Speech Based Automatic Lie Detection

Mahmoud E. Gadallah* Matar A. Matar* Ayman F. Algezawi**
* Egyptian Armed Forces ** National Defense Center

Abstract:

This work studies the effect of the emotions that is experienced due to guilt situation on different vocal parameters in attempt to identify whether or not the suspect is lying. The homomorphic speech processing is applied to extract the vocal parameters related to the source excitation such as: pitch, pitch power and vowel duration and those related to the vocal tract such as: formant frequencies and its gain. Also the energy as a global vocal parameter is computed. The vocal parameters are extracted from normal speech utterances and from stressed utterances for the same suspect in order to determine the most significant vocal parameters that can be affected by emotional stress. Correlation coefficients were investigated between the pitch power and the digitized smoothed output of the Psychological Stress Evaluator (PSE). More than 0.8 correlation coefficient has been found.

Six cases of real time criminal suspects cases were investigated throughout this work. Traditional lie detection questioning techniques were used to develop questionnaires for these criminal cases. More over, a case for an actor simulating different emotional states (downloaded from the Internet) was investigated for the effect of different emotions on the vocal parameters. Speech vocal parameters and the PSE Hirch&Wiegele scoring method were investigated for stress (due to anxiety or guilt). Pitch contour exhibits the most significant sensitivity for speech-based stressed/unstressed classification.

1. Introduction:

Deception as a psychological process (from an evolutionary perspective), may be viewed as the action of human trying to hide informations from being showed up because the consequences, if these informations are shown up, may be against himself or some body that he is covering up. Such deceiving behavior could be called lying (The CIA's Secret manual on Coercive Questioning [1]).The tools for detection of deception, for purposes of justice, are being developed since in the middle of the century. Scientific techniques that can distinguish the changes in human physiological functions in case of lie or intention to deceive were initiated and continued to evolve.

The inventors have introduced a lie detector machine that can measure the human physiological changes (such as: blood pressure and pulse rate) during an interrogation, this device is known as the Polygraph, which depends on sensors attached to the suspect during an investigation to measure the physiological variance through it. These attachments to the suspects were the reason for a set of constrains which make the polygraph test complex. Later trials from the inventors to overcome the problem of attaching equipments to the suspect were directed to the speech, since the speech conveys mental balance and the general state of functioning of the entire organism, thus a speech based lie detectors were introduced to the market in attempt to replace the polygraph. These speech based lie detectors namely were: the Psychological Stress Evaluator (PSE) and the Voice Stress Analyzer (VSA). The PSE was developed by DEKTOR Corporation of Springfield, Virginia (1970) [2]. The PSE response is defined by Smith [3] *as a rise of the pen from its zero base-line, followed by a number of pulses of varying amplitude at the fundamental frequency.* B.F.Fuller [4] claimed that stress arousal causes a less amplitude variations as shown in Fig.1:

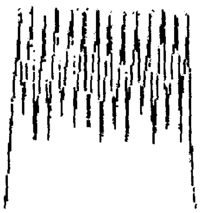

Fig.1 Stressed utterance output from the PSE

Hirsch & Wiegele [6] has developed a scoring method to enhance the poor reliability (which was reported as low as 0.38 Nachshon, et. [5]) of the PSE. Fig. 2 illustrates Hirsch & Wiegele scoring technique at which they count the number of adjacent vocal pulsations that do not differ in height and divides this number by the total number of pulsations in the voice sample. Hirsch & Wiegele have reported a reliability of 0:74 for their scoring technique [6],[7]. In Fig.2 the scoring result for the left pattern is 6/4 = 1.5, the scoring for the right pattern is 1/3 = 0.33, i.e. the right pattern seems to be a stressed utterance.

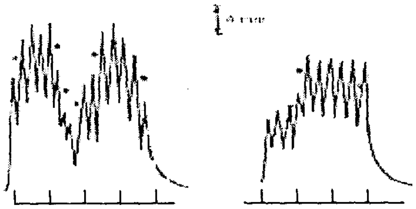

Fig. 2 Hirch&Wiegele scoring method

2.Speech and Emotions

Carl Williams & Kenneth Stevens [8] reported that respiration is frequently a sensitive indicator in certain: emotional situations, such as, *conscious attempts at deception,* and conflict. The respiratory pattern is frequently disturbed in anxiety states. An increase in respiration rate would presumably result in an increased subglottal pressure during speech. This height ended subglottal pressure would give rise to a higher pitch frequency Fo during voiced sounds in speech. The increased respiration rate could also lead to shorter durations of speech between breaths, with a consequent effect on the basic temporal pattern of speech. They also claimed that other relevant physiological effects of certain emotions are dryness of the mouth often observed under conditions of emotional excitement, anticipation, fear, and anger, and tremor and disorganization of motor response, observed under conditions of emotional conflict. These effects can have an influence on various components of the speech system, including the larynx, which is directly involved in the control of Fo.

Muscle activity in the larynx and the condition of the vocal cords are likely to have a more direct influence on the sound output and, in particular, on the fundamental frequency, than changes in muscle activity in other parts of the speech generating system, such as the tongue, lips, and jaw. The reason is that the vibrating vocal cords have a direct effect on the volume velocity through the glottis, whereas the other muscles and vocal tract components simply shape the resonant cavities for sound that is generated at the vocal cords. Such physiological changes as increased subglottal pressure, excessive dryness or salivation, and decreased smoothness of motor control can have an influence on the waveform of the pulses from the vocal cords, as well as on their frequency. For example, increased subglottal pressure generally gives rise to a narrowing of individual glottal pulses, and hence to a change in the spectrum of the pulses. Under some circumstances, such as excessive salivation, there may be irregularities in the waveform of the glottal output from one pulse to the next (this was noticed during the data collection for this work that for the excessive salivation pitch contour is down). It has been assumed by Bonner, Jones [9] that a *rise in frequency-rate above that normally basic is an indication of fear.* They said that *if voice rises in pitch then this thereby a loudly proclaim to the emotional tension.*

With regard to the consonant, the terms easy and hard, applied to the attack and release of the hyphen, refer in a very general and non-mathematical way to compare the length of the consonants beginning and end [9]. Bonner and Jones supposed that most individuals under emotional stress would speak more jerkily than is normal, i.e. would attack and release the hyphen more abruptly. They reported that since speech is a physiological process, based on the functioning of the organism at large, it is reasonable that they should find no definite one-directional trends in any of the attributes of speech studied under emotional tension. They also reported that there is nearly always a fairly marked change in the attributes of speech under emotive strain, but the nature of the change shows wide individual differences, as has also been found in the physiological factors of pulse rate and respiration. The same results has been obtained from the studies of the physiological factors of pulse rate, respiration, etc. Darrow says that strong emotion may cause a fall rather than a rise of blood pressure. It is reported in [9] that some researchers concluded that rate of breathing under fear increases in some people and decreases in others.

Since phonetic events play an important role in the transmission of the emotional modes, Lieberman [10] wondered whether pitch frequency plays a secondary role in the presence of phonetic information and whether pitch information is immaterial or negligible in the presence of a correct and complete phonetic description of the speech material. Lieberman concluded that phonetic content, gross changes in fundamental frequency, the fine structure of the fundamental frequency, and the speech envelope amplitude, are all contributing to the transmission of the emotional modes. The different emotional modes did not all depend to the same degree on all the acoustic parameters.

3.Vocal Parameters Extraction:

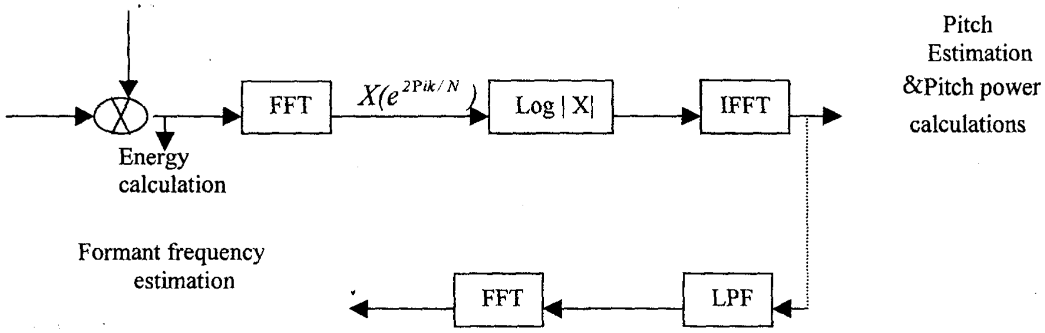
Pitch Estimation &Pitch power calculations
FFT
$X(e^{2Pik/N})$
Log | X|
IFFT
Energy calculation
Formant frequency estimation
FFT
LPF

Fig.3 vocal parameters estimation

An algorithm for vocal parameters extraction is developed in this work and it is based upon the homomorphic technique. We choosed this technique because it is accurate in pitch calculations as reported by M.Nool [11]. Fig. 3 shows the block diagram of the vocal parameters extraction via homomorphic technique. The speech signal is classified into voiced and unvoiced intervals. The voiced intervals are segmented into frames of 256 samples (23 ms) and the following steps are applied to these frames:

- Step 1:

1. Multiply the frame (frame i) by a Hamming window:

$$y_i(j) = \text{s(j)}.\text{w}(j) \text{ where w(j) } = (0.54 + 0.46\cos(2\pi\,\text{j}/255))$$

2. Compute the energy of the frame:

$$E_i = \sum_{j=0}^{256} y_{ij}^{2} \quad ,i = 1 \ldots., N$$

Where N is the total number of the frames in an utterance .

3. Compute the FFT for the frame:

$$Y_i(f) = \mathrm{Real}\{\, \mathrm{FT}(y_i[j])\} \quad .$$

4. Compute the logarithm of the magnitude spectrum of the frame :

$$Y_i(f) = 20 \log |\, Y_i(f)\,| \quad .$$

5. Compute the Inverse FFT for the frame to get the output of the homomorphic operation (Cepstrum).

$$y_i(j) = \mathrm{IFT}\{Y_i(f)\} \quad .$$

This output contains information of both the vocal tract (lower part of the time scale) and the excitation (high part of the time scale). This could be shown in Fig. 4.

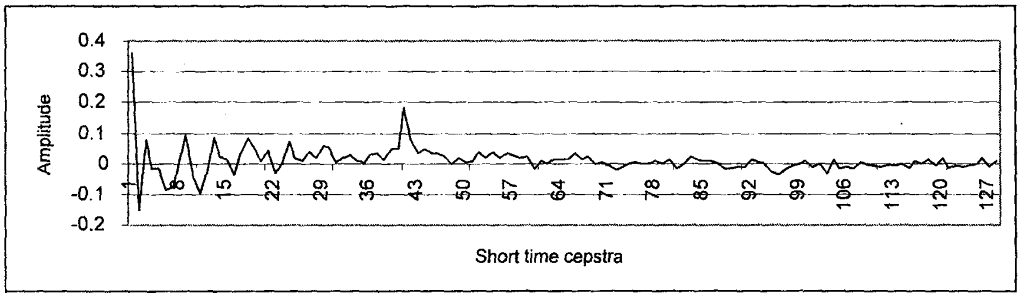

Fig. 4 typical output after IFFT via homomorphic analysis

6. Beginning from a high time sample location (40 for male, 20 for women and children) seek for the peak in the remainder of the frame.

$$X_i = \underset{\substack{j=40 \\ \text{or } 20}}{\overset{128}{\mathrm{Max}}}\ y_i(j) .$$

$$P_i = \mathit{index}\ \{Xi\}.$$

$$F\,P_i = i.$$

$$Ai = \{Xi\}.$$

Where P_i is the pitch period at frame i, $F\,P_i$ is the frame number whose pitch period is P_i and Ai is the amplitude of the signal at P_i .

- Step 2: For the whole voiced interval do the following :

1. Compute the mean energy for the utterance.

$$\overline{E} = (\sum_{i=1}^{N} E_i\,)\,/\,N \quad , \quad N = \text{total number of frames}.$$

2. Compute the maximum pitch Amplitude for the utterance.

$$A_{\max} = \underset{i=1}{\overset{N}{\mathrm{MAX}}}(A_i) \ .$$

3. Compute the vowel duration in msec [11].

$$Vd = 0.23 \sum_{i=1}^{N} (V/2) \quad ,V = \begin{cases} 1 & \text{for } v \geq s \\ 0 & \text{elsewhere} \end{cases}$$

Where s is a percentage threshold of Amax

4. Compute the mean of the pitch amplitudes:

$$\overline{AP} = (\sum_{i=0}^{an} A_i)/an \quad , A_i = \begin{cases} A_i & \text{for } P_i(A_{max}) + 5 > P_i(A_i) > P_i(A_{max}) - 5 \\ 0 & \text{elsewhere} \end{cases}$$

, an .... the number of P_i's that that lie in the range[11] :

$$(P_i(A_{max}) + 5) > P_i(A_i) > (P_i(A_{max}) - 5).$$

5. Compute the H&W scores [6]. The chart of the PSE height is 40 mm (the distance allowed for the heat pen to vary the peak from its base zero-level to the maximum) and the counting score for H&W method is computed if the difference of 4 mm or more between the peak and its successor peak occurred . It is assumed that a 0.1 of the maximum pitch amplitude to be the threshold difference between every pitch amplitude peak and its successor peak.

$$H \& W = 2(\sum_{i=0}^{v} k)/v \; , \; k = \begin{cases} 1 & \text{for } |(A_i - A_{i+1})| > 0.1 A_{max} \\ 0 & \text{elsewhere} \end{cases}$$

4.Case Studies

The vocal parameters, which could be affected by different emotions, are:

- The pitch contour.
- The pitch amplitude.
- The vowel duration.
- Formant frequencies displacement.
- Amplitude at the first formant frequency.

In this work we have studied the effect of stress on the five features mentioned above.
Seven case studies were collected and investigated for stress indications through relevant vocal parameters. The first case represents an actor simulating different emotional utterances was downloaded from the Internet. Cases two to seven were recorded from real crime situations. All the utterances that were collected during tests were sampled at 11 kHz and each sample is represented in 8 bits. In this analysis the speech signals are divided into frames of length 256 samples (.023 sec.).

The results of our investigations were compared with the confessions made by the suspects after the recordings were made. During tests, the suspects were asked with a set of prepared questions that should have an answer of "No" (a medium loud "لا" in Arabic language), in the same traditional ways used for the POLYGRAPH or the PSE investigations.

5.Test Generation

A convenient environment should be established for recording a session with the person-under-test. Quiet, in terms of corrupting acoustical noises, comfortable and equipped with good quality recording aids. As to the questionnaire, the asked questions should be well organized as to get the best results. It shouldn't touch the suspect's dignity as such a question may engage the reaction for any one. The suspect shouldn't be physically harmed either (the general state of the suspect has so much to do with the results [12]. The answers under consideration in this work has been collected using different questioning techniques such as the Peak Of Tension (POT) questioning technique. This technique is to perform a set of similar questions and among them is the direct question. For example a stolen case with 200$ inside the questions are formulated as the following:

■Did you know that 20$ were inside the stolen case ?

■Did you know that 40$ were inside the stolen case ?

■Did you know that 100$ were inside the stolen case ?

■Did you know that 200$ were inside the stolen case ?

■Did you know that 300$ were inside the stolen case ?

■Did you know that 350$ were inside the stolen case ?

The question is to be repeated 3 successive times with swapping for the direct question with another one in the middle. If the suspect gives a reaction (A difference in the vocal parameters show up) in the three times, then he should be involved in the robbery.
One of the cases that were investigated for stress is presented in the following:
In this case the suspect was accused by theft. It could be observed in Fig. 5, which represents the pitch contour plots, that the suspects pitch contours of the answers to the relevant questions is significantly higher than the mean of the pitch contours of the answers to irrelevant questions.

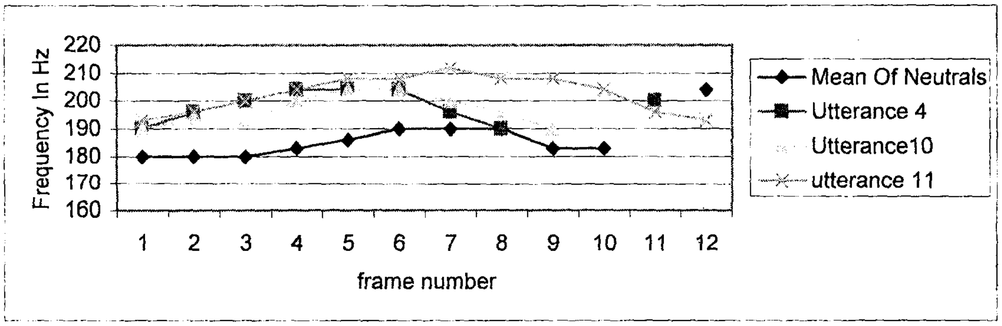

Fig. 5 Pitch contours plot for a suspect.

Fig. 6 represents the pitch power for the same utterances of Fig. 5. Also, considerable variations could be seen in the pitch power as shown in Fig. 6. The formant frequencies haven't shown a significant variation neither in the gain nor in the frequency locations.

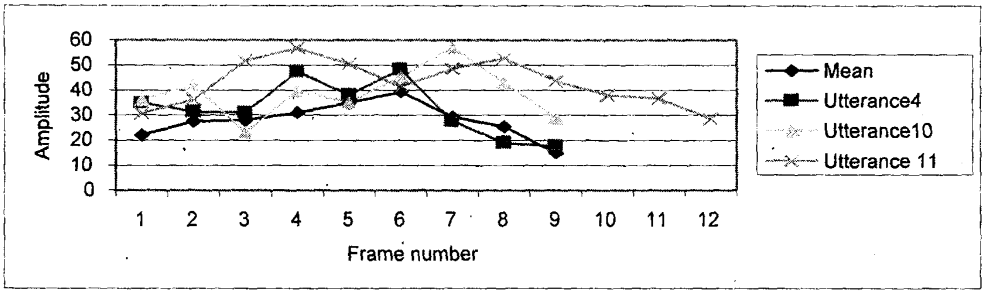

Fig. 6 Pitch Power

The efficiency of the different features for lie detection has been evaluated for all the cases considered in this work. This accuracy is computed by initially determining the ensured utterances that represents lie, then we calculate the ratio between the answers to irrelevant questions that contains significant deviations of the vocal parameters and the total number of the ensured lie utterances. Table 1 shows the lie detection accuracy for the different features. From this table, we can observe that the pitch contour is the most sensitive vocal parameter to the emotional stress.

Table 1 Lie detection accuracy for different features.

Vocal Measure	*Accuracy*
Mean energy	61%
Mean pitch amplitudes	78%
Max. pitch amplitudes	86%
Pitch contour	93%
Vowel duration	82%
H&W scores	74%

6.Pitch Contour-Based Automatic Stress Detection:

From the previous section, it is concluded that the pitch contour is the feature that could significantly reflect the emotional state. Therefore we propose an automatic stress detection algorithm based upon the pitch contour estimation. Such algorithm comprises the following steps:

- Using the pitch estimation technique described in section 3, calculate the pitch feature for all the utterances (answers) corresponding to the irrelevant questions.

$$\overline{Fo} = \text{average}(Fo).$$

- For each relevant utterance find its deviation from the mean $\overline{Fo}$

$$\Delta F = |Fo - \overline{Fo}|$$

- The relevant answer is classified as stressed one if its Fo is greater than or equal 5 Hz from the mean of the neutral Fo, i.e. $\Delta F \geq 5$ Hz.

Note: the pitch deviation 5 Hz is an empirical value taken via many tests.

7. Conclusions:

Homologue speech based stress evaluators: PSE and CVSA are actually classified by some researchers as stress arousal detectors. The CVSA may have a better chance since the operation is computerized and consequently less sensitive to the analyst experience compared to the PSE.

It is noteworthy that the polygraph doesn't distinguish anxiety or indignation from guilt and suffers from the lie complexities [4] and [13]. Consequently, the success of the results depends on the skills of convenience of analyst and the questionnaire.

The addressed problem of speech-based lie detection has been investigated. We have relied upon the claim that lie does not cause a known distinctive physiological reaction [1] and [13] (no distinct characteristics or patterns) but it could be identified (if there are no other reason for the person-under test) in response to the relevant questions. Therefore, what we have implemented actually is a stress detector.

Several case studies have been elaborated and automatic stress detector has been proposed on the basis of a pitch contour classification criterion. Classification accuracy scored higher than 90% for the implemented automatic stress detector.

References

[1] Furedly,J.J., Davis,C., and Gurevich,M.(1988) Differentiation of deception as a psychological process: A psychphysiological approach. Psychophysiology,25:683-688.

[2] Dektor Conterintelligence and Security ,Inc. (1973) Psychological Stress Evaluator , users manual.

[3] Smith, G.A (1977) Voice analysis for the measurement of anxiety. British Journal of Medical Psychology, 50, 367-373.

[4] Fuller,B.F. (1984) Reliability and validity of an interval measure of vocal stress. Psychological Medicine, 14(1), 159-166.

[5] Michael Noll ,(1966), Cepstrum Pitch Determination. Journal of acoustical society of america,41(2) ,293-309.

[6] Wiegele, T.C , Hirsch, L. (1981). Methodological aspects of voice stress analysis. New directions for methodology of social and behavioral science 7, 89-103 .

[7] Wiegele, T.C (1978). The psychophysiology of elite stress in international crises: a preliminary test of voice measurement technique. International studies Quarterly 22,467-511.

[8] Carl Williams (1972),Emotions and speech: some acoustical correlates. Journal of acoustical society of america,52(4) ,1238- 1250.

[9] Bonner,R., "changes in the speech pattern under emotional tension." American Journal of Psychology, 56, 1943, pp 262-273.

[10] Lieberman, P., and Michaels, S.B., Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. The Journal of the Acoustical Society of America, 34, 1962, pp 922-927.

[11] Michael Noll, (1966), "Cepstrum Pitch Determination. Journal of Acoustical Society of America, 41(2),293-309.

[12] Iacono,W.(1991) Another critical assault on Lie Detection.[Review of theories and applications in the detection of deception : A psychophysiological and international perspective]. Contemporary Psychology36:862-864.

[13] Elaad,E.(1990). Detection of guilty knoledge in real-life criminal investigations. Journal of Applied Psychology,75:521-29.