ORIGINAL ARTICLE

Evaluation of Vocal Recognition for Early Detection of Mild Cognitive Impairment

Mayumi Horie, Toshihide Harada, Tadayuki Iida, Satomi Aoi, Hiromi Ikeda

ABSTRACT

Objective: **As there are few means to stop the progression of dementia after its onset, it is important to delay or prevent the transition from mild cognitive impairment (MCI) to dementia. However, early detection methods for MCI have not yet been established; therefore, preventive interventions for dementia remain unavailable. In this study, we examined whether vocal recognition evaluation, which uses the time from speech presentation to judging the emotion of speech as an evaluation, could be effective for MCI screening.**

Design: **A clinical study.**

Materials and Methods: **Female of middle age or older 77 participants were recruited from the division of public relations in a city in Japan. The Japanese version of the Montreal Cognitive Assessment (MoCA-J) was used to screen for MCI. Participants were divided into two groups by neurologist consultation and MoCA-J scores: those with MoCA-J scores of 25 or below were categorized as MCI suspects, and those with a score of 26 or above as non-MCI suspects. Voice samples expressing four different emotions (acceptance, rejection, bluffing, and joking) were used in vocal recognition experiments.**

Results: **A one-second delay in the time to recognize emotions through voices was associated with an approximately 34-fold increase in the risk of MCI.**

Conclusion: **Using vocal recognition assessment was considered to be effective for early screening of MCI.**

KEY WORDS

mild cognitive impairment, ultra-early screening, vocal recognition assessment

INTRODUCTION

In Japan, the number of dementia patients reached 4.62 million in 2012, and the number of people with pre-dementia (MCI), who are at high risk of developing dementia, is estimated to be 4 million. As there are few measures to stop the progression of dementia after onset, it is important to take measures to delay or prevent the transition to dementia at the MCI stage. As such, development of methods for the early detection of MCI is necessary.

Previous studies on language function in dementia have indicated that as dementia progresses, the ability to comprehend spoken language declines[1]. However, nonverbal communication, such as facial expressions and eye contact, remains relatively intact even as dementia progresses[2,3]. Even so, the ability to read the feelings of others from their facial expressions is said to decline in dementia patients compared with young and normal elderly people[4,5]. Many studies on nonverbal communication in dementia patients have been conducted to understand the emotions of others, focusing on facial expressions, which are classified as visual information[6]. However, in communicative situations, it is important to read visual and auditory information, such as emotions, in a person's speaking voice. No studies have been conducted on vocal recognition to read these emotions in patients with dementia or MCI, to date.

However, the relationship between understanding another person's emotions in speech using auditory cues, such as speech prosody (intonation, rhythm, loudness, pitch, pitch loss), and MCI, is unknown. In the present study, we aimed to clarify the relationship between understanding others' emotions in speech as a form of vocal recognition assessment and the risk of MCI, with the aim of early detection of MCI. We further developed a vocal recognition assessment tool that can be used in conjunction with existing MCI screening methods.

OBJECTIVE

We conducted this study to implement vocal recognition assessment in individuals belonging to the group of MCI suspects and non-MCI suspects who were divided by neurologist consultation and the Montreal Cognitive Assessment (MoCA-J)[7,8], which is considered as a screening test for MCI to evaluate the effectiveness of early detection of a group of MCI suspects.

MATERIALS AND METHODS

1. Participants

This study included healthy female volunteers of middle age or older recruited by the division of public relations in a city in Japan. These women were informed regarding the objective of the study and the policy that participation was voluntary, and those who presented their consent were included in this study as participants. Those who had no medical history, neuropsychological abnormalities, or medical illnesses as determined by interview with a neurologist were considered

Received on June 3, 2022 and accepted on July 1, 2022
Faculty of Health and Welfare, Prefectural University of Hiroshima
1-1 Gakuen-cho, Mihara City, Hiroshima 723-0053, Japan
Correspondence to: Mayumi Horie
(e-mail: horie@pu-hiroshima.ac.jp)

ORCID ID:
Mayumi Horie: 0000-0003-4155-1581

Table 1: Four types of literal meanings and emotions expressed by voices and spoken phrases

		emotion expressed by voice			
		Positive		Negative	
Linguistic meaning	Positive	**Acceptance(6)**		**Bluffing(6)**	
		I give this to you	I understand it	I give this to you	I understand it
		Ok	Let's play	Ok	Let's play
		Let's go	Let's do it	Let's go	Let's do it
	Negative	**Joking(6)**	**Rejection(6)**		
		I don't know	I can't do it	I don't know	I can't do it
		I don't understand	Stop it	I don't understand	Stop it
		I hate it	No	I hate it	No

Table 2: Comparison of the group of MCI suspects and the group of non-MCI suspects with vocal expression of four emotions

	group of MCI suspects 16participants Median (25%-75%)	group of non-MCIspects 16participants Median (25%-75%)	P-value
Rejection	1732.8 (1586.5-1894.1)	1236 (1141.5-456.5)	< 0.001
Joking	2088.2 (1618.9-3248.5)	1345.8 (1161.9-549.3)	0.002
Bluffing	1861.8 (1386.3-2275.8)	1472.4 (1122.1-637.1)	0.026
Acceptance	1282.4 (1047.4-1383.0)	1082.5 (962.9-1233.2)	0.160

Table 3: The risk of MCI and the ability to recognize four voices

		(adjusted odds ratio)	
	odds ratio	95%CI	P value
Rejection	33.8	2.90—393.6	0.005
Joking	5.14	1.22—21.7	0.026
Bluffing	2.69	1.02—7.11	0.046
Acceptance	1.85	0.864—3.97	0.113

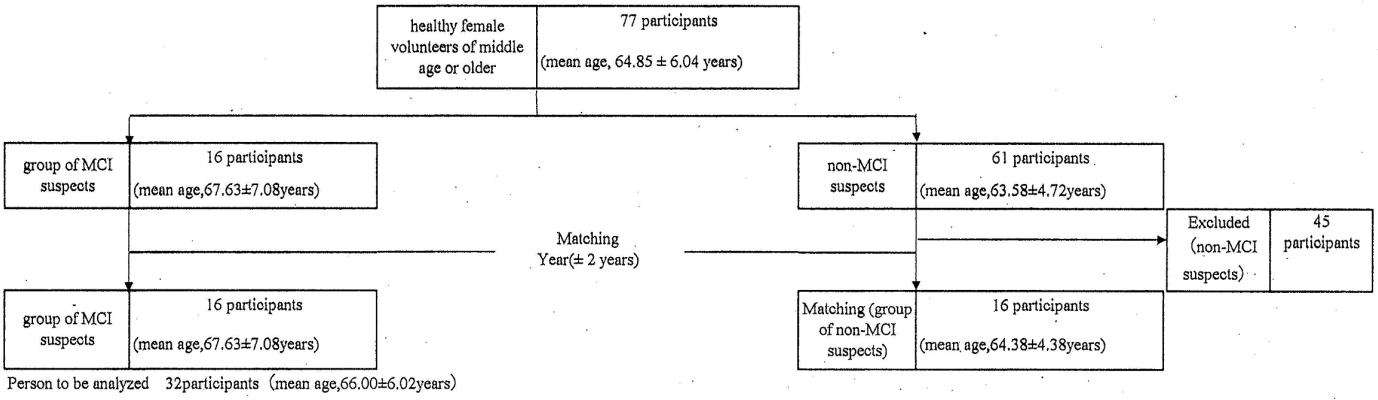

Figure 1: Person to be analyzed.

study collaborators. Furthermore, audiometry was performed in the participants to confirm that they had a hearing acuity of 1,000 Hz 30-40 dB (Pediatric Audiometer (PAS), Interacoustics Inc.). A total of 77 participants in whom the absence of significant problems with hearing acuity could be confirmed (aged 56 1 years) were included as study collaborators.

Among the 77 study collaborators (mean age, 64.85 ± 6.04 years), 16 participants with a MoCA-J score of 25 points or less and who were diagnosed with suspected MCI using the DSM-5 diagnostic criteria and CDR diagnostic criteria by a neurologist were assigned to a group of MCI suspects[7,9] (mean age, 67.63 ± 7.08 years). and 16 participants whose MoCA-J score was 26 points or higher and diagnosed as suspected non-MCI by a neurologist were matched for age (± 2 years) and assigned to a group of non-MCI suspects (mean age,64.38 ± 4.38years). These 32 participants (mean age, 66.00 ± 6.02 years) were included in the analysis (Fig. 1). This study was approved by the Ethical Committee of the Prefectural University of Hiroshima, in accordance with the Declaration of Helsinki.

2. Survey items

2-1. Cognitive Function Assessment

The MoCA-J was used to assess cognitive function as a screening test for MCI[7,8]. This test is a screening tool for MCI that assesses multidisciplinary cognitive functions in approximately 10 min. Participants receive a score out of 30 points, and it has been reported that a score of 26 points or higher in the Japanese version is within the normal range[8].

2-2. Vocal Recognition Time

A PC Tuxoscope was used to teach emotional speech, and the vocal recognition reaction time from presentation to response was used as a form of evaluation. Because there was no significant difference between the groups, with 94.3% of the MCI-suspect group and 91.9% of the non-MCI-suspect group responding correctly to vocal recognition. Therefore, the vocal recognition reaction time was measured.

2-2-1. Emotional Voice Sample

To judge emotions from Japanese voices, six phrases with a positive linguistic meaning were used to prepare 12 speeches with two different prosodies: "Acceptance," which indicates a positive emotion; "Bluffing," which indicates a negative emotion (example: using a phrase; and "OK," which has a positive linguistic meaning. We prepared a speech expressing "Acceptance" as a positive emotion when the response is "I am really OK and feels pleasure with it" and another speech expressing "Bluffing" as a negative emotion when the response is "I am OK because there is no other choice, but I do not really want to do it" containing inconsistency between emotion and linguistic meaning). In the same way, six phrases with negative linguistic meanings were used to prepare 12 speeches that have two different prosodies; "Joking," which shows positive emotion, and "Rejection," which shows negative emotion. That is, we used a total of 24 speeches expressing four types of emotion that serve as stimuli (hereinafter referred to as vocal recognition exercises)

The voice samples were also used in a previous study wherein vocal recognition assessments were conducted on people on the autism spectrum[10] (Table.1). We employed common phrases encountered in daily life that do not affect the judgment of participants depending on their ability to understand language, so that the participants could judge emotions through the prosody of voices alone.

To select a speaker who produces the vocals, we asked four women who have experience playing to produce emotional speeches with four different prosodies as they do in their daily lives.

In preparation for vocal recognition exercises, we recorded the speech after connecting an audio interface (Steinberg UR AR58125) and condenser microphone (Audio-Technica BP892cW) to a PC in a soundproof chamber. The distance between the microphone and lips was set at approximately 10 cm, and the speeches were recorded at 44.1 kHz and 16 bits.

To verify the validity of the vocal recognition exercises used to understand other people's emotions through voices, 25 adult university students (mean age of 21.5 years) were asked to listen to the speeches produced by the four women, and speeches produced by one speaker for which the mean correct answer rate for identifying the speaker's emotions expressed by voices was the highest was selected as speech that would be reproduced to the study collaborators. Among 24 vocal recognition exercises as vocal stimuli that the adult university students heard, all the students answered correctly to 22 exercises, while for the remaining two exercises of "Joking", one to two students answered incorrectly.

2-2-2. Procedures for speech hearing

As a method of presenting vocal stimulus exercises, speeches were reproduced through a PC using an audio interface (Steinberg UR AR58125) and a speaker (ADAM A5X) in a room with an equivalent noise level of 49.0 dB while the participants were instructed to keep quiet. When exercises for practice were presented, the volume of speech was adjusted for each participant to fully hear the speech.

To implement exercises, a program wherein vocal and image stimuli can be presented with timing control was used. A vocal stimulus exercise was presented, and "a picture of face expressing pleasure and a picture of face expressing anger while rejecting something" was used when speeches expressing "Joking" and "Rejection" were presented. In addition, "a picture of face expressing joy and a picture of face expressing unsatisfaction with reluctance to accept something" were used when speeches expressing "Acceptance" and "Bluffing" were presented, and "Which face is shown by the speaker when the speaker produces the speech" was asked to the participants while presenting these pictures to let the participants choose an appropriate picture. Exercises for practice that are separate from study exercises were implemented in advance to confirm whether the participants could understand the correct matching of speakers' feelings when the speeches were produced and pictures of facial expressions, and participants who could understand the correct matching participated in the study.

3. Analysis Method

Each MoCA-J, MMSE, and vocal recognition assessment were standardized, and the standardized values were used for analysis.

Mann—Whitney"s U-test was used to compare the MCI and non-MCI suspect groups, with the time taken to judge acceptance, rejection, bluffing, and joking.

To examine the extent to which each vocal index influenced cognitive decline, we used MoCA-J judgments as the objective variable, and examined the influence of risk factors using binomial logistic regression analysis. Before selecting the explanatory variables, the correlations between acceptance, rejection, bluffing, and joking were examined using Spearman's correlation coefficients. None of the variables showed a strong correlation ($r < 0.7$), and were therefore considered appropriate as explanatory variables.

A level of significance of 5% or lower was defined as statistically significant for all tests.

RESULTS

1. Comparison of the group of MCI suspects and the group of non-MCI suspects

The ability to recognize the vocal expression of four emotions (acceptance, rejection, bluffing, and joking), was compared between the group of MCI suspects and the group of non-MCI suspects using the Mann-Whitney's U test. There was a significant difference in the ability to recognize the vocal expressing rejection ($p = 0.0001$), joking ($p = 0.002$), and bluffing ($p = 0.026$) between the groups, and the time to react to these voices was longer in the MCI suspects group. However, there was no significant difference in vocal expressing acceptance ($p = 0.16$) between the groups (Table 2).

2. Impact of decreased cognitive functions on vocal recognition

Binomial logistic regression was performed to obtain the adjusted odds ratio and its 95% confidence interval to evaluate the relationship between the risk of MCI and the ability to recognize four voices. The results of the analysis demonstrated that recognition of emotions of rejection, joking, and bluffing was related to a risk of MCI, with an adjusted odds ratio of 33.80 ($p = 0.005$) for rejection, 5.14 ($p = 0.026$) for joking, and 2.69 ($p = 0.046$) for bluffing. There was no relationship between recognition of acceptance and the risk of MCI (Table 3).

DISCUSSION

The effectiveness of vocal recognition assessment to detect the risk of MCI was verified in addition to that of MoCA-J, and it was revealed that the speed of judgment of emotions (rejection, joking, and bluffing) through voices as auditory perception ability was significantly lower in the group of MCI suspects than in the group of non-MCI suspects.

However, in the judgment of emotion of acceptance, in which linguistic meaning and emotion were consistently positive, there was no difference between the group of MCI suspects and the group of non-MCI suspects, indicating that speeches expressing acceptance were easily understandable. There was a trend that the speed of judging the emotion of acceptance through voices was the highest among the four emotions in both groups. Therefore, the speed of judging the emotion of acceptance through voices cannot be an effective indicator in screening for MCI.

It was revealed that the speed of judging emotion of rejection is most sensitive to the screening of the risk of MCI. In previous studies, when a speech with an angry voice expressing negative emotion was presented to healthy elderly people (mean age of 70.6 years), the time to understand the meaning expressed in the speech was significantly prolonged, which suggests that information provided with an angry voice may inhibit the function of processing semantic information because of the strong impact of the emotion[11]. The results of this study demonstrated that the speed of judging emotion of rejection, which expresses negative emotion, was lower than that of acceptance, which expresses positive emotion in both groups (Table 2). However, in terms of the ability to detect the risk of MCI, it was revealed that the risk of MCI increases by 34 times as the time to recognize the emotion of rejection through voices delays by one second.

Furthermore, it was demonstrated that the risk of MCI can be detected by the speed of judging emotions of joking and bluffing through voices. The risk of MCI increases by approximately five times as the time to recognize the emotion of joking delays by one second;

and the risk of MCI increases by approximately three times as the time to recognize the emotion of bluffing delays by one second. The speed of judging emotions of joking and bluffing though voices did not have a higher ability to detect the risk of MCI compared to that of rejection, but the speed of judging these emotions can be used to screen for the risk of MCI.

Based on the above, it was suggested that measurement of the speed of judging emotions of rejection, joking, and bluffing using voices expressing these emotions as vocal recognition assessment is effective for early identification of people belonging to the group of MCI suspects.

In the future, we will consider the possibility of gauging the ability to understand the intentions of others using speech prosody as a screening tool for MCI detection.

Dementia includes Alzheimer's disease (AD) and cerebrovascular dementia (VD), but AD is the most common, accounting for about 70% of cases.

In recent years, Amyloid imaging with PET is used as an ultra-early diagnosis of AD. Amyloid plaques associated with AD accumulate in the cerebral cortex even before the onset of dementia. An amyloid PET agent, [C-11]PiB, accumulates in the posterior cingulate gyrus, precuneus, and prefrontal cortex in AD[12]. It has been reported that the [C-11] PiB test for MCI shows a positivity rate of approximately 60 0%; this has been noted as the first finding suggesting a transition to AD[13,14].

These parts of the brain are closely associated with circuits such as the hippocampus, fornix, prethalamic area, posterior cingulate gyrus, and parahippocampal gyrus, which are involved in memory, and the frontal association area. It is also involved in circuits such as the amygdala, medial dorsal nucleus of the thalamus, anterior cingulate gyrus, prefrontal cortex, and temporal lobe pole, which are involved in emotional function.

In previous studies on emotions, Studies in healthy adults have shown that the dorsomedial prefrontal cortex is involved in understanding the intentions of others, such as the theory of mind[15-19], and that the dorsomedial prefrontal cortex is activated in the same way, not only to understand others' intentions in visual presentation tasks, but also in audio prosody presentation[20].

From these, it is speculated that the deterioration of brain function related to memory and emotion is a factor that delays the judgment of voice prosody recognition in individuals with MCI.

One limitation of this study was the small number of MCI cases. This is because MCI feels that her memory has deteriorated by herself, but she rarely visits a medical institution because it does not interfere with her life. In this study, we discussed 21% of healthy middle-aged and elderly volunteer women detected as suspected MCI. Therefore, to explain the relationship between MCI and the results of vocal recognition assessment, it is necessary to investigate many clinical cases in which physicians have made a definitive diagnosis of MCI / healthy subjects.

REFERENCES

1. Jeffrey L, Cummings D, Benson F, Hill MA, Read S. Aphasia in dementia of the Alzheimer type. AAN 1985; 35: 394-397.
2. Kohase Y, Aihara A, Nakamura A, Ogasawara A, Inoue T. "Research Project on Improving the Quality of Care for the Elderly with Dementia and Their Families" Report. Dementia Care Research and Training Obu Center 2006; Heisei 18nd Year Dementia Nursing Care Research Report: 61-66.
3. Kohase Y, Aihara A, Nakamura A, Ogasawara A. Research project report on communication in dementia care. Dementia Care Research and Training Obu Center 2007; Heisei 19nd Year Dementia Nursing Care Research Report:1-11.
4. Henry JD, Ruffman T, McDonald S, *et al*. Recognition of disgust is selectively preserved in Alzheimer's disease. Neuropsychologia 2008; 46: 1363-1370.
5. Maki Y, Yoshida H, Yamaguchi T, Yamaguchi H. Relative preservation of the recognition?of positive facial expression "happiness" in Alzheimer disease. Int Psychogeriatr 2013; 25: 105-110.
6. Koumi H, Kishikawa Y, Sonoda K, lshii H, Narumoto J. A study for facial recognition of mild Alzheimer's Disease patients. Bulletin of Aino Gakuin 2006; 20: 9-23.[in Japanese, English abstract].
7. Ziad SN, Natalie AP, Valérie B, Simon C, *et al*. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005; 53: 695-699.
8. Fujiwara y, Suzuki H, Yasunaga M, *et al*. Brief screening tool for mild cognitive impairment in older Japanese: validation of the Japanese version of the Montreal Cognitive Assessment. Geriatr Gerontol Int 2010; 10: 225-232.
9. Fage BA, Chan CH, Gill SS, *et al*. Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a community setting. Cochrane Database Syst Rev 2015; 2: CD010860.
10. Horie M, Okamura H.Exploring a method for evaluation of preschool and school children with autism spectrum disorder through checking their understanding of the speaker's emotions with the help of prosody of the voice. Brain Dev 2017; 39(10): 836-845.
11. Kohase Y, Aihara A, Nakamura A, *et al*. Research project report on multidimensional approach and rehabilitation for improving QOL of elderly people with dementia in institutions. Dementia Care Research and Training Obu Center 2012; Heisei 24nd Year Dementia Nursing Care Research Report:1-22.
12. Matsuda H, Imabayashi E: Molecular neuroimaging in Alzheimer's disease. Neuroimaging Clin N Am 2012; 22: 57-65.
13. Rowe CC, Ng S, Ackermann U, *et al*. Imaging beta-amyloid burden in aging and dementia. Neurology 2007; 68: 1718-1725.
14. Forsberg A, Engler H, Almkvist O *et al*. PET imaging of amyloid deposition in patients with mild cognitive impairment. Neurobiol Aging 2008; 29: 1456-1465.
15. Fletcher PC, Happe F, Frith U, *et al*. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 1995; 57: 109-128.
16. Gallagher HL, Happe F, Brunswick N, *et al*. Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia 2000; 38: 11-21.
17. Ferstl EC, Cramon DY. What does the frontomedian cortex contribute to language processing:Coherence or theory of mind?; NeuroImage 2002; 17: 1599-1612.
18. Castelli F, Happ F, Frith U, Frith C. Movement and mind:A functional imaging study of perception and interpretation of complex intentional movement patterns. NeuroImage 2000; 12: 314-325.
19. Calder AJ, Lawrence AD, Keane J, *et al*: Reading the mind from eye gaze. Neuropsychologia 2002; 40: 1129-1138.
20. Homma M, Imaizumi S, Ozawa Y, Maruishi M, Muranaka H. Neural Mechanism for Inferring a Speaker's Mind from Speech. Jpn. J. Logop. Phoniatr 2007; 48(1): 9-18. [in Japanese, English abstract].