Aalborg Universitet

Automatic Smoker Detection from Telephone Speech Signals

Poorjam, Amir Hossein; Hesaraki, Soheila; Safavi, Saeid; Van hamme, Hugo; Bahari, Mohamad Hasan

Published in:
Speech and Computer

DOI (link to publication from Publisher):
10.1007/978-3-319-66429-3_19

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Poorjam, A. H., Hesaraki, S., Safavi, S., Van hamme, H., & Bahari, M. H. (2017). Automatic Smoker Detection from Telephone Speech Signals. In A. Karpov, R. Potapova, & I. Mporas (Eds.), *Speech and Computer: 19th International Conference, SPECOM 2017, Hatfield, UK, September 12-16, 2017, Proceedings* (pp. 200-210). Springer. https://doi.org/10.1007/978-3-319-66429-3_19

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: June 18, 2025

Automatic Smoker Detection from Telephone Speech Signals

Amir Hossein Poorjam[1], Soheila Hesaraki[2], Saeid Safavi[3],
Hugo van Hamme[4], and Mohamad Hasan Bahari[4]

[1] Audio Analysis Lab, AD:MT, Aalborg University, Aalborg, Denmark
ahp@create.aau.dk
[2] Department of Electrical Engineering, Faculty of Engineering,
Ferdowsi University of Mashhad, Mashhad, Iran
soheila.hesaraki@alumni.um.ac.ir
[3] ECE Division, Information Engineering and Processing Architectures Group,
School of Engineering and Technology, University of Hertfordshire, Hatfield, UK
s.safavi@herts.ac.uk
[4] Center for Processing Speech and Images (PSI), KU Leuven, Leuven, Belgium
hugo.vanhamme@kuleuven.be, mohamadhasan.bahari@esat.kuleuven.be

Abstract. This paper proposes an automatic smoking habit detection from spontaneous telephone speech signals. In this method, each utterance is modeled using i-vector and non-negative factor analysis (NFA) frameworks, which yield low-dimensional representation of utterances by applying factor analysis on Gaussian mixture model means and weights respectively. Each framework is evaluated using different classification algorithms to detect the smoker speakers. Finally, score-level fusion of the i-vector-based and the NFA-based recognizers is considered to improve the classification accuracy. The proposed method is evaluated on telephone speech signals of speakers whose smoking habits are known drawn from the National Institute of Standards and Technology (NIST) 2008 and 2010 Speaker Recognition Evaluation databases. Experimental results over 1194 utterances show the effectiveness of the proposed approach for the automatic smoking habit detection task.

Keywords: Smoker detection · i-Vector · Non-negative factor analysis · Score fusion · Logistic regression

1 Introduction

Speech signals carry speaker's important information such as age, gender, body size, language, accent and emotional/psychological state [1–5]. Automatic identification of speaker characteristics has a wide range of commercial, medical and forensic applications such as interactive voice response systems, service customization, natural human-machine interaction, recognizing the type of pathology of the speakers, and directing the forensic investigation process. In this research, we focus on speaker's smoking habit detection, which is an ingredient of speaker profiling systems and behavioral informatics. The effect of smoking habits also on different speech analysis systems such as speaker gender detection, age estimation, intoxication-level recognition and emotional state identification shows the importance of an automatic smoking habits detection system and motivates the analysis of the smoking habit effects of speech signals. Experimental studies show that many acoustic features of the speech signal such as fundamental frequency, jitter and shimmer are influenced by cigarette smoking [6,7]. Although experimental studies reveal the effect of smoking on different acoustic characteristics of speech, the relation of these acoustic cues with speaker smoking habits is usually complex and affected by many other factors such as speaker age, gender, emotional condition and drinking habits [3]. Furthermore, technical factors such as speech duration, recording device and channel conditions also influence the estimation accuracy and make smoking habit detection very challenging for both humans and machines.

In this paper, we propose an automatic smoker detection from the telephone speech signals. To our knowledge, this is the first work on this condition and thus the result of no baseline system is reported in this paper. However, we adopt and apply state-of-the-art techniques developed within speaker and language recognition fields.

Modeling speech recordings with Gaussian mixture model (GMM) mean supervectors is considered as an effective approach to convert variable duration signals into fixed dimensional vectors to be used as features in support vector machines (SVM) [8]. This technique has been successfully applied to different speech processing tasks such as speaker's age estimation [9]. While effective, GMM mean supervectors are of a high dimensionality resulting in high computational cost and difficulty in obtaining a robust model in the context of limited data. Consequently, dimension reduction through PCA-based methods has been found to improve performance in age estimation from GMM mean supervectors [9]. In the field of speaker and language recognition, recent advances using i-vector framework [10], which provide a compact representation of an utterance in the form of a low-dimensional feature vector, have considerably increased the classification accuracy [10,11]. I-vectors successfully replaced GMM mean supervectors in speaker age estimation too [12]. We have recently introduced a new framework for adaptation and decomposition of GMM weights based on a factor analysis similar to that of the i-vector framework [13]. In this method, namely non-negative factor analysis (NFA), the applied factor analysis is constrained such that the adapted GMM weights are non-negative and sum to unity. This method, which yields new low-dimensional utterance representation approach, was applied to speaker and language/dialect recognition successfully [13,14]. In this paper, we propose a hybrid architecture of NFA and i-vector frameworks for smoker habit detection. This architecture consists of two subsystems based on i-vectors and NFA vectors and score-level fusion of i-vector-based and NFA-based recognizers is considered to improve the classification accuracy. The performance of the proposed method is evaluated on a spontaneous telephone speech signals of National Institute of Standards and Technology (NIST) 2008 and 2010

Speaker Recognition Evaluation (SRE) databases. Experimental results confirm the effectiveness of the proposed approach.

2 System Description

2.1 Problem Formulation

In the smoking habit estimation problem, we are given a set of training data $D = \{\boldsymbol{\nu}_i, y_i\}_{i=1}^{N}$, where $\boldsymbol{\nu}_i \in \mathbb{R}^d$ denotes the i^{th} utterance and y_i denotes the corresponding smoking habits. The goal is to approximate a classifier function g, such that for an utterance of an unseen speaker, $\boldsymbol{\nu}_{\text{tst}}$, the probability of the estimated output classified in the correct class get maximum. That is, the estimated label, $\hat{y} = g(\boldsymbol{\nu}_{\text{tst}})$, is as close as the true label.

2.2 Utterance Modeling

First, we convert variable-duration speech signals into fixed-dimensional vectors, which is performed by fitting a GMM to acoustic features extracted from each speech signal. The parameters of the obtained GMMs characterize the corresponding utterance. Due to lack of data, fitting a separate GMM for a short utterance can not be performed accurately, specially in the case of GMMs with a high number of Gaussians. Therefore, parametric utterance adaptation methods are usually applied to adapt a universal background model (UBM) to characteristics of utterances in training and testing databases. In this paper, i-vector framework for adapting UBM means and NFA framework for adapting UBM weights are applied.

Universal Background Model and Adaptation: Consider a UBM with the following likelihood function of data $\boldsymbol{\mathcal{X}} = \{\mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_\tau\}$.

$$p(\mathbf{x}_t|\lambda) = \sum_{c=1}^{C} b_c p(\mathbf{x}_t|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$$
$$\lambda = \{b_c, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c\},\ c = 1, \ldots C, \tag{1}$$

where $\mathbf{x}_t$ is the acoustic vector at time t, b_c is the mixture weight for the c^{th} mixture component, $p(\mathbf{x}_t|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ is a Gaussian probability density function with mean $\boldsymbol{\mu}_c$ and covariance matrix $\boldsymbol{\Sigma}_c$, and C is the total number of Gaussians in the mixture. The parameters of the UBM –λ– are estimated on a large amount of training data from smoking and non-smoking speakers.

i-vector Framework: One effective method for speaker age estimation involves adapting UBM means to the speech characteristics of the utterance. Then the adapted GMM means are extracted and concatenated to form Gaussain mean supervectors. This method have been shown to provide a good level of performance [9]. Recent progress in this field, however, has found an alternate method of modeling the class dependent GMM mean supervectors that provides superior recognition performance [3]. This technique referred to as total variability modeling [10], assumes the GMM mean supervector, $\mathbf{M}$, can be decomposed as:

$$\mathbf{M} = \mathbf{u} + \mathbf{T}\mathbf{v}, \tag{2}$$

where $\mathbf{u}$ is the mean supervector of the UBM, $\mathbf{T}$ spans a low-dimensional subspace (400 dimensions in this work) and $\mathbf{v}$ are the factors that best describe the utterance-dependent mean offset $\mathbf{T}\mathbf{v}$. The vector $\mathbf{v}$ is treated as a latent variable with the standard normal prior and i-vector is its maximum-a-posteriori (MAP) point estimate. The subspace matrix $\mathbf{T}$ is estimated via maximum likelihood in a large training dataset. An efficient procedure for training $\mathbf{T}$ and for MAP adaptation of i-vectors can be found in [15]. In the total variability modeling approach, i-vectors are the low-dimensional representation of an audio recording that can be used for classification and estimation purposes.

The NFA Framework: The NFA is a new framework for adaptation and decomposition of GMM weights based on a constrained factor analysis [14]. This new low-dimensional utterance representation approach was applied to speaker and language/dialect recognition tasks successfully [13,14]. The basic assumption of this method is that for a given utterance, the adapted GMM weight supervector can be decomposed as:

$$\mathbf{w} = \mathbf{b} + \mathbf{L}\mathbf{r}, \tag{3}$$

where $\mathbf{b}$ is the UBM weight supervector (2048 dimensional vector in this paper). $\mathbf{L}$ is a matrix of dimension $C \times \rho$ spanning a low-dimensional subspace. $\mathbf{r}$ is a low-dimensional vector that best describes the utterance-dependent weight offset $\mathbf{L}\mathbf{r}$. In this framework, neither subspace matrix $\mathbf{L}$ nor subspace vector $\mathbf{r}$ are constrained to be non-negative. However, unlike the i-vector framework, the applied factor analysis for estimating the subspace matrix $\mathbf{L}$ and the subspace vector $\mathbf{r}$ is constrained such that the adapted GMM weights are non-negative and sum up to one. The procedure of calculating $\mathbf{L}$ and $\mathbf{r}$ involves a two-stage algorithm similar to EM. In the first stage, $\mathbf{L}$ is assumed to be known, and we try to update $\mathbf{r}$. Similarly in the second stage, $\mathbf{r}$ is assumed to be known and we try to update $\mathbf{L}$. The subspace matrix $\mathbf{L}$ is estimated over a large training dataset and is used to extract a subspace vector $\mathbf{r}$ for each utterance in train and test datasets. The obtained subspace vectors representing the utterances in train and test datasets are used to estimate the smoking habits of speakers in this paper.

2.3 Classifiers

Logistic Regression (LR): Logistic regression (LR) is a widely used classification method [16], which assumes that

$$y_i \sim Bernoulli(f(\boldsymbol{\theta}^{\top} \boldsymbol{\nu}_i + \theta_0)) \tag{4}$$

where $^{\top}$ represents a transpose, y_is are independent, $\boldsymbol{\theta}$ is a vector with the same dimension of $\boldsymbol{\nu}$, θ_0 is a constant and $f(\cdot)$ is a logistic function and defined as:

$$f(\cdot) = \frac{1}{1 + e^{-(\cdot)}} \tag{5}$$

The output of the logistic function, is a value between zero and one. In the problem of smoker detection, we intend to model the probability of a smoker speaker given his/her speech. That is, $P(Smoker|\boldsymbol{\nu}_i) = f(\boldsymbol{\theta}^{\top} \boldsymbol{\nu}_i + \theta_0)$, where $\boldsymbol{\nu}_i$ is the feature vector corresponding to the i^{th} utterance. Vector $\boldsymbol{\theta}$ and constant θ_0 are the model parameters, which are found through the maximum likelihood estimation (MLE).

Naive Bayesian Classifier (NBC): Bayesian classifiers are probabilistic classifiers working based on Bayes' theorem and the maximum posteriori hypothesis. They predict class membership probabilities, i.e., the probability that a given test sample belongs to a particular class. The Naive Bayesian classifier (NBC) is a special case of Bayesian classifiers, which assumes class conditional independence to decrease the computational cost and training data requirement [17]. In this paper, class distributions are assumed to be Gaussian.

Gaussian Scoring (GS): This classification approach, labeled as GS in this paper, assumes that each category has a Gaussian distribution and full covariance matrix is shared across all categories [18]. In this method, GS score of the test vector $\boldsymbol{\nu}_{\text{test}}$ for the l^{th} class is calculated as:

$$s_l = \boldsymbol{\nu}_{\text{test}}^{\top} \mathit{\Psi}^{-1} \bar{\boldsymbol{\nu}}_l - \frac{1}{2} \bar{\boldsymbol{\nu}}_l^{\top} \boldsymbol{\Psi}^{-1} \bar{\boldsymbol{\nu}}_l, \tag{6}$$

where $\bar{\boldsymbol{\nu}}_l$ is the mean of the vectors for the l^{th} class in the training dataset and $\boldsymbol{\Psi}$ is the common covariance matrix shared across all categories.

Von-Mises-Fisher Scoring (VMF): This classification approach, labeled as VMF in this paper, works based on simplified VMF distribution [19]. In this method, VMF score of the test vector $\boldsymbol{\nu}_{\text{test}}$ for the l^{th} class is calculated as:

$$s_l = \boldsymbol{\nu}_{\text{test}}^{\top} \bar{\boldsymbol{\nu}}_l, \tag{7}$$

2.4 Training and Testing

The proposed smoking habit detection approach is depicted in Fig. 1. During the training phase, each utterance is mapped onto a high dimensional vector using one of the mentioned utterance modeling approaches described in Sect. 2.2. The obtained vectors of the training set are then used as features with their corresponding smoking habit labels to train a classifier. During the testing phase, the utterance modeling approaches are applied to extract high dimensional vectors from an unseen test utterance and the smoking habit is recognized using the trained classifier.

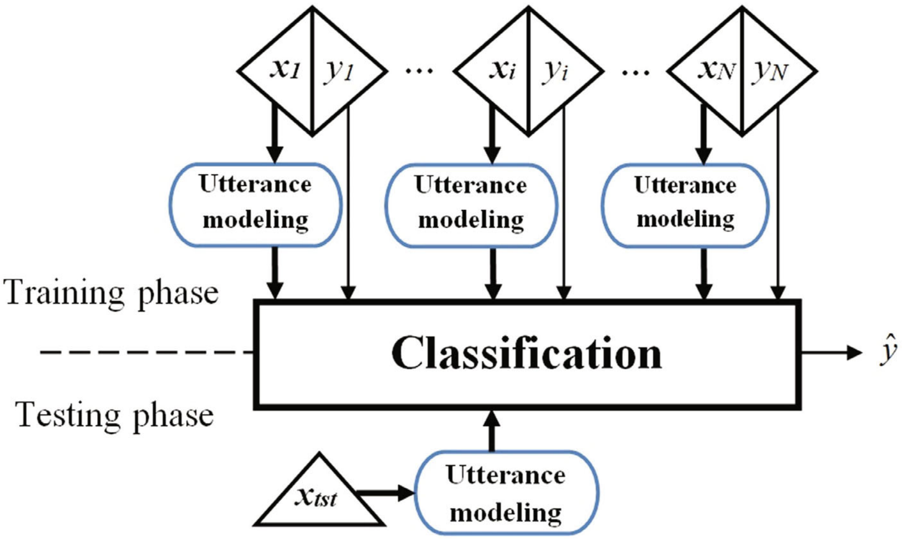

Fig. 1. Block-diagram of the proposed smoker detection approach in training and testing phases

3 Experimental Setup

3.1 Database

The National Institute for Standard and Technology (NIST) have held annual or biannual speaker recognition evaluations (SRE) for the past two decades. With each SRE, a large corpus of telephone (and more recently microphone) conversations are released along with an evaluation protocol. These conversations typically last 5 min and originate from a large number of participants for whom additional meta data is recorded including age, height, language and smoking habits. The NIST databases were chosen for this work due to the large number of speakers and because the total variability subspace requires a considerable amount of development data for training. The development data set used to train the total variability subspace and UBM includes more than 30,000 speech recordings and is sourced from the NIST 2004–2006 SRE databases, LDC releases of Switchboard 2 phase III and Switchboard Cellular (parts 1 and 2). For the purpose of smoker detection, telephone recordings from the common protocols of the recent NIST 2008 and 2010 SRE databases are used. Speakers of NIST 2008 and 2010 SRE databases are divided into three disjoint parts such that 60%, 20% and 20% of all speakers are used for training, development and testing, respectively. The smoking habits histogram of male and female utterances (there might be multiple utterances from each speaker) of training, development and testing databases are depicted in Fig. 2. As depicted in the figure, the problem is dealing with an unbalanced datasets which can make the problem of classification more difficult. The effect of unbalancing in the database can be slightly alleviated by considering the distribution of each class of the training set into consideration during the training phase.

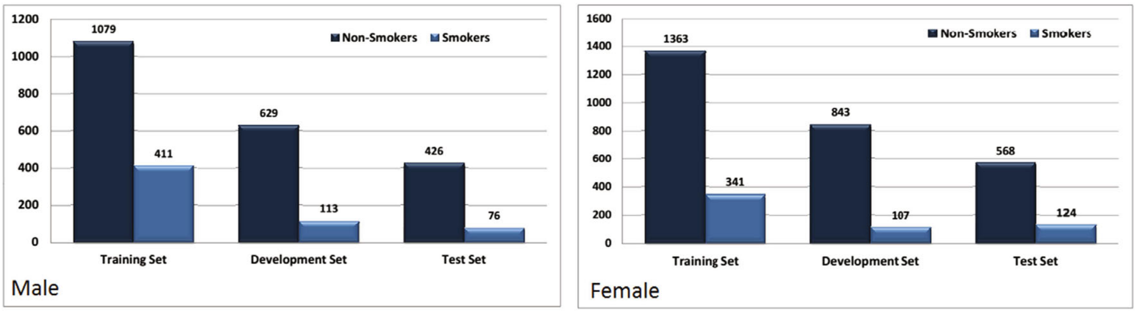

Fig. 2. The smoking habit histograms of the male and female speakers in training, development and test datasets

3.2 Performance Metric

Two performance metrics, namely minimum log-likelihood-ratio cost $C_{\text{llr,min}}$ and area under the receiver operating characteristic curve are considered. In this section, the applied performance measure methods are described briefly.

Log-Likelihood Ratio Cost: Log-Likelihood Ratio Cost ($C_{\text{llr,min}}$) is a performance measure for classifiers with soft, probabilistic decisions output in the form of log-likelihood-ratios. This performance measure is an application-independent since it is independent of the prior distribution of the classes [20]. $C_{\text{llr,min}}$ represents the minimum possible C_{llr} which can be achieved for an optimally calibrated system [20]. In this study, in order to calculate $C_{\text{llr,min}}$, the FoCal Multiclass Toolkit [21] is utilized.

Area Under the ROC Curve (AURC): Receiver operating characteristic (ROC) curve is a widely used approach to measure the efficiency of classifiers. In a ROC curve the true positive rate (sensitivity) is plotted versus the false positive rate (1-specificity) for different operating points. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A classifier with perfect discrimination has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test [22]. Therefore, classifiers can be evaluated by comparing their area under the ROC curves (AURCs). The AURC takes a value between 0 and 1. This value for a perfect classifier is 1, and for a useless classifier, which its posterior is equal to its prior, is 0.5.

4 Results and Discussion

This section presents the results of the proposed smoking habit detection approach. The acoustic feature consists of 20 Mel-frequency cepstrum coefficients (MFCCs) [23] including energy appended with their first and second order derivatives, forming a 60 dimensional acoustic feature vector. This type of feature is very common in state-of-the-art i-vector based speaker and language recognition systems [24]. To have more reliable features, Wiener filtering, speech activity detection [25] and feature warping [26] have been considered in front-end processing. The obtained $C_{\text{llr,min}}$ and AURC of applying different classifiers over the i-vector based and the NFA based classifiers are reported in Table 1.

Table 1. The $C_{\text{llr,min}}$ and AURC of applying different classifiers over the i-vector and NFA frameworks

Utterance modeling	$C_{\text{llr,min}}$					AURC				
	LR	VMF	GS	NN	NBC	LR	VMF	GS	NN	NBC
i-vector	0.86	0.90	0.98	0.90	0.93	0.74	0.51	0.56	0.70	0.66
NFA-vector	0.90	0.91	0.97	0.93	0.98	0.68	0.65	0.59	0.66	0.56

We can observe that LR yields more accurate results compared to other applied classifiers. Thus, this classifier is used in the rest of experiments in this paper. It is also shown that i-vector framework, which works based on Gaussian means, is more accurate than NFA framework working based on Gaussian weights. Different studies show that GMM weights, which entail a lower dimension compared to Gaussian mean supervectors, carry less, yet complimentary, information to GMM means [5,27]. For example, Zang et al. applied GMM weight adaptation in conjunction with mean adaptation for a large vocabulary speech recognition system to improve the word error rate [27]. In [5], a feature-level fusion of i-vectors, GMM mean supervectors, and GMM weight supervectors is applied to improve the accuracy of accent recognition. To enhance the smoking habit detection accuracy we apply a score-level fusion of the i-vector and the NFA classifiers. The fusion is performed by training a logistic regression on the outputs of the classifiers using the development data. The $C_{\text{llr,min}}$ and AURC of obtained results after fusion are 0.845 and 0.754, respectively. The relative improvements of $C_{\text{llr,min}}$ obtained by the proposed fusion scheme compared to

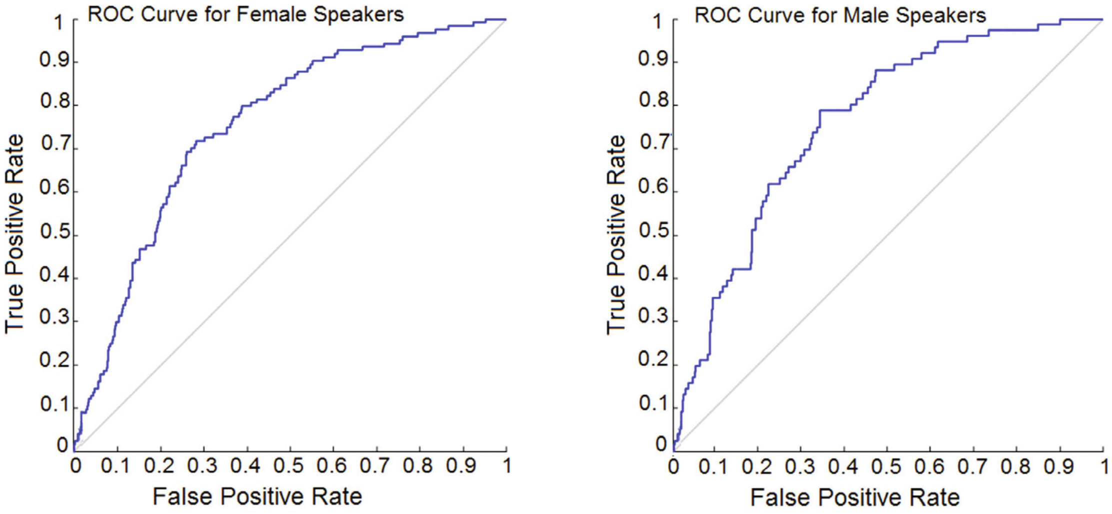

Fig. 3. The ROC curve of the proposed method for female and male speakers

the i-vector and the NFA frameworks are 1.8% and 6.5%, respectively. The relative improvements of AURC after fusion compared to the i-vector-based and the NFA-based systems are 1.9% and 11%, respectively. The ROC curves of the proposed fusion for male and female speakers are illustrated in Fig. 3.

5 Conclusions

In this paper, we proposed a new approach for automatic smoking habit detection from telephone speech signals. In this method, utterances were modeled using the i-vector and the NFA frameworks, which are based on the factor analysis on GMM means and weights, respectively. Then, several classifier were employed to discriminate smokers and non-smokers. To improve the performance, the score-level fusion of the i-vector-based and the NFA-based systems was considered. The proposed method was evaluated on telephone speech signals of NIST 2008 and 2010 SRE databases. Experimental results over 1194 utterances demonstrated the effectiveness of the proposed approach in automatic smoker detection.

References

1. Poorjam, A.H., Bahari, M.H., van Hamme, H.: A novel approach to speaker weight estimation using a fusion of the i-vector and NFA frameworks. J. Electr. Syst. Signals **3**(1), 47–55 (2017)
2. Poorjam, A.H., Bahari, M.H., van Hamme, H.: Multitask speaker profiling for estimating age, height, weight and smoking habits from spontaneous telephone speech signals. In: Proceedings of International Conference on Computer and Knowledge Engineering, pp. 7–12 (2014)
3. Bahari, M.H., van Hamme, H.: Speaker age estimation using hidden Markov model weight supervectors. In: Proceedings of 11th International Conference on Information Science, Signal Processing and their Applications (2012)

4. Poorjam, A.H., Bahari, M.H., Vasilakakis, V., van Hamme, H.: Height estimation from speech signals using i-vectors and least-squares support vector regression. In: Proceedings of International Conference on Telecommunications and Signal Processing (2015)
5. Bahari, M.H., Saeidi, R., van Hamme, H., van Leeuwen, D.: Accent recognition using i-vector, gaussian mean supervector and gaussian posterior probability supervector for spontaneous telephone speech proceedings. In: Proceedings of ICASSP, pp. 7344–7348 (2013)
6. Sorensen, D., Yoshiyuki, H.: Cigarette smoking and voice fundamental frequency. J. Commun. Disord. **15**(2), 135–44 (1982)
7. Gonzalez, J., Carpi, A.: Early effects of smoking on the voice: a multidimensional study. Med. Sci. Monit. **10**(12), 49–56 (2004)
8. Campbell, W., Sturim, D., Reynolds, D.: Support vector machines using GMM supervectors for speaker verification. IEEE Signal Process. Lett. **13**(5), 308–311 (2006)
9. Dobry, G., Hecht, R.M., Avigal, M., Zigel, Y.: Supervector dimension reduction for efficient speaker age estimation based on the acoustic speech signal. IEEE Trans. Audio Speech Lang. Process. **19**(7), 75–85 (2011)
10. Dehak, N., Kenny, P., Dehak, D., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. IEEE Trans. Audio Speech Lang. Process. **19**(4), 788–798 (2011)
11. Poorjam, A.H., Saeidi, R., Kinnunen, T., Hautamäki, V.: Incorporating uncertainty as a quality measure in i-vector based language recognition. In: Proceedings of Odyssey, pp. 74–80 (2016)
12. Bahari, M.H., McLaren, M., van Hamme, H., van Leeuwen, D.A.: Age estimation from telephone Speech using i-vectors. In: Proceedings of Interspeech (2012)
13. Bahari, M.H., Dehak, N., van Hamme, H., Burget, L., Ali, A.M., Glass, J.: Non-negative factor analysis of Gaussian mixture model weight adaptation for language and dialect recognition. IEEE/ACM Trans. Audio Speech Lang. Process. **22**(7), 1117–1129 (2014)
14. Bahari, M.H., Dehak, N., van Hamme, H.: Gaussian mixture model weight supervector decomposition and adaptation. Technical report. Computer Science and Artificial Intelligence Laboratory (CSAIL) of MIT (2013)
15. Kenny, P., Ouellet, P., Dehak, N., Gupta, V., Dumouchel, P.: A study of inter-speaker variability in speaker verification. IEEE Trans. Audio Speech Lang. Process. **16**(5), 980–88 (2008)
16. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
17. Yager, R.: An extension of the naive Bayesian classifier. Inf. Sci. **176**(5), 577–588 (2006)
18. Martinez, D., Plchot, O., Burget, L., Glembek, O., Matejka, P.: Language recognition in i-vectors space. In: Proceedings of Interspeech, pp. 861–864 (2011)
19. Singer, E., Torres-Carrasquillo, P., Reynolds, D., McCree, A., Richardson, F., Dehak, N., Sturim, D.: The mitll NIST LRE 2011 language recognition system. In: Proceedings of Odyssey, pp. 209–215 (2012)
20. Brummer, N., van Leeuwen, D.A.: On calibration of language recognition scores. In: Proceedings of Odyssey (2006)
21. Brummer, N.: FoCal multi-class: toolkit for evaluation, fusion and calibration of multi-class recognition scores (2007)
22. Zweig, M., Campbell, G.: Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. **39**, 561–577 (1993)

23. Deller, J.R., Hansen, J.H.L., Proakis, J.G.: Discrete-Time Processing of Speech Signals, 2nd edn. IEEE Press, New York (2000)
24. Lee, K.A., et al.: The 2015 NIST language recognition evaluation : the shared view of I2R, Fantastic4 and SingaMS. In: Proceedings of Interspeech, pp. 3211–3215 (2016)
25. McLaren, M., van Leeuwen, D.: A simple and effective speech activity detection algorithm for telephone and microphone speech. In: Proceedings of NIST SRE Workshop
26. Pelecanos, J., Sridharan, S.: Feature warping for robust speaker verification. In: Proceedings of Odyssey, pp. 213–218 (2001)
27. Zhang, X., Demuynck, K., van Hamme, H.: Rapid speaker adaptation in latent speaker space with non-negative matrix factorization. Speech Commun. **55**, 893–908 (2013)