Unsupervised Learning Using Time-Domain and Frequency-Domain Features of Audio Signals for the Classification of Mild Cognitive Impairment

Hong-Han (Hank) Chau and Yawgeng Chau

Department of Electrical Engineering, Yuan Ze University, Taiwan, R.O.C.

Abstract— **In this research, we conducted a study on MCI detection with unsupervised learning. We collected a total of 104 audio samples from Mandarin-speaking test subjects. The study sample contains 72 MCI patients and 32 normal test subjects diagnosed by a clinical psychiatric doctor. Time-domain and frequency-domain features of the mid- and short-term audio signals are extracted and their accuracy performances are analyzed. With unsupervised learning based on frequency-domain features, the accuracy of MCI detection can reach 73%.**

Keywords— **Mild Cognitive Impairment, Unsupervised Learning, Time-Domain, Frequency-Domain, Audio Signal.**

I. Introduction

The World Health Organization (WHO) estimates that 50 million people are currently living with dementia worldwide and this figure will nearly triple by 2050. Mild Cognitive Impairment (MCI) has been used to categorize the transition between the mentally healthy stage and dementia. MCI presents a syndrome of clinical importance for the early detection of dementia or Alzheimer disease (AD) [1]-[3]. As addressed in [3], the annual rate of MCI deteriorating into dementia is from 10% to 15%, and about 50% of MCI patients develop dementia within three to five years [4]. Thus, early detection is extremely important for the treatment of MCI to prevent unnoticed deterioration into dementia. However, there is not yet a simple and automatic method to detect MCI [5].

Traditional methods of MCI or AD detection require procedures that are inconvenient, time-consuming, and costly for some potential patients. By taking advantage of machine learning (ML), MCI detection using audio data has attracted many researchers' interests. Past studies for MCI or AD detection with audio data mostly use supervised learning, such as the schemes in [6]-[8]. In [6], three different supervised classifiers of Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF) were studied for MCI detection using linguistic features of 48 MCI patients and 38 healthy controls (HC). In [7], SVM, Nearest Neighbor (NN), NB, and Decision Trees (DT) were employed with linguistic features for the detection of AD based on 28 AD patients. In [8], four supervised ML models, SVM, K-NN, Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN), were used for MCI detection based on speech (linguistic) fluency and voice quality features of 40 MCI and 60 HC samples. In [9], different supervised ML models, including Xgboost, SVM, DT, RF, and MLP were examined for the classification of MCI based only on patients' speech data. In [10], various audio features of voice quality and speech fluency were identified for distinguishing MCI patients from HC based on audio recordings from 26 MCI patients and 29 healthy controls. In [11], unsupervised learning with linguistic features was applied for AD detection. From the results published in [6]-[11], ML with audio data could really benefit MCI classifications. On the other hand, in the published literature, the study sample sizes are often limited due to obstacles for the collection of sensitive clinical data. Moreover, the MCI detection with features generated from speech or linguistics alone may be inaccurate as addressed in [10]. Furthermore, almost all results in published literature are based on phonemes of alphabet-type languages instead of Eastern languages such as Mandarin.

In the study presented in this paper, under the cooperation with the psychiatry doctor from Far-Eastern Memorial Hospital in New Taipei City, we collected 104 audio data from Mandarin-speaking people. In the collected data, there are 72 MCI patients and 32 normal testees according to the clinical diagnosis. The information of potential patients or testee is given in Table 1 below. The approval code for voice recording of potential patients is FEMH-105147-E. Each audio data is a 3~8 minutes recording, in which a potential patient describes the same set of three pictures shown to them. We extracted various time-domain and frequency-domain features from the audio recordings, and applied these features for MCI classification with unsupervised learning.

Table 1 The information of potential patients

Age	Number of Testees
55-60	5
61-65	2
66-70	14
71-75	23
76-80	29
81-85	21
86-90	10

(a) Age distribution

Ground-Truth	Gender	Number of Testees
Normal	Male	19
	Female	13
MCI	Male	31
	Female	41

(b) Gender distribution

The remaining parts of the paper are organized in the following way. In Section II, we address various time-domain and frequency-domain features of mid-term audio signals. In Section III, we present the corresponding feature extraction and analysis. In Section IV, the unsupervised learning model is proposed and relevant classification results are illustrated. Section V summarizes the conclusions.

II. Feature Extraction

We first divide the collected audio signal into mid-term segments, where short-term processing is performed on each mid-term segment to extract the time-domain and frequency-domain features [12].

In Fig. 1, we illustrate the architecture of feature extractions used in the paper, where each mid-term audio signal is divided into short-term frames. In our experiments, we considered the cases of 2-second (2s), 4s, and 8s frames, and each frame is sub-divided into 23-millisecond (23ms) sub-frames for short-term feature extraction.

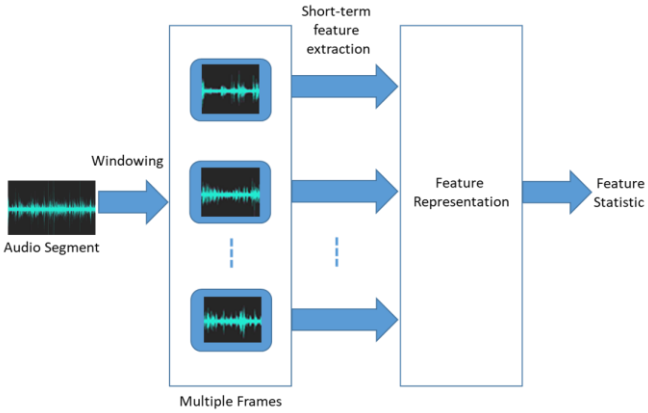

Fig. 1 The architecture of feature extraction

In the context of this research, we analyzed the following time-domain and frequency-domain features.

A. Time-Domain Features

Zero-Crossing Rate (ZCR): This is the rate at which audio signal changes sign (+/–). It may identify noise, silence, pause, etc. The *ZRC* is given by

$$\text{ZCR} = \frac{1}{W}\sum_{i=1}^{W} 1_{\{s_i s_{i-1}<0\}}(s_i s_{i-1}),\ 1_{\{s_i s_{i-1}<0\}} = \begin{cases} 1 & \text{if } s_i s_{i-1} < 0 \\ 0 & \text{if } s_i s_{i-1} \geq 0 \end{cases} \quad (1)$$

where W is the short-term window size (i.e., the number of samples), s_i denotes the sampled audio signal at the i^{th} instance, and $1_{\{s_i s_{i-1}<0\}}$ is an indicator function.

Root-Mean-Squared (RMS) Energy: This is the audio "signal strength" at a specific point in time to identify loudness. The *RME* is given by

$$\text{RMS Energy} = \sqrt{\frac{1}{W}\sum_{i=1}^{W}|s_i|^2} \quad (2)$$

As examples, we plot *ZCR* in Fig. 2 and *RME* in Fig. 3 for the first 2s frame.

B. Frequency-Domain Features

Spectral Centroid (SC): This is the center of the spectrum 'gravity'. Let S_k be the k^{th} coefficient of a Fast Fourier Transform (FFT). The *SC* is given by

$$\text{SC} = \frac{\sum_{k=1}^{W_F} kS_k}{\sum_{k=1}^{W_F} S_k} \quad (3)$$

where W_F is the short-term window size in frequency-domain, i.e., the number of FFT coefficients for an audio short-term signal.

Spectral Spread (SS): This is the normalized variance of spectral distribution. It is given by

$$\text{SS} = \frac{\sum_{k=1}^{W_F} (k - SC)^2 S_k}{\sum_{k=1}^{W_F} S_k} \quad (4)$$

As some examples, in Fig. 4 and Fig. 5, we show the SC and SS for the first 2s frame, respectively.

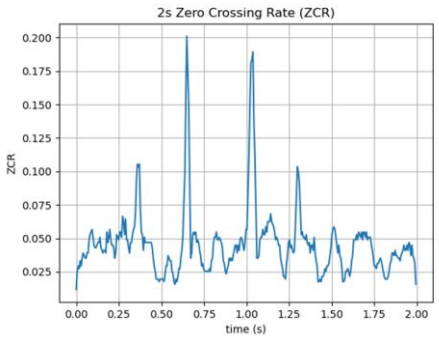

Fig. 2 Example of ZCR for the first 2s frame

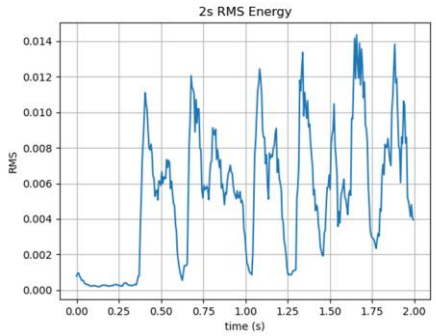

Fig. 3 Example of RMS energy for the first 2s frame

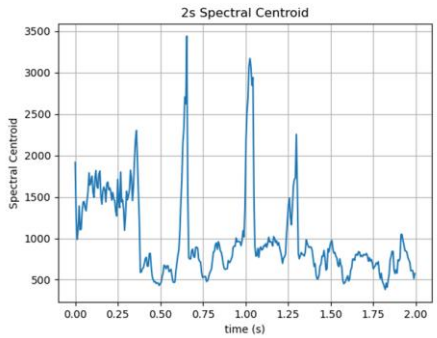

Fig. 4 Example of SC for the first 2s frame

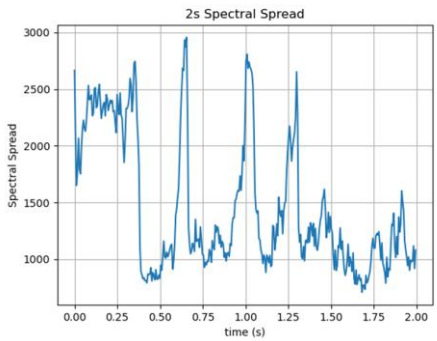

Fig. 5 Example of SS for the first 2s frame

C. Feature Extraction

We first divide each 2s mid-term frame into 23ms short-term frames, and compute the average of every aforementioned feature of the short-term frames. The average is considered as one single sample. Then, the mean and the standard deviation (std) of all samples are evaluated to obtain the final feature statistics for each testee as illustrated in Fig. 1.

III. Results of Unsupervised Learning for MCI Detection

Based on the above time-domain and frequency-domain statistics, we implement the unsupervised clustering with K-means method [13] for relevant experiments. The ground-truth of a MCI or normal subject is obtained through professional diagnosis using magnetic resonance imaging for testee brains. With the ground-truth, we would like to understand if we may separate the two cases with K-means using each feature. In the experiments, the binary classification accuracy is used as a performance index. In the following plots, the feature statistics have been normalized along x-y coordinators.

In Fig. 6, we illustrate the clustering result based on the ZCR. Fig. 7 gives the results based on the RMS energy. The clustering results based on frequency-domain features are given as SC in Fig. 8 and SS in Fig. 9 respectively. Across Fig. 6 to Fig. 9, the SC feature yields the best accuracy of 73%, while the ZRC feature gives a similar accuracy. Moreover, from Fig. 6 with ZCR and Fig. 8 with SC, we notice that there are clearly two separated clusters, one for MCI cases and another for normal cases.

From the experiment results, based on either the ZCR or SC feature, or both, the derived mean and std may be used to classify the MCI cases with a decision boundary.

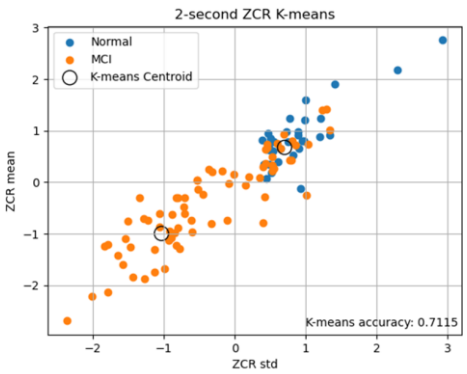

Fig. 6 Clustering result based on ZRC

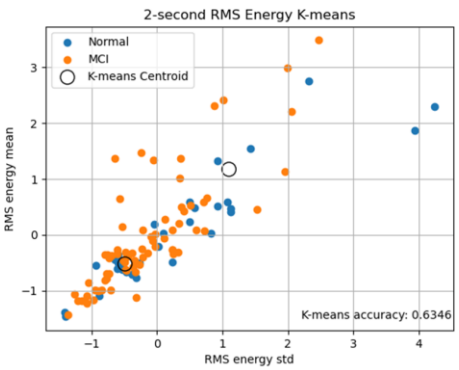

Fig. 7 Clustering result based on RMS energy

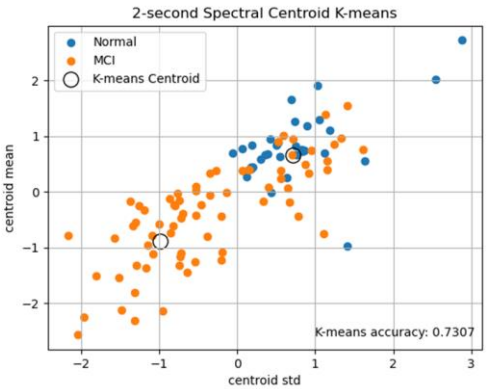

Fig. 8 Clustering result based on SC

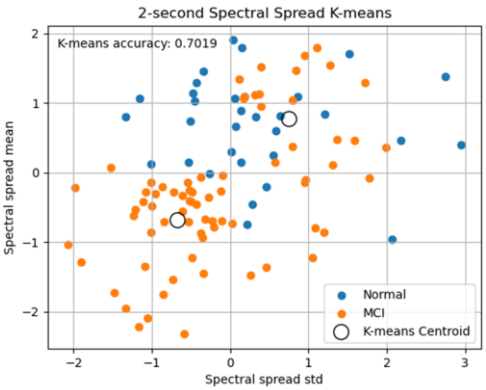

Fig. 9 Clustering result based on SS

IV. Conclusions

In this paper, we studied unsupervised learning with the K-means method for automatic MCI detection based on different time-domain and frequency-domain features. Although some other features, such as the spectral entropy, have been examined, the accuracy index (not shown in the paper) is not as good as that of the features studied in the context. The relevant mean and std of each individual feature were evaluated. Using these different features, we compared the corresponding accuracy metrics. For new potential patients, we may apply their audio data for preliminary automatic MCI classification. More various time-domain and frequency-domain features can be considered in future works. Furthermore, the combinations of different features may also be explored. The benefits that such a machine learning model could bring include a fast preliminary diagnosis from a remote site, which may save time and avoid physical contact.

Acknowledgment

This work is supported by Qualcomm Technologies, Inc. through Research Collaboration Agreement Number YUA-487823. The authors appreciate the audio data supplied by Doctor Yi-Fang Chuang with Far-Eastern Memorial Hospital and Doctor Yi-Chien Liu with Cardinal Tien Hospital.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. A. P. Porsteinsson, e. al., “Diagnosis of early Alzheimer’s disease: Clinical practice,” J. Prev Alzheimers Dis, vol. 8, pp. 535-562, June 2021.
2. C. R. Jr Jack, et al., “NIA-AA research framework: Toward a biological definition of Alzheimer's disease,” Alzheimers Dement. vol. 14, no. 4, pp. 535-562, Apr. 2018.
3. S. T. Farias, et al., “Progression of Mild Cognitive Impairment to Dementia in Clinic-vs Community-based Cohorts,” Archives of Neurology, vol. 66, no. 9, pp. 1151-1157, 2009.
4. K. Glynn, et al., “Clinical Utility of Mild Cognitive Impairment Subtypes and Number of Impaired Cognitive Domains at Predicting Progression to Dementia: A 20-year Retrospective Study,” Int. J. Geriatric Psych., vol. 36, pp. 31-37, 2020.
5. F. Jessen, et al., “Prediction of dementia by subjective memory impairment: Effects of severity and temporal association with cognitive impairment,” Archives of General Psychiatry, vol. 67, pp. 414-422, 2010.
6. A. Khodabakhsh, et al., “Evaluation of linguistic and prosodic features for detection of Alzheimer’s disease in Turkish conversational speech,” Eurasip J. Audio Speech Music Processing, vol. 9, Mar. 2015.
7. L. Toth, et al., “A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech,” Curr. Alzheimer Res., vol. 15, pp. 130-138, 2018.
8. K. López-de-Ipiña, et al., “On the analysis of speech and disfluencies for automatic detection of mild cognitive impairment,” Neural Comput. Appl. vol. 32, pp. 15761-15769, 2020.
9. C. Themistocleous, M. Eckerström, and D. Kokkinakis, “Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls,” PLoS One, vol. 5, 2020.
10. Y. Kumar, et al., “ML-based analysis to identify speech features relevant in predicting Alzheimer’s disease,” arxiv: 2110.13023, Oct. 2021
11. A. Khodabakhsh, et al., “Detection of Alzheimer’s disease using prosodic cues in conversational speech,” Proceedings of the 22nd Signal Processing and Commun. App. Conf., pp. 1003-1006, April 2014.
12. T. Giannakopoulos and A. Pikrakis, *Introduction to Audio Analysis: A MATLAB Approach*, Academic Press, 2014.
13. https://en.wikipedia.org/wiki/K-means_clustering