Automatic Assessment of Intelligibility in Noise in Parkinson Disease: Validation Study

Gemma Moya-Galé[1*], PhD; Stephen J Walsh[2*], PhD; Alireza Goudarzi[3*], PhD

[1]Department of Communication Sciences & Disorders, Long Island University, Brooklyn, NY, United States

[2]Department of Mathematics and Statistics, Utah State University, Logan, UT, United States

[3]Factorize, Tokyo, Japan

[*]all authors contributed equally

Corresponding Author:
Gemma Moya-Galé, PhD
Department of Communication Sciences & Disorders
Long Island University
1 University Plaza
Brooklyn, NY, 11201
United States
Phone: 1 718 780 4125
Email: gemma.moya-gale@liu.edu

Abstract

Background: Most individuals with Parkinson disease (PD) experience a degradation in their speech intelligibility. Research on the use of automatic speech recognition (ASR) to assess intelligibility is still sparse, especially when trying to replicate communication challenges in real-life conditions (ie, noisy backgrounds). Developing technologies to automatically measure intelligibility in noise can ultimately assist patients in self-managing their voice changes due to the disease.

Objective: The goal of this study was to pilot-test and validate the use of a customized web-based app to assess speech intelligibility in noise in individuals with dysarthria associated with PD.

Methods: In total, 20 individuals with dysarthria associated with PD and 20 healthy controls (HCs) recorded a set of sentences using their phones. The Google Cloud ASR API was used to automatically transcribe the speakers' sentences. An algorithm was created to embed speakers' sentences in +6-dB signal-to-noise multitalker babble. Results from ASR performance were compared to those from 30 listeners who orthographically transcribed the same set of sentences. Data were reduced into a single event, defined as a success if the artificial intelligence (AI) system transcribed a random speaker or sentence as well or better than the average of 3 randomly chosen human listeners. These data were further analyzed by logistic regression to assess whether AI success differed by speaker group (HCs or speakers with dysarthria) or was affected by sentence length. A discriminant analysis was conducted on the human listener data and AI transcriber data independently to compare the ability of each data set to discriminate between HCs and speakers with dysarthria.

Results: The data analysis indicated a 0.8 probability (95% CI 0.65-0.91) that AI performance would be as good or better than the average human listener. AI transcriber success probability was not found to be dependent on speaker group. AI transcriber success was found to decrease with sentence length, losing an estimated 0.03 probability of transcribing as well as the average human listener for each word increase in sentence length. The AI transcriber data were found to offer the same discrimination of speakers into categories (HCs and speakers with dysarthria) as the human listener data.

Conclusions: ASR has the potential to assess intelligibility in noise in speakers with dysarthria associated with PD. Our results hold promise for the use of AI with this clinical population, although a full range of speech severity needs to be evaluated in future work, as well as the effect of different speaking tasks on ASR.

KEYWORDS

XSL•FO
RenderX

Introduction

Parkinson disease (PD) is the second most common neurodegenerative disease, following Alzheimer disease [1]. Approximately 1 million individuals are estimated to be affected by the disease in the United States [2], and its prevalence surpasses 6 million people worldwide [3], with numbers projected to increase in the future [2]. Close to 90% of individuals with PD evidence problems with voice or speech, an impairment known as hypokinetic dysarthria, which has a latency that averages 7 years post–disease onset [4]. This motor speech disorder is characterized by hypophonia (ie, reduced loudness), monopitch, monoloudness, articulatory imprecision, reduced stress, short rushes of speech, and variable rate [5]. As a result, many individuals affected by the disease complain of intelligibility problems (ie, their ability to be understood by others) [6], especially in noisy environments (eg, when dining out at a restaurant). Additionally, the presence of background noise has been shown to negatively affect even speakers with mildly dysarthric speech [7]. Overall, these speech deficits substantially reduce speakers' social participation and overall quality of life [8], as their inability to effectively communicate with others increases their frustration and social isolation.

The application of artificial intelligence (AI) in the medical field has brought promising results to enhance communication and, ultimately, quality of life [9] in a wide range of individuals. For example, voice-assisted technology, which is used in devices such as Siri or Alexa, has become increasingly more present among individuals with a neurodegenerative disease, such as those with PD [10], and has gradually been incorporated as a potential available tool for health professionals, such as speech and language pathologists [11]. The development of automatic speech recognition (ASR) technologies has substantially advanced in the past 40 years, especially given the onset of deep learning mechanisms [12]. Most crucially, the use of ASR has been shown to be effective in estimating speakers' intelligibility deficits for different clinical populations who may present with speech impairments [13], such as those resulting from a laryngectomy [14], a cleft palate [15], or head and neck cancer [16]. Additionally, the clinical validity of ASR has also been explored in individuals with apraxia of speech and aphasia with promising results [17,18]. Project Euphonia has achieved a large-scale data set with over 1 million recordings of disordered speech, with the ultimate goal to personalize ASR models to enhance communication in individuals who experience speech and language difficulties [19,20]. Despite the great advancements that these findings represent, however, research on the application of ASR for individuals with the motor speech disorder of dysarthria has been more limited [21-23], and it has underscored the high degree of variability that characterizes dysarthric speech [13], especially with increased speech severity levels [24]. Dimauro et al [25] explored the use of ASR with 28 individuals with dysarthria associated with PD, 22 healthy older adults, and 15 healthy young controls. In their study, the speech-to-text system focused on the recognition error rates of words from different speech tasks. Although their results upheld the use of AI as a promising resource for clinical populations, it is important to note, however, that their experiment was conducted in quiet conditions, which may not reflect the real-life challenges speakers with PD face in everyday communication. More recently, Gutz et al [26] used the Google Cloud ASR API for intelligibility measurement with 52 speakers with dysarthria associated with amyotrophic lateral sclerosis and 20 healthy controls. Additionally, the authors used noise-augmented ASR to assist the AI system in discriminating between healthy speech and mildly dysarthric speech. Results from their study showed high variability and poor internal validity of machine word recognition rate, suggesting that this technology may have limited clinical applicability for this population at this time.

Our previous pilot work examined ASR performance in multitalker babble noise to measure speech intelligibility from a reading task in 5 speakers with PD and 5 healthy adults [27]. Preliminary results supported the feasibility of AI technologies to simulate real-life challenges posed by ambient noise. Our current study was aimed at expanding our previous work with speakers with dysarthria associated with PD to preliminarily validate the use of ASR in noise with this clinical population. To that end, this study reports on the development, pilot-testing, and validation of a web-based app, *Understand Me for Life* [27], to assess speech intelligibility in noise using the Google Cloud ASR API in speakers with dysarthria associated with PD. Specifically, our aims were to (1) examine how ASR compared to human transcription, the current gold standard, when determining intelligibility accuracy scores for speakers with hypokinetic dysarthria associated with PD; and (2) determine the extent to which ASR could accurately discriminate between speakers with dysarthria and healthy controls.

Methods

Ethics Approval

This study was approved by the Institutional Review Board at Long Island University, Brooklyn (21/01-002-Bkln).

Speakers

In total, 20 individuals with PD (12 women and 8 men; mean age 73.3 years; age range 62-81 years) and 20 age- and sex-matched neurologically healthy adults participated in the speech recordings for this study. Individuals with PD had to meet the following inclusion criteria: (1) having a medical diagnosis of PD, (2) having experienced changes in their voice that represented a current concern, (3) having a stable anti-Parkinsonian medication, (4) passing the Montreal Cognitive Assessment [28], and (5) being a native speaker of English. Exclusion criteria included having received intensive voice-focused treatment in the past 2 years prior to the study and having received deep brain stimulation. Neurologically healthy speakers (12 women and 8 men; mean age 70.5 years; age range 59-84 years) with no history of motor speech impairments served as controls. Table 1 presents the speakers' biographical details and clinical characteristics.

Dysarthria severity ranged from mild to moderate in these speakers and was assessed from a conversation sample by an experienced speech and language pathologist. Consensus with a second speech and language pathologist was obtained for the final dysarthria severity estimates [29].

Table 1. Speakers' biographical details and clinical characteristics.

Speaker	Age (years)	Sex	YPD[a]	Dysarthria severity	Patient's voice complaint
P1[b]	77	Female	9	Mild	Voice is softer and sounds are not as well-articulated
P2	77	Male	1	Mild-moderate	Voice is softer
P3	70	Female	6	Mild	Hoarseness
P4	72	Female	4	Mild	Less control over shaping words, changes in loudness, and occasional rapid breathing
P5	72	Female	7	Mild-moderate	Voice is much lower and softer and reduced intelligibility
P6	80	Female	8	Mild-moderate	Increased fatigue, hoarseness, and lack of clarity
P7	80	Female	8	Mild	Reduced fundamental frequency range for singing and "scratchy feeling" in throat
P8	67	Female	9	Mild-moderate	Lower pitch, hoarseness, voice is much softer, and reduced intelligibility
P9	65	Female	5	Mild	Recent coughing, softness of voice, and voice sounds rougher and softer than usual.
P10	78	Female	7	Mild	Slurring, voice is softer, and intelligibility has been affected.
P11	60	Female	8	Mild	Occasional reduction in loudness
P12	66	Male	7	Mild	Fluctuations in voice and voice is much softer
P14	73	Male	8	Mild	Occasional reduction in loudness and stuttering
P14	80	Female	7	Mild-moderate	Voice is softer
P15	73	Male	13	Mild-moderate	Voice is softer and more strained
P16	78	Male	4	Mild	Voice is softer, trouble finding words, and sometimes intelligibility is affected
P17	62	Male	13	Moderate	Voice is very soft, problems with intelligibility, and fast speaking rate
P18	81	Male	8	Mild-moderate	Voice is softer, breathiness, and have to clear throat more often
P19	80	Female	8	Mild	Voice is softer
P20	76	Male	7	Moderate	Soft voice and hoarseness
HC1[c]	68	Female	N/A[d]	N/A	N/A
HC2	71	Male	N/A	N/A	N/A
HC3	64	Female	N/A	N/A	N/A
HC4	67	Male	N/A	N/A	N/A
HC5	72	Female	N/A	N/A	N/A
HC6	77	Female	N/A	N/A	N/A
HC7	72	Male	N/A	N/A	N/A
HC8	71	Male	N/A	N/A	N/A
HC9	67	Female	N/A	N/A	N/A
HC10	78	Male	N/A	N/A	N/A
HC11	59	Female	N/A	N/A	N/A
HC12	61	Male	N/A	N/A	N/A
HC13	75	Female	N/A	N/A	N/A
HC14	66	Female	N/A	N/A	N/A
HC15	63	Female	N/A	N/A	N/A
HC16	63	Male	N/A	N/A	N/A
HC17	84	Female	N/A	N/A	N/A
HC18	84	Male	N/A	N/A	N/A

Speaker	Age (years)	Sex	YPD[a]	Dysarthria severity	Patient's voice complaint
HC19	65	Female	N/A	N/A	N/A
HC20	83	Female	N/A	N/A	N/A

[a]YPD: years postdiagnosis.

[b]P: patient (speaker with dysarthria associated with Parkinson disease).

[c]HC: healthy control.

[d]N/A: not applicable.

Speech Stimuli and Recording Procedures

A set of 100 grammatically and semantically correct sentences was created for this study. Sentences differed in length, from 5 to 9 words (eg, "Take care of my house while I am away"), and contained high frequency words in the English language (The English Lexicon Project) [30]. The data set was then divided into 4 different blocks of 25 randomized sentences each, with blocks having an equal number of sentences from each sentence length. Each speaker was randomized to 1 block of stimuli for speech recordings, so that each block was read by 10 different speakers. Recordings were self-paced and conducted in a quiet room in the speakers' homes using a customized web-based app, *Understand Me for Life* [27], that the speakers could access from their mobile phones. The first author met with speakers over the Zoom videoconferencing platform (Zoom Video Communications) to explain the recording procedure and address any potential questions. Careful directions were provided to ensure a constant 8-cm (3.15 inches) mouth-to-microphone distance [31,32]. Given the possibility of PD-related motor impairments hindering adequate recordings (eg, tremors), care partners were recruited to assist speakers when necessary. Speakers were allowed to rerecord a sentence in cases of extraneous noise in the background. A brief familiarization phase was provided at the beginning of the recording session so that speakers could practice using the interface. Feedback from speakers was obtained for later app optimization.

For each recorded sentence, the app automatically embedded the speakers' voice signal into +6-dB signal-to-noise multitalker babble noise [33] to provide an intelligibility score, defined as the percentage of words accurately understood by the ASR system. Automatic feedback on performance was provided at the end of the recording session and not after each sentence to avoid any potential priming effects that could influence sentence production on subsequent items [34].

Multitalker Babble Noise

Multitalker babble is thought to be the most common type of environmental noise experienced by listeners [35], which, therefore, makes it more ecologically valid in speech perception experiments. For this study, 10-second sample recordings from National Public Radio were used. Audio files were manually checked to control for sudden changes in the speech signal (eg, increase in vocal intensity). Prolonged silences (ie, over 500 ms) were trimmed, followed by the equalization of the audio spectrum in a moving window. An equal number of male and female speakers was implemented in the creation of background noise [36]. The equalized audios were finally combined to render 10-talker babble [33].

Listeners

In total, 30 neurologically healthy adults (25 women and 5 men; mean age 23.1 years; age range 18-31 years) participated as listeners in the study. Listeners were recruited via flyers and word of mouth across the New York City area. Inclusion criteria for participation required listeners to be native speakers of English; have no history of speech, language, or communication impairment; have no prior experience with motor speech disorders; and pass a bilateral pure-tone hearing screening at 25-dB hearing level at 500, 1000, 2000, and 4000 Hz [37]. Listeners were paid US $20 for their participation in the study.

Human Transcription

Listeners completed the intelligibility assessment task free field (ie, without headphones) in a quiet space at the Long Island University campus, in Brooklyn, New York. The task was accessible through the *Understand Me for Life* portal on a MacBook Pro laptop (Apple Inc). Listeners maintained a distance of 85 cm from the loudspeakers (Logitech Z150), and the loudspeakers were placed 31 cm from each other. Listener-to-loudspeaker distance represented the typical distance between conversational partners [38]. The task took approximately 30-40 minutes to complete.

A brief familiarization phase was presented before the start of the experiment and contained 3 sentences produced by a neurologically healthy adult male speaker. Listeners were instructed to write down word by word what they heard and not worry about punctuation marks. Each listener was randomly assigned to 1 speaker per block, with block presentation being random across listeners. Therefore, each listener heard a total of 4 speakers and 100 sentences. Sentences were presented in multitalker babble, hence replicating the AI condition. To avoid abrupt onsets and offsets of stimuli, 400 ms of noise were inserted at the beginning of each sentence, and each sentence was followed by 50 ms of babble noise [39]. To obtain an average score for subsequent transcription accuracy calculations, each speaker was assigned to 3 listeners. None of the listeners required a break during the completion of this task.

Data Analysis

Automatic Intelligibility Assessment

Automatic intelligibility assessment (AIA) was conducted using the Google Cloud ASR API, a speech-to-text AI system with documented low word error rate for individuals with healthy speech that is thought to be the best platform to handle dysarthric speech, although software performance is still dependent on speech severity, with high word error rates in cases of more severely affected speech [40].

For a given produced utterance (*S*) and the corresponding target sentence (*T*), stimuli were suitably padded with whitespace to ensure that both *S* and *T* were of equal length (*L*). Each word in *S* was codified with w_s and each word in *T* with w_t, where *s* and *t* were numbers from 0 to *L* − 1. Accuracy was calculated by the formula as follows:

$$f(S,T) = argmax_{s,t} \frac{100}{L} \sum_{s=0}^{L-1} \sum_{t=0}^{L-1} \sigma(w_s, w_t) \qquad (1)$$

where $\sigma(w_s, w_t) = 1$ if $w_s = w_t$, and 0 otherwise. This step was implemented to avoid providing a score to words that appeared in both *S* and *T* but were out of order [27].

Manual Intelligibility Assessment

Transcription accuracy scores were calculated as the percentage of words correctly transcribed. Orthographic transcriptions are considered the most objective measure to assess intelligibility in dysarthria [33]. Listeners' orthographic transcripts had to match the target to be accepted as correct [32,41]. Obvious spelling errors or errors involving homonyms did not impact calculation scores and were assessed as correct responses. Omissions or additions of morphemes (eg, flower for flowers) were coded as an error.

Statistical Analysis

The goal of the first phase of statistical analysis was to assess the degree to which the AIA could score as well or better than the average human transcriber (ie, listener). As described above, 3 listeners orthographically transcribed sentences from the same speakers, and their data were condensed into a *percentage accuracy* measure for each sentence, which summarized the percentage of words the human listener correctly transcribed. For each question, the average percentage accuracy, denoted as $\hat{a}_{ij, \text{human avg}}$, was computed for each sentence *j* within each speaker *i* to reduce intralistener variability. The AIA system also received a percentage accuracy measure for each sentence or speaker, which we denoted as $\hat{a}_{ij, \text{AIA}}$. The success of the AIA system was defined as follows:

$$S_{ij} := \begin{cases} 1 & \text{if } \hat{a}_{ij,\text{AIA}} \geq \hat{a}_{ij,\text{human avg}} \\ 0 & \text{else} \end{cases} \qquad (2)$$

The AIA system was considered to give a successful transcription if its percentage accuracy score was at least as good as the average of the human listeners' accuracies for sentence *j* within each speaker *i*. The data were then condensed up to the speaker level by computing the proportion of successes of the AIA system over the *j* = 1 , ... , 25 sentences read by speaker *i* as follows:

$$\hat{p}_i = \frac{1}{25} \sum_{i=1}^{25} S_{ij} \qquad (3)$$

This procedure provided an estimate of the probability of success of the AIA system transcription for randomly selected speakers. Standard binomial statistics were used to quantify uncertainty in this analysis and present the results with appropriate statistical summaries and CIs. We investigated whether data provided evidence that the AIA transcriber success differed whether the system was transcribing a healthy control (HC) or a speaker with dysarthria associated with PD and whether sentence length had an effect on AIA success, via a logistic regression analysis.

The goal of the second phase of statistical analysis was to compare the ability of the resulting AIA transcription data summaries to discriminate between healthy controls and speakers with dysarthria. To investigate this goal, we applied linear discriminant analysis to identify optimal discrimination thresholds for both the listener transcriptions and the AIA transcriptions and summarized the discrimination ability of each via typical confusion matrices and correct percentage classification summaries. All statistical analyses were conducted in R statistical software (version 4.1.1; R Foundation for Statistical Computing) [42] and a discriminant and classification analysis was conducted via the *lda* function in the *MASS* package [43].

Intralistener reliability was assessed via percentage agreement on several (approximately 10) duplicate speaker sentences. Interlistener reliability was controlled for in this assessment by condensing each of the 3 listeners' percentage accuracy measures for each speaker or sentence into the average.

Results

A summary of intrarater reliability is shown in Figure 1. The average percentage agreement of repeated responses of this study's listeners was 80%.

The success summaries of the AIA transcriber at the speaker level are presented in Figure 2. The figure shows estimates of the probability of success for each speaker (ordered by score) with a 95% CI. The mean probability of success is indicated by the red horizontal line. The figure illustrates that the expected success probability of the AIA transcriber for a randomly selected speaker was approximately 0.8 (95% CI 0.65-0.91), with the AIA system scoring 80% of target sentences as well or better than the human transcribers for half (22/40, 55%) of the study's speakers. The success probability estimates stratified by speaker group (HC or speaker with dysarthria) are shown in Figure 3. The figure suggests that the AIA transcriber had a slightly more difficult time accurately transcribing the sentences read by speakers with dysarthria, with a slight decline in the estimate of probability of success for speakers #14, #18, and #19.

Figure 1. Distribution of intrarater percentage agreement across the 30 listeners.
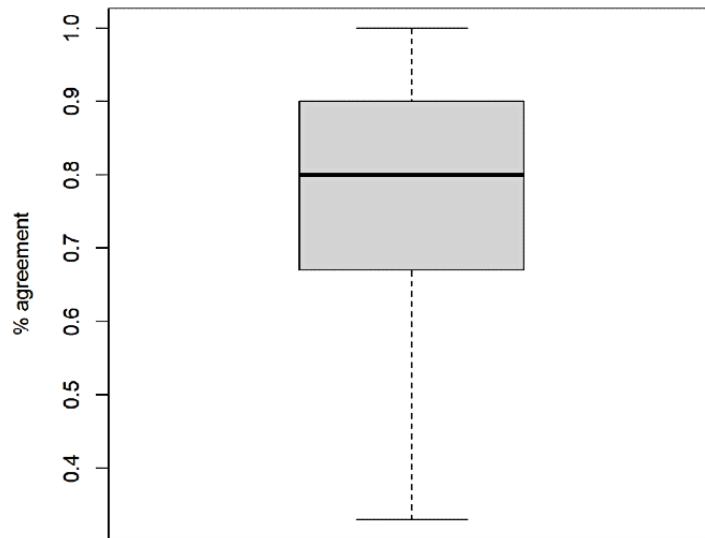

Figure 2. Estimates of the probability that the automatic intelligibility assessment transcriber will be as accurate as human transcribers for each speaker. The vertical bands are 95% CIs on the estimate of probability of success. Black dotted line=0.5 and red dotted line=median AI probability of success. AI: artificial intelligence; C: control; P: patient with dysarthria.
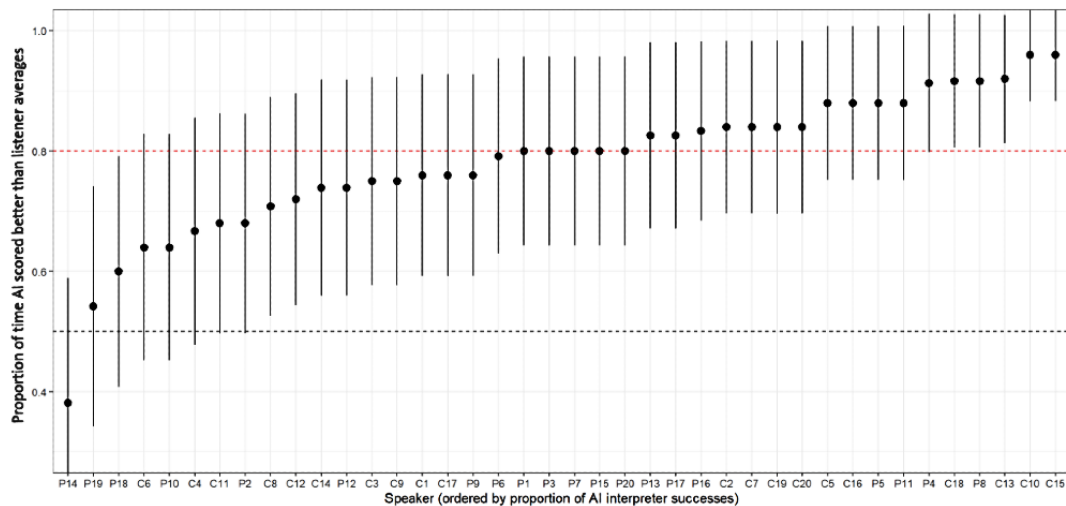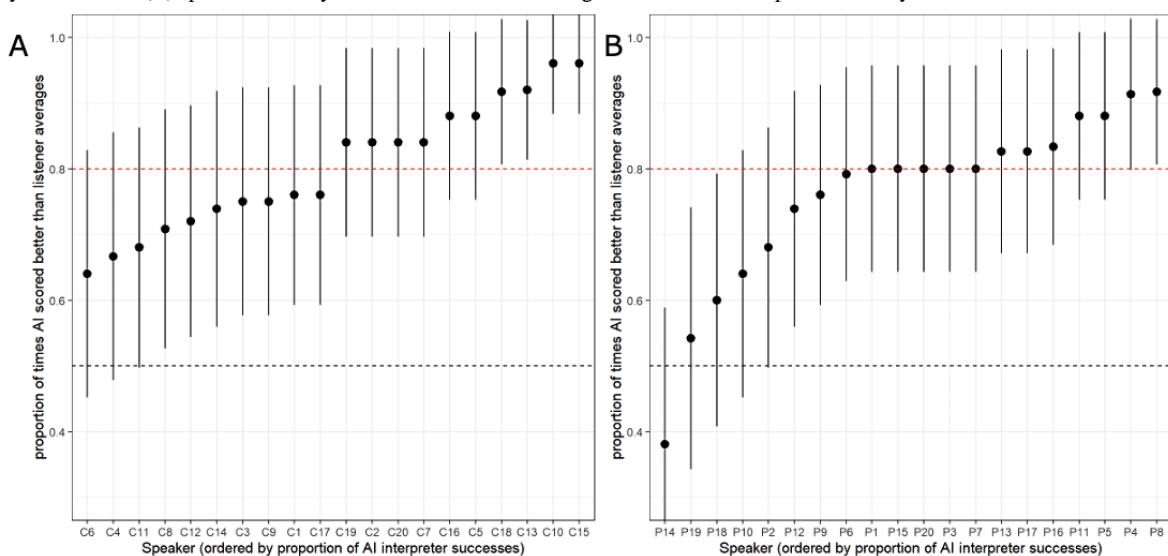

Figure 3. Estimates of the probability that the automatic intelligibility assessment transcriber will be as accurate as human transcribers for each speaker: (A) healthy controls and (B) speakers with dysarthria. AI: artificial intelligence; C: control; P: patient with dysarthria.

We further analyzed these data via a logistic regression model. The response was the (logit) probability of AI success and the predictors were speaker group (HC or speakers with dysarthria) and sentence type. Speaker-to-speaker variance was controlled for by including speaker as a random effect. The fitted model estimates are presented in Table 2. The advantage of this approach is that each row provides a significance test for each term *provided we have controlled for the effects of the other terms*. In this regard, after controlling for speaker and sentence length, we see that these data provide weak evidence that AI success differs significantly by speaker group (ie, between HC and speakers with dysarthria; *P*=.23). Further, sentence length was found to have a significant negative impact on AI success (*P*<.001). The results are represented in an effects plot in Figure 4. The left panel illustrates that an estimate of the probability of AI success for speakers with dysarthria is 0.78, but this value is not significantly different from the estimate of the probability

of AI success for HCs (0.82; *P*=.23). The right panel illustrates an estimated dependence of the probability of AI success on sentence length, with each increase in sentence length decreasing AI success probability by an estimated 0.03.

Percentage accuracy distributions by transcriber (human or AIA system) and speaker group are presented in Figure 5. The box plots in Figure 5 indicate that the median accuracy score for speakers with dysarthria was farther from the median accuracy score for healthy controls as compared to the distance between the 2 medians for the human transcriber data. This finding suggests that the AIA system data may offer better discrimination and classification ability for speaker group.

Confusion matrices recording the classification rates of discriminants based on human transcription data and AIA system data are presented in Table 3.

Table 2. Fitted logistic regression model coefficients.

Effect	Estimate	SE	*z* value	*P* value
Intercept	3.14414	0.44774	7.022	<.001
Speaker group	−0.25525	0.21156	−1.207	.23
Sentence length	−0.23658	0.05763	−4.105	.001

Figure 4. Estimated effects and CIs from the logistic regression of probability of AI success as a function of (A) speaker group, (B) sentence length, and speaker random effect. AI: artificial intelligence; HC: healthy controls.
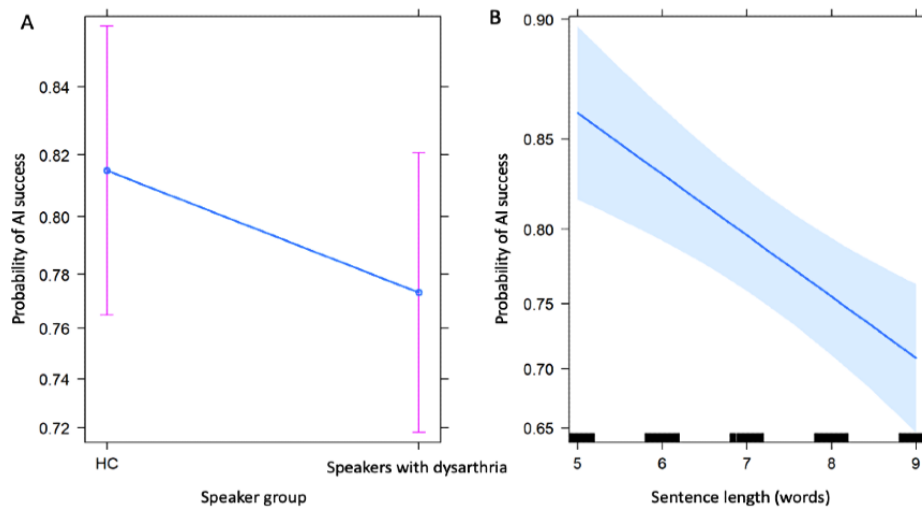
Figure 5. Box plots of the estimates of AIA system success by speaker category and transcriber: (A) human listener and (B) AIA system. AI: artificial intelligence; AIA: automatic intelligibility assessment; HC: healthy controls.
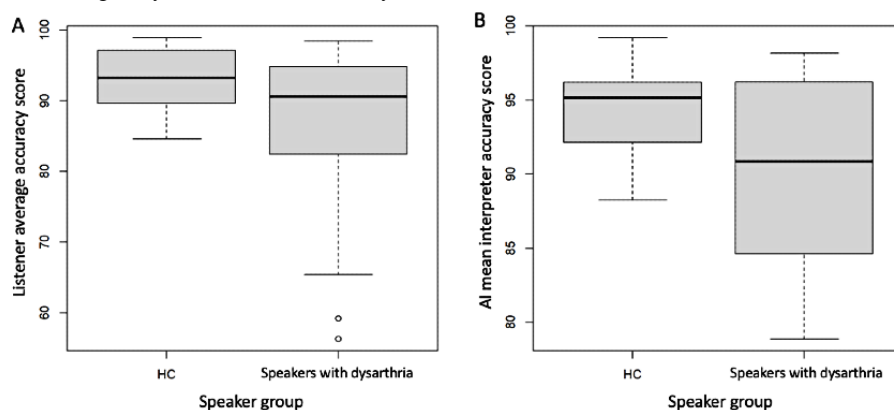
Table 3. Classification summary of the speakers based on linear discriminants fit to the human transcription data and automatic intelligibility assessment system data.

True group	Classified group via discriminant			
	Discriminant from human listener average data (overall predictive accuracy: 0.6)		Discriminant from artificial intelligence data (overall predictive accuracy: 0.675)	
	HC[a]	PD[b]	HC	PD
HC	15	5	15	5
PD	11	9	8	12

[a]HC: healthy control.

[b]PD: Parkinson disease.

Discussion

Principal Findings

This study aimed to develop, pilot-test, and validate the use of a web-based app, *Understand Me for Life*, to automatically measure speech intelligibility in noise in speakers with hypokinetic dysarthria associated with PD. Additionally, a secondary objective of the study was to determine whether ASR could discriminate between the speech of healthy controls and that of speakers with dysarthria.

Literature on ASR performance on clinical populations, especially those with motor speech disorders, is still sparse. To validate the use of speech-to-text technology to determine intelligibility accuracy scores for speakers with dysarthria, ASR performance was benchmarked relative to that of human transcribers [19]. Results showed that the ASR system had an 80% chance of performing as well as or better than a human transcriber on any random speaker. The potential capacity of ASR to outperform human listeners has been shown in recent studies [19], although further work is required with longer utterances and different speech tasks, as summarized in the limitations section below. Our findings also echo those reported with other clinical populations, such as those with a diagnosis of apraxia of speech and aphasia [17,18]. Additionally, our data provided no evidence that the mean probability of ASR success differed between the 2 groups of speakers, either a speaker with dysarthria or a healthy control. Thus, the success of the speech-to-text system did not depend on whether the speaker was neurologically healthy or presented with hypokinetic dysarthria associated with PD. It is important to acknowledge, however, that our speakers did not evidence dysarthria across all severity ranges; this limitation will be addressed in future work. Sentence length did influence ASR, with a decrease in accuracy observed for longer sentences, which was an expected result and is in agreement with prior literature [19,26].

The second aim of the study was to determine whether ASR could accurately discriminate between speakers with dysarthria and healthy controls. Results showed that both the human and the AIA system data provided the same classification rates for healthy controls (15/20, 75% correctly classified and 5/20, 25% incorrectly classified as speakers with dysarthria), hence evidencing equal specificity (ie, 75%). The AIA system data, however, yielded a slightly better classification success for speakers with dysarthria (12/20, 60% correct PD classifications compared to the human transcription data that only yielded 9/20, 45% correct PD classifications), which suggests stronger sensitivity than the one obtained for human transcribers (ie, 60% vs 45%). In traditional studies using human listeners, performance on intelligibility assessments has not shown significant differences between speakers with mild dysarthria secondary to PD and healthy controls [33], hence suggesting that group classification based on intelligibility scores may depend on speech severity. In our study, AI correctly classified 12 speakers with dysarthria (out of 20), a result that could be explained by the severity levels of our sample ranging from mild to mild-to-moderate only.

Limitations and Future Work

The study's limitations warrant future work in this research area. It should be noted that our sample of speakers with dysarthria did not include those with more severe speech deficits. Therefore, these results offer a preliminarily promising, albeit not conclusive, clinical tool for measuring intelligibility in individuals with dysarthria associated with PD. Nevertheless, ASR performance with a more diverse speech severity range in speakers with dysarthria associated with PD should be explored. It is likely that increased speech severity in individuals with PD would impact ASR, as this increase was also found in speakers with dysarthria associated with amyotrophic lateral sclerosis [26]. An additional limitation from this study is that the speech stimuli were derived from read sentences rather than from conversational speech. Although sentences rendered a higher level of predictability and, thus, control, conversational speech would have greater ecological validity. Finally, we should also acknowledge that previously reported studies used different ASR methodology compared to this study and that, as discussed in Jacks et al [18], ASR technology is in constant and rapid evolution, rendering any results on ASR in need of systematic reevaluation for the proper and valid use of ASR-assisted clinical tools.

Our ongoing work is motivated by the concept of self-management, which, in the context of a chronic illness such as PD, has become increasingly relevant. Self-management relates to the patient's ability to identify a given behavior (eg, voice changes) and react or problem-solve in accordance with such observation [44]. Having the knowledge on how to respond to the worsening of disease symptoms and when to seek medical advice has been shown to be crucial contributors to patients' well-being [45]. The implementation of ASR in speech intelligibility assessment, therefore, can potentially serve to

establish preventative measures before the onset of speech and intelligibility degradation and control measures (eg, referral to a speech therapist) if speech deficits already exist.

Conclusions

This study validated the use of ASR to measure intelligibility in real-life settings (ie, using background noise) in speakers with mild-to-moderate dysarthria associated with PD. Therefore, our preliminary data show that ASR has the potential to assess intelligibility in noise in this clinical population. Results hold promise for the use of AI as a future clinical tool to assist patients and speech and language therapists alike, although the full range of speech severity needs to be evaluated in future work, as well as the effect of different speaking tasks on ASR.

Conflicts of Interest

None declared.

References

1. Dorsey ER, Constantinescu R, Thompson JP, Biglan KM, Holloway RG, Kieburtz K, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 2007 Jan 30;68(5):384-386. [doi: 10.1212/01.wnl.0000247740.47667.03] [Medline: 17082464]

2. Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, Parkinson's Foundation P4 Group. Prevalence of Parkinson's disease across North America. NPJ Parkinsons Dis 2018 Jul 10;4:21 [FREE Full text] [doi: 10.1038/s41531-018-0058-0] [Medline: 30003140]

3. Dorsey ER, Sherer T, Okun MS, Bloem BR. The emerging evidence of the Parkinson pandemic. J Parkinsons Dis 2018 Dec 18;8(s1):S3-S8 [FREE Full text] [doi: 10.3233/JPD-181474] [Medline: 30584159]

4. Müller J, Wenning GK, Verny M, McKee A, Chaudhuri KR, Jellinger K, et al. Progression of dysarthria and dysphagia in postmortem-confirmed Parkinsonian disorders. Arch Neurol 2001 Feb 01;58(2):259-264. [doi: 10.1001/archneur.58.2.259] [Medline: 11176964]

5. Duffy JR. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. 4th ed. Amsterdam, the Netherlands: Elsevier Mosby; 2020.

6. Moya-Galé G, Rossi A, States RA. A community-based program for exercise and social participation for individuals with Parkinson's disease: a multidisciplinary model. Perspect ASHA SIGs 2020 Oct 23;5(5):1290-1296. [doi: 10.1044/2020_persp-20-00031]

7. Chiu Y, Forrest K. The impact of lexical characteristics and noise on intelligibility of Parkinsonian speech. J Speech Lang Hear Res 2018 Apr 17;61(4):837-846. [doi: 10.1044/2017_JSLHR-S-17-0205] [Medline: 29587306]

8. McAuliffe MJ, Baylor CR, Yorkston KM. Variables associated with communicative participation in Parkinson's disease and its relationship to measures of health-related quality-of-life. Int J Speech Lang Pathol 2017 Aug 27;19(4):407-417 [FREE Full text] [doi: 10.1080/17549507.2016.1193900] [Medline: 27347713]

9. Derosier R, Farber RS. Speech recognition software as an assistive device: a pilot study of user satisfaction and psychosocial impact. Work 2005;25(2):125-134. [Medline: 16131742]

10. Duffy O, Synnott J, McNaney R, Brito Zambrano P, Kernohan WG. Attitudes toward the use of voice-assisted technologies among people with Parkinson disease: findings from a web-based survey. JMIR Rehabil Assist Technol 2021 Mar 11;8(1):e23006 [FREE Full text] [doi: 10.2196/23006] [Medline: 33704072]

11. Kulkarni P, Duffy O, Synnott J, Kernohan WG, McNaney R. Speech and language practitioners' experiences of commercially available voice-assisted technology: web-based survey study. JMIR Rehabil Assist Technol 2022 Jan 05;9(1):e29249 [FREE Full text] [doi: 10.2196/29249] [Medline: 34989694]

12. Kodish-Wachs J, Agassi E, Kenny PIII, Overhage JM. A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech. AMIA Annu Symp Proc 2018 Dec 05;2018:683-689 [FREE Full text] [Medline: 30815110]

13. Tu M, Wisler A, Berisha V, Liss JM. The relationship between perceptual disturbances in dysarthric speech and automatic speech recognition performance. J Acoust Soc Am 2016 Nov;140(5):EL416-EL422 [FREE Full text] [doi: 10.1121/1.4967208] [Medline: 27908075]

14. Schuster M, Haderlein T, Nöth E, Lohscheller J, Eysholdt U, Rosanowski F. Intelligibility of laryngectomees' substitute speech: automatic speech recognition and subjective rating. Eur Arch Otorhinolaryngol 2006 Feb 7;263(2):188-193. [doi: 10.1007/s00405-005-0974-6] [Medline: 16001246]

15. Maier A, Haderlein T, Stelzle F, Nöth E, Nkenke E, Rosanowski F, et al. Automatic speech recognition systems for the evaluation of voice and speech disorders in head and neck cancer. EURASIP J Audio Speech Music Process 2010 Dec 01;2010(1):1-7 [FREE Full text]

XSL·FO

RenderX

16. Maier A, Nöth E, Nkenke E, Schuster M. Automatic assessment of children's speech with cleft lip and palate. 2006 Oct Presented at: IS-LTC 2006: 5th Slovenian and 1st International Conference on Language Technologies; October 9-10, 2006; Ljubljana, Slovenia URL: http://nl.ijs.si/is-ltc06/proc/06_Maier.pdf

17. Ballard KJ, Etter NM, Shen S, Monroe P, Tien Tan C. Feasibility of automatic speech recognition for providing feedback during tablet-based treatment for apraxia of speech plus aphasia. Am J Speech Lang Pathol 2019 Jul 15;28(2S):818-834. [doi: 10.1044/2018_AJSLP-MSC18-18-0109] [Medline: 31306595]

18. Jacks A, Haley KL, Bishop G, Harmon TG. Automated speech recognition in adult stroke survivors: comparing human and computer transcriptions. Folia Phoniatr Logop 2019 May 22;71(5-6):286-296 [FREE Full text] [doi: 10.1159/000499156] [Medline: 31117105]

19. Green JR, MacDonald RL, Jiang PP, Cattiau J, Heywood R, Cave R, et al. Automatic speech recognition of disordered speech: personalized models now outperforming human listeners on short phrases. 2021 Presented at: Interspeech 2021; August 30 to September 3, 2021; Brno, Czechia p. 4778-4782. [doi: 10.21437/interspeech.2021-1384]

20. MacDonald RL, Jiang PP, Cattiau J, Heywood R, Cave R, Seaver K, et al. Disordered speech data collection: lessons learned at 1 million utterances from Project Euphonia. 2021 Presented at: Interspeech 2021; August 30 to September 3, 2021; Brno, Czechia p. 4833-4837. [doi: 10.21437/interspeech.2021-697]

21. Christensen H, Cunningham S, Fox C, Green P, Hain T. A comparative study of adaptive, automatic recognition of disordered speech. 2012 Presented at: Interspeech 2012: 13th Annual Conference of the International Speech Communication Association (ISCA); September 9-13, 2012; Portland, OR p. 1776-1779. [doi: 10.21437/interspeech.2012-484]

22. Sharma HV, Hasegawa-Johnson M. Acoustic model adaptation using in-domain background models for dysarthric speech recognition. Comput Speech Lang 2013 Sep;27(6):1147-1162. [doi: 10.1016/j.csl.2012.10.002]

23. Vásquez-Correa JC, Orozco-Arroyave JR, Bocklet T, Nöth E. Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. J Commun Disord 2018 Nov;76:21-36. [doi: 10.1016/j.jcomdis.2018.08.002] [Medline: 30149241]

24. Le D, Licata K, Mower Provost E. Automatic quantitative analysis of spontaneous aphasic speech. Speech Commun 2018 Jun;100:1-12. [doi: 10.1016/j.specom.2018.04.001]

25. Dimauro G, Di Nicola V, Bevilacqua V, Caivano D, Girardi F. Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. IEEE Access 2017 Oct 17;5:22199-22208. [doi: 10.1109/access.2017.2762475]

26. Gutz SE, Stipancic KL, Yunusova Y, Berry JD, Green JR. Validity of off-the-shelf automatic speech recognition for assessing speech intelligibility and speech severity in speakers with amyotrophic lateral sclerosis. J Speech Lang Hear Res 2022 Jun 08;65(6):2128-2143. [doi: 10.1044/2022_JSLHR-21-00589] [Medline: 35623334]

27. Goudarzi A, Moya-Galé G. Automatic speech recognition in noise for Parkinson's disease: a pilot study. Front Artif Intell 2021 Dec 22;4:809321 [FREE Full text] [doi: 10.3389/frai.2021.809321] [Medline: 35005616]

28. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]

29. Fletcher AR, McAuliffe MJ, Lansford KL, Liss JM. Assessing vowel centralization in dysarthria: a comparison of methods. J Speech Lang Hear Res 2017 Feb 01;60(2):341-354 [FREE Full text] [doi: 10.1044/2016_JSLHR-S-15-0355] [Medline: 28124069]

30. Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, et al. The English Lexicon Project. Behav Res Methods 2007 Aug;39(3):445-459. [doi: 10.3758/bf03193014] [Medline: 17958156]

31. Levy ES, Moya-Galé G, Chang YM, Campanelli L, MacLeod AAN, Escorial S, et al. Effects of speech cues in French-speaking children with dysarthria. Int J Lang Commun Disord 2020 May 20;55(3):401-416. [doi: 10.1111/1460-6984.12526] [Medline: 32077196]

32. Moya-Galé G, Keller B, Escorial S, Levy ES. Speech treatment effects on narrative intelligibility in French-speaking children with dysarthria. J Speech Lang Hear Res 2021 Jun 18;64(6S):2154-2168. [doi: 10.1044/2020_JSLHR-20-00258] [Medline: 33719503]

33. Chiu YF, Neel A. Predicting intelligibility deficits in Parkinson's disease with perceptual speech ratings. J Speech Lang Hear Res 2020 Feb 26;63(2):433-443. [doi: 10.1044/2019_JSLHR-19-00134] [Medline: 32097080]

34. Maas E, Robin DA, Austermann Hula SN, Freedman SE, Wulf G, Ballard KJ, et al. Principles of motor learning in treatment of motor speech disorders. Am J Speech Lang Pathol 2008 Aug;17(3):277-298. [doi: 10.1044/1058-0360(2008/025)] [Medline: 18663111]

35. Fontan L, Tardieu J, Gaillard P, Woisard V, Ruiz R. Relationship between speech intelligibility and speech comprehension in babble noise. J Speech Lang Hear Res 2015 Jun;58(3):977-986. [doi: 10.1044/2015_JSLHR-H-13-0335] [Medline: 25809922]

36. Moya-Galé G, Goudarzi A, Bayés À, McAuliffe M, Bulté B, Levy ES. The effects of intensive speech treatment on conversational intelligibility in Spanish speakers with Parkinson's disease. Am J Speech Lang Pathol 2018 Feb 06;27(1):154-165. [doi: 10.1044/2017_AJSLP-17-0032] [Medline: 29351354]

37. ANSI S3.6-2004: specifications for audiometers. American National Standards Institute. 2004 May 13. URL: https://webstore.ansi.org/Standards/ASA/ansis32004 [accessed 2022-10-03]

38.   Hall ET. The hidden dimension. Leonardo 1973;6(1):94. [doi: 10.2307/1572461]

39.   Levy ES, Moya-Galé G, Chang YHM, Freeman K, Forrest K, Brin MF, et al. The effects of intensive speech treatment on intelligibility in Parkinson's disease: a randomised controlled trial. EClinicalMedicine 2020 Jul;24:100429 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100429] [Medline: 32639484]

40.   De Russis L, Corno F. On the impact of dysarthric speech on contemporary ASR cloud platforms. J Reliable Intell Environ 2019 Jul 6;5(3):163-172. [doi: 10.1007/s40860-019-00085-y]

41.   Cannito MP, Suiter DM, Beverly D, Chorna L, Wolf T, Pfeiffer RM. Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. J Voice 2012 Mar;26(2):214-219. [doi: 10.1016/j.jvoice.2011.08.014] [Medline: 22209057]

42.   R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.r-project.org/ [accessed 2022-10-03]

43.   Venables WN, Ripley BD. Modern Applied Statistics with S (MASS). 4th ed. New York, NY: Springer; 2002.

44.   Lorig K. Self-management of chronic illness: a model for the future. Generations 1993 Oct 03;17(3):11-14 [FREE Full text]

45.   Hayes C. Identifying important issues for people with Parkinson's disease. Br J Nurs 2002 Jan 17;11(2):91-97. [doi: 10.12968/bjon.2002.11.2.9309] [Medline: 11823736]

Abbreviations

AI: artificial intelligence
AIA: automatic intelligibility assessment
ASR: automatic speech recognition
HC: healthy control
PD: Parkinson disease