Towards Detecting Simultaneous Fear Emotion and Deception Behavior in Speech∗

Safa Chebbi and Sofia Ben Jebara

University of Carthage, SUP’COM, LR11TIC01 COSIM Research Lab, 2083, Ariana, Tunisia

Keywords: Speech Modality, Fear Emotion, Deception Behavior, Decision-level Fusion, Belief Theory.

Abstract: In this paper, we propose an approach to detect simultaneous fear emotion and deception behavior from speech analysis. The proposed methodology is the following. First, two separate classifiers to recognize fear and deception are conceived based on adequate voice features using K-Nearest Neighbors’ algorithm. Then, a decision-level fusion based on the belief theory is applied to infer whether the studied emotion and behavior are detected simultaneously or not as well as their degree of presence. The proposed approach is validated on fear/non-fear emotional and deception/non-deception databases separately. Results for separate classifiers reach an accuracy rate in the range of 95% with 24 features for fear recognition and 75% using 8 features for deception detection.

1 INTRODUCTION

Recently, speech modality has been receiving a growing interest by the scientific community as it is one of the most fundamental human communication mean. Indeed, it is considered as a leaky channel providing useful information about the speaker’s state (Fairbanks and Hoaglin, 1941). Therefore, several applications based on speech analysis have been conceived in the field of human-computer interaction in different research areas including psychology, cognitive science, artificial intelligence, computer vision, and many others (Ververidis et al., 2004) (Cowie and Cornelius, 2003) (Lee et al., 2005) (Pantic and Rothkrantz, 2003).

In this context, several studies based on speech analysis have made many achievements in the last years according mainly to emotion recognition and deception detection (El Ayadi et al., 2011) (Koolagudi and Rao, 2012) (Graciarena et al., 2006). Thus, acoustic properties hidden in speech have been investigated to identify behaviors and emotions. To do it, different vocal features have been explored such as prosodic ones modeling the accent and the intonation of the voice (Cowie et al., 2001), spectral properties (Rong et al., 2009), voice quality features (Scherer, 1986) and perceptual features (Haque et al., 2005).

Different emotions have been studied along with automatic emotion recognition patterns including fear, anger, happiness, disgust, surprise, boredom, sadness. It is mainly useful for man-machine interaction applications where the speaker’s emotions play an important role such as medical diagnostic tools (France et al., 2000), tutoring systems (Schuller et al., 2004), call centers (Ma et al., 2006), etc. Moreover, deception behavior has been one of the main tasks addressed based on automatic speech analysis which can be exceedingly helpful in forensic applications including mainly law enforcement and national security agencies (Wang et al., 2004) (Kirchhübel and Howard, 2013) (Bond, 2008). We relate for example detecting deception in the statements of suspects or witnesses or evaluating whether or not an individual is hiding information or providing incomplete information.

In this study, we attempt to detect simultaneous fear emotion and deception behavior in speech. This system could be especially useful in high stake situations related to security or delinquency issues. Fear and deception have been especially targeted since criminals are likely to be very fearful of being discovered and if they have to speak, they will logically attempt to deceive security agents or police investigators.

The adopted approach consists in conceiving two...
Towards Detecting Simultaneous Fear Emotion and Deception Behavior in Speech

systems separately for recognizing fear and deception through speech modality. The voice is analyzed by extracting a set of 72 pitch-based features. Then, relevant ones, that well discriminate between fear and non-fear classes as well as deception and truth classes, are selected. These features are fed to the related classifier for fear or deception detection based on K-Nearest Neighbors (KNN) algorithm. Local decisions related to fear and deception detection are taken and their outputs, in terms of probabilities are used as inputs of the decision level fusion module. Afterward, speech-based decisions captured from fear and deception recognition systems are merged to get the final decision about the degree of simultaneous fear and deception detection. The main contribution in this research consists in the use of the belief theory in order to integrate fear emotion and deception behavior together.

The rest of this paper is organized as follows: in the next section, the adopted approach is presented and described. Section III will provide a detailed description about the proposed decision-level fusion. Finally, section IV is devoted to present the fear and deception classification results as well as the fusion results for simultaneous fear and deception detection.

2 THE PROPOSED APPROACH FOR SIMULTANEOUS FEAR AND DECEPTION DETECTION IN SPEECH

The block diagram, depicted in Fig. 1, illustrates an overview of the adopted speech phases analysis leading to simultaneous fear and deception detection. The proposed approach consists of 5 steps: feature extraction, feature selection, fear and deception classification, decision level fusion and finally simultaneous fear and deception detection. These steps are described in details below.

2.1 Feature Extraction

After acquiring speech, acoustic features are first extracted from the voice. We have been interested in vocal-folds related features and more precisely in pitch which represents its opening-closing frequency. The pitch has been widely investigated in several studies for emotion recognition as well as deception behavior detection and has proved its usefulness (Ekman et al., 1976). Therefore, a set of 72 pitch-based features is considered. They are classified into four groups: 12 usual measures (mean, max, deviation, ...), 28 features related to pitch’s derivative and second derivative (as they are linked to vocal folds’ vibration speed and acceleration), 14 features related to speech voicing and 18 varied others. The whole set of features, explored in this study, is displayed in Tab. 1 below.

2.2 Feature Selection

After features extraction, relevant feature selection is a crucial step which should be performed in order to avoid the curse of dimensionality phenomenon. Highly informative features are selected based on the Fisher Discriminant Ratio (Theodoridis and Koutroumbas, 2009). The latter is used to quantify the discriminatory power between fear and non-fear classes as well as deception and truth classes. Thus, features are ranked in descending order according to their FDR importance. Then, the classification is manipulated, separately for deception and fear classifiers, by adding at each iteration one feature from the ranked list. Finally, the features considered as relevant are the ones providing the best classification result with the minimum feature number.

2.3 Classification Step

After selecting relevant features which best discriminate between fear and non-fear classes as well as deception and truth classes independently, they are fed to the related classifier (fear or deception). These classifiers infer the emotion/behavior class most likely expressed (fear or non-fear for fear classifier and deception or truth for deception classifier) as well as the probability of each identified status. In this study, the classification has been manipulated using K-Nearest Neighbors (KNN) algorithm. KNN has been chosen according to previous study dealing with a comparison between many classifiers (KNN, decision tree, support vector machine, and subspace discriminant analysis). The latter has revealed that KNN gives an adequate tradeoff between classification performance and features dimensionality. The classification quality was judged using many complementary criteria (Sokolova and Lapalme, 2009):
- Accuracy: it stands for the overall effectiveness of a classifier.
- Precision: it represents the class agreement of the data labels with the positive labels given by the classifier.
- Recall: it stands for the effectiveness of a classifier to identify positive labels.
- F1 Score: it stands for the relation between data’s positive labels and those given by a classifier. Indeed, the precision and recall measures are combined providing a single measurement which is the F1 Score.

2.4 Fear and Deception Decision-level Fusion

After passing fear and deception probabilities to the fusion unit, they are combined together to infer a final decision about the simultaneous fear and deception detection. We adopted an advanced approach based on belief theory for decisions merging (Shafer, 1976).

The main advantage of fusion belief theory is that it takes into consideration the imprecision and uncertainty of fear and deception classifiers and returns, as a result, the probabilities of detecting simultaneous fear and deception. Hence, the belief theory requires, as inputs, the probabilities of being a fear or no fear sequence as well as the probabilities of being a deceptive or truthful sequence. It returns as output the probabilities of three levels of simultaneous fear and deception detection (high, low and not). The adopted fusion methodology is described in detail in the next section.

2.5 Simultaneous Fear and Deception Detection

Finally, the probabilities of simultaneous fear and deception detection levels are passed to the decision unit. The level with the maximum probability is the one retained as final decision. It is then decided whether there are simultaneous fear and deception or not as well as their intensity degree (high, low or no fear-deception).

3 DECISION-LEVEL FUSION APPROACH

The goal of this section is to present in more detail the fusion level approach and how to make a decision about whether fear and deception are detected simultaneously. This fusion approach takes as inputs fear and deception classifiers responses.

3.1 Basic Beliefs Assignment

This level corresponds to quantifying our beliefs using a credibility function based on prior knowledge about the reliability of fear and deception classifiers. In more detail, it consists in the following steps.

i) Modeling fear and deception discernment frames denoted \( \Omega_f \) and \( \Omega_d \) respectively. They contain all possible outputs of their classifiers: fear (F) and non-fear (\( \bar{F} \)) are the two possible outputs according to fear classifier and deception (D) and non-deception (\( \bar{D} \)) are those according to deception detector one. Formally, \( \Omega_f = \{F, \bar{F}\} \) and \( \Omega_d = \{D, \bar{D}\} \).

ii) Deducing the power sets of the discernment frames denoted \( 2^{\Omega_f} \) and \( 2^{\Omega_d} \). They contain all the possible combinations of the hypotheses. Formally: \( 2^{\Omega_f} = \{\emptyset, \{F\}, \{\bar{F}\}, \{F \cup \bar{F}\}\} \) and \( 2^{\Omega_d} = \{\emptyset, \{D\}, \{\bar{D}\}, \{D \cup \bar{D}\}\} \), where \( \{F\} \) (resp. \( \{\bar{F}\} \)) stands for the hypothesis that fear is true (resp. false), \( \emptyset \) represents the conflict between fear and non-fear, and \( \{F \cup \bar{F}\} \) which represents the uncertainty area between fear or non-fear hypothesis. The same principle of definitions are available for \( 2^{\Omega_d} \).

iii) Based on the reliability of fear and deception classifiers, a mass is assigned to each subset of the power sets \( 2^{\Omega_f} \) and \( 2^{\Omega_d} \), reflecting their beliefs. These masses take values in the interval \([0,1]\). That is to
say, these masses translate the certainty degree and ignorance of the problem. \( m_f \) and \( m_d \) are fear and deception belief masses respectively. In our case, \( m_f \) is defined as follows (similarly for \( m_d \) but replacing \( f \) with \( d \)):

\[
m_f(x) = \begin{cases} 
0 & \text{if } x = \emptyset \\
\alpha_f \cdot \text{score}_f & \text{if } x = \{F\} \\
(1 - \text{score}_f) \cdot \alpha_f & \text{if } x = \{\bar{F}\} \\
1 - \alpha_f & \text{if } x = \{F \cup \bar{F}\}
\end{cases}
\]

where:

- \( \alpha_f \) and \( \alpha_d \) are fear and deception classifiers accuracy rates.
- \( \text{score}_f \) and \( \text{score}_d \) are posterior probabilities of being fear and deception sequences respectively.

Indeed, a null mass is assigned for \( \emptyset \) as a sequence can be only fear or non-fear. The fear mass \( (m_f(\{F\})) \) is defined as the product of the fear classifier prior probability by the probability of having a fear sequence \( \text{score}_f \). The non-fear mass is defined as the product of the fear classifier prior probability by the probability of having a non-fear sequence \( (1 - \text{score}_f) \). The fear and non-fear union mass is assigned the probability of having faulty predictions by the fear classifier (represented by \( 1 - (m_f(\{F\}) + m_f(\{\bar{F}\})) = 1 - \alpha_f \)).

### 3.2 Categorization and Decision Level

The second level of fusion belief theory corresponds to decision making with 3 levels of detecting fear and deception simultaneously: high-fear-deception, low-fear-deception, no-fear-deception. It is constructed as follows:

1. Combining fear and deception discernment frames using cartesian product between \( \Omega_f \) and \( \Omega_d \).
The obtained discernment frame, denoted $\Omega_{f,d}$, is defined as:
$\Omega_{f,d} = \Omega_f \times \Omega_d = \{F, \bar{F}\} \times \{D, \bar{D}\} = \{(F, D), (F, \bar{D}), ...\}$. Hence, four couples of emotion/behavior are obtained.

ii) Transiting from emotion/behavior couples to simultaneous fear and deception detection problem has been carried by creating the set $\Omega_{FD} = \{FD, FD_{low}, FD_{high}\}$. FD stands for not detecting deception and fear simultaneously, it corresponds to the subset $\{(\bar{F}, \bar{D})\}$, where neither fear nor deception is detected. $FD_{low}$ is defined as low level of simultaneous fear and deception detection. It contains two subsets $FD_{low} = \{(F, \bar{D}), (\bar{F}, D)\}$, which correspond to the identification of only one emotion/behavior. $FD_{high}$ is defined as high level of simultaneous fear and deception detection $FD_{high} = \{(F, D)\}$ since both fear and deception are detected.

iii) The last step of fusion belief theory consists in decision making. In our study, the pignistic probability, noted $BetP$, has been used as the decisive criterion [12]. This latter consists in equiprobably distributing the beliefs of the hypotheses. Indeed, the pignistic probability of each hypothesis from $\Omega_f$ is calculated using the following formula:

$$BetP_{\Omega_f} = \left\{ \begin{array}{l}
BetP(F) = m_f(F) + \frac{m_f(F,D)}{2} \\
BetP(F) = m_f(F) + \frac{m_f(F,\bar{D})}{2}
\end{array} \right.$$  \hspace{1cm} (2)

The same kinds of formula are available for deception. The probabilities of the three levels of detecting fear and deception simultaneously are calculated as follows:

$$\begin{align*}
prob(FD) &= prob(\bar{F}, \bar{D}) = prob(\bar{F}) \cdot prob(\bar{D}) \\
prob(FD_{low}) &= prob(F, \bar{D}) + prob(\bar{F}, D) \\
&= prob(F) \cdot prob(\bar{D}) + prob(\bar{F}) \cdot prob(D) \\
prob(FD_{high}) &= prob(F, D) = prob(F) \cdot prob(D)
\end{align*}$$  \hspace{1cm} (3)

As a result, the hypothesis with the maximum probability is the one retained as the final decision. It is then deduced whether fear and deception are detected simultaneously as well as their intensity degree (high, low, not).

4 EXPERIMENTAL RESULTS

4.1 Corpus

The fear emotion recognition approach is tested over the combination of three audio emotional databases: EMO database (Burkhardt et al., 2005), SAVEE database (Jackson and Haq, 2014) and RAVDESS database (Livingstone and Russo, 2018). The objective is to get a large and varied scenario by combining many languages and emotions. In total, a large database including 2247 sequences elaborated by 38 actors simulating 8 emotion types (fear, anger, happiness, disgust, surprise, boredom, neutral and sadness) is obtained. The classes repartition through the corpus is the following: 14% for fear and 86% for non-fear class.

The deception behavior recognition has been investigated with a real-life trial deception detection dataset (Pérez-Rosas et al., 2015). It is an audio-visual database consisting of videos collected from public court trials. Statements provided by defendants and witnesses in courtrooms are accumulated and labeled based on judgment outcomes and police investigations. It consists of 196 video clips: 53% of them are deceptive and 47% are truthful ones.

4.2 Fear and Deception Detection Results

The objective of this subsection is to present the performance of fear emotion and deception classifiers separately. As mentioned previously, the features are ranked according to the FDR importance then the classification is performed by adding one feature from the ordered list at each iteration.

Fig. 2 (resp. Fig. 3) illustrates the classification criteria evolution according to features’ dimensionality. The x-axis gives the number of features used in classification, from the ranked list by fisher discriminant ratio (FDR). The classification criteria are illustrated in the y-axis.

Based on Fig. 2, one can notice that all criteria, except recall, present the same evolution: an increasing then decreasing variation over 30 features vector size. The evolution of the recall measure presents an increasing variation and then stabilization. The best values have been obtained for a feature dimensionality between 20 and 30. The intersection between all measures seems to be the best tradeoff: features number= 24, accuracy= 95%, precision= 95%, recall= 95%, F1score= 95%, TPR= 95%, TNR= 95%. The list of most relevant features corresponds to 6 usual features, 3 features related to the derivative and second derivative, 3 ones from the speech voicing family and 12 others.

Dealing with deception, one can notice from Fig. 3 that criteria evolution fluctuates in the ascending and descending order and results are worse than those of fear. It is perhaps due to the database size (196 se-
Moreover, the classification criteria are closer for a dimensionality of less than 10. When dealing with a tradeoff between all classification measures, the feature group with 8 vector size seems to be the most adequate one (features number= 8 (3: usual measures, 3: derivative and second derivative family, 2: others), accuracy= 75%, precision= 77%, recall= 73%, F1score= 75%, TPR= 73%, TNR = 78%).

4.3 Fusion Results

As there is no ground truth dealing with simultaneous fear and deception, we propose to validate this study using the previous databases. For the four separate subsets (fear, non-fear, deception, and truth), we calculate the probabilities of the 4 possible couples of emotion/behavior. For each test sequence of the 4 classes, the probabilities of belonging to the fear and deception classes are calculated separately and then the fusion method is applied. The 4 couples probabilities \( (\text{prob}(F,D); \text{prob}(F,\bar{D}); \text{prob}(\bar{F},D); \text{prob}(\bar{F},\bar{D})) \) are obtained and based on them the degree of fear and deception detection is deduced. Indeed if \( \text{prob}(F,D) \) is the highest one, it corresponds to a high level of simultaneous fear and deception \( (FD_{\text{high}}) \). Else if \( \text{prob}(\bar{F},\bar{D}) \) is the highest one, it corresponds to the absence of simultaneous fear and deception \( (FD) \). Otherwise, it corresponds to a low level of simultaneous fear and deception detection \( (FD_{\text{low}}) \). Boxplots of the couples probabilities distribution have been drawn for fear, non-fear, deception and truth test subsets.

Based on Fig. 4 and Fig. 6, one can notice that
the boxplot of the couple probability $\text{prob}(F,D)$ has a higher range value compared to other couples for fear sequences as well as deception ones. Thus, it confirms our hypothesis assuming the correspondence of fear emotion and deception behavior in speech.

Based on Fig. 5, we can see a clear difference in the distribution of $\text{prob}(\bar{F},\bar{D})$ compared to other couples probabilities, which is an expected result as we deal with non-fear sequences. Also, this result confirms the absence of deception as the non-fear subset is conceived around primary emotions (anger, neutral, disgust, happiness, boredom, and sadness).

According to truth subset and based on Fig. 7, the distribution of $(\bar{F},D)$ probability presents higher range values compared to others. Although, it is expected to have higher probabilities for the couples
(F, D) and (F, D).

After calculating the probabilities of the couples of emotion/behavior, we deal with the probabilities of detecting simultaneous fear and deception. We hence calculate the probabilities of the three considered levels: high, low, and not (prob(FD_{high}), prob(FD_{low}), and prob(FD)).

Fig. 8 represents the simultaneous Fear-Deception degrees for the test subset. For each sequence, the probabilities to detect a high, low, or non-Fear-Deception level are illustrated. One can note that the highest probabilities are obtained for the high Fear-Deception class then the low Fear-Deception class, which is an expected result as the highest probabilities are obtained for one of the couples (Fear, Deception) or (Fear, non-deception).

Tab. 2 presents the rate of Fear-Deception levels for each test database subset (fear, non-fear, deception and truth). According to fear and deception test subsets, the majority of sequences have been judged as a high-Fear-Deception level, some of them as a low-Fear-Deception level and few cases as non-Fear-Deception. According to non-fear subset, the majority of sequences have been judged as No-Fear-Deception. For truthful ones, the majority of them have been judged as a low-Fear-Deception level.

5 CONCLUSIONS

The goal of this study was to detect both deception behavior and fear emotion in speech. The proposed approach operates by fusing decisions from fear and deception classifiers based on the belief theory. The performance of separate classifiers reaches 95% and 75% as an accuracy rate for fear and deception respectively. Then our approach has been validated on fear/non-fear emotional and deception/truth databases. Future work will consider other emotions and behaviors whose detection may be of high importance in forensics applications. Other modalities may as well be interesting to explore such as body gesture and facial expressions.

REFERENCES