 방해요소 빈도 [간투사, 수정, 반복 등], 최대 문장길이)이고, 나머지 1개는 발화의 내용 전달 측면을 평가하기 위해 활용되는 CIU 개수이다. 총 발화 수 계산을 위한 발화 분류는 Kim(1997)에서 제시한 발화 구분 원칙에 근거하여 실시하였다.

사람이 실시한 발화 수기분석은 다음과 같이 진행되었다. 총 2명의 평가자(제1, 2저자)가 분석에 참여하였으며, 발화 분석 지표를 계산하기 전 함께 발화 분류 기준을 숙지한 뒤 각자 독립된 환경에서 기준에 근거한 발화 분류를 실시하였다. 이때 사전에 사람이 음성파일을 듣고 수기전사한 결과를 기반으로 발화 분류가 이루어졌다. 그 후 두 평가자는 서로의 발화 분류 내용을 검토하고, 차이가 있는 경우 합의를 통해 결과를 통일하는 과정을 거쳤다. 발화 분류가 마무리된 후, 발화 분석 지표 계산이 진행되었다. 사전에 분석 지표 6개의 개념 및 계산 방법을 숙지하는 과정을 거친 후 각자 독립된 환경에서 분석을 진행하였다. 각자 분석 지표 계산을 완료한 후 서로의 분석 결과를 상호 검토 및 수정하는 과정을 거쳤다. 최종적으로 합의된 발화 분류 및 지표 분석 결과는 사람의 수기분석 결과로 도출되었다.

ChatGPT-4o를 활용한 발화 자동분석은 다음과 같은 절차로 진행되었다. 먼저, 자동분석을 위한 사전 학습 단계로서, 분석 지표의 개념 및 채점 방법을 학습시키는 과정을 거쳤다. 이 과정은 유창성 지표 분석과 CIU 분석 간 필요한 학습 내용의 차이를 고려하여 서로 독립된 대화창에서 이루어졌다. 유창성 지표 분석의 경우, 발화 구분 원칙(Kim, 1997)을 텍스트 형태로 제공한 뒤 "발화 분류는 Kim(1997)의 발화 구분 원칙에 근거하여 실시할 예정입니다. 지금 제공한 원칙을 잘 확인해주세요"의 프롬프트를 함께 제공하였다. 이어서 5개 분석 지표의 개념 및 채점 방법 등을 학습시키는 과정을 거쳤다(예, '최대 문장길이'는 대상자의 발화 중 최대 어절 길이로 이루어진 문장이 총 몇 개의 어절로 구성되어 있는지 확인하는 지표입니다. 발화 구분 원칙에 따라 분류된 발화 중 가장 많은 어절을 포함하는 문장의 어절 수를 계산해주세요). CIU 분석에서도 동일하게 CIU의 정의 및 채점 방법 등을 학습시킨 뒤, 그림설명 과제에서 사용된 그림을 ChatGPT-4o에 업로드하고, 해당 그림에서 추출 가능한 CIU 목록을 스스로 도출하도록 하는 과정을 거쳤다(예, 그림설명과제에서 사용한 그림 이미지입니다. 제시한 그림에서 표현 가능한 CIU 목록을 모두 정리해주세요).

위와 같이 각 사전 학습 단계를 마친 후, ChatGPT-4o가 발화 구분 및 분석을 자동적으로 수행하게 하였다. 자동분석을 위해 필요한 대상자의 발화 원본은 연구자가 수기전사한 결과를 기반으로 텍스트 형식으로 제공되었다. 이때 분석을 지시하는 프롬프트(예, 대상자 발화를 확인하고, 유창성 지표 5가지를 분석해주세요)를 제시하고 도출된 최초 분석 결과를 '자동분석 1차 시도'로 정의하였다. 만약 자동분석 결과에서 오류가 확인될 경우, 오류를 스스로 인지하고 자체적으로 수정 분석을 시행할 수 있도록 추가적인 수정 프롬프팅을 제공하였다. 수정 프롬프팅은 Appendix 1에 제시한 예시를 기반으로 ChatGPT-4o가 제시한 분석 결과의 오류 유형에 따라 선별적으로 제공되었으며, 최대 3회까지 추가적으로 제공한 후 최종 수정된 결과를 확인하였다. 이 결과를 '자동분석 4차 시도' 결과로 지칭하여, 수정 프롬프팅 제공에 따른 결과 향상 여부를 확인하기 위해 결과 비교에 포함하였다.

3. 통계 분석

본 연구의 통계 분석은 SPSS 29.0(Statistical Package for the Social Science, Version 29.0) 프로그램(IBM Corp., Armonk, NY)로 실시하였다. 사람의 발화 전사 결과에 대한 클로바노트의 전사일치율을 비교하기 위해 집단별로 독립표본 t-검정을 실시하였다. 또한, 사람과 ChatGPT-4o의 1차 시도, 4차 시도 결과 간 차이를 확인하기 위해 집단별로 반복측정 분산분석(repeated measures ANOVA)을 실시하고, 통계적으로 유의한 차이가 있는 경우 Scheffe 사후검정을 실시하였다.

Ⅲ. 연구 결과

1. 사람과 클로바노트의 전사일치율 비교

장년층 발화에 대한 사람과 클로바노트의 전사일치율은 80.60%(8.52)였고, 노년층 발화에 대한 사람과의 전사일치율은 78.92%(12.51)로 나타났다. 장년층 집단과 노년층 집단에서 산출된 전사일치율은 유의한 차이가 없었다(p=.606).

2. 사람과 ChatGPT-4o 간 발화 분석 결과 비교

발화 분석에 대한 검사자(사람)의 수기분석 결과와 ChatGPT-

4o의 1차 및 4차 시도 자동분석 결과를 장년층(Table 2)과 노년층(Table 3)에서 각각 비교하였다. 그 결과, 두 집단 모두에서 6가지 분석 지표 중 ChatGPT-4o의 1차 시도와 4차 시도가 사람의 수기분석과 유의한 차이가 없었던 항목은 5항목(총 발화 수, 총 어절 수, 총 음절 수, 유창성 방해요소 빈도, 최대 문장길이)이었다. 그러나 CIU 개수에 대한 자동분석 결과에서는 장년층 발화의 경우 1차 시도에서 사람의 수기분석 결과와 유의한 차이를 보였으나, 수정 프롬프팅을 제시한 4차 시도에서는 사람과 차이가 없었다. 반면, 노년층의 경우 1차 시도와 4차 시도 모두 사람의 수기분석 결과와 통계적으로 유의한 차이를 보였다.

Table 2. Comparison of middle-aged adults' discourse analysis results between human and GPT

Variable	Human	GPT_1 (Trial 1)	GPT_4 (Trial 4)	*F*	post-hoc
Total number of utterances	11.72 (3.64)	10.89 (3.58)	11.78 (4.44)	.887	
Total number of words	52.67 (17.63)	48.22 (16.99)	49.39 (17.82)	1.041	
Total number of syllables	135.67 (47.89)	134.72 (61.68)	135.39 (62.78)	.003	
Disfluency frequency	12.06 (11.30)	8.56 (3.52)	8.22 (3.26)	1.558	
Maximum sentence length	9.11 (2.95)	8.33 (1.65)	8.61 (1.98)	.821	
Number of CIU	20.89 (9.01)	18.78 (7.64)	20.28 (8.55)	8.647***	H > GPT_1* GPT_1 > GPT_4*

Note. Values are presented as mean (*SD*).
CIU=correct information unit; H=human.
*p<.05, ***p<.001

Table 3. Comparison of older adults' discourse analysis results between human and GPT

Variable	Human	GPT_1 (Trial 1)	GPT_4 (Trial 4)	*F*	post-hoc
Total number of utterances	13.90 (8.42)	11.43 (5.31)	12.60 (6.81)	1.678	
Total number of words	60.52 (41.99)	50.00 (20.39)	57.88 (49.17)	.803	
Total number of syllables	152.64 (104.05)	121.02 (49.05)	124.14 (55.64)	2.567	
Disfluency frequency	13.36 (14.93)	12.40 (6.88)	12.81 (7.35)	.113	
Maximum sentence length	9.12 (3.68)	8.20 (2.26)	8.07 (2.25)	2.086	
Number of CIU	22.12 (10.03)	19.88 (8.61)	21.31 (9.28)	13.702***	H > GPT_1*** H > GPT_4* GPT_1 > GPT_4***

Note. Values are presented as mean (*SD*).
CIU=correct information unit; H=human.
*p<.05, ***p<.001

Ⅳ. 논의 및 결론

본 연구는 장노년층의 발화를 바탕으로 음성인식 AI인 클로바노트를 발화 자동전사에 활용하여 사람의 수기분석 결과와의 전사일치율을 확인하고, 생성형 AI인 ChatGPT-4o를 발화 자동분석에 활용하여 사람의 수기분석 결과와 수정 프롬프팅 제공에 따른 자동분석 결과를 비교하였다.

먼저, 사람과 클로바노트 간 전사일치율을 확인한 결과, 장년층 발화에 대한 평균 전사일치율은 80.60%였고, 노년층 발화에 대한 평균 전사일치율은 78.92%로 나타났으며, 집단에 따른 전사일치율에서 유의한 차이가 확인되지 않아, 클로바노트가 연령층과 관계없이 비슷한 수준의 전사 성능을 보이는 것으로 나타났다. 이러한 일치율은 네이버의 음성인식 API 서비스인 클로바 음성인식(Clova speech recognition: CSR)의 음성인식률을 확인한 선행연구에서 보고된 결과(85.99%, 77.51%)와 유사하다(Choi et al., 2020; Yoo et al., 2020). 본 연구에서 활용된 클로바노트는 웹사이트 이용 기반 기술이므로 이러한 결과는 음성 전사를 위한 AI의 사용에서 웹사이트 이용과 API 이용 기술 간의 정확도가 유사함을 보여준다.

클로바노트의 자동전사 결과를 질적으로 확인하였을 때, 다음의 두 가지 측면에서 정확하지 않은 전사 능력을 보이는 것이 나타났다. 첫째, 언어적 비유창성 요소 대부분을 올바르게 전사하지 않는 경우가 빈번하였다. 간투사(예, 음, 어, 뭐 등), 삽입어(예, 이제 등) 또는 반복 표현을 일부 생략하거나 전사하지 않았으며(예, 뭐, 고, 고, 고, 공원에서→공원에서), 다른 간투사나 단어로 대치하여 전사되는 경우(예, 엄마는, 어, 식사 준비하고→엄마는 뭐 식사 준비하고 / 그, 개가 싸우는거고→이건 개가 싸우는거고)가 확인되었다. 둘째, 대상자의 발화를 들리는 소리 그대로 전사하지 않고 자동으로 음소를 수정하는 오류를 보였다. 오조음을 정조음으로 변환하거나(예, 돼지→돼지, 캐→개), 구어적으로 허용되는 표현을 변환하거나(예, 인제→이제, 케일→케이크), 또는 사투리를 표준어로 변환하는(예, 어무니→어머니, 인자[이제]→이제) 양상을 보였다. 이러한 전사 오류 유형은 학령기 일반 아동과 언어학습장애 아동을 대상으로 CSR을 사용한 선행 연구(Yang et al., 2023)에서 보고된 결과 내용과 일치한다. 일반적인 상황에서 클로바노트의 사용 용도는 일상에서 대화나 회의 내용 등을 빠르게 기록하는 것에 최적화되어 있기 때문에 이러한 자동 수정 기능이 유용할 수 있다. 반면에, 언어병리학 분야에서는 오조음이나 비유창성 요소 등이 언어능력 평가에서 핵심적인 분석 항목이 되기 때문에, 대상자의 발화를 청지각적으로 들리는 그대로 전사하는 것이 중요하며 자동 수정 기능이 분석의 정확성을 저해할 수 있다. 이러한 제한점에도 불구하고, 임상 현장에서 대상자의 발화를 처음부터 청지각적으로만 듣고 수기전사를 실시하는 것보다, 클로바노트를 활용해 시각적 전사 자료를 제공받고 이를 사람이 수정하는 방식이 시간과 노력을 더욱 절약할 수 있게 한다(Scott et al., 2022; Yang et al., 2023). 예비 연구(Kang et

al., 2024)를 통해 사람과 3가지 음성인식 AI의 전사 소요시간을 비교한 결과, 사람은 400음절의 읽기 음성 파일에 대해 약 8~13분 정도의 전사 시간이 소요된 반면, 음성인식 AI는 모두 1분 이내(최소 약 5~최대 51초)의 시간이 소요되었다. 클로바노트는 이용 시간이 한 달간 최대 600분으로 제한된다는 제약을 가지나, 임상 현장에서의 발화 전사가 주로 평가 시 수집된 대상자의 발화를 분석하기 위해 시행되는 점과 전사 결과를 온전히 신뢰하기에는 클로바노트의 성능이 아직 제한점이 있음을 고려할 때, 상대적으로 길고 오랜 시간 산출된 발화에 대해서 시간 절약을 목적으로 사용한다면 충분히 유용하게 활용 가능할 것으로 사료된다. 즉, 클로바노트를 발화 전사를 위한 보조 도구로 활용한다면, 전사 작업의 효율성을 높이는 동시에, 사람의 청지각적 판단을 보완하는 수단으로서 임상 및 연구 현장에서 유용하게 사용될 수 있을 것이다.

다음으로 사람의 수기분석과 ChatGPT-4o의 1차, 4차 자동분석 결과를 비교하였다. 그 결과, 6개의 분석 지표 중 발화의 유창성 측면을 평가하기 위한 지표 5가지(총 발화 수, 총 어절 수, 총 음절 수, 유창성 방해요소 빈도, 최대 문장길이)에 대한 ChatGPT-4o의 최초 분석(1차 시도) 및 최종 분석(4차 시도) 결과는 연령층에 관계없이 사람과 유사한 수준의 능력을 보이는 것으로 나타났다. 그러나 본 연구에서는 발화 분석 지표의 빈도를 기준으로 사람과 ChatGPT-4o 간 결과를 양적으로만 비교하였기 때문에, 실제 자동분석 내용을 질적으로 확인해보면 사람의 수기분석 결과와 비교하여 오류나 누락이 존재하는 경우가 다소 있었다. 예를 들면, ChatGPT-4o가 분류한 발화 목록을 확인했을 때 1차와 4차 시도 결과 모두 수기분석을 통해 분류한 발화 목록과 동일하지 않은 경우가 빈번하였다. 이로 인해 측정을 위해 정확한 발화 분류가 필요한 지표(총 발화 수, 최대 문장길이)에 대한 자동분석 결과도 양적으로는 수기분석 결과와 동일하나 세부 분석 내용을 확인해보면 차이가 있는 경우가 있었다. 이는 ChatGPT-4o가 텍스트 형식으로 제공된 대상자의 발화를 '문장 단위와 문맥, 주제 및 의미 흐름'을 기준으로 분류하기 때문으로 해석된다. 그러나 언어병리학 분야에서의 발화 단위는 문장뿐만 아니라 단어, 구 등 다양한 형태이다. 따라서 이러한 분류를 위해서는 대상자의 발화 사이 쉼(pause)이나 억양의 변화 등을 고려해야 한다. 하지만 현시점에서 ChatGPT-4o가 음성 파일(.mp3, .wav 등)을 직접 인식하거나 분석을 수행하는 데에는 어려움이 있다. 즉, 호흡이나 운율 등의 청지각적 요소를 고려한 분석을 실시할 수 없기 때문에 사람의 발화 분류 결과와 일치하지 않는 경우가 발생한 것으로 볼 수 있다.

ChatGPT-4o는 어절과 음절의 개수 계산에서도 반복적으로 오류를 보였다. 이러한 오류의 원인에 대해서는 다음의 두 가지 가능성을 고려해 볼 수 있다. 첫째, 선행된 프롬프팅의 내용이나 반복 제공 방식이 후속 분석에 영향을 미쳤을 가능성이 있다. 생성형 AI는 입력된 프롬프트에 민감하게 반응하며, 동일한 의미라도 표현 방식이나 구조에 따라 상이한 결과를 도출할 수 있다(Chun, 2025). 특히 분석을 반복적으로 진행하는 과정에서 프롬프트를 지나치게 구체화하거나 일관되지 않게 제공하는 경우, 이전 정보와의 혼합이 발생하여 오히려 결과의 정확성이 저하될 수 있다(Han et al., 2025). 둘째, ChatGPT의 언어 처리 방식에서 기인한 오류일 가능성이 있다. ChatGPT는 입력된 문장을 '토큰(token)'이라는 단위를 기준으로 처리한다. 토큰은 단어의 일부나 전체, 혹은 두 단어가 결합된 형태로 인식되기도 하는 등 불규칙한 구조를 가진다. 이러한 방식은 영어처럼 낱말 간 띄어쓰기가 명확하고 형태 변화가 비교적 단순한 언어에 최적화되어 있지만, 교착어(agglutinative language)인 한국어에는 적절하지 않을 수 있다. 한국어는 의미를 갖는 어근에 여러 문법형태소(예, 조사, 어미 등)가 결합된 복합적인 형태를 이루는 경우가 많아, ChatGPT가 이를 정확하게 분리하지 못하는 경우가 발생할 수 있다(Jeon, 2022; Wang, 2023). 예를 들어, "학교에 갔다"라는 표현에서 ChatGPT는 이를 '학교', '에', '갔', '다'처럼 나누어 인식해 결과적으로 어절이나 음절 수를 실제보다 많거나 적게 계산하는 오류를 보일 수 있다. 따라서 자동전사 결과와 마찬가지로, 자동분석 결과 역시 사람의 추가적인 검토 및 수정 과정이 필요하다는 점을 시사한다.

한편, CIU에 대한 자동분석 결과를 살펴보면 두 집단 모두에서 사람의 수기분석 결과와 1차 시도가 유의하게 차이가 나는 양상을 보였다(Table 2, 3). 이는 나머지 5가지 지표와 달리 CIU가 내용 전달의 측면을 분석하는 지표로서 단순히 개수뿐만 아니라, 과제 또는 발화 상황에 적절한 표현이 맞는지 판단해야 하는 더 복합적인 분석을 요구하기 때문에 나타난 결과로 추측된다. CIU에 대한 자동분석 결과를 질적으로 확인하였을 때 관찰된 양상은 다음과 같다. 첫째, 1차 시도에서는 하나의 발화 내에서 반복된 동일 표현을 한 개의 CIU로만 계산하는 경우가 있었다. 둘째, 하나의 CIU로 간주해야 할 표현을 두 개로 분리해 분석하는 오류도 확인되었다(예, "개싸움이네"라는 표현에서 '개싸움'을 '개'와 '싸우다'로 분리하여 각각 별개의 CIU로 계산). 그러나 이러한 오류는 이후 수정 프롬프팅(Appendix 1)을 통해 오류 내용을 재확인하고 더욱 정교하게 분석하도록 추가적으로 지시함으로써, 1차 시도에 비해 분석된 CIU 목록이 사람과 유사하게 변경되는 것을 확인할 수 있었다. 이는 수정 프롬프팅을 통해 CIU에 대한 ChatGPT-4o의 분석 성능을 개선시키는 것이 가능함을 시사한다.

CIU에 대한 자동분석 결과에서 주목할 점은 4차 시도의 경우 집단별로 분석 능력이 다르게 나타났다는 것이다. 장년층의 발화에 대해서는 4차 시도에서 유의한 개선을 보이면서 사람과 유사한 수준의 분석이 가능하였다(Table 2). 이는 프롬프트의 내용이나 형식에 따라 생성형 AI의 응답이 달라질 수 있으며, 추가적인 프롬프팅 과정을 통해 분석 정확도를 향상시킬 수 있음을 보여주는 결과이다(Chun, 2025). 본 연구에서는 Appendix 1에서 제시한 예시와 같이 비교적 제한된 형태의 프롬프트를 사용하였고, 재시도 횟수도 최대 3회로 한정하였다. 그럼에도 재시도 이후 사람의 수기분석 결과와 유사한 수준의 분석 결과가 도출된 점은 언어병리학 분야에서 필요한 분석을 적절히 유도 가능하도록 수정 프롬프팅을 더욱 정교히 지속적으로 제공할 경우, ChatGPT가 사람에 준하는 수준의 분석을 이루어낼 가능성이 있음을 시사한다.

그러나 노년층의 발화에서는 이러한 4차 시도에 따른 개선이

나타나지 않아서 여전히 사람과는 분석 수준의 차이가 관찰되었다. 일반적으로 CIU는 내용상 적절하고 올바른 정보를 제공하는 낱말을 모두 포함하므로, 중복되는 표현이더라도 모두 개별적으로 계산해야 한다. 그러나 ChatGPT-4o는 하나의 발화 내에서 동일한 표현이 여러 번 산출될 때 CIU를 1개로 일괄적으로 산정하는 양상을 빈번하게 보이며, 수정 프롬프트를 제공하더라도 2회 반복된 CIU는 올바르게 계산하거나 재수정하는 반면, 3회 이상 반복되는 경우에는 여전히 중복된 개수만큼 CIU 개수로 산정하지 못하는 양상이 관찰되었다. 이에 노년층의 발화를 질적으로 살펴보았을 때, 장년층에 비하여 발화 내 반복 표현의 산출 빈도가 더 높았기 때문에 이러한 산정 오류가 더욱 빈번히 발생했을 가능성이 있다.

본 연구의 결과를 종합해보면, 장노년층 집단의 자발화 자료에 대하여 클로바노트는 대상자의 발화를 일차적으로 전사하는 데에 활용 가능한 수준의 자동전사 성능을 보이긴 하지만, 정확한 전사를 위해서는 사람이 마지막 단계에서 검토 및 보완하는 과정이 필요하다. ChatGPT-4o는 유창성 측면의 분석 지표에 대해 사람과 유사한 수준의 개수 산정이 가능한 듯 보이지만, 세부적으로 살펴보면 분석 내용이 사람의 수기분석 결과와 다르거나 누락된 경우가 있으므로 어절 또는 음절 수 세기에서 다소 오류를 보이는 것으로 확인되었다. 따라서 현시점에서는 유창성 분석에 있어 ChatGPT-4o를 단독으로 사용하는 것에 주의를 요한다. 내용 전달 측면의 분석 지표인 CIU의 경우, 발화 내 반복이 많지 않은 경우 수정 프롬프팅을 통한 재시도를 거치게 되면, 개수 산정에서 사람과 유사한 수준으로 분석이 가능함이 확인되었다. 본 연구는 고령 인구의 증가와 더불어 언어 및 인지 능력의 평가에 대한 수요가 급격히 증가하고 있는 현 상황에서, 이러한 자동화 기술을 활용할 때 주의해야 할 점이 무엇인지를 확인하는 것에 그 의의가 있다.

본 연구는 다음과 같은 제한점을 가진다. 첫째, 음성인식 및 생성형 AI를 각 1종(클로바노트, ChatGPT-4o)만 활용하였기 때문에, 다양한 AI 간 성능 차이에 대한 비교·분석이 이루어지지 않았다. 둘째, 자동전사 및 자동분석 성능을 평가하기 위한 수기전사 및 분석 값의 기준이 2인의 수기전사 및 분석 결과에 근거하였다. 그러나 사람의 수기분석은 주관적 판단에 따라 오류 개입의 가능성이 있어, 이를 기준으로 AI의 성능을 평가하는 것은 객관성과 정확성 측면에서 한계를 지닌다. 셋째, ChatGPT-4o를 활용한 자동분석 과정에서 제공된 사전 학습 및 수정 프롬프팅이 실제 분석 능력 향상에 기여했는지에 대한 명확한 인과 관계를 검증하지 못하였다.

본 연구의 제한점을 보완하고 인공지능 기반 언어 평가의 임상적 활용성을 더욱 확장하기 위하여, 향후에는 다음과 같은 후속 연구가 이루어질 필요가 있다. 첫째, AI의 경우 적용되는 언어모델에 따라 결과가 달라질 수 있으므로 보다 다양한 유형의 음성인식 및 생성형 AI 각각의 전사 및 분석 성능을 비교하여 임상적 적용의 실제로써 활용할 뿐만 아니라, 중재 도구로서의 활용 가능성까지 함께 탐색하는 연구가 필요하다. 둘째, 기존의 양적 분석에 더하여, AI를 활용한 질적 분석이 가능해진다면 쉼이나 억양과 같은 초분절적인 요소와 더불어 발화의 의미나 구조적 측면 등에 대해 다면적인 탐색을 시도하는 연구가 필요하다. 셋째, 생성형 AI의 분석 정확도를 향상시키기 위하여 언어병리학적 프롬프트 제공 방식에 대한 표준화된 매뉴얼이나 가이드라인을 개발하고, 해당 지침의 타당성과 효과를 검증하는 후속 연구가 요구된다.

Reference

Cho, Y.-I. (2023). Super-giant AI and generative AI. *ICT Standard Weekly, 1145,* 1-9.

Choi, M. A., Kim, S. H., Jo, M. A., Park, D. Y., Kim, Y. H., & Yoon, J. H. (2020). Development and enhancement of automatic caption generation system based on speech-to-text for the hearing impaired. *In Proceedings of 2020 the Korean Institute of Broadcast and Media Engineers Summer Conference,* 465-468.

Chun, H. (2025). Study on prompting in generative AI: A creative writing tool for the digital era. *Journal of Digital Contents Society, 26*(2), 275-285. doi:10.9728/dcs.2025.26.2.275

Chung, B. (2013). The relationship of sample sizes and mean length of utterances for typically developing children aged 2 to 4 years. *Journal of the Korean Association for Persons with Autism, 13*(3), 39-51. uci:G704-SER000008951.2013.13.3.008

Du, Y., & Juefei-Xu, F. (2023). Generative AI for therapy? Opportunities and barriers for ChatGPT in speech-language therapy. Tiny Papers @ ICLR 2023. https://openreview.net/forum?id=cRZSr6Tpr1S

Han, J., & Lee, M. (2024). Problem-based learning and ChatGPT: Explorative analysis of the relationship between chatgpt prompts and problem-solving skills. *The Journal of General Education, 26,* 111-145. doi:10.24173/jge.2024.01.26.4

Han, J., Park, S.-A., Ha, J., Lee, C., Jung, K., Han, M. L., ... Cho, G. (2025). Design of an automated framework for applying generative AI-based source code obfuscation techniques. *Journal of the Korea Society of Computer & Information, 30*(1), 73-85. doi:10.9708/jksci.2025.30.01.073

Han, J. H. (2023). The opening of generative AI era. *Media Issues & Trends, 55,* 6-17.

Huh, J., Park, S., Lee, J. E., & Ye, J. C. (2023). Improving medical speech-to-text accuracy using vision-language pre-training models. *IEEE Journal of Biomedical & Health Informatics, 28*(3), 1692-1703. doi:10.1109/jbhi.2023.3345897

Hwang, S.-I., & Kim, M. (2019). An analysis of artificial intelligence (A.I.)_related studies' trends in Korea focused on topic modeling and semantic network analysis. *Journal of Digital Contents Society, 20*(9), 1847-1855. doi:10.9728/dcs.2019.20.9.1847

Jeon, T. (2022). A linguistic study on tokenization methods for Korean text. *Language Facts & Perspectives, 55,* 309-354. doi:10.20988/lfp.2022.55..309

Kang, H. W., Kang, J. K., Lee, S. B., & Sim, H. S. (2022). Applications and performances of artificial intelligence in assessment and diagnosis of communication disorders: A systematic review of the literatures. *Communication Sciences*

& Disorders, 27(3), 703-722. doi:10.12963/csd.22923

Kang, J. K., Kang, H. W., Lee, S. B., & Sim, H. S. (2022). Research trends on the use of artificial intelligence in intervention for communication disorders. *Journal of Speech & Hearing Disorders, 31*(2), 107-115. doi:10.15724/jslhd.2022.31.2.107

Kang, M. G., Lee, S. B., Lee, D. H., Jeon, J. A., & Yoon, J. H. (2024). Exploring transcription accuracy of speech recognition AI in stuttering. *In Proceedings of the 23rd Korean Speech-Language & Hearing Association Conference, 23,* 242-244.

Kang, Y. W., Jang, S. M., & Na, D. L. (2012). *Seoul neuropsychological screening battery-second edition* (SNSB-II). Seoul: Human Brain Research & Consulting Co.

Kang, Y. W., Na, D. L., & Hahn, S. H. (1997). A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. *Journal of the Korean Neurological Association, 15*(2), 300-308.

Kee, B. S. (1996). A preliminary study for the standardization of geriatric depression scale short form-Korea version. *Journal of Korean Neuropsychiatric Association, 35*(2), 298-307.

Kim, H., Kwon, M. S., Na, D. L., Choi, S. S., Lee, K. H., & Chung, C. S. (1998). Decision making in fluency measures of aphasic spontaneous speech. *Korean Journal of Communication Disorders, 3*(1), 5-19.

Kim, J. W., Kang, Y., Lee, H.-Y., Kim, J., & Yoon, J. H. (2019). Changes in naming and cognitive abilities as the effects of semantic feature analysis treatment in middle-aged and older adults. *Communication Sciences & Disorders, 24*(1), 172-185. doi:10.12963/csd.18582

Kim, J. W., Kim, H., Namkoong, K. Kim, S. J., & Kim, D. Y. (2006). Spontaneous speech traits in patients with Alzheimer's disease. *Korean Journal of Communication Disorders, 11*(3), 82-98. uci:G704-000725.2006.11.3.005

Kim, Y. T. (1997). Study on utterance length in 2-4 year-old Korean children. *Korean Journal of Communication Disorders, 2*(1), 5-25.

Kwon, B. (2025). Analysis of generative AI research trends in South Korea using keyword network and BERTopic. *The Journal of Society for e-Business Studies, 30*(1), 167-187. doi:10.7838/jsebs.2025.30.1.167

Kwon, M. S., Kim, H., Choi, S. S., Na, D. L., & Lee, K. H. (1998). A study for analyzing spontaneous speech of Korean adults with CIU scoring system. *Korean Journal of Communication Disorders, 3,* 35-49.

Lee, S. B., & Yoon, J. H. (2025). Current applications and perceptions of generative AI in speech-language pathology clinical practice and research. *Journal of Speech-Language & Hearing Disorders, 34*(1), 155-169. doi:10.15724/jslhd.2025.34.1.155

Lee, Y. M., & Kim, H. (2001). An utterance analysis of conversations and picture description tasks of Korean adults. *Korean Journal of Communication Disorders, 6*(1), 1-11.

Ministry of Health and Welfare. (2025). 2023 dementia epidemiological and status survey results. Retrieved from https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&act=view&list_no=1484959&tag=&nPage=1

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research, 36*(2), 338-350. doi:10.1044/jshr.3602.338

Pavelko, S. L., Owens, R. E., Jr., Ireland, M., & Hahs-Vaughn, D. L. (2016). Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, Speech, and Hearing Services in Schools, 47*(3), 246-258. doi:10.1044/2016_lshss-15-0044

Scott, A., Gillon, G., McNeill, B., & Kopach, A. (2022). The evolution of an innovative online task to monitor children's oral narrative development. *Frontiers in Psychology, 13,* 903124. doi:10.3389/fpsyg.2022.903124

Sim, H. S., Kim, Y. T., Lee, Y. K., Kim, M. S., Kim, S. J., Lee, E. J., … Yoon, M. S. (2019). *Diagnosis & evaluation in communication disorders* (2nd ed.). Seoul: Hakjisa.

Statistics Korea. (2025). 2025 Population projection for Korea. Korean Statistical Information Service. Retrieved from https://kosis.kr/search/search.do?query=%EB%85%B8%EC%9D%B8%EC%9D%B8%EA%B5%AC

Wang, G. Q. (2023). Natural language analysis of Korean texts of AI-based chatbots and exploration of Korean education utilization: Focusing on ChatGPT and New-Bing. *Culture & Convergence, 45*(5), 1-17. doi:10.33645/cnc.2023.05.45.05.01

Yang, H. J., Oh, E.-B., & Kim, J.-M. (2023). Comparison of automatic speech recognition system for school-aged children's narratives: Naver Clova speech and Google speech-to-text. *Communication Sciences & Disorders, 28*(1), 30-38. doi:10.12963/csd.23952

Yoo, H.-J., Kim, M.-W., Park, S.-K., & Gim, G.-Y. (2020). Comparative analysis of Korean continuous speech recognition accuracy by application field of cloud-based speech recognition open API. *The Journal of Korean of Communications & Information Sciences, 45*(10), 1793-1803. doi:10.7840/kics.2020.45.10.1793

Appendix 1. Examples of revision prompt for utterance analysis

유창성 지표 분석에 대한 수정 프롬프트 예시	CIU 분석에 대한 수정 프롬프트 예시
1. 언어병리학 분야의 전문가라고 가정하고 발화 구분 원칙을 다시 확인하여 더욱 정교하게 분석해주세요	1. 언어병리학 분야의 전문가라고 가정하고 CIU를 더욱 정교하게 분석해주세요.
2. 언어병리학 분야에서 연구를 위해 실시되는 중요한 분석입니다. 대상자의 발화와 각 지표별 분석 방법을 다시 확인하고 자세히 분석해주세요.	2. 언어병리학 분야에서 연구를 위해 실시되는 중요한 분석입니다. CIU 분석을 자세히 실시해주세요.
3. 어절 수와 음절 수 계산을 정확하게 다시 실시해주세요.	3. "○○"은 CIU가 아닙니다. 그림을 다시 확인하고, 그림에 적절한 표현인지 확인하여 CIU 목록을 다시 정리해주세요.
4. 발화 구분 원칙을 다시 숙지하고, 대상자의 전체 발화에서 제외 기준에 해당하는 표현을 확인해 발화 분류와 분석을 진행해주세요.	4. CIU가 일부 누락되었습니다. 발화를 다시 확인해서 세밀하게 분석해주세요.
	5. 대상자의 발화에서 "○○"이 몇 번 표현되었는지 확인해보고, 중복된 개수만큼 CIU에 모두 포함시켜 주세요.

언어치료연구 2025년 제34권 제3호 001-011
ISSN 1226-587X / eISSN 2671-7158
JSLHD (Journal of Speech-Language & Hearing Disorders)

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2025.34.3.001
http://jslhd.org/

음성인식 및 생성형 AI의 발화 전사 및 분석 능력 탐색: 장노년층 발화를 중심으로

강민국[1], 이소빈[1], 조은별[2], 나덕렬[3], 윤지혜[4*]

[1] 한림대학교 일반대학원 언어병리청각학과 석사과정
[2] 해피마인드 의원 연구원
[3] 성균관대학교 의과대학 삼성서울병원 신경과 교수
[4] 한림대학교 언어청각학부 교수

목적: 본 연구는 장노년층 발화를 대상으로 음성인식 AI(클로바노트)를 이용한 자동전사와 생성형 AI(ChatGPT-4o)를 이용한 자동분석의 임상적 활용 가능성과 유용성을 탐색하고자 하였다.
방법: 만 50~90세 사이의 장노년층 성인 60명(장년층 18명, 노년층 42명)을 대상으로 그림 설명 과제를 통해 설명담화를 수집하였다. 전사 결과는 사람과 클로바노트 간 전사일치율을 비교하였고, 발화 분석은 사람과 ChatGPT-4o(1차, 4차 시도) 간 분석 지표 6가지(총 발화 수, 총 어절 수, 총 음절 수, 유창성 방해요소 빈도, 최대 문장길이, CIU 개수)에 대한 분석 결과를 비교하였다.
결과: 첫째, 클로바노트의 평균 전사일치율은 장년층 발화에서 80.60%, 노년층 발화에서 78.92%로 활용 가능한 수준의 자동전사 성능을 보였으나, 비유창성 요소(반복, 간투사 등)나 오조음을 생략 및 자동 수정하는 경향이 있어 정확한 전사를 위해서는 사람의 추가 검토 및 보완이 필수적이었다. 둘째, ChatGPT-4o는 유창성 관련 지표 분석 시, 장년층과 노년층 집단 모두에서 사람과 통계적으로 유사한 수준의 분석 능력을 보였다. 그러나 자동분석 결과를 세부적으로 확인해보면 발화 분류나 어절·음절 수 계산에서 수기분석 결과와 차이가 있었다. 셋째, ChatGPT-4o는 CIU 분석 시, 장년층 집단에서는 수정 프롬프팅 제공(4차 시도)을 통해 사람과 유사한 수준의 분석 능력을 보였다. 그러나 노년층에서는 수정 프롬프팅을 제공하는 경우에도 개선이 관찰되지 않았다.
결론: 본 결과는 언어병리학 분야에서 음성인식과 생성형 AI 기술의 잠재적 임상 활용 가능성 및 한계를 보여주며, 임상적 유용성을 높이기 위한 후속 연구의 필요성을 제언한다.

검색어: 인공지능, 자동음성인식, 생성형 AI

교신저자 : 윤지혜(한림대학교)
전자메일 : j.yoon@hallym.ac.kr

게재신청일 : 2025. 05. 30
수정제출일 : 2025. 07. 11
게재확정일 : 2025. 07. 31

이 논문은 2025년도 한림대학교 교비연구비에 의하여 연구되었음(HRF-202504-004).

ORCID
강민국
https://orcid.org/0009-0004-3996-8104
이소빈
https://orcid.org/0009-0009-9491-6899
조은별
https://orcid.org/0000-0002-0327-7035
나덕렬
https://orcid.org/0000-0003-1572-7862
윤지혜
https://orcid.org/0000-0003-1403-2276

참고문헌

강민국, 이소빈, 이다현, 전진아, 윤지혜 (2024). 말더듬 발화에 대한 음성인식 AI 의 전사 정확도 탐색. **한국언어치료학회 2024년도 제23회 학술발표대회 논문집, 23,** 242-244.

강연욱, 나덕렬, 한승혜 (1997). 치매환자들을 대상으로 한 K-MMSE 의 타당도연구. **대한신경과학회지, 15,** 300-307.

강연욱, 장승민, 나덕렬 (2012). **서울신경심리검사 2판**. 서울: 휴브알엔씨.

강진경, 강혜원, 이수복, 심현섭 (2022). 의사소통 장애 중재에서의 인공지능 활용에 대한 동향 연구. **언어치료연구, 31**(2), 107-115.

강혜원, 강진경, 이수복, 심현섭 (2022). 의사소통장애의 평가 및 진단에서 인공지능 적용과 성과에 관한 체계적 문헌고찰. **Communication Sciences & Disorders, 27**(3), 703-722.

권미선, 김향희, 최상숙, 나덕렬, 이광호 (1998). 한국 성인의 자발화 분석에 관한 연구. **언어청각장애연구, 3,** 35-49.

권보람 (2025). 키워드 네트워크와 BERTopic 을 활용한 국내 생성형 인공지능 연구 동향 분석. **한국전자거래학회지, 30**(1), 167-187.

기백석 (1996). 한국판 노인 우울 척도 단축형의 표준화 예비연구. **신경정신의학, 35**(2), 298-307.

김영태 (1997). 한국 2-4세 아동의 발화길이에 관한 기초연구. **언어청각장애연구, 2**(1), 5-25.

김정완, 강연욱, 이호영, 김재현, 윤지혜 (2019). 의미자질분석 중재에 따른 장노년층의 이름대기 및 인지능력 변화. **Communication Sciences & Disorders, 24**(1), 172-185.

김정완, 김향희, 남궁기, 김세주, 김덕용 (2006). 알츠하이머형 치매환자의 발화특성. **언어청각장애연구, 11**(3), 82-98.

김향희, 권미선, 나덕렬, 최상숙, 이광호, 정진상 (1998). 실어증환자 자발화의 유창성 연구. **언어청각장애연구, 3,** 5-19.

보건복지부 (2025). 2023년 치매역학조사 및 실태조사 결과 발표. https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&act=view&list_no=1484959&tag=&nPage=1

심현섭, 김영태, 이윤경, 김미선, 김수진, 이은주, ... 윤미선 (2019). **의사소통장애의 진단과 평가(2판).** 서울: 학지사.

양희재, 오은별, 김정미 (2023). 학령기 아동의 이야기 전사를 위한 자동음성

인식 프로그램 비교: Naver Clova와 Goolge STT를 중심으로. **Communication Sciences & Disorders, 28**(1), 30-38.

왕감경 (2023). AI 기반 챗봇 한국어 텍스트의 자연어 분석 및 한국어 교육 활용 모색: 챗GPT(ChatGPT)와 뉴빙(New-Bing)을 중심으로. **문화와융합, 45**(5), 1-17.

유현재, 김명화, 박상길, 김광용 (2020). 클라우드 기반의 음성인식 오픈 API의 응용 분야별 한국어 연속음성인식 정확도 비교 분석. **한국통신학회논문지, 45**(10), 1793-1803.

이소빈, 윤지혜 (2025). 언어병리 임상과 연구에서의 생성형 AI 활용 현황 및 인식. **언어치료연구, 34**(1), 155-169.

이영미, 김향희 (2001). 대화와 그림설명과제를 통한 한국성인 발화의 비교 분석. **언어청각장애연구, 6**(1), 1-11.

전태희 (2022). 한국어 텍스트의 토큰화 방법에 관한 언어학적 연구: fastText 단어 임베딩을 이용하여. **언어사실과 관점, 55,** 309-354.

전현주 (2025). 디지털 시대의 창의적인 글쓰기 도구, 생성형 AI의 프롬프팅에 관한 연구. **디지털콘텐츠학회논문지, 26**(2), 275-285.

정부자 (2013). 2-4세 일반아동의 자발화 표본크기와 평균발화길이의 비교. **자폐성장애연구, 13**(3), 39-51.

조영임 (2023). 초거대 AI와 생성형 인공지능. **ICT Standard Weekly, 1145,** 1-9.

최미애, 김승현, 조민애, 박동영, 김용호, 윤종후 (2020). 청각장애인을 위한 음성-자막 자동 변환 시스템 개발 및 음성 인식률 고도화. **한국방송미디어공학회 2020년 하계학술대회 논문집,** 343-346.

통계청 (2025). 장래인구추계, 2025. https://kosis.kr/search/search.do?query=%EB%85%B8%EC%9D%B8%EC%9D%B8%EA%B5%AC

한정훈 (2023). 생성형 AI 시대의 개막. **미디어 이슈 & 트렌드, 55,** 6-17.

한지훈, 박승아, 하준서, 이창민, 정경미, 한미란, ... 조금환 (2025). 생성형 AI 기반 난독화 기법 적용 자동화 프레임워크 설계. **한국컴퓨터정보학회논문지, 30**(1), 73-85.

한진영, 이민정 (2024). 문제중심학습과 챗GPT: 프롬프트와 문제해결력에 대한 탐색. **교양학 연구, 26,** 111-145.

황서이, 김문기 (2019). 국내 인공지능분야 연구동향 분석: 토픽모델링과 의미연결망분석을 중심으로. **디지털콘텐츠학회논문지, 20**(9), 1847-1855.