ReLiDSS: Novel Lie Detection system from speech signal

Hanen Nasri*, Wael Ouarda and Adel M. Alimi
Research Groups in Intelligent Machines, University of Sfax,
National School of Engineers (ENIS), BP 1173, Sfax, 3038, Tunisia
Email: hanen.nasri.tn@ieee.org, wael.ouarda@ieee.org,adel.alimi@ieee.org

***Abstract*—Lying is among the most common wrong human acts that merits spending time thinking about it. The lie detection is until now posing a problem in recent research which aims to develop a non-contact application in order to estimate physiological changes. In this paper, we have proposed a preliminary investigation on which relevant acoustic parameter can be useful to classify lie or truth from speech signal. Our proposed system in is based on the Mel Frequency Cepstral Coefficient (MFCC) commonly used in automatic speech processing on our own constructed database ReLiDDB (ReGIM-Lab Lie Detection DataBase) for both cases lie detection and person voice recognition. We have performed on this database the Support Vector Machines (SVM) classifier using Linear kernel and we have obtained an accuracy of Lie and Truth detection of speech audio respectively 88.23% and 84.52%.**
Keywords speech signal, lie detection, stress, person recognition, SVM, MFCC, pitch.

I. Introduction

Man is the only species able to generate particular sound signals to form utterances qualified by speech. Speaking is the act of producing sounds consciously in order to be understood, to expressing feeling, listening, sharing transmitting, influencing others, questioning and persuading, etc. The speech signal contains not only linguistic and expressive information but also biological and organic data. Now, it is used to reveal physiological changes behind different behaviors like lie, stress, emotion, pathologies, etc.

A liar is someone who intends to deceive or mislead the other part even by a true information and with not bad objective. Lying is a form of deception which was always classed as a instinct part of human nature. Therefore, thinking about techniques to catch it is not a recent domain. Human voice analysis can be useful to solve the problem of incapacity to insure the statement veracity in meetings, interviews and mobile calls.

In this work we are aiming to propose a new technique for lie detection from speech signal. Our main objective is to find the appropriate features that can characterize lie on human voice. We have started by building a new database named ReLiDDB (ReGIM-Lab Lie Detection DataBase) which contain false and true declarations recorded in indoor and outdoor environment. Our proposed system is based on the Mel Frequency Cepstral Coefficient (MFCC) commonly used in automatic speech processing on our own database ReLiDDB. We have performed on this database the Support Vector Machines (SVM) classifier using Linear kernel.

This paper is organized as follows: We will start with the state of art in order to present an overview of lie detection approaches and stress detectors and the heart rate estimators from voice signal. In the second part we will describe the different phases of our proposed system ReLiDSS and the experimental results. Conclusively, we will cite our work prospects.

II. State of art

A. Lie detection

Decades of research have given rise to three principal ways to discover liars: physiological responses measure, speech analysis and behaviors recognition based on multiple traits like faces, facial expression, etc [1].

1) Physiological approach: Polygraphs are started by physiological indices analysis of several features like: heart rate, respiratory rate, skin conductivity and blood pressure measurement of the subject to detect its deception [2]. This way involves deducing deception by analyzing the physiological responses to a structured set of questions. Several test strategies have been developed like the Control Question Test (CQT) and the Guilty Knowledge Test (GKT) [1].

The CQT technique procedure is formed on a comparison between responses to a set of relevant and controlled questions. These latter, are about similar studied misdeeds. It is supposed that the accused person responds better to controlled questions than to relevant questions, because the first are about an action that they didn't commit. The test is noted inconclusive if there is no difference to be found [3].

GKT test is constituted of multiple-choice questions around information that only a guilty could know. A greater physiological response to the correct choice is regarded as a sign of deception. The disadvantage of this method is that it demands to investigators to have information that only the guilty would know and it can have a false interpretation when no reaction for the correct response is for a lack of knowledge and not for innocence [4].

2) Verbal approach: Later, researchers have based their study of lie detection on the analysis of verbal speech which is more useful because it do not require a direct contact between person and acquisition system such as heart rate, blood pressure, etc. Some works were lunched in this context like Criteria based Content Analysis (CBCA) and Reality Monitoring (RM) [2].

978-1-5090-4320-0/16/$31.00 ©2016 IEEE

The objective of the CBCA is to looking for 19 sincerity criteria in written transcripts of witness interviews to assess the allegations credibility of sexual abuse made by children [1]. RM is based on the personal experience of the narrator put in the test of the human memory and essayed to find difference between external and internal memories [2].

The SCAN technique (Scientific content Analysis) is another verbal veracity assessment method developed in Israel consists of a contextual analysis of the suspect statement. These techniques help to detect lies but not with the assistance of stress, anxiety, sadness, nervousness, fear and shame [2].

However, recent research has demonstrated that the limitation of ancient lie detectors is due to the false hypothesis that specific behavioral indicators appear when people are lying. This finding has generated an alteration in this domain which has adapted two new approaches to judge the veracity: the active interviewer technique and contextual information consideration [5].

3) Active interviewer approach:

Time Restricted Integrity-Confirmation: The absence of an automatic behavioral changes for deception pushed research to seek a process to provoke and increase it. It is the principal idea of active interviewer approach where the interviewer should have an active role to manage the interaction. Therefore, interviewing strategies founded on cognitive psychology and theoretical models of lie formulation were developed.

The Activation Decision-Construction Model (ADCM) is one of these models which was the basis of the development of TRI-Con (Time Restricted Integrity-Confirmation) for deception detection. TRi-Con objective is to prevent the mentally preparation of response and to reduce the cognitive effort to do. It has started with general questions to activate the truth in the interviewee memory and let relevant questions for the last moment. Besides, the questions structure imposes to listen until the last word of the question to understand it and demands two or three words to be answered. Open-ended questions have to be avoided and the interviewee is asked to answer instantaneously to improve the cognitive load measurement. The inconsistencies are revealed by including any related questions examining the same subject.

Response time (in milliseconds), related questions divergence (detected from recorded responses), pupil dilation and eye movements (computed with an eye tracker) are the considered variables identifying the cognitive load. Rapidly answering, less inconsistencies and pupil dilation and more eye motion are signs of sincere truth-telling [5].

Induced Cognitive Load: Induced Cognitive Load is another approach from which the idea was to induce the cognitive load during the interview, in order to make difficult the task of lying and to expose observable signs of cognitive overload. The psychologist Aldert Vrij and colleagues tested it by consulting 80 persons (50% guilties and 50% innocent). The participants had to recite the incident in reverse order which demand a considerable cognitive exertion.

More movements of hand and finger (in the normal order), auditory details, contextual implanting and speech tone (in reverse order) are discriminative indicators for truth-telling.

Contrariwise, hesitating, making errors, blinking eyes and trembling leg or foot are the deceiving signs of liying when they recount the scene in reverse. The problem still to differentiate if the foot movement and eye blinks are caused by cognitive overload or lying action [6].

Strategic Use of Evidence: The two previous techniques are based on the role of the interviewer to activate the behavioral changes while the SUE is established from the idea that deception is more complex than truth-telling. The idea is to exploit cleverly the proofs, during the interview, and consider the consistency in his declaration the sign of his innocence. proposed the following steps for SUE technique interview. First, the interviewer must treat the events' progression. Second, the interviewee must be informed about the crime and requested about his position at the crime's time with details. After that, a set of specific questions must be asked (based on the incriminating data). Finally, the inquirer repeats to the interlocutor his saying to let him correct if it has anything wrong and to explain the inconsistences if it exists [7].

Unexpected questions: Vrij and her colleagues have tested the method of posing unlooked for questions when there are two or more guilties. This technique is mainly based on the consistency between subjects declarations. The idea was that innocents stories should coincide (they describe the reality). Contrariwise, guilties proclamations, even if they prepare it, will not be corresponded when unexpected questions are asked, especially when they are separated.

The test of Vrij et al. was realized on 80 persons (in pairs) where 40 were convicted and the rest were innocent. The attainment began by describing the restaurant layout. Then, easy questions were asked. Finally, unlooked for queries concerning temporal and spatial information were posed. The first and latter questions were discriminant unlike anticipated questions [8].

Drawings: Later, in 2012 Leins, Fisher and Vrij investigated another technique by characterizing the place and positions of present people. They concluded that truth-telling description contains more contextual details. This is because, liars are not ready to this type of questions or that they only focused on the present persons, in order to prepare responses, or not focused on the place or the contrary [9].

4) Contextual signs approach: The second orientation besides the active interviewer technique is to follow contextual index of lie. Park et al. have done in 2002 an ingenious test to 202 persons to look for the exactitude of the supposition that lie can be detected from behavioral indices in the real world. Candidates were invited to recount the manner in which they discovered lie in real situation that they lived. The answers showed that indicia were generally: information from another part, physical evidence or confessions. Park et al. concluded that the absence of this external information, judged as the most useful for deception detection, prevent scientific studies to be extrapolated to the real world.

Conclusively, the two approaches are complementary. Even if behavioral signs are significant to detect lie, contextual cues

Fig. 1. The upper front part of the brain: the forelock

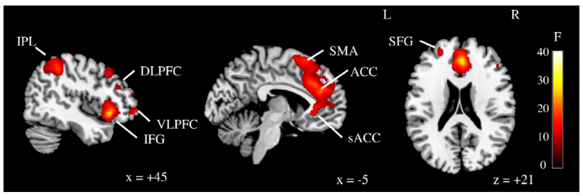

Fig. 2. The brain parts responsible for lie [11]

TABLE I
THE BRAIN PARTS RESPONSIBLE FOR LIE

Spontaneous truth-telling	High activity in the fronto-parietal network (right DLPFC, VLPFC and IPL)
Spontaneous lying	High activity of the sACC, right IFG, left SMA, ACC, IPL and SFG
Instructed lying	right IFG, left SMA, ACC, IPL and SFG

can be used by the interviewee to establish the veracity [10].

5) Brain activity: Scientists have recently proved that the center of lying is the frontal lobe. A significant activity occurs in the forelock when the man lies or makes errors (Figure 1). Many studies, focused on the examination of deception neural correlates due to the functional magnetic resonance imaging (fMRI) evolution, have shown that the anterior cingulate cortex (ACC) and the prefrontal cortex were more activated during lying. Yin et al. has discovered that it has great differences in brain activity when a man lie spontaneously, when he planned his lie and when he says the truth [11] (Figure 2 and table I).

6) Lie detection technologies: Besides, it exists lie detection technologies applying pattern recognition in face like Facial Action Coding System (FACS) and Face Reading Technology for Lie Detection from Ugail [2]. Similarly, a few lie detection application were designed basing on voice signal like Voice stress analysis (VSA), the psychological stress evaluator (PSE), and The Diogenes digital voice stress analysis, Layered Voice Analysis (LVA) in addition to software applications, considered inadequate, in the App Store on iTunes or on Android. Table II present a desciprion of a set of these technologies provided by vendors on their website and one field survey for evaluation of the narratives provided by field of its use. Notes: TP= True positive, TN= True Negative, FP= False positive, FN= False Negative [12].

TABLE II
RELATED WORKS ON LIE DETECTION FROM SPEECH SIGNAL TECHNOLOGIES

Year	Device	Observation	Results
2007	LVA	Poor validity in detecting stress and/or deception	TP= 15% TN= 95% FP=15% FN= 73%
2007	CVSA	Poor validity in detecting stress and/or deception	TP= 8% TN= 90% FP= 35% FN= 65%
2002	PSE	Not effective in detecting deception	No results indicated
2002	Sonogram measured voice pitch, intensity and duration	neither reliable nor useful	at chance
2002	Vericator	Detected stress but not deception	No results indicated
2002	Diogenes	Detected stress but not deception	No results indicated
1995	CVSA	failed to detect stress	Accuracy=38.7% chance= 25%
1994	Audio pitch analysis and spectrum decomposition	No voice measure reliably indicated deception	No signficant differences
1990	Mark II Voice Analyser	Chinese males showed higher level for prepared lies only no effects for females	Interaction with type of lie and sex more stress detected for males with prepared lies
1973	VSA	Did not distinguish deception	TP=36% chance=33%
1973	PSE	Did not distinguish deception	TP=32% chance=33%

The percentage satisfaction still the problem with lie detectors using pattern recognition speech and face [2].

B. Stress detection

There are few works in literature which focus on Lie detection due to its high correlation with incoming information form blood, brain and stress states. Researchers still understand what does mean a liar? How can we detect a liar based on his speech? For this reason, we present in our state of the art an overview of related work based on stress detection coming from audio signal. Physiological analysis demonstrated that the stress changes the hormonal levels of the body which provokes reactions like increased heart rate and blood pressure, muscle activation and respiration fastness [2]. Consequently,

TABLE III
RELATED WORKS ON STRESS DETECTION FROM SPEECH SIGNAL

Ref	Features	Classification	Accuracy
[14]	380 acoustic features MFCC, TEO	SVM	86%
[15]	Pitch, Spectral centroid, High frequency, ratio, Speaking rate, MFCCs, TEO-CB-AutoEnv	GMM	81% indoor 76% outside
[16]	Pitch, MFCC, RASTA PLP	SVM	92.4%
[17]	F0, Fhi, Flo, F0 variation, Jitter, sAPQ, DSH, NHR, DUV	LDA	84%
[18]	Pitch, MFCC, Delta and delta-delta coefficients,	SVM	78%% ‘

TABLE IV
SCENARIOS OF DATABASE RECORDING FOR STRESS FROM SPEECH DETECTION

Ref	Database
[14]	Students in exams period speaking two languages: Mandarin: 2: 42: 39 hours English: 3: 2: 28 hours
[15]	14 students: Job Interviews, Marketing jobs and Neutral Task (indoor and outdoor)
[16]	using Stroop test, TSST test and TMCT test
[17]	Emergency Calls Database
[18]	18 students aged 18 to 39. Many physiological and behavioral signals were recorded
[20]	24 different male speaking Czech two parts: stress from exam and few days later (repeat the same text)
[21]	18 persons (15 males and 3 females): 5 days of data. 90%: training and 10%: test

the muscles controlling the respiratory system are affected and process of speech production changes during stress [13]
Related work (see III) have shown that stress may be characterized from speech by the pitch variation, the higher considered frequency bands of energy, the speak rate and voice intensity increase [13].
Zuo and Fung have intended to study the gender and language dependence for stress detection from speech signal. A bilingual database as collected (Chinese Mandarin and English) from 25 university students 12 males and 15 females for the Mandarin database and 17 males and 14 females for the English database suffering from stress of exams. Twelve questions were posed commencing by the unstressed one to induce stress later. All responses were annotated by two professional annotators. They arrived to construct a stress classifier trained by SVM with a performance of 86% and to prove the gender dependence and the relative language independence on stress revelation. They proved also that from 380 acoustic features extracted using Open SMILE TEO and MFCC features are the more important features than pitch for stress detection [14].
Lu et al. presented Stress Sense a real-time system using smartphones for stress identification across multiple individuals in diverse conversational situations. Fourteen students, with a mean age of 22.86, participated to register the database. A job Interview and marketing job test were designed to activate stress in addition to neutral scenarios to obtain unstressed speech (reading text). Stress Sense features are: pitch, MFCCs, spectral centroid, High frequency ratio, speaking rate and TEO-CB-Auto Env. The performance obtained is about 81% for indoor and 76% for outdoor environments [15].
The MFCC and pitch pertinence to detect stress was confirmed by Kurniawan et al. in the hand of RASTA PLP features. In this work, they purposed to unify Galvanic Skin Response (GSR) signals and speech for stress detection from speech and achieved 92% accuracy with SVM classifiers. Despite acoustic extracted features have higher predictive power it presents more person dependent than GSR features which limited the generalization capacity. Combining classifiers of both features do not upgrade significantly the generalization accuracy [16].
The pitch is latterly designed as a significant factor to percept emotional stress in Demenko et al. 9 acoustic features (see table III) were used to categorize male and female stress state from an emergency calls database that consists of recordings of police intervention requests and crime notifications. The higher accuracy obtained was 84% [17].
Stress ID is another mobile application aiming to detect stress from voice by MFCC and pitch features with delta and delta-delta coefficients [18]. The experiment established with the Trier Social Stress Test (TSST), composed of a public speaking and cognitive task [19] giving rise to multiple a behavioral and physiological signals. They benefit from only voice signal to develop Stress ID. The classification used SVM to give a best accuracy about 78%.
There are many other examples presented in [13], [20] working acoustic features for stress detection. As far as, speech signal is operated for pathology and emotion detection and even for heart rate estimation.
The main objective of pathology detection from voice is not to assess disorders presence but also to accelerate and the diagnosis process [22]. Panek et al. sought for four types of pathologies (hyper functional dysphonia, functional dysphonia, laryngitis, vocal cord paralysis) which were tested using the ”a”, ”i” and ”u” vowels at high, low and normal pitch with female/male cases separation. The 100% accuracy obtained announce that the kPCA and NLPCA methods can perform a step towards vocal folds pathology detection [22].
The database used to develop these systems are described in IV.

C. Heart rate estimation

The heart rate (HR) is the number of times the heart beats in one minute. It varies among adults and can be affected by medical conditions, medications and psychological state. HR is one of the vital signs used to evaluate our physical and psychological condition. Accepting that when a man is lying he must be stressed proved that the increase of HR is a sign

TABLE V
RELATED WORKS ON HEART RATE ESTIMATION FROM SPEECH SIGNAL

Ref	Features	Classification	Accuracy
[23]	Formants	SVM	95%
[24]	RMMSD, HF, entropy, complexity, Pulse waveform (HR)	HMM	96.4%

of lying. However, a professional liar can be relaxed, on the other hand a man can be stressed although he is saying truth. Mesleh et al. have proposed a new non-contact method for HR extraction from vowel. The relationship modeling between vowel speech and HR show that a short evolution of the vowel speech should be caused by a heart beating. Further, the HR is estimated from maximum peaks of formant in the short-time Fourier transform (STFT). The proposed technique achieved 95% as average accuracy compared to a contact pulse oxy-meter [23].

III. PROPOSED SYSTEM

According to the state of the art, deception detection is mainly based on heart rate increasing, stress measurement and brain activity. We began by testing speech signal features able to distinguish lie from human voice.
For those purposes, we started by constituting a new database: the procedure was that every candidate was requested to try to deceive us by recounting false and true stories and let us provide the truthfulness of every story. The objective was to push them to make effort to mislead us. This first scenario was recorded in a sound studio. MFCC and pitch, are designed as best features for stress detection from speech signal, were extracted and tested.
Later, we thought to pass on outdoor environment and test the same scenario to look for the environment influence on these features. We are now recording a new speech database using the active interview technique to verify the difference between spontaneous and non-spontaneous lie.
Automatic speech processing is based on the identification of the suitable features to be extracted from the audio signal. The chosen features should be able to classify the signal content and discard the other stuff like background noise.
Like any system of speech classification, the first step in ReLiDSS is the sound acquisition. ReLiDSS is composed of two main phases. First, the train phase to extract and classify voice signal features. Second, the test phase of to detect deception from new data (Figure 3).
We have choosen to begin our inspection of relevant features of speech signal to classify lie with MFCCs and pitch which have proved their efficiency in automatic speech recognition and stress detection. Two scenarios were followed to build our own database. We have used the toolbox Voice Box of Matlab to extract MFCC and pitch from obtained records. Hereafter, we have trained our model lie classification with SVM. This model will be used to classify new data [25], [26].

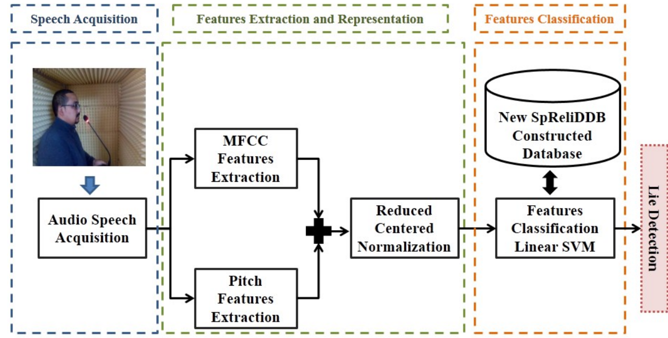

Fig. 3. Proposed System for Lie Detection using new constructed database

A. Mel Frequency Cepstral Coefficient (MFCC)

We have adopted to extract the MFCC (introduced by Davis and Mermelstein in the 1980's) which is widely used in automatic speech processing, to detect lie from speech signal. To compute MFCC from a speech signal the following step:

- The signal should be framed into short frames (20-40 ms. 25ms is the standard and in our system we have chosen 30 ms).

$$Y[n] = X[n] - 0.95x[n-1] \quad (1)$$

- The estimated periodigram of the power spectrum should be calculated for each frame. *

$$[Y[n] = X(n) - w(n) \quad (2)$$

- Then, we have applied the bank filter to the power spectra and we sum the energy in each filter.

$$w(n) = 0.54 - 0.45\cos(\frac{2\pi n}{N-1})\ (0 \le n \le (N-1)) \quad (3)$$

$$Y(w) = FFT[h(t) * X(t)] = H(w) * X(w) \quad (4)$$

- We have computed later the logarithm of all bank filter energies.

$$F(mel) = [2595 * \log_{10}[1+f]700] \quad (5)$$

- Finally, we have taken the DCT of the log filter bank energies.

Figure 4 illustrates the application of MFCC on voice speech.

B. Pitch

The pitch or the fundamental frequency noted "F0" represents the vibration produced by the vocal cords opening and closing speed under the pressure of the air. F0 is measured by the ratio between the sound wave speed and the sound length divided by two ones. The pitch variations over time are the intonation of the speech produced.
The increase of the standard deviation, mean value and/or the range of pitch and decrease of pitch jitter were ranked as a stress indication. Other features were also employed like minimum and maximum of pitch.
Pitch was extracted using the fxrapt function of Voice Box

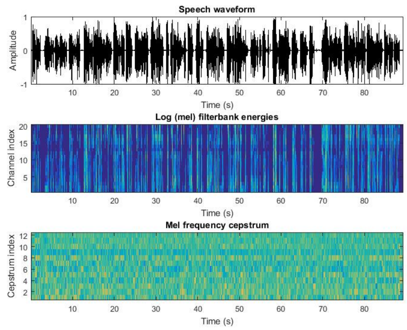

Fig. 4. MFCC implementation steps

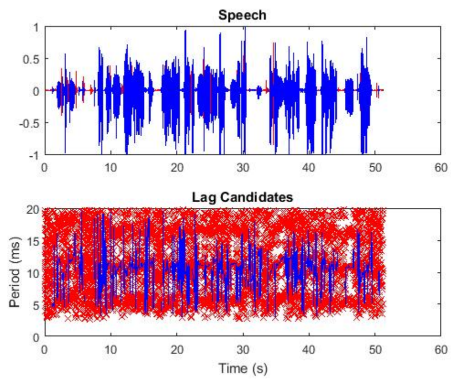

Fig. 5. Pitch extraction

which is illustrated in figure 5. Before the normalization of our features, we will proceed as mentioned in figure 3, to the concatenation of VMFCC and VPitch into the same vector x. The normalization of two features vectors having different range values like Pitch and MFCC descriptors is very helpful to change the unit of features. The RCN using following equation was carried out on features vector x obtained by the concatenation of VMFCC and VPitch issued from Pitch and MFCC techniques. Furthermore, all new normalized features will have the same mean and dispersion which can enhance classification task.

C. Features Classification Techniques

Support Vector Machine is a set of supervised learning techniques to solve discrimination and regression problems [27], [28], [29]. SVM can be used when the data has exactly two classes. For classification task, SVM separates data points in each class by finding the best hyper LAN that maximize the margin between the two classes [30], [31].
Many kernels were used in literature in order to resolve the problem of high dimensional data. Table VI below shows some function that can be used [28], [32]. Corresponding Kernel Functions Used to Perform the Support Vectors Machines.

TABLE VI
CORRESPONDING KERNEL FUNCTIONS USED TO PERFORM THE SUPPORT VECTOR MACHINES

SVM kernels	**Kernel Functions**
Linear	W.x
Radial Basis Function (RBF)	$e^{\frac{-\left\|x-x'\right\|^2}{2\delta^2}}$
MultiLayers Perceptron (MLP)	$\tan(x.x+c)$

Fig. 6. A candidate recording speech false statement

D. New constructed database ReGIM-Lab Lie Detection DataBase (ReliDDB)

ReLiDDB was collected from 40 volunteers (females and males), most of them are university and doctoral students. This speech corpus was recorded in unconstrained acoustic environments. First, data took place in a sound studio with a professional microphone. Later, we have decided to include outdoor records for the sake of environment influence research. To help them to be inspired, we have requested every candidate to recount false and true stories in a free order and let us detecting when he is lying. The duration vary between persons and between the same person's records. The range is about 1 min 20 seconds and 2 minutes per story).

E. ReliDDBI: ReGIM-Lab Lie Detection DataBase database based on the interview technique

We have started to build another database using the interview technique. We began with people we know well to facilitate the management of the dialogue progress.
We have persuaded the following approach:

- General questions: name, age, profession, etc.
- Personal questions: here we have tried to let them preparing and estimating which questions will be asked about that.
- Unexpected questions: suddenly questions about other domain were inquired.

We have exploited our knowledge about these people to provoke any subjects that we know they do not say veracity. Finally, the candidate has to cite false and true affirmation without having to say the real version.

TABLE VII
EVALUATION OF THE PERFORMANCE OF MFCC, PITCH AND OUR PROPOSED SYSTEM IN LIE DETECTION ON OUR NEW CONSTRUCTED DATABASE.

	Number	Accuracy Lie	Accuracy Truth
MFCC	13	79.32%	80.10%
Pitch	1	37.78%	43.33%
ReLiDSS	14	88.23%	84.52%

TABLE VIII
CONFUSION MATRIX OF PROPOSED SYSTEM FOR LIE DETECTION.

	Lie	Truth
Lie	88.23%	11.77%
Truth	15.84%	84.52%

IV. EXPERIMENTAL RESULT

A. Experimental result for Lie detection

Our proposed system called ReLiDSS was carried out on our new constructed database ReliDDB. This section shows the performance of proposed solution to detect lie and truth in human speech. Table VII illustrates the high performance of proposed scheme for lie detection in comparison with MFCC and Pitch features on the same database. Evaluation of the performance of MFCC, Pitch and our proposed system in Lie detection on our new constructed database.

The database analyse and annotation was made manually: eliminate silence and noise. The signal should be framed into short frames between 20 and 40 ms. We have chosen 30 ms which is judged as enough time to detect lies by human. After testing different SVM kernels, our data seem relevant and linearly separable. Our constructed database for learning with Linear SVM contains following features vectors:

- 175116 featured vectors of truth samples (37% from dataset)
- 286156 featured vectors of lie samples (68% from dataset)

Table VIII shows the confusion matrix generated after the training of our constructed dataset based on the normalization and the combination between the MFCC and Pitch features. Our proposed system is a first investigation on Lie detection using speech voice of persons. It has shown a competitive performance compared to lie detectors based on behavioral (stress), physiologic (blood pressure, heart rate) which are till now unable to detect efficiently a liar.

B. Experimental result for Lie detection for Person voice recognition

Automatic identificatin of persons from their voice may be useful to evolve the ReLiDSS aibility to detect lie. In this context, we suggested to induce the person voice recognition in our proposed system.

In order to achieve this, we started by testing the effectiveness of the MFCC features, pitch as well as their fusion. Our database incorporates for learning with SVM 137640 vectors related for 40 persons (candidats who participated on the record of ReLiDDB). The obtained results (Figure 7 and 8) have shown the efficiency of features fusion approach.

Our proposed approach was a valuable tool to recognize people from their voice. Table IX describe the performance evaluation of person recognition via the voice signal based on MFCC, Pitch and our system ReLiDSS descriptors on our new database ReliDDB.

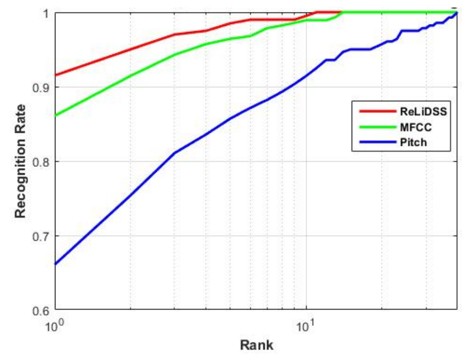

Fig. 7. CMC curves of ReLiDSS, MFCC and Pitch For Person Voice Recognition

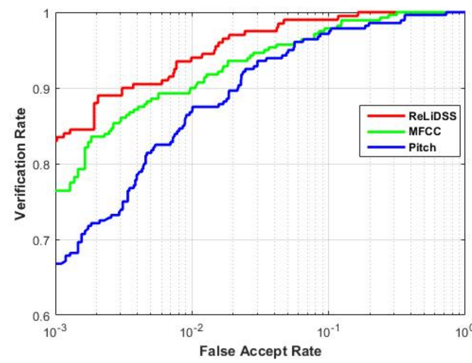

Fig. 8. ROC curves of ReLiDSS, MFCC and Pitch For Person Voice Recognition

TABLE IX
PERFORMANCE EVALUATION OF VOICE RECOGNITION BASED ON RELIDSS, MFCC AND PITCH DESCRIPTORS ON OUR NEW DATABASE "RELIDDB".

Approaches	MFCC	Pitch	ReLiDSS
% Equal Error Rate (EER)	16.23	10.05	4.21
% Rank-One Recognition Rate	66.90	87.23	91.87
% Verification Rate at FAR=0.001	66.87	78.43	84.12
% Verification Rate at FAR=0.01	88.59	90.21	94.74
% Verification Rate at FAR=0.1	97.02	97.23	99.51

V. Conclusion

The main contribution of our work consists on a preliminary investigation on Lie detection based on speech and voice. All related works were based on behavioral and physiologic features like heart rate, blood pressure, etc. These traditional methods present a disagreement of client because it requires a direct contact between consumer and system of acquisition. This investigation shows that the combination of MFCC and Pitch features was very helpful in order to classify a speech signal coming from human voice into lie or truth classes.

In order to evaluate our system, we have constructed a new database ReliDDB which contains 37% and 68% samples of lie. We have obtained as accuracy of lie and truth classification respectively 88.23% and 84.52%. As a future work, we will try to look into the environment influence on the deception detection from voice signal. Later, we will study the effectiveness of the active interviewer technique to increase the accuracy and to trust the results. Also, we will intend to study the relationship between voice, stress, heart rate and brain activity and produce a multimodal system. First, a new data base for lie detection will be built containing the speech, EEG and ECG and using the interview technique. Second a stress from speech detector will be implemented. Third, we will create an estimator of HR and BA from the speech signal. Finally, all this subsystems will be merged to give an efficient lie detector.

Acknowledgment

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

References

[1] A. Vrij, K. Edward, K. P. Roberts, and R. Bull, "Detecting deceit via analysis of verbal and nonverbal behavior," Journal of Nonverbal behavior, 2000, vol. 24(4), pp. 239-263.

[2] P. A. N. Cardona, "A compendium of pattern recognition techniques in face, speech and lie detection," IJRRAS, September 2015, vol. 24, pp. 529-551.

[3] G. Ben-Shakhar, "A critical review of the Control Questions Test (CQT," Handbook of polygraph testing, 2002, pp. 103-126.

[4] G. Ben-Shakhar and E. Elaad, "The Guilty Knowledge Test (GKT) as an application of psychophysiology: Future prospects and obstacles," Handbook of polygraph testing, 2002, pp: 87-102.

[5] J. Masip, and C. Herrero," New Approaches IN Deception Detection II. Active Interviewing Strategies and Contextual Information," Papeles del Psiclogo, June 2015, vol. 36(2), pp.96-108.

[6] A. Vrij, S. Mann, R.Fisher, S. Leal, B. Milne and R. Bull, "Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order," Law and Human Behavior, 2008, vol.32, pp. 253-265.

[7] P. A. Granhag, L. Strmwall and M. Hartwig. The SUE-technique: The way to interview to detect deception. Forensic Update, 2007, vol. 88, pp. 25-29.

[8] A. Vrij, S. Leal, P. A. Granhag,S. Mann, R. Fisher, J.Hillman and K. Sperry. "Outsmarting the liars: The benefit of asking unanticipated questions," Law and Human Behavior, 2009, vol. 33, pp. 159-166.

[9] D. Leins, R. Fisher and A.Vrij, "Drawing on liars' lack of cognitive flexibility: Detecting deception through varying report modes," Applied Cognitive Psychology, vol. 26, pp. 601-607, 2012.

[10] H. S. Park, T. R. Levine, S. A. McCornack, K. Morrison and S. Ferrara, "How people really detect lies," Communication Monographs, 2002, vol. 69, pp. 144-157.

[11] L. Yin, M. Reuter and B. Weber, "Let the man choose what to do: Neural correlates of spontaneous lying and truth-telling," Brain and cognition, vol. 102, pp. 13-25, 2016.

[12] S. Bhatt and S. E. Brandon, "Review of voice stress based technologies for the detection of deception", Retrieved February, 2008, vol. 20, pp. 2012.

[13] A. Alberdi, A. Aztiria and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," Journal of biomedical informatics, 2016, vol. 59, pp. 49-75.

[14] X. Zuo, and P. N. Fung, "A cross gender and cross lingual study on acoustic features for stress recognition in speech," In Proceedings 17th International Congress of Phonetic Sciences (ICPhS XVII), Hong Ko, 2011.

[15] H. Lu, A. T. Campbell, and D. Gatica-perez, "StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones, 2012," Ubicomp, pp. 351-360.

[16] H. Kurniawan, A. V. Maslov and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," In Computer-Based Medical Systems (CBMS), IEEE 26th International Symposium, June 2013, pp.209-214.

[17] G. Demenko, M. Jastrzebska, "Analysis of voice stress in call centers conversations," In: Proceedings of the 6th International Conference on Speech Prosody, 2012, vol.1, pp.3-6.

[18] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and M. O. Mozos, "Stress detection using wearable physiological sensors," In Artificial Computation in Biology and Medicine, 2015, pp. 526-532.

[19] R. Miller and C. Krischbaum, "Trier Social Stress Test," In Encyclopedia of Behavioral Medicine, Springer New Yok, 2013, pp.2005-2008.

[20] M. Stanek and M. Sigmund, "Psychological Stress Detection in Speech Using Return-to-opening Phase Ratios in Glottis," Elektronika ir Elektrotechnika, 2015, vol. 21(5), pp. 59-63.

[21] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," In Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on, September 2013, pp. 671-676.

[22] D.Panek, A. Skalski, J. Gajda and R. Tadeusiewicz, "Acoustic analysis assessment in speech pathology detection," International Journal of Applied Mathematics and Computer Science, 2015, vol. 25(3), pp. 631-643.

[23] A. Mesleh, D. Skopin, S. Baglikov and A. Quteishat, "Heart rate extraction from vowel speech signals," Journal of computer science and technology, 2012, vol. 27(6), pp. 1243-1251.

[24] X. Li, Z. Chen, Q. Liang, Y. Yang, "Analysis of mental stress recognition and rating based on hidden Markov model," J. Comput. Informa. Syst. 10 (18), 2014, vol. 27, pp 7911-7919.

[25] W. Ouarda, H. Trichili, A. M. Alimi and B. Solaiman, "MLP Neural Network for face recognition based on Gabor Features and Dimensionality Reduction techniques," Multimedia Computing and Systems (ICMCS), 2014 International Conference on, Marrakech, 2014, pp. 127-134.

[26] T. Dhieb, W. Ouarda, H. Boubaker, M. Ben Halima and A. M. Alimi, "Online Arabic writer identification based on Beta-elliptic model", 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), Marrakech, 2015, pp. 74-79.

[27] W. Ouarda, H. Trichili, A. M. Alimi, B. Solaiman, "Towards A Novel Biometric System For Smart Riding Club ", Journal of Information Assurance and Security, 2016, Vol. 11 Issue 4, p201-213.

[28] W. Ouarda, H. Trichili, A. M. Alimi and B. Solaiman, "Bag of face recognition systems based on holistic approaches," 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), Marrakech, 2015, pp. 201-206.

[29] T. Dhieb, W. Ouarda, H. Boubaker, A. M. Alimi, "Beta-Elliptic Model for Writer Identification from Online Arabic Handwriting", Journal of Information Assurance and Security, 2016, Vol. 11 Issue 5, p263-272.

[30] W. Ouarda, H. Trichili, A. M. Alimi and B. Solaiman, "Face recognition based on geometric features using Support Vector Machines," Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of, Tunis, 2014, pp. 89-95.

[31] I. Jarraya, W. Ouarda and A. M. Alimi, "A Preliminary Investigation on Horses Recognition Using Facial Texture Features," Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, Kowloon, 2015, pp. 2803-2808.

[32] W. Ouarda, H. Trichili, A. M. Alimi and B. Solaiman, "Combined local features selection for face recognition based on Nave Bayesian classification," Hybrid Intelligent Systems (HIS), 2013 13th International Conference on, Gammarth, 2013, pp. 240-245.