The application of fractional Mel cepstral coefficient in deceptive speech detection

Xinyu Pan, Heming Zhao, Yan Zhou

The inconvenience operation of EEG P300 or functional magnetic resonance imaging (FMRI) will be overcome, when the deceptive information can be effectively detected from speech signal analysis. In this paper, the fractional Mel cepstral coefficient (FrCC) is proposed as the speech character for deception detection. The different fractional order can reveal various personalities of the speakers. The Linear Discriminant Analysis (LDA) model which has the ability of global optimal vector mapping is introduced, and the performance of FrCC and MFCC in deceptive detection is compared when all the data are mapped to low dimensional. Then the hidden Markov model (HMM) is introduced as a long term signal analysis tool. 25 male and 25 female participants are involved in the experiment. The results show that the clustering effect of optimal fractional order FrCC is better than that of MFCC. The average accuracy for male and female speaker is 59.9% and 56.2% respectively by using the FrCC under LDA model. When MFCC is used, the accuracy is reduced by 3.2% and 5.9% respectively for male and female. The accuracy can be increased to 71.0% and 70.2% for male and female speaker when HMM is used. Moreover, some individual accuracy is increased over 20%, or even more than 85%, when FrCC is introduced. The results show that the deceptive information is indeed hidden in the speech signals. So speech based psychophysiology calculating may be a valuable research field.

The Application of Fractional Mel Cepstral Coefficient in Deceptive Speech Detection

Pan Xinyu[1, 2] Zhao Heming[1] and Zhou Yan[1]

1、(*School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu*, 215006, *China*)

2、(*School of Electronics & Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu*, 215009, *China*)

Abstract: The inconvenience operation of EEG P300 or functional magnetic resonance imaging (FMRI) will be overcome, when the deceptive information can be effectively detected from speech signal analysis. In this paper, the fractional Mel cepstral coefficient (FrCC) is proposed as the speech character for deception detection. The different fractional order can reveal various personalities of the speakers. The Linear Discriminant Analysis (LDA) model which has the ability of global optimal vector mapping is introduced, and the performance of FrCC and MFCC in deceptive detection is compared when all the data are mapped to low dimensional. Then the hidden Markov model (HMM) is introduced as a long term signal analysis tool. 25 male and 25 female participants are involved in the experiment. The results show that the clustering effect of optimal fractional order FrCC is better than that of MFCC. The average accuracy for male and female speaker is 59.9% and 56.2% respectively by using the FrCC under LDA model. When MFCC is used, the accuracy is reduced by 3.2% and 5.9% respectively for male and female. The accuracy can be increased to 71.0% and 70.2% for male and female speaker when HMM is used. Moreover, some individual accuracy is increased over 20%, or even more than 85%, when FrCC is introduced. The results show that the deceptive information is indeed hidden in the speech signals. So speech based psychophysiology calculating may be a valuable research field.

Key words: Deceptive speech detection, Fractional Mel Cepstral Coefficient (FrCC), Linear Discriminant Analysis (LDA), Hidden Markov model (HMM), Psychophysiology

Introduction

Deception detection is regarded as an ancient and mysterious topic in the long history of human science, and there have also accumulated many useful results. The mechanism of modern polygraph is based on the changes of EEG signals, due to the contribution of psychophysiology research. Event Related Potential (ERP) of P300 and Functional Magnetic Resonance Imaging for brain (FMRI) are widely used in polygraph technology [1, 2], since they can intuitive reflect the process of psychological and physiological changes (such as the memories) during lying. Speakers' facial expressions and postures can also be an auxiliary way for lie detection in some special circumstance [3]. The application of these techniques achieved good performances. The complex measurement process and the needs of the participant's cooperation make their further promotion limited. Some information can be derived from speech including the speaker's gender, age, emotional and mental states. Therefore speech signals, as the carrier of deception information, may provide the basis for lie detection. Due to the complicated process from psychological activity to physiological reaction, voice based

39 polygraph still exists the following problems. First, there is lack of physiological experimental results
40 demonstrating the theory basis. Second, the study of hearing mechanism also cannot provide clearly
41 conclusions. Third, the deception detection results are not as intuitive as that of speech recognition or speaker
42 identification, because lying is a process. Psychological stress evaluators (PSE) [4] and Voice stress analyzers
43 (VSA) [5] are used to measure the voice tremors, which are considered as the reflection of stress. Layered voice
44 analysis (LVA) [6] claims that the link between the certain types of brain activity and lie are discovered.
45 Though there are some controversial, these methods are effective to some extent. The recognition accuracy in
46 deceptive detection is significant low. Bond [7] claimed that people only achieve an average of 54% correct lie-
47 truth judgments. Hirschberg [8] reports a classification accuracy of 66.4% versus a chance baseline of 60.2%. A
48 study by Graciarena [9] reports an accuracy of 64.0% versus a chance baseline of 60.4%. Most of the studies
49 are focus on the traditional speech features screening, and few mechanism analysis of the deceptive speech
50 features has been made. Paper [10] reports that short time energy, pitch trace and formant F1, F2, F3 did not
51 show clear correlation with deceptive information. Owing to the lack of robust feature analysis, the authors
52 still give positive prospect for speech deceptive detection.

53   The MFCC parameters are used as the main characteristics in the state-of-the-art speech recognition
54 systems. The standard extraction process of MFCC makes it suitable for standard pronunciation pattern
55 classification. Since lying is a kind of the behaviors of the individual differences, the standard characteristics
56 cannot fully reflect the personality of each speaker. The conventional Fourier analysis cannot fully reveal the
57 deceptive information hidden in the voice message. Most of the researches show the performance of GMM and
58 SVM systems in lie detection, but there are few paper reports the changes of phonetic features (including
59 MFCC) after linear transform, and there is also very few linear classification system is used in such experiment.
60 It is very important to evaluate the performance of some linear algorithms in feature extraction or classification,
61 when lie detection research is in the preliminary stage. The results can provide the mathematical foundation for
62 the further use of complex models.

63   In this paper, Fractional Fourier Transform (FrFT) is introduced in deceptive speech feature extraction,
64 Linear Discriminant Analysis model (LDA) and hidden Markov model (HMM) are proposed for classification.
65 Fractional Fourier transform is regarded as the angle rotation from traditional Fourier analysis. Many current
66 literature reports the Fractional Fourier transform is used for the analysis of linear frequency modulated
67 waveform [11,12,13]. Linear frequency modulated signal can be transformed into an impulse signal under certain
68 angle by FrFT. The application of FrFT in speech signal processing field is gradually increasing, and the
69 accuracy of speech recognition and speaker identification is improved when FrFT is introduced in feature
70 extraction [14, 15]. The fractional order linear canonical transform algorithm also obtains a good result in speech
71 signal reconstruction [16]. There are many successful applications with LDA or HMM in the field of speech
72 signal processing. Behzad used LDA and its modified algorithm to reduce the speech recognition error rate [17],
73 and Ana and Jordi used LDA to achieve the phonetic features analysis of snoring patients [18, 19]. Rabiner and
74 Matthias successfully used HMM in speech recognition [20,21].

75   In this paper, the Mel Cepstral Coefficient in fractional domain (called Fractional Mel Cepstral
76 Coefficient, FrCC) is extracted for voice spectrum analysis. Then LDA and HMM model are used to
77 distinguish between the truth and deception. Different fractional order spectrum analysis in Mel domain can
78 further refine the pronunciation characteristics of all liars. The rotation of MFCC parameters in the Mel
79 domain is introduced, and the details of each individual difference in speech signal may reinforce to a certain

80   extent. The LDA algorithm can help us to find the best rotation angle, and obtain the best division result. The

81   HMM based time series model can reveal the psychological and physiological changes when lying, and

82   increase the recognition accuracy. The FrCC parameter corresponding with the highest accuracy can be called

83   optimal order FrCC. The deceptive detection accuracy of all optimal order FrCC is higher than that of MFCC,

84   so the acoustic characteristics of speech signals can provide some support for lie detection.

85   The arrangement of the full text is as follows. The first section introduces the study of distinguish features

86   of deceptive speech, and provides the application basis of fractional Fourier analysis in this subject. The

87   second section presents the calculation of Fractional Mel cepstral coefficients. The LDA algorithm and HMM

88   model are introduced in the third section. Experiment results and analysis are described in the fourth section.

89   The conclusions are drawn finally.

90

1   Deceptive speech   feature introduction   and Fractional   Fourier Transform application foundation

93   1.1 The introduction of speech signals based polygraph application

94   Lying is a process from the psychological activity to physiological execution. Firstly, people decide to lie

95   from Conscious mind, and then organize the language to cover the real content. Finally control vocal organs to

96   form the voice. Tommaso et al. studied the deception detection result among the people of different personality

97   by the means of pattern recognition. The conclusion is that the outgoing personality groups are easier to be

98   identified [22]. It is also proved that lie detection is individual differences. Lamb et al. reported that the

99   frequency of words with different part of speech appear in the trial process may also be a way to analyze the

100   lying possibilities [23]. In the field of natural language processing, Christie et al. showed that the result is

101   influenced by lexical, syntax, sentence length and motivation [24]. Furthermore, the organization of text and

102   voice can be used in anti-phishing detection system, preventing people from being cheated in instant

103   messengers [25]. These researches discuss the deceptive speech detection results are in psychology and natural

104   language processing fields, including personality difference, pronunciation difference, language expression,

105   and sentence organization and so on. Then Sofia et al. researched the speaker's differences under normal state

106   and tense state [26]. Cheryl et al. summarized that the person under tension will show the following case:

107   adrenaline increasing, higher blood pressure, sympathetic excitement, bronchiectasis, and cricothyroid muscle

108   tension. Some people tended to show increased pitch frequency and voice trembling, but not everyone like this

109   [27]. These physiological studies have proved the existence of the difference between normal and deceptive

110   state, and provides some simple basis in physiological field for speech deception detection.

111   The above conclusion from psychology, linguistics and physiological aspects are relatively consistent, but

112   the researches in acoustic field have different results. Gopalan et al. claim that the trace of pitch and first

113   formant, which are processed by AM-FM model and Teager operator, presented the definite difference

114   between normal and deceptive speech [28]. However, Christin et al. gave the opposite conclusion [10]. They

115   executed several experiments, and the results are presented on a range of speech parameters including

116   fundamental frequency, overall amplitude and mean vowel formants F1, F2, F3. There could not establish a

117   significant correlation between deceptiveness and truthfulness. The two results appeared opposite. Pitch and

118   formants are susceptible to the influence of speaker's pronunciation habit, language content, and coarticulation.

119   The parameters will also be changed due to different extraction algorithm, so it is not a good choice for using
120   them as the speech features for lie detection. There is little research to reveal the process from psychological
121   activity to speech production. Muhammad et al. use bark spectrum as speech features, and use neural network
122   as the classification model to identify the truthfulness and deceptiveness [29]. So using robust acoustic
123   characteristics   for   deceptive   speech   identification   should   be   more   reliable.   And
124   there is still plenty of scope for more progress.

125

126   1.2 The Fractional Fourier Transform application foundations
127   All the current research has not investigated the feature difference between normal speech and lie.
128   Physiological studies also could not provide any explanation, whether there are specific changes of articulators
129   or not when people are lying. The existing information is only obtained from speech features statistical
130   research result or classification conclusion by traditional pattern recognition models. The speech features in
131   frequency domain are mostly achieved by power spectrum transform. If the liar's psychological and
132   physiology changes indeed impact the frequency of speech, the deceptive information can be extracted by
133   short-time frequency analysis for voice signals. But it does not work, if there is only the speech phase changed.
134   Therefore, a speech feature which can express both the change of frequency and phase is needed. The
135   fractional Fourier analysis is applicable to the task.
136   That being the case,we take the cosine signal as an example to compare the difference between the
137   traditional Fourier transform and fractional Fourier transform. (Please refer to the next section for the detail
138   description of FrFT.)

139   $$x(t) = \cos(\varpi_0 t) \quad \Leftrightarrow \quad |X(\varpi)| = \pi \left(\delta(\varpi - \varpi_0) + \delta(\varpi + \varpi_0) \right) \quad (1)$$

140   $$x(t) = \cos(\varpi_0 t + \theta) \quad \Leftrightarrow \quad |X(\varpi)| = \pi \left(\delta(\varpi - \varpi_0) + \delta(\varpi + \varpi_0) \right) \quad (2)$$

141   Through (1) and (2), it is shown that the Fourier amplitude-frequency response can't reflect the phase
142   difference of the two signals. According to trigonometric formula, Eq. (2) can be expanded and transformed by
143   FrFT as the Eq. (3).

144   $$x(t) = \cos(\varpi_0 t + \theta) = \cos(\varpi_0 t)\cos(\theta) - \sin(\varpi_0 t)\sin(\theta) \ldots$$

145   $$\overset{frft}{\Leftrightarrow} \quad X_\alpha(u) = \cos(\theta) X_\alpha[\cos(\varpi_0 t)](u) - \sin(\theta) X_\alpha[\sin(\varpi_0 t)](u) \quad (3)$$

146   Then:

147   $$X_\alpha[\cos(\varpi_0 t)](u) = \sqrt{1 + j\tan(\alpha)}\,\cos(u\varpi_0 \sec(\alpha)) \exp(-\frac{j}{2}(\varpi_0^2 + u^2)\tan(\alpha)) \quad (4)$$

148   $$X_\alpha[\sin(\varpi_0 t)](u) = \sqrt{1 + j\tan(\alpha)}\,\sin(u\varpi_0 \sec(\alpha)) \exp(-\frac{j}{2}(\varpi_0^2 + u^2)\tan(\alpha)) \quad (5)$$

149

$$X_\alpha(u) = \cos(\theta)\sqrt{1+j\tan(\alpha)}\cos(u\varpi_0\sec(\alpha))\exp(-\frac{j}{2}(\varpi_0^2+u^2)\tan(\alpha))...$$

$$-\sin(\theta)\sqrt{1+j\tan(\alpha)}\sin(u\varpi_0\sec(\alpha))\exp(-\frac{j}{2}(\varpi_0^2+u^2)\tan(\alpha)) \tag{6}$$

$$=\sqrt{1+j\tan(\alpha)}\exp(-\frac{j}{2}(\varpi_0^2+u^2)\tan(\alpha))(\cos(\theta)\cos(u\varpi_0\sec(\alpha))+\sin(\theta)\sin(u\varpi_0\sec(\alpha)))$$

$$=\sqrt{1+j\tan(\alpha)}\exp(-\frac{j}{2}(\varpi_0^2+u^2)\tan(\alpha))\cos(u\varpi_0\sec(\alpha)+\theta)$$

151   Eq. (6) expressed the FrFT result of $\cos(\varpi_0 t+\theta)$. The phase θ still exists in $|X_\alpha(u)|$, so Eq. (6) can

152   reserve the phase information.

153   The use of speech signal for lie detection is only at the preliminary stage. If the difference between

154   truthfulness and deceptiveness is really expressed by the amplitude and phase of speech spectrum,

155   the fractional Fourier transform analysis method should be an effective way to reveal the distinction. So as to

156   suit for the diversity of speakers, variety orders of FrFT should be involved. The difference between normal

157   speech and lie can be enhanced due to some orders of FrFT.

158

2. Fractional Mel Cepstral Coefficient (FrCC) extraction

160   The FrCC parameters are modified based on MFCC. First step is short-time analysis, then transform time

161   domain samples to frequency domain by FrFT under a set of rotation angles. The following step is to divide

162   the signals into Mel frequency domain by triangular filters, then conduct by a DCT at last. So the whole

163   process is shown in figure 1.

164   The details of FrCC calculation steps are shown as follows:

165   (A) The fractional Fourier transform for speech signals is shown in (7).

$$S_\alpha(u) = F_p[s(t)] = \int_{-\infty}^{+\infty} s(t)K_\alpha(t,u)dt \tag{7}$$

167   Here, $\alpha = p\frac{\pi}{2}$, p is the set of real numbers, the order of FrFT. $K_\alpha(t,u)$ is the primary function of FrFT,

168   and its specific expressions is presented in Eq. (8).

$$K_\alpha(t,u) = \begin{cases} \sqrt{\frac{1-j\cot\alpha}{2\pi}}\exp(j\frac{t^2+u^2}{2}\cot\alpha - jtu\csc\alpha), & \alpha \neq n\pi \\ \delta(t-u), & \alpha = 2n\pi \\ \delta(t+u), & \alpha = (2n\pm1)\pi \end{cases} \tag{8}$$

170   According to the properties of (8), when $\alpha = \frac{\pi}{2}$ the fractional Fourier transform is equal to the

171   traditional Fourier transform.

172   (B) Eq. (9) provides the spectrum mapping operator from fractional domain to Mel frequency domain.

$$M(u) = 1125\ln(1+u/700) \tag{9}$$

174 The Mel frequency band is based on the human auditory characteristics, it should also meet such requirement
175 in fractional domain. So Eq. (10) presents the frequency projection operator.

176
$$u = f \times \sin \alpha \qquad (f \text{ is linear frequency}) \qquad (10)$$

177 The output of each triangular filter is $|S_\alpha^M(u)|$, M refers to M-th Mel component. Figure 2 shows the

178 fractional Mel frequency schematic.
179

180 (C) The fractional Mel cepstral coefficients (FrCC) can be achieved after a DCT of $|S_\alpha^M(u)|$.

181
$$FrCC_n = \sqrt{\frac{2}{N}} \sum_{k=1}^{M} Log \, | S_\alpha^M(u)| \cos[\pi(k-0.5)n/M] \qquad (11)$$

182
183 The FrCC parameters can be calculated according to the above equations. And FrCC is equal to MFCC

184 when $\alpha = \dfrac{\pi}{2}$.

185

3. LDA and HMM Model

187 3.1 LDA Algorithm
188 The main function of Linear Discriminant Analysis (LDA) is to project the high-dimensional samples
189 onto a low dimensional space. It is aimed to maximize the distance between classes, and minimize the distance
190 in the class. So LDA is suitable for two group classification task, such as truthfulness and deceptiveness

191 division. The $S = \{s_1, s_2, \ldots\ldots s_n\}$ refers to the training voice set, and s_i is belong to ω_1 or ω_2, which

192 represents for normal speech or lie respectively. A projection operator w defined as the best vector may map

193 x_i to the one-dimensional y.

194
$$y = w^T s_i \qquad (12)$$

195 Then, it is very easy to make a decision by a simple comparison.

196
$$\begin{cases} s_i \in \omega_1 & y \geq t \\ s_i \in \omega_2 & y < t \end{cases} \qquad (t \text{ is a threshold}) \qquad (13)$$

197 So the main task of LDA is to calculate the optimal mapping vector w.

198 The mean of each category can be expressed as $\mu_i = \dfrac{1}{N_i} \sum_{s \in \omega_i} s_i \qquad (14)$

199 The mean value is changed after projection.

200
$$\tilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y = \frac{1}{N_i}\sum_{s\in\omega_i} w^T s = w^T \mu_i \tag{15}$$

201 The distance between two means is $D(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T(\mu_1 - \mu_2)|$ (16)

202 And the variance after projection is $\tilde{d}_i = \sum_{y\in\omega_i}(y-\tilde{\mu}_i)^2$ (17)

203 Define the objection function $J(w)$, when reaching the max ratio of the distance between these two categories
204 and the variance within the classes, the w is the best vector.

205
$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{d}_1 + \tilde{d}_2} \tag{18}$$

206 The Eq. (17) can be decomposed into (19).

207
$$\tilde{d}_i = \sum_{y\in\omega_i}(y-\tilde{\mu}_i)^2 = \sum_{s\in\omega_i}(w^T s - w^T \mu_i)^2 = \sum_{s\in\omega_i} w^T(x-\mu_i)(x-\mu_i)^T w \tag{19}$$

208 Then define $d_i = \sum_{s\in\omega_i}(s-\mu_i)(s-\mu_i)^T$ (20)

209 And we may have $d_w = d_1 + d_2$ (21)

210 So $\tilde{d}_1 + \tilde{d}_2 = w^T d_w w$ (22)

211 And $(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (w^T\mu_1 - w^T\mu_2)^2 = w^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w = w^T d_k w$ (23)

212 The Eq. (18) can be written as $J(w) = \dfrac{w^T d_k w}{w^T d_w w}$ (24)

213 The optimal vector is $w = d_w^{-1}(\mu_1 - \mu_2)$ (25)

214 The test voice set can be easily divided into two groups by Eq. (12) and (13) when the w is obtained.
215
216 3.2 Hidden Markov Model
217   The hidden Markov model (HMM) can be considered as a generalization of a mixture model. The hidden
218 variables are related through a Markov process, and the observation is controlled by the hidden state. So the
219 state is not directly visible in a HMM, but output observation is visible and dependent on the state. Each state
220 has a probability distribution corresponding to the possible output. Therefore the output sequence generated by
221 an HMM presents some information about the sequence of invisible states.
222   The random variable $x(t)$ presents the hidden state at time t (with $x(t) \in \{x_1, x_2, x_3\}$). The random
223 variable $y(t)$ is considered as the speech observation at time t (with $y(t) \in \{y_1, y_2, y_3, y_4\}$). According to the
224 basic theory of Markov process, it is clear that the conditional probability distribution of the hidden variable $x(t)$
225 only depends on the value of the $x(t-1)$. The values at time t-2 and before have no influence. The value of the
226 speech observation $y(t)$ also depends on the value of the $x(t)$.

227   Two types of parameters called transition probabilities and output probabilities are contained in a HMM.
228   The hidden state at time t is determined by hidden state at time t-1 according to the transition probabilities.
229   There is also a set of output probabilities to describe the distribution of the observed variable.
230   There are some important parameters in a HMM.
231   1. N, the number of states in the model.
232   2. M, the number of observation symbols per state.
233   3. The transition probability distribution.

$$a_{ij} = P(x_{t+1} = s_j \mid x_t = s_i) \tag{26}$$

234

235   4. The output observation probability distribution.

$$b_j(k) = P(y_t = o_k \mid x_t = s_j) \tag{27}$$

236

237   5. The initial state distribution.

$$\pi_i = P(x_t = s_i) \tag{28}$$

238

239   The famous forward-backward algorithm, EM algorithm, and Viterbi algorithm can be used to train the
240   models and solve the recognition task [20,21]. Though the HMM is a traditional method used in recognition
241   systems, it is still a suitable model in deceptive speech detection with time series speech signals.
242

4. Experiment results and analysis

243

244   4.1 Speech database

245   The liar's appearance has a direct relationship with individual personality,
246   culture background, conversation content and the cost of being seen through the lie. So the speech sources
247   should be collected in a real circumstance. According to the concealed information test theory [30], we designed
248   an interesting game, and the speech database is selected from the game records. There are two groups in the
249   game called A-group and B-group. Every person in A-group should tell a story, and the persons in B-group can
250   ask all kinds of questions according to the story. Due to the different stories told by peoples in A-group, the
251   questions and answers are different from each other. Since the persons in B-group do not know whether the
252   story is experienced the storyteller himself/herself, they should decide the true or false through the teller's
253   answers. If the persons in B-group speculate the correct result, they win the game and obtain some reward.
254   Otherwise, the person in A-group wins. So if the story is a lie, the teller should try his/her best to keep the
255   secret from every one to win the game. Peoples in B-group should ask as many questions as possible
256   (generally more than 10 questions) to make the liars nerves and mistakes. Here, we select the every fake stories
257   and fake answers as the deceptive speech samples. Then the corresponding tellers should record a set of
258   normal speech in a calm environment, the topics can include such as self-introduction, hobbies, and daily life
259   topics and so on. These records should long enough to cover as much as possible syllables in Mandarin
260   Chinese. So the normal speech samples are collected.
261   At last, we reserved useful records of 50 participants, including the 25 men and 25 women. Due to some
262   limitations, the participants are mainly 25 to 35 years old. The SNR of all the samples are more than 25dB. All
263   speech is mono sampled at 8 kHz and quantified with 16 bits. The frame length is 20 ms, and the overlap is 10

264   ms when the speech is under short-time analysis. The data set is divided into two parts, namely the training set
265   and test set. The experiment is under a unified standard to divide the data set due to the different length of the
266   every people's speech sample. The training set contains 30% of whole record, and the remaining data are
267   regarded as the test set.

Human Ethics

269   This research was approved by the Institutional Review Boards of Soochow University School of
270   Electronics and Information Engineering, and Suzhou University of Science and Technology School of
271   Electronics and Information Engineering. The speech set recording is carried out in a game style, so all the
272   participants are confirmed with verbal consent.

273

274   4.2 The experiment results for LDA model
275   The experiment step is expressed as follows:
276   (A) Divide the speech signals into short frames with the length of 20 ms and overlap of 10 ms.
277   (B) Extracted the FrCC parameters from every frame, and 12 FrCC coefficients and 12 delta FrCC coefficients

278   are used as the FrCC vector from one speech frame. The range of angle is $\alpha \in (0, \pi)$,   with the 0.01π as

279   the step. So there are 100 FrCC vector groups in every frame.
280   (C) Select 30% of the total data as the train set, and use LDA algorithm to calculate the optimal vector w.
281   (D) Take the remaining 70% of data as test set. Map the test set into low-dimensional space by w.
282   Then use Eq. (12) and (13) to make the decision and the statistical accuracy can be obtained at last.
283   In this experiment, the recognition results of MFCC parameters are taken as a benchmark to compare with
284   that of FrCC parameters. The results are shown in table 1, 2 and figure 3, 4. The first line of these tables
285   denotes the people's ID. The second line indicates the corresponding α of highest accuracy for FrCC. The
286   accuracy of FrCC and MFCC are shown in line 3 and 4 respectively.
287   In order to further refine the improvement of FrCC, the vector variance is introduced to compare the
288   clustering performance of the two parameters. The vector variance is shown in Eq. (29) and (30).

289
$$R_1 = \frac{\sum_{i=1}^{N}(normfrcc_i - \overline{normfrcc})^2}{\sum_{i=1}^{N}(normmfcc_i - \overline{normmfcc})^2} \qquad (29)$$

290
$$R_2 = \frac{\sum_{j=1}^{M}(decfrcc_j - \overline{decfrcc})^2}{\sum_{j=1}^{M}(decmfcc_j - \overline{decmfcc})^2} \qquad (30)$$

291   Here, the $normfrcc_i$ presents the FrCC of normal speech, and the $\overline{normfrcc}$ presents the mean vector.

292   The $normmfcc_i$ and $\overline{normmfcc}$ present the MFCC of normal speech and MFCC mean vector respectively.

293   The R_1 in Eq. (29) denotes the vector variance ratio between FrCC and MFCC of normal speech. The R_2 in

294   Eq. (30) denotes the vector variance ratio between FrCC and MFCC of deceptive speech. The results are

295   presented in Table 3 and 4.

296

297   4.3 The experiment results of HMM model

298   There are many sophisticated tools for HMM training and testing, such as HTK or Matlab software

299   package. The speech signals should also be divided into short frames. Then the speech characters such as FrCC

300   and MFCC can be regarded as the observations, the psychophysiology status is regarded as the invisible states.

301   The 30% of the total data is regarded as the train set, and the remaining data is test set. The speech characters

302   changed frame to frame, and the hidden Markov chain can present the process of the psychophysiology

303   changes. The experiment results are shown is the table 5 and 6.

304

305   4.4 Results Analysis and Discussion

306   In the section 4.2 and 4.3, the experiment results show that the identification accuracy of FrCC

307   parameters under certain angles is higher than that of MFCC parameters. The FrCC coefficients introduced to

308   the LDA model make the clustering performance much better. The accuracy will be increased when HMM

309   model is used to enhance the contextual information. The following paragraphs give some brief explanation to

310   the experiment results.

311   (A) In the LDA recognition system, the men groups' average accuracy of FrCC with best angle is 59.9%, and

312   MFCC is 56.3%. The average of best angle is $\overline{\alpha} = 0.51\pi$, and the variance of α is $D(\alpha) = 0.23\pi$. The

313   women groups' average accuracy of FrCC with best angle is 56.2%, and MFCC is 50.3%. The average of best

314   angle is $\overline{\alpha} = 0.59\pi$, and the variance of α is $D(\alpha) = 0.22\pi$. The best angle of 10, 22 in men's group and 8

315   in women's group is $\dfrac{\pi}{2}$. In these cases, the FrCC coefficients are equal to MFCC coefficients. In the other

316   cases, the identification performance of FrCC under LDA model is better than that of MFCC. The accuracy

317   increased from 36.7% to 56.0% when FrCC is introduced in the 16th men. And the accuracy of many people is

318   increased over 10%. Due to individual differences, the accuracy is only a little increased with some people.

319   Overall, when FrCC coefficients are involved, the average accuracy is increased by 3.6% in men group and 5.9%

320   in women group respectively. Although the average growth of accuracy is not very large, there are great

321   progresses with some individuals. The FrCC parameters can therefore improve the deceptive detection

322   performance.

323   (B) The performance of FrCC parameters may be changed with different α of FrFT. Due to the diversity and

324   non-stationary characteristics of speech, and personality difference of the speakers, it is impossible to

325   determine the optimal α before the experiment. The best α is selected by the highest accuracy. So the

326   mechanism of selection algorithm should be further studied.

327   (C) Most of the R_1 and R_2 is less than 1 in table 3 and 4. It is shown that the clustering performance of certain

328   FrCC is much better than that of MFCC. The FrCC data is concentrated to the clustering center. The existence

329   of α enhanced the appearance of the change of articulator, when people are lying or under stress. These faint

330   details could not be reflected by MFCC. These conclusions may be explained by phase change statement
331   described in section 1.2, and should be further verified by physiology research.
332   (D) The experiment results show that the performance of FrCC is better than that of MFCC. That is to say the
333   new character may be more suitable in some speech based psychophysiology information processing field.

334   There are 25 men and 25 women participants, and the range of angle is $\alpha \in (0, \pi)$, with the 0.01π as the

335   step. So there are 2500 recognition accuracy results in each men set and women set. The distribution of FrCC
336   (with all angle α) based man and woman deceptive speech recognition accuracy is presented in the Fig. 7 and
337   8 respectively. The black solid line is accuracy distribution, and the red dotted line is the average accuracy of
338   MFCC. According to the statistical result, the men groups' average accuracy of MFCC is 56.3%, so
339   approximately 20.7% (518/2500) of the FrCC based recognition accuracy is higher than 56.3%. The women
340   groups' average accuracy of MFCC is 50.3%, so approximately 60.8% (1520/2500) of the FrCC based
341   recognition accuracy is higher than 50.3%.
342   (E) Each speech frame is regarded as the basic unit for deceptive speech detection under LDA model. It reveals
343   the advantages of FrCC coefficients for this classification task. The vector w is a global optimal vector, but the
344   speech flow is a time-varying process. If the context information can be introduced in the identification system,
345   and map the speech signal onto the state flows. Then use statistical model for classification, there may be a
346   better result.
347   (F) The speech is a time series signal, and the contextual information may be hidden among the speech frames.
348   So the HMM model can mining this information and reveal the relations among the adjacent speech frames. In
349   the HMM recognition system, the men groups' average accuracy of FrCC with best angle is 71.0%, and MFCC
350   is 66.4%. The women groups' average accuracy of FrCC with best angle is 70.2%, and MFCC is 65.0%. When
351   HMM model is involved, the average accuracy is increased by 11.1% and 14.0% respectively in two groups.
352   The highest accuracy of FrCC is 82.0% in men set, and 85.4% in women set. The largest individual accuracy
353   increase is 31.2% in women set (from 39.9% to 71.1%, 13th woman) and 18.5% in men set (from 45.5% to
354   64.0%, 12th man). The FrCC based deceptive detection accuracy comparison between LDA and HMM are
355   shown in Fig. 9 and 10.
356   (G) The ROC curve [31] is usually used to analysis the performance of the identification system. Here, the
357   deceptive detection is a binary classification problem, in which the outcomes are labeled either as positive (p)
358   or negative (n). There are only four outcomes from a binary classifier, the true positive (TP), false
359   positive (FP), true negative (TN) and false negative (FN). So we select two parameters, the true positive rate
360   (TPR, sensitivity) and true negative rate (TNR, specificity) to evaluate the performance of the LDA and HMM
361   model. The sensitivity defines how many correct positive results occur among all positive samples during the
362   experiments. Specificity defines how many correct negative results occur among all negative samples during
363   the experiment. The definition equations are shown in (31) and (32). The statistical results are presented from
364   Fig.11 to Fig.14. The difference between the sensitivity and specificity of every participant is not large, so
365   LDA and HMM are the suitable tool for dividing the normal or deceptive speech.

$$sensitivity = \frac{TP}{TP + FN}$$

366   (31)

$$specificity = \frac{TN}{FP+TN}$$

367

(32)

368

369   In summary, the experimental results at least reflected that the acoustic feature is effective for lie
370   detection. Faint difference between truthfulness and deceptiveness can be expanded under some improved
371   acoustic features such as FrCC, and these characteristics may play an important role in deceptive speech
372   identification.

373

374   5 Conclusion and prospect

375   Lie detection based on Speech signal analysis is affected by many factors, such as the psychological
376   quality of the subjects, the way of speaking, interference of environment, and the cost of being exposed, etc. So
377   the development of this technology is relatively difficult. The lack of psychological and physiological research
378   basis also makes less progress in this field. In this paper, fractional Mel cepstral coefficient (FrCC) has been
379   proposed as the speech feature, Linear Discriminant Analysis model (LDA) and hidden Markov model (HMM)
380   are introduced as the classifier. The experiment results show that the clustering effect of FrCC under optimal
381   angles is better than that of MFCC, and the truthfulness/deceptiveness identification accuracy of FrCC is
382   higher than that of MFCC through LDA or HMM. The successful application has demonstrated that the FrCC
383   parameter can be used in deceptive speech detection. And it gives some further experiment evidence in this
384   field.

385   The future work mainly focuses on the following aspects. First, establish a unified optimal angle search
386   mechanism, and achieve complete extraction algorithm of FrCC. Second, further deep mining related features,
387   construct data fusion model, enhance the useful property, and compress the redundant information and
388   interference. Third, deep mining the time series model, and enhance the contextual information for deceptive
389   speech detection. Speech based deceptive detection may be an important aid for the traditional neuroimaging
390   methods.

391

392   **References**

393   [1] Gao JF, Zhang WJ, Yang Y, Hu JJ, Tao CY and Guan JA. (2014), *Lie Detection Study Based on P300*
394   *and Extreme Learning Machine*, Journal of University of Electronic Science and Technology of China, 43,
395   2,301-305.
396   [2] Dong SS,Chen FY,He HJ. (2013), *Lie Detection Using Brain Imaging Technology and Its*
397   *Psychophysiological Basis*, ACTA BIOPHYSICA SINICA, 29, 2, 94-104.
398   [3] Anton N. (2012), *Computational deception and noncooperation*, IEEE Intelligent Systems, 60-75.
399   [4] Eriksson A, Lacerda F. (2007), *Charlatanry in Forensic Speech Science: a Problem to be Taken Seriously*,
400   International Journal of Speech Language and the Law, 14, 2, 169-193.
401   [5] Harnsberger J D, Hollien H, Martin C A, Hollien KA. (2009), *Stress and Deception in Speech: Evaluating*
402   *Layered Voice Analysis*, Journal of Forensic Science, 54, 3, 642-650.
403   [6] Anolli L, Ciceri R. (1997), *The Voice of Deception: Vocal Strategies of Naïve and Able Liars*, Journal of

404   Nonverbal Behavior, 21, 4, 259-284.

405   [7] Bond, C F, DePaulo B M. (2006), *Accuracy of Deception Judgments*, Personality and Social Psychology
406   Review, 10, 3, 214-234.

407   [8] Hirschberg J, Benus S, Brenier J, Enos F, Friedman S, Gilman S, Girand C, Graciarena M, Kathol A,
408   Michaelis L. (2005), *Distinguishing Deceptive from Non-Deceptive Speech* , Proc. Eurospeech, Lisbon, 1833-
409   1836.

410   [9] Graciarena M, Shriberg E, Stolcke A, Enos F, Hirschberg J, Kajarekar S. (2006), *Combining Prosodic,*
411   *Lexical and Cepstral Systems for Deceptive Speech Detection*, Proc. IEEE ICASSP, Toulouse, 1033-1036.

412   [10] Christin K, David M. (2012), *Detecting suspicious behavior using speech: Acoustic correlates of*
413   *deceptive speech - an exploratory investigation*, Applied Ergonomics, 1-9.

414   [11] Pang CS, Liu L, Shan T. (2014), *Time-Frequency Analysis Method Based on Short-Time Fractional*
415   *Fourier Transform*, ACTA ELECTRONICA SINICA, 42, 2, 347-352.

416   [12] Zhu JD, Zhao YJ, Tang J. (2013), *Periodic FRFT Based Detection and Estimation for LFMCW Signal*,
417   Journal of Electronics & Information Technology, 35, 8, 1827-1833.

418   [13] Zhu WT, Su T, Yang T, Zheng JB, Zhang L. (2014), *Detection and Parameter Estimation of Linear*
419   *Frequency Modulation Continuous Wave Signal*, Journal of Electronics & Information Technology, 36, 3, 552-
420   558.

421   [14] Yin H, Xie X, Kuang JM. (2012), *Acoustic features based on auditory model and adaptive fractional*
422   *Fourier transform for speech recognition*, ACTA Acustica, 37, 1, 97-103.

423   [15] Pawan K A, and Raghunath S H. (2013), *Fractional Fourier transform based features for speaker*
424   *recognition using support vector machine*, Computers and Electrical Engineering, 39, 550-557.

425   [16] Qiu W, Li BZ, Li XW. (2013), *Speech recovery based on the linear canonical transform*, Speech
426   Communication, 55, 40-50.

427   [17] Behzad Z, Ahmad A, Babak N, Azarakhsh J. (2011), *Optimized discriminative transformations for speech*
428   *features based on minimum classification error*, Pattern Recognition Letters, 32, 948-955.

429   [18] Ana M B, Rubén F P, Doroteo T T, José L B M, Eduardo L G, Luis H G. (2014), *Analysis of voice*
430   *features related to obstructive sleep apnoea and their application in diagnosis support*, Computer Speech and
431   Language, 28, 434-452.

432   [19] Jordi S C, Cristian M, Oriol C M, Ferran B, Carlos Q, José A, Joaquín D C. (2014), *Detection of severe*
433   *obstructive sleep apnea through voice analysis*, Applied Soft Computing, in press.

434   [20] Rabiner R, Schafer W. (2007), *Introduction to digital speech processing*. Foundations and Trends in
435   Signal Processing, 1, 1, 1-194.

436   [21] Matthias W, John M. (2009), *Distant speech recognition*. Wiley, 283-316.

437   [22] Tommaso F, Fabio C and Massimo P. (2013), *The effect of personality type on deceptive communication*
438   *Style*, European Intelligence and Security Informatics Conference, 1-6.

439   [23] Lamb C E, and Skillicorn D B. (2013), *Detecting deception in interrogation settings*, IEEE Intelligence
440   and Security Informatics. Seattle, 160-162.

441   [24] Christie M F, David P B and Dursun D. (2008), *Exploration of feature selection and advanced*
442   *classification models for high-stakes deception detection*, Proceeding of the 41st Hawaii international
443   conference on system sciences, 1-8.

444   [25] Mohammed M A, Lakshmi R. (2012), *Deceptive Phishing Detection System*, Proceedings of the

445   International Conference on Pattern Recognition, Informatics and Medical Engineering, 458-465.

446   [26] Sofia H, Pekka S, Elisabeth L, Eeva S, Susanna S. (2013), *The Association between Possible Stress*

447   *Markers and Vocal Symptoms*, Journal of Voice, 27, 6, 787.e1-787.e10.

448   [27] Cheryl L G, Kirk W B, Jennifer B C, Keith F C, Winter A S. (2013), *Vocal Indices of Stress: A Review*,

449   Journal of Voice, 27, 3, 390.e21-390.e29.

450   [28] Gopalan K and Wenndt S. (2007), *Speech analysis using modulation based features for detecting*

451   *deception*, The 15th International Conference on Digital Signal Processing, 619-622.

452   [29] Muhammad S and Kaliappan G. (2013), *Deception Detection in Speech Using Bark Band and*

453   *Perceptually Significant Energy Features*, IEEE 56th International Midwest Symposium on Circuits and

454   Systems, 1212-1215.

455   [30] Verschuere B, Ben-shakhar G, Meijer E (2011), *Memory Detection: Theory and Application of the*

456   *Concealed Information Test*, Cambridge University Press.

457   [31] Fawcett T. (2006), *An introduction to ROC analysis*, Pattern Recognition Letters, 27, 861–874.

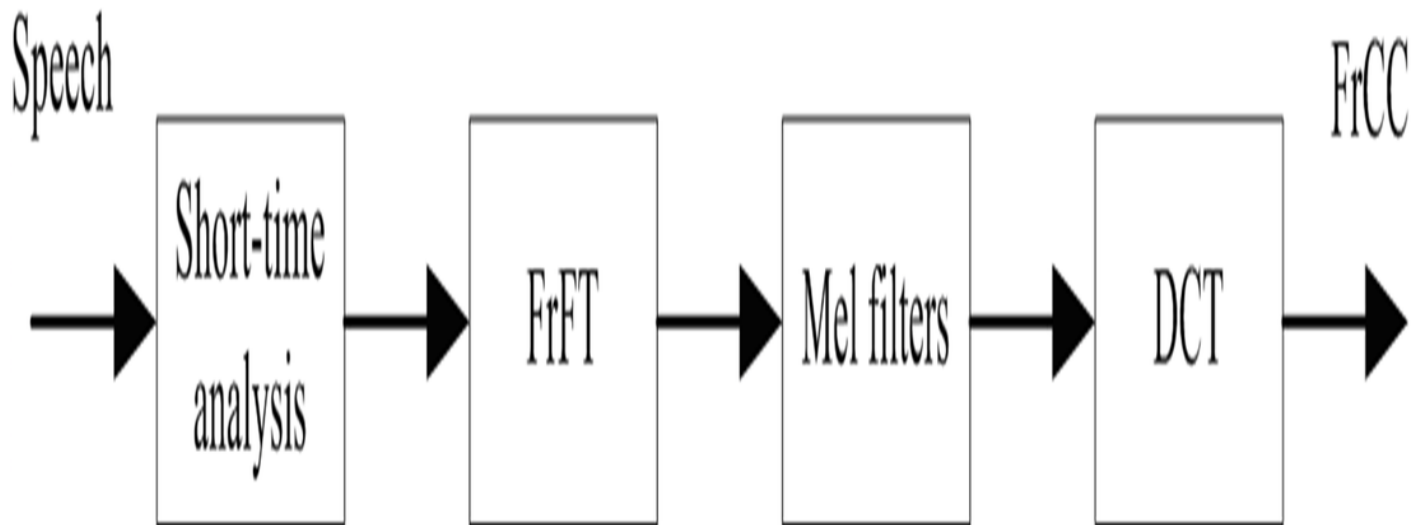
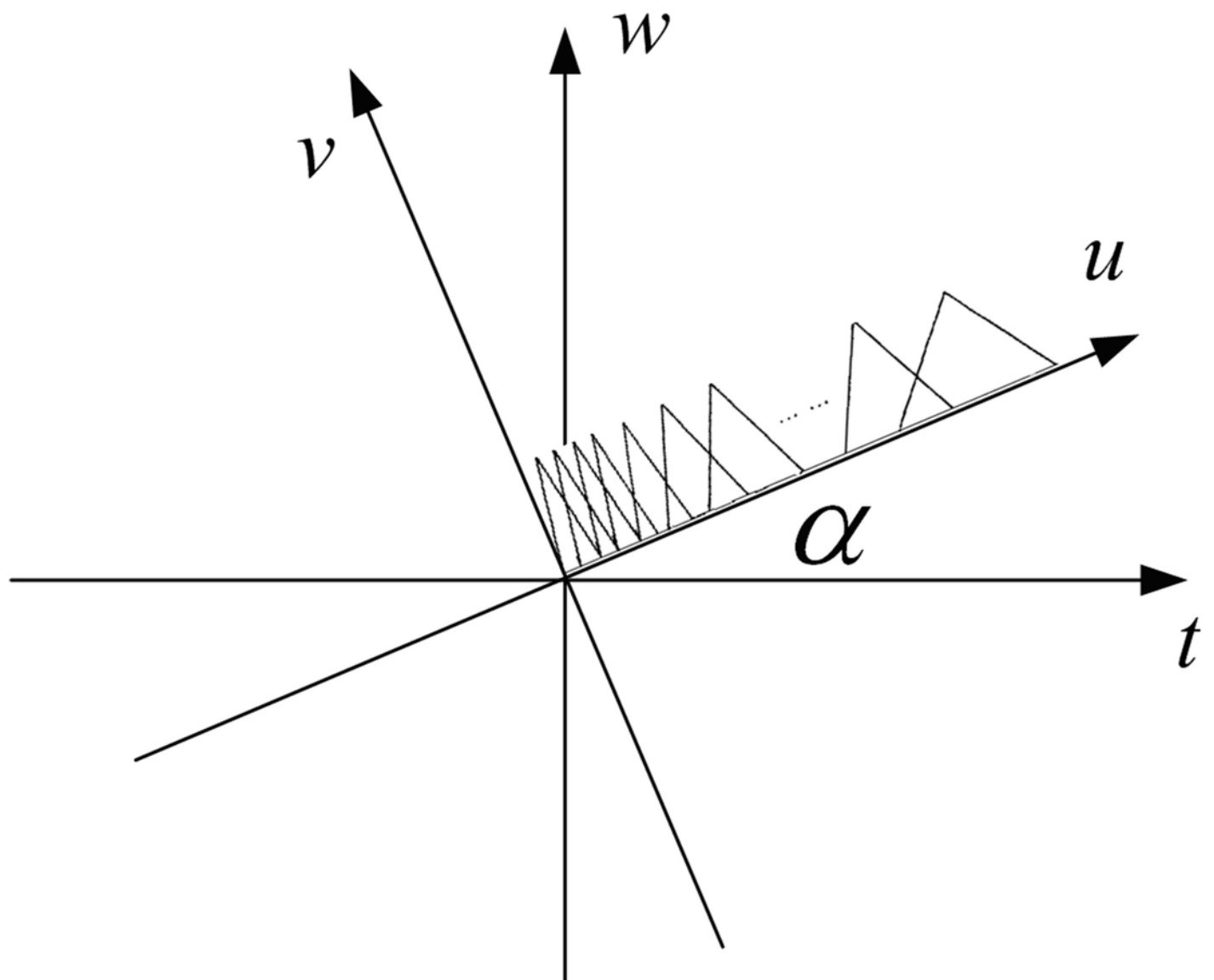
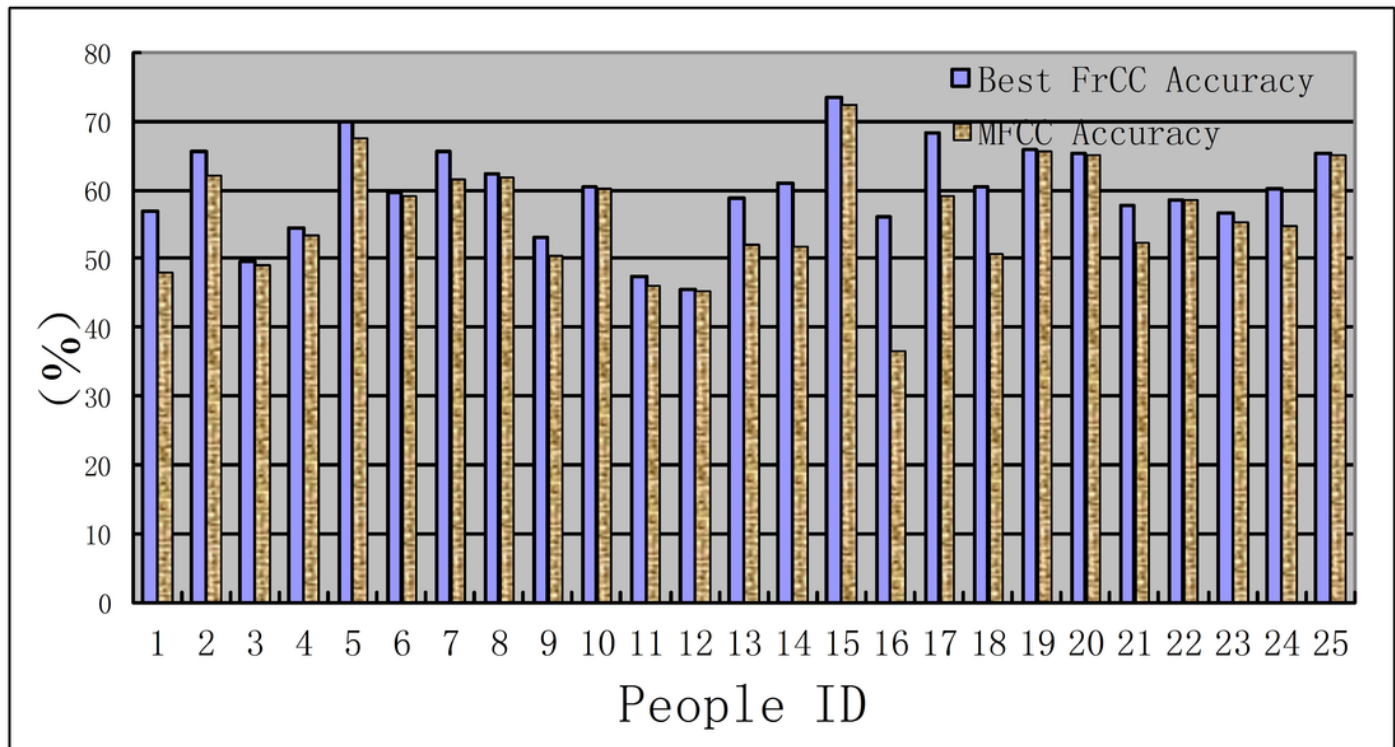
458

459

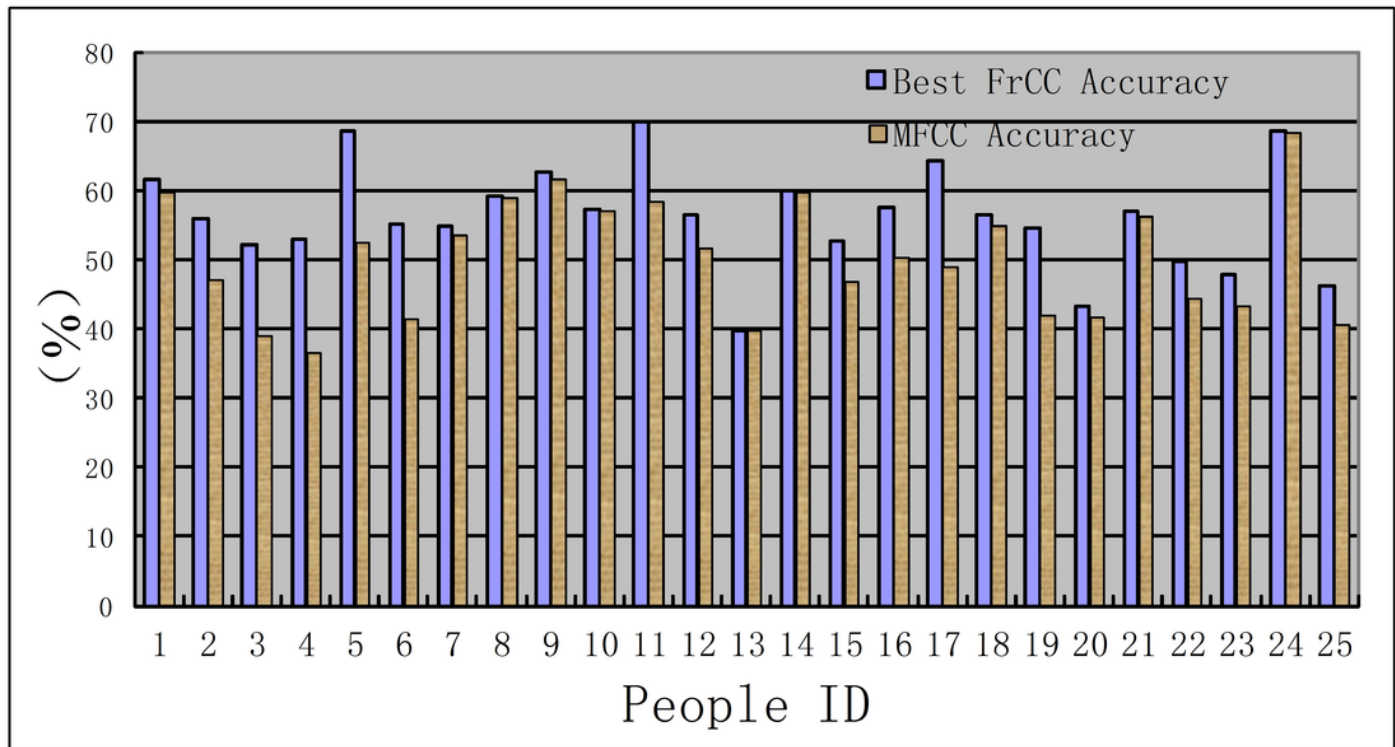
460

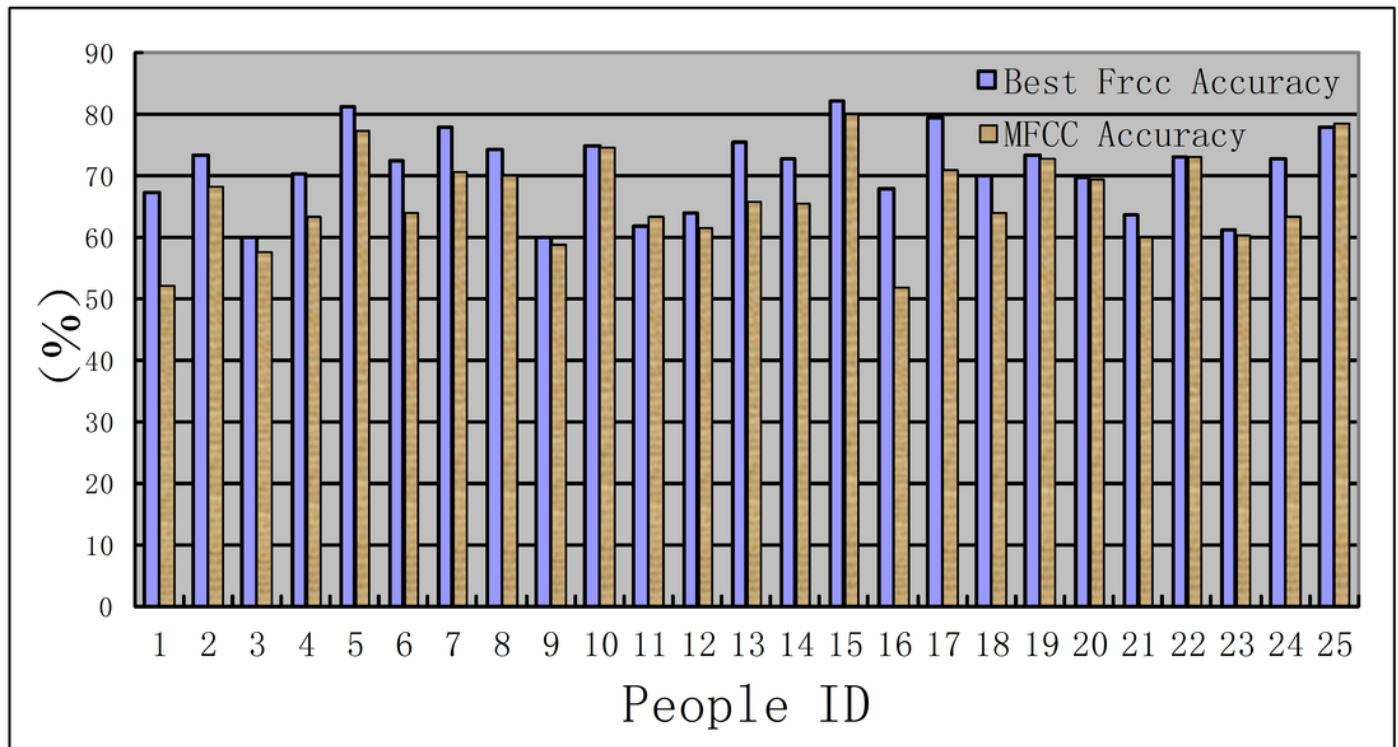
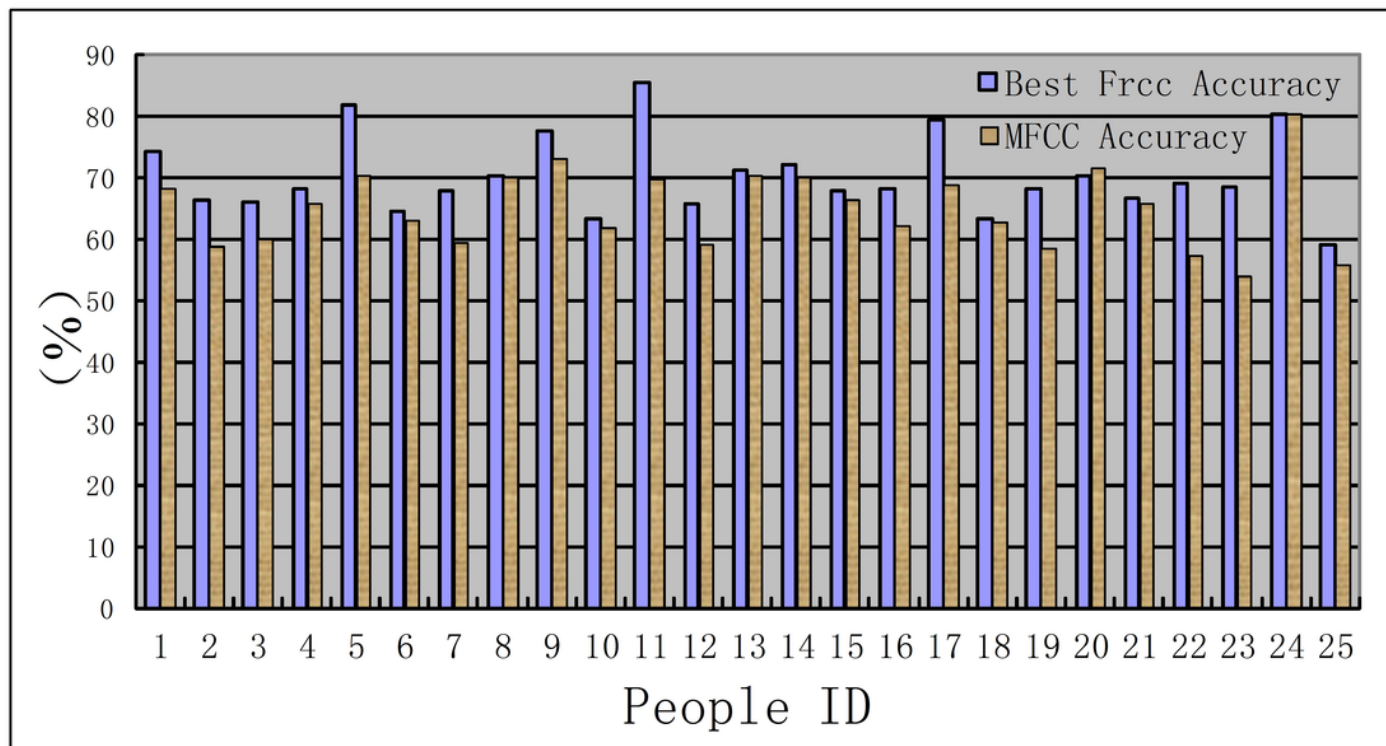
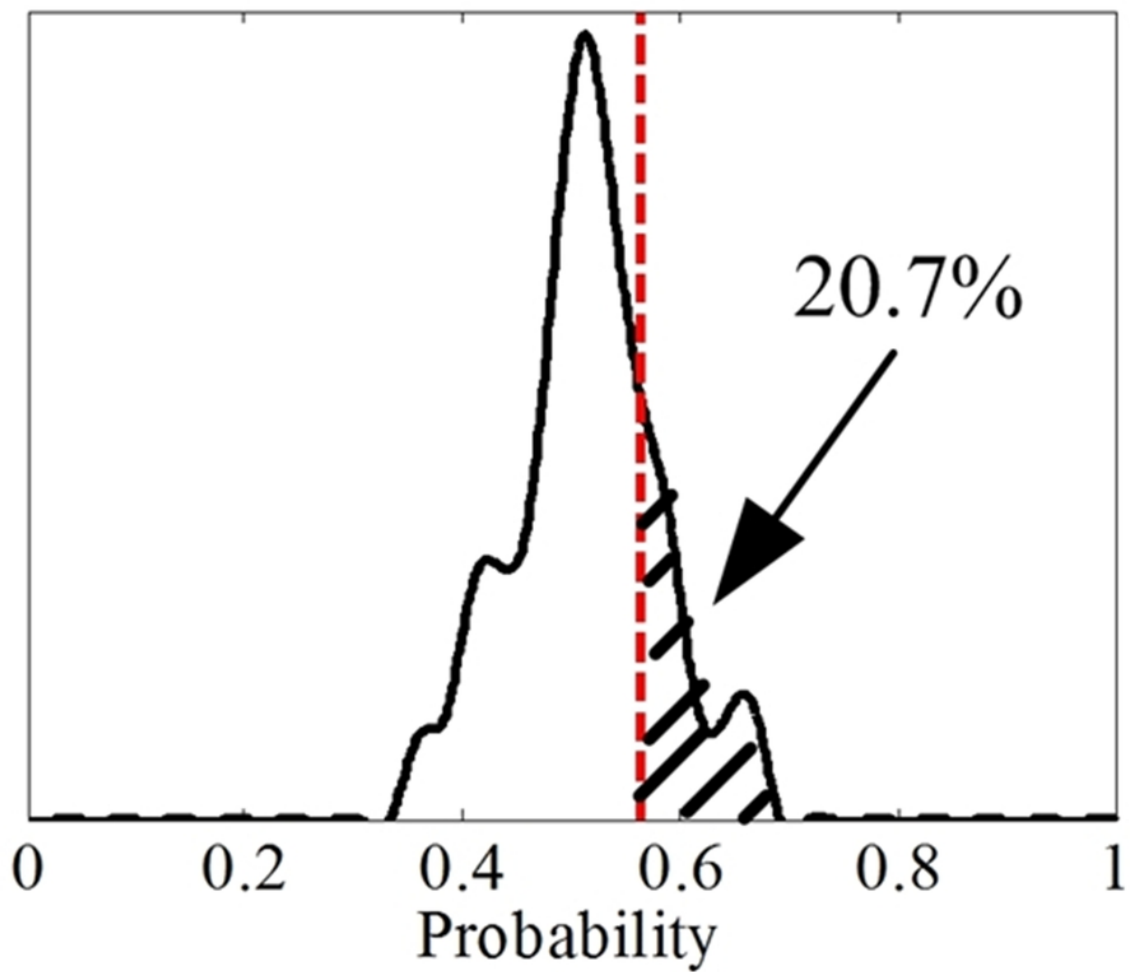
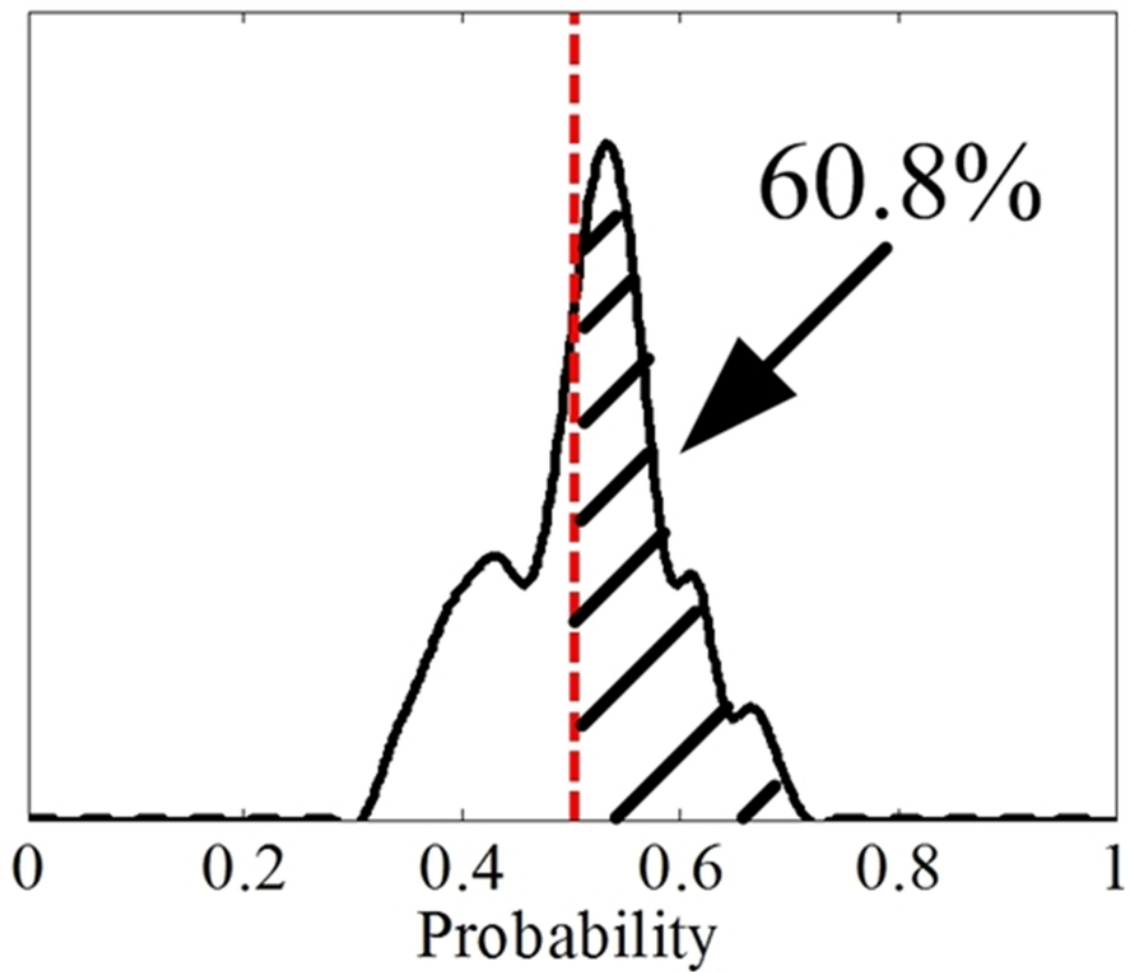
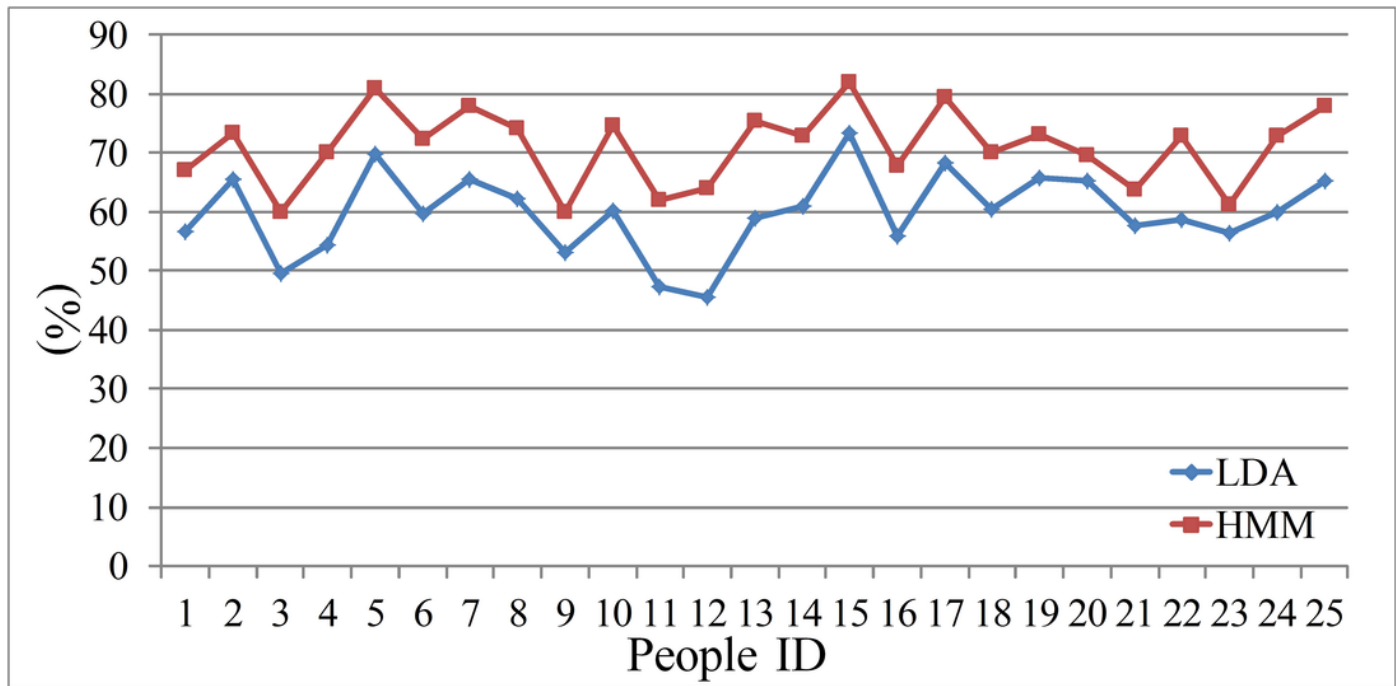
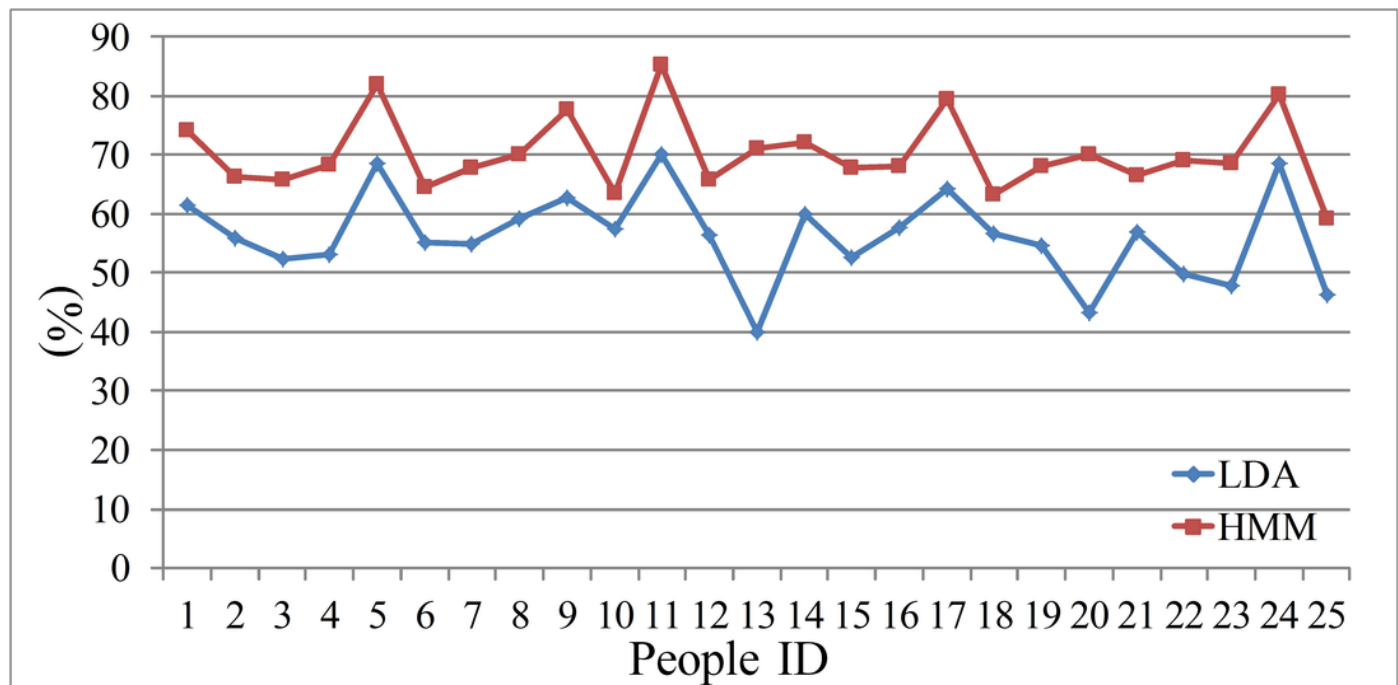
461

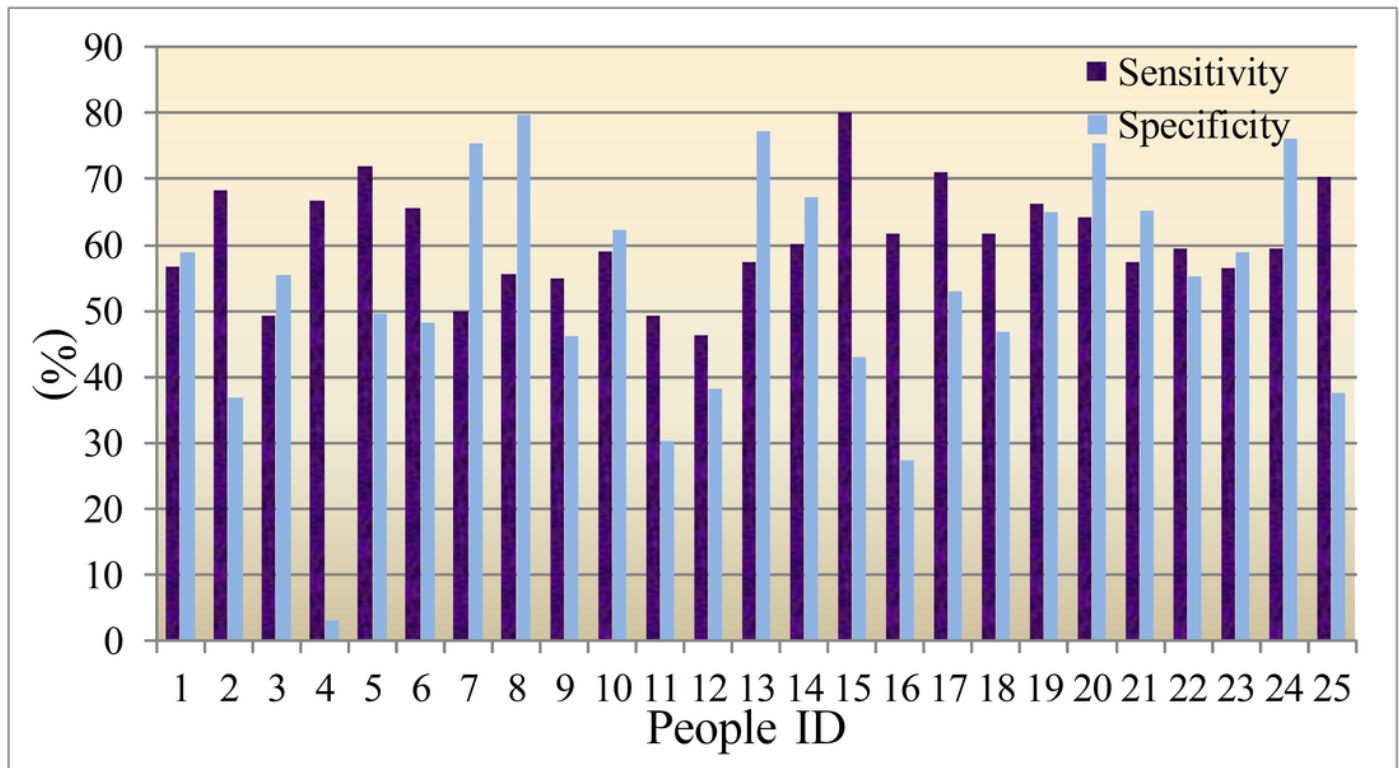
1

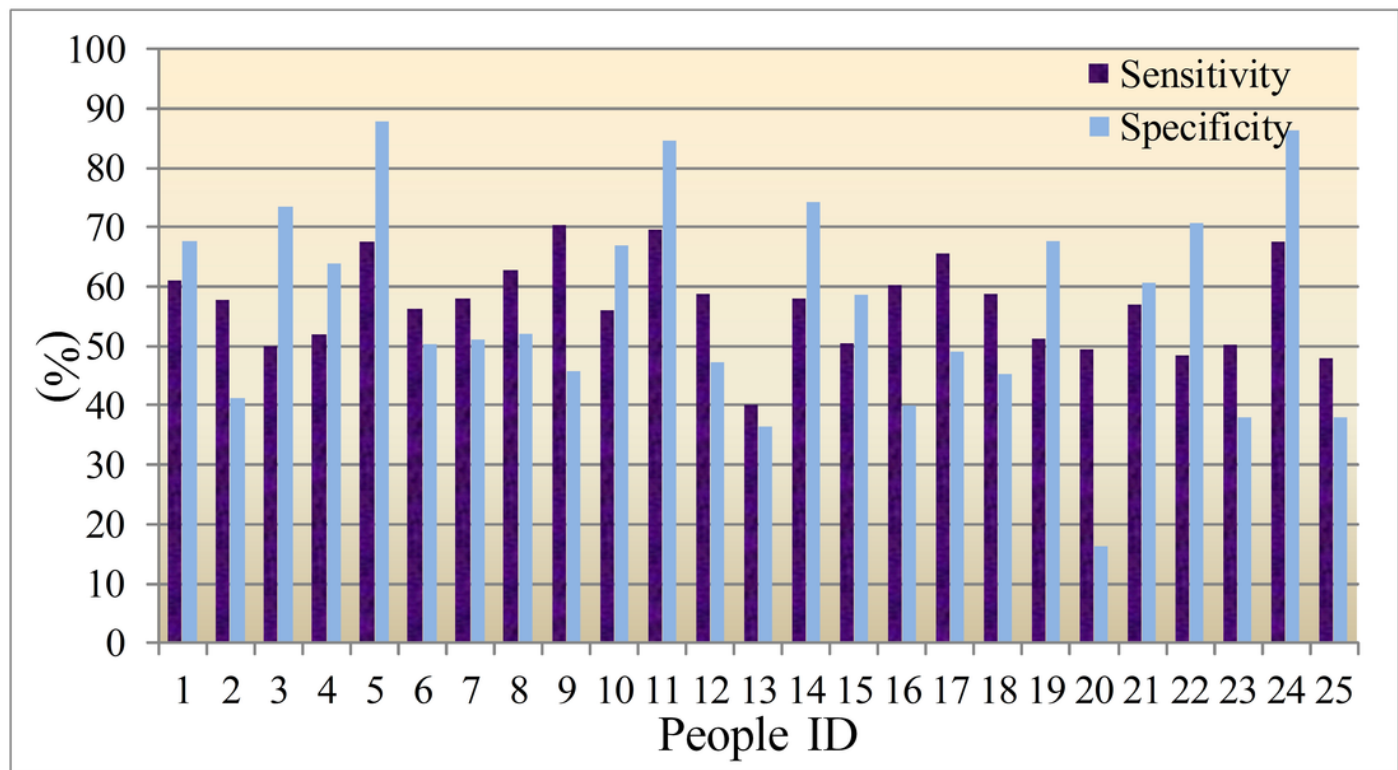
FrCC extraction process

The fractional Mel triangular filter schematic

Accuracy of men set under LDA model

4

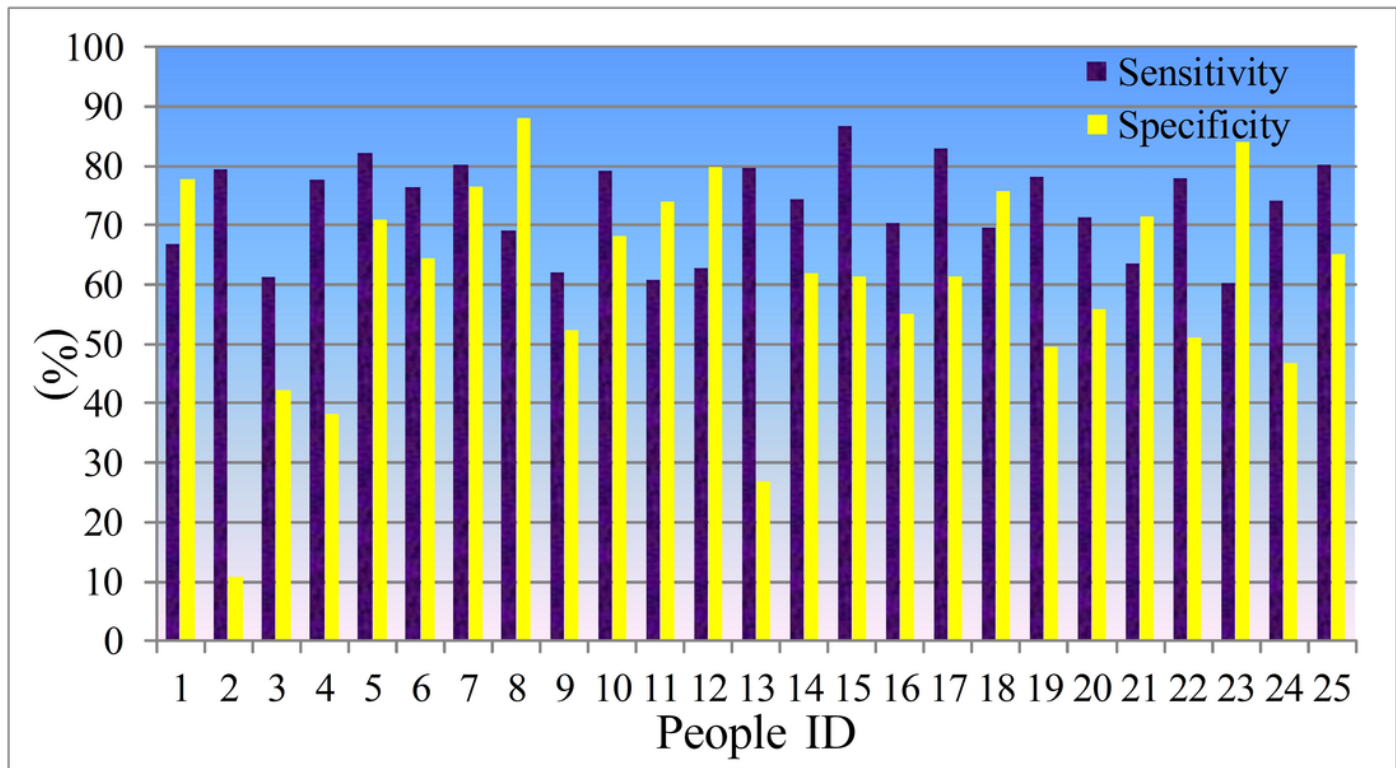
Accuracy of women set under LDA model

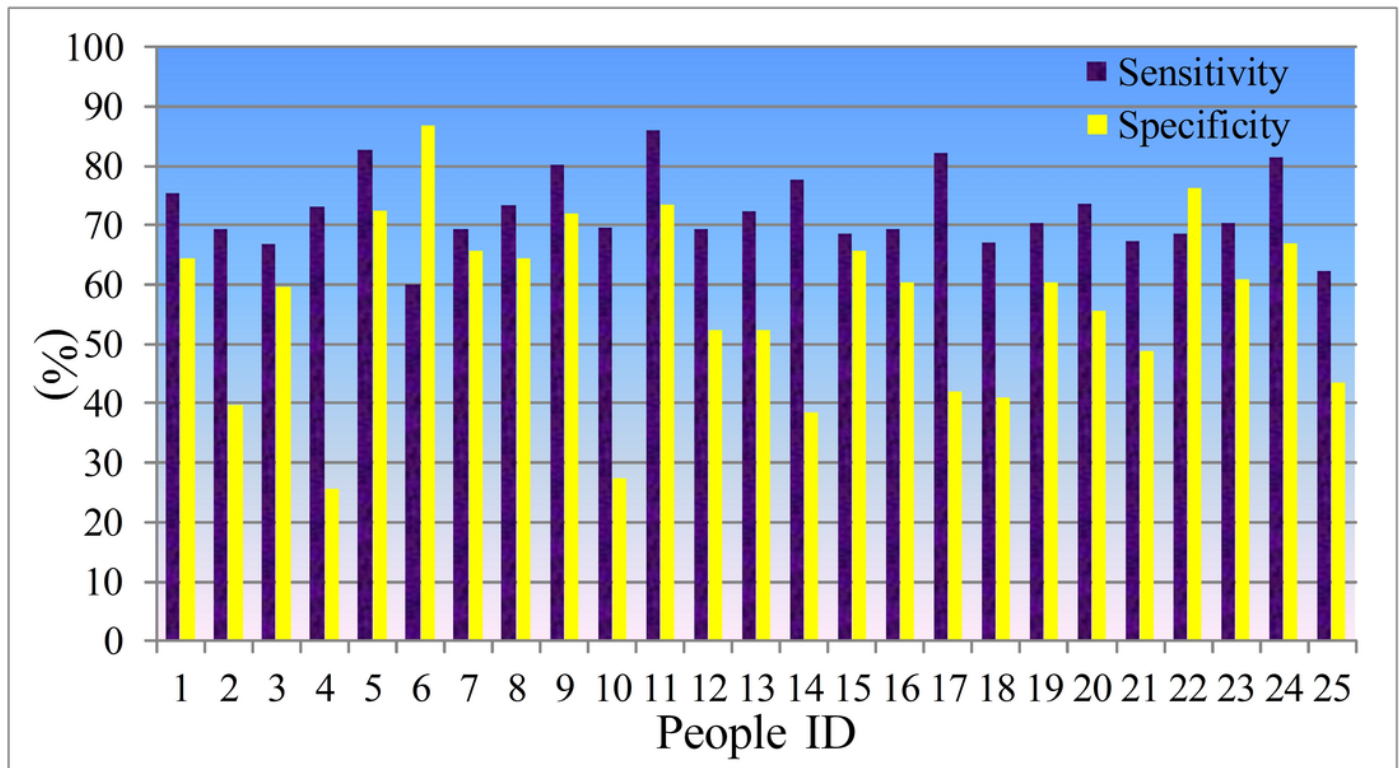
5

Accuracy of men set for HMM model

6

Accuracy of women set for HMM model

The distribution of FrCC based deceptive speech recognition accuracy by LDA in men set

The distribution of FrCC based deceptive speech recognition accuracy by LDA in women set

The FrCC detection accuracy of LDA and HMM in men set

The FrCC detection accuracy of LDA and HMM in women set

11

The sensitivity and specificity of men set in LDA model

12

The sensitivity and specificity of women set in LDA model

13

The sensitivity and specificity of men set in HMM model

14

The sensitivity and specificity of women set in HMM model

Accuracy of men set for LDA model

2
3

Table 1. Accuracy of men set for LDA model

People ID	1	2	3	4	5	6	7	8	9	10	11	12
α($\times\pi$)	0.83	0.70	0.08	0.90	0.52	0.52	0.88	0.49	0.62	0.50	0.37	0.52
FrCC Accuracy (%)	56.8	65.5	49.5	54.4	69.8	59.7	65.6	62.2	53.2	60.3	47.4	45.5
MFCC Accuracy (%)	48.0	62.1	49.1	53.6	67.7	59.2	61.6	61.9	50.5	60.3	46.1	45.4

Continued Table1. Accuracy of men set for LDA model

People ID	13	14	15	16	17	18	19	20	21	22	23	24	25
α($\times\pi$)	0.01	0.99	0.52	0.15	0.20	0.34	0.51	0.51	0.02	0.50	0.48	0.97	0.51
FrCC Accuracy (%)	58.9	60.9	73.4	56.0	68.2	60.5	65.9	65.2	57.6	58.6	56.5	60.1	65.4
MFCC Accuracy (%)	52.1	51.9	72.6	36.7	59.3	50.8	65.7	65.1	52.5	58.6	55.5	54.8	65.3

7

Accuracy of women set for LDA model

2
3

Table 2. Accuracy of women set for LDA model

People ID	1	2	3	4	5	6	7	8	9	10	11	12
α ($\times\pi$)	0.65	0.77	0.96	0.97	0.75	0.97	0.48	0.50	0.60	0.57	0.77	0.44
FrCC Accuracy (%)	61.5	56.0	52.3	53.1	68.7	55.2	55.0	59.1	62.7	57.4	70.1	56.4
MFCC Accuracy (%)	59.8	47.3	39.0	36.6	52.5	41.6	53.7	59.1	61.7	57.2	58.4	51.9

Continued Table 2. Accuracy of women set for LDA model

People ID	13	14	15	16	17	18	19	20	21	22	23	24	25
α ($\times\pi$)	0.51	0.52	0.21	0.36	0.09	0.99	0.99	0.51	0.45	0.01	0.68	0.51	0.60
FrCC Accuracy (%)	39.9	60.1	52.6	57.7	64.3	56.6	54.7	43.3	57.0	49.9	47.8	68.6	46.3
MFCC Accuracy (%)	39.8	60.0	47.0	50.4	49.1	55.0	42.2	41.7	56.5	44.6	43.3	68.6	40.8

7
8
9
10
11

Vector variance ratio of men set

2
3
4

Table 3. Vector variance ratio of men set

People ID	1	2	3	4	5	6	7	8	9	10	11	12
α (×π)	0.83	0.70	0.08	0.90	0.52	0.52	0.88	0.49	0.62	0.50	0.37	0.52
R_1	0.75	0.79	0.55	0.55	0.99	1.01	0.63	0.99	0.87	1.00	0.70	0.98
R_2	0.74	0.70	0.53	0.62	1.00	0.97	0.66	0.99	0.84	1.00	0.72	0.97

Continued Table 3. Vector variance ratio of men set

People ID	13	14	15	16	17	18	19	20	21	22	23	24	25
α (×π)	0.01	0.99	0.52	0.15	0.20	0.34	0.51	0.51	0.02	0.50	0.48	0.97	0.51
R_1	0.60	0.59	0.99	0.69	0.70	0.74	0.99	0.99	0.57	1.00	1.01	0.71	1.00
R_2	0.63	0.50	0.98	0.64	0.67	0.67	1.00	0.98	0.52	1.00	1.01	0.59	0.99

8
9
10
11
12
13
14
15
16
17

Vector variance ratio of women set

2
3
4

Table 4. Vector variance ratio of women set

People ID	1	2	3	4	5	6	7	8	9	10	11	12
α ($\times\pi$)	0.65	0.77	0.96	0.97	0.75	0.97	0.48	0.50	0.60	0.57	0.77	0.44
R_1	0.82	0.68	0.59	0.49	0.66	0.54	0.99	1.00	0.98	0.98	0.61	0.96
R_2	0.81	0.66	0.49	0.54	0.66	0.57	0.98	1.00	0.89	0.99	0.63	0.94

Continued Table 4. Vector variance ratio of women set

People ID	13	14	15	16	17	18	19	20	21	22	23	24	25
α ($\times\pi$)	0.51	0.52	0.21	0.36	0.09	0.99	0.99	0.51	0.45	0.01	0.68	0.51	0.60
R_1	0.99	0.99	0.67	0.79	0.48	0.54	0.46	1.00	0.94	0.43	0.72	0.99	0.84
R_2	0.99	0.99	0.67	0.76	0.48	0.56	0.44	0.98	0.96	0.45	0.69	0.99	0.84

8
9
10
11
12
13
14
15
16

Accuracy of men set for HMM

2
3
4
5
6

Table 5. Accuracy of men set for HMM

People ID	1	2	3	4	5	6	7	8	9	10	11	12
α ($\times\pi$)	0.83	0.70	0.08	0.90	0.52	0.52	0.88	0.49	0.62	0.50	0.37	0.52
FrCC Accuracy (%)	67.1	73.3	59.9	70.1	81.1	72.3	77.9	74.2	60.0	74.7	61.9	64.0
MFCC Accuracy (%)	52.3	68.2	57.8	63.4	77.5	64.0	70.6	70.0	58.8	74.7	63.3	61.7

Continued Table 5. Accuracy of men set for HMM

People ID	13	14	15	16	17	18	19	20	21	22	23	24	25
α ($\times\pi$)	0.01	0.99	0.52	0.15	0.20	0.34	0.51	0.51	0.02	0.50	0.48	0.97	0.51
FrCC Accuracy (%)	75.5	72.8	82.0	67.7	79.5	70.0	73.2	69.6	63.7	73.0	61.2	72.8	77.9
MFCC Accuracy (%)	66.0	65.4	80.1	52.0	71.0	64.1	72.9	69.6	60.1	73.0	60.4	63.3	78.7

10
11
12
13
14
15
16
17

Accuracy of women set for HMM

2
3
4
5

Table 6. Accuracy of women set for HMM

People ID	1	2	3	4	5	6	7	8	9	10	11	12
α ($\times\pi$)	0.65	0.77	0.96	0.97	0.75	0.97	0.48	0.50	0.60	0.57	0.77	0.44
FrCC Accuracy (%)	74.1	66.3	65.9	68.2	81.9	64.5	67.7	70.2	77.6	63.4	85.4	65.7
MFCC Accuracy (%)	68.2	59.0	60.1	66.0	70.5	63.2	59.5	70.2	73.1	61.9	69.9	59.2

Continued Table 6. Accuracy of women set for HMM

People ID	13	14	15	16	17	18	19	20	21	22	23	24	25
α ($\times\pi$)	0.51	0.52	0.21	0.36	0.09	0.99	0.99	0.51	0.45	0.01	0.68	0.51	0.60
FrCC Accuracy (%)	71.1	72.1	67.7	68.1	79.4	63.2	68.0	70.2	66.6	69.0	68.5	80.3	59.1
MFCC Accuracy (%)	70.5	70.0	66.6	62.1	68.8	62.8	58.6	71.7	65.9	57.4	53.9	80.3	55.7

9
10
11
12
13
14
15
16
17
18
19