RESEARCH ARTICLE
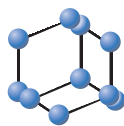
Automatic Detection of Cognitive Impairments through Acoustic Analysis of Speech

Ryosuke Nagumo[1,*], Yaming Zhang[1], Yuki Ogawa[1], Mitsuharu Hosokawa[1], Kengo Abe[1], Takaaki Ukeda[1], Sadayuki Sumi[1], Satoshi Kurita[2], Sho Nakakubo[2], Sangyoon Lee[2], Takehiko Doi[2] and Hiroyuki Shimada[2]

[1]*Panasonic Corporation, Osaka, Japan;* [2]*Department of Preventive Gerontology, Center for Gerontology and Social Science, National Center for Geriatrics and Gerontology, Aichi, Japan*

Abstract: ***Background***: Early detection of mild cognitive impairment is crucial in the prevention of Alzheimer's disease. The aim of the present study was to identify whether acoustic features can help differentiate older, independent community-dwelling individuals with cognitive impairment from healthy controls.

Methods: A total of 8779 participants (mean age 74.2 ± 5.7 in the range of 65-96, 3907 males and 4872 females) with different cognitive profiles, namely healthy controls, mild cognitive impairment, global cognitive impairment (defined as a Mini Mental State Examination score of 20-23), and mild cognitive impairment with global cognitive impairment (a combined status of mild cognitive impairment and global cognitive impairment), were evaluated in short-sentence reading tasks, and their acoustic features, including temporal features (such as duration of utterance, number and length of pauses) and spectral features (F0, F1, and F2), were used to build a machine learning model to predict their cognitive impairments.

Results: The classification metrics from the healthy controls were evaluated through the area under the receiver operating characteristic curve and were found to be 0.61, 0.67, and 0.77 for mild cognitive impairment, global cognitive impairment, and mild cognitive impairment with global cognitive impairment, respectively.

Conclusion: Our machine learning model revealed that individuals' acoustic features can be employed to discriminate between healthy controls and those with mild cognitive impairment with global cognitive impairment, which is a more severe form of cognitive impairment compared with mild cognitive impairment or global cognitive impairment alone. It is suggested that language impairment increases in severity with cognitive impairment.

1. INTRODUCTION

Dementia refers to the loss of cognitive functions severe enough to disrupt daily life, and Alzheimer's Disease (AD) is the most common type of dementia. As a neurodegenerative condition, AD affects multiple regions of the brain as it progresses, and more than one cognitive domain can be impaired. In 2017, 15% of the Japanese population above the age of 65 was reported to have AD [1], and there is still no effective cure available. Mild Cognitive Impairment (MCI) is defined as the pathological stage between healthy aging and dementia. It is believed that individuals with MCI are at an increased risk of developing AD. Therefore, it is important to detect cognitive impairment at an early stage, so that the progression of MCI can be delayed.

Although the impairment of cognitive functions is mostly measured through neuropsychological screening tests, for example, the Mini-Mental State Examination (MMSE) [2] and Montreal Cognitive Assessment (MoCA) [3], these tests usually need to be administered by trained professionals. In recent years, researchers have attempted to detect cognitive impairments through assessment methods that are non-invasive and less time-consuming. Memory impairment has been regarded as the early marker of amnestic MCI and early-stage AD, and patients usually suffer from episodic memory loss due to the shrinkage of the hippocampus, located in the medial temporal lobe and crucial to memory function. Meanwhile, studies have also demonstrated that language impairment is prevalent among individuals with AD, and noticeable changes can appear at an early stage [4]. For instance, patients with early-stage AD often have difficulties in picture naming and category fluency tasks, indicat-

*Address correspondence to this author at the Panasonic Corporation, 1048, Oaza-Kadoma, Kadoma City, Osaka 571-8501, Japan; Tel: +81-6-6908-5800; Fax: +81-6-6900-2807; E-mail: nagumo.ryosuke@jp.panasonic.com

ing impairment in semantics. At the same time, several studies have suggested impairments in phonetics and phonology [5, 6]. Specifically, the verbal fluency is often influenced. It was reported that the oral reading of patients with AD is marked by temporal features such as reduced speech and articulation rate, low effectiveness of phonation time, and increases in the number and proportion of pauses [7].

Some studies have been conducted to investigate the potential of acoustic features about serving as markers for detecting cognitive impairment. For instance, König *et al.* detected MCI and AD from healthy controls (HCs) by employing temporal acoustic features, including duration and ratio of pauses, with an accuracy of 79% and 87%, respectively [8]. Themistocleous *et al.* further reported that it is possible to differentiate people with MCI from healthy individuals by using acoustic features, including vowel duration, vowel formant frequency, and fundamental frequency, as predictors [9]. The underlying cause of these acoustic changes in speech is still not clear, and they could be attributed to greater cognitive processing and longer planning time for utterance.

Language is a complex function, and multiple regions of the brain are involved in the execution of speech function. As cognitive impairment often leads to compromised motor performance, it potentially has an impact on speech function. Recent studies have reported that patients with AD have a high incidence of apraxic disorders in oral and speech functions [10]. As speech production requires coordinated movements of articulatory organs, impaired speech planning has a potential influence on articulation speed. It has been well documented that individuals with MCI and AD displayed markedly reduced articulatory agility during an articulatory oral motor task, in which the participants were asked to produce sequential speech as fast as possible [11, 12].

In this study, we aimed to investigate the possibility of differentiating individuals with MCI, global cognitive impairment (GCI, defined as a Mini Mental State Examination score of 20-23 and no clinical indications of dementia), and MCI with GCI (a combined status of MCI and GCI) from HCs by extracting acoustic features from speech. To achieve the research goal, we employed four speech tasks to collect the utterance data, placing a relatively low cognitive burden on the participants.

The present study is part of a community-based cohort study, conducted by the National Center for Geriatrics and Gerontology-Study of Geriatric Syndrome, between September 2017 and July 2018, which aimed at establishing a screening system for preventing geriatric syndromes.

2. METHODS

2.1. Participants

A total of 10041 elderly people above the age of 65 years were recruited from Tokai and Toyoake cities, Aichi, Japan. They all received a health check, which included a face-to-face interview, physical and cognitive assessments, and a speaking test.

Our inclusion criteria were as follows [13, 14]: all participants were aged 65 years or older at the time of examination and resided either in Tokai or Toyoake city. The exclusion criteria were as follows: health problems such as AD, stroke, Parkinson's disease, depression, self-reported MCI, and other brain diseases, for example, brain tumor and epilepsy (n = 1118); inability to perform basic activities of daily living (ADLs), including eating, grooming, walking, bathing, and climbing up and down stairs (n = 19); MMSE score < 20, which indicates moderate dementia (n = 104) [15, 16]; and missing data regarding the exclusion criteria, cognitive assessments, and other measurements (n = 21). After applying these exclusion criteria, 8779 participants (mean age 74.2 ± 5.7 in the range of 65-96, 3907 males and 4872 females) out of the initial 10041 respondents were analyzed.

2.2. Cognitive Evaluation

All participants completed both the MMSE [2, 17] and the National Center for Geriatrics and Gerontology-Functional Assessment Tool (NCGG-FAT) [18]. The MMSE assesses global cognitive abilities including orientation, registration, attention, calculation, recall, language, and copying. Participants got MMSE scores in the range of 0-30, the lower which shows more severe cognitive impairments. We used MMSE to classify participants with GCI, in accordance with previous findings [19], and participants with a score of 20-23 on the MMSE and no clinical indications of dementia were considered to have GCI.

On the other hand, the NCGG-FAT evaluates cognitive functions in the following four domains: word memory (immediate recognition and delayed recall), attention (Trail Making Test (TMT)-part A), executive function (TMT-part B), and processing speed (Digit Symbol Substitution Test). Participants were given about 20 minutes to complete the battery. These four tests in the NCGG-FAT assessed cognitive impairments according to the standardized thresholds to define impairment in the corresponding domain (score < 1.5 standard deviations below the age-specific mean) previously established in population-based cohorts [18]. Therefore, participants were scored in the range of 0-4 in the NCGG-FAT, score 0 means no impairment in any of the four cognitive tests while 1, 2, and 3 mean impairment in 1 test, 2 tests, and 3 tests, respectively. The participants with a score of 1-4 were considered to have MCI, and a higher score shows multiple domain impairments of the cognitive function. In the GCI group, participants whose cognitive performance also deviated from HCs in NCGG-FAT were considered to have more severe cognitive impairment, thus were classified as MCI with GCI.

Given these two scores of cognitive assessments, our operational definitions were as follows [20]: HC: MMSE of 24-30, NCGG-FAT of 0; MCI: MMSE of 24-30, NCGG-FAT of 1-4; GCI: MMSE of 20-23, NCGG-FAT of 0; MCI with GCI: MMSE of 20-23, NCGG-FAT of 1-4.

Although MCI was divided into two categories, that is, single domain and multiple domains, according to the degree of impairments in the previous study [20], this was not the case in the present study because we wanted to maintain the large sample size in each category to validate the machine learning model, which demands big size data.

Table **1** shows the participant characteristics across all classes (*i.e.,* HC, MCI, GCI, and MCI with GCI). Of the 8779 participants, 6343 (72.3%), 1601 (18.2 %), 367 (4.2 %), and 468 (5.3 %) were part of the HC, MCI, GCI, and MCI with GCI groups, respectively. Table **1** also includes the results of the MMSE and NCGG-FAT. In the case of impairments assessed with the NCGG-FAT, because the same participant can belong to more than one impairment category, the summed number of participants in all four categories can exceed the total number of each group.

2.3. Voice Recording

Reading tasks involves simultaneous processing of orthographic, phonological, and semantic information, thus have been widely used to evaluate language function in individuals with MCI and AD [21, 22]. In this study, we adopted the following four sentence reading tasks to measure cognitive impairments: a) Vowel utterances b) Tongue twister c) Diadochokinesis d) Short sentences (see Appendix **A**).

The recording procedure was as follows. The recordings were conducted in a closed space surrounded by sound-absorbing materials, supervised by trained staff. If participants missed the sentence, they were guided to retry the task. All the vocal tasks were recorded using an Audio Technica H6 Handy Recorder with an Audio Technica SGH-6 Shotgun Mic Capsule connected to a PC. The microphone was set on a stand placed 10 cm away from the participant's mouth. While the sampling frequency was 44 kHz, the gain level was adjusted by the staff to ensure that the recording power would not saturate.

Then, recorded voices were processed with handwritten Python programs. Readymade vocal processing software (*e.g.,* Praat [23]) was not used given that we wanted to deploy the fully automatic vocal processing system running in the server. With regard to the computational calculations, participants' voices were first converted to the time series data of power, fundamental frequency, and formant frequency. Then, signals were smoothed by simple moving average calculation for noise reduction.

2.4. Acoustic Features

After the pre-processing, acoustic features were extracted from each task. We mainly focused on the acoustic features reflecting the temporal characteristics of speech, while features in the frequency domain were also extracted in the vowel utterance task.

2.4.1. Vowel Utterance

This task required the maintaining of articulatory organs in the target position and could reflect the speaker's articulation accuracy [24]. It has been reported that the acoustic features of vowels can be used to differentiate individuals with MCI from HCs [9]. Therefore, in this task, the mean and standard deviations of vowel formant frequency (F1 and F2), and fundamental frequency (F0), were measured. Besides, the triangular vowel space area and vowel articulation index were calculated as possible markers of articulatory impairments [24].

2.4.2. Tongue Twister

This task required fast movement between different places of articulation, reflecting one's attention and executive function. Wutzler *et al.* suggested that the tongue twister task may be able to indicate mild cognitive dysfunction [25]. In this task, we extracted features that reflect the speaker's speech tempo, including utterance duration, number of pauses, length of pauses, number of uttered syllables, the mean and standard deviation of syllable duration, together with the mean and standard deviation of the power of maximum and minimum.

2.4.3. Diadochokinesis

Diadochokinesis is a task that has been widely used for checking speech planning abilities. Previous studies have suggested its potential to detect individuals with MCI from HCs [11, 12]. In this task, we extracted the same features used in the tongue twister task.

2.4.4. Short Sentences

In the final task, the participants were required to read two short sentences three times at normal speed. Given that this assessment does not tap participants' memory function, it was considered to involve a lower cognitive burden than the sentence repeating task conducted in a previous study [8].

Previous studies have reported temporal changes in the speech of individuals with dementia in reading tasks [7, 8]. Moreover, Lowit *et al.* reported that individuals with MCI showed significantly smaller changes in articulation rate under habitual, fast, and slow conditions compared to HCs [26]. Therefore, in the present task, together with the temporal features measured in the previous two tasks, we also created the features of the mean utterance time ratio of tasks b) and d).

2.5. Machine Learning

Successively to the above procedures, the voices of all the participants were converted into numerical feature values in a table-data format, given that machine learning models were able to predict the cognitive classes from such features. We adopted logistic regression as a machine learning model of binary classification [27]. The objective function of this model was squared error with L2 norm regularization, which was intended to decrease the variance of the model. The fitted model was evaluated by the area under the receiver operating characteristics curve (ROC-AUC), which is robust to the imbalance of sample size for each class and independent from the threshold of models [28]. As shown in Table **1**, the numbers of participants in each cognitive class were highly imbalanced.

As the objective variable, we adopted the cognitive impairment class (*e.g.,* MCI, GCI, and MCI with GCI) or the impaired domain (*e.g.,* word memory, attention, executive function, and processing speed) assessed by the NCGG-FAT as the positive label, and HCs as the negative label. As explanatory variables, all features calculated in the procedure described in section 2.4 were adopted. We conducted no feature selections, for example, statistical tests, before splitting the data into train and test to avoid leakage leading to

Table 1.   Characteristics of participants in the four cognitive groups: HC, MCI, GCI, and MCI with GCI.

-	HC (n = 6343)	MCI (n = 1601)	GCI (n = 367)	MCI with GCI (n = 468)
Demographic Variables				
Age (Years)	73.5 ± 5.4	$74.9 \pm 5.6^{*}$	$77.2 \pm 5.9^{*,\dagger}$	$78.5 \pm 5.9^{*,\dagger,\ddagger}$
Education (Years)	12.0 ± 2.4	$11.6 \pm 2.4^{*}$	$10.8 \pm 2.2^{*,\dagger}$	$10.5 \pm 2.2^{*,\dagger}$
Male, n (%)	2658 (41.9)	757 (47.3)	230 (62.6)	262 (56.0)
Cognitive Performance				
MMSE (Score)	27.7 ± 1.9	$26.8 \pm 1.9^{*}$	$22.4 \pm 0.8^{*,\dagger}$	$22.0 \pm 1.0^{*,\dagger,\ddagger}$
Impairments Assessed with the NCGG-FAT				
Word memory, n (%)	-	548 (34.2)	-	215 (45.9)
Attention, n (%)	-	474 (29.6)	-	123 (26.3)
Executive function, n (%)	-	924 (57.7)	-	351 (75.0)
Processing speed, n (%)	-	160 (10.0)	-	112 (23.9)

*, †, ‡: statistically significant ($p < 0.01$) from HC, MCI, and GCI, respectively.

imprecise evaluations [29]. The following describes our machine learning procedure.

1. Selection of the target label, for example, MCI and HCs as the positive and negative labels, respectively.

2. 3-fold cross-validation at 10 random seeds. These different seeds were adopted to estimate the variance of data splitting because our dataset was small. Besides, we conducted 5-fold cross-validation in each splitting to optimize the hyper-parameters of the model using only train data, that is, inner cross-validation. Such a nested cross-validation ensures the relevance of validation, avoiding optimistic evaluations [30]. We adopted stratified 3-fold splitting to ensure the equal data size of each class in each splitting.

3. The calculation of 30 ROC-AUC values as a result of 3-fold cross-validation with 10 random seeds. The mean and standard deviations of the ROC-AUC values for each target were evaluated to assess the prediction ability and the model variance. Besides, we show all 30 ROC curves in one figure to illustrate the prediction ability of the machine learning models and the variance due to data subsampling with reference to the previous work [8].

3. RESULTS

3.1. Classification of Cognitive Class

The mean values and their standard deviations of the ROC-AUC values for 30 models were as follows: MCI: 0.6128 (±0.0097); GCI: 0.6731 (±0.0238); MCI with GCI: 0.7712 (±0.0180). The differences between all the prediction scores were statistically significant ($p < 0.001$). Fig. (**1**) also shows the ROC curves derived from 30 models for each target: a) MCI, b) GCI, and c) MCI with GCI, respectively.

3.2. Classification of Cognitive Impairments Assessed by the NCGG-FAT

For a more detailed analysis, we predicted the four impaired domains assessed by the NCGG-FAT from HC group through the acoustic features with machine learning models. The results of the mean and standard deviation of ROC-AUC scores for each target were 0.6641 (±0.0168), 0.6474 (±0.0164), 0.6928 (±0.0102), and 0.7586 (±0.0160) for the word memory, attention, executive function, and processing speed impairments, respectively. The results of statistical tests were as follows: word memory and attention: not significant ($p = 0.269$); executive function and word memory/attention: significant ($p < 0.001$); processing speed and all the rest: significant ($p < 0.001$). Fig. (**2**) also illustrates the ROC curves of all 30 models for each target: a) word memory, b) attention, c) executive function, and d) processing speed.

3.3. Classification of Individual Tasks

To clarify the contribution of each task to the detection of cognitive impairments, all impairments were predicted by machine learning models using acoustic features derived from Tasks a)-d) described in section 2.3. Table **2** reports ROC-AUC values for cognitive impairment class (MCI, GCI, and MCI with GCI) predicted by each task. Table **3** shows ROC-AUC values for cognitive domains assessed by the NCGG-FAT (word memory, attention, executive function, and processing speed) predicted by each task.

4. DISCUSSION

Our results indicate that sentence reading tests based on speech analysis can differentiate individuals with MCI with GCI from HCs with higher accuracy than those with MCI or
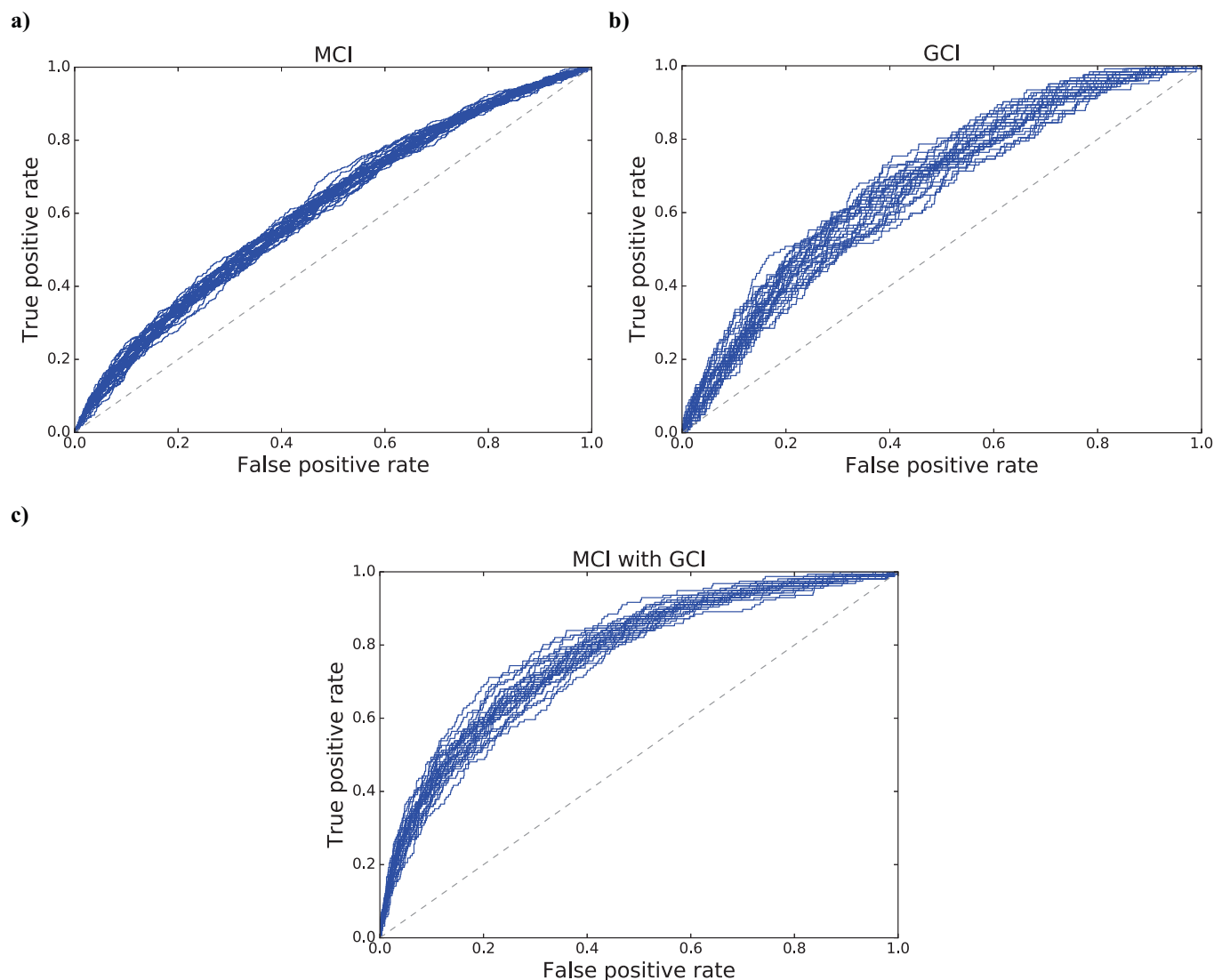
a)

b)

c)

Fig. (1). ROC curves derived from 30 models for the prediction of cognitive impairments: **a)** MCI, **b)** GCI, and **c)** MCI with GCI. The mean values and their standard deviations of the ROC-AUC values were as follows: **a)** 0.6128 (±0.0097); **b)** 0.6731 (±0.0238); **c)** 0.7712 (±0.0180). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

GCI alone. Given that MCI with GCI is a more severe cognitive impairment than MCI [18], the result is in concordance with the previous study describing that language impairment increases in severity with cognitive impairment [31].

With regard to such reading tasks, all the participants were instructed to read words and sentences at either their usual speed or faster to ensure the involvement of the executive function and information processing abilities [32, 33], leading to the high prediction scores of execution and processing impairments reported in section 3.2. However, given that all the participants were allowed to see the reading materials, the short-term memory function was not employed, which is likely the reason behind the low prediction ability of word memory impairments. In addition, considering that this reading task did not force the participants to either select their attention (as opposed to the TMT-part A) or sustain their attention until the stimuli were shown in the display (in contrast to the Continuous Performance Test [34]), the atten-

tion impairments prediction score also remained low. Furthermore, the low prediction scores of word memory and attention impairments may have led to the low MCI prediction score, as memory function is one of the risk factors for AD [35].

As shown in Tables **2** and **3**, we achieved relatively high prediction scores in tasks of tongue twisters, diadochokinesis, and short sentences. Specifically, these tasks made large contributions to the high predictability of MCI with GCI and processing speed impairments.

It can be noticed that the reading materials in both the tongue twister and diadochokinesis tasks consisted of multiple plosive consonants. We believe that compared with spontaneous speech tasks used in other studies [6, 36], both these tasks tapped the speaker's ability to rapidly change between different places of articulation, thus having a high speech planning load. This result aligns with many previous studies
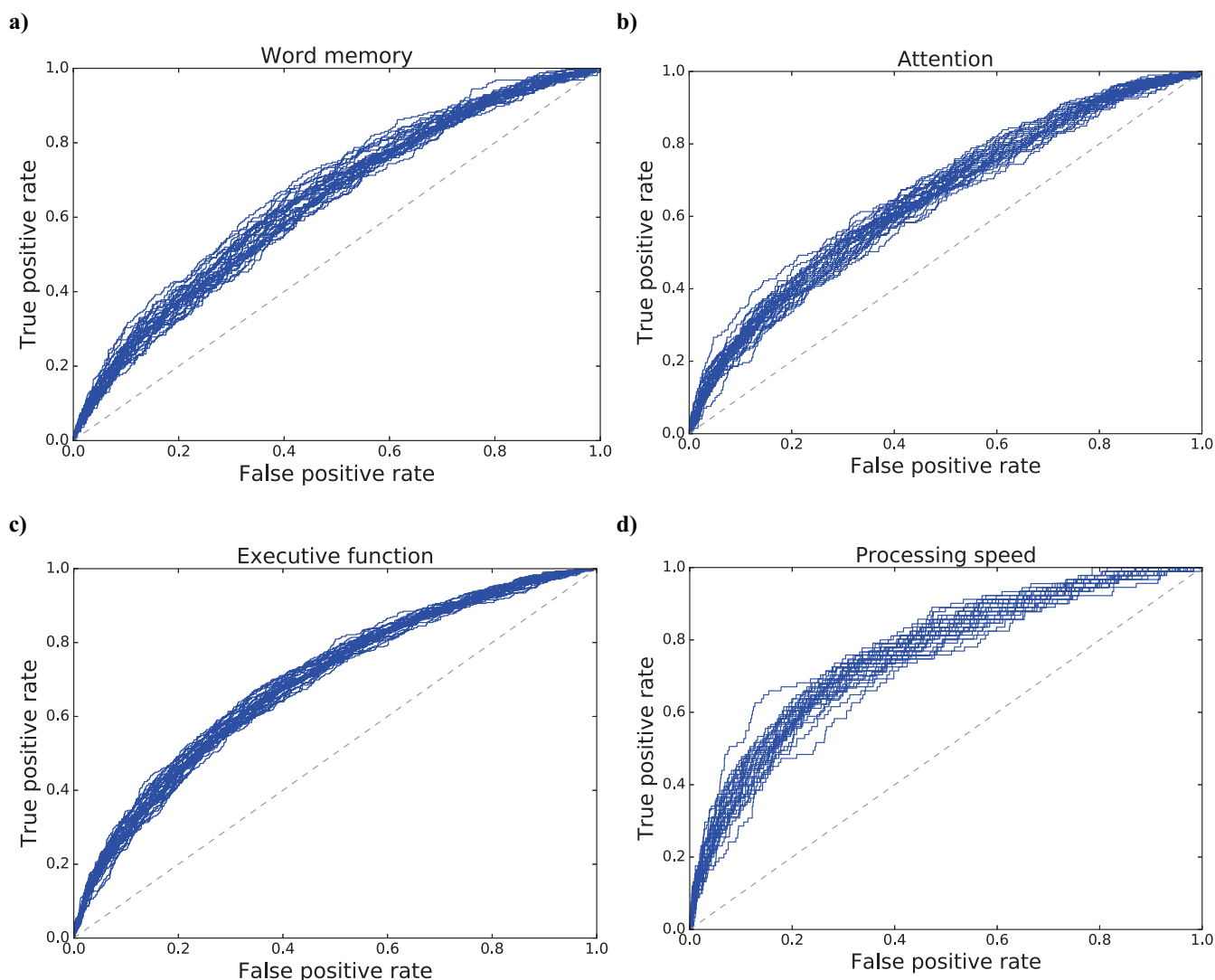
Fig. (2). ROC curves derived from 30 models for the prediction of impaired individuals' performance in four cognitive domains: **a**) word memory, **b**) attention, **c**) executive function, and **d**) processing speed. The results of the mean and standard deviation of ROC-AUC values were as follows: **a**) 0.6641 (±0.0168); **b**) 0.6474 (±0.0164); **c**) 0.6928 (±0.0102); **d**) 0.7586 (±0.0160). *(A higher resolution / colour version of this figure is available in the electronic copy of the article).*

Table 2.   ROC-AUC values for the prediction of cognitive impairment class (MCI, GCI, and MCI with GCI) using acoustic features derived from single tasks.

-	MCI	GCI	MCI with GCI
Vowel utterance	0.5471 (±0.0116)	0.6206 (±0.0202)[*]	0.6315 (±0.0225)[†]
Tongue twister	0.5975 (±0.0095)	0.6237 (±0.0193)[*]	0.7057 (±0.0132)[†‡]
Diadochokinesis	0.5826 (±0.0114)	0.6113 (±0.0227)[*]	0.6951 (±0.0175)[†‡]
Short sentence	0.5853 (±0.0127)	0.6345 (±0.0160)[*]	0.7116 (±0.0190)[†‡]

[*], [†], [‡]: statistically significant difference (*p* < 0.001) between MCI and GCI; MCI and MCI with GCI; and GCI and MCI with GCI, respectively.

on temporal changes in the speech of individuals with cognitive impairment [11, 12, 25]. It is possible that the lower speech tempo and more hesitation time in utterances is a consequence of low language processing speed and disordered speech planning abilities.

However, the prediction score based on the vowel utterance task remained low. In the task, acoustic features in the frequency domain did not seem effective in detecting individuals with cognitive impairment. It is suggested that articulation accuracy is still preserved in the early stages of

Table 3. ROC-AUC values for the prediction of cognitive impairments in different domains assessed by the NCGG-FAT (word memory, attention, executive function, and processing speed) using acoustic features derived from single tasks.

-	Word memory	Attention	Executive function	Processing speed
Vowel utterance	0.5794 (±0.0150)	0.5669 (±0.0185)	0.5864 (±0.0121) [‡]	0.5842 (0.0203)
Tongue twister	0.6180 (±0.0106)	0.6289 (±0.0226)	0.6463 (±0.0113) [*]	0.7136 (±0.0220) [†,§,‖]
Diadochokinesis	0.6112 (±0.0149)	0.6158 (±0.0157)	0.6346 (±0.0087) [*,‡]	0.6929 (±0.0203) [†,§,‖]
Short sentence	0.5966 (±0.0139)	0.6234 (±0.0164)	0.6520 (±0.0122) [*,‡]	0.7025 (±0.0238) [†,§,‖]

[*], [†], [‡], [§], [‖]: statistically significant difference ($p < 0.001$) between word memory and executive function; word memory and processing speed; attention and executive function; attention and processing speed; and executive function and processing speed, respectively.

cognitive impairment, and thus cannot be used as a marker in the detection of MCI.

According to Swets's interpretation of AUC scores, in this study, we achieved a moderate accuracy of 0.7712 (±0.0180) in separating MCI with GCI from the control group, while the prediction score for MCI and GCI remained low [37]. Given that MCI with GCI is the most severe cognitive impairment among the three groups, we believe that our test can be used as a first-step screening tool to pick out the individuals with the highest risk of developing dementia in a community. The main strengths of this study include a large sample size and the language-independent signal processing method we adopted. Unlike some previous studies which extracted linguistic features from spontaneous speech manually [6, 36], the present study only focuses on acoustic features and all the features were extracted automatically. Furthermore, compared with memory tasks and sentence repeating tasks, the reading tasks we adopted are cognitively less demanding, thus is less influenced by the participant's stress level.

This study has some limitations. The way to classify MCI in this study is based only on the cognitive tests but no assessment of instrumental ADLs, and also the exclusion criteria included only impairment of basic ADLs which are not sensitive to evaluate for functional independency that is one of the core criteria for MCI. These factors may make the classification of MCI unreliable.

It is suggested that the low MCI prediction score might be the consequence of insufficient memory and attention load. Further studies must involve a higher prediction ability of word memory and attention impairments by adding tests assessing them. One of these candidates is movie delayed recall. For example, Tóth *et al.* used immediate and delayed recall of animated movies to detect MCI [38]. It is hoped that adding these recall tests to our tasks may lead to high prediction power of MCI and GCI.

CONCLUSION

To conclude, this study was conducted to determine whether it is possible to detect individuals with cognitive impairment (MCI, GCI, and MCI with GCI) from HCs by using acoustic features extracted from speech during vocal tasks. As the research method, we employed the signal processing technique and machine learning algorithm. The results demonstrated that machine learning models can differentiate MCI with GCI from HCs with high accuracy, while MCI and GCI were not well differentiated. This might have been due to our task design, which only tapped executive function and processing speed. In order to improve accuracy, further studies must include tasks tapping memory function and attention.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study protocol was approved by the Ethics Committee of the National Center for Geriatrics and Gerontology, Japan (Approval #: 1067).

HUMAN AND ANIMAL RIGHTS

No animals were used in this study. All humans research procedures followed were in accordance with the guidelines of the Declaration of Helsinki of 1975, as revised in 2013 (http://ethics.iit.edu/ecodes/node/3931).

CONSENT FOR PUBLICATION

Both a written and verbal explanation of the present study was provided to all the participants, who provided written informed consent.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

CONFLICT OF INTEREST

NR, ZY, OY, HM, AK, UT, and SS are salaried employees of Panasonic Corporation. SH was the recipient of research funding from Panasonic Corporation to conduct this joint research.

ACKNOWLEDGEMENTS

Declared none.

APPENDIX A

Appendix A.   Detail description of four sentence reading tasks to measure cognitive impairments: a) Vowel utterances b) Tongue twister c) Diadochokinesis d) Short sentences.

Task	Description
Vowel utterances	Pronounce the five Japanese vowels for three seconds each, in the order of /i/, /e/, /a/, /o/, and /u/.
Toungue twister	Read the following short sentence including plosives (*e.g.,* /k/ and /t/) as fast as possible three times:/ki-ta-ka-ra-ki-ta-ka-ta-ta-ta-ki-ki/.
Diadochokinesis	Repeat the following syllables consisting of plosives and vowels as fast as possible for five seconds: /pa/, /ta/, and /ka/.
Short sentences	Read the following short sentences including plosives three times at the usual conversational speed: /ta-n-ke-n-ka-wa-bo-u-ke-n-ga-da-i-su-ki-de-su/ and /ki-ta-ka-ze-to-ta-i-yo-u-ga-de-te-i-ma-su/.

REFERENCES

[1]   Cabinet Office. Annual Report on the Aging Society 2017. Summary 2018.

[2]   Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12(3): 189-98 (1975).
http://dx.doi.org/10.1016/0022-3956(75)90026-6 PMID: 1202204

[3]   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 53(4): 695-9 (2005).
http://dx.doi.org/10.1111/j.1532-5415.2005.53221.x   PMID: 15817019

[4]   Morris RG. The Cognitive Neuropsychology of Alzheimer-Type Dementia. New York: Oxford University Press 1996.

[5]   Biassou N, Grossman M, Onishi K, Mickanin J,Hughes E, Robinson KM, *et al.* Phonologic processing deficits in Alzheimer's disease. Neurology 45(12): 2165-9 (1995).
http://dx.doi.org/10.1212/WNL.45.12.2165 PMID: 8848186

[6]   Hoffmann I, Németh D, Dye CD, Pákáski M, Irinyi T, Kálmán J. Temporal parameters of spontaneous speech in Alzheimer's disease. Int J Speech Lang Pathol 12(1): 29-34 (2010).
http://dx.doi.org/10.3109/17549500903137256 PMID: 20380247

[7]   Martínez-Sánchez F, Meilán JJG, García-Sevilla J, Carro J, Arana JM. Oral reading fluency analysis in patients with Alzheimer disease and asymptomatic control subjects. Neurologia 28(6): 325-31 (2013).
http://dx.doi.org/10.1016/j.nrleng.2012.07.017 PMID: 23046975

[8]   König A, Satt A, Sorin A, Hoary R, Toledo-Ronen O, Derreumaux A, *et al.* Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst) 1(1): 112-24 (2015).
http://dx.doi.org/10.1016/j.dadm.2014.11.012 PMID: 27239498

[9]   Themistocleous C, Eckerström M, Kokkinakis D. Identification of mild cognitive impairment from speech in Swedish using deep sequential neural networks. Front Neurol 9: 975 (2018).
http://dx.doi.org/10.3389/fneur.2018.00975 PMID: 30498472

[10]   Cera ML, Ortiz KZ, Bertolucci PHF, Minett TSC. Speech and orofacial apraxias in Alzheimer's disease. Int Psychogeriatr 25(10): 1679-85 (2013).
http://dx.doi.org/10.1017/S1041610213000781 PMID: 23742823

[11]   Östberg P, Bogdanović N, Wahlund LO. Articulatory agility in cognitive decline. Folia Phoniatr Logop 61(5): 269-74 (2009).
http://dx.doi.org/10.1159/000235649 PMID: 19696488

[12]   Watanabe Y, Arai H, Hirano H, Morishita S, Ohara Y, Edahiro A, *et al.* Oral function as an indexing parameter for mild cognitive impairment in older adults. Geriatr Gerontol Int 18(5): 790-8 (2018).
http://dx.doi.org/10.1111/ggi.13259 PMID: 29380503

[13]   Shimada H, Tsutsumimoto K, Lee S, Doi T, Makizako H, Lee S, *et al.* Driving continuity in cognitively impaired older drivers. Geriatr Gerontol Int 16(4): 508-14 (2016).
http://dx.doi.org/10.1111/ggi.12504 PMID: 25953032

[14]   Shimada H, Doi T, Lee S, Makizako H. Reversible predictors of reversion from mild cognitive impairment to normal cognition: a 4-year longitudinal study. Alzheimers Res Ther 11(1): 24 (2019).
http://dx.doi.org/10.1186/s13195-019-0480-5 PMID: 30867057

[15]   National Institute for Health and Care Excellence. Alzheimer's disease - donepezil, galantamine, rivastigmine and memantine (TA217) NICE technology appraisal guidance 2011.

[16]   Shimada H, Makizako H, Tsutsumimoto K, Doi T, Lee S, Suzuki T. Cognitive frailty and incidence of dementia in older persons. J Prev Alzheimers Dis 5(1): 42-8 (2018).
PMID: 29405232

[17]   Teng EL, Chui HC. The modified mini-mental state. (3MS) examination. J Clin Psychiatry 48(8): 314-8 (1987).
PMID: 3611032

[18]   Makizako H, Shimada H, Park H, Doi T, Yoshida D, Uemura K, *et al.* Evaluation of multidimensional neurocognitive function using a tablet personal computer: test-retest reliability and validity in community-dwelling older adults. Geriatr Gerontol Int 13(4): 860-6 (2013).
http://dx.doi.org/10.1111/ggi.12014 PMID: 23230988

[19]   O'Bryant SE, Humphreys JD, Smith GE, Ivnik RZJ, Graff-Radford NR, Petersen RC, *et al.* Detecting dementia with the mini-mental state examination in highly educated individuals. Arch Neurol 65(7): 963-7 (2008).
http://dx.doi.org/10.1001/archneur.65.7.963 PMID: 18625866

[20]   Shimada H, Makizako H, Doi T, Tsutsumimoto K, Lee S, Suzuki T. Cognitive impairment and disability in older Japanese adults. PLoS One 11(7): e0158720 (2016).
http://dx.doi.org/10.1371/journal.pone.0158720 PMID: 27415430

[21]   Noble K, Glosser G, Grossman M. Oral reading in dementia. Brain Lang 74(1): 48-69 (2000).
http://dx.doi.org/10.1006/brln.2000.2330 PMID: 10924216

[22]   Pattamadilok C, Chanoine V, Pallier C, Anton J-L, Nazarian B, Belin P, *et al.* Automaticity of phonological and semantic processing during visual word recognition. Neuroimage 1149: 244-55 (2017).
http://dx.doi.org/10.1016/j.neuroimage.2017.02.003

[23]   Boersma P, Weenink D. Praat: doing phonetics by computer Available at: www.fon.hum.uva.nl/praat

[24]   Skodda S, Grönheit W, Schlegel U. Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. PLoS One 7(2): e32132 (2012).
http://dx.doi.org/10.1371/journal.pone.0032132 PMID: 22389682

[25]   Wutzler A, Becker R, Lämmler G, Haverkamp W, Steinhagen-Thiessen E. The anticipatory proportion as an indicator of language impairment in early-stage cognitive disorder in the elderly. Dement Geriatr Cogn Disord 36(5-6): 300-9 (2013).
http://dx.doi.org/10.1159/000350808 PMID: 24022211

[26]   Lowit A, Brendel B, Dobinson C, Howell P. An investigation into the influences of age, pathology and cognition on speech production. J Med Speech-Lang Pathol 14: 253-62 (2006).
PMID: 19330040

[27]   Bishop C. Pattern Recognition and Machine Learning. Springer 2006.

[28]   Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 27: 861-74 (2006).
http://dx.doi.org/10.1016/j.patrec.2005.10.010

[29] Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci USA 99(10): 6562-6 (2002).
http://dx.doi.org/10.1073/pnas.102102699 PMID: 11983868

[30] Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 11: 2079-107 (2010).

[31] Alegret M, Peretó M, Pérez A, Valero S, Espinosa A, Ortega G, *et al.* The role of verb fluency in the detection of early cognitive impairment in Alzheimer's disease. J Alzheimers Dis 62(2): 611-9 (2018).
http://dx.doi.org/10.3233/JAD-170826 PMID: 29480180

[32] Wingfield A, Poon LW, Lombardi L, Lowe D. Speed of processing in normal aging: effects of speech rate, linguistic structure, and processing time. J Gerontol 40(5): 579-85 (1985).
http://dx.doi.org/10.1093/geronj/40.5.579 PMID: 4031406

[33] Salthouse TA, Coon VE. Influence of task-specific processing speed on age differences in memory. J Gerontol 48(5): 245-55 (1993).
http://dx.doi.org/10.1093/geronj/48.5.P245 PMID: 8366270

[34] Beck LH, Bransome ED Jr, Mirsky AF, Rosvold HE, Sarason I. A continuous performance test of brain damage. J Consult Psychol 1956; 20(5): 343-50.

http://dx.doi.org/10.1037/h0043220 PMID: 13367264

[35] Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, *et al.* Current concepts in mild cognitive impairment. Arch Neurol 58(12): 1985-92 (2001).
http://dx.doi.org/10.1001/archneur.58.12.1985 PMID: 11735772

[36] Mueller KD, Koscik RL, Hermann BP, Johnson SC, Turkstra LS. Declines in connected language are associated with very early mild cognitive impairment: results from the Wisconsin Registry for Alzheimer's prevention. Front Aging Neurosci 9: 437 (2018).
http://dx.doi.org/10.3389/fnagi.2017.00437 PMID: 29375365

[37] Swets JA. Measuring the accuracy of diagnostic systems. Science 240(4857): 1285-93 (1988).
http://dx.doi.org/10.1126/science.3287615 PMID: 3287615

[38] Tóth L, Hoffmann I, Gosztolya G, Vincze V, Szatloczki G, Banreti Z *et al.* A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. Curr Alzheimer Res 15(2): 130-8 (2018).
http://dx.doi.org/10.2174/1567205014666171121114930 PMID: 29165085