

Comparison of Speech Recognition Accuracy of AI Speakers and Listener's Intelligibility for Dysarthric Speech

Joo Hyoung Kim¹, Ji Youn Kim^{2*}

¹ Dept. of Speech Therapy, Woosong College, Professor

² Dept. of Speech-Language Therapy & Aural Rehabilitation, Woosong University, Professor

Purpose: The purpose of this study is to investigate whether there is a difference in the intelligibility of listeners' speech intelligibility and the recognition accuracy of AI speakers according to the linguistic unit.

Methods: Dysarthric speakers were asked to read 4 sentences and their voices were recorded. An audio file was created by dividing the each of the 4 sentences into words. The listeners were asked dictate the speech to obtain syllable intelligibility and to rate it on a 5-point scale. The speech recognition accuracy of a Kakao Mini AI speaker was obtained by playing the speech and converting it into text, and this was compared with the intelligibility of the listener.

Results: First, the speech recognition accuracy of the Kakao Mini was significantly lower for the dysarthria speaker than for the normal speaker. Second, the syllable intelligibility for dysarthric speech was significantly higher than the speech recognition accuracy of the Kakao Mini. The human understood sentences better than words, but there was no significant difference in the speech recognition accuracy of the sentence and word by the Kakao Mini. Third, there was a significant correlation between scale intelligibility, syllable intelligibility, and speech recognition accuracy of the Kakao Mini and the syllable intelligibility of the listener for dysarthric speech.

Conclusions: The results show that the speech recognition accuracy of Kakao Mini for dysarthric speakers is still insufficient at the sentence level. However, it showed that the AI speaker can be used as one of the objective intelligibility measurement methods to evaluate the intelligibility of a speaker with speech impairment by showing a similar pattern to the human ear in the word unit.

Correspondence: Ji Youn Kim, PhD

E-mail: bluejiyou@hotmail.com

Received: August 31, 2022

Revision revised: September 29, 2022

Accepted: October 31, 2022

This work was supported by Kakao Corp (2020).

ORCID

Joo Hyoung Kim

<https://orcid.org/0000-0002-8927-6921>

Ji Youn Kim

<https://orcid.org/0000-0003-4259-778X>

Keywords: Dysarthria, intelligibility, AI speakers, speech recognition

1. 서 론

인공지능이 우리 생활에 익숙해지면서 음성인식 기술을 이용한 다양한 상품들이 대중화되었다. 음성기반 플랫폼을 사용하게 되면 손을 이용하지 않고도 편리하게 기기를 관리하거나 제어할 수 있어서 인공지능 스피커를 통해 사용자의 음성을 인식하여 검색 기능을 사용할 수 있고, 메시지 전송이나 음악 재생과 사진 촬영 등이 가능해졌다. 이러한 음성인식을 활용하는 분야는 점점 확대될 것이며 앞으로 다양한 상품들이 개발될 것으로 예상된다. 현재까지 음성인식 기술은 일반인들의 음성 데이터를 기반으로 개발되었기 때문에 음성인식을 활용하지 못하는 사람들도 있다. 말장애를 가진 사람들은 일반 화자와 다른 음성 특징을 가지고 있어 자동 말소리 인식 기기들은 그들의 말을 인식하는 데

제한이 있을 수밖에 없다.

마비말장애를 가진 사람들은 중추 혹은 말초신경계 손상으로 인해 말산출에 필요한 근육들이 약화되어 조음, 공명, 호흡, 발성, 운율 등에 문제를 나타낸다. 말운동체계의 손상으로 인해 정확한 음소 산출에 어려움을 보일 뿐 아니라 과다비성, 음질저하, 부자연스러운 운율 문제를 보이므로 대화상대자는 마비말장애 화자들의 말을 이해하기 어렵다. 청자가 화자의 말소리를 알아들을 수 있는 정도를 나타내는 지표로 말명료도(speech intelligibility)가 사용되는데, 말명료도는 여러 요인에 의해 영향을 받는다. 화자의 자음정확도나 공명, 운율 등과 같은 말산출과 직접적으로 관련된 요인들이 영향을 주기도 하지만 전달하고자 하는 메시지의 길이나 예측 가능성, 문맥적 단서나 주변 환경의 소음의 정도, 화자 말에 대한 청자의 청취 경험 등도 말명료도 판단에 차이를 가져올 수 있다(Hustad, 2008).

말명료도를 측정하는 방법에는 평정법과 음절 명료도가 있다. 평정법은 척도를 이용하여 전반적인 명료도를 판단하는 방법이고, 음절 명료도는 화자의 발화를 청자가 받아 적은 후 화자가 의도한

음절수에 대하여 청자가 정확하게 이해한 음절수의 비율을 말한다 (Kim & Shin, 2020). 이러한 방법들은 모두 청자의 주관적인 판단에 의한 것이라 연구자들은 신뢰도가 높으면서 객관적으로 말명료도를 평가할 수 있는 방법에 관심을 가져왔다. 말명료도가 저하된 화자들의 중재 효과를 입증하기 위해서도 보다 객관적인 자료가 필요하기 때문이다(McHenry & LaConte, 2010).

최근 인공지능 스피커가 상용화되면서 말명료도를 객관적으로 측정하기 위한 방안으로 음성인식 기술이 제안되고 있다. 하지만 사람의 귀와 마찬가지로 음성인식 시스템 역시 다양한 요인에 의해 음성인식의 정확도가 달라질 수 있다. 들려주는 단위가 단어인지 문장인지에 따라, 어휘 사용 빈도에 따라, 발화가 읽기 과제인지 자발화 과제인지에 따라, 마이크 성능이나 주변의 소음대 배음비(SNR)에 따라 음성인식 정확도에 차이가 있다고 보고되었다(Kitzing et al., 2009; Mustafa et al., 2015; Young & Mihailidis, 2010).

마비말장애 화자들의 말소리는 지각적으로 명료도가 저하되고 마비말장애 유형에 따라 말산출 변이가 커서 음성인식 정확도가 낮고 (Green et al., 2021), 중증도에 따라서도 명료도나 음성인식에 차이를 보인다(Mustafa et al., 2015)고 보고되었다. 마비말장애 화자들의 말은 정상 청자들도 잘 알아듣지 못할 뿐 아니라 음성인식 시스템 역시 그들의 말을 잘못 인식할 가능성이 높다는 것을 의미한다. 그렇다면 마비말장애 화자들의 말에 대한 음성인식 능력이 인공지능 스피커와 사람의 귀가 얼마나 유사한지 알아볼 필요가 있다. 말장애 화자들에 대한 선행 연구들은 인공지능 스피커의 음성인식 정확도와 청자의 말명료도를 비교하였을 때 상반된 결과를 보여주고 있다. 일부 선행 연구들에서는 말장애 화자의 말소리에 대한 청자의 말명료도와 음성인식 시스템의 정확도가 높은 상관성이 나타난다고 보고하고 있다(Doyle et al., 1997; Hwang et al., 2011; McHenry & LaConte, 2010; Schuster et al., 2006; Tu et al., 2016). 다른 연구들에서는 일반 화자의 경우 청자의 말명료도 평가와 음성인식 스피커의 정확도는 크게 차이가 나지 않지만, 말장애 화자의 경우 음성인식 스피커가 청자의 귀보다 정확도가 떨어지는 경향이 있다고 보고하였다(Hwang et al., 2011; Schuster et al., 2006).

본 연구에서는 마비말장애의 말소리에 대하여 정상 청자의 이해 정도와 인공지능 스피커의 음성인식 정확도를 비교함으로써 음성인식에 영향을 미치는 요인들을 구체적으로 확인하고자 한다. 음성인식 시스템을 활용한 중재가 이루어진다면 마비말장애 화자의 말명료도를 높이는 데에도 기여할 수 있어 말장애를 가진 화자들의 의사소통 향상에 도움이 될 것이다. 마비말장애를 가진 사람들의 말소리를 음성인식 시스템이 인식할 수 있다면 그들의 삶의 질에도 긍정적으로 기여할 것이다. 본 연구는 음성인식 시스템이 언어치료 평가 및 치료에 기여할 수 있는 기초 자료를 제공하는 데 목적을 두고 있다.

II. 연구 방법

1. 연구 대상

1) 음성녹음 화자

연구 대상으로 후천적 마비말장애로 진단받았으며 MMSE 24점 이상으로 문장 읽기가 가능한 화자 10명을 선정하였다. 10년 이상의 경력과 마비말장애 환자의 평가와 치료를 5년 이상 경험한 1급 언어재활사 2명이 화자의 음성을 듣고 중증도를 판단하였다. 마비말장애 화자에 대한 일반적인 정보는 Table 1에 제시하였다.

선정된 마비말장애 화자와 성별과 연령이 동일하고 말하기와 읽기에 어려움이 없는 정상 화자 10명도 포함하였다.

Table 1. Participants' information

	Age	Sex	Etiology	Type	Speech severity
1	35	Female	CI	Spastic	Moderate
2	48	Male	CI	UUMN	Moderate
3	50	Male	CI	Flaccid	Mild
4	51	Female	CI	Flaccid	Moderate
5	51	Male	CI	Flaccid	Severe
6	53	Female	ICH	Spastic	Moderate
7	53	Male	ICH	Spastic	Moderate
8	62	Male	CI	Spastic	Mild
9	64	Male	CI	Hypokinetic	Mild
10	77	Male	CI	Flaccid	Severe

Note. CI=cerebral infarction; ICH=intracerebral hemorrhage; UUMN=unilateral upper motor neuron.

2) 일반 청자

마비말장애 화자의 음성을 들어본 경험이 없는 정상 청력을 가진 20대 대학생 남녀 30명을 청자로 구성하였다.

2. 연구 절차

1) 화자의 음성녹음

대상자들의 동의를 구해 소음이 적은 조용한 공간에서 발화 대상자의 입과 마이크 거리를 약 10~15cm로 유지하여 LOTTE Linear PCM Voice Recorder LVR-533로 녹음하였다. 읽기 자료는 굴림체, 35pt로 한 장에 한 문장씩 적어 A4 용지 가로 방향으로 출력하여 제공하였고, 각 대상자들에게 4개의 문장을 읽게 하였다. 한 호흡에 편안하게 읽을 수 있는 4~5어절로 문장을 구성하고, 10~17음절로 각 문장의 길이를 달리하였다. 첫음절은 모음이나 비음으로 하고 마지막 음절은 모음으로 하여 음향적으로 문장의 시작과 끝을 명확하게 알 수 있도록 하였다.

우리말소리의 다양한 자음을 포함하고 화자가 읽기 편안하도록 문장 형태를 구성하였다. 각 문장은 다음과 같다. (1)외로울 때면 마음맞는 사람을 찾아보세요. (2)오늘 가나요, 내일 가나요? (3)아름다운 민아는 마음도 예뻐요. (4)낙엽이 떨어지는 가을길을 걸어요.

2) 음성자료

녹음된 마비말장애 화자 음성을 언어학적 단위에 따른 음성인식과 말명료도 차이를 비교하기 위해 Praat을 활용해서 문장 단위와 어절 단위로 분리하여 음성파일을 저장하였다. 마비말장애 화자 10명이 산출한 4개의 문장은 17개의 어절로 구분되었다. 10명의 마비말장애 화자가 읽은 문장은 40개이다. 문장파일 40개와 각 문장을 어절 단위로 나눈 어절 파일이 170개로 총 210개의 음성 파일이 만들어졌다. 청자가 말명료도를 평가할 때 동일한 문장에 대한 반복 학습 효과를 제거하고자 한명의 청자가 문장 4개의 명료도만 평가하도록 10개의 세트로 구분하였다. 즉, 한 세트에는 서로 다른 4명의 화자가 산출한 4개의 문장 파일과 각 문장을 어절로 구분한 17개의 어절 파일을 포함하여 총 108음절로 이루어진 21개의 음성파일이 포함되었다. 예를 들어, 세트 1에는 A 화자의 (1)번 문장과 (1)번 문장의 어절 다섯 개, B 화자의 (2)번 문장과 (2)번 문장의 어절 4개, C 화자의 (3)번 문장과 그 문장의 어절 4개, D 화자의 (4)번 문장과 그 문장의 어절 4개가 포함되었다. 세트 2에는 E, F, G, H 화자의 문장과 어절이, 세트 3에는 I, J, A, B 화자의 문장과 어절이 포함되었다. 청자들은 10개 중 하나의 세트를 듣고, 받아 적기를 하고 말명료도를 척도로 평정한다. 10개 세트에 대해 30명의 청자가 명료도를 측정하므로 각 세트는 청자 3명이 듣고 평가하게 된다.

일반 화자 10명의 음성은 4개 문장과 각 문장을 17개 어절로 분리하여 총 210개 음성 파일로 나누어 '헤이카카오'가 받아 적을 수 있도록 하였다.

3) 청자 명료도 평가

말명료도를 측정하기 위해 Praat에서 제공되는 청지각 실험용 스크립트인 Experiment MFC (collection)를 이용하여 실험용 플랫폼을 구성하였다. 조용한 환경에서 노트북 컴퓨터(LG15U53)를 통해 동일한 헤드폰(Sony MDRZX100)을 착용하고 실험자극을 듣도록 하였다. 청자는 10개의 세트 중 하나에 배정되어 음성파일을 듣게 된다. 학습 효과를 방지하기 위해 어절을 먼저 듣고 문장의 순서로 들려주었으며 어절 파일과 문장 파일 내에서 제시 순서를 임의로 배정하여 예측성을 줄이도록 프로그램을 설계하였다.

실험 방법과 평가 기준을 간단히 설명한 후, 종이에 받아 적게 한 후 각 어절 및 문장의 명료도를 척도로 판단하여 화면에 제시된 1~5의 숫자 중 하나를 클릭하도록 하였다. 문장은 2회 반복해서 들을 수 있게 설정하였다.

청자들이 전사한 어절 및 문장은 목표 어절 및 문장을 기준으로 각 문장이나 어절에 사용된 총 음절수로 나누어 음절 명료도율 %로 계산하였다. 예를 들어, '낙엽이 떨어지는 가을길을 걸어요'를 청자가 '딸기를 찾으러 갈 길을 걸어요'라고 전사했다면

청자가 정확하게 받아 적은 음절수 5를 목표음절수 14로 나누어 해당 문장의 음절 명료도율은 약 35.7%가 된다.

말명료도 척도 평가는 1~5의 5점 리커트 척도로 이루어졌다 (1=전혀 못 알아듣겠다. 2=잘 못 알아듣겠다. 3=그럭저럭 알아들을 만하다. 4=못 알아듣는 부분이 약간 있다. 5=매우 잘 알아듣겠다).

4) 인공지능 음성인식 정확도

녹음된 마비말장애 화자 음성을 컴퓨터에 연결된 카카오톡미니스피커로 재생하였다. 휴대폰에 카카오톡 미니스피커를 연결하고 '헤이카카오'를 설치하고, 스피커와 10cm 거리를 두고 음성파일을 재생하였다. '헤이카카오'의 받아쓰기 기능을 활성화하여 음성파일을 텍스트로 변환하였다. 음성이 재생되는 동시에 텍스트로 변환되고, 문장이 끝나면 문맥에 맞게 오인식을 수정하는 작업이 이루어지기 때문에 문장이나 어절 재생을 마친 후 헤이카카오가 최종적으로 받아쓰기를 한 결과를 전사하였다. 목표 어절 및 문장과 비교하여 음성인식 받아쓰기 결과로 얻어진 어절과 문장의 음절당 정확도를 계산하였다. 음성파일은 2회 재생되었고, 음절당 정확도가 더 높은 수치를 인공지능 스피커의 음성인식 정확도로 사용하였다.

3. 실험설계

인공지능 스피커에서 인식한 음절 정확도와 일반 청자의 말명료도의 차이를 확인하기 위하여 SPSS 25.0 프로그램을 이용하였다. 정상 화자와 마비말장애 화자 집단에 대한 언어학적 단위에 따른 인공지능 스피커의 음성인식 정도를 비교하기 위해 이요인 분산분석(two-way ANOVA)을 실시하였다. 마비말장애 음성에 대한 인공지능 스피커의 음성인식 정확도와 청자의 음절 명료도 차이를 알아보기 위해 각 문장과 어절에 대한 청자의 음절 명료도 평균을 구하여 반복측정 분산분석(repeated measure ANOVA)을 실시하였다. 청자의 명료도 측정 방법 간의 상관관계를 알아보기 위해 척도 명료도와 음절 명료도 간의 스피어만 상관분석(Spearman's rho)으로 상관계수를 구하였고, 마비말장애 음성에 대한 청자의 음절 명료도와 인공지능 스피커의 음성인식 정확도의 상관관계를 알아보기 위하여 스피어만 상관분석(Spearman's rho)을 실시하였다.

III. 연구 결과

1. 정상 화자와 마비말장애 화자에 대한 인공지능 스피커의 음성인식 정확도 비교

정상 화자와 마비말장애 화자 집단의 언어학적 단위(문장, 어절)에 대한 인공지능 스피커의 음성인식 정확도는 Table 2와 같다. 각 집단과 언어학적 단위에 따른 음성인식 정확도에 차이가 있는지 살펴보기 위해 이요인 분산분석을 실시한 결과 정상 화자 집단에 대한 음성인식 정확도가 마비말장애 화자 집단보다 높은 수치

를 보이며 집단 간에는 유의한 차이를 보였다($p<.05$). 그러나 언어학적 단위에 따른 차이나 집단과 언어학적 단위의 상호작용에는 유의한 차이가 나타나지 않았다.

Table 2. Comparison of Kakao-mini's speech recognition (%)

	Sentence	Word	Total
Control	95.84	85.20	87.22
Dysarthria	46.73	44.00	44.52
Total	71.28	64.60	65.87

2. 마비말장애 화자에 대한 청자의 음절 명료도와 인공지능 스피커의 음성인식 정확도

마비말장애 화자의 음성에 대해 문장과 어절을 모두 포함한 청자의 평균 음절 명료도는 55%, 인공지능 스피커의 음성인식 정확도는 45%로 통계적으로 유의한 차이가 있었다($p<.001$). 즉, 마비말장애 화자의 음성을 인공지능 스피커보다 청자가 더 잘 알아들었다.

언어학적 단위에 따른 청자의 음절 명료도는 문장에서 77.34%, 어절에서 49.23%이었다. 인공지능 스피커의 음성인식 정확도는 문장과 어절에서 각각 46.83%, 44.14%이었다. 각 문장과 어절에 대한 청자의 음절 명료도와 인공지능 스피커의 음성인식 정확도는 Table 3과 같다.

Table 3. Listener's intelligibility and Kakao-mini's speech recognition (%) of dysarthric speech

Sentence (Number of syllable)	Listener	Kakao- mini	Word (Number of syllable)	Listener	Kakao- mini
1 (17)	85.10	55.29	11 (3)	21.11	16.67
			12 (2)	41.67	35.00
			13 (4)	50.00	65.00
			14 (3)	52.22	56.00
			15 (5)	79.33	48.00
2 (10)	82.67	52	21 (2)	41.67	50.00
			22 (3)	58.89	60.00
			23 (2)	56.67	65.00
			24 (3)	50.00	46.67
3 (13)	78.71	44.62	31 (4)	73.33	37.50
			32 (3)	34.44	46.67
			33 (3)	32.22	26.67
			34 (3)	85.56	70.00
4 (14)	62.86	35	41 (3)	38.89	13.33
			42 (4)	37.50	60.00
			43 (4)	40.00	50.00
			44 (3)	43.33	33.33
Total	77.34	46.73		49.23	44.14

언어학적 단위에 대해 청자의 음절 명료도와 인공지능 스피커의 음성인식 정확도를 반복측정 분산분석으로 비교한 결과(Table 4), 청자의 음절 명료도와 인공지능 스피커의 음성인식 정확도에 유의한 차이가 나타났다. 이러한 차이는 언어학적 단위와 상호작용 효과를 나타냈는데 문장 수준에서 두 방법 간에 유의한 차이가 나타났다. 언어학적 단위에 따라 청자의 음절 명료도에는 유의한 차이($p<.001$)가 있는 반면, 인공지능 스피커는 문장과 어절 간 유의한 차이를 보이지 않았다. 사람들은 어절에 비해 문장에서의 음절 명료도가 더 높게 나타났지만 인공지능 스피커는 언어학적 단위(문장, 어절)에 따른 음성인식 정확도에 차이가 없었다.

Table 4. Repeated measures ANOVA of speech recognition according to linguistic units

	SS	df	MS	F
Linguistic unit	15249.24	1	15249.24	.004**
Error	378104.86	208	1817.81	
Accuracy	206183.48	1	206183.48	<.001***
Accuracy × Linguistic unit	10551.97	1	10551.97	<.001***
Error	135767.76	208	652.73	
Total				

** $p<.01$, *** $p<.001$

3. 청자의 음절 명료도와 척도 명료도와의 관계

마비말장애 화자가 읽은 4개의 문장에 대한 청자의 척도 명료도 결과는 Figure 1과 같다. 4개 문장에 대한 척도 명료도의 평균은 3.28이었고, 어절의 척도 명료도 평균은 2.59로 문장과 비교하여 어절 단위의 척도 명료도는 낮은 편이었다. 각 문장 별로 살펴보면, 1번 문장이 3.63로 척도 명료도가 가장 높았고, 4번 문장이 3.07로 가장 낮은 척도 점수를 보였다. 이는 청자의 음절 명료도 패턴과 유사하였다. 어절 단위에서는 음절 명료도와 척도 명료도 간에 다소 차이를 보이기는 하였지만 스피어만 상관분석 결과, 명료도 측정 방법인 척도 명료도와 음절 명료도는 $r=.582(p<.001)$ 로 비교적 높은 상관이 나타났다.

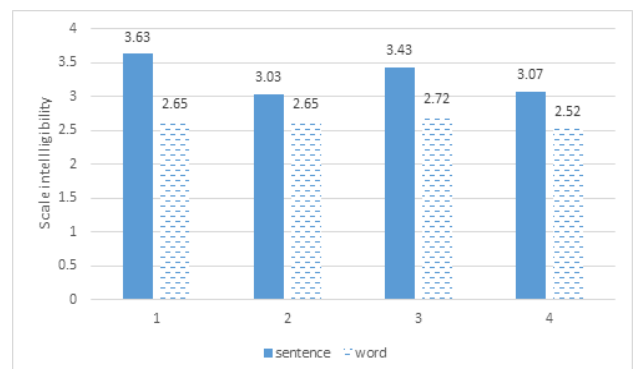


Figure 1. Listener's rating scale

4. 청자의 명료도와 인공지능 스피커의 음성인식 정확도 관계

청자가 판단한 명료도(음절 명료도, 척도 명료도)와 인공지능 스피커의 음성인식 정확도를 비교하기 위하여 스피어만 상관분석을 실시하였다.

청자의 음절 명료도와 인공지능 스피커의 음성인식 정확도를 비교한 결과, 문장 단위에서 $r=.585(p<.001)$ 로 높은 상관을 보였고 어절 단위에서 $r=.465(p<.001)$ 로 유의한 상관을 보였다.

청자의 척도 명료도와 인공지능 스피커의 음성인식 정확도는 $r=.393(p<.001)$ 으로 유의한 상관관계를 나타내었다.

IV. 논의 및 결론

본 연구는 말소리 산출에 어려움을 보이는 마비말장애 화자의 음성을 인식하는 능력에서 문장과 어절의 언어학적 단위에 따른 인공지능 스피커와 일반 청자 간에 차이가 있는지 알아보고자 하였다. 연구의 주요 결과에 대한 요약과 논의는 다음과 같다.

첫째, 사람이 인공지능 스피커 음성인식보다 마비말장애 화자의 음성을 더 잘 알아들었다. 인공지능 스피커의 음성인식은 정상 화자의 음성에 대해서는 사람이 귀로 듣는 음절 명료도와 유사한 수준의 음성인식 능력을 가지고 있지만, 마비말장애 화자의 음성에 대해서는 사람이 듣는 것에 비해 음성인식 능력이 상당히 낮았다. 이와 같은 결과는 정상 화자와 마비말장애 화자에 대한 청자의 말 명료도와 컴퓨터 음성인식 정확도를 비교한 Menigistu와 Rudzicz (2011)의 연구 결과와 유사하다. 마비말장애 화자의 음성에 대해 사람보다 컴퓨터의 음성인식 정확도가 떨어지는 결과에 대해 연구자들은 마비말장애 화자의 음성은 음향학적 정보들이 왜곡되었을 뿐 아니라 음절이나 과다비성과 같은 음성, 부적절한 숨이나 운율적인 부자연스러움 등이 나타나는데, 이러한 말의 특징은 사람의 귀가 말소리를 알아듣는 데에는 간접 효과가 크지는 않았지만 음성인식 시스템이 알아듣는 데에는 방해 요인으로 작용했을 것이라 해석하였다.

또한 음성인식 시스템에 따라 화자의 음성을 인식하는 방식에 차이가 있으나 본 연구에서 사용된 카카오톡의 ‘헤이카카오’가 받아 적은 마비말장애 화자의 음성인식 정확도는 다른 인공지능 스피커를 사용해서 측정한 수치와 큰 차이가 없었다. 아마존의 알렉사, 애플의 시리, 구글의 구글 어시스턴트와 같은 음성인식 시스템에 마비말장애 화자의 문장을 들려주고 인식하도록 하였을 때 47~58% 정도의 음성인식 정확도를 나타냈다는 결과(Ballati et al., 2018)와 유사한 수준이었다.

둘째, 언어학적 단위에 따른 청자의 말명료도와 인공지능 스피커의 음성인식 정확도를 비교하여 살펴보았다. 사람은 문장에서 음절 명료도가 높은 반면 어절에서는 명료도가 상당히 낮아 문장과 어절 간 음절 명료도에서 통계적으로 유의한 차이를 보였다. 인공지능 스피커는 문장에서 약간 더 정확도가 높긴 하였으나 문장과 어절 간의 음성인식 정확도의 차이가 크지 않았다. 문맥의 영향이 적은 어절 단위에서는 인공지능 스피커와 청자의 음성인식 정도의

차이가 크지 않음을 고려할 때 사람은 문장을 이해하는데 맥락을 더 많이 활용하고 있음을 알 수 있다. 그러나 이와 같은 결과는 청자의 음절 명료도에서 단어와 문장 간 차이가 크지 않았던 McHenry와 LaConte(2010)의 연구와 차이가 있다. McHenry와 LaConte(2010)는 3명의 마비말장애 화자 음성을 청자와 컴퓨터 음성인식 시스템에 들려주고 인식 정도를 비교하였는데 단어 수준에서 컴퓨터 음성인식 정확도의 범위는 3~6%이었으며 문장 수준은 2~4%이었다고 하였다. 이에 반해 240명의 청자들에게 들려주었더니 단어 수준에서는 85~88%, 문장 수준에서는 62~83%를 알아들었다고 하였다. 본 연구와 선행 연구 결과의 차이는 마비말장애 화자의 말장애 중증도 및 어절 분석의 방법 차이에 기인한 것으로 보인다. McHenry와 LaConte(2010)의 연구에서는 마비말장애 화자 3명의 중증도가 각각 경도, 중도, 중·심도이었으며, 화자가 단어와 문장을 각각 따로 녹음하여 분석하였으나 본 연구에서는 문장 읽기 자료를 어절 단위로 잘라 분석하였기 때문에 의미를 지닌 하나의 단어가 인식될 때와 문장 내에서 조사가 포함된 목적어나 부사어 등의 어절로 사용될 때의 인식 차이 때문이라 생각된다.

셋째, 청자의 음절 명료도와 마찬가지로 척도 명료도 역시 어절보다 문장에서 높은 수치가 나타났다. 각 문장별로 살펴보면, 음절 명료도와 척도 명료도 모두 1번 문장이 가장 높았고, 4번 문장이 낮은 척도 명료도와 음절 명료도를 보이며 명료도 측정 방법으로 얻은 두 결과가 비슷한 양상을 보였다. 상관분석 결과, 음절 명료도와 척도 명료도 간에 유의한 상관관계를 보였다. 문장에 비해 어절에서는 척도 명료도와 음절 명료도의 차이가 크지 않았다. 그러나 (3)번 문장 내 ‘민아는’ 어절에서 척도 명료도에 비해 음절 명료도 수치가 더 낮았는데, 이는 문장 내에 고유명사가 포함되면서 청자의 음절 명료도가 낮아지는 요인으로 작용한 것으로 해석할 수 있다. 이러한 결과는 명료도 측정 방법 간의 차이를 보일 수 있음을 밝힌 선행연구와 비슷한 결과로(Hwang et al., 2011) 명료도 측정에서 음절 명료도와 척도 명료도 사이에는 차이가 있을 수 있음이 설명된다.

마지막으로 마비말장애 화자의 음성에 대한 인공지능 스피커의 음성인식 정확도와 청자의 음절 명료도 간에 유의한 상관관계를 보였다. 인공지능 스피커가 말소리 산출에 어려움이 있는 마비말장애 화자의 음성을 인식하는 데에 아직까지는 사람의 귀보다는 인식 정확도가 떨어지기는 하지만 그래도 상당히 근접하게 음성인식이 가능하고 사람이 귀가 듣는 것과 비슷한 양상을 보인다고 해석할 수 있다. 그러나 이러한 결과는 5명의 마비말장애를 대상으로 음성인식 시스템과 청자들의 명료도를 비교했던 선행연구와는 차이가 있다. Hwang 등(2011)은 5명의 마비말장애 화자를 대상으로 단어 수준에서의 음성인식 시스템과 청자들의 발화 명료도 간에 높은 상관관계를 나타냈으나 컴퓨터의 음성인식 정확률이 사람의 발화 명료도보다 비교적 더 높은 결과를 보인다고 하였다. 하지만 선행연구의 연구대상에는 5명의 뇌성마비 화자 중 의사소통에 불편함이 없는 화자가 2명이 포함되어 있어 비교적 명료도가 높은 화자를 대상으로 하였음을 알 수 있고 문장 내 어절 단위가 아닌, 단어로 음성을 제시하였기에 음성인식이 훨씬 수월했으리라 사료된다.

본 연구에서는 마비말장애 화자의 음성에 대한 청자의 말명료도

와 인공지능 스피커를 통한 음성인식 정도를 비교해봄으로써 인공지능 스피커가 말소리장애의 객관적인 평가도구로 사용될 수 있는지의 가능성을 살펴보았다. 정상 화자에 대한 인공지능 스피커의 음성인식 정확도는 높은 편이었으나 아직 마비말장애 화자의 음성을 정확하게 알아듣는 데에는 제한이 있었다. 마비말장애 화자의 음성이라 할지라도 어절 단위에서는 사람의 귀로 듣는 명료도와 유사한 수치를 나타냈으나 문장 수준에서는 청자의 음절 명료도에 비해 낮은 음성인식 정확도를 보였다. 사람은 문맥을 고려하여 말소리장애를 가진 화자의 음성에서 의미를 유추할 수 있는 데 반해 인공지능 스피커는 의미를 고려하기는 하지만 아직은 근접한 소리를 인식하는 데 그친다는 것으로도 해석할 수 있다.

청력검사 분야에서 음성인식 관련한 프로그램이 논의되어 왔으나 인공지능 음성분석 및 음성인식과 관련된 평가 및 증재 연구는 아직 부족한 실정이다(Bae et al., 2015). 말소리장애인들과 언어재활사들이 음성인식 시스템을 언어치료 평가 및 치료에 적용할 수 있다면 의사소통장애인에 이르기까지 인공지능 스피커의 대중화가 이루어질 수 있을 것이고, 의사소통장애인의 삶의 질 향상에도 기여할 것이다. 그 한계를 극복하기 위해서는 말소리장애인 음성에 대한 많은 데이터를 확보해야 한다. 따라서 현재 인공지능이 어느 정도까지의 발음을 인식할 수 있는지를 확인하고 말소리장애인의 음성인식 시스템을 구축하기 위한 기초 연구가 더 필요하다. 연구들이 뒷받침된다면 인공지능 스피커의 음성인식 정확도는 말명료도를 객관적으로 측정하는 수치로 활용될 수 있다. 또한 말소리장애 대상자들의 조음 능력이 인공지능 시스템이 인식할 수준으로 향상될 수 있도록 언어치료 도구로 음성인식 시스템이 활용될 수 있고, 인공지능 시스템의 음성인식 정확도를 증재의 목표 기준으로 삼을 수 있을 것이다(Jeon et al., 2019). 말소리장애를 가진 대상자들의 말소리를 인공지능 스피커가 인식할 수 있도록 음성인식 시스템의 음성인식 능력을 높이고, 말소리장애가 심한 대상자들에게는 음성인식이 가능할 만큼 구어 능력을 향상시켜 인공지능 음성인식 시스템이 의사소통장애인의 말을 인식할 수 있도록 한다면 말소리장애를 가진 화자들도 음성인식 기능을 이용할 수 있으리라 생각된다.

마비말장애 화자의 수와 음성자료가 적어 본 연구 결과를 모든 말소리장애 음성에 대한 인공지능 스피커의 음성인식 능력으로 확대하여 해석하기에는 제한이 있다. 적은 수의 일반 젊은 청자들에 한정되어 말명료도를 측정하였으므로 마비말장애의 중증도에 따라 음성인식의 차이가 있고 청자마다 마비말장애의 말소리를 인식하는 데 개인차가 컸다. 또한 본 연구는 문장 읽기 상황만을 사용하였으므로 자연스러운 말하기 상황에서의 차이가 있을 수 있다. 그러므로 다양한 말소리장애를 가진 화자의 음성으로 더 많은 수의 청자나 전문가 집단의 청자를 대상으로 발화과제를 달리하여 다른 인공지능 스피커와의 비교 등으로 확대하여 추후 연구를 지속할 필요가 있다.

Reference

- Bae, I. H., Kim, G. H., Park, H. J., Kim, J. D., & Kwon, S. B. (2015). Development of a speech recognition threshold application using voice recognition technology: Pilot study. *Journal of Speech & Hearing Disorders*, 24(4), 103-112. doi:10.15724/jslhd.2015.24.4.010
- Balliti, F., Corno, F., & De Russis, L. (2018). "Hey Siri, do you understand me?" : Virtual assistants and dysarthria. *Intelligent Environments*, 23, 557-566.
- Doyle, P. C., Leeper, H. A., Kotler, A., Thonas-Stonell, N., O'Neill, C., Dylke, M., & Rolls, K. (1997). Dysarthric speech: A comparison of computerized speech recognition and listener intelligibility. *Journal of Rehabilitation Research and Development*, 34(3), 309-316.
- Green, J. R., MacDonald, R. L., Jiang, P., Cattiau, J., Heywood, R., Cave, R., ... & Tomanek, K. (2021). Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases. *Proceedings of Interspeech*, 2021, 4778-4782.
- Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 51(3), 562-573.
- Hwang, Y. M., Kim, S. H., & Chung, M. H. (2011). Comparison between intelligibility and speech recognition accuracy of dysarthric speech. *Proceedings of HCI Korea*, 2011, 355-358. Retrieved from <http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01880425>
- Jeon, J., Lee, H., & Hwang, S. (2019). An AI speech recognition application for the articulation disorder. *Proceedings of the KCC*, 2019, 1854-1856.
- Kim, S. J., & Shin, J. Y. (2020). *Speech sound disorders*. Seoul: Sigmapress.
- Kitzing, P., Maier, A., & Ahlander, V. L. (2009). Automatic Speech Recognition (ASR) and its use as a tool for assessment or therapy of voice, speech, and language disorders. *Logopedics Phoniatrics Vocology*, 34(2), 91-96.
- McHenry, M. A., & LaConte, S. M. (2010). Computer speech recognition as an objective measure of intelligibility. *Journal of Medical Speech-Language Pathology*, 18(4), 99-103.
- Megnistu, K. T., & Rudzicz, F. (2011). Comparing humans and automatic speech recognition systems in recognizing dysarthric speech. *Proceedings of the 24th Canadian Conference on Artificial Intelligence*, St. John's, Canada, 291-300.
- Mustafa, M. B., Rosdi, F., Salin, S. S., & Mughal, M. U. (2015). Exploring the influence of general and specific factors in the recognition accuracy of an ASR system for dysarthric speaker. *Expert Systems with Applications*, 42(8), 3924-3932.
- Schuster, M., Haderlein, T., Noth, E., Lohscheller, J., Eysholdt, U., & Rosanowski, F. (2006). Intelligibility of laryngectomees' substitute speech: Automatic speech recognition and subjective rating. *European Archives of Oto-Rhino-Laryngology and Head & Neck*, 263, 188-193.
- Tu, M., Wisler, A., Berisha, V., & Liss, J. M. (2016). The relationship between perceptual disturbances in dysarthric speech and automatic speech recognition performance. *Journal of Acoustic Society of America*, 140(5), 416-422.

Young, V., & Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and implications for

speech-based applications used by the elderly: A literature review. *Assistive Technology*, 22(2), 99-112.

마비말장애 화자에 대한 인공지능 스피커의 음성인식 정확도와 청자의 말명료도 비교

김주형¹, 김지연^{2*}

¹ 우송정보대학 언어치료과 교수

² 우송대학교 언어치료·청각재활학과 교수

목적: 이 연구는 마비말장애 화자의 음성을 청자의 명료도와 인공지능 스피커의 인식 정확도를 언어학적 단위에 따라 차이가 있는지를 알아봄으로써 인공지능 스피커가 말명료도 평가 및 중재에 활용될 수 있는 기초 자료를 제공하고자 한다.

방법: 마비말장애 화자 10명에게 4개의 문장을 읽게 하고 음성을 녹음하였다. 4개의 문장과 각 문장을 어절로 구분하여 음성파일을 만든 뒤 30명의 청자에게 들려주고 받아 적게 하여 음절 명료도를 구하였고, 5점 척도로 평정하게 하였다. 카카오톡에게 음성을 들려주고 텍스트로 변환하여 인공지능 스피커의 음성인식 정확도를 구하였고 이를 청자의 명료도와 비교하였다.

결과: 첫째, 인공지능 스피커의 음성인식 정확도는 정상 화자에 비해 마비말장애 화자에 대해서는 유의하게 낮은 편이었다. 둘째, 마비말장애 화자 음성에 대한 청자의 음절 명료도가 인공지능 스피커의 음성인식 정확도보다 유의하게 높았다. 사람의 귀는 어절에 비해 문장 단위를 더 잘 알아들었지만 인공지능 스피커는 문장과 어절의 음성인식 정확도에 유의한 차이가 없었다. 셋째, 청자의 명료도 측정 방법에 따른 차이를 살펴본 결과, 척도 명료도와 음절 명료도 간에 유의한 상관관계를 나타냈다. 넷째, 마비말장애 화자의 말소리에 대한 인공지능 스피커의 음성인식 정확도와 청자의 음절 명료도 간에는 유의한 상관을 보였다.

결론: 마비말장애 화자 음성에 대한 인공지능 스피커의 음성인식 정확도가 문장 단위에서는 아직 부족하지만 어절 단위에서는 사람의 귀와 비슷한 양상을 보임으로써 말소리장애 화자의 명료도 평가에 인공지능 스피커가 객관적인 명료도 측정방법의 하나로 사용될 수 있음을 보여주었다.

교신저자: 김지연(우송대학교)

전자메일: bluejiyoun@hotmail.com

게재신청일: 2022. 08. 31

수정제출일: 2022. 09. 29

게재확정일: 2022. 10. 31

이 연구는 2020년 카카오의 지원을 받아 수행된 연구임.

ORCID

김주형

<https://orcid.org/0000-0002-8927-6921>

김지연

<https://orcid.org/0000-0003-4259-778X>

검색어: 마비말장애, 말명료도, 인공지능 스피커, 음성인식

참고 문헌

김수진, 신지영 (2020). **말소리장애**. 서울: 시그마프레스.

배인호, 김근호, 박희준, 김진동, 권순복 (2015). 음성인식을 이용한 어음인지 역치검사 어플리케이션 개발을 위한 기초연구. **언어치료연구**, 24(4), 103-112.

전재민, 이혁, 황성수 (2019). 조음장애인들의 음성인식을 위한 AI 어플리케이션.

2019 **한국컴퓨터종합학술대회 논문집**, 1854-1856.

황유미, 김선희, 정민화 (2011). 마비 말 장애의 발화 명료도와 음성 인식을 비교. 2011 **한국 HCI 학회 학술대회 논문집**, 355-358.
<http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01880425>