

THE
TrusterPro™
TECHNOLOGY
RELIABILITY
TEST

CONDUCTED BY:
SHLOMO BRUCK

URAN LTD.

Polygraph, Reliability, Graphology.

Member of the Israel Polygraph Examiners Association

Member of the American Polygraph Association

Tel: 972-3-6875424 Fax: 972-3 -5374486

e-mail: uranus@inter.net.il

This research and the patent-pending Truster technology it reviews are copyrighted, with all rights reserved.

This contents of this research may not be copied in any form.

To receive more copies or other information, please contact Uran LTD.

Reliability Research for the TrusterPro Vocal Lie Detection System

Introduction

The idea of vocal lie detection was born in the early 1940s, when a vocal lie detection experiment was conducted, using only hearing, within the framework of research conducted by Fay and Middleton. In this experiment, 47 test subjects were asked to answer a set of questions by either lying or telling the truth, while a team of 60 examiners attempted to identify truth or lies, using only their hearing. The research results showed chances slightly above mere guesses, in detecting lies (60.99%) and probability equal to guesses in identifying the truth (50.05%).

A later experiment (Mottley, 1974) separated the pitch from the sound wave, in an attempt to detect stress indicative of lies. The experiment was conducted on 20 female test subjects who were asked to answer “No” to several questions about selected issues (*Peak of Tension Test*). The analysis process examined density, main frequency, duration, format structure and harmonics. This examination failed to find differences between truthful and deceitful answers in all parameters except duration. The final analysis findings matched the findings reported by Fay and Middleton, i.e.- slightly above the probability of guessing.

Additional research was conducted by other scientists with the objective of characterizing the emotional content level using the level of pitch. Research conducted by Lieberman and Michaels in 1962 separated the pitch level from the sound wave of the speakers and the processed wave was played back to several examiners. The examiners reported not being able to identify the emotional level of the speaker using their hearing, when the processed wave was heard, as opposed to 85% success in identifying the original wave. The conclusion of the experiment was that there exists a certain correlation between the pitch level in the voice and the emotional intensity of the speaker.

In additional studies conducted by Stellar, Kraus, Geller, Olson and Apple (1977), the researchers claimed that the central frequency of the speaker (FO) rose when they were lying. Tolkemit and Scheller (1986) reported that the FO was actually less sensitive than the FO floor and claimed that it served as a better indicator of stress (the FO floor rises under stress). The FO floor was defined as the last FO value to be obtained during the speaker’s declaration.

An additional method of diagnosis (better known as PSE), was tested to varying degrees of success by Barland (1978) as well as Brenner, Bearnscombe and Schwartz (1979). This technology utilizes the identification of changes only in the speech frequency modulation (FM), usually called “microtrimmers”. This method was never accepted by most researchers and the assumed relation between the FM and the body’s autonomous system was never established. In a controlled experiment conducted by Horvath (1982), the system failed to identify lies at a probability higher than blind guesses.

Description of the Proposed Technology:

The present technology utilizes several indices derived from the sound wave (details below) and the construction of a multi-dimensional profile of the voice of the speaker in a neutral situation (on the basis of truth). The present technology is not limited to analyzing deception only in Yes or No answers and is in fact capable of analyzing any word or sentence.

The present technology utilizes a mathematical model for the automatic (computerized) analysis of the speech flow and may be implemented in the course of the actual conversation, while supplying indications such as True, False or others, in real time.

Assumptions at the Basis of the Theory:

1. The psychological structure of individuals (various test subjects) is unique and differs among various individuals. The condition of deception will render certain individuals emotional, create stress among other individuals and affect the cognitive level of other subjects.
2. Different testing situations may generate different reactions during analysis. When the deception is premeditated, the test subject may react in a certain way. On the other hand, when the subject must lie unexpectedly, the reaction may be different. Different reactions may also be obtained when the subject lies due to necessity or in order to challenge the interrogator.
3. The sound wave carries a great amount of information regarding the subject's mental, psychological and physical state.
4. The sound wave may be consciously altered by the subject (the subject may consciously change his voice).
5. There is no possibility of conducting an examination under true conditions and in an ideal environment.
6. There is no uniformity in the quality of equipment that will be employed in conducting the tests in the future.
7. The system does not directly measure physiological indices, and focuses on psychological indices that are directly expressed in the voice.

The indices that we will identify in the sound wave:

1. **SPT** (Samples per Thorn) - A pure value (unitless) expressing the presence of the relatively high frequencies in the analyzed sample. (The existing technology utilizes this value to characterize the emotional level of the subject).
2. **SPJ** (Samples per Jump) - A pure value (unitless) expressing the presence of the relatively low frequencies in the analyzed sample. (The existing technology utilizes this value to characterize the cognitive level of the subject).
3. **AVJ** (Average Jump) - A mathematical value expressing the average frequency in the range of low frequencies (characterized by SPJ). The existing technology utilizes this value to examine the degree of thought invested in the spoken words.
4. **JQ** (Jumps Standard Error) - A mathematical value expressing the standard deviation in the range of low frequencies in the examined sample. The existing technology utilizes this value to determine the degree of stress in the subject.
5. **MainF** (Main Frequency) - The value of the main frequency in the voice of the speaker (as a percentage of the overall contribution to the wave) and in a precise spectrum analysis of 16 frequencies, differing between men and women.
6. **FX** (Frequency Extras) - The number of additional frequencies in the analyzed frequency window, that possess relatively high values in relation to the MainF.
7. **FFLIC** - The number of direction changes (flicks) in the analyzed frequency window that possess a contribution exceeding a set minimum of the MainF.
8. **SFLC** (Sample Flicks) - The overall number of flicks in the sum of the values SPT and SPJ the sample windows of the original sound wave. In this technology, this value expresses the anticipation of a reaction from the interrogator constitutes a numerical value for the degree of uncertainty in the speech.
9. **ANT** (Anticipation) - The sum value in percentages of the two highest frequencies in the analyzed frequency window. The basic hypothesis of the system claims a direct correlation between the value of these values and the subject's degree of anticipation.

The Calibration Stage: The Neutral Emotional Voice Pattern

During the calibration stage, the system will perform several continuous examinations of different speech segments defined as the truth. The values described above will be collected in a table for concluding the calibration stage and constructing the neutral speech pattern. The calibration stage may last for anything between a single sample to six consecutive samples - as desired by the examining operator.

While taking into considering the basic assumption that each and every test subject is psychologically unique and different, the system will set an average value for each of the parameters mentioned above upon the completion of the calibration stage. An additional value expressing the standard deviation from this average and an additional value that - with the completion of the analysis process - will ultimately be used to determine the degree of reliability of such variables.

The system possesses minimal demands regarding the calibration samples and consequently, samples suspected of being lies will automatically be screened out and not weighed in at the

end of the calibration process. These minimal demands include both permanent and variable demands as the calibration process advances.

The Examination Stage: Compatibility with the Neutral Voice Pattern

The basic model states that any deviation from the profile that was determined during calibration requires analysis. Deviation on a single plane will express one type of diagnosis, while deviation on several parameters will lead to a different diagnosis.

For example: Processed SPT values that are considerably different than the Neutral Profile values will indicate a high degree of excitement, while a considerable deviation in SPJ values will be diagnosed as hesitation or insecurity. Deviation in both planes together will be diagnosed as a certain degree of lying (a small deviation is inaccuracy, a severe deviation expresses a lie).

The compatibility process is performed as follows:

The numerical value of all indices will be expressed as a percentage of the value set for the Neutral Profile and will be “softened” by the value that was identified as the standard deviation of the same series of values during the calibration process. A deviation from a value that was relatively stable during calibration will obviously be more significant than a value that tended to fluctuate more during calibration.

It should be mentioned that pursuant to the research model and the basic assumptions, the obtained diagnoses may appear to be without an expected logical sequence, yet the system’s diagnoses will nevertheless express the subject’s sensations as the words were said.

Lie Detection

The lie detection process consists of several moves:

1. Identifying the variance of the present values and the base values for each separate plane.
2. Generating summarized estimators for the overall degree of deviation, according to lie-related indices and stress-related indices.
3. Obtaining a specific diagnosis for each separate segment.
4. In the event that the diagnosis suspects a lie, the lie/truth-supporting estimators will initiate a lie verification process.
5. A comprehensive analysis of the entire sequence of diagnoses over time and in accordance with the psychological mold of the entire conversation. (Open conversation, stressed conversation, interrogation, job interview). This research will implement the molds/patterns relating to a stressful conversation and an interrogation (where the lies are premeditated).
6. Analysis of the sequence of results by a human expert.

The research assumes that it is impossible to detect deception among all test subjects in the same manner - since the situation of deception may cause excitement (emotional level), deviation (high excitement or negative excitement) among certain test subjects, while in others (or even the same subjects under different psychological situations) a situation of deception may lead to high stress (identified by a different estimator). On the other hand, additional reactions in a “deception situation” may also be extreme in other indices. For example, under a high degree of tension or very low degree of tension in the subject, the system identifies tension

using the MainF. The value of the Main Frequency, when analyzed as described above normally lies in the 25%-35% range for 95% of all the test subjects we ever examined.

According to the research assumptions, a high degree of tension would be a degree where the MainF value obtains values over 35%. Consequently, the greater the levels of tension, the higher the MainF value. A low degree of tension is characterized by a MainF lower than 25% and consequently, the lower the tension (negative), the lower the Main Frequency (MainF). According to the research assumption, any deviation in the tension level (upward or downward) may indicate a lie, where it is possible to differentiate between two “types” of lies: A premeditated lie where the subject would be asked to lie and an unplanned lie (more consistent with day-to-day situations). The system weighs the values of the deviation from the normal values and uses the results for obtaining the final findings. Rising values for FX, FFLIC and SFLC will serve as supporting indices for the detection of deception (regardless of the Neutral Voice Pattern). According to the examined technology, the higher these values, the higher the probability of deception.

Method:

The Subjects: The experiment consisted of a group of 60 individuals (each of which was tested under three different test conditions with an average age of 19. The group consisted mainly of men with a small number of women. The group was heterogeneous in terms of ethnic origin, education and social-economic levels. All the participants were healthy.

Location and Equipment:

The experiment was conducted in ordinary Examination rooms. The Examination room included: a desk, two chairs, a Lafayette polygraph instrument and a digital tape recorder (DAT) with a lapel microphone.

One of the Examination rooms also included an Olivetti Echos P-90 laptop PC, with TrusterPro software (alpha version). The computer included a sound card to which the microphone was connected.

Room layout: The walls of the room are bare. The room itself is acoustic.

All the inspections: The polygraph tests and TrusterPro examinations were recorded.

The Experiment

Five polygraph examiners and one TrusterPro examiner conducted the experiment. The Experiment Manager (EM) greeted the test subjects every morning. The EM staged a mock crime with the test subjects. The selection of the participants in the mock “crime” was totally random and created a situation whereby there were “guilty” subjects and “innocent” subjects.

The “crimes” included the following:

1. Theft of money
2. “Assault”
3. Narcotics

The test subjects were subsequently divided into Examination rooms.

The examiners in each room received the details of the mock crime incident from the EM, without receiving information regarding the involvement or innocence of the “suspect” test

subjects in these “crimes”.

At this stage, the test subjects were given a standard polygraph examination. Upon the completion of the examination, the polygraph examiners analyzed the polygraph output.

At this stage the TrusterPro examiner entered the Examination room and conducted a TrusterPro examination on the same test subjects.

Instructions Given to the Test Subjects

The instructions that were given to the test subjects were as follows:

- This system analyzes your voice to determine whether you are lying or telling the truth.
- We will now analyze the same issue that was tested by polygraph.

The inspection will be conducted in three stages:

- Stage A:* Tell me a little about yourself, details such as: name, last name, where you grew up...
- Stage B:* Tell me what you are suspected of doing, or what you know about the event on account of which you are undergoing this examination.
- Stage C:* I will now ask you several questions regarding the event you are suspected of having committed.

The Experiment Conditions

1. A polygraph examination conducted according to polygraph examination procedures, employing control questions, according to the relevant topic being examined.
2. ONLINE examination using TrusterPro. The subjects related in their own words what they were suspected of having committed and denied any involvement in the event. This was done in a continuous conversation whose average duration was approximately 40 seconds.
3. INTERVIEW examinations using TrusterPro, where the subjects were asked relevant questions along with control questions that were formulated in a format similar to polygraph examinations. The subjects answered only “Yes” or “No” to these questions (Appendix A - examination questionnaire).

Upon the completion of each stage - the results of that stage were analyzed.

Upon completion of Stage A, the polygraph examiners analyzed their results and submitted their findings to the EM.

Upon the completion of Stages B and C, the TrusterPro examiner analyzed the results of both stages and submitted his findings to the EM. At this stage, the EM submitted the “actual” results and the correlation was analyzed between the “reality”, the polygraph results and both stages of the TrusterPro examination results.

Upon the completion of the analysis, the EM thanked the subjects who participated in the tests and asked them not to divulge the objective of the examination and its essence to their friends,

during the next three weeks. The participants were also made to understand that both systems were successful in determining which of them were ‘truthful’ and which ‘lied’.

Results:

The five polygraph examiners who participated in the research evaluated the results of their tests, using the numerical scoring method (Barland & Raskin).

This method compares the relevant question with the control question, according to the examination routes. NBZ and Yankee methods were employed in the examination, when the comparison is always to the strong control in the chart. Result of the numerical scoring ranging between (+3) to (-3) were inconclusive, while findings beyond this range were either truth (over +3) or deception (below -3).

The numerical scoring results appear in Part A of Table-1.

The evaluations of the TrusterPro Online test results were performed automatically by the software. The program was set to regard an "Honesty Rate" of 95% - 100% as truthful, 90% - 95% as "Inconclusive" and under 90% as deception. The TrusterPro Online results appear in Part B of Table-1.

The results of TrusterPro in the Interview Mode were based on the indices obtained from the software, where the decoding method compared Truth Stress and Global Stress indices to the questions relevant to control questions and the overall decision refers to "larger than..." or "smaller than..." parameters only and not to numerical differences.

When the differences were below 10% the result was "Inconclusive".

The findings of this decoding appear in Part 3 of Table-1.

After the decoding, the "real" results were submitted by the EM and were then compared with the decoding findings.

Table-2 presents the distribution of the polygraph examinations results, as compared with the "real" results.

The distribution the TrusterPro results in the Online Mode, as compared with the "real" results, appear in Table-3.

The distribution of the TrusterPro results in the "Interview Mode", as compared with the "real" findings, appear in Table-4.

Table 1:

Examined by polygraph	n=59			
Polygraph errors (excluding Inc)	n=10	Of which:	4=F.P.	6=F.N.
	%=17		40%	60%
Examined by TrusterPro (Online)	n=50			
TrusterPro O.L. errors (excluding Inc)	n=17	Of which:	8=F.P.	9=F.N.
	%=34		47%	53%
Examined by TrusterPro (Interview)	n=50			
TrusterPro Interview errors (excluding Inc)	n=10	Of which:	7=F.P.	3=F.N.
	%=20		70%	30%

Table 2:

Distribution of polygraph examination results:

Examiner results	Innocent Subjects			Deceptive Subjects		
	DI	NDI	INC	DI	NDI	INC
N=	4	22	5	19	6	3
%	13	71	16	68	21	11
% (Excluding Inc)	15	85	-	76	24	-

Table 3:

Distribution of TrusterPro Online examination results:

Examiner results	Innocent Subjects			Deceptive Subjects		
	DI	NDI	INC	DI	NDI	INC
N=	8	16	0	15	9	2
%	33	67	0	58	35	7
% (Excluding Inc)	33	67	0	63	37	-

Table 4:

Distribution of TrusterPro Interview examination results:

Examiner results	Truthful Test Subjects			Lying Test Subjects		
	DI	NDI	INC	DI	NDI	INC
N=	2	15	2	17	3	6
%	29	63	8	65	12	23
% (Excluding Inc)	32	68	-	85	15	-

Discussion

The present study was conducted with the intention of analyzing the reliability of TrusterPro (TP) in two of its modules: TrusterPro Online (OL) and TrusterPro Interview (Int). The reliability test was conducted on cases of “mock crimes” and was compared with polygraph examinations conducted on the same test subjects and as compared with the “real” index - all test subjects were subjects who volunteered for the study. In this chapter of the article, I will attempt to clarify several points that influenced the results in my opinion.

The TrusterPro examinations were conducted *after* the polygraph examinations and were therefore somewhat “contaminated” by the polygraph examinations. This might have affected the TP findings.

No pre-test interview was conducted in the course of the TP examinations (as normally done during polygraph examinations). In the Interview mode, the questions were presented to the test subject during their writing, while in the Online mode, the test subject was asked to describe the degree of his involvement in the mock crime, in his own words. This decision, to conduct no pre-test interview, was based on the fact that a test subject, having just undergone a polygraph examination, does not require additional stimulation. In fact, we took advantage of the influence of the polygraph test.

In comparing the data between the polygraph tests and the TrusterPro results, we see a slight difference “in favor” of the polygraph in the results of both these instruments. Such a difference that may originate from the degree of contamination as stated above, i.e.- the subject’s motivation was higher before his/her polygraph examination (Davidson, 1968). Upon completing the examination, a certain general relaxation occurred. This relaxation was primarily expressed in the TP Int mode (possibly since this examination is similar in structure to the polygraph examination).

When employing the OL mode, we see a relatively high percentage of errors. In hindsight, after the study, it became evident that the test subjects had in fact told a “cover story” for their actual presence in the area. The cover story, which included details such as: “I work here in the building...”, “I was just passing through...”, etc. - was also a lie and was constructed by the EM in order to offer an apparently logical explanation for the test subject’s suspicion, thereby affecting the software’s automatic deciphering.

Our hypothesis that during the TrusterPro OL examination, certain truthful participants were diagnosed with a certain lying percentage, due to the fact that their cover story was a lie, despite the fact that their actual participation in the event was “innocent”.

(In the improved OL system, the examiner can mark the relevant sentence for analysis, thereby preventing the reception of irrelevant data and disruption of the relevant findings. The improved software will be tested in a similar research framework, as the one under which the basic software was tested).

We are assuming that the results of both the polygraph and the TrusterPro were less accurate than would have been obtained in a real situation, due to the problematics in examining mock crimes. These problematics originate from the lack of motivation of the test subjects on one hand, as opposed to a real situation, where the participants are highly motivated (the “guilty”

subject tries to avoid discovery, while the “innocent” subject fears from the false result).

An additional difference concerns the motivation level of the examiner, who feels greater personal stress in mock situations. This is because the results constitute immediate feedback, without the ability to “hide” behind technical or other explanations for the false results. Such a situation, which hints at the professional level of the examiner, often creates undue stress that affects the course of the examination in one way or another.

The distribution of errors between false positives (FP) and false negatives (FN) in the TP and the polygraph examinations is an additional case in point. It is evident that in the OL mode, the distribution between FP and FN is nearly 50%. In effect, there is no skewness in either direction in terms of the error. A possible explanation for this issue concerns the minimal involvement of the examiner in the examination process, the deciphering process and the decision making regarding the results.

In the Int mode, the distribution of error leans in the direction of FP, i.e. 70% of the total errors occurred among truthful test subjects, while only 30% of the total errors were made in the examination of guilty test subjects. If we take into consideration that the study was conducted on examinations of mock crimes, then this result serves to reinforce the diagnostic ability of the Int mode, since this distribution resembles the distribution that exists in real situations. If we compare this to the polygraph examination, we will see that the reverse is true, i.e. 60% of the total errors occurred in the examination of guilty test subjects (FN), while only 40% of the total errors occurred in the examination of truthful subjects (FP). This situation is the reverse of reality and may be explained by the fact that the examination was conducted on mock crimes. In fact, the level of motivation of the guilty test subject was low. In other words, their level of fear of discovery was low since there was no “penalty” associated with the discovery. In the control question method where the polygraph examinations were conducted, there is a tendency toward a greater number of errors in diagnosing guilty subjects during examinations conducted on mock crimes. Such a result is the typical effect of the control questions on test subjects that are allegedly “guilty” of mock crimes. This is in comparison with real situations, where the diagnosis of guilty subjects is better than that of truthful subjects.

In light of the above, it is evident that TrusterPro, in both tested modules, offers a degree of reliability that indicates that the developed instrument is suitable for field work, under reasonable restrictions and when being employed intelligently, in a manner compatible with the circumstances and events.

An additional study is planned for the future, which will examine a range of influences of various stress levels on the TrusterPro findings, while examining whether TrusterPro measures changes in the stress level of the test subject accordingly.

Legend:

D.I. – Deception indicated

N.D.I. – No deception indicated

INC – Inconclusive decision

F.P. – False positive (Deception was detected when the subject was truthful)

F.N. – False negative (No deception indicated when the subject was guilty)

References

“An analysis of voice responses for the detection of deception” By Victor L. Cestaro , P.H.D
& Andrew B. Dollins, P.H.D.

Barland, G.H. & Raskin, D.C. Validity and reliability of polygraph examination of criminal suspects (Report No. 76-1, U.S. department of justice).

Barland, G.H. & Raskin, D.C. An evaluation of field techniques in detection of deception. *Psychophysiology*, 1975, 12.321-330.

Bersh, P.J. A validation of polygraph examiner judgments. *Journal of Applied Psychology*. 1969, 53.388-403.

Horvath, F.S. & Reid, J.E. The reliability of polygraph examiner diagnoses of truth and deception.

Davidson, P.O. Validity of the guilty knowledge technique. The effect of motivation. *Journal of Applied Psychology*, 1968, 52.62-65.