JSLHD 2025. Vol. 34, No. 1. 075-085
ISSN 1226-587X / eISSN 2671-7158
Journal of Speech-Language & Hearing Disorders 언어치료연구

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2025.34.1.075
http://jslhd.org/

Speech Characteristics in Detecting Cognitive Decline

Mai Ahmed[1], Soon Bok Kwon[2*]

[1] *Dept. of Linguistics, Pusan National University, Doctor's Course, Korean Language and Literature Dept Ain Shams University, Assistant Teacher*
[2] *Dept. of Language and Information, Pusan National University, Professor*

Purpose: Cognitive health affects production of speech as the speech production process relies on memory systems for word retrieval and sentence composition. This study aims to explore the relationship between temporal acoustic features of speech and cognitive decline.
Methods: Speech samples were collected from 32 elderly individuals at different stages of cognitive health and divided into three groups: normal cognitive (n=13), mild cognitive impairment (n=10), and dementia (n=9). Mean age of population was 83 years. And speech samples were collected using two speech tasks: Picture description task, and Interview task. The samples were then analyzed using Praat to derive acoustic and suprasegmental parameters. Extracted variables were finally analyzed using one-way ANOVA to identify statistical significance of differences and connections between specific speech features and cognitive decline.
Results: The findings indicate that temporal feature changes are perceptible in the early stages of cognitive decline, especially articulation rate, speech rate, and filler duration. Acoustic features, on the other hand, did not show any significant connection to cognitive decline except for intensity (SPL [dB]). The results also show that the type of speech task affects the significance of the speech variable.
Conclusions: Speech suprasegmental parameters are detectable in the speech of cognitive decline, even in the early stages. Future studies should consider examining other speech characteristics related to semantics and their relationship with prosodic features and cognitive ability.

Keywords: Cognitive decline, mild cognitive impairment, speech analysis, acoustic features of speech, suprasegmental parameters of speech

Correspondence : Soon Bok Kwon, PhD
E-mail : sbkwon@pusan.ac.kr

Received : November 27, 2024
Revision revised : December 30, 2024
Accepted : January 31, 2025

This work was supported by (2024) Specialization Project of Pusan National University

ORCID
Mai Ahmed
https://orcid.org/0000-0003-2470-9666
Soon Bok Kwon
https://orcid.org/0000-0002-9424-0077

Ⅰ. Introduction

Dementia (De) presents a significant challenge in the current times where increased life expectancy is increasing and leading to a growing elderly population. This demographic shift in population necessitates greater attention to the unique needs of older adults, particularly in addressing cognitive decline, a key concern of aging. Since there is no definitive cure for cognitive decline (especially Alzheimer [AD]) to this day, the best course of action is the early detection of cognitive decline and the implementation of memory training intervention methods in hopes of delaying or managing the decline process (Riley et al., 2022).

Given the importance of early detection, a lot of research focus is now directed at finding a cost-effective non-invasive way of cognitive decline detection. Research in this topic covers a wide area from motor skill to eye movement to speech detection. The latter of which is what this paper is concerning with, that is the detection of cognitive decline in elderly speech through analyzing acoustic and suprasegmental parameters of speech. Since speech signals can be considered as biomarkers of the progression of cognitive decline as Ahmed et al. (2013), König et al. (2015), and Fraser et al. (2016) pointed out, many studies have tried to analyze speech voice signal characteristics that are statistically significant in distinguishing between different cognitive decline stages and normal cognitive health as well (López-de-Ipiña et al., 2013; Meilán et al., 2020; Nagumo et al., 2020; Sumali et al., 2020; Wang et al., 2022). Spontaneous speech, in particular, offers a unique perspective on cognitive health, as it demands memory retrieval and semantic and syntactic clarity (De Looze et al., 2021).

While previous research has advanced automated detection systems for cognitive decline using machine learning, these studies often prioritize the development of technological tools over in-depth analysis of the specific effects of cognitive decline on speech production. Consequently, there remains a gap in understanding the acoustic and prosodic characteristics associated with cognitive decline and how these features vary across its stages.

Existing literature has reported on the speech features that were the most statistically significant when automatically sorting the speech of the cognitively impaired. For speech temporal features, Sadeghian et al. (2021) pointed out the features such as, proportion of pause length and the proportion of phonation length to the whole audio along with speech rate. Silent or pause-related features were similarly highlighted by König et al. (2015), Toth et al. (2018), Da Cunha et al. (2022), and Liu et al. (2023). While Martínez-Nicolás et al. (2022) and Yamada et al. (2021) have also added phonation measures as well as a significant feature.

Regarding acoustic features on the other hand, Themistocleous et al. (2020) has mentioned that shimmer, center of gravity and cepstral peak prominence are also significant in differentiating healthy and cognitively declining groups. And Hall et al. (2019) has also mentioned jitter and shimmer, while Gonzalez-Moreira et al. (2015) pointed to the significance of standard deviation and mean of fundamental frequency. It is note worthy to mention though that the first two research have also pointed out suprasegmental features along with the acoustic features for the automatic screening of cognitive decline. Indeed, Ahmed & Kwon (2024) have concluded that temporal features were more reported to be significant in the cognitive decline screening more than acoustic variables. This is due to the association of cognitive decline with disfluency in speech which results in more silent or pause segments.

This paper seeks to bridge the gap in the literature by moving beyond machine learning applications to conduct a detailed analysis of the acoustic and prosodic speech characteristics linked to cognitive decline. The study aims to identify which speech features are most indicative of cognitive decline across its different stages and to determine their relative significance in the context of early detection and assessment. A particularly novel aspect of this research is its focus on Korean native speakers, a population for which studies on the speech characteristics of cognitive decline remain scarce.

Previous research on Korean speakers, such as Ha et al. (2009), has focused on speech disfluencies (e.g., fillers and substitutions), while Choi et al. (2013) concentrated on linguistic features of participants' speech. More recently, Park et al. (2024) incorporated acoustic analysis into their study; however, they did not highlight the significance of the specific speech measures analyzed, nor did they account for prosodic characteristics—factors frequently reported in the international literature as being strongly associated with cognitive decline. By addressing these gaps, this study not only broadens the understanding of cognitive decline's impact on speech production but also contributes valuable insights specific to Korean-speaking populations, paving the way for culturally and linguistically tailored approaches to early detection and intervention.

II. Methods

1. Participants

Each of the volunteering participants in this analysis took the Cognitive Impairment Screening Test (CIST), which is a cognitive ability screening tool developed in South Korea, and then consequently was placed in one of three groups according to the score of the test. The three groups represent stages of cognitive health deterioration, normal cognitive group (NC, *n*=13, female=12, male=1), mild cognitive impairment group (MCI, *n*=10, feamle=9, male=1), and dementia suspected group (De, *n*=9, female=9), with total sample of 32 participants. Due to the influence of age and education on CIST scores, group placement was determined on a case-by-case basis, considering individual participant characteristics. Specifically, participants were categorized into three groups (NC: Participants scoring within the range of healthy cognitive ability for their age and education level, MCI: Participants scoring within 6 points below the expected range for healthy cognition, considering their age and education, De: Participants scoring 7 or more points below the expected range for healthy cognition, considering their age and education.)

The groups did not differ much based on age mean or education mean, but there seemed to be a significant difference between the three groups regarding the CIST score for each group. Table 1 below showcases age education and CIST mean and standard deviation calculations for each group.

Table 1. Demographic information by group

Category	NC	MCI	De	*p*-value
n (%)	13 (40.6%)	10 (31.25%)	9 (28.1%)	-
Age (*SD*)	82.5 (5.38)	84.7 (5.33)	81.9 (6.83)	.529
Education (*SD*)	9.76 (4.26)	8.00 (2.00)	7.30 (5.20)	.348
CIST (*SD*)	22.15 (3.62)	12.60 (2.41)	6.20 (4.40)	<.001***

Note. De=dementia; MCI=mild cognitive impairment; NC=normal cognitive; *SD*=standard deviation
****p*<.001

2. Speech tasks

Two tasks were performed to collect spontaneous speech from the participants, the picture description task and a question-and-answer task. Picture description task has been used in much research before to record spontaneous speech of patients of cognitive decline to analyze the feasibility of using acoustic features in automatic diagnose of cognitive decline (Kumar et al., 2022; Liu et al., 2023; Themistocleous et al., 2020). The most commonly used picture in the picture description task used in previous research (Forbes-McKay & Venneri, 2005; Fromm et al., 2024; Hall et al., 2019; Yamada et al., 2022) and this one as well is the “cookie theft” picture developed as a tool in the Boston Aphasia Examination. Fraser et al. (2019) has argued that the “cookie theft” picture has been used in studies analyzing data of different languages (Japanese, Norwegian, Chinese etc.) which makes it a very good candidate for cross-linguistic comparative studies.

The second task of question-and-answer was also conducted to record spontaneous speech of the participants. The questions were about memories of the participants and were comprised to cover happy, sad, and neutral emotional states representing memories. The question-and-answer task was aimed at collecting the speech of the participants using their long-term memory and recollection.

3. Acoustic and suprasegmental speech features

Acoustic features related to fundamental frequency (F0) were reported to have been significant in differentiating the speech of NC and cognitive impairment in studies such as Gonzalez-Moreira et al. (2015), De Stefano et al. (2021), and Kumar et al. (2022). In this study however only the coefficient of variation of fundamental frequency (F0 cov) is calculated as F0 measurements are known to be different between men and women (Yoshii et al., 2023). We are not putting a lot of emphasis on pitch related features as Tanaka et al. (2017) has explained that Pitch related features are more connected to emotions than cognitive impairment. That being said we are not ruling out all pitch related features altogether. We are also considering Jitter and Shimmer measures as Lin et al. (2020) and Hall et al. (2019) have pointed out the importance of jitter and shimmer measures in the speech of cognitive impairment. This study is analyzing jitter (local), jitter (rap), jitter (ppq5), jitter (ddp) and shimmer (local), shimmer (apq3), shimmer (apq5), shimmer (apq11) shimmer (dda). Moreover, mean autocorrelation and harmonic-to-noise ratio (HNR) are also calculated and analyzed as acoustic features of speech. Other features under analysis includes speech intensity measuring speech pressure level (dB), and voice quality assessing cepstral peak prominence (CPP).

For temporal suprasegmental features, the study is considering pause segment related speech characteristics such as no. of pauses, total pause duration, and pause rate. Studies such as Toth et al. (2018), König et al. (2015), Liu et al. (2023), and Da Cunha et al. (2022) have all reported on the significance of pause related features such as rate, length and duration in analyzing cognitive impairment speech. More features related to phonation are also considered in this study; features such as total phonation time, phonation rate, no. of syllables per utterance, speech rate, and articulation rate. Boschi et al. (2017) has emphasized how phonation time is also an important temporal feature that can indicate Cognitive impairment together with pause. Other features such as total duration of speech, and time of reaction are also considered in this analysis. As for duration of response time Yoshii et al. (2023) pointed out its significance in differentiating cognitive impairment speech. Features such as filler (um, ung, a etc.) duration is also important in analyzing speech disfluencies and there for it is also a feature considered in the analysis.

4. Analysis tools and statistics

Speech recording was done using TASCAM Portacapture X8 (TEAC AMERICA, INC. CA, USA) which has a sampling rate of 192KHz and quantization of 32-bit. The recordings were saved to file type “wav”, and acoustic feature analysis was done using Praat. The extracted features were statistically analyzed for significance using One-way ANOVA, and post hoc analysis was done using Least significant Difference (LSD). This study has utilized

LSD post-hoc due to its higher sensitivity in detecting smaller differences.

Ⅲ. Results

1. Acoustic variables

The result of analyzing speech acoustic variables showed that no acoustic features are related to cognitive impairment speech in the speech derived from the interview task. Only the variable of sound pressure level (SPL dB) analyzed in the data of the picture description task samples had a *p*-value of .048 that is lower than *α* =.05, which indicates a significance in the difference of intensity values between the groups. It is noteworthy to mention that the *p*-value of the variable of SPL (dB) that is analyzed in the data of the interview task is just marginally over the *α* value. This could indicate that SPL (dB) mean is a significant variable in analyzing the speech of cognitive impairment. However, in a follow up post hoc LSD test the significance of difference between the groups was found to be in De-NC (LSD=4.42〈ABS=5.18) data. Indicating that the difference in SPL (dB) is associated more with the later stages of cognitive decline.

2. Suprasegmental features

The analysis of the speech temporal characteristics analysis results reveal that variables such as duration of fillers, speech rate and articulation rate are influenced by cognitive health decline. As shown in Table 3 the variables filler duration, speech rate, and articulation rate are mostly related to interview task data. And in all the three variables Filler duration was the most statistically significant with a *p*-value of .0005. On the other hand, in the picture description task, only articulation rate resulted in statistical significance (*p*=.004).

Post hoc analysis also affirms the significance of the data derived from the interview task as the analysis results suggest that the data derived from the interview task can differentiate between different cognitive health stages. That is, both filler duration and articulation rate data retrieved in the interview task have shown statistical significance of differences between De-MCI (LSD=.858〈ABS=1.1, LSD=2.54〈ABS=5.2 respectively) and MCI-NC (LSD=.79〈ABS=1.2, LSD=2.33〈ABS=4.1 respectively). Whereas even though speech rate was confirmed to be of significance in the One-way ANOVA test, the post hoc analysis shows that difference between groups is in the De-MCI (LSD=.98〈ABS=1.12) and De-NC (LSD=.93〈ABS=1.14) data only which suggest that this variable is

Table 2. Statistics analysis of acoustic variables

Variable	Interview				Picture description			
	M	*SD*	*F*	*p*-value	*M*	*SD*	*F*	*p*-value
F0cov	.34	.16	1.609	.218	.32	.13	2.839	.075
Jitter (local)%	2.37	.54	.067	.935	2.32	.59	.123	.885
Jitter (rap)%	1.07	.30	.317	.730	1.04	.32	.350	.708
Jitter (ppq5)%	1.18	.28	.121	.887	1.16	.32	.201	.819
Jitter (ddp)%	3.22	.89	.318	.730	3.12	.97	.350	.708
Shimmer	11.18	1.86	1.124	.339	10.57	1.91	1.111	.343
Shimmer (local)dB	1.07	.15	1.070	.356	1.03	.15	1.197	.317
Shimmer (apq3)%	4.45	1.14	2.247	.124	4.20	.99	.934	.404
Shimmer (apq5)%	6.45	1.42	1.501	.240	5.99	1.41	1.099	.347
Shimmer (apqII)%	11.48	2.15	.104	.901	10.77	2.45	.956	.396
Shimmer (dda)%	13.35	3.42	2.247	.124	12.61	2.98	.934	.404
Mean autocorrelation	.88	.03	.875	.427	.89	.04	.137	.872
HNR	11.41	1.66	.824	.449	12.51	2.19	.118	.889
CPP	13.93	1.41	.812	.454	13.29	1.51	.788	.464
SPL (dB)	64.15	5.53	3.146	.058	63.29	5.35	3.374	.048*

Note. CPP=cepstral peak prominence; HNR=harmonic-to-noise ratio; SPL=sound pressure level.
*p〈.05

also connected to later stages of cognitive decline. For the picture description data articulation rate was found to be also significantly different between De-MCI (LSD=1.15〈 ABS=2.04) and De-NC (LSD=1.09〈ABS=1.25). Figure 1 illustrates the connection between the statistically significant features and the speech of the three groups.

3. Task type and derived parameter

Table 4 demonstrates the effect of the type of task on

Table 3. Statistical analysis of speech duration data

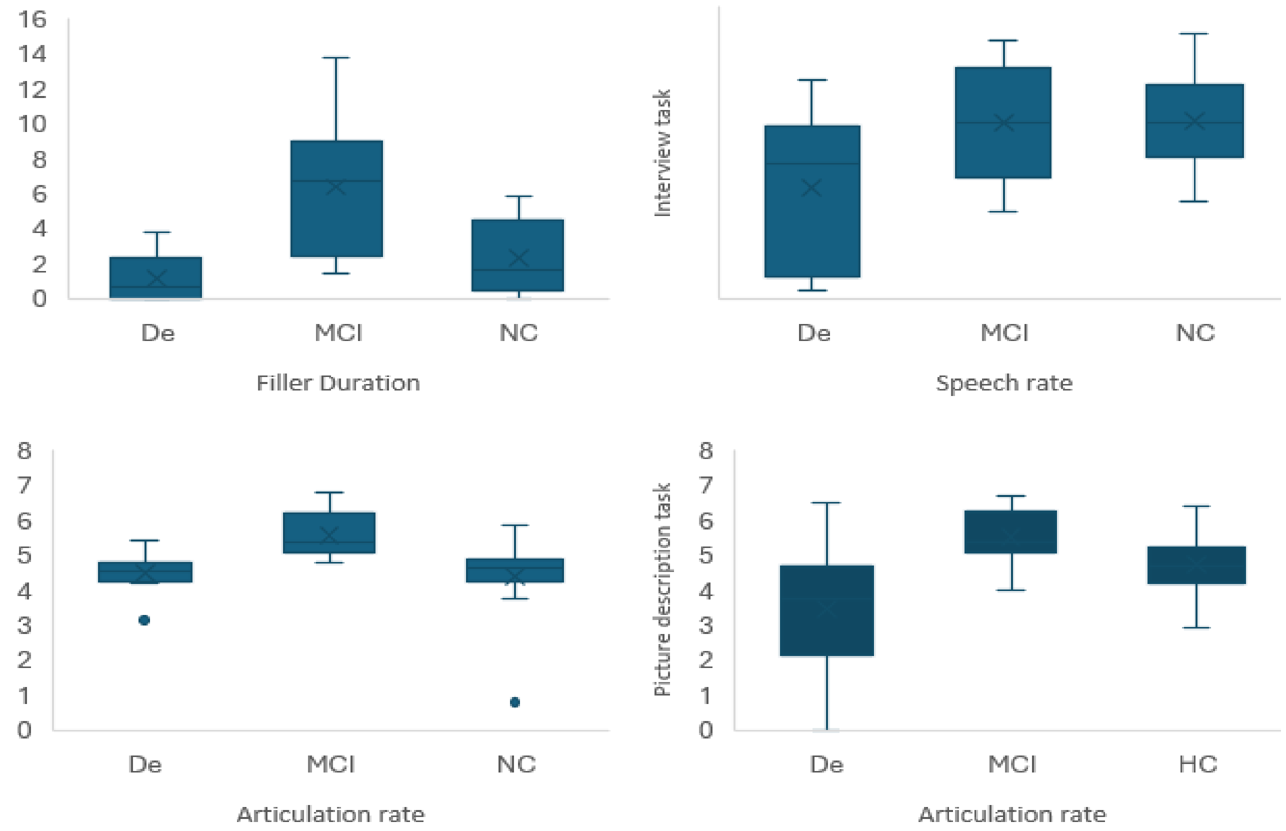
Variable	Interview				Picture description			
	M	*SD*	*F*	*p*-value	*M*	*SD*	*F*	*p*-value
Speech time	119	167	1.44	.25	53.9	22.0	2.64	.08
Reaction time	5.8	5.4	.24	.79	5.1	3.6	.30	.74
Pause no.	38.0	60.2	1.33	.28	13.5	6.9	2.00	.15
Pause duration	41.2	58.3	.68	.52	27.6	9.2	1.03	.36
Pause rate	.6	1.3	.49	.61	.5	.2	.76	.48
Phonation duration	79.2	121.0	2.71	.08	19.3	10.9	1.50	.23
Phonation proportion	67.1	78.8	1.47	.24	34.6	15.2	.26	.76
Filler duration	3.3	3.4	10.1	.0005***	2.1	3.0	2.06	.14
Syllable no.	367	595	1.81	.18	93.8	55.4	1.34	.27
Speech rate	2.7	1.1	3.79	.034*	1.7	.8	.66	.52
Articulation rate	4.8	1.1	5.54	0.009**	4.6	1.5	6.64	0.004**

$^{*}p$〈.05, $^{**}p$〈.01, $^{***}p$〈.001

Table 4. *t*-test analysis of the effect of task on parameter

Variable	De			MCI			NC		
	M	*SD*	*p*-value	*M*	*SD*	*p*-value	*M*	*SD*	*p*-value
F0 cov	.32	.13	.08	.29	.14	.19	.36	.15	.44
Jitter (local)%	2.40	.71	.98	2.30	.60	.91	2.35	.41	.53
Jitter (rap)%	1.13	.40	.90	1.03	.32	.98	1.03	.22	.51
Jitter (ppq5)%	1.22	.37	.97	1.17	.32	.98	1.14	.23	.67
Jitter (ddp)%	3.38	1.21	.91	3.09	.95	.98	3.08	.67	.51
Shimmer	11.4	2.60	.44	10.7	1.90	.81	10.5	1.18	.03*
Shimmer (local)dB	1.11	.18	.61	1.03	.17	.99	1.02	.10	.04*
Shimmer (apq3)%	4.79	1.53	.41	4.21	.89	.62	4.09	.68	.20
Shimmer (apq5)%	6.77	2.08	.48	6.13	1.29	.92	5.90	.75	.02*
Shimmer (apqII)%	11.5	3.07	.74	11.2	2.52	1.00	10.7	1.41	.009**
Shimmer (dda)%	14.3	4.59	.41	12.6	2.67	.62	12.2	2.03	.20
Mean autocorrelation	.88	.04	.71	.89	.04	.62	.88	.03	.03*
HNR	11.8	2.04	.43	12.2	2.49	.51	11.8	1.59	.007**
CPP	13.2	1.96	.67	13.6	1.28	.03	13.8	1.23	.49
SPL (dB)	60.8	7.37	1.00	63.0	4.25	.26	66.2	3.22	.73
Speech time	53.8	26.5	.06	71.7	41.9	.03*	120	184	.11
Reaction time	4.69	4.55	.71	6.07	5.02	.52	5.52	4.36	.92
Pause no.	13.8	12	.16	22.7	14.7	.02*	36.3	66.6	.10
Pause duration	28.4	14.9	.21	29	13.6	.82	42.6	63.3	.30
Pause rate	.56	.23	.37	.44	.18	.007**	.69	1.42	.52
Phonation duration	21.7	21	.17	35	31.5	.02*	79	134	.03*
Phonation proportion	36.8	24.8	.39	43.9	16.2	.007**	65	87	.09
Filler duration	1.22	1.40	.92	5.04	4.41	.18	1.93	1.98	.33
Syllable no.	99.6	99.3	.19	194	172	.01*	349	658	.06
Speech rate	1.68	1.18	.44	2.46	1.04	.01*	2.38	1.02	〈.001
Articulation rate	3.99	1.45	.15	5.56	.74	.84	4.56	1.04	.41

Note. De=dementia; MCI=mild cognitive impairment; NC=normal cognitive.
$^{*}p$〈.05, $^{**}p$〈.01, $^{***}p$〈.001



Note. De=dementia; MCI=mild cognitive impairment; NC=normal cognitive.

Figure 1. Boxplot of statistically significant suprasegmental features

the analyzed parameters for each group. As shown in the table, the data for De was not affected by the type of task performed. Only MCI and NC groups exhibited statistically significant differences between tasks. In the case of the differences in variable calculations between tasks for acoustic variables were only observed in the NC group, not in the other two. These differences were more pronounced in amplitude-related measures. Suprasegmental parameters, however, showed significant differences between dependent derivations of the measures in both NC and MCI groups. Upon closer examination, the parameters exhibiting significant differences based on the task were all related to response time. This may be attributed to the fact that the interview task requires a longer response time compared to the picture description task. Which widens the range of the data to be analyzed for these variables and increases the probability of detection of cognitive impairment.

Ⅳ. Discussion and Conclusion

This study aims to analyze the association between cognitive decline and acoustic and temporal characteristics of spontaneous speech. Subjects were divided into three groups according to cognitive health, and their speech was recorded and analyzed for temporal measures and acoustic variables. Each subject was given two tasks to complete, a picture description task and a questions and answers interview task, through which the speech samples for each participant were collected for analysis.

The results of the analysis show that cognitive decline can be detected in speech through the analysis of different acoustic and speech temporal variables. The results suggest that there may be a stronger connection between cognitive decline and temporal speech characteristics than there is with acoustic speech characteristics. That is, in our result it was confirmed that suprasegmental parameters such as filler duration and articulation rate have statistically significant differences between De-MCI and MCI-NC. On the contrary for acoustic variables the statistical significance only indicated a connection to De by showing difference between De-NC.

Our finding of the importance of articulation rate in the speech samples of MCI is consistent with the findings by Themistocleous et al. (2020). Themistocleous et al. (2020) have compared the speech of cognitively healthy individuals and MCI and found that a few speech features including articulation rate are significant in differentiating the speech of NC and MCI. The other features which showed significance for their data counter this study's findings, features such as shimmer and average speaking time to be of significance whereas these features have not amounted to any significance in our data. Also, another study by Yamada et al. (2021) reported that pause duration, speech rate and phonation time to be significant in differentiating NC and MCI speech, in all these variables we have found the speech rate to be the only significant variable in our data. Nevertheless these differences we can attribute to two causes; the first one is that both the studies mentioned above have had a larger sample for assessment than ours (Themistocleous et al., 2020; Yamada et al., 2021), and the second is that the studies either depended on one task only or on different tasks to derive the speech samples for analysis (Themistocleous et al., 2020, picture description; Yamada et al., 2021, interview questions about daily life, counting backward, subtraction, phonemic and semantic verbal fluency, and picture description). And in the case of the picture description task, it is noteworthy to mention that this study's sample has found some difficulty in identifying the characteristics of the picture as the "cookie theft" picture content slightly differs from everyday native life of our sample.

Many previous studies have reported the importance of silence or pause related features in distinguishing the speech of cognitive impairment including König et al. (2015). König et al. (2015) concluded that the pause length variable was the most significant to distinguish between NC, MCI, and AD, but their study did not consider the frequency of pauses and instead analyzed pause length and rate. Martínez-Nicolás et al. (2022) has argued that the number of pauses in speech and phonation time are significant features that distinguish the speech of NC, MCI and AD patients. In this study a paragraph reading task was used to collect samples of speech from 400 participants. Vincze et al. (2021) has reported that the no. of pauses, or the proportion of pauses in relation to total speech time increases with the progression of De. However, the analysis of our data did not find any importance of the pause related features but for the statistical significance of filler duration in the data of the interview task. Yoshii et al. (2023) has reported that both filler duration and filler proportion are both significant in discriminating the speech of NC and MCI. Albeit these features could not amount to a significant effect size, they were derived from the speech of everyday conversation, that is spontaneous speech. This coincides with our finding that the filler duration can distinguish the speech of MCI from De and NC in spontaneous speech.

Regarding acoustic features, statistical analysis results have not indicated any significant values in our data but for intensity. But in analyzing acoustic features in the speech of cognitive decline many studies consider and report on the significance of Jitter and shimmer instead of intensity's SPL (dB) specifically. Nishikawa et al. (2022), Liu et al. (2023), Kumar et al. (2022), Themistocleous et al. (2020), Yoshii et al. (2023), Yamada et al. (2021, 2022), and Hall et al. (2019) have all considered jitter and shimmer features in their analysis but only Yoshii et al. (2023), Themistocleous et al. (2020), Hall et al. (2019), and Yamada et al. (2022) have reported the significance of these variables in the analysis of cognitive impairment. Even in that case, the reports seldom included all the measures of jitter and shimmer as significant, only Yoshii et al. (2023) has reported the importance of all the jitter and shimmer features we analyze in this study when measuring them in spontaneous speech, Yamada et al. (2022) has also reported the significance of jitter and shimmer in discriminating the speech of NC, MCI and De when derived from picture description speech, but they did not specify which specific features they used in their analysis. On the other hand, Themistocleous et al. (2020) only reported on shimmer in analyzing NC and MCI speech, and Hall et al. (2019) only reported the importance of jitter and shimmer (local and apq3) in distinguishing the speech of NC, MCI, and AD. Jitter and shimmer, however, did not show any significance of differences between the three groups in our analysis. This may be due to differences of analysis samples in our study and the other studies. In our study we have calculated the acoustic measures in a sample of about 20 seconds of continuous speech for each participant in each task. It is not clear how the samples in other studies were defined and measured which may cause difference in findings between our analysis and theirs.

Ultimately, what our result confirms is that speech temporal variables are more related to the early stages of cognitive impairment. This is disclosed in the result of our LSD post hoc analysis. Our analysis discovered that acoustic features are not statistically significant enough to differentiate between different stages of cognitive decline save for the De stage when compared to NC, while suprasegmental features such as filler duration and articulation rate are statistically substantial in differentiating between the three groups.

This is due to how cognitive decline affects memory functions such as word retrieval, which significantly impacts speech production. When individuals struggle to access the appropriate words, they experience pauses and hesitations. To compensate for these delays, they may increase their reliance on filler words (such as "um," "uh," "like") to fill the silence and maintain conversational flow. This compensatory strategy can lead to a decrease in articulation rate. In essence, as speakers require more time to retrieve the intended words, their overall speech becomes slower, characterized by fewer syllables per utterance and a higher frequency of filler words.

However, these variables' significance depends on the type of task used to derive speech. This means that there is a relation between the suprasegmental variable change and the type of task the participant undergoes to produce speech. These cases of speech feature statistical significance based on the task of speech the features are derived from has also been reported in other studies. For example, Yoshii et al. (2023) has noted that only 5 of 17 features derived from the speech recorded of the Mini Mental State Exam tasks had significant difference between groups, compared to 16 of these features showing significance when derived from spontaneous speech recordings of everyday life question task. Yamada et al. (2022) has also reported the significance of the features that differentiate the speech of NC and De in relation to the type of speech task they are derived in; for example, they reported that proportion of pause duration, jitter, shimmer, and phoneme rate all showed significant differences in the speech recorded of picture description task, while proportion of pause duration, and pitch variation were the significant features in subtraction task and proportion of pause duration and shimmer only were significant in semantic fluency task speech.

In the picture description task, participants only needed to observe a presented image and recall the words used to describe its elements. Conversely, the interview task demanded greater cognitive effort, requiring participants to recollect a personal life event, retrieve relevant words, and then construct sentences to describe that experience. This increased cognitive load likely explains why the variables derived from the interview task exhibited more statistical significance, even after post-hoc analysis, compared to those from the picture description task.

Several limitations were present in this study: First, small sample size: The limited number of participants may have constrained the ability to generalize our findings. A larger sample size would be required to confirm the robustness of our observations. Second, picture description task: The "cookie theft" picture, while used in prior cognitive impairment speech studies, may have been unfamiliar to our participants. This potential unfamiliarity could have influenced their performance and the quality of the collected speech data. In future studies, employing a more culturally relevant picture and comparing the results with this study would help assess the validity of using the "cookie theft" picture. Third, limited interview questions: The interview was restricted to three questions, potentially limiting the range of emotional expressions and the depth of insights. To broaden the scope of the analysis, future studies should incorporate a wider array of questions covering diverse life situations. Fourth, focus on acoustic and suprasegmental features: This study primarily focused on acoustic and suprasegmental features of speech. Future research should incorporate linguistic features such as semantics and syntax into the analysis. Furthermore, investigating the combined effect of semantic, utterance length, and acoustic-suprasegmental parameters in identifying cognitive health could yield valuable insights.

Our findings also suggest that the time required to complete each task significantly influences the nature of the collected data. Since the interview task necessitates greater participant engagement and a longer response time than the picture description task, the data derived from the interview task is richer in information, providing a more comprehensive assessment of speech fluency and the overall richness of utterances. Consequently, the speech signals obtained during the interview task are more suitable for investigating the association between cognitive abilities and speech characteristics.

Reference

Ahmed, M., & Kwon, S. B. (2024). A systematic literature review on acoustic of speech variables for measuring cognitive function. *Journal of Speech-Language & Hearing Disorders, 33*(1), 1-11. doi:10.15724/jslhd.2024.33.1.001

Ahmed, S., Haigh, A.-M. F., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain, 136*(12), 3727-3737. doi:10.1093/brain/awt269

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected speech in neurodegenerative language disorders: A review. *Frontiers in Psychology, 8,* 269. doi:10.3389/fpsyg.2017.00269

Choi, H., Kim, J. H., Lee, C. M., & Kim, J. I. (2013). Features of semantic language impairment in patients with amnestic mild cognitive impairment. *Dementia and Neurocognitive Disorders, 12,* 33-40. doi:10.12779/dnd.2013.12.2.33

Da Cunha, E., Plonka, A., Arslan, S., Mouton, A., Meyer, T., Robert, P., & Gros, A. (2022). Logogenic primary progressive aphasia or Alzheimer disease: Contribution of acoustic markers in early differential diagnosis. *Life, 12*(7), 933. doi:10.3390/life12070933

De Looze, C., Dehsarvi, A., Crosby, L., Vourdanou, A., Coen, R. F., Lawlor, B. A., & Reilly, R. B. (2021). Cognitive and structural correlates of conversational speech timing in mild cognitive impairment and mild-to-moderate Alzheimer's disease: Relevance for early detection approaches. *Frontiers in Aging Neuroscience, 13,* 637404. doi:10.3389/fnagi.2021.637404

De Stefano, A., Di Giovanni, P., Kulamarva, G., Di Fonzo, F., Massaro, T., Contini, A., & Cazzato, C. (2021). Changes in speech range profile are associated with cognitive impairment. *Dementia and Neurocognitive Disorders, 20*(4), 89-98. doi:10.12779/dnd.2021.20.4.89

Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Science, 26*(4), 243-254. doi:10.1007/s10072-005-0467-9

Fraser, K. C., Fors, K. L., & Kokkinakis, D. (2019). Multilingual word embeddings for the assessment of narrative speech in mild cognitive impairment. *Computer Speech & Language, 53,* 121-139. doi:10.1016/j.csl.2018.07.005

Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease, 49*(2), 407-422. doi:10.3233/JAD-150520

Fromm, D., Dalton, S. G., Brick, A., Olaiya, G., Hill, S., Greenhouse, J., & MacWhinney, B. (2024). The case of the cookie jar: Differences in typical language use in dementia. *Journal of Alzheimer's Disease, 100*(4), 1417-1434. doi:10.3233/JAD-230844

Gonzalez-Moreira, E., Torres-Boza, D., Kairuz, H. A., Ferrer, C., Garcia-Zamora, M., Espinoza-Cuadros, F., & Hernandez-Gómez, L. A. (2015). Automatic prosodic analysis to identify mild dementia. *BioMed Research International.* doi:10.1155/2015/916356

Ha, J.-W., Jung, Y. H., & Sim, H. S. (2009). The functional characteristics of fillers in the utterances of dementia of Alzheimer's type, questionable dementia, and normal elders. *Korean Journal of Communication Disorders, 14*(4), 514-530.

Hall, A. O., Shinkawa, K., Kosugi, A., Takase, T., Kobayashi, M., Nishimura, M., & Yamada, Y. (2019). Using tablet-based assessment to characterize speech for individuals with dementia and mild cognitive impairment: Preliminary results. *AMIA Joint Summits on Translational Science Proceedings, 2019,* 34-43.

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., & David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1*(1), 112-124. doi:10.1016/j.dadm.2014.11.012

Kumar, M. R., Vekkot, S., Lalitha, S., Gupta, D., Govindraj, V. J., Shaukat, K., & Zakariah, M. (2022). Dementia detection from speech using machine learning and deep learning architectures. *Sensors, 22*(23), 9311. doi:10.3390/s22239311

Lin, H., Karjadi, C., Ang, T. F. A., Prajakta, J., McManus, C., Alhanai, T. W., & Au, R. (2020). Identification of digital voice biomarkers for cognitive health. *Exploration of Medicine, 1,* 406-417. doi:10.37349/emed.2020.00028

Liu, J., Fu, F., Li, L., Yu, J., Zhong, D., Zhu, S., & Li, J. (2023). Efficient pause extraction and encode strategy for Alzheimer's disease detection using only acoustic features from spontaneous speech. *Brain Sciences, 13*(3), 477. doi:10.3390/brainsci13030477

López-de-Ipiña, K., Alonso, J.-B., Travieso, C. M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., & Martinez de Lizardui, U. (2013). On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. *Sensors, 13*(5), 6730-6745. doi:10.3390/s130506730

Martínez-Nicolás, I., Llorente, T. E., Ivanova, O., Martínez-Sánchez, F., & Meilán, J. J. G. (2022). Many changes in speech through aging are actually a consequence of cognitive changes. *International Journal of Environmental Research and Public Health, 19*(4), 2137. doi:10.3390/ijerph19042137

Meilán, J. J. G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T. E., & Carro, J. (2020). Changes in the rhythm of speech difference between people with nondegenerative mild cognitive impairment and with preclinical dementia. *Behavioural Neurology.* doi:10.1155/2020/4683573

Nagumo, R., Zhang, Y., Ogawa, Y., Hosokawa, M., Abe, K., Ukeda, T., & Shimada, H. (2020). Automatic detection of cognitive impairments through acoustic analysis of speech. *Current Alzheimer Research, 17*(1), 60-68. doi:10.2174/1567205017666200213094513

Nishikawa, K., Akihiro, K., Hirakawa, R., Kawano, H., & Nakatoh, Y. (2022). Machine learning model for discrimination of mild dementia patients using acoustic features. *Cognitive Robotics, 2,* 21-29. doi:10.1016/j.cogr.2021.12.003

Park C.-Y., Kim, M., Shim, Y. S., Ryoo, N., Choi, H., Jeong, H. T., & Youn, Y. C. (2024). Harnessing the power of voice: A deep neural network model for Alzheimer's disease detection. *Dementia and Neurocognitive Disorders, 23*(1), 1-10. doi:10.12779/dnd.2024.23.1.1

Riley, C. O., McKinstry, B., & Fairhurst, K. (2022). Accuracy of telephone screening tools to identify dementia patients remotely: Systematic review. *JRSM Open, 13*(9). doi:10.1177/20542704221115956

Sadeghian, R., Schaffer, J. D., & Zahorian, S. A. (2021). Towards an automatic speech-based diagnostic test for Alzheimer's disease. *Frontiers in Computer Science, 3,* 624594.

doi:10.3389/fcomp.2021.624594

Sumali, B., Mitsukura, Y., Liang, K., Yoshimura, M., Kitazawa, M., Takamiya, A., & Kishimoto, T. (2020). Speech quality feature analysis for classification of depression and dementia patients. *Sensors, 20*(12), 3599. doi:10.3390/s20123599

Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., & Nakamura, S. (2017). Detecting dementia through interactive computer avatars. *IEEE Journal of Translational Engineering in Health and Medicine, 5,* 2200111. doi:10.1109/JTEHM.2017.2752152

Themistocleous, C., Eckerström, M., & Kokkinakis, D. (2020). Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls. *PLoS ONE, 15*(7), e0236009. doi:10.1371/journal.pone.0236009

Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., & Kalman, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research, 15*(2), 130-138. doi:10.2174/1567205014666171121114930

Vincze, V., Szatlóczki, G., Tóth, L., Gosztolya, G., Pákáski, M., Hoffmann, I., & Kálmán, J. (2021). Telltale silence: Temporal speech parameters discriminate between prodromal dementia and mild Alzheimer's disease. *Clinical Linguistics & Phonetics, 35*(8), 727-742. doi:10.1080/02699206.2020.1827043

Wang, H.-L., Tang, R., Ren, R.-J., Dammer, E. B., Guo, Q.-H., Peng, G.-P., & Wang, G. (2022). Speech silence character as a diagnostic biomarker of early cognitive decline and its functional mechanism: A multicenter cross-sectional cohort study. *BMC Medicine, 20,* 380. doi:10.1186/s12916-022-02584-x

Yamada, Y., Shinkawa, K., Kobayashi, M., Nishimura, M., Nemoto, M., Tsukada, E., . . . Arai, T. (2021). Tablet-based automatic assessment for early detection of Alzheimer's disease using speech responses to daily life questions. *Frontiers in Digital Health, 3,* 653904. doi:10.3389/fdgth.2021.653904

Yamada, Y., Shinkawa, K., Nemoto, M., Ota, M., Nemoto, K., & Arai, T. (2022). Speech and language characteristics differentiate Alzheimer's disease and dementia with Lewy bodies. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 14*(1), e12364. doi:10.1002/dad2.12364

Yoshii, K., Kimura, D., Kosugi, A., Shinkawa, K., Takase, T., Kobayashi, M., & Nishimura, M. (2023). Screening of mild cognitive impairment through conversations with humanoid robots: Exploratory pilot study. *JMIR Formative Research, 7,* e42792, doi:10.2196/42792

Appendix 1. Definitions of variables under analysis

Variable	Definition
Response time	Total time of answer including silent segments
Reaction time	Total time from the last syllable of question to the first syllable of answer
Number of pauses	Total number of silences (more than .3s)
Proportion of pause	Total number of pause segments divided by total response time
Phonation time	Total time of all syllables produced
Proportion of phonation	Total time of all syllables produced divided by total response time
Filler duration	Total time of speech dysfluencies produced (such as: um, ung etc.)
Speech rate	Total number of syllables divided by total time of response
Articulation rate	Total number of syllables produced divided by total time of phonation
Number of syllables	Total number of syllables produced in articulation
F0 cov	Coefficient of variation of the fundamental frequency
Jitter (local, rap, ppq5, ddp)	Fluctuations in pitch
Shimmer (local, apq3, apg5, apq11, dda)	Fluctuations in volume