Literature Survey of Structural Weighting of Polygraph Signals:

Why Double the EDA?

Raymond Nelson

Abstract

Electrodermal activity (EDA) is a useful and important source of information in psychophysiological detection of deception (PDD) testing. A variety of methods have been used to evaluate the relative contribution of EDA data, and results have been consistent throughout several decades of research in different laboratories. Information indicates that EDA data contributes approximately half of the information used to make effective classifications of deception or truth-telling in the CQT paradigm. This paper is a survey of existing literature on the correlation and weighting coefficients for EDA and other recorded data relative to PDD.

Introduction

Why double the electrodermal activity (EDA) scores in a polygraph setting? EDA is a useful and important source of information in psychophysiological detection of deception (PDD) testing. Scientific experiments are often used to study the strength of association between an unknown phenomena of interest and some observable data that serve as a proxy for the unknown phenomena of interest. In addition to studying the strength of association, test developers are also concerned with determining the optimal weighted combination of the measures/data that will maximize test effects such as test sensitivity, and specificity, or false-negative and false-positive errors. In the literature review that follows, summarized in Table 1, one can trace the development of the technology and study designs leading to the currently used weighting coefficients for EDA and other sensor data in the PDD testing context. Table 1 contains the studies in temporal order, the fitting technique type, the coefficients for the non-normalized data and the coefficients for the normalized data which leads to the current weighting system.

Methods

Published literature was surveyed for information relating measures of respiration, electrodermal (EDA), cardiovascular activity and vasomotor activity to psychophysiological detection of deception (polygraph) test outcomes. These signals have emerged as useful in the PDD testing context because they have been shown to be correlated with differences in deception and truth-telling, while contributing unique or additional variance to accurate classifications when used in combination. Criterion coefficients of interest, which describe the relationship between the data and the criterion state of deception or truth-telling, are summarized in Table 1. A number of different types of coefficients have been reported in various studies, including correlation coefficients, discriminate and regression coefficients and other structural coefficients. The optimal weights for each of the physiological measures was determined by the data from each of the included studies. The relative contributions to the optimal weights, as seen over time, converge to a scoring system with roughly half the weight given to EDA, i.e. twice that given to the other recorded signals.

Also included in Table 1 are the normalized coefficients. Normalization of values ensures that all information can be compared on a similar scale for which the sensor values will sum to one (1). The actual structural models can be expected to differ somewhat from this simplification. However, normalization in this manner can help readers to easily appreciate the relative strength of contribution of the different recording sensors.

Literature review

Kubis (1962) described an early study, funded by the U.S. Air Force, on the feasibility of using computers in polygraph testing. Discriminate functions were calculated for each of three polygraph test charts, the means of the three discriminate function are shows in Table 1. EDA accounted for over 60% of the diagnostic variance in that study. Kubis concluded that there was sufficient validity to the comparison question test – premised on the loading of physiological responses to relevant and comparison stimuli – to warrant confidence in the polygraph as a possible aide to a decision about whether or not to interrogate. Kubis noted the high degree of subjectivity and variability in manual polygraph feature extraction at that time, especially with regard to respiration and cardiovascular activity. Additionally, he commented on the weaker available computing technology at that time and recommended against attempts to develop computerized polygraph systems that would be intended to make on-the-spot decisions. Kubis recommended the use of computers in studying and developing polygraph features and polygraph feature extraction. One difference between the instrumentation used in this study and that of a modern polygraph instrument is that cardiovascular activity was recorded with an electronically amplified fingertip (mechanical circumference) plethysmograph, which is different than both the arterial pressure cuff and the photoelectric plethysmograph in use today.

Kubis (1964), reported the results of a second study on the feasibility of using computers in polygraph testing, also funded by the U.S. Air Force. This study also used a mechanical/circumference fingertip plethysmograph in lieu of the arterial cardio cuff sensor. Results of manual/numerical scores were reported along with the results of an objective measurement approach to feature extraction. Data shown in Table 1 are the concordance rate or frequency deceptive and truthful scores that concurred with the criterion states of the sample cases. Normalizing the proportions resulted in an EDA coefficient of .41, for summed numerical scores and .39 for discriminate scores.

Kircher (1981) used a sample of community participants to study the use of computers in the evaluation of PDD test data. Statistical and structural coefficients were reported for a variety of potential physiological measurements in PDD testing and analysis, including the three-chart means of standardized discriminate function coefficients for EDA, cardio and respiration data. Coefficients were included for both manual scores and computer/measurement scores. Computer scores also included a vasomotor sensor. EDA accounted for 50% of the variance in manual scores and 41% of the variance in computer scores for the guilty and innocent participants. Kircher concluded that the computer may be capable of offering improvements in polygraph data analysis.

Kircher (1983) reported the results of another study on the use of computers in lie detection and reported point-biserial correlation coefficients for both manual and computerized feature extraction with respiration, EDA, cardiovascular and vasomotor sensors. In that study, EDA accounted for 28% of the variance in manual scores and nearly 47% of the variance in computer scores for guilty and innocent participants. Kircher also reported the results of a discriminate function, for which the normalized coefficient for EDA was .46. Kircher reported accuracy ranging from 87.5% to 93.7%, with the computer algorithm slightly more effective at identifying programmed guilty subjects and manual scores slightly more effective at programmed innocent subjects, suggesting that physiological response patterns may be qualitatively different for deception and truth-telling. All point biserial coefficients in Table 1 are shown as r_{pb2} to facilitate more intuitive comparison of results from different studies.

Kircher and Raskin (1988) reported a comparison of human and computerized evaluations of polygraph data, including respiration, EDA, cardio and vasomotor sensors. Point-biserial correlations are shown as r_{pb2} in Table 1. For manual scores the EDA accounted for 26% of the variance. However, EDA scores may be strongly covariant with vasomotor activity, also an indicator of autonomic activity, and the vasomotor scores also accounted for 26% of the diagnostic variance. For computer scores, for which the vasomotor sensor did not contribute additional information and was not included in the structural model, the EDA accounted for 46% of the observed variance in deceptive and truthful outcomes for programmed guilty and innocent subjects.

Raskin, Kircher, Honts and Horowitz (1988) completed a field study that was funded by the National Institute of Justice, and reported point-biserial correlation coefficients for respiration, EDA, and cardiovascular activity. Point-biserial correlations are shown as rpb2 in Table 1. EDA data alone accounted for 53% of the observed variance in deceptive and truthful outcomes for the guilty and innocent examinees in the field sample.

Raskin and Kircher (1990) completed a study on the use of computer algorithms in polygraph data analysis and countermeasure detection. They reported the coefficients for a discriminate function that included respiration, EDA, and cardiovascular activity. EDA accounted for 55% of the information in the normalized discriminate function that optimized the separation of guilty and innocent study participants. Discriminate coefficients and the normalization are shown in Table 1.

Capps and Ansley (1992) completed a survey of scores assigned by field examiners for each recording sensor and each scoring feature using a seven-point numerical scale. As shown in Table 1, over 40% of the points were assigned to the EDA sensor data.

Harris and Olson (1994), in a patent filing for the Polyscore algorithm, described the coefficients of a logistic regression that included EDA, cardiovascular activity, and respiration (also included pulse line length and pulse rate though these are not shown in Table 1). The EDA weighting coefficient was 49% after normalizing only the information that is most similar to traditional polygraph feature extraction (respiration, EDA and cardiovascular activity).

Ansley and Krapohl (2000) completed a survey of examiner scores using the seven-position scoring method with confirmed field cases. Table 1 shows that EDA scores accounted for 55% of all assigned scores in the sample data.

Honts, Amato and Gordon (2000) completed a study on the outside issue question. The study report included correlation coefficients for each of four scorers for respiration, EDA and cardiovascular activity, in addition to the vasomotor sensor. Table 1 shows the correlations in the form of a coefficient of determination (r^2). The coefficient of determination is an estimate of the observed variation in deceptive and truthful outcomes for guilty and innocent examinees that is explained by each sensor alone (without the addition of the other sensors). The mean of the r^2 coefficients for the four scorers is shown in Table 1. The coefficient of determination for EDA was .39. After normalizing the coefficients to compare their relative strength, EDA produced a coefficient of .41 when the vasomotor data was included and .57 when vasomotor data was excluded.

Kircher, Krisjianson, Gardner and Webb (2005) studied the validity of various scoring criteria that were previously taught at the U. S. government polygraph school and other accredited polygraph training programs in the past. This study concluded that some of the scoring features in the past were unnecessary – leading to a reduction of scoring features by the Department of Defense (2006) to only those features that are supported by scientific evidence – and suggested that primary scoring features accounted for a majority portion of the diagnostic variance that is extracted from recorded polygraph data. Table 1 shows the point biserial correlations were reported for information extracted from each recording sensor. EDA data produced an r^2 of .45. When normalized with the other sensor coefficients the relative weight for the EDA was .51.

Nelson, Krapohl & Handler (2008). Described the development of the Objective Scoring System, version 3 (OSS-3), and reported the coefficients from a discriminate analysis. Table 1 shows the normalized structural weighting for the EDA data was .53.

Nelson (2018) described the evolution and development of auto-centering EDA solutions for field polygraph instruments and reported the point-biserial correlation coefficient for of EDA and the criterion state of deception and truth-telling. The r^2 (coefficient of determination) for auto-centered EDA data was .49 and is shown in Table 1. The r^2 for manually centered EDA was .48, suggesting that EDA accounts for approximately half of the variation in deceptive and truthful scores for guilty and innocent examinees.

Nelson [in press] reported the results of a structural weighting function for respiration, EDA, and cardiovascular activity, computed with a simple genetic algorithm. A genetic algorithm is a simple form of machine learning (also known as artificial intelligence) based in the principles of genetics and evolution: random solutions, survival of the fittest, recombination, mutation, and generational

Table 1. Criterion coefficients for respiration, EDA, cardiovascular activity and vasomotor activity.

	Type of coefficient	Non-normalized coefficients				Normalized coefficients			
		Respiration	EDA	Cardio	Vasomotor	Respiration	EDA	Cardio	Vasomotor
Kubis (1962)[1]	df	.18	1.0	.419	-	.12	**.62**	.26	-
Kubis (1964)[1] summed	Prop. correct	.55	.94	.82	-	.24	**.41**	.35	-
Kubis (1964)[1] discriminate	Prop. correct	.56	.85	.76	-	.26	**.39**	.35	-
Kircher (1981) Numerical	df	.28	.65	.37	-	.22	**.50**	.28	-
Kircher (1981) Computer	df	.35	.73	.43	-.27	.20 (.23)	**.41 (.48)**	.24 (.28)	.15
Kircher (1983) Numerical	r_{pb}^2	.32	.37	.28	.36	.24 (.33)	**.28 (.38)**	.21 (.29)	.27
Kircher (1983) Computer	r_{pb}^2	.24	.56	.19	.19	.20 (.24)	**.47 (.57)**	.16 (.19)	.16
Kircher (1983) Table 9	df	.37	.68	.17	.25	.25 (.30)	**.46 (.56)**	.12 (.14)	.17
Kircher & Raskin (1988) Manual	r_{pb}^2	.57	.61	.53	.60	.25	**.26**	.23	.26
Kircher & Raskin (1988) Computer[3]	r_{pb}^2	.30	.59	.37	-	.24	**.47**	.29	-
Raskin, Kircher, Honts & Horowitz (1988)	r_{pb}^2	.15	.53	.48	-	.13	**.46**	.41	-
Raskin & Kircher (1990)	df	-.26	.78	.37	-	.18	**.55**	.26	-
Capps & Ansley (1992)	total scores	2537	3805	3074	-	.27	**.40**	.33	-
Harris & Olsen (1994)[2]	β	-2.6	5.5	3.1	-	.23	**.49**	.23	
Ansley & Krapohl (2000)	frequency	3,455	10,109	4966	-	.19	**.55**	.26	-
Honts Amato & Gordon (2000)	r^2	.12	.39	.17	.26	.13 (.18)	**.41 (.57)**	.18 (.25)	.28
Kircher, Krisjianson, Gardner & Webb (2005)	r_{pb}^2	.18	.45	.26	-	.20	**.51**	.29	-
Nelson, Krapohl & Handler (2008)	df	.629	1.753	.920	-	.19.	**.53**	.28	-
Nelson & Handler (2013)	r_{pb}^2	.19	-	-	-	-	-	-	-
Nelson (2018)	r_{pb}^2	-	.49	-	-	-	-	-	-
Nelson [in press]	proportion	-	-	-	-	.12	**.54**	.34	-

1 These studies used an electronically amplified fingertip plethysmograph instead of the traditional cardio cuff sensor. The fingertip plethysmograph is superseded in contemporary polygraph systems by the photoelectric plethysmograph that is similar to a medical pulse oximeter. Whereas a medical pulse oximeter can acquire and record information on pulse rate, respiration rate and oxygen saturation, the photoelectric plethysmograph in polygraph testing is used to record changes in vasomotor activity.

2 The logistic function also included pulse and pulse line length, though these are not shown in order to simplify the illustration of the relative contribution of EDA, cardiovascular and respiration activity.

3 The multivariate weighting function also included EDA recovery time (duration) and EDA burst frequency (complexity). The discriminate function was: EDA amplitude = .77, EDA recovery time = .27, EDA burst frequency = .28, cardio amplitude = .22 and respiration = -.40.

improvement. Results from optimization with the genetic algorithm are shown in Table 1. When applied to respiration, EDA and cardiovascular data, EDA data accounted for 54% of the diagnostic variance in a sample data of confirmed field polygraphs.

Discussion

This paper is a literature survey of the development of structural weighting coefficients for respiration, EDA, cardiovascular, and vasomotor activity signals used in PDD testing. Nineteen different coefficient functions from 14 different studies are shown in Table 1, along with the results of two additional studies that reported the criterion coefficients for individual sensors. Also shown in Table 1 are the normalized coefficients. Normalized coefficients are a simplification of multivariate structures but offer the advantage of easier and more intuitive comparison of different types of coefficients.

The included studies cover a wide time span, from 1962 to the present. Published studies have employed a variety of methods to evaluate the strength of relationship and structural contributions of scores from different PDD recording sensors and the criterion states of deception and telling. Although there is some variability in different estimates of weighting coefficients there is obvious consistency in that EDA data accounted for greater proportion of diagnostic variance than other sensors for all studies included in Table 1. The mean of all normalized coefficients for the EDA data in Table 1 was .45. when the vasomotor sensor was included in the normalization, and .49 without the vasomotor sensor.

Vasomotor activity sensors have been used inconsistently and are not included in many studies. When it is included, the normalized proportion of the vasomotor sensor has varied from .15 in the numerical scores of Kircher (1981) to .28 in the normalized correlations of Honts, Amato and Gordon (2000). Vasomotor activity was not included in the discriminate function of Kircher and Raskin (1988), though the point-biserial coefficient was reported for the manual scores in this study. This suggests that vasomotor activity may not have contributed additional diagnostic variance to the computer model even though the vasomotor data is correlated with differences between deception and truth-telling in the PDD testing context. Reasons for this are not completely understood and may be incompletely explored. It is reasonable to assume that vasomotor activity would have been included in a computer function if it contributed additional diagnostic variance. It is possible that vasomotor activity may covary strongly with both cardiovascular activity, and EDA, and this may be related to the absence of the vasomotor data in the discriminate function. Further research is needed in this area.

Of the 19 normalized functions, 16 of them produce an EDA weighting coefficient over .4. None of the normalized coefficients for the other sensors exceeded that of the EDA. However, one study, involving the manual scores of Kircher and Raskin (1988) included a vasomotor coefficient that equaled that of the EDA (.26). One other study, involving the manual scores of Kircher (1983) reported a vasomotor coefficient (.27) that nearly equaled that of the EDA (.28).

An obvious limitation of this project is that no attempt was made to test the significance of observed differences between the included studies. Also, no attempt was made to test the difference between the sensor data within the included studies. Another, necessary, caution is in order when attempting to interpret normalized correlation coefficients. This is because correlation coefficients are calculated for individual sensors and does not account for covariance between the sensors. Some of the included studies did report a structural model, and these may offer better information than simple normalization of the reported coefficients.

Finally, it is important to remember that neither EDA, nor any of the other sensor data, are expected to be a perfect, deterministic, indicator of deception and truth-telling. EDA is often discussed in the context of the sweating metaphor. However, just as EDA is not synonymous with deception, EDA is also not synonymous with sweating. Both EDA and sweating are associated with increased activity in the autonomic nervous system, and the array of polygraph recording sensors is almost uniformly autonomic. However, neither sweating nor autonomic activity are synonymous with or deterministic of deception. That is autonomic activity and sweating can occur for other reasons. Sweating is merely a convenient metaphor for EDA.

EDA data in field polygraph testing is measured using electrical measurements: Ohms or Siemens. However, EDA is not synonymous with electrical resistance or electrical conductance. That is, non-human objects can have electrical resistance and electrical conductance without EDA. EDA is a complex phenomenon for which electrical measurements are a convenient and expedient form of data acquisition. Just as electrical resistance and electrical conductance are a proxy for EDA, EDA itself is a proxy for autonomic activity, while autonomic activity is a proxy for deception. Proxy information is not adequate alone and may be more adequate when combined with other information. Other measurement technologies, besides resistance and conductance, exist for EDA data.

If the normalized coefficients from these studies are interpreted as an indicator of the proportion of test scores and test results that is explained by each recording sensor relative to the other sensors, cardiovascular activity data may account for approximately 30% of observed PDD results, while respiration data may account for approximately 20% of observed results. EDA data may account for approximately 50% of observed PDD scores and test results. Automated computer scoring algorithms have frequently made use of the differences in the contributions of the different recorded PDD signals. Also, differences in structural weight or contribution can be observed in manual PDD scoring methods based on the seven-position Likert-type scale (Likert, 1932). However, manual scoring methods that make use of the unweighted three-position scale may be ignoring some of the diagnostic variation it recorded PDD test data.

One important, and sometimes easily overlooked, difference between seven-position and three-position scales in PDD test data analysis is that the seven-position scale is a Likert-type scale – intended to transform subjective information to numerical values – whereas the three-position scale can be characterized as an objective rank scale. In other words, differences in seven position scale values are subjective or arbitrary (i.e., without mathematical proof as to the selection of differences in scale values) whereas three-position scale values can be subject to objective mathematical proof as to all differences in scale values. Although seven position scores can achieve a similar approximation, seven-position scores are likely to remain less reliable than three-position scores in field practice, due to the lack of theoretical and mathematical proof as to differences in seven-position scale values – leading to either subjectivity or arbitrariness (i.e., arbitrary use of mathematical ratios) in score assignment, or to a non-trivial optimization problem that will require a volume of high quality data and analytic effort. Doubling the EDA scores of three-position manual scores is a simple and objective way to closely and objectively approximate the optimal structural solution that can be achieved through more complex statistical methods.

Weighted three-position EDA scores make use of long-standing knowledge about differences in the structural contribution of different PDD signal, do so in a manner that does not introduce additional subjectivity, arbitrariness and unreliability to the analytic process. It is therefore not surprising that some non-parametric feature extraction and numerical transformation methods, such as the Objective Scoring System (Krapohl, 2002; Krapohl & McManus, 1999) and the ESS/ESS-M (Nelson, Krapohl & Handler, 2008; Nelson et al., 2011; Nelson, 2017), have reported some advantages in manual scoring when weighting the EDA data more than the other sensor data. Although there are some known advantages to data analytic and machine-learning methods that are deliberately naive as to the structural contribution of different signals – especially in the early stages of the development of analytic solution – multivariate solutions that can make use of available knowledge about the relative strength of different signals have ultimately tended to be more powerful or effective. Continued interest is warranted in the differences in correlation and structural contribution of different PDD recording sensors and the potential for optimization and improvement of PDD test effectiveness that may be achieved through the strategic use of naive and weighted structural solutions in PDD analytics.

References

Ansley, N. & Krapohl, D.J. (2000). The frequency of appearance of evaluative criteria in field polygraph charts. *Polygraph, 29,* 169-176.

Capps, M. H. & Ansley, N. (1992). Numerical scoring of polygraph charts: What examiners really do. *Polygraph, 21,* 264-320.

Department of Defense (2006). *Federal psychophysiological detection of deception examiner handbook.* [Reprinted in Polygraph, 40(1), 2-66.]

Harris, J. C. & Olsen, D.E. (1994). *Polygraph Automated Scoring System. Patent Number: 5,327,899.* U.S. Patent and Trademark Office.

Honts, C. R., Amato, S. & Gordon, A. (2000). *Validity of outside-issue questions in the control question test.* Final Report on Grant No. N00014-98-0725.

Kircher, J. C. (1981, June). *Computerized chart evaluation in the detection of deception.* [Masters Thesis.] University of Utah. Salt Lake City.

Kircher, J. C. (1983).C*omputerized decision making and patterns of activation in the detection of deception.* [Doctoral dissertation.] University of Utah, Salt Lake City.

Kircher, J. C., Kristjansson, S., Gardner, M. K., & Webb, A. K. (2012). *Human and computer decision making in the psychophysiological detection of deception. Polygraph, 41(2),* 77-126.

Kircher, J. C. & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73,* 291-302.

Krapohl, D. J. (2002). Short report: Update for the objective scoring system. *Polygraph, 31,* 298-302.

Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph, 28,* 209-222.

Kubis, J. F. (1962). *Studies in Lie Detection: Computer Feasibility Considerations.* Final report, RADC-TR 62-205, Contract AF 30(602)-2270. Air Force Systems Command, U.S. Air Force, Griffiss Air Force Base. New York: Rome Air Development Center.

Kubis J. F. (1964). *Analysis of Polygraphic Data.* Final report, U.S. Air Force contract RADC-TDR-64-101, AF30(602)-2634. Air Force Systems Command, U.S. Air Force, Griffiss Air Force Base. New York: Rome Air Development Center.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140, 5-55.*

Nelson, R. (2017). Multinomial reference distributions for the Empirical Scoring System. *Polygraph & Forensic Credibility Assessment, 46 (2).* 81-115.

Nelson, R. (2018). Electrodermal signal processing: a correlation study of auto-centered EDA and manually-centered EDA with the Criterion State of Deception and Truth-telling. *Polygraph & Forensic Credibility Assessment, 47(1)*, 53-65.

Nelson, R. [in press]. Structural Optimization of Respiration, EDA and Cardio Activity Using a Genetic AI. *Polygraph & Forensic Credibility Assessment,* [in press].

Nelson, R., & Handler, M. (2013). Pneumograph signal processing and feature extraction. *Polygraph, 41(3),* 163-172.

Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B. & Oelrich, M. (2011). Using the Empirical Scoring System. *Polygraph, 40,* 67-78.

Nelson, R., Krapohl, D., & Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph, 37,* 185-215.

Raskin, D. C. & Kircher, J.C. (1990, May 2). *Development of a Computerized Polygraph System and Physiological Measures for Detection of Deception and Countermeasures: A Pilot Study.* Contract 88-L655300-000. Scientific Assessment Technologies, Inc. [Reprinted in Polygraph,30, (3), 153-162.]

Raskin, D. C., Kircher, J. C., Honts, C. R. & Horowitz, S.W. (1988). *A study of the validity of polygraph examinations in criminal investigations*. Final Report, National Institute of Justice, Grant No. 85-IJ-CX-0040. [Reprinted in Polygraph & Forensic Credibility Assessment , 2019, 48 (1)].