PURPOSE-LED
PUBLISHING™

PAPER • OPEN ACCESS

Lie Detection using Speech Processing Techniques

View the article online for updates and enhancements.

Lie Detection using Speech Processing Techniques

E P Fathima Bareeda[1], B S Shajee Mohan[2] and K V Ahammed Muneer[3]

[1] M.Tech Signal Processing, Government Engineering College, Kozhikode.
[2] Department of Applied Electronics and Instrumentation Engineering, Government Engineering College, Kozhikode.
[3] Department of Applied Electronics and Instrumentation Engineering, Government Engineering College, Kozhikode.

E-mail: `fathimabareeda@gmail.com, smohan@geckkd.ac.in, ahammedcet@gmail.com`

Abstract. The objective of this paper is to design and set up an efficient lie detector that can detect lies in spoken sentences in an object language. Usually lie detection make use of invasive methods that extract different physiological signals such as EEG *etc.,* from the subject's body and analyze them to decide on the presence of lie in the utterances. This work proposes an efficient lie detection technique based on the non-invasive method that involves only recording the subject's speech utterances. Discriminative and meaningful features are extracted from the speech and classifiers are built based on SVM to discriminate between truth and lie. The classifier will be trained efficiently so that a better performance can be obtained compared with existing lie detectors. This paper aims to exploit the psycho-neural aspects and dependence of speech signals to predict and detect the presence of lies in isolated speech utterances.

1. Introduction

Human beings are the only species having a lot of virtual masks in their behavior. The tone of a person's voice has an amazing effect to express feeling such as emotions, anger, sadness, *etc.,* Different aspects of a person's social position including physical states, psychic states, physical-psychic states, experiences are unintentionally conveyed through the speech. Speech is a powerful tool to measure psycho-neural changes within the human being.

Lying could be a defense reaction to avoid hassle with the law, bosses, or authority figures. The occurrence of deception is common in daily life. Hiding information or misrepresenting it sometimes leads to huge problems. The importance of lie detection is increasing day by day. Decades of research proved that three principal ways of deception detection: physiological responses, behavioral analysis, and verbal analysis. Physiological responses are based on the assumption that when an individual experiences apprehension, fear, anxiety, *etc.,* the physiological parameters like blood pressure, galvanic skin resistance will increase considerably. This approach consists of mainly three techniques: Polygraph [1], interview, and guilty knowledge test [2]. On the other hand, behavior analysis [3] is based on the analysis of faces, facial expressions, *etc.,* But these two approaches require trained experts and it is a time-consuming process. Later, researchers focused their study on the verbal analysis of speech [1]. This approach has the advantage that it doesn't require direct contact between the subjects and the acquisition system. Reality monitoring, scientific content analysis, and criteria-based content analysis [1] are the three techniques in these approaches. Recently, computer-based detection of deception gained popularity, because of its capability to analyze complex datasets. The main problem with computer-based methods is that they require distinctive features to classify truth or lie.

This work proposes a novel approach for lie detection using only speech signals. This method particularly deals with extracting discriminative features such as MFCC, that play a crucial role in detecting deception. From the speech signal, MFCC features are extracted, and the mean of the MFCC of the entire frames of a complete speech utterance is used to represent the true or lie utterance. Finally, SVM based classifier is employed to classify truth or deceptive voices. The audio sequence from the dataset called "Real life trial data" [4] that contains 61 deceptive and 60 truthful video clips are used in this study. Here, the lie detection is performed by using the speech signal, so speech signals are extracted from the video clips. Truth versus lie classification is performed by training the binary classifier called SVM using the proposed features. After several rounds of experimentations, the obtained result shows that the proposed method can achieve an overall accuracy of 81%.

The remaining part of the paper is structured as follows: the overview of proposed lie detection is described in Section 2, while dataset description is in Section 3, the experimental details are given in Section 4, and the proposed modification using GMM based method is explained in Section 5, followed by conclusions.

2. Overview of proposed lie detection system

The internal state of individuals is expressed through the voice. This technique proposes an efficient lie detection technique based on non-invasive methods that involve only recording the subject's speech utterances. Here, machine learning-based approaches are used to assess the performance of extracted features from the real-life dataset, that includes truthful and deceptive speeches. These speeches are extracted from the video clips of the dataset. Lie detection is performed by using the SVM based classifier.
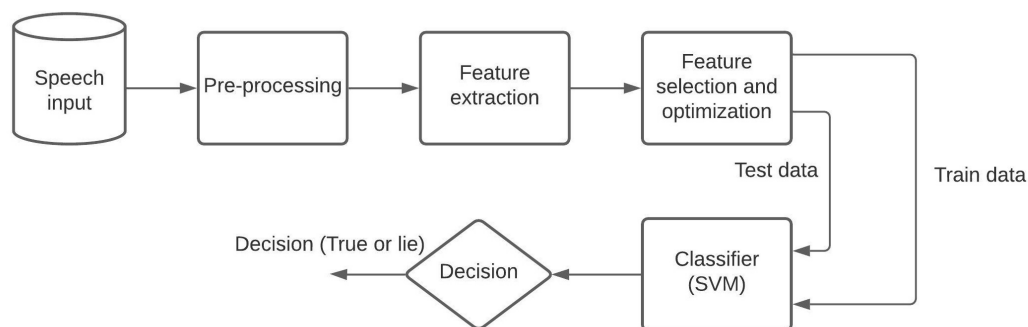
Figure 1. Block diagram of the proposed lie detection technique

For the speech classification, the first step in the proposed method (Fig. 1) is speech acquisition. Pre-processing of the acquired speech is done to get a fine-grained form of speech signals to ensure the accurate prediction of the lie. Noise is removed through the STFT based thresholding technique in this stage. Features are extracted from the pre-processed signals. Here, an inspection of relevant features is done by choosing MFCCs. Then the features are given to classifiers for better discrimination of lie from the truth and based on this a decision is made.

2.1. Pre-processing

Pre-processing is a crucial part of preparing data to achieve high accuracy and efficiency in the subsequent classification. Audio signals are cleaned in this phase to remove the background noises, silent portion, and other irrelevant information using the STFT based thresholding method.

Figure 2. Block diagram of noise removal using STFT based thresholding method

2.1.1.   STFT based noise reduction Noise reduction can be achieved by using STFT based thresholding method. It consists of mainly three steps (Fig. 2):

(i) Compute the STFT of noisy signal

$$S(m,w) = STFTx(n) \tag{1}$$

(ii) Threshold the STFT

$$S2(m,w) = g(S(m,w)) \tag{2}$$

where *g(a)* is a thresholding function

(iii) Compute the inverse STFT

$$y(n) = STFT^{-1}S2(m,w) \tag{3}$$

Here, the thresholding operation is non-linear. Hence, STFT-thresholding is a non-linear noise reduction technique.

2.2. Feature extraction

Feature extraction refers to the process of transforming a large set of data into numerical features while still preserving information in the original data with sufficient accuracy. High computational power and a large amount of memory are required for analyzing raw data. Feature extraction helps to reduce the number of attributes to describe the original data. The features that are used for classifying lie consists of MFCC (Mel frequency cepstral coefficients), mean of MFCC of all frames.

2.2.1.   Mel frequency cepstral coefficients (MFCC) Mel frequency cepstral coefficients [5] are the widely used representation in various speech processing applications because it represents the original signal in a better way. To compute MFCC (Fig. 3) from a given audio signal, the steps are:
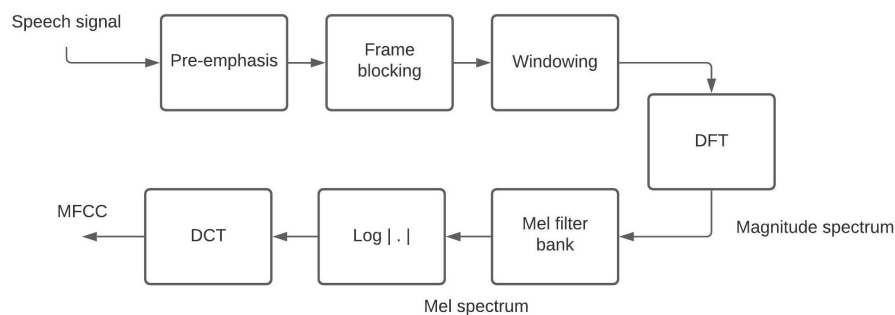
Figure 3. Block diagram of MFCC extraction process

(i) Pre-emphasis: This step boosts the amount of energy in high frequencies by using filters. The main purpose of the pre-emphasis filters is to balance the speech signal frequency. The input/output relationship of the pre-emphasis filter in the time domain is given by:

$$y_1(n) = x_1(n) - 0.95\, x_1(n-1) \tag{4}$$

(ii) Frame blocking: In these stages, speech samples are blocked into small frames. Usually, the length of each frame is in the range of 20 to 40 msec (25msec is chosen for the proposed framework). *i.e.,* speech signal is split into N frames of M samples.

(iii) Windowing: Long sound signals are analyzed in these stages by taking a sufficiently representative section. The main purpose of this stage is to reduce the edge effect, smooth the edges, and to enhance the harmonics while taking the DFT on the signal. Usually, the Hamming window is used. If the window $w_1(n)$ is defined for $0 \le n \le M\text{-}1$:

$$y_1(n) = x_1(n) * w_1(n) \tag{5}$$

for Hamming window, $w_1(n)$ is given by:

$$w_1(n) = 0.54 - 0.46\, cos(\frac{2\pi n}{M-1}) \quad (0 \le n \le (M-1)) \tag{6}$$

(iv) DFT spectrum: This stage converts each windowed frame to a magnitude spectrum using DFT.

$$X(k) = \sum_{i=0}^{N-1} x(i) e^{\frac{-j2\pi ik}{N}} \quad 0 \le k \le N-1 \tag{7}$$

(v) Mel spectrum: The Fourier transformed signal is squared to get the DFT power spectrum and it is passed through the band-pass filters called Mel-filter bank to obtain mel spectrum.

$$f_{mel} = 2595\, log_{10}\, (1 + \frac{f}{700}) \tag{8}$$

$$y(n) = \sum_{i=0}^{N-1} [|X(i)|^2 W_j(i)] \quad 0 \le j \le J-1 \tag{9}$$

where j is the total number of mel weighing filters

(vi) Log: In these phases, take the log of power at each of mel frequencies.

(vii) DCT: Log mel spectrum is converted into the time domain by this stage. After converting it to the time domain, MFCCs (Fig. 4) are obtained. This set of coefficients are called acoustic vectors.
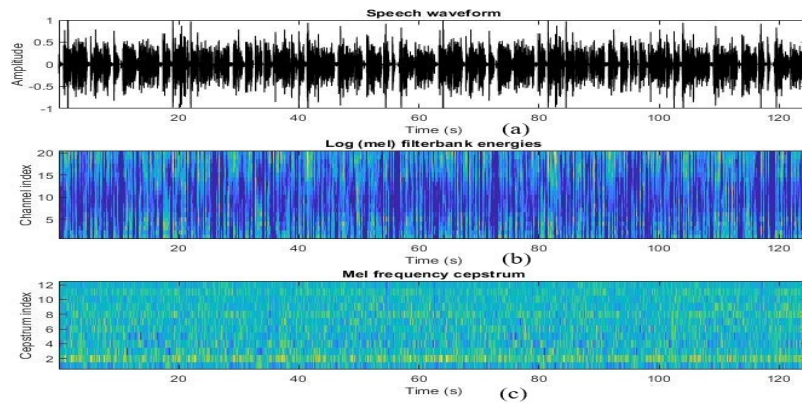
Figure 4. Plots for filter bank energies (b) and mel frequency cepstrum (c) for a speech utterence in (a)

2.3. Feature selection and optimization

Feature selection (variable subset selection) involves choosing a subset of the relevant features (variables) to construct the model. It reduces the number of features for minimizing the computational complexity and cost. The pros of feature selection are:

- Lesser training times are required

- Curse of dimensionality can be avoided

- Overfitting can be reduced

Feature selection can be done by using principal component analysis (PCA).

2.3.1. Principal component analysis

The objective of principal component analysis (PCA) is to reduce the dimensionality of the feature vectors comprising of a large number of uncorrelated variables, while maintaining most of the variations in the features as much as possible. This is done by transforming into a new variables set called principal components. Principal components preserve most of the data variations that are present in the original features. Eigen vectors of covariance matrices are called principal components. Thus, the eigen decomposition of the data covariance matrix is used to compute principal components.

2.4. Classifiers

Classification of speech into true and the lie is performed using SVM classifiers. It is a supervised learning technique mainly used for regression and classification problems. The main aim of SVM is to find an optimal hyperplane that separates truth and lie classes. The kernels used for the classification problem are Gaussian, and polynomial kernels as shown in Table 1. Finally, a decision is made based on the classification results.

Table 1. Kernel functions used for SVM classifier

Kernels	Kernel functions
Gaussian kernel	$exp(-\frac{1}{2\sigma^2}\lVert\mathbf{x}-\mathbf{x}_i\rVert^2)$
Polynomial kernel	$(\mathbf{x}^T\mathbf{x}_i + 1)^d$

3. Dataset

In this work, the dataset called "Real life trial data" [4] is considered. It was created by Rada Mihalcea, Veronica Perez-rosas, and friends. The dataset consists of videos collected from public court trial, where the truthful and deceptive statements are easily observed and verified. In most of the clip duration, the defendant or witness in the video should be visible. Also, the audio quality is adequate to hear and comprehend what the person is stating. Defendant and witnesses are taken into account for labeling the videos. The dataset comprised 121 videos of which 61 deceptive and 60 truth video clips. For deceptive and truthful clips, the average video length is 27.7 seconds and 28.3 seconds respectively. Dataset consists of 21 unique female and 35 unique male speakers. The ages of speakers are in the range of 16 to 60 years. Video files converted into audio files. With an average of 66 words per transcriptions, 8,055 words are involved in the final set of transcriptions.

4. Experimental setup and results

4.1. Feature description

Lie classification was carried out by using LibSVM classifier [6] and by using two different representations as shown in Table 2. In the first method, concatenated MFCC features of all frames, and in the second method, mean MFCC values of all frames in an utterance are considered. For the MFCC calculation, the signals are framed as short frames of 25msec and for each frame 13 MFCC coefficients are computed. The number of frames for each audio signal depends on the length of the audio clips. Later, the MFCC of all the truth samples and lie samples are concatenated separately. Finally, the mean of MFCC was taken from concatenated truth samples and also from lie samples.

Table 2. Feature Description

	Feature vector size	
	Truth	Lie
MFCC	13×3960	13×3390
Concatenated MFCC	13×334026	13×330445
Mean of MFCC	60×13	61×13

4.2. Classifier implementation

For data classification, SVMs are useful techniques. Implementation of the lie classification system was done by using LibSVM [6]. Training and testing sets are usually involved in classification tasks. The training set consists of examples and their label information. Based on the training data, SVM produces a model. The target value of the test data is predicted by the model given only the attributes of test data.

Given a training set $(\mathbf{z}_i, y_i)$, $i = 1,2$ . . . , N where $\mathbf{z}_i \in \mathbb{R}^N$ and $y \in (1, -1)^N$, the SVM find the optimal values of weight vector $\mathbf{w}$ and b by solving the following objective function:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w^T}\mathbf{w} + C\sum_{i=1}^{N} \xi_i \tag{10}$$

subject to the constraints:

$$y_i(\mathbf{w}^T\mathbf{z}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{11}$$

The transformation of input data $\mathbf{x}_i$ from data space of a lower dimension to feature space of higher dimensional space is done by using a non-linear transformation ϕ. In feature space, SVM finds a linearly seperable hyperplane that maximize the margin. C is a user-defined positive parameter and ξ_is are the slack variables. Furthermore, kernel function is defined as $\mathrm{K}(\mathbf{z}_i,\mathbf{z}_j) = \mathbf{z}_i^T \mathbf{z}_j$.

4.3. Performance analysis

Table 3 presents the performance of the standard SVM classifier using MFCC features (for polynomial and Gaussian kernels) for the Real-life trial data. The performance is observed to be better for the polynomial kernel of degree 3. The performance drops for Gaussian kernel with $\sigma = 0.01$. On the other hand, almost the same performance is noted for the polynomial kernel with degree 2 and Gaussian kernel with $\sigma = 0.02$.

Table 3. Performance of true, lie classification system using SVM (with polynomial and Gaussian kernels)

Kernel	Kernel parameter	Classification accuracy
Gaussian	$\sigma = 0.01$	75.000000 ± 8.316526
	$\sigma = 0.02$	78.703704 ± 6.543408
Polynomial	Degree $=2$	78.703704 ± 1.814815
	Degree $= 3$	81.481481 ± 4.801549

5. Proposed modification using GMM based method

The present work can be modified by using GMM based method for true-lie classification (Fig. 5). Two classes of speech signals exhibit different probability density distributions. This difference in the two categories of speech signals motivated to extends the proposed method in the following direction. Initially, modeling of true and lie classes using the optimum number of GMM components. The parameters comprising mean, variance, and weight of the GMM components will be used for the representations of true and lie in the classification process. *i.e.,* feature extraction involves estimation of separate GMMs for true and lie classes. Using this learnt GMM, the likelihood probability of any test data can be determined.
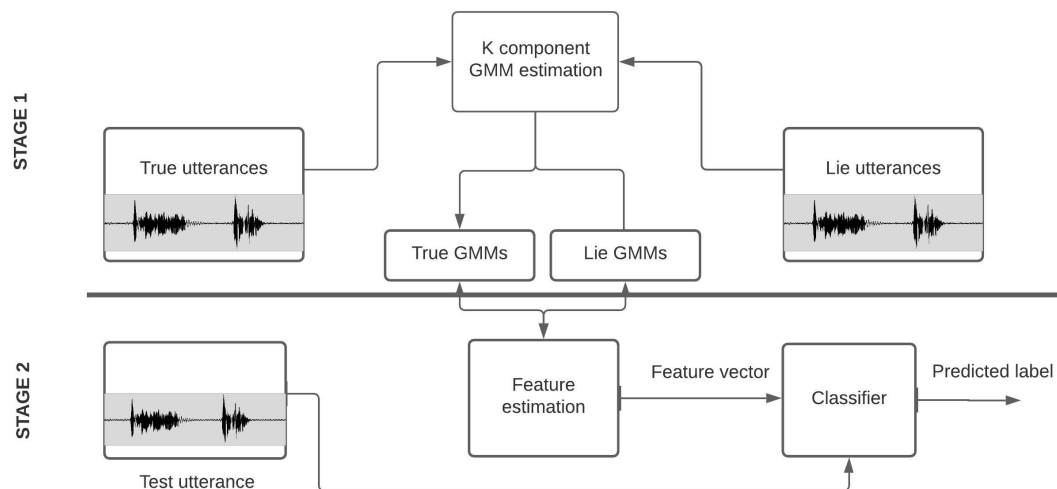
Figure 5. Block diagram of the proposed GMM based method

6. Conclusions

This work proposes a novel true-lie detection technique for isolated speech utterances. Related works are focused on analyzing psychological and behavioral measures for detecting deception. This work exploits the fact that psycho-neural changes are highly dependent on speech signals. MFCCs and mean of MFCCs are used to classify truth or lie with the help of SVM based classification algorithm. The system performance was validated for the dataset called "Real life trial data" that contains 61 lie samples and 60 truth samples. Experiments show that overall classification accuracy of 81% and 78% are achieved for lie and truth classes using polynomial and Gaussian kernels respectively.

References

[1] Cardona P A N 2015 A compendium of pattern recognition techniques in face, speech and lie detection *IJRRAS* **24** pp 529-51

[2] Ben-Shakhar G and Elaad E 2002 *The Guilty Knowledge Test (GKT) as an Application of Psychophysiology: Future Prospects and Obstacles* Handbook of polygraph testing pp 87- 102

[3] Vrij A, Edward K, Kim P Roberts and Ray Bull 2000 Detecting deceit via analysis of verbal and nonverbal behavior *J. of Nonverbal behavior* **24(4)** pp 239-63

[4] Veronica Perez-rosas, Mohamed Abouelenien, *et al.,* 2015 Deception Detection using Real-life Trial Data *ICMI'15: Proc. of the 2015 ACM on Int. Conf. on Multimodal Interaction* pp 59–66

[5] Wael Ouarda, Adel M Alimi, *et al.,* 2016 ReLiDSS: Novel lie detection system from speech signal *IEEE/ACS 13th Int. Conf. of Computer Systems and Applications (AICCSA)* (Agadir, Morocco)

[6] Chang C and Lin C J 2011 LIBSVM: a library for support vector machines *ACM Transactions on Intelligent Systems and Technology* **2(3)** pp 27:1–27:27

[7] Sami Ekic, Turgut Kavas, *et al.,* 2017 Deception Detection From Speech Signals *Int. Conf. on Advances and Innovations in Engineering (ICAIE)* (Elazig, Turkey) pp 1016-21

[8] Faryal Amber, Adeel Yousaf, *et al.,* 2019 P300 based deception detection using convolutional neural network in 2019 2nd Int. Conf. on Communication, Computing and Digital Systems (C-CODE) (Islamabad: IEEE) pp 201–4

[9] Mahmoud E Gadallah, Matar A Matar, *et al.,* 1999 Speech Based Automatic Lie Detection *Proc. of the 16th National Radio Science Conference* (Cairo, Egypt) pp C33/1-C33/8

[10] Sondhi S, Vijay R, *et al.,* 2016 Voice analysis for detection of deception *11th Int. Conf. on Knowledge Information and Creativity Support Systems* (Yogyakarta, Indonesia)

[11] Kurniawan H, Maslov A V and *et al.,* 2013 Stress detection from speech and galvanic skin response signals *Proc. of the 26th IEEE Int. Symp. on Computer-Based Medical Systems* (Porto, Portugal) pp 209-14

[12] Veronica Perez-rosas, Mohamed Abouelenien, *et al.,* 2015 Verbal and Nonverbal Clues for Real-life Deception Detection *Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing* pp 2336–46

[13] Sushma Venkatesh, Raghavendra Ramachandra, *et al.,* 2019 Robust Algorithm for Multimodal Deception Detection *2019 IEEE Conf. on Multimedia Information Processing and Retrieval (MIPR)* (San Jose, CA, USA) pp 534–37

[14] Smith G A 1977 Voice analysis for the measurement of anxiety *British J. of Medical Psychology* **50(4)** pp 367-73

[15] Savita Sondhi, Munna Khan, Ritu Vijay and Ashok K. Salhan 2015 Acoustic analysis of speech under stress *Int. J. of Bioinformatics Research and Applications* **11(5)** pp 417–32