 doi:10.20944/preprints201901.0299.v1

Type of the Paper (Article)

A Framework for Improving the Interpersonal Relationship of the Elderly with Mild Cognitive Impairment by Using Speaker Recognition and Social Network Platforms

Tan-Hsu Tan [1], Fu-Rong Jean [1], Chia-Jung Lin [1], Tsung-Yu Liu [2] and Yung-Fa Huang [3,*]

[1] Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan, R.O.C.; thtan@ntut.edu.tw

[2] Department of Multimedia and Game Science, Lunghwa University of Science and Technology, New Taipei, Taiwan, R.O.C.; tyliu@gmail.com

[3] Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung, Taiwan, R.O.C.; yfahuang@cyut.edu.tw

* Correspondence: yfahuang@cyut.edu.tw; Tel.: +886-4-2332-3000

Abstract: This study aims to develop an elderly care system for improving the interpersonal relationship of the elderly with mild cognitive impairment (MCI) by employing the speaker recognition technique and association functionality of social network platforms. Firstly, the speaker recognition units based on the Gaussian Mixture Model (GMM) and Gaussian Mixture Model-Universal Background Model (GMM-UBM) are implemented to identify the visitor via individual input utterance. After the visitor is identified, the proposed system will be linked to the private database and social network platforms to extract the associated message of two parties. Experimental results indicate that the speaker recognition unit based on GMM-UBM achieves the best performance. Finally, five elderly persons are invited to measure the usability of the proposed system. A questionnaire is used to survey the five elderly persons, and the result indicates that the proposed system is highly potentially applicable in improving the interpersonal relationship of the elderly with MCI.

Keywords: **m**ild cognitive impairment (MCI); speaker recognition; Gaussian Mixture Model (GMM); Universal Background Model (UBM)

1. Introduction

Mild Cognitive Impairment (MCI) refers to a person whose memory ability is more severely degraded than a normal person of the same educational level and age, but who has not yet shown other manifest dementia symptoms [1]. Previous works have indicated that MCI may be an early symptom of dementia [2]. The elderly person with dementia will face many difficulties in daily life without the help of a caregiver. As the number of the elderly with dementia increases, the resulting care expense becomes a heavy financial burden for a family and government. Dementia development can be delayed or even reversed (in case of reversible-dementia) if the elderly can be properly treated in the early stage of dementia, such as MCI. Thus, how to slow down the memory lapses is critical for the elderly with MCI, since this can postpone the development of dementia.

This study aims to improve the memory of elderly persons with MCI by applying the speaker recognition technology together with the association functionality of the social network platform, thereby enhancing their interpersonal relationships and slowing down the development of dementia. Firstly, this study will develop a speaker recognition unit to identify the identities of the visitors of the elderly through their voices. After identifying the identity of the visitor, the system will be linked to self-built database as well as the social network platform to retrieve historical data associated with both parties. In this way, the associated message will allow the elderly with MCI immediately to go back to the good old days and share beautiful things of both parties and thus achieves the effect of the reminiscence therapy. Moreover, the proposed system can stimulate the memory of the elderly with MCI, hence alleviating the tendency of cognitive decline.

The speech recognition technology has been used in elderly care and health care [3]. Recently, speaker recognition technology [4] has also been used to develop the elderly care system. A framework of the ubiquitous healthcare system based on cloud computing and speaker recognition for monitoring an elderly life has been presented in the work of Ou et al. [4]. However, the application of speaker recognition to elderly care is still very limited. With our best knowledge, the proposed system would be the first work exploring speaker recognition and a social network platform in order to improve the interpersonal relationship of the elderly with MCI.

2. System framework

Figure 1 shows a framework of the proposed system. The execution procedure of the system is described as follows:

1. The elderly with MCI holds a tablet PC to receive the utterance of the visitor, and then the tablet PC captures the feature parameters of the utterance.
2. The tablet PC uploads the feature parameters to the server.
3. The server performs the task of speaker recognition.
4. According to the recognition result, the identity and background information of the visitor will be shown on the PC screen, and then the message associated with two parties obtained from the self-built database as well as the social network platform will be sent to the tablet PC and displayed on the PC screen.

3. Proposed Methods

3.1. Speaker Recognition

Human physiological characteristics are intrinsically beneficial to recognize personal identity credentials. As human voices are easily generated, captured, and transmitted, they can be easily applied to intelligent home automation, voice-controlled user interfaces, and personal identification. As a result, speaker recognition has been becoming an intuitive and valuable technology for identifying a person by his/her voice.

The category of speaker recognition can be classified as either speaker verification or speaker identification [5,6]. When an unknown speaker claims an identity, the purpose of the speaker verification task is to confirm whether this claim is true or not. On the contrary, the task in speaker identification is to recognize an unknown speaker from a group of recognized persons (denoted as closed or in-set speaker identification). The main purpose of this study is to assist the elderly persons with MCI quickly identifying their visitors and all the speech features of the visitors have been already included in the database server in advance. This study is therefore classified as an application of speaker identification.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

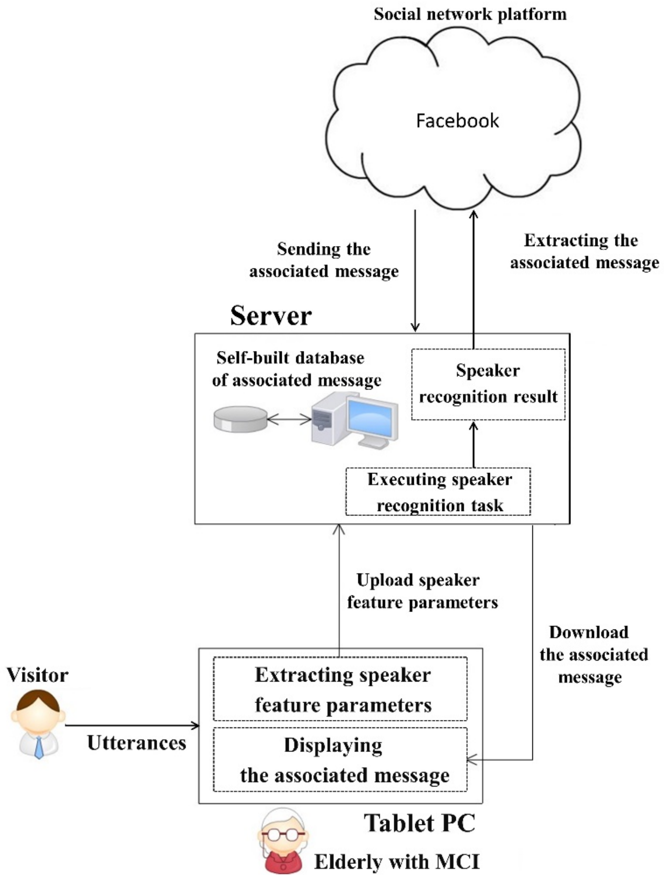

Figure 1. Framework of the proposed systems

3.2. Speech Features (Mel-Frequency Cepstrum Coefficients, MFCCs)

One of the most popular short-term acoustic features are the Mel-frequency cepstral coefficients (MFCCs) which are derived based on the auditory model of the human ear. They have been well-known to be very effective in speech recognition. Therefore, the MFCC features are also commonly employed in speaker recognition. Since human have higher resolution in low frequency bands, we can partition the frequency range in the frequency domain with the Mel- frequency scale for matching the auditory characteristics of the human ear. For the Mel-frequency scale it is almost linear spacing below 1000 Hz and logarithmic spacing above 1000 Hz.

Figure 2 shows the procedure of obtaining the MFCC coefficients from a speech sampling stream. The speech sampling stream is first high-pass filtering for pre-emphasis and then multiplied by a window function for calculating the short-term Fourier power spectrum. A linearly spaced Mel-frequency filter-bank analysis is performed on the spectrum to produce the spectrum energy (also known as the filter-bank energy coefficients) in each Mel-frequency channel. Finally, MFCC coefficients are obtained by performing discrete cosine transform (DCT) on the logarithmic-scale energy spectrum and preserving a number of leading coefficients. In the experiments, twelve MFCCs and one log energy were used as the static feature vector.

 doi:10.20944/preprints201901.0299.v1

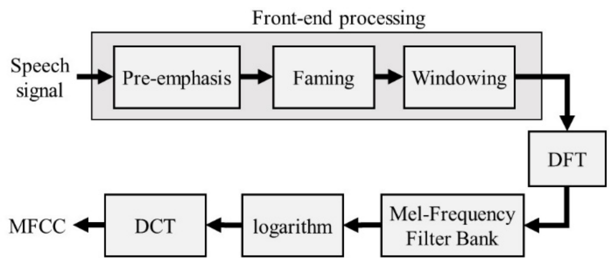

Figure 2. The procedure of extracting the MFCCs from a speech signal

3.3. Gaussian Mixture Model (GMM)

The Gaussian mixture model (GMM) has been widely used in speech and speaker recognition [6,7]. There are three essential parameters in each GMM so that a GMM can be mathematically expressed as

$$\lambda = \{w_i, \overline{\boldsymbol{\mu}}_i, \Sigma_i \mid i = 1, \dots, M\} \tag{1}$$

where M is the number of mixture components, w_i represents the weight, $\overline{\boldsymbol{\mu}}_i$ the mean vector, and Σ_i the covariance matrix of the i-th mixture component. Based on this GMM parameters we can calculate the probability density for any speech features that are produced as a linear combination of the M mixture components.

We employ the K-means algorithm to initialize the model's parameters and then apply the Expectation-Maximization algorithm (EM) to progressively optimize the parameters of the GMM. For a sequence of T training feature vectors $\boldsymbol{X} = [\boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \dots \;\; \boldsymbol{x}_t \dots \;\; \boldsymbol{x}_T]$, $1 \le t \le T$, the GMM likelihood can be computed as the text following an equation need not be a new paragraph. Please punctuate equations as regular text.

$$p(\boldsymbol{X}|\lambda) = \prod_{t=1}^{T} p(\boldsymbol{x}_t|\lambda). \tag{2}$$

For a common speaker identification task, in the beginning a GMM is obtained for each speaker. During testing, the speech feature sequence for an unknown speaker is compared against every GMM, and the most likely speaker (with the highest-scoring GMM) is selected as the identified speaker as

$$\hat{k} = \arg\max_{1 \le k \le S} p(\lambda_k|\boldsymbol{X}) \tag{3}$$

where S is the total number of known speakers, and $\hat{k}$ represents the identified speaker. Applying the Bayesian theorem to Eq. (3), we obtain

$$\hat{k} = \arg\max_{1 \le k \le S} \frac{p(\boldsymbol{X}|\lambda_k)p(\lambda_k)}{p(\boldsymbol{X})} \tag{4}$$

Assume that the probability of occurrence for each speaker is the same, all the feature vectors are mutually independent, and the likelihood of the GMM λ_k given the observed feature sequence $\boldsymbol{X}$ is calculated in logarithmic scale. Eq. (4) can then be replaced by

$$\hat{k} = \arg\max_{1 \le k \le S} \sum_{t=1}^{T} \log p(\boldsymbol{x}_t|\lambda_k) \tag{5}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

According to the recognition result, the identity and background information of the visitor will be shown on the PC screen, and then the message associated with two parties obtained from the self-built database as well as the social network platform will be sent to the tablet PC and displayed on the PC screen.

3.4. Universal Background Model (UBM) and GMM-UBM Model.

The universal background model (UBM) [6,7] is basically a large GMM trained to represent features distribution for all speakers. Therefore, we can use the UBM to solve the problem of insufficient training corpus. When applying the UBM to speaker recognition, at first we don't exploit the EM algorithm to train each speaker's model. Instead, we use the EM algorithm and most of speaker's voices to train a UBM, and then adapt each speaker model by using his/her individual corpus based on the maximum a-posterior (MAP) adaptation criterion.

This GMM-UBM approach utilizes the UBM as an alternate speaker model and as an initial model for each speaker since it is capable of tracing most of the speech utterance variation but cannot be covered by using only a few corpus to obtain. As shown in Figure 3, in this approach a speaker's GMM is derived and adapted from the UBM by updating the well-trained UBM parameters. As this GMM-UBM model can more truly reflect the distribution of each speaker's features, performance degradation due to the drawback of insufficient training corpus can then be compensated.

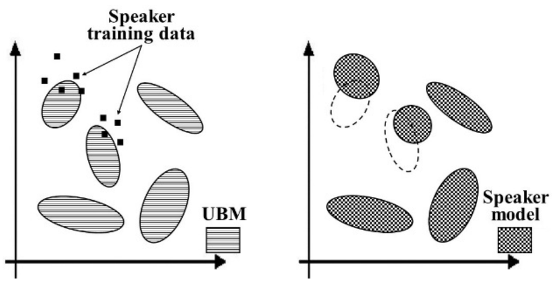

Figure 3. A speaker's GMM derived and adapted from the UBM

This is an example of an equation: Given a UBM, $\lambda_{UBM} = \{w_i, \bar{\mu}_i, \Sigma_i \mid i = 1, \ldots, M\}$ λUBM={ wi, μi, ∑i|i=1,...,M}, and a speaker's feature sequence, $\boldsymbol{X} = [\boldsymbol{x}_1 \ \ \boldsymbol{x}_2 \ldots \ \ \boldsymbol{x}_t \ldots \ \ \boldsymbol{x}_T]$, $1 \leq t \leq T$, the steps of adapting this speaker's GMM-UBM model are summarized as follows.

A. The probability of the observed feature vector with respect to the UBM mixture components can be calculated as

$$p(i|\boldsymbol{x}_t) = \frac{w_i p_i(\boldsymbol{x}_t)}{\sum_{j=1}^{M} w_j p_j(\boldsymbol{x}_t)} \tag{6}$$

where $i = 1, \cdots, M; t = 1, \cdots, T$.

B. Next, we use all the pre-determined values of $p(i|\boldsymbol{x}_t)$ to calculate the weight, the mean, and the covariance parameters as the followings:

$$n_i = \sum_{t=1}^{T} p(i|\boldsymbol{x}_t), \tag{7}$$

$$E_i(\boldsymbol{X}) = \frac{1}{n_i} \sum_{t=1}^{T} p(i|\boldsymbol{x}_t)\, \boldsymbol{x}_t, \tag{8}$$

and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

$$Var_i(\boldsymbol{X}) = \frac{1}{n_i}\sum_{t=1}^{T} p(i|\boldsymbol{x}_t)\boldsymbol{x}_t\boldsymbol{x}_t^{T} \tag{9}$$

C. Finally, adjust the model weight, model mean, and model covariance by using the MAP adaptation update equations as follows.

$$\widehat{w}_i = [\frac{\alpha_i^w n_i}{T} + (1-\alpha_i^w)w_i]\gamma, \tag{10}$$

$$\widehat{\boldsymbol{\mu}}_i = \alpha_i^m E_i(\boldsymbol{X}) + (1-\alpha_i^m)\boldsymbol{\mu}_i, \tag{11}$$

and

$$\widehat{\Sigma}_i = \alpha_i^v Var_i(\boldsymbol{X}) + (1-\alpha_i^v)(\Sigma_i + \widehat{\boldsymbol{\mu}}_i\widehat{\boldsymbol{\mu}}_i^{T}) - \widehat{\boldsymbol{\mu}}_i\widehat{\boldsymbol{\mu}}_i^{T} \tag{12}$$

where $\alpha_i^\rho = \frac{n_i}{n_i + r^\rho}$, $\rho \in \{w, m, v\}$ are adaptation coefficients, γ is a scaling factor that ensures the sum of the mixture weights equal to unity, and r^ρ is a fixed factor related to the parameter ρ. In the experiments, we set $r^\rho = 16$.

In the GMM-UBM approach, we usually calculate the score (a matching measure) for a target speaker in the logarithmic scale as follows.

$$S(\boldsymbol{X}) = log\, p(\boldsymbol{X}|\lambda_{GMM}) - log\, p(\boldsymbol{X}|\lambda_{UBM}), \tag{13}$$

where $S(\boldsymbol{X})$ is the logarithmic score, $\boldsymbol{X}$ is the test sentence spoken by an unknown speaker, λ_{UBM} is the UBM, and λ_{GMM} is the target speaker's GMM adapted from the UBM.

The architecture of the speaker identification system based on the GMM-UBM approach is shown in Figure 4. At the training stage, we use a large amount of background speech corpus to train the global UBM. Next, each speaker's GMM model is derived from the UBM and adapted by using his or her individual corpus based on the MAP adaptation update procedure. At the test stage, the logarithmic score must be calculated for every known person, and the person with the highest score should be chosen as the identified speaker.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

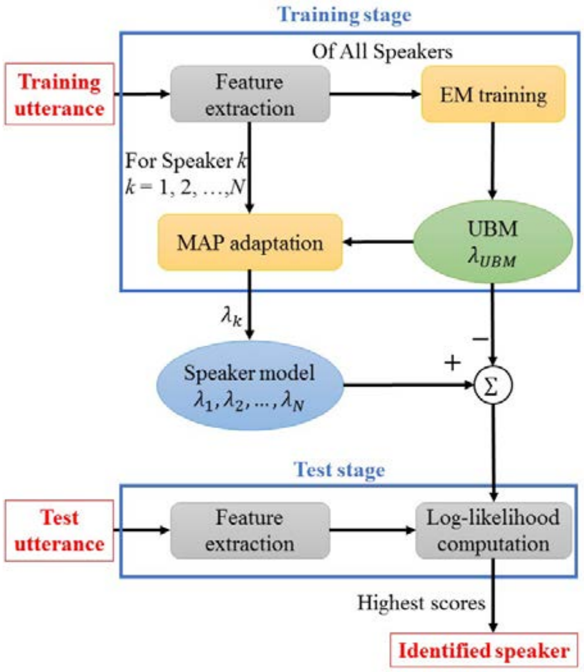

Figure 4. The speaker identification system based on the GMM-UBM approach

3.5. Facebook Graph API.

The OAUTH [8], a third-party authentication platform, provides an open protocol to allow secure authorization with a simple and standard method for web, mobile and desktop applications. The authorization flow and each step of the process of obtaining an access token is shown in Figure 5. OAUTH allows the Resource Owner to provide a token rather than his/her account name and password. By using this secure manner, it enables the third-party users to access some certain information that the Resource Owner has stored in a particular server.

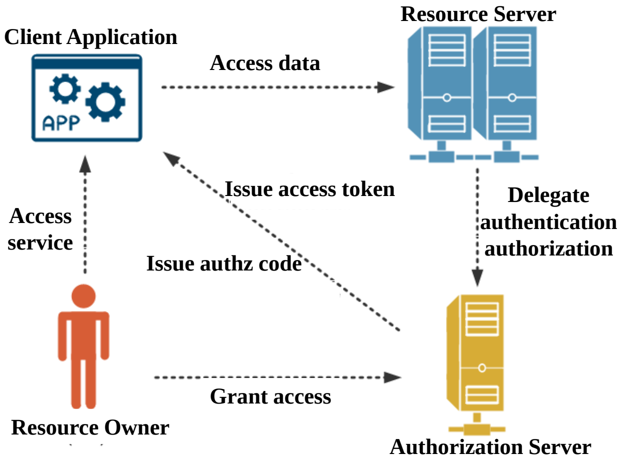

Figure 5. The authorization flow for the OAUTH 2.0 protocol

Facebook is a popular online social networking service platform that let users share and send text or multimedia messages to one another. The Facebook Graph API is a programming framework based on the low-level HTTP-based API for conveniently accessing and writing Facebook information. The Facebook Graph API also makes use of the OAUTH 2.0 protocol to offer secure paths that developers can access to users' information. In this study we use the Facebook Graph API to retrieve related photos of identified visitors and elderly with MCI.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

4. Experimental Results

A series of experiments was conducted using a speech corpus, and speaker models based on GMM and GMM-UBM. The details are described as follows.

4.1. Speaker Corpus.

An open speech corpus provided by Chunghwa Telecom was employed [9]. This corpus consists of isolated Mandarin numeral with a sampling frequency of 10 kHz. The corpus is composed of 50 male and 50 female speakers; each one recorded a single Mandarin numeral from 0 to 9, each numeral was repeated 6 times. According to our survey, the number of visitors is generally less than 20. Therefore, 20 speakers are randomly selected from this corpus, including 10 male and 10 female speakers, respectively, to form a 20-person corpus. This new version is denoted as corpus-1 for simplicity. The speech feature is represented by a 39-dimensional vector, which is composed of MFCC, Delta MFCC, and Delta-Delta MFCC. The frame length and the frame shift are 25 *ms* and 10 *ms*, respectively.

4.2. Training and Test Corpus

The corpus-1 is further divided into training and test corpus. The training corpus is used to train the Universal Background Model (UBM) and individual speaker model. The training procedure using corpus-1 is described as follows. The first four utterances of the speaker are used as training data, and the remaining two utterances serve as test data. Among them, the UBM is obtained by training all 20 speakers' utterances, and then two utterances were randomly selected from the training data of each speaker to obtain the target speaker's GMM-UBM model by use of the maximum a-posterior (MAP) algorithm [10].

4.3. Results of Speaker Recognition

Table 1 shows the resultant accuracy of speaker recognition by means of the Gaussian Mixture Model (GMM) and the GMM-UBM Model with different number of Gaussian mixture components. Table 1 indicates that when the Gaussian mixture is 64, the recognition rate of the speaker recognition system based on the GMM model is the highest, which is 85.5%. When the Gaussian mixture is 128, the recognition rate is 85.25%. The recognition rate of the speaker recognition system based on the GMM-UBM model is 95% when the Gaussian mixture is 64, and the recognition rate is 96.5% when the Gaussian mixture is 128, which is the best performance. Therefore, the GMM-UBM model far outperforms the GMM model. As a result, the speaker recognition system developed in this study will use a GMM-UBM model with a Gaussian mixture of 128 for the speaker model.

Table 1. Speaker recognition rate for corpus-1

Gaussian mixture	GMM (%)	GMM-UBM (%)
16	80.5	93.25
32	84.25	94
64	85.5	95
128	85.25	96.5

4.4. Results of Care System Experiments with Speaker Recognition

The system developed in this research combines speaker recognition technology with multimedia information sharing on the social networking site (Facebook). Five elderly persons with MCI, consisting of 2 male persons and 3 female persons, have been invited to participate the experiments in their residences. Their ages are between seventy-six and eighty-two years old. Firstly, the voices of visitors (eight persons for each case) are recorded to train the speaker models, which are then stored on the server. When the visitor speaks to the elderly, his/her speech feature parameters will be extracted by the tablet PC and then are transmitted to the server for speaker

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

recognition. As the identity of the visitor is identified, the server will request the third-party authentication platform (Facebook Graph API in this study) to retrieve the associated message of the two parties, such as photos. Then the associated message is sent to the tablet PC to display on the screen. The speaker recognition rates of all cases are higher than 96% when the GMM-UBM model is employed with a mixture of 128.

Table 2 shows the information about the five elderly people who joined the experiment. They are represented by the codes U1~U5. Three women and two men, aged 65 or older, have met the definition of the elderly family. In addition, in the interview before the experiment, the elderly people said that they usually have poor memory or take a long time to think about things. This phenomenon may be one of the symptoms of MCI, so it is the requirement for the experiment in this study.

Table 2. The information about the five elderly people

User ID	U1	U2	U3	U4	U5
Gender	Female	Male	Female	Female	Male
Ages	82	80	77	76	79

Before the system starts to perform the identification work, we recorded the voices of all visitors/friends and then train the voice to get the speaker model and stored it in the server. During the experiment, the users only need to connect to the network and upload the characteristic parameters of the voice of the visitor/friends to be identified to the server to obtain the identification result. At the same time, after obtaining the user authorization, the server requests the photo from the third-party authentication platform of Facebook Graph API. Then, the photos of the two parties were displayed sequentially in the form of a slide show.

The purpose of the care system developed in this study is to improve the memory of the MCI elderly people. Therefore, we invite five elderly people to use the platform in their own homes. When relatives/visitors visit, the elderly people use the handheld mobile device to perform the previous procedure. After the test is completed, a questionnaire survey is conducted to understand the feedback and satisfaction of the users to the system platform.

Table 3. The information about the five elderly people

SUS Scores	SUS Grade	Percentage (%)	SUS Scores	SUS Grade	Percentage (%)
84.1 ~ 100	A+	96 ~ 100	71.1~72.5	C+	60 ~ 64
80.8 ~ 84	A	90 ~ 95	65~71	C	41 ~ 59
78.9 ~ 80.7	A−	85 ~ 89	62.7~64.9	C−	35 ~ 40
77.2 ~ 78.8	B+	80 ~ 84	51.7~62.6	D	15 ~ 34
74.1 ~ 77.1	B	70 ~ 79	0~51.7	F	0 ~ 14
72.6 ~ 74	B−	65 ~ 69			

The questionnaire for this experiment is divided into two parts: system operation and satisfaction of use. The first part refers to the System Usability Scale (SUS) [10] created by John Brooke in 1986. The scale consists of 10 topics, including the positive statements of odd items and the negative statements of even items. Each question was evaluated using a 5-level likert scale. The degree of consent is marked with 1 to 5 points: 1 point represents very disagree, 2 points means disagreement, 3 points means no opinion, 4 points means consent, and 5 points means very agree. Table 3 is a scoring standard for system usability evaluation [11]. The second part is based on the user's assessment of the system's benefits, based on the personal experience to assess whether the system used provides effective assistance.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

Table 4 shows the results of the first part of the questionnaire filled out by five elderly people. There are 10 questions in total, of which "Q1-A" is the first question of the first part of the questionnaire. For example: "Q1-1" is the first part of the questionnaire. For part of the first question, the degree of consent is indicated by 1 to 5 points, 5 points means very agree, and 1 point means very disagree.

Table 5 is the percentage of SUS obtained according to the scores given by the system for each user according to Table 4. From Table 5, it is observed that the SUS scores of the five silver hairs are all above 77 and the average value is 81.5. According to Table 3, the SUS grade of 81.5 is A, and the percentage grade is about 91. This result shows that the system is more usable than the other 90% of the system, that is, the score of the system is the top 10%, which means that the user has a high degree of willingness to recommend the system to friends. In summary, the five elderly people gave a good evaluation of the use of this care system.

Table 4. Survey results in the first part of the questionnaire

Items / User ID	Q1-1	Q1-2	Q1-3	Q1-4	Q1-5	Q1-6	Q1-7	Q1-8	Q1-9	Q1-10
U1	4	1	5	3	4	2	5	1	4	2
U2	4	2	5	2	5	2	5	1	4	2
U3	5	2	4	4	5	1	5	1	4	4
U4	5	2	4	3	5	2	5	2	5	3
U5	5	1	5	3	4	1	4	2	4	2

Table 5. Evaluation results in the first part of the questionnaire

User ID / Evaluation Item	U1	U2	U3	U4	U5	Average
SUS Score	82.5	85	77.5	80	82.5	81.5
SUS Grade	A	A+	B+	A−	A	A
Percentage	92	97	81	88	92	91

Table 6 shows the results of the second part of the questionnaire filled by the five elderly people. There are 8 questions in total, of which "Q2-A" is the second question of the second part of the questionnaire. From Table 6, it is observed that among the 40 points, the users gave a score of 35 or more, with an average of 36.4 points, which means that the elderly people think that using the system is quite helpful to improve the quality of care.

Table 6. Survey results in the second part of the questionnaire

Question / No.	Q2-1	Q2-2	Q2-3	Q2-4	Q2-5	Q2-6	Q2-7	Q2-8	Scores
U1	4	4	5	5	5	4	5	4	36
U2	5	4	4	4	5	5	4	4	35
U3	5	5	4	5	5	4	5	5	38
U4	5	5	4	5	5	4	5	4	37
U5	4	5	5	5	5	4	5	3	36
Average	4.6	4.6	4.4	4.8	5	4.2	4.8	4	36.4

Figure 6 illustrates a scenario of the experiment held in a residence. In the experiments, five elderly persons are invited to measure the usability of the proposed system [10,11]. From the score results of the questionnaire used to survey the five elderly persons, it exhibits a very positive feedback of system usability, which reaches top 10%. Therefore, the proposed system is highly potentially applicable in improving the interpersonal relationship of the elderly with MCI.

Figure 6. A scenario of the experiment held in a residence

5. Conclusions

This study has presented a care system to improve the interpersonal relationship of the elderly with MCI by employing the speaker recognition technique and association functionality of social network platforms. In the care system, two speaker recognition units based on the GMM and GMM-UBM is implemented to identity the visitor via individual input utterance. After the visitor is identified, the care system can be linked to the private database and social network platforms to extract the associated information of two parties to show on the tablet PC held by the elderly with MCI. Moreover, via the effect of the reminiscence therapy using association data, they can easily go back to the good old days and share beautiful things. Besides warming up their interaction in a very short time, this method stimulates the memory of the elderly with MCI, hence alleviating the tendency of cognitive decline. Experimental results indicate that the speaker recognition unit based on GMM-UBM obtains the best performance. Finally, five elderly persons are invited to measure the usability of the proposed system. A questionnaire is used to survey the five elderly persons, and it exhibits a very positive feedback of system usability, which reaches top 10%. Therefore, the proposed system is highly potentially applicable in improving the interpersonal relationship of the elderly with MCI.

Author Contributions: "conceptualization, TH. Tan and FR. Jean; methodology, FR. Jean; software, CJ. Lin.; validation, TH. Tan, FR. Jean. and YF. Huang; investigation, CJ. Lin.; resources, TH. Tan and TY. Liu.; data curation, TY. Liu; writing—original draft preparation, FR. Jean; writing—review and editing, TH. Tan and YF. Huang; project administration, TH. Tan; funding acquisition, TH. Tan and YF. Huang".

Funding: This research was funded by Ministry of Science and Technology of Taiwan under Grant MOST 105-2221-E-027-112, MOST 106-2218-E-027-017, MOST 106-2221-E-324-020 and MOST 107-2221-E-027-106.

Conflicts of Interest: The authors declare no conflict of interest.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 January 2019 doi:10.20944/preprints201901.0299.v1

References

1. Akl, A.; Taati, B.; Mihailidis, A. Autonomous unobtrusive detection of mild cognitive impairment in older adults. *IEEE Trans. Biomed. Eng*. **2015**, 62(5), 1383–1394.
2. Ritchie, K.; Artero, S.; Touchon, J. Classification criteria for mild cognitive impairment: a population-based validation study. *Neurology*, **2001**, 56(1), 37-42.
3. Zhao, Y. Speech-recognition technology in health care and special-needs assistance. *IEEE Signal Processing Magazine* **2009**, 87-90.
4. Ou, Y.-Y.; Shih, P.-Y.; Chin, Y.-H.; Kuan, T.-W.; Wang, J.-F.; Shih, S.-H. Framework of Ubiquitous Healthcare System Based on Cloud Computing for Elderly Living, Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Kaohsiung, Taiwan, 29 Oct.-1Nov. 2013.
5. Chang, S.-W. Performance Evaluation and Improvement of Speaker Recognition under Multiple Environments, Master Thesis, Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan, July 31 2003.
6. Togneri, R. & Pullella, D. (2011). An overview of speaker identification: accuracy and robustness issues," IEEE Circuits and Systems Magazine, 23-61.
7. Reynolds, D. A.; Quatieri, T. F.; Dunn, R. B. Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, **2000,** 10(1-3), 19-41.
8. OAuth 2.0 official website: https://oauth.net/ (accessed on 20 Jan. 2017).
9. Chunghwa Telecom Laboratories website: http://www.chttl.com.tw/en/index/ (accessed on 2 Jan. 2017).
10. Bangor, A.; Kortum, P.; Miller, J. Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies* **2009**, 4(3), 114-123.
11. Sauro, J. Measuring Usability with the System Usability Scale (SUS), 2 Feb. 2011. Available online: https://measuringu.com/sus/ . (accessed on 20 Jan. 2017).