Development of a Physiological Monitoring System for Credibility Assessment

Hana K. Ulman
Department of Chemical and Biomedical Engineering, West Virginia University

SPECIFIC AIMS: Humans tend to operate under the default presumption of honesty, which enables efficient communication and trust for our species; however, the tendency to presume honesty makes us vulnerable to acts of deceit. The existence of deception is traditionally identified through acts of cognitive and physical arousal that occur due to behaviors characteristic of intentional lying[19]. Further, the monitoring of fluctuations in an individual's physiology can provide more objective means of measure than human intuition alone. The advent and proliferation of physiological monitoring technologies has enabled user-centered design solutions through the establishment of a biofeedback loop, wherein physiological data are garnered from a device(s), processed, and fed as input parameters for classification. The polygraph is a biofeedback system that serves to quantify deception-induced stress via the measurement of heart rate (HR), respiration rate (RR), electrodermal activity (EDA) and blood pressure (BP)[32]; however, it has demonstrated anecdotal success, with an inconclusive rate of up to 20%[24]. Despite the advancements in wearable and off body monitoring devices, little progress has been made to integrate multiple data streams in a biofeedback system that detects truth from lie. The long-term goal is to collect physiological and questionnaire data (N=100) that will serve as inputs for a semi-autonomous model that can effectively distinguish truth vs. lie, with higher accuracy than human intuition (~54%)[3] or polygraph (~90%)[24] and provide a percent-based likelihood of truthfulness. The overall objective is to appropriately characterize the physiologic responses that are a result of deception induced stress, validate the hallmark metrics obtained by polygraph against wearable and off-body technologies, and develop a model that accurately classifies truth from lie. I hypothesize that a multi-modal approach which utilizes less obtrusive sensors than the polygraph will result in higher model accuracy, through the incorporation of sensors that aim to capture nonstrategic behaviors and exploits the proposition of a more exacerbated response from baseline due to the notion that one is being less monitored. The *rationale* is motivated by the need for credibility assessment in non-traditional laboratory settings, where the feasibility to deploy conventional techniques requires an overly invasive sensor setup, a highly controlled environment, and experienced personnel[22]. To develop this system and thereby attain the overall objective, I will pursue the following specific aims:

Aim 1: Collection and Characterization of Physiological Signals Related to Deception-Induced Stress:

Investigate how deception related stress drives the elicitation of cardiorespiratory, musculoskeletal, and thermal responses and oculomotor behavior with respect to the related regulatory autonomic components and collinearities between various signals and metrics that occur during active deceit. The following sub-aims are detailed in Table 1 below and intend to capture metrics reflective of cardiorespiratory, thermal, musculature and oculomotor responses of the following through the implementation of the appropriate signal processing and analysis techniques.

Table 1. Sub-aims used to characterize differences in physiologic stress response

Aim	Measure	Previous Findings During Active Deceit
1A: Cardiorespiratory	HR, RR, heart rate variability (HRV)	Inc. in HR[58] and dec. in respiration rate[11].
1B: Thermal Response	Surface skin temperature, skin conductance level (SCL), skin conductance response (SCR) components, and frequency.	Inc. in EDA and phasic frequency[34,38,51] and inc. in surface skin temperature in the periorbital region.
1C: Muscular Activation	Postural sway, muscle contraction and activation.	Inc. in postural rigidity[5,31].
1D: Oculomotor Behavior	Pupil diameter, blink rate, fixation points, saccadic velocity	Inc. in pupil diameter[20] and differences in blink patterns[12,60].

Aim 2: To Evaluate the Accuracy of Minimal Contact & Off-Body Systems Against Traditional Polygraph.

There exist a multitude of suggested alternatives to polygraph, however, the validation and subsequent efficacy of each sensor to its gold-standard counterpart is often not assessed. Determination of a sensor's accuracy will provide insight into the utility of its incorporation in the final system.

Exploratory Aim: To determine whether an individual's perception of self or behavioral motivational systems impact the corresponding changes in physiology which may be more or less pronounced.

Aim 3: To Develop a Multi-Modal Deception Detection Algorithm.

The capture of multiple congruent signals provides an added layer of redundancy to ensure detection of the various compensatory, heighted states of arousal that occur during active pursuits of deceit that may vary across individuals[10]. These instances will be identified through an anomaly detection algorithm, such as a long-short term memory model.

The development of such a system significantly advances the field of credibility assessment, with respect to deployment in field-based security screenings federally and within local communities. Further, the strategies used for anomaly detection can be extrapolated to other instances of biomedical state classification.

RESEARCH STRATEGY
I. SIGNIFICANCE

For centuries, humankind has searched for an infallible lie detector; however, it should come as no surprise that there are no fool-proof methods to determine if an individual is lying. Our ancestral environment did not prepare us to be astute lie detectors, at least, not without the aid of tools and techniques to provide more accurate means of measure than intuition alone. In fact, human deception detection is typically no better than chance itself[42]. For instance, when asked to discern truth from falsehood, humans accurately predict lies as false a mere 47% of the time and truth as nondeceptive 61% of the time[7].

Deception is a complex social behavior driven by higher order cognitive functions that are modulated by personality, environmental, and situational factors. Theories of deception detection posit that it is more cognitively demanding to actively deceive than to tell the truth, evoking a chain of mechanisms that results in psychological arousal that in turn creates a physiological response. Further, the cognitive processes associated with deception demand resources and hence produce somatic and autonomically driven changes that can be captured and used as an indirect indicator of deception-induced stress. The detectable patterns that occur during states of active deceit are termed, leakage, which refers to the involuntary physiological or behavioral cues that are found when lying. Leakage is quantified through the deviations in certain autonomic variables such as, increase in HR and skin conductance, or decrease in RR. Thus, the identification and characterization of the inadvertent behaviors or responses that reliably detect deception, is the primary focus of deception research.

The polygraph and concealed information test (CIT) are deemed the current gold-standards for deception detection and are widely used in law enforcement, criminal investigations, pre-employment screenings, and national security systems. Polygraph relies on a *strong* emotional response to a stimulus and subsequently measures deviations in the *arousal* of involuntary bodily reactions related to skin conductance, respiration, blood pressure, and cardiac activity[57]. It presupposes that the individual will have a consistent and measurable physiologic response when actively trying to deceive the operator. Traditional polygraph measures changes in an individual's autonomic arousal by measuring parameters such as HR, RR, BP, and EDA; these measurements are obtained using electrocardiography (ECG), pneumography, blood pressure cuffs, and finger plates. The CIT utilizes the same sensors as polygraph, but incriminating information is dispersed amongst irrelevant information, that should hypothetically evoke a response in guilty, but not innocent individuals. It is postulated that this form of lie detection elicits responses that are a byproduct of the physiological reactivity associated with *orienting* a response in accordance with the perceptible quality of the answer to a relevant question. When administered in highly controlled settings, the polygraph has an estimated accurate rate of 0.90; however, many studies have demonstrated that the polygraph is subjugated to an overestimation of test accuracy, with high false positive and false negative rates[33].

Additionally, there are some limitations surrounding the applicability of polygraph in general settings; for instance, polygraph requires professional experts to score the data and attach the sensors; the sensors require the subject to be tethered to the system at a fixed length and can cause heightened levels of anxiety from the onset of the interview. Furthermore, the polygraph suffers from inadequate construct validity, susceptibility to countermeasure, test-retest reliability, inter-rater reliability between examiners, and the variability associated with the interaction between the examinee and the polygraph examiner. The drawbacks of polygraph have incentivized the substitution of other technologies that provide more reliable and autonomized means of measure. The implementation of other methodologies such as functional magnetic resonance imaging[17] and event-related potentials (ERPs)[41] aimed at measuring neural activity associated with deception have been explored; however, both techniques require a profoundly controlled environment, amongst highly specialized and expensive equipment and technicians to interpret the results.

Countermeasures describe the method used by prevaricators, wherein the individual may deliberately alter their normative behavioral pattern to subsequently affect their body's psychophysiological response and thus, confounds, or at least significantly complicates deception detection strategies. The two most common forms of countermeasure-based approaches are "probes as low salient" and "irrelevant as high salient", that involve the subject dampening their response when the incriminating stimulus is present and heightening their arousal during irrelevant responses, respectively. A previous study determined that if appropriate countermeasures are taken, nearly 50% of guilty individuals can successfully fool a polygraph by either consciously increasing their cognitive effort (e.g., counting backwards from 100 by seven) or state of physical arousal (e.g., biting the tongue) during irrelevant responses[21]. These methods are reliant upon artificially counteracting the neural and bodily response that are typically evoked during deceit, which is dependent on volitional, conscious control[4].

To circumvent the challenges found in countermeasures, several studies have proposed an approach that involves rapid serial visual presentation (RSVP), where questions or stimuli are presented on the fringe of awareness, at a rapid rate (<10 seconds), to confound strategies based upon volitional control[4]. Understanding the rate at which information is processed and resources are consumed provides insight into the cognitive or operational demand needed to answer a question either truthfully or untruthfully. Despite the encouraging findings surrounding the implementation of RSVP, it is only efficacious with methods and metrics that are high in temporal resolution and modulated by *instantaneous stress conditions*, whereas other physiological responses exhibit a delay at stimulus presentation. For instance, rapid stimulus presentation would be inconsequential when measuring metrics such as HR, BP, or EDA, which are more significantly modulated by *sustained stress conditions*. These parameters are regulated by a downstream signaling pathways that are not instantaneous processes, which is thus reflected as gradual fluctuations in physiology.

Characterization of both *sustained and instantaneous* stress responses is crucial for successful deception detection strategies. However, measuring *instantaneous stress responses* presents quite the challenge for a multitude of reasons. Foremost, determining the appropriate metric that provides such detailed granularity is difficult to come by. Further, the sensor to obtain the desired metric must be sophisticated enough to be able to detect the minute fluctuations that occur within an acute time window. ERPs have demonstrated preliminary success at detecting deception but are constrained to laboratory environments and do not translate well to field conditions due to extraneous noise and need for a highly specialized expert. In recent years, thermal imaging (TI) and eye tracking techniques have gained traction as viable detection deception strategies because they can capture both forms of stress (*sustained and instantaneous)* and can be monitored within an acute time window, with reported classification accuracies of 91.7%[38] and 90%[3], respectively, when used alone. However, to our best knowledge no studies have examined the efficacy of both sensors used in unison or in addition to minimal contact, wearable devices.

Therefore, there exists a **critical need** for a system that integrates multiple physiological monitoring technologies that expounds upon the current methodologies used in deception detection through the incorporation of additional, unobtrusive sensors, that aim to measure instantaneous stress responses. Further, it is hypothesized that the inconspicuous nature of wearable and off-body sensors may result in a more dampened baseline, which in turn, results in a more exaggerated deviation from normative behavior and thus, easier anomaly detection. The integration of multiple bio-sensing technologies adds a layer of redundancy, with the intent to capture the compensatory behaviors or responses that occur as a result of deception related stress that may appear differently across individuals or remain unaccounted for in strategies of countermeasures. No system will be 100% foolproof, especially when dealing with the variabilities associated with human behavioral decision-making; nevertheless, the intent is to gain a definitive statistical advantage in every encounter to help facilitate decision-making surrounding an individual's truthfulness.

II. INNOVATION

There has been a renewed interest in lie detection strategies that has been largely motivated by the technological advancements made in physiological monitoring systems[37]. The innovation of the following approach challenges current deception detection paradigms by leveraging the technological advancements made in less-intrusive physiological monitoring systems and by utilization of:

- A study design that overtly encourages participants to answer deceptively, thus increasing the net magnitude of **physiological reactivity**. Based upon prior research, four categories of physiological parameters that are hypothesized to suggest truthfulness will be evaluated, including cardiorespiratory activity, thermal activity, musculature activation, and changes in oculomotor behavior.
- The **integration** of correlative **personality research,** with traditional, **experimental,** deception detection research provides a more wholesome perspective of inter-individuals differences that drive deception strategies and outcomes.
- A sophisticated sensor suite that includes traditional polygraph sensors for **validation**, and wearable / off-body monitoring counterparts that purport to measure the physiologic measures less obtrusively. While some deception detection strategies have addressed the need for more innovative solutions than

polygraph, few have extensively validated the efficacy of their sensors against measures traditionally captured by polygraph.

- Advanced **signal processing** and **feature extraction** techniques that implement mathematical formulations (time domain measures, wavelet analysis, spectral analysis) to obtain information that is not inherently obvious but provides the underlying psychophysiological context that may aid in understanding the processes that drive deception induced stress.
- A systematic approach to determine which **type** of machine learning or deep learning **model** and with what **combinations of signals** yield the highest classification accuracy.
- The development of a **semi-autonomous model** that not only classifies truth from lie but provides a percent-based likelihood of truthfulness. Furthermore, the autonomy of this setup provides a more objective means of measure, that removes operator bias which could be affected by gender, race, culture, age, etc.

The current methodology set forth by polygraph requires a highly controlled-for environment, extensive sensor setup, and subsequent hyper-awareness of the participation, associated with the invasive sensor setup. Incorporation of less-intrusive, wearable and off-body monitoring sensors may elude the participant to believe that they are being 'less monitored' and thus have a more exaggerated response from their normative baseline. The assumed trade-off between accuracy and convenience with existing polygraph methodologies will be assessed, with the opportunity to increase both accuracy and utility with the models developed herein. In summary, this proposal unites the fields of physiology, psychology, biomedical engineering, and computer science to help facilitate critical decision making regarding the implementation of such a system that addresses a critical gap in credibility assessment scenarios.

III. APPROACH

Overview of Approach: The specific aims will be achieved in a 2-year timeframe (please refer to Figure 6 for further detail) based on the approach outlined in Figure 1. Successful completion of the project will require the following: data collection (N=100), preparation (segmentation, signal processing, etc.), validation of off-body and minimal contact sensors against their traditional counterpart, identification of physiologic variables that are elicited during deception induced stress, algorithm selection, optimization and tuning, training, validation, and testing.

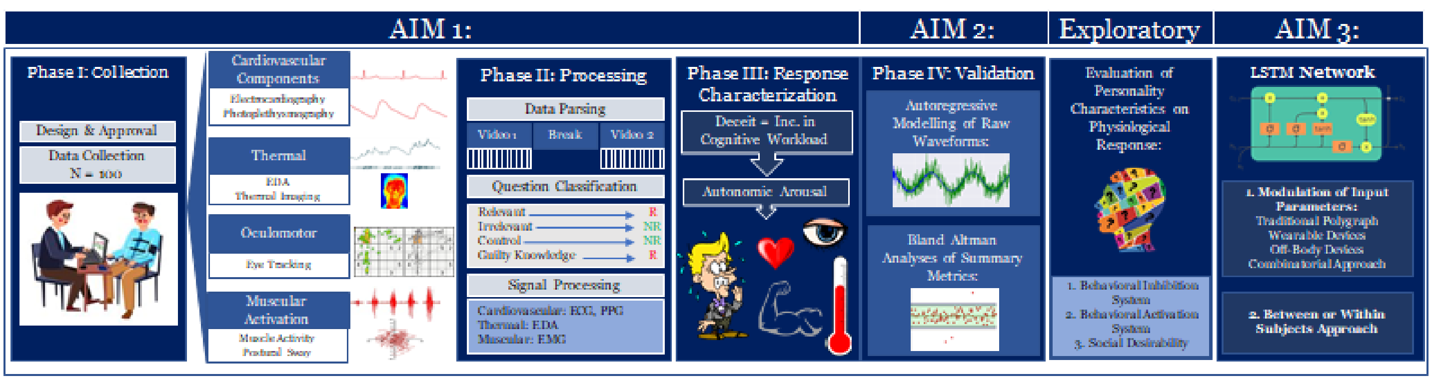

Figure 1. Proposed Approach for Data Collection, Processing, Validation, Analysis, and Model Development

Aim 1. Collection and Characterization of Physiological Signals Related to Deception-Induced Stress

The **ancillary objective** of this aim is to collect and determine the presence of physiologic patterns that are characteristic of deception induced stress. Based upon previous research, these four physiological systems combine novelty with a high potential for accuracy, suggesting promising options for inclusion into a future lie detection system. Further, existing research has demonstrated the utility of these types of parameters in reflecting behavioral changes associated with lying. Recent advances in sensor technology augment the implementation of improved sensors during credibility assessments; however, the simultaneous utilization of numerous sensors for multi-system monitoring, that are capable of improving existing assessment options, have been poorly researched.

Hypotheses:

- Deception is a complex social response; therefore, we should expect that individuals will exhibit different types of compensatory actions to conceal their lies; this could result in the attenuation or amplification of a signal or physiologic parameter.
- When answering a question deceptively, participants will show patterns of behavioral and physiological responses that significantly differ from that of nondeceptive answers. Expected outcomes related to deceit related behavior include increased pupil diameter, vasomotor tone, heart rate, EDA response, and decrease in HRV and body movement.
- Some individuals may try to utilize countermeasure strategies to deter lie detection which may result in a FP or FN in some signals; however, the incorporation of additional sensors that covertly collect data may pick up the compensatory behaviors or cues that these individuals are unaware of doing.

Data Parsing: Prior to the calculation of the various metrics found below, the data must first be pre-processed and parsed according to the 30 second window response time allotted for each question. Correspondingly, the categorization of each question will be determined for each response time to one of the following: control, relevant, irrelevant, and guilty knowledge.

Physiological Systems of Assessment: Each of the following sub-aims found below explore the relevant physiological responses that could be modulated by active deceit, the sensors used to acquire the data, and the appropriate signal processing methodologies that should be implemented to acquire a clean signal. Effective pre-processing techniques enable the obtention of a more refined version of the signal, devoid of noise or unwanted artifacts, which in turn amplifies the relevant physiological response of interest. Physiological signals represent non-stationary, stochastic processes that cannot be cleaned with traditional noise cleaning methodologies. Each of the following sensor types requires its own unique signal processing technique; however, the main objective for processing remains the same, categorization of the peaks that relate to an underlying physiologic response.

Aim 1A: Cardiorespiratory Activity:

Heart Rate (HR; pulse rate (PR)) and Heart Rate Variability (HRV; pulse rate variability (PRV)): HR (and PR) and HRV (and PRV) are involuntary responses of the ANS and purport to provide a quantification of cardiovascular system dynamics, homeostatic balance, acute stress, recovery, and dysfunction[47,48]. Modulation of HR over time are indicative of stress adaption, where lower values suggest the body is at rest (negligible stressors) and higher values reflect activities that require higher metabolic demand related to physical, cognitive, or emotional stressors[52]. The minimum, maximum, average, and phasic change (delta) in HR, and the magnitude of the accelerative and decelerative HR response will be calculated for each 30-second epoch. Similar to HR, HRV is driven by the ANS but represents more complex neurocardiac functions related to regulation of the parasympathetic and sympathetic nervous systems that are modulated by, but not limited to, acute stress, environmental and psychological challenges, and metabolism. Traditionally, HRV parameters are calculated in 5 min epochs; however, ultra-short-term measures of HRV within 30 second windows have been accepted for certain measures of HRV; namely, the root mean square of successive differences (RMSSD)[47]. Of late, RMSSD has gained extensive popularity, as it is the only HRV metric that is uninfluenced by respiratory rate[14,43] and thus less affected by strategic countermeasures. RMSSD values tend to decrease in the presence of acute stress[6,56]. I hypothesize that when answering deceitfully, there will be an increase in HR and decrease in RMSSD.

As previously stated, HR and HRV metrics can be obtained using two different types of sensor modalities that require vastly different methodological approaches as a consequence of the shape and characteristics of the obtained signal.

Electrocardiography (ECG): ECG devices measure the electrical activity (specifically, atrioventricular depolarizations) of the heart via electrodes placed on the skin's surface. The signal captured by ECG outputs as a series of temporally driven PQRST waveforms, which can be used to derive an individual's HR, inter-beat-interval (IBI), and HRV[48]. The R peak is the most characteristic component of an ECG waveform and denotes a heartbeat, whereas the time between subsequent R peaks represents is used to obtain IBIs. Despite the widespread use of ECG in clinical and consumer settings, the signals are subjugated to noise related to baseline wander (BW; low frequency artifact that arises from breathing or electrode interference)[26], powerline interference (movement in leads that are attached to the electrodes), muscle activity, and movement. Removal of BW is a standard practice and involves a linear, time-invariant, high pass filter with a cut-off frequency under 0.05 Hz[23]. Powerline interference is characterized by 50-60 Hz sinusoidal interference that superimposes the low frequency P and T waves of the ECG signal[25]. It can be removed by applying a second-order finite impulse response filter with a relatively selective notch, that narrows the band of admissible frequencies[13]. The most challenging aspect of ECG filtering is removal of muscle activity that tends to completely overlap the PQRST complex. However, due to the repetitive nature of ECG and the limitation of movement in this present study design, the signal can be denoised via ensemble averaging. Following artifact removal, a peak detection and tracking algorithm can be used to identify and quantify the number and time between successive R peaks.

Photoplethysmography (PPG): PPG devices measure volumetric changes in distal arterial blood flow to infer HR dynamics and is very similar in nature to ECG; however, the signal outputted by the PPG device is more sinusoidal in nature and lacks distinct R peaks. PPG devices are often worn on distal locations of the body such as the wrist or finger, and are therefore, significantly corrupted by motion artifact. Luckily, another fundamental sensor incorporated within PPG based devices is an accelerometer, which can be used to remove motion due to noise. First the PPG signal will be band-pass filtered using a 4^{th} order Butterworth Infinite Impulse Response (IIR) filter from the ranges of 0.3-5 Hz[53]. The accelerometer data will then be ingested as an input for singular value decomposition to be used as a reference for an adaptive least-mean square (LMS) filter followed[1]. Following categorization of artifact due to movement, the smoothed signal will then undergo peak detection. However, due to the nature of this study, I do not expect to see the PPG signal contaminated by significant amounts of noise.

Aim 1B: Thermal Responses:

Changes in temperature related to the amount of infrared radiation emitted or variance in conductance provide insight regarding an individual's state of physical arousal. During active deceit, surface skin temperature in the periorbital region and skin conductance of the hands, tends to increase.

Thermography**:** Thermal imaging (TI) captures infrared radiation (heat) emitted by an object. It is theorized that TI can bridge the gap between the behavior aspects of deception detection (emotion changes, facial expressions, etc.) and traditional ANS measures obtained in polygraph[38]. Traditionally, TI is used to assess the temperature of an object; however, thermodynamical modeling of the images can also be used to estimate subcutaneous and cutaneous blood flow, breathing rate, relative tidal volume, and cardiac pulse[27,50]. In recent years, a growing body of research has examined thermal imaging as a non-contact alternative for monitoring deception. The notion behind using TI is driven by the physiological reaction due to the acute stress associated with deception. During a state of sympathetic arousal, there is a rapid redirection of blood flow to subcutaneous vasculature, which results in a surface skin temperature change[11]. Previous studies have found that skin temperature increases with guilt, specifically around the two periorbital regions of the face[35,50]. I hypothesize that increases in blood flow as a result of active deceit should cause blood from the core regions of the body to flow to the periorbital regions, resulting in greater temperature changes. To appropriate analyze the TI data, the periorbital region must be identified within the images by placing a bounding box that appropriately identifies this region of interest (ROI). In order to track the ROI across images, over time, the Kanade-Lucas-Tomasi (KLT) algorithm was utilized[54]. The KLT algorithm identifies prominent feature points within the ROI by using the minimum eigenvalue algorithm developed by Shi and Tomasi[49]. These features are then identified and tracked across subsequent frames through a 2D geometric transformation that estimates the degree to which the new points have translated, rotated, and scaled from the old points. Figure 2 to the right, demonstrates successful implementation of the aforementioned process. Once the ROI has been identified and segmented from the original image, the maximum and minimum temperature for each pixel in the images within a 0.33 second epoch will be determined using the peak-hold and valley-hold algorithm, resulting in a total of 90 total maximum and minimum temperature values for each question in addition to the calculation of average pixel intensity of the left and right hemiface and periorbital region. A mixed factor 'maximum amplitude' and 'minimum amplitude' ANOVA will be conducted using the mean temperature from the periorbital and each hemiface. A third ANOVA will be performed on the difference between the average temperature and the corresponding minimum and maximum temperatures within the periorbital regions. Independent variables will include group (rocket or bomb) and question type (control, relevant, irrelevant, or guilty knowledge).

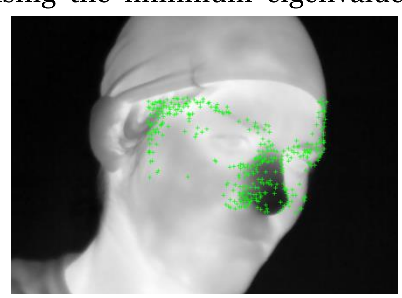

Figure 2. Identification of Prominent Facial Features Using KLT algorithm

Electrodermal Activity (EDA): EDA has been widely used to provide insight into psychological processes including, but not limited to, emotional response and cognitive states[45] and has been widely used in the field of deception detection[7]. EDA signals are comprised of tonic and phasic components that characterize skin conductance responses that result from a sympathetic response to a stimulus[9]. Tonic components relate to the slower acting components of the signal, such as the skin conductance level (SCL) which reflects the general amount of autonomic arousal. The phasic component relates to the faster moving components of the signal in response to a stimulus and is termed the skin conductance response (SCR), which can be subdivided into latency, rise-time, amplitude, and half-recovery time components. In addition, the number of EDA peaks per response time can point to the intensity of a stress response, as numerous studies have suggested that EDA peak frequency increases with emotional evoking, stressful situations[51]. Previous studies have been able to achieve high accuracies using skin conductance alone with reported classification accuracies of 76% and 83% for guilty and innocent examinees, respectively[28]. Detection of the tonic and phasic components of EDA are driven by peak detection. Signal processing of EDA signals has been well established in the literature and involves median filtering over the original EDA signal, followed by a low-pass Butterworth filter (cutoff frequency ~5Hz) on the phasic data to remove any powerline noise. Manual identification of the peak amplitude threshold (~0.005) is used to determine whether or not a value meets peak criterion[36].

Aim 1C: Musculature Activation:

Muscular activation will be quantified from a granular and broad perspective through the evaluation of the right trapezius muscle activity using EMG and the body's postural sway when sitting on a chair placed under a force plate, respectively.

Muscle Activity: EMG is used to determine the initiation of a movement by measuring the electrical potentials of muscles generated during contraction. Commonly used EMG variables include the average rectified value (ARV), root mean square value (RMS), median frequency (MDF), mean frequency (MNF), and conduction velocity (CV). ARV and RMS values are calculated on pre-processed data and quantify the square root of the average power of the EMG signal and the area under the rectified (negative voltages made positive) EMG signal, respectively[15,39]. MNF and MDF represent spectral components of the signal and are the gold standards used to detect muscle fatigue are traditionally obtained by a Fast Fourier Transformation (FFT), but autoregressive (AR) techniques have also been implemented[15,39].

EMG signals exemplify nonstationary behavior during dynamic and sustained muscle contractions. Surface EMG devices will pick up activity from all muscles in the vicinity, and thus contains crosstalk from adjacent muscles. From a processing standpoint, fidelity of the signal is influenced by the signal-to-noise ratio (SNR) and signal distortion. EMG SNR can be affected by nearby electrical equipment, ambient EMG radiation, motion artifact, and inherent instability of the signal[59]. During a polygraph or CIT, participants are not undergoing any type of physical activity, therefore, we do not expect the signal to be substantially corrupted. The Teager-Kaiser energy-tracking operator (TEKO) is a simplistic algorithm used to determine the initiation of a movement or twitch, by squaring the current timepoint and subtracting off the product of the previous and future timepoint, this suppresses ambient noise and enlarges the signal[61].

All variables will be calculated within 30 second epochs, which align with the given time that a participant has to answer a question. To my knowledge, this is the first study to examine the effects of deception on surface muscle activity.

Postural Sway: Changes in movement pattern related to deception related stress could be a compensatory effect related to the cognitive burden or control associated with fabrication of a believable lie. Video-based detection deception has suggested that distinctive movements are able to differentiate truthfulness from deceit[30]; however, these techniques may not be sensitive to minute or occluded movements which can be overcome by static posturography analyses. Previous studies have examined the changes in center of pressure (COP) parameters during truth and deception related responses, which revealed that there were significant differences in COP ellipse eccentricity (EE), major axis angle (A_{MA}), and median frequencies (f_{ML})[30], indicative of increased postural rigidity during acts of deceit[55]. These findings could be attributed to the cognitive load associated with lying. For instance, the increase in cognitive load associated with deception may result in neglect of body language, or the decrease in movement may be attributed to the hyperawareness of deceitful individuals to their environment and result in a compensatory strategy to overcontrol their movements to attempt to hide what may be perceived as deceptive behavior. Furthermore, highly motivated truth-tellers exhibited an increase in postural fidgeting[5]. COP time series in the medial-lateral (ML) and anterior-posterior (AP) directions will be used to obtain mean displacement in each direction, COP path length (L_{path}), EE, A_{MA}, major axis length (L_{MA}),and f_{ML} in the ML and AP directions. I hypothesize that when answering deceptively, individuals will exhibit increased postural rigidity, followed by an adjustment in the ML or AP direction to compensate for their stillness following the question. Furthermore, I believe that postural sway provides a more generalized and objective means of measure than other components measured by polygraph, because it is not directly influenced by medications that affect ANS functioning.

Aim 1D: Oculomotor Behavior Through Eye Tracking:

Oculomotor behavior tracking has gained significant popularity in the application of deception detection. Three different mechanisms can be used to identify deception: fixation points, blinks, and pupil dilation. Previous studies have demonstrated that fixation patterns of deceptive participants differ significantly than innocent participants, while blink rate and blink latency have been founded to be statistically different when subjects were presented with an object related to an incriminating action[12]. Furthermore, pupil diameter has been deemed a reliable indicator of cognitive workload and emotional arousal and has been shown to increase during deception[20]. In addition to the aforementioned metrics, we will assess visual saccadic eye movements and spontaneous saccadic eye movements. Traditional visual research evaluates saccadic movements and their role in inspecting a visual scene or object; however, saccadic eye movements can also occur independent of these instances, often involuntarily. These movements are largely associated when individuals attempt to retrieve information stored in long-term memory[44]. Previous literature suggests that individuals display more saccadic eye movements when telling a spontaneous lie than a truth[58]. I hypothesize that eye tracking will be one of the most reliable indicators of deception induced stress, due to the inconspicuous way in which participants are monitored and the difficulty surrounding voluntary control of one's eye.

Statistical Analyses: Following acquisition of these metrics from the post-processed signals, a principal components analysis (PCA) will be conducted to remove highly correlated variables from the analysis. Highly correlated data results in overly redundant information that ultimately reduce the computational efficiency of the model. The PCA will serve to reduce the dimensionality of the data by maximizing the variance of the parameters that are uncorrelated. Furthermore, a mixed factor Analysis of Variance (ANOVA) will be conducted using the aforementioned physiologic metrics as the dependent variable, and the group (rocket or bomb) and question type (control, relevant, irrelevant, and guilty knowledge) as independent variables.

Aim 2: To Evaluate the Efficacy of Minimal Contact and Off-Body Monitoring Systems Against Current Gold Standards Used in Polygraph.

This proposal will utilize two methodological approaches to validate the efficacy of wearable and off-body systems against their gold-standard counterparts. The implementation of either technique is driven by the output of each device and more specifically, whether a raw waveform is attainable. Comparison of the raw waveform provides a more detailed analysis regarding sensor efficacy, because every datapoint in time can be compared, whereas the alternative is to compare summary metrics averaged across an epoch of time. Implementation of the former is limited to accessibility of the raw data output. Some of the devices utilized in this study do not output raw data; similarly, other sensors, like thermography aim to indirectly measure a parameter such as EDA; however, the traditional waveform of EDA cannot be directly compared an infrared image.

1. **Autoregressive Modelling of Raw Waveforms:** Autoregressive (AR) modelling uses regression to predict future values based on past values. The main assumption for AR models is signal stationarity, which is violated by the stochastic nature of physiological signals; however, the application of AR modeling in this instance will not be used for forecasting. Rather, I intend to exploit the properties of the AR model to capture the correlation parameters, by treating the traditional sensor as the actual measure and the wearable or off-body sensor as the 'forecasted' measure. From there, the differences between the error terms will be used to ascertain the level of agreement between the two signals.

2. **Quantification of Agreement Between Two Summary Statistics:** AR provides a robust methodology to measure the raw waveform of a given physiologic signal. However, as stated previously, we do not always have a direct correlate for each type of sensor. For example, the Polar H10 and Bittium Faros are both ECG devices whose outputs can be used to derive parameters related to HR and HRV; however, unlike the Bittium Faros, data exported from the Polar H10 does not contain the raw ECG waveform; therefore, AR techniques would be ill-reputed. Consequently, a different technique must be implemented to assess the level of congruency between the two signals. This can be achieved by taking the post-processed signal and calculating metrics that each signal purports to provide (e.g., HR, RMSSD, etc.). A Bland Altman (BA) analysis is the standard way to appraise the level or limits of agreement between two instruments by measuring the mean difference and standard deviation of the two measures[18]. Following post-processing of the data, summary metrics will be calculated between the two sensors of interest for every 30 second epoch corresponding to the time in which the participant has to answer the question. Ideally, the two instruments should yield the same results; however, any measurement of variables implies some degree of error due to inherent analytical imprecisions. If the variation in the differences between the two instruments is only linked to analytical imprecisions, then the average of the differences should sum to 0. In cases where the mean difference is not 0, the BA test can point to which sensor and by how many units a sensor deviates from its respective counterpart. For example, if the difference between sensor A and B is positive, then it suggests that sensors A tends to overestimate the variable by X number of units than sensor B.

Exploratory Aim. Understanding the Behavioral Components that Drive Deception Detection.

Deception is a complex and multi-faceted social behavior that involves higher cognitive functioning processes. Therefore, it is crucial to categorize the behavioral component that drives deception related stress.

Rationale: Baseline personality characteristics can serve as a modulator of deception related response. Therefore, accounting for the inter-individual differences that arise from personality discrepancies can provide a means of control for these differences and provide an indirect means of validity of responses during the interview. Furthermore, if the model is unsuccessful in accurately classifying truth from lie, the results from the questionnaires can aid in the understanding of how variations in uncontrollable measures (such as personality) can affect model success.

The behavioral inhibition system (BIS) and behavioral activation system (BAS) refer to the neurophysiological processes that are activated in the presence of punishment and reinforcement, respectively, where the former drives avoidance-related behaviors and emotions (withdrawal, catastrophizing, anxiety), while the latter facilitates approach-related behaviors, motivated by reward (impulsivity, excitement, self-efficacy)[2]. This questionnaire aims to assess personality qualities that point towards the dominance or sensitivity of one system over the other. Henceforth, scores may serve as a covariate to account for inter-individual differences and could serve as a correlate for expected physiologic patterns expected in some individuals with higher inhibition or activation baselines. Individuals with higher BIS scores demonstrate a more profound BIS activation and may be more responsive to cues of retribution or punishment, compared to individuals with lower BIS sensitivity. Previous studies have demonstrated that there was a significant positive correlation in individuals with a higher BIS score and those identified as being deceptive. I hypothesize that individuals with inflated BIS scores or deflated BAS scores will significantly affect the classification accuracy of the model. This posits the following questions:

1. Do BIS/BAS scores correlate with physiologic measures from guilty or innocent responses (Metric X Group (Rocket or Bomb) X BIS or BAS interaction)?
2. How do inflated or deflated BIS/BAS scores affect the model accuracy?
3. Can an underbalance in underactive BIS and over reactive BAS provide insight into whether an individual will have an enhanced physiological reactivity when *reward* or punishment is insight? This could provide insight into whether an operator should incentivize the participant with punishment (longer prison sentence) or reward (polygraphs are difficult to beat, but intelligent people can beat it).

Marlow-Crowne Social Desirability Scale (MCSDS) aims to assess how individuals will distort their answers to obtain more desirable scores than would be achieved by responding honestly. Evidence has suggested that personality variables correlate with performance outcomes, especially regarding decisions made in high stakes situations[29]. Understanding the differences in personality that motivate more socially desirable responses can provide insight into the measure of validity of participants' responses during the interview or be used to control for individual differences. Higher scores suggest that an individual may present themselves in an unrealistically favorable manner. I hypothesize that individuals with higher MCSDS scores will be more likely to lie during relevant questions which are more likely to put their social character in jeopardy.

1. Does a higher or lower MCSDS score correlate with physiologic measure from guilty or innocent responses?
2. Do individuals with higher or lower MCSDS scores tend to lie more during relevant questions?

Statistical Analyses: A series of Multivariate Analysis of Variance (MANOVAs) tests will be conducted with envelope content as the independent variable and physiologic metrics and BIS / BAS / Marlow Crowne scales as dependent variables.

Aim 3: To Develop a Multi-Modal Deception Detection Algorithm.

Data garnered from this study will be used to train a classifier using deceptive and nondeceptive responses based on either a between-person or within-person approach, using a long short-term memory (LSTM), deep learning (DL) model. DL has gained significant traction over recent years, with one of its significant advantages being that it is able to execute feature engineering by itself. Furthermore, these models are extremely robust at delivering high accuracies and eliminating the need for data labeling. The major pitfall of DL based strategies is the 'black-box' effect that it has on the data, more specifically, we are unable to determine the underlying physiological mechanisms or what parameters drive model success. Despite the black-box effect of DL, the physiologic parameters that drive deception induced will be characterized through the successful completion of Aim 1.

Hypothesis: Model accuracy will be driven by the transient change in an individuals' physiology that significantly differ when answering deceptively. I hypothesize that this ephemeral difference will be detectable by determining the physiological anomalies that drive deception induced stress.

Questions:

- What of the following types of sensor suites or combination of sensors will yield the highest model accuracy when using a long short-term memory deep learning approach?
    - Traditional polygraph (Cardiorespiratory, and thermal response)
    - Wearable devices (Musculature, cardiorespiratory, and thermal response)
    - Off-Body Measures (Thermal and oculomotor response)
- Will a between subjects or within subjects approach yield higher model accuracy?

Classification Algorithms: Deception detection classification will be driven by the concepts that underly biomedical anomaly detection, where the principal task is to identify datapoints that do not fit the overall data distribution. DL has become an increasingly popular architecture for medical anomaly detection due to its ability to model data nonlinearly and automatically detect features through hierarchal learning[16]. Recurrent neural network (RNN) architectures such as long short-term memory (LSTM) can model relationships that are dependent on the temporal evolution of a signal. Like all neural networks, at each node, the model performs a calculation using the input parameters and returns an output value; however, in RNNs the given output is then used as an input for the next sequence. This technique is termed back propagation, where the gradient at any given time is dependent upon the prediction of previous timesteps. Traditionally, RNNs are limited at modelling long-term dependencies; however, LSTM circumvents this challenge through the inclusion of an internal state or 'memory cell' wherein information can be stored and retrieved over many time steps. The input parameters, previous timestep, and internal states are modulated by gates[16]. LSTM networks contain three gates, that control informational flow within a node. The first gate is the 'forget gate', that uses a sigmoid activation function to output a value between 0 (completely forget) and 1 (keep) for each number in the cell state. The second gate, 'input' is used to determine what new information will be stored in the cell state, which is dependent on a sigmoid activation function and a tanh layer. The tanh layer creates new candidate values to the existing cell state while determining what information from past iterations is still meaningful. Lastly, the 'output' gate determines what relevant pieces of information should be outputted to the next layer. These gates work together to ultimately 'learn' to make a prediction by adding or removing relevant or irrelevant information to the state cell. Figure 3 below demonstrates the basic architecture of a simple LSTM network. Post-processing of the output layer for the LSTM network will determine the probability that a given response to a question is either truth or lie. SoftMax is an exponentially increasing function that normalizes the input values to the output value between 0 and 1; the sum of the final output layer is 1. Therefore, the SoftMax function can be used to determine the likelihood of truthfulness.

Figure 3. General architecture of a LSTM network.

Between or Within Subjects Approach

The between person's approach uses the responses from all but one person as the training set, while the left-out person is used as the test set, whereas the within-subject approach looks for deception cues specific to one person and how their responses differ from their own baseline[38]. In total there will be a total of 3,200 responses (800 of control, relevant, irrelevant, and guilty knowledge responses) and 32 responses per subject. Previous deception detection studies have used as little as 400 responses[38].

Between-Subjects Approach. To understand whether deceptive patterns generalize across a population, a between-subject's approach will be implemented. This approach utilizes the behavioral profiles of all but one subject as the training set, where the one person is used as the test set. It is done in a round-robin way, where each person eventually serves as the test set. The classification accuracy is determined by taking an average of each instance. I hypothesize that this approach will not be very efficacious, due to the behavioral tendencies that deception tends to evoke, I believe that greater anomaly detection will be achieved when a subject serves as their own baseline.

Within-Subjects Approach. This approach compares behavioral profiles against an individual's normative patterns. To implement this model, a specific subject's data will be partitioned into n folds using stratified sampling. The training dataset is then constructed by concatenating n-1 folds, where the remaining folds are used to test the model, using a round-robin approach to generate a total of n local models and n total test sets. The percentage of correctly classified responses across all participants is then determined to evaluate overall accuracy. The appropriate number of samples to 'hold-out' will be determined as a apart of model accuracy.

Determination of Sample Size and Power

Conventional machine learning or deep learning strategies utilize approximately 10,000 samples for model training. However, the purpose of this project is to serve as a pilot study for such a system that utilizes minimal contact or off-body monitoring technologies. Following the completion of this pilot study, we will be able to determine which sensors can be kept in the final sensor suite and then more robustly collect and test data from a larger sample size. To determine the appropriate sample size for this pilot study, a power analysis will be performed. However, the conventional strategies to determine the appropriate N are not valid because there is not an applicable way in which to determine the effect size. Therefore, this study will be powered by a confusion matrix and setting the sensitivity to 0.9. The appropriate sample size will be determined by adding subjects until the 0.9 threshold is reached. A post-hoc power analysis can be computed following collection and analysis of the data that can drive selection of the appropriate sample size for future models.

Research Design and Methods.

Human Subject Procedures.

Recruitment: A total of 100 participants, between the ages of 18-42 will be recruited from Morgantown, West Virginia. Subjects will be recruited via study advertisements and complete a qualification survey to determine eligibility for participation.

Inclusion and Exclusion Criterion: Previous literature has suggested that polygraph is more susceptible to false positives or false negatives if a subject has a condition or uses medication that affects the ANS[7]. The following exclusion criteria would disqualify an individual from participation in the study: currently taking any medication that significantly affects natural blood pressure regulation, under the influence of any prescription stimulant, under the influence of alcohol or illicit drugs, currently undergoing hormone therapy, currently has any cold-like symptoms, pregnant or trying to become pregnant, has ever been diagnosed with having a mood disorder, schizophrenia, depression, or bipolar disorder, has Vitiligo or other depigmenting conditions of the skin, has hearing difficulties, blindness, or is unable stay seated for approximately 40 minutes.

Addressing Relevant Biological Variables: Throughout data collection, participants age, biological sex, and skin color will be tracked to ensure that an adequate representation of the population is obtained. Furthermore, following consent, participants will fill out an intake survey that captures caffeine intake, sleep quality, etc.

Sensor Suite: All data will conform to the compliance framework outlined by HIPAA and be deidentified and stored on encrypted hard drives.

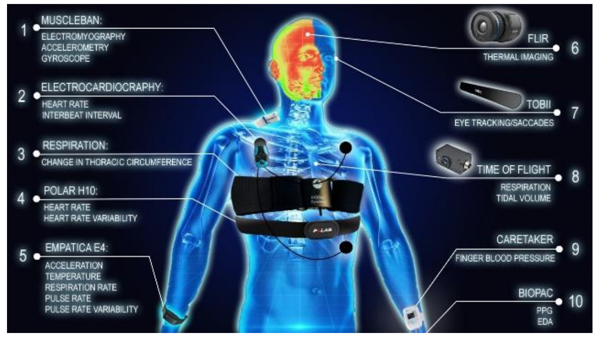

Figure 4. Sensor Setup and Placement

The **MuscleBanBE** is a single-channel electromyography (EMG) sensor, and a triaxial accelerometer and magnetometer that enables real-time acquisition of muscle activity at a 12-bit, 1000 Hz sampling rate. The sensor attaches on the participant's right trapezius via two electrodes. Data exports as a .h5 file and synchs to system time following collection.

As shown in Figure 5, electrodes from the **Bittium Faros** (FDA 501(k) cleared, class IIa medical device), a multi-lead ECG device, will be placed according to Einthoven's triangle on the chest. The device samples at 1000 Hz and incorporates a triaxial accelerometer that samples at 100 Hz. Data exports as an .edf file and synchs to system time following collection.

The MP160 **Biopac** system is a customizable amplifier suite that incorporates PPG, EDA, and a respiration sensor. respiration belt, EDA sensors, and a finger-based PPG sensor in their amplifier suite. The EDA sensors are attached to two electrodes that attach to the participant's first and second finger, and the finger-based PPG sensor is attached to the third finger. The aforementioned electrodes attach to the participant's nondominant hand and sample at 35 and 100 Hz, measuring skin conductance level / response and PR/PRV, respectively. The respiration belt will be placed around the thoracic circumference of the individuals and uses a mechanical strain gauge that responds to chest expansion during inhalation. The amplifier transmits packets of information via Bluetooth to AcqKnowledge (proprietary software used with Biopac systems) that is manually synched with system time upon the start of data collection.

The **Polar H10** is a wearable chest-based device, placed at the xiphoid process and has been deemed as the gold-standard ECG, wearable device for multiple HR validation studies. Traditionally, the device synchs to the Polar Flow- phone application; however, this does not readily provide raw inter-beat interval (IBI) data. Thus, the Elite HRV application will be used to obtain IBI data. Outputs will provide IBI times in milliseconds (ms) in a .txt file.

The **Empatica E4** is a wrist-worn device that provides the raw signal outputs and measures pulse rate, pulse rate variability, EDA, temperature, and triaxial accelerometry (ACC). The raw signals sample at 64 Hz, 4 Hz, 4 Hz, and 32 Hz for PPG, EDA, temperature, and ACC, respectively. Data are exported as .csv files and synch with the computer system time.

The **FLIR A700** camera is a thermal imaging camera that measures temperatures between a range of -4°F and 248°F. Infrared resolution of the images are 16-bit, 640 x 480 and sampled at a frame rate of 30 Hz.

The **Tobii Nano Pro** is an off-body, screen-based, eye tracker which captures data at 60 Hz, designed for fixation-based studies and is placed on the laptop in which the participant views the video. Data is exported as a .dat file and contains information surrounding left and right eye X and Y position and pupil diameter. Similar to the FLIR camera, the Tobii Nano Pro is connected to a router, that collates the data in LabView.

The **Bertec** is a force plate that provides 1000 Hz raw positional, moment, and center of pressure data and is placed underneath of the participant's chair. Data outputted by this device does not readily provide timestamps, therefore, upon the initiation of collection, therefore; manual timestamp determination will be performed.

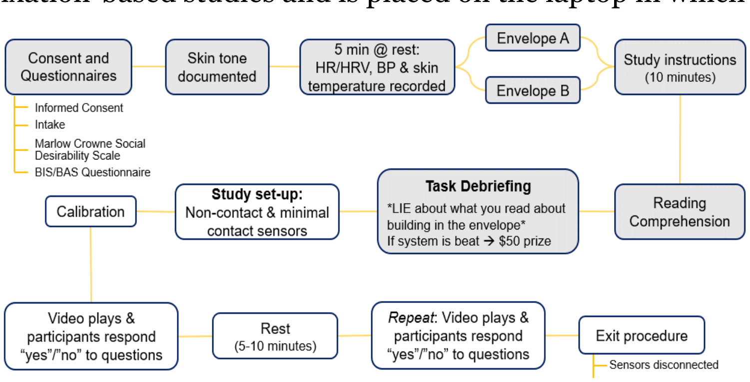

Figure 5. Study Design

Study Design and Methods. The experimental design is outlined in Figure 5. Following determination of inclusion, eligible participants will be consented in a separate room from data collection and be asked to complete three questionnaires: 1.) General intake questionnaire 2.) BIS / BIA 3.) Marlow Crowne Social Desirability. Participant's skin tone will be documented according to the Fitzpatrick scale and a resting blood pressure and pulse rate will be taken. Study personnel will instruct the participant to choose between two sealed envelopes: A or B. The envelopes either contain a mock-scenario surrounding the construction of a model rocket or a pipe bomb, with the latter intended to create a more heightened state of arousal as it contains incriminating information. Participants will be randomized into one of the two groups prior to data collection and will ultimately not be given the 'choice to choose' between the two hypothetical objects as both envelopes will contain the same contents. Following envelope selection, participants will be given time to comprehend the contents before they are debriefed on the task, where participants are incentivized to 'beat the system' by answering untruthfully to questions that may divulge the identity of the hypothetical object, but told to answer truthfully, otherwise. Study personnel inform participants that if they 'beat the system', they will win a $50 gift card, otherwise, they will receive no payment.

After being debriefed on the task, the participant is brought into a private room where sensors outlined in the aforementioned section are attached and calibrated. A prerecorded video containing 16 questions will be played for the participant in which they will be asked to answer either 'yes' or 'no'. Participants will be given a total of 30 seconds to answer each question. Question types include relevant (related to the topic of investigation, e.g.) Did the object you build have a fuse?), irrelevant (questions that should elicit little to no emotional or physiological response, e.g.) Are you currently inside of a building?), control (questions used as comparisons to other questions, e.g.) Have you ever lied to a person of higher authority?), and guilty knowledge questions (contain relevant to specific details of the investigation and *should* be answered truthfully by individuals in the model rocket group and untruthfully by individuals in the bomb group, e.g.) Did you intend to harm the safety of others by building your object?). The video will be presented two times, with a 10-mintue break in between. Following the termination of both videos, the subjects will be asked to fill out the NASA TLX workload questionnaire and informed that there was in fact, not system to beat.

Timeline.

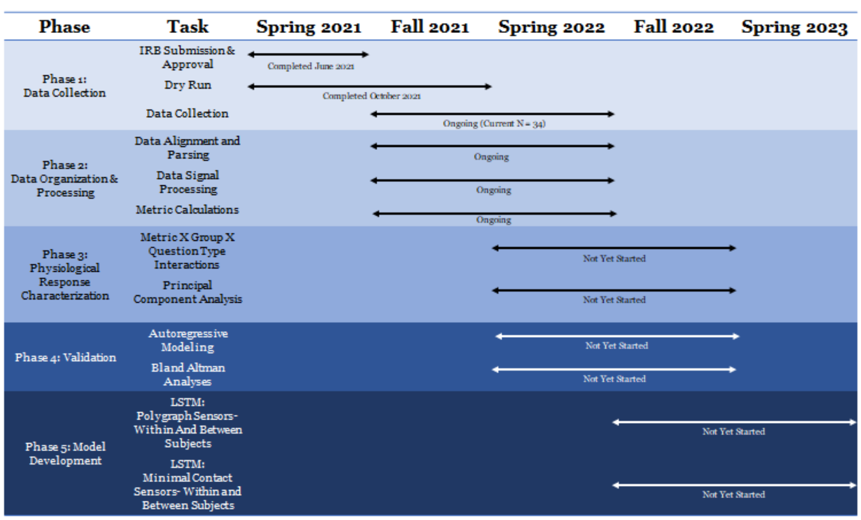

Figure 6. Proposed timeline for completion of events.

Concluding Remarks.

The intent of this dissertation research proposal is to highlight my proposed strategy for the successful completion of the aforementioned aims that attempt to objectively and comprehensively evaluate the efficacy of wearable and off-body monitoring technologies in a deception detection model. More specifically, this project serves as a stepping stone for our funding agency to fulfill their ancillary objective of developing an off-body physiologic monitoring system for credibility assessment.

References.

[1] Badhan, A. B. (2018). *Non-invasive Wearable Wristband for Heart Rate and Respiration Rate Estimation Using Photoplethysmography with Motion Artifact Removal*. University of Maryland, Baltimore County.

[2] Bjørnebekk, G. (2007). Reinforcement sensitivity theory and major motivational and self-regulatory processes in children. *Personality and Individual Differences*, *43*(8), 1980-1990.

[3] Bond, G. D. (2008). Deception detection expertise. *Law and Human Behavior*, *32*(4), 339-351

[4] Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., & Su, L. (2014). Countering countermeasures: detecting identity lies by detecting conscious breakthrough. PloS one, 9(3), e90595

[5] Burgoon, J. K., Hamel, L. M., Blair, J. P., & Twyman, N. W. (2020). Factors that facilitate or impair kinesic and vocalic nonverbal behaviors during interpersonal deception. In *Social Intelligence and Nonverbal Communication* (pp. 79-117). Palgrave Macmillan, Cham.

[6] Chand, T., Li, M., Jamalabadi, H., Wagner, G., Lord, A., Alizadeh, S., ... & Sen, Z. D. (2020). Heart rate variability as an index of differential brain dynamics at rest and after acute stress induction. Frontiers in neuroscience, 14, 645.

[7] Cook, L. G., & Mitschow, L. C. (2019). Beyond the polygraph: deception detection and the autonomic nervous system. *Federal Practitioner*, *36*(7), 316.

[8] Corr, P. J. (2002). JA Gray's reinforcement sensitivity theory: Tests of the joint subsystems hypothesis of anxiety and impulsivity. *Personality and individual differences*, *33*(4), 511-532.

[9] Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, *8*(2), 132-142.

[10] De Turck, M. A., & Miller, G. R. (1985). Deception and arousal: Isolating the behavioral correlates of deception. *Human Communication Research*, *12*(2), 181-201.

[11] Derakhshan, A., Mikaeili, M., Gedeon, T., & Nasrabadi, A. M. (2020). Identifying the optimal features in multimodal deception detection. *Multimodal Technologies and Interaction*, *4*(2), 25.

[12] D. C. Derrick, K. Moffitt and J. F. Nunamaker, Jr., Eye Gaze Behavior as a Guilty Knowledge Test: Initial Exploration for Use in Automated, Kioskbased Screening, 43rd Annual Hawaii International Conference on System Sciences, Computer Society Press, Poipu, Hawaii, 2010.

[13] Dhillon, S. S., & Chakrabarti, S. (2001, October). Power line interference removal from electrocardiogram using a simplified lattice based adaptive IIR notch filter. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 4, pp. 3407-3412). IEEE.

[14] Esco, M. R., & Flatt, A. A. (2014). Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. Journal of sports science & medicine, 13(3), 535.

[15] Farina, D., & Merletti, R. (2000). Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions. *Journal of Electromyography and Kinesiology*, *10*(5), 337-349.

[16] Fernando, T., Gammulle, H., Denman, S., Sridharan, S., & Fookes, C. (2020). Deep Learning for Medical Anomaly Detection--A Survey. *arXiv preprint arXiv:2012.02364*.

[17] Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003)

[18] Giavarina, D. (2015). Understanding bland altman analysis. Biochemia medica, 25(2), 141-151.

[19] Gombos, V. A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, social, and general psychology monographs*, *132*(3), 197-214.

[20] Heilveil, I. (1976). Deception and pupil size. *Journal of Clinical Psychology*.

[21] Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, *70*(1), 177.

[22] J. Knight, "The truth about lying", *Nature*, vol. 428, no. 6984, pp. 692-694, 2004.

[23] Kaur, M., & Singh, B. (2011, February). Comparison of different approaches for removal of baseline wander from ECG signal. In *Proceedings of the international conference & workshop on emerging trends in technology* (pp. 1290-1294).

[24] Krapohl, D. J. (2006). Validated polygraph techniques. *Polygraph*, *35*(3), 149-155.

[25] Levkov, C., Mihov, G., Ivanov, R., Daskalov, I., Christov, I., & Dotsinsky, I. (2005). Removal of power-line interference from the ECG: a review of the subtraction procedure. *BioMedical Engineering OnLine*, *4*(1), 1-18.

[26] Lenis, G., Pilia, N., Loewe, A., Schulze, W. H., & Dössel, O. (2017). Comparison of baseline wander removal techniques considering the preservation of ST changes in the ischemic ECG: a simulation study. *Computational and mathematical methods in medicine*, *2017*.

[27] Lewis, G. F., Gatto, R. G., & Porges, S. W. (2011). A novel method for extracting respiration rate and relative tidal volume from infrared thermography. Psychophysiology, 48(7), 877-887.

[28] MacLaren, V. V. (2001). A quantitative review of the Guilty Knowledge Test. *Journal of applied psychology*, *86*(4), 674.

[29] Mesmer-Magnus, J., Viswesvaran, C., Deshpande, S., & Joseph, J. (2006). Social desirability: The role of over-claiming, self-esteem, and emotional intelligence. Psychology Science, 48(3), 336.

[30] Michael, N., Dilsizian, M., Metaxas, D. N., & Burgoon, J. K. (2010). Motion profiles for detecting deception using visual cues. In 11th European conference on computer vision, Crete, Greece.

[31] Mullin, D. S., King, G. W., Saripalle, S. K., Derakhshani, R. R., Lovelace, C. T., & Burgoon, J. K. (2014). Deception effects on standing center of pressure. *Human movement science*, *38*, 106-115.

[32] National Research Council. (2003). *The polygraph and lie detection*. National Academies Press.

[33] Nelson, R., & Handler, M. (2013). A brief history of scientific reviews of polygraph accuracy research. APA Magazine, 47(6), 22-28.

[34] Pennebaker, J. W., & Chew, C. H. (1985). Behavioral inhibition and electrodermal activity during deception. *Journal of personality and social psychology*, *49*(5), 1427.

[35] Pollina, D. A., Dollins, A. B., Senter, S. M., Brown, T. E., Pavlidis, I., Levine, J. A., & Ryan, A. H. (2006). Facial skin surface temperature changes during a “concealed information” test. *Annals of biomedical Engineering*, *34*(7), 1182-1189.

[36] Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in electrodermal activity data collection and signal processing: A systematic review. *Sensors*, *20*(2), 479.

[37] Rogers, R., Boals, A., & Drogin, E. Y. (2011). Applying cognitive models of deception to national security investigations: Considerations of psychological research, law, and ethical practice. *The Journal of Psychiatry & Law*, *39*(2), 339-364.

[38] Rajoub, B. A., & Zwiggelaar, R. (2014). Thermal facial analysis for deception detection. *IEEE transactions on information forensics and security*, *9*(6), 1015-1023.

[39] Reaz, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological procedures online*, *8*(1), 11-35.

[40] Rogers, R., Boals, A., & Drogin, E. Y. (2011). Applying cognitive models of deception to national security investigations: Considerations of psychological research, law, and ethical practice. *The Journal of Psychiatry & Law*, *39*(2), 339-364.

[41] Rosenfeld, J. P., Ben-Shakhar, G., & Ganis, G. (2012). Detection of concealed stored memories with psychophysiological and neuroimaging methods.

[42] Roulin, N., & Ternes, M. (2019). Is it time to kill the detection wizard? Emotional intelligence does not facilitate deception detection. *Personality and Individual Differences*, *137*, 131-138.

[43] Saboul, D., Pialoux, V., & Hautier, C. (2013). The impact of breathing on HRV measurements: Implications for the longitudinal follow-up of athletes. European journal of sport science, 13(5), 534-542.

[44] Salojärvi, J., Kojo, I., Simola, J., & Kaski, S. (2003, September). Can relevance be inferred from eye movements in information retrieval. In *Proceedings of WSOM* (Vol. 3, No. 01, p. 2003).

[45] Sequeira, H., Hot, P., Silvert, L., & Delplanque, S. (2009). Electrical autonomic correlates of emotion. *International journal of psychophysiology*, *71*(1), 50-56.

[46] Serrano-Ibáñez, E. R., López-Martínez, A. E., Ramírez-Maestre, C., Esteve, R., & Jensen, M. P. (2019). The behavioral inhibition and activation systems and function in patients with chronic pain. *Personality and Individual Differences*, *138*, 56-62.

[47] Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*, 258.

[48] Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in psychology*, *5*, 1040.

[49] Shi, Jianbo and Carlo Tomasi. "Good Features to Track," *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[50] Srivastava, N., & Dubey, S. (2018, March). Deception detection using artificial neural network and support vector machine. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1205-1208). IEEE.

[51] Ströfer, S., Noordzij, M. L., Ufkes, E. G., & Giebels, E. (2015). Deceptive intentions: Can cues to deception be measured before a lie is even stated?. *PloS one*, *10*(5), e0125237.

[52] Stone, J. D., Ulman, H. K., Tran, K., Thompson, A. G., Halter, M. D., Ramadan, J. H., ... & Hagen, J. A. (2021). Assessing the Accuracy of Popular Commercial Technologies That Measure Resting Heart Rate and Heart Rate Variability. *Frontiers in Sports and Active Living*, *3*, 37.

[53] Tanweer, K. T., Hasan, S. R., & Kamboh, A. M. (2017, May). Motion artifact reduction from PPG signals during intense exercise using filtered X-LMS. In *2017 IEEE international symposium on circuits and systems (ISCAS)* (pp. 1-4). IEEE.

[54] Tomasi, Carlo and Takeo Kanade. *Detection and Tracking of Point Features*, Computer Science Department, Carnegie Mellon University, April, 1991.

[55] Twyman, N. W., Proudfoot, J. G., Schuetzler, R. M., Elkins, A. C., & Derrick, D. C. (2015). Robustness of multiple indicators in automated screening systems for deception detection. *Journal of Management Information Systems*, *32*(4), 215-245.

[56] Verkuil, B., Brosschot, J. F., & Thayer, J. F. (2014). Cardiac reactivity to and recovery from acute stress: Temporal associations with implicit anxiety. International Journal of Psychophysiology, 92(2), 85-91.

[57] Verschuere, B., Crombez, G., Koster, E., & Uzieblo, K. (2006). Psychopathy and physiological detection of concealed information: A review. Psychologica Belgica, 46(1-2).

[58] Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research*, *22*(4), 544-562.

[59] Wang, J., Tang, L., & Bronlund, J. E. (2013). Surface EMG signal amplification and filtering. *International Journal of Computer Applications*, *82*(1).

[60] Webb, A. K., Honts, C. R., Kircher, J. C., Bernhardt, P., & Cook, A. E. (2009). Effectiveness of pupil diameter in a probable-lie comparison question test for deception. *Legal and Criminological Psychology*, *14*(2), 279-292.

[61] Xia, Y., Gao, Q., & Ye, Q. (2015). Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models. *Biomedical Signal Processing and Control*, *18*, 254-262.