Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 1016–1022
Marseille, 20-25 June 2022

Elderly Conversational Speech Corpus with Cognitive Impairment Test and Pilot Dementia Detection Experiment Using Acoustic Characteristics of Speech in Japanese Dialects

Meiko Fukuda[1], Maina Umezawa[2], Ryota Nishimura[1], Yurie Iribe[2], Kazumasa Yamamoto[3], Norihide Kitaoka[4]
[1]Tokushima University, Tokushima, Japan
[2]Aichi Prefectural University, Aichi, Japan
[3]Chubu University, Department of Computer Science, Kasugai, Japan
[4]Toyohashi University of Technology, Toyohashi, Japan
[1]fukuda.meiko@tokushima-u.ac.jp

Abstract

There is a need for a simple method of detecting early signs of dementia which is not burdensome to patients, since early diagnosis and treatment can often slow the advance of the disease. Several studies have explored using only the acoustic and linguistic information of conversational speech as diagnostic material, with some success. To accelerate this research, we recorded natural conversations between 128 elderly people living in four different regions of Japan and interviewers, who also administered the Hasegawa's Dementia Scale-Revised (HDS-R), a cognitive impairment test. Using our elderly speech corpus and dementia test results, we propose an SVM-based screening method which can detect dementia using the acoustic features of conversational speech even when regional dialects are present. We accomplish this by omitting some acoustic features, to limit the negative effect of differences between dialects. When using our proposed method, a dementia detection accuracy rate of about 91% was achieved for speakers from two regions. When speech from four regions was used in a second experiment, the discrimination rate fell to 76.6%, but this may have been due to using only sentence-level acoustic features in the second experiment, instead of sentence and phoneme-level features as in the previous experiment. This is an on-going research project, and additional investigation is needed to understand differences in the acoustic characteristics of phoneme units in the conversational speech collected from these four regions, to determine whether the removal of formants and other features can improve the dementia detection rate.

Keywords: conversation, super-elderly, speech corpus, dementia

1. Introduction

Although drugs are available to slow the progression of dementia, fundamental treatments are still under development, so it is important to diagnose individuals with the disease as early as possible. The World Health Organization (WHO) estimates that global rates of dementia will triple from current levels by 2050 (World Health Organization, 2021). In Japan, the number of elderly people with dementia is expected to reach 7 million by 2025, or about one in five people aged 65 or older. As a result, the Japanese government has made the early detection of dementia a national priority(Ministry of Internal Affairs and Communication, Japan, 2015).

The Mini-Mental State Examination (MMSE) is widely used internationally as a neuropsychological screening test for dementia, while Hasegawa's Dementia Scale-Revised (HDS-R) (Imai and Hasegawa, 1994) is used more often in Japan and East Asia. These tests involve having subjects answer simple, verbal questions and perform some task. These screening tests are convenient and high sensitivity but not sufficient to make a definitive diagnosis of dementia. A combination of medical examinations and tests is needed to accurately diagnose dementia and differentiate it from other diseases with similar symptoms. These diagnostic methods include detailed interviews with doctors, blood tests, and CT or MRI imaging. Early diagnosis and treatment are essential to slow the progression of the disease, however, many older people are reluctant to undergo dementia testing in a hospital setting.

Figure 1: Regions of Japan where we recorded the speech of elderly speakers.

In order to encourage more elderly people to undergo dementia testing, care must be taken not to offend or upset them. To lower the mental burden during assessment, methods of screening for dementia using

Table 1: Number of Tokushima speakers in each age group, their gender and dementia test results

Dementia tendency	Sex	60-69	70-79	80-89	90-99	Subtotal
Positive	Male	0	1	5	0	6
	Female	0	1	6	4	11
Negative	Male	0	1	1	1	3
	Female	0	0	10	2	12
	Total	0	3	22	7	32

Table 2: Number of Aichi speakers in each age group, their gender and dementia test results

Dementia tendency	Sex	60-69	70-79	80-89	90-99	Subtotal
Positive	Male	0	0	0	1	1
	Female	0	1	3	4	8
Negative	Male	1	2	2	1	6
	Female	0	0	6	1	7
	Total	1	3	11	7	22

Table 3: Number of Chiba speakers in each age group, their gender and dementia test results

Dementia tendency	Sex	60-70	70-79	80-89	90-99	Subtotal
Positive	Male	0	1	2	0	3
	Female	0	0	6	0	6
Negative	Male	0	0	1	0	1
	Female	0	2	2	0	4
	Total	0	3	11	0	14

Table 4: Number of Mie speakers in each age group, their gender and dementia test results

Dementia tendency	Sex	60-70	70-79	80-89	90-99	Subtotal
Positive	Male	0	0	1	1	2
	Female	0	0	4	0	4
Negative	Male	0	1	0	0	1
	Female	0	1	4	1	6
	Total	0	2	9	2	13

only everyday conversational speech have been developed (Satt et al., 2014; Orimaye et al., 2014; Luz et al., 2018). Aramaki analyzed the results of automated speech recognition and reported a relationship between certain linguistic features and dementia (Aramaki et al., 2016). Tanaka et al. (Tanaka et al., 2017) and Ujiro et al. (Ujiro et al., 2018) used avatars to ask subjects questions, and then modeled the acoustic features, linguistic features and dialogue characteristics of their replies. They observed that the acoustic and linguistic features of their subjects' speech were often affected by the use of regional dialects.

Dialects are thought to have different developments in phonology, grammar, and vocabulary in one region than in other regions, and have a language system unique to one region. Elderly people tend to use regional dialects more often than younger people, and to use these dialects in a purer form, thus it is likely (National Institute for the Humanities, National institute for Japanese Language and Linguistics, 2011), when using spontaneous speech to screen the elderly for dementia, that the results will be influenced by their use of the dialect of the region where they grew up or lived. There has been no research on the use of speech for dementia testing that takes the effect of Japanese dialects into account; mainly, we focus on effects of acoustic features in this study. According to Misao Tojo, there are 16 dialects in Japan (Kudo et al., 1996). Cooperating nursing homes were found in four regions of Japan; Chiba, Aichi, Mie and Tokushima prefectures (Figure 1). To develop more effective dementia screening methods and explore the effect of regional dialects on dementia test results, we developed a speech corpus consisting of the conversational speech of elderly people from these four regions. The Japanese dialects spoken in Chiba, Aichi, Mie and Tokushima are classified as the Kanto, Tokai-Tosan, Kinki and Shikoku dialects, respectively. The interviewees were also screened for dementia using the HDS-R test, and the speech data was annotated with the results.

In order to develop more accurate screening methods, we first attempted to detect signs of dementia in the speech of elderly people in Aichi and Tokushima, excluding certain acoustic features to minimize the effect of the difference in dialects. In a second experiment, we attempted to detect signs of dementia in speech which included four regional dialects. The rest of this paper is organized as follows. We describe our new corpus of conversational elderly speech in Section 2. Our proposed method for dementia screening using conversational speech, our experimental procedure, and our results when screening speech in two Japanese dialects are explained in detail in Section 3. In Section 4, we investigate the performance of our proposed method when widening the dialectal area to include four regions. Finally, we conclude this paper in Section 5.

2. Collection of Conversational Elderly Speech

Our research project was approved by the ethics committees of Nagoya University, Tokushima University and Toyohashi University of Technology in Japan. All of the participants provided written, informed consent for the recording of their speech. Participants with dementia were asked to participate in the recording only with the consent of their families as well as themselves. Personal information is protected when used for research.

2.1. Participants and Recording Regions

We collected read speech from elderly Japanese participants for automatic speech recognition, and constructed a speech corpus of super-elderly speech (Fukuda et al., 2020). At the same time, we also recorded conversational speech of 159 subjects for dementia detection in the form of free conversation between the participants and our interviewers. However, we did not include the speech of all of the HDS-R session participants in our super-elderly corpus. The medical histories of the participants except dementia were not taken into consideration during the selection process.

Based on participant's HDS-R scores, each subject was classified as either suffering from dementia or as not having symptoms of dementia. As a result, 26 participants (19.1%) of our 128 participants in Tokushima and Aichi were judged to be impaired by dementia (Tables 1 and 2). However, since there were 102 elderly participants without symptoms of dementia, and only 26 participants with dementia, we decided to only use the conversational speech data of 28 participants without dementia (selected randomly) and of 26 participants with dementia in Tokushima and Aichi for our investigation. We then extracted acoustic features from the conversational speech of these 54 participants and compared the speech characteristics of from the impaired and unimpaired groups. Table 1-4 describe their age, gender, participants with dementia tendency distribution. Participants' ages ranged from 66 to 98.

2.2. Speech Recording

2.2.1. Recording Devices and Environment

The devices used to record the read speech of the study participants were a lapel microphone (Sony ECM-88B), a desktop microphone (Audio-Technica AT9930) and an 8-track field recorder (Tascam DR-680MKII). However, in Aichi a linear PCM recorder (Tascam DR-05 ver. 2) and the same Audio-Technica desktop microphone were used The WAV format was used for the recorded speech files (16kHz, 16-bit, Mono). Most of the speech recording sessions were held at elderly care facilities, but some were recorded at a university. Since we did not use soundproof rooms for recording, some background noise, such as the sound of an air conditioner or the voices of people outside the room, are included in the recorded speech files. The level of background noise is approximately 40-45 dB.

2.2.2. Recording Procedure

Before recording, we explained the purpose of the study and our data collection procedure. Only those who agreed to participate after listening to our explanation took part in the recording sessions. Participants were first asked to read phonetically rich sentences aloud, and the speech of those who could do so successfully was included in our super-elderly speech corpus (Fukuda et al., 2020). An interviewer then evaluated the participants for symptoms of dementia using the HDS-R. The free conversation between the participants and interviewers after the HDS-R evaluations was also recorded. It is the speech recorded during the HDS-R evaluations and the free conversation that followed which was used in this study on early detection of dementia using the acoustic features of speech.

2.2.3. The Hasegawa's Dementia Scale-Revised (HDS-R)

The HDS-R (Kato, 1991; Imai and Hasegawa, 1994) is the most widely used dementia screening test in Japan than the Mini-mental status examination (MMES) (Kurlowicz and Wallace, 1999). Examinees respond verbally to nine spoken questions eliciting information such as the subject's age, the current date or year, and location, performing simple mathematical calculations, recalling the names of five everyday objects mentioned earlier in the evaluation, etc. These interviews take only five to ten minutes. Unlike the MMSE, participants are not asked to draw a picture or transcribe sentences as performance tests. Therefore, the HDS-R evaluation might do not impose much of a mental burden on participants but is more sensitive than the MMSE (Kim et al., 2005). The highest possible HDS-R score is 30, and a score of 20 or lower is considered to be an indication of dementia. The scoring method is defined in the HDS-R testing materials, and our examiners scored the participants' answers using the suggested method. In this study, we confirmed each participant's dementia diagnosis with the staff of the nursing homes where they resided, when applicable, to verify

Table 5: Average duration of silent pauses in phonemes

Position of silent pause	Average length (sec) Dementia	W/o dementia
before /z/	0.11	0.02*
before /q/	0.05	0.01*
before /hy/	0.12	0.04*

$p<0.01$:**, $p<0.05$:*

Table 6: Average duration of silent pauses in sentences

Position of silent pause	Average length (sec) Dementia	W/o dementia
before /z/	0.73	0.21*
before /w/	7.11	3.25**
before /t/	27.0	16.1**
before /n/	25.6	15.0**
before /hy/	1.57	0.42**
before /N/	5.19	2.28**
before /K/	25.0	16.1*

$p<0.01$:**, $p<0.05$:*

impairment due to dementia. Since our analysis target is conversational speech, we chose to use the HDS-R since it only uses verbal responses for dementia diagnosis.

3. Preliminary Test of Proposed Method

3.1. Pre-processing of Speech Data

The recorded speech data was divided into sentence units, the entire corpus was transcribed, and the transcription of the speech data was verified and edited manually by trained employees who listened to the recorded speech data. Forced phoneme alignment was performed on the conversational speech of Aichi and Tokushima using Julius, the Japanese speech recogniser (Lee and Kawahara, 2009), based on acoustic models trained using the S-JNAS corpus (Baba et al., 2001).

3.2. Acoustic Features Related to Dementia

In our preliminary test (Umezawa et al., 2019), acoustic features related to dementia were targeted in the elderly speech collected in Aichi and Tokushima prefectures. Using the widely recognized acoustic symptoms of dementia, we extracted 608 features from phoneme units and 6 features from sentence units in the conversational speech. Examples of features selected from phoneme units are the number and duration of silent pauses, differences in the gradient and elevation of pitch and power at transitions from consonants to vowels, MFCCs, ΔMFCCs, formants, and differences in the gradients and elevation of pitch and power at the onset of consonants. The features selected from sentence units are the number and duration of silent pauses, speaking rate (morpheme/speaking time), pitch at the end of the sentence, amplitude deviation of the third from the first formants, and coefficient variation in pitch and power.

To verify that there were correlations between the selected speech features and symptoms of dementia, Welch's t-tests were applied to the features mentioned above when comparing the speech of participants with and without symptoms of dementia. Significant differences in the speech of impaired and unimpaired participants were found in 44 of the acoustic features of phoneme and sentence units in the Aichi data, and in 51 features in the Tokushima data. When using a combination of speech data from the two regions, significant differences were observed in 64 acoustic features. Examples of features with significant differences in phoneme units are the number and duration of silent pauses, gradient and elevation differences in pitch and power at the transitions from consonants to vowels, gradient and elevation differences in pitch and power at the onset of consonants, MFCCs and ΔMFCCs. Acoustic features with significant differences in the sentence units were number and length of silent pauses, and coefficient variation in pitch and power. As a clear example, the results for duration and number of silent pauses inserted before consonants are shown in Tables 5 and 6. These results show that short pauses inserted in front of certain consonants tended to be longer and more frequent in the group with dementia, confirming that people with dementia have difficulty speaking smoothly. As another example, the power gradients at the transitions from consonants to vowels are shown in Table 7. As for the voiceless friction consonants, /sh/ and /s/ showed an increase in the power difference in dementia-prone subjects. For /ch/, which is also a voiceless friction consonant, there was a tendency to increase the power difference, although it was not significant.

Table 7: Comparison of Δ power between consonant and subsequent vowel

Position of silent pause	Average Dementia	W/o dementia
after /s/	11.4	9.24**
after /sh/	7.90	6.82*
after /ch/	11.35	9.78
after /m/	3.09	3.45*

$p<0.01$:**, $p<0.05$:*

3.3. Dementia Detection Experiment

Our main dementia detection experiment using the selected speech features was then conducted. Our classifier used a linear kernel of Support Vector Machine (SVM), and the statistical method used for estimation was 10-fold cross-validation. Because the units and scales of the speech features differed, the features were normalized before classification, allowing us to achieve very high classification accuracy. First, we evaluated the speech data from Aichi and Tokushima separately, then we combined the data of both groups of participants to perform a second evaluation. Compared to when the speech from each region was evaluated separately, the dementia detection accuracy rate fell by 27% when the speech data of elderly participants who spoke different dialects of Japanese were combined. Since the speech data of the Aichi and Tokushima participants were recorded under the same conditions, this drop in the accuracy rate suggests a significant dialect difference between the speech data of the two groups.

3.4. The Differences Between Dialects of Aichi and Tokushima

We then analyzed differences between the dialects of the Aichi and Tokushima participants who had not been diagnosed with dementia, especially differences which would affect the acoustic features of their speech. Figures 2 and 3 show spectrograms of the Japanese word "densha" ("train") as spoken by elderly people from Aichi and Tokushima, respectively. Kisibe et al. (Kishie and Yoshihiro, 2006) reported that the beginning of the pronunciation of the Da line in the Tokushima dialect is a nasal sound. In a previous study (Hattori et al., 1958), The nasalized vowel had a stronger component around 250 Hz, so it was thought that the same characteristics would appear in consonants. Looking at our Tokushima speaker's /d/ spectrogram, red square in Figure 3, we can see formants at low frequencies, so this formant was considered a nasal formant. Another example is that the Aichi speaker devoices the vowel /u/ (data not shown). This feature is not appear in Tokushima speaker. Thus, these dialects' deference might appear in the information of voice tract, and MFCC, ΔMFCC, and formants may be the acoustic features primarily responsible for differences in pronunciation between the Aichi and Tokushima dialects. Based on this observation, we then performed Welch's t-tests between various acoustic features extracted from the speech of our elderly Aichi and Tokushima participants without symptoms of dementia, in order to eliminate the influence of dementia on the observed acoustic features. Our results revealed significant differences in MFCC, ΔMFCC and formants of elderly speakers from Aichi and Tokushima. Table 8 shows the results for the formants.

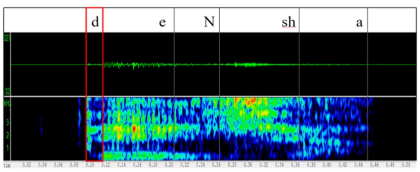

Figure 2: Spectrogram of "densha" ("train") in the Aichi dialect

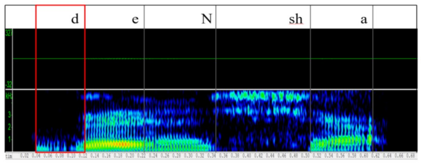

Figure 3: Spectrogram of "densha" ("train") in the Tokushima dialect

Table 8: Formant exhibiting significant difference in Aichi and Tokushima by Welch's t-tests

	Aichi	Tokushima
F1	348	311**
F2	1768	1923*
F3	3030	3463**
F4	4346	4927**

p<0.01:**, p<0.05:*

3.5. Exclusion of Dialect-Associated Features

We then repeated our dementia detection experiment, but this time we excluded the dialect-associated acoustic features (MFCC, ΔMFCC and formants) from the other acoustic features described in Section 3.2. The results of this experiment are shown in Table 10. When the formants, MFCC and ΔMFCC, which reflect the articulation-related information, were removed individually or in combination, the highest dementia detection accuracy rate (weight average: 91.3%) was achieved when only the formant features were removed. This is an improvement of 16.9% compared to when no features were removed shown in Table 9 and 10.

4. Detection of Dementia in Speech from Four Regions

We also attempted to detect signs of dementia in speakers from the other two dialectal regions of Japan, i.e., in Chiba and Mie prefectures, in addition to Aichi and Tokushima prefectures. Here, to simplify the experiment, we only used 54 sentence-based acoustic features, which did not require phoneme-level alignment. Among these features, significant differences between speakers with and without dementia were found in silent pauses, mean vectors of 10- and 12-dimensional MFCCs, the mean vectors of the first two dimensions of ΔMFCCs, and the standard deviation vector of 11-dimensions of ΔMFCC (Table 11). These are all features for which significant differences were found

Table 9: Dementia detection rates with formant

Dementia tendency	Precision	Recall	F-measure
Positive	0.760	0.731	0.745
Negative	0.759	0.768	0.772
Weight average	0.759	0.759	0.759

Table 10: Dementia detection rates without formants

Dementia tendency	Precision	Recall	F-measure
Positive	0.862	0.962	0.909
Negative	0.960	0.857	0.857
Weight average	0.913	0.907	0.907

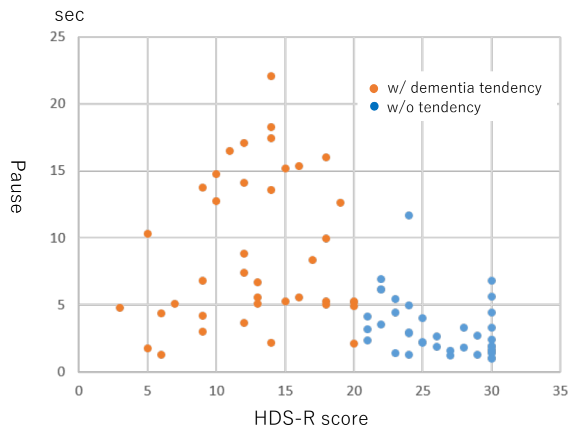

Figure 4: Scatter plot showing relationship between duration of silent pauses and HDS-R score. Each dot represents one speaker.

even when only evaluating speakers from Aichi and Tokushima. For example, the scatter plot in Figure 4 shows negative correlation between the duration of silent pauses and a negative HDS-R dementia diagnosis (correlation co-efficient = -0.498). However, regarding the participants whose HDS-R score is around 5, their pause duration is shorter than those whose HDS-R score is about 10 to five. The cause of this tendency is that they quickly say "No" against the question of HDS-R without thinking about these questions. In this experiment, only the feature representing the difference between the amplitudes of the first and third formants was excluded to eliminate the influence of dialect.

4.1. Experimental Result

When speech from the study participants in all four regions was combined, the dementia detection rate fell to 76.6% (Table 12). However, the acoustic features of the phoneme units representing dialectical differences were not excluded in this experiment. In future work, we would like to examine differences in the acoustic characteristics of phoneme units in speech collected from all four of these regions, to determine whether the removal of phoneme formants and other dialect-related features improves accuracy when detecting speakers with symptoms of dementia.

5. Conclusion

We recorded interviews with 128 elderly subjects whose ages ranged from 66 to 99, in four regions in Japan, and constructed a corpus of conversational, elderly speech. When comparing the speech of elderly speakers who scored from 0 to 20 on the HDS-R dementia test with those who scored 21 or higher, significant differences were observed in 64 acoustic features, including an increase in the duration and frequency of silent pauses. We then used these features to classify elderly study participants from two regions of Japan where different dialects of Japanese are spoken into two groups, those impaired by dementia and those who were not impaired, and achieved dementia detection accuracy rate of 91.3%. We found that formants, MFCC and ΔMFCC, which are related to articulation-related information, were different in the speech of residents living in each area reflecting dialectical differences. By excluding these dialect-related features, accuracy in dementia detection was improved by almost 15%. Next, we attempted to detect speakers with symptoms of dementia using speech we collected from four regions of Japan using sentence-based features. In this experiment, only the feature representing the difference between the amplitudes of the first and third formants was excluded, in order to eliminate the influence of dialects. When speech from the four regions was combined, the detection rate when attempting to identify speakers with symptoms of dementia was just 76.6%.

6. Acknowledgements

This study was partially supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research, Grant Numbers, 19H01125 and 19K12022, Japan and by ROIS NII Open Collaborative Research 2019, Grant Numbers 19S0403, Japan.

7. Bibliographical References

Aramaki, E., Shikata, S., Miyabe, M., and Kinoshita, A. (2016). Vocabulary size in speech may be an early indicator of cognitive impairment. *PloS one*, 11(5):e0155195.

Baba, A., Yoshizawa, S., Yamada, M., Lee, A., and Shikano, K. (2001). Elderly acoustic model for large vocabulary continuous speech recognition.

Fukuda, M., Nishizaki, H., Iribe, Y., Nishimura, R., and Kitaoka, N. (2020). Improving speech recognition for the elderly: A new corpus of elderly japanese speech and investigation of acoustic modeling for speech recognition. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6578–6585.

Hattori, S., Yamamoto, K., and Fujimura, O. (1958). Nasalization of vowels in relation to nasals. *The Journal of the Acoustical Society of America*, 30(4):267–274.

Imai, Y. and Hasegawa, K. (1994). The revised hasegawa's dementia scale (hds-r)-evaluation of its usefulness as a screening test for dementia. *Hong Kong Journal of Psychiatry*, 4(2):20.

Kato, S. (1991). Development of the revised version

Table 11: Acoustic features in sentence unit in four regions participants

Feature	average	
	Dementia	w/o dementia
silent pause	9.651	3.203**
mean MFCC (10 dimension)	-4.409	-3.024*
mean MFCC (12 dimension)	-3.266	-1.868*
mean ΔMFCC (2 dimension)	0.0007	0.014**
standard deviation of ΔMFCC (11 dimension)	1.261	1.305*

$p<0.01$:**, $p<0.05$:*

Table 12: Dementia detection rates using four regions speech

Dementia tendency	Precision	Recall	F-measure
Positive	76.2	78.0	77.1
Negative	76.9	75.0	75.9
Weight average	76.6	76.5	76.5

of hasegawa's dementia scale (hds-r). *Jpn Geriatr Psychiatr Med*, 2:1339–1347.

Kim, K., Lee, D., Jhoo, J., Youn, J., Suh, Y., Jun, Y., Seo, E., and Woo, J. (2005). Diagnostic accuracy of mini-mental status examination and revised hasegawa dementia scale for alzheimer's disease. *Dementia and geriatric cognitive disorders*, 19(5-6):324–330.

Kishie, S. and Yoshihiro, A. (2006). On-glide nasal [d] and [g] in the shikoku dialects. a focus on the tokushima dialect (in japanese). 10(1):49–59.

Kudo, I., Nakama, T., Watanabe, T., and Kameyama, R. (1996). Data collection of japanese dialects and its influence into speech recognition. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, volume 4, pages 2021–2024. IEEE.

Kurlowicz, L. and Wallace, M. (1999). The mini-mental state examination (mmse).

Lee, A. and Kawahara, T. (2009). Recent development of open-source speech recognition engine julius. In *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, pages 131–137. Asia-Pacific Signal and Information Processing Association, 2009 Annual ....

Luz, S., de la Fuente, S., and Albert, P. (2018). A method for analysis of patient speech in dialogue for dementia detection. *arXiv preprint arXiv:1811.09919*.

Ministry of Internal Affairs and Communication, Japan. (2015). New orange plan. `https://www.mhlw.go.jp/file/06-Seisakujouhou-12300000-Roukenkyoku/nop1-2_3.pdf`. Last accessed 27 December 2021.

National Institute for the Humanities, National institute for Japanese Language and Linguistics. (2011). Research project report on the actual situation of languages and dialects in crisis(in japanese). `https://www.bunka.go.jp/seisaku/kokugo_nihongo/kokugo_shisaku/kikigengo/jittaichosa/index.html`. Last accessed 20 April 2022.

Orimaye, S. O., Wong, J. S.-M., and Golden, K. J. (2014). Learning predictive linguistic features for alzheimer's disease and related dementias using verbal utterances. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From linguistic signal to clinical reality*, pages 78–87.

Satt, A., Hoory, R., König, A., Aalten, P., and Robert, P. H. (2014). Speech-based automatic and robust detection of very early dementia. In *Fifteenth Annual Conference of the International Speech Communication Association*.

Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., and Nakamura, S. (2017). Detecting dementia through interactive computer avatars. *IEEE journal of translational engineering in health and medicine*, 5:1–11.

Ujiro, T., Tanaka, H., Adachi, H., Kazui, H., Ikeda, M., Kudo, T., and Nakamura, S. (2018). Detection of dementia from responses to atypical questions asked by embodied conversational agents. In *Interspeech*, pages 1691–1695.

Umezawa, M., Iribe, Y., Kitaoka, and Norihide. (2019). Construction of speech corpus for elderly japanese dementia detection. In *2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*. IEEE.

World Health Organization. (2021). Risk reduction of cognitive decline and dementia. `https://www.who.int/publications/i/item/risk-reduction-of-cognitive-decline-and-dementia`. Last accessed 27 December 2021.