Desarrollo de Características: Investigación de la Latencia Mínima de Respuesta para Señales Electrodermales y de Cardio

Raymond Nelson

Abstract

Un paso fundamental en el análisis de datos de pruebas poligráficas es la extracción de características, y se guía por parámetros como son la ventana de evaluación (WOE)*, la ventana de inicio de respuesta (ROW) y la latencia mínima de respuesta (MRL). Para la actividad electrodérmica (EDA) la MRL convencional es de 0.5 segundos a partir del inicio del estímulo, pero algunos profesionales se han preguntado si ampliar este umbral podría reducir los artefactos y mejorar la precisión. De igual manera, no existe una norma para la MRL para los datos de actividad cardiovascular (CA). Utilizando dos conjuntos de datos de archivo, este estudio evaluó cómo los diferentes ajustes de la MRL afectan el tamaño del efecto estadístico de las características del EDA y CA. Los resultados mostraron que ampliar la MRL más allá de 0.5 segundos reduce la cantidad de datos utilizables y debilita los tamaños del efecto. Los hallazgos respaldan el mantenimiento de una MRL de 0.5 segundos para el EDA e introducen un apoyo empírico para un umbral de MRL en los datos de CA, lo que subraya la importancia de la selección de parámetros basada en datos para el análisis poligráfico.

*Nota del traductor: Para el presente documento se conservan todas las abreviaturas en inglés como son ROW, WOE, EDA y todos los demás del presente artículo, ya que todos los poligrafistas de habla hispana así los utilizan y están familiarizados con ellos.

Introducción

La extracción de características es el paso inicial del análisis de los datos de prueba en la detección psicofisiológica del engaño (PDD), seguido de la transformación numérica, la reducción de datos, el cálculo de un clasificador estadístico y la interpretación. La extracción de características involucra la identificación de los puntos de inicio y finalización de los cambios fisiológicos que se producen en respuesta a los estímulos de la prueba. Los procedimientos utilizados para la extracción de características se basan en una definición conceptual de cada característica de respuesta. Tanto para *la actividad electrodérmica* (EDA) como para *la actividad cardiovascular* (CA), una característica de respuesta se define como la distancia vertical (eje y) dentro de los gráficos de series temporales posteriores a cada inicio de estímulo, medida desde el inicio de una respuesta fisiológica hasta su pico o finalización. *El desarrollo de características* se refiere al proceso de derivar conceptos basados en la evidencia y prácticas de campo para la extracción de características.

La adhesión a principios científicos como la objetividad y la reproducibilidad requiere minimizar, en la medida de lo posible, la dependencia en la subjetividad y la intuición. Para lograrlo, deben cumplirse dos condiciones esenciales para poder atribuir razonablemente que los cambios en la actividad fisiológica se deben a los estímulos de la prueba. En primer lugar, cada estímulo de la prueba debe ser el único cambio que se produzca en el entorno inmediato al tiempo de su presentación. Si cualquier otro estímulo ambiental se produce simultáneamente, se confundirán las atribuciones causales relacionadas con la respuesta fisiológica.

Además de la ausencia de estímulos que compitan, el cambio fisiológico debe producirse con una adecuada temporalidad, relacionada con el estímulo de prueba, ni antes ni mucho después de su presentación. Sin embargo, evaluar la puntualidad de los cambios fisiológicos observados plantea retos, ya que está sujeta a la variabilidad entre evaluadores y a diferencias tanto de opinión como de procedimiento.

Se han desarrollado tres dispositivos conceptuales para proporcionar una estructura y definición operacional para la evaluación de la puntualidad de respuesta. La *ventana de evaluación* (WOE; National Center for Credibility Assessment, 2017) se define como el período fijo que comienza con el inicio del estímulo y que se extiende con una duración predeterminada. Se asume típicamente que los cambios fisiológicos que se producen dentro de esta ventana están relacionados con el estímulo. La duración del WOE se expresa en segundos y se ha descrito con longitudes que varían entre 15 y 20 segundos (Kircher y Raskin, 1988; Nelson, 2020; Nelson y Handler, 2013; Timm, 1982).

Otro dispositivo conceptual para evaluar la rapidez de la respuesta es la

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx

ventana de inicio de la respuesta (ROW), que comienza con el inicio del estímulo y se extiende hasta cinco segundos después de que el examinado proporciona una respuesta verbal (Bell et al., 1999; National Center for Credibility Assessment, 2017; Nelson, 2020, 2024). Esta ventana delimita el intervalo de tiempo en el que debe producirse el inicio de una respuesta para que se tome en cuenta para efectos de puntuación.

El tercer dispositivo conceptual para evaluar la puntualidad de la respuesta es la restricción *de latencia mínima de respuesta* (MRL). La latencia de respuesta, o período de latencia, se refiere al intervalo entre el inicio del estímulo y el inicio de una respuesta fisiológica. La restricción MRL es un *parámetro* de análisis que define el umbral para determinar si una respuesta fisiológica será incluida en el análisis, mientras que los datos de entrada observados consisten en las señales fisiológicas registradas durante la prueba. En el caso de los datos EDA, el valor MRL se fija tradicionalmente en los primeros 0.5 segundos después del inicio del estímulo (Bell et al., 1999) y sirve como umbral para determinar la puntualidad de la respuesta. Los cambios fisiológicos que se producen dentro de este periodo no están a tiempo y se excluyen de la puntuación. Las investigaciones previas acerca del desarrollo de características no han descrito el uso de un parámetro MRL para los datos CA.

Utilizando estos parámetros, la característica de respuesta tanto para EDA como para CA se extrae mediante la amplitud del aumento en los valores de los datos (eje y) que comienza después del período MRL para cada presentación de los estímulos. El inicio de la respuesta se identifica por el inicio de un segmento de pendiente positiva dentro del ROW, mientras que el pico de respuesta o el final de la respuesta se identifica por la conclusión de un segmento de pendiente positiva, marcado por el inicio de un segmento de pendiente negativa (Bell et al., 1999; National Center for Credibility Assessment, 2017, Nelson 2020). El valor de respuesta extraído se expresa en unidades adimensionales, que representan la distancia máxima entre un pico de respuesta y el inicio de la respuesta anterior dentro del ROW.

Aunque el parámetro convencional de MRL se define como el periodo inicial de 0.5 segundos posterior al inicio del estímulo (Bell et al., 1999), el debate entre los expertos en PDD (Krapohl et al. 2019; Krapohl y Nix, 2019) ha cuestionado si se justifica un MRL más largo, es decir, si el uso de un MRL más largo mejoraría el tamaño del efecto en la PDD. El argumento a favor de ampliar el umbral del MRL parte de la base de que las respuestas fisiológicas requieren más tiempo para desarrollarse, bajo la premisa de que las respuestas tempranas pueden representar artefactos o fluctuaciones autónomas no específicas, en lugar de reacciones que realmente fueron provocadas por el estímulo. Benedek y Kaernbach (2010) sugirieron que las respuestas fásicas rara vez comienzan antes de 1 segundo a partir del inicio del estímulo. En cuanto a los datos de CA, no se encontró ninguna especificación publicada ni norma de práctica de campo para el parámetro MRL. Los profesionales de campo han debatido la posible necesidad de un

parámetro MRL para la extracción de características con datos de CA.

Es importante destacar que, aunque Sjouwerman y Lonsdorf (2019) describieron un «límite» (es decir, un límite superior de una ventana de puntuación óptima) de 3 segundos, basado en +/- 2 desviaciones estándar alrededor de la media para las latencias de respuesta de la conductividad de la piel al utilizar estímulos auditivos, no está claro si su uso del término «latencias» corresponde con la noción de latencia de inicio de la respuesta, como se usa en las pruebas de PDD, o si simplemente se refieren a la latencia entre el inicio del estímulo y el pico de respuesta de la conductividad de la piel. No se pudo encontrar ningún análisis objetivo publicado anteriormente sobre la restricción del MRL para el inicio de la respuesta EDA en los datos de las pruebas de PDD. Dado que la selección de los parámetros de análisis es en última instancia, una pregunta para un estudio empírico, este estudio tiene como objetivo investigar los efectos de diferentes configuraciones de parámetros MRL dentro del contexto de las pruebas PDD con preguntas de comparación. Uno de los objetivos de este proyecto es proporcionar apoyo empírico para la selección de un valor de parámetro MRL que maximice los efectos de los resultados de las pruebas de PDD tanto para los datos EDA como para los CA.

Método

Datos

Este estudio utilizó dos conjuntos de datos de archivo, cada uno de los cuales comprendía exámenes PDD realizados en respuesta a un problema conocido o alegado. Una de las muestras de archivo comprendía 40 exámenes PDD, realizados con el formato Utah de 3 preguntas, después de un simulacro de delito llevado a cabo durante un curso básico de formación en poligrafía acreditado por la APA. Los examinadores eran 10 poligrafistas en entrenamiento que examinaron a otros cuatro estudiantes que habían sido programados como culpables o inocentes. La mitad de los examinados fueron programados como culpables y la otra mitad como inocentes.
que evaluaron a otros cuatro estudiantes que habían sido programados como culpables o inocentes. La mitad de los examinados fueron programados como culpables y la otra mitad como inocentes.

Cada uno de los exámenes de la muestra de laboratorio constaba de tres preguntas relevantes (RQs), tres preguntas de comparación (CQs) y cinco gráficos, utilizando el conjunto estándar de sensores de registro requeridos en PDD (respiración torácica y abdominal, actividad electrodérmica y actividad cardiovascular, además de un sensor de actividad). Estos exámenes también incluían el sensor vasomotor opcional. Todos los exámenes de la muestra de laboratorio se realizaron utilizando sistemas poligráficos computarizados Lafayette (dispositivos LX4000 y LX5000) y la plataforma LXsoftware.

La otra muestra de archivo consistió en 44 exámenes PDD que se realizaron dentro del contexto de investigaciones criminales llevadas a cabo por

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx

agencias policiales que enviaron los exámenes para su inclusión en el archivo de casos confirmados de 2002 del Departamento de Defensa de los Estados Unidos. Estos exámenes se realizaron utilizando diferentes formatos comunes de pruebas PDD, que incluyen el formato de Prueba de Comparación de Zonas Federal, el formato Federal You-Phase, los formatos de Prueba de Preguntas Generales Modificadas de la Fuerza Aérea de los Estados Unidos, que incluyen dos, tres y cuatro RQs, y de dos a cuatro CQs, y el formato de Prueba de Preguntas Generales Modificadas del Ejército de los Estados Unidos, realizada con tres o cuatro RQs y dos o tres CQs. Algunos de los exámenes incluían el registro de cuatro gráficos de prueba. Este análisis solo incluyó los tres primeros gráficos registrados. Sin embargo, un caso solo incluía dos gráficos.

Todos los exámenes de la muestra de campo se realizaron durante la década de 1990, antes de que se exigiera el uso de un sensor de actividad del asiento en 2007, y constan de una serie estándar de sensores PDD puntuados numéricamente. Los examinadores eran examinadores de polígrafo de campo que trabajaban apoyando a investigadores policiales. Los examinados eran sospechosos identificados de delitos, la mitad de los cuales fueron posteriormente declarados culpables y la otra mitad inocentes. La muestra se difundió durante varios años como parte del material promocional que acompañaba a las actualizaciones de software del sistema de polígrafo computarizado Axciton. No se ha conservado la identidad real de los examinados para la muestra de campo, y no se dispone de información sobre los resultados de los casos, salvo la etiqueta del estado confirmado para cada caso.

Todos los datos de ambas muestras se exportaron al formato ASCII de la NCCA (Equipo editorial, 2019). El formato ASCII de la NCCA es un formato de datos de pruebas PDD basado en texto para el que todos los instrumentos poligráficos que cumplen con la Norma de la APA para Instrumentos Poligráficos (Asociación Americana de Poligrafía, 2023a) incluyan la capacidad de exportar e importar desde este formato de datos PDD.

Análisis

Todo el análisis se realizó en el Lenguaje y Ambiente R para Cómputo Estadístico (R Core Team, 2022), incluyendo el procesamiento de señales, la extracción de características y otras tareas de puntuación, la representación gráfica y los cálculos estadísticos.

Para la muestra de laboratorio, se evaluaron 2400 puntos de análisis para el EDA, incluyendo 1200 puntuaciones tanto para los datos del EDA y del CA [3 RQs + 3 CQs) * 5 gráficos * 40 casos = 1200 puntos de análisis]. Había un total de 600 presentaciones de las CQs y 600 presentaciones de las RQs en los datos de la muestra de laboratorio. Hubieron 183 (15%) segmentos de estímulo del EDA para los que no fue posible extraer características. El número real de respuestas extraídas varió en función del período MRL. Para los datos cardíacos,

se extrajeron 472 respuestas en RQs y 486 respuestas en CQs en los datos CA, con 242 (20%) de segmentos de estímulo en los que no fue posible extraer características.

La muestra de campo constó de 766 presentaciones de RQs y CQs, que incluyeron 389 presentaciones de RQs y 377 presentaciones de CQs. Ciento sesenta y siete (167) (22 %) de las presentaciones de estímulos no produjeron ninguna respuesta CA utilizable. Ciento treinta y ocho (138) (18 %) de las respuestas EDA no fueron utilizables. El número real de respuestas extraídas varió con el período MRL.

Aunque ni la muestra de laboratorio ni la de campo son adecuadas para estimar los tamaños del efecto del polígrafo, como son la sensibilidad, la especificidad o las tasas de error de la prueba - los cuales contribuyen con los efectos de la precisión de la prueba - ambas muestras tienen estados de criterio claramente etiquetados y, por lo tanto, son ventajosas para evaluar los efectos de diferentes grupos de parámetros de análisis, como el MRL. Para cada una de las muestras, todos los casos se puntuaron con una versión totalmente automatizada del Sistema de Puntuación Empírica - Multinomial (ESS-M), que incluye la extracción automatizada de características, la transformación numérica y la reducción de datos automatizada, la comparación automatizada de las puntuaciones de las pruebas con el modelo de referencia estadístico y la ejecución automatizada de reglas de decisión para clasificar los resultados como estadísticamente significativos para el engaño o la veracidad, o como inconclusos si no son estadísticamente significativos.

Para ambas muestras - la de laboratorio (n = 40) y la de campo (n = 44) - se calculó la latencia de respuesta para todas las respuestas en todas las CQs y todas las RQs. Uno de los objetivos de este estudio era investigar la hipótesis de que pueden existir algunas diferencias en los tiempos de latencia de respuesta para las CQs y las RQs en los exámenes de laboratorio y de campo. Otro objetivo de este proyecto era investigar la hipótesis de que una MRL superior a 0.5 segundos podría aumentar la contribución de los datos de EDA y CA para lograr clasificaciones correctas de engaño y veracidad - indicadas por la fuerza de la asociación con el estado de culpabilidad o inocencia fuertemente etiquetado en las muestras de archivo de laboratorio y de campo.

Las respuestas serían rechazadas si la latencia de respuesta no excediera el parámetro MRL. Ya que es posible que, tras el rechazo de una respuesta no utilizable durante el MRL, se produzca una segunda respuesta utilizable dentro del ROW, el análisis no era tan simple como rechazar las respuestas que comenzaban durante el periodo MRL. En su lugar, sería necesario ejecutar iterativamente los procedimientos de extracción de características, para todas las respuestas y todos los casos, utilizando un vector de parámetros MRL. En otras palabras, el análisis no consistía simplemente en excluir progresivamente las respuestas basándose en un parámetro MRL creciente. En cambio, todos los casos se reevaluaron utilizando un vector de

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx

parámetros MRL tanto para los datos EDA como para los CA. Aunque el cálculo de esto sería prohibitivo en un paradigma de puntuación manual, es más fácil de lograr utilizando una versión totalmente automatizada - y por lo tanto objetiva - del método ESS-M para TDA (Nelson, 2024).

Todos los casos de ambas muestras se puntuaron 11 veces, utilizando un algoritmo ESS-M automatizado, con un vector de parámetros MRL de 0.0 segundos a 5.0 segundos. Para cada iteración se guardó el espacio del parámetro [0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0], del inicio de la respuesta, el pico de respuesta y la magnitud de la respuesta para las respuestas del EDA y CA para cada presentación de las CQs y RQs. Las relaciones R/C se calcularon después de seleccionar los puntos de análisis de acuerdo con los procedimientos publicados para los formatos de prueba. Las relaciones R/C se sometieron a transformaciones logarítmicas naturales para corregir la asimetría de la relación y permitir el cálculo de los coeficientes de correlación biserial de punto. Posteriormente se transformaron en puntuaciones ESS-M enteras las relaciones log(R/C) tanto para las respuestas del EDA como del CA. Se sumaron por separado las puntuaciones ESS-M para las puntuaciones EDA y CA de cada una de las muestras archivadas. Y entonces se calcularon las correlaciones biseriales puntuales para las puntuaciones ESS-M de EDA y CA y los estados criterio etiquetados.

Se planteó la hipótesis de que aumentar la duración del MRL podría aumentar los efectos de precisión de la prueba, por lo que se pensó que las puntuaciones enteras del ESS-M podrían proporcionar una comprensión más directa o intuitiva del efecto del MRL. Por esta razón, se calculó la proporción de puntuaciones ESS-M correctas para las puntuaciones EDA y CA para cada una de las 11 iteraciones de puntuación sobre los diferentes parámetros del vector MRL desde 0.0 y hasta 5.0 segundos. Finalmente, ya que también se supuso que el aumento en la longitud del parámetro MRL podría dar lugar a una disminución en los efectos deseados de precisión como consecuencia del aumento de la proporción de respuestas para las que no es posible la extracción de características, también se calculó la proporción de puntuaciones distintas a cero para cada iteración del espacio de parámetros MRL.

Resultados

Tiempo de latencia de respuesta. Mientras que el MRL se refiere al parámetro mínimo de latencia de respuesta, el tiempo de latencia de respuesta se refiere al tiempo transcurrido desde el inicio del estímulo hasta el inicio de la respuesta. Para la muestra de laboratorio, el tiempo medio combinado de latencia de respuesta fue de 3.0 segundos, con una mediana de 2.7 segundos. La muestra de campo produjo una latencia de respuesta media combinada de 2.4 segundos, con una mediana de 2.5 segundos. La tabla 1 muestra un resumen de las medias, medianas y cuantiles, junto con los límites superior e inferior para las CQs y RQs de las muestras de campo y

de laboratorio.

Tabla 1. Tiempo de latencia de respuesta para la muestra de laboratorio y la muestra de campo.

Tabla 1. Tiempo de latencia de respuesta para muestras de laboratorio y de campo.

Muestra de laboratorio (n = 40)		Mín*	1.°cuartil	Mediana	Media	3.°cuartil	Máx
	RQ	0.5 s	1.6 s	2.6 s	2.7 s	3.6 s	8.5 s
	CQ	0.5 s	1.9 s	2.8 s	3.2 s	4.3 s	8.9 s
	Combinado	0.5 s	1.7 s	2.7 s	3.0 s	3.9 s	8.9 s
Muestra de campo (n=44)		Mín*	1.°cuartil	Mediana	Media	3.°cuartil	Máx
	RQ	0.5 s	1.7 s	2.6 s	2.5 s	3.1 s	9.0 s
	CQ	0.5 s	1.6 s	2.3 s	2.6 s	3.2 s	9.7 s
	Combinado	0.5 s	1.6 s	2.4 s	2.5 s	3.2 s	9.7 s

* Esta parte del análisis se completó utilizando el MRL tradicional, lo que significa que no se extrajeron las respuestas con tiempos de latencia inferiores a 0.5 s.

El tiempo medio de latencia de respuesta fue de 2.6 segundos para las RQs tanto en las muestras de campo como en las de laboratorio. La latencia media de respuesta fue de 2.8 segundos para la muestra de laboratorio y de 2.3 segundos para la muestra de campo. Sin embargo, esta diferencia puede atribuirse a las diferencias entre las condiciones de campo y de laboratorio, y puede deberse a diferencias entre los tipos de CQs. Los casos de laboratorio se llevaron a cabo utilizando CQs de mentira directa (DLCQ), mientras que los casos de campo se llevaron a cabo utilizando CQs de mentira probable (PLCQ).

La variación en las latencias de respuesta entre las muestras de campo y de laboratorio, así como entre los DLCQ y los PLCQ, sugiere posibles diferencias sistemáticas, aunque se requiere una investigación más profunda. Además, dada la imposibilidad de aislar el origen de la diferencia observada (laboratorio frente a campo o DLCQ frente a PLCQ), los datos que se muestran en la tabla 1 indican que la distribución de las diferencias en los tiempos de latencia de respuesta está sesgada y no se ajusta a los supuestos lineales que requiere un ANOVA paramétrico de dos vías. Por esta razón, no se completó ningún análisis 2x2 para investigar más a fondo estas diferencias. Los desequilibrios en el tamaño de las muestras también complican un análisis más profundo. En futuros trabajos se podría emplear un ANOVA de permutación - procedimiento que podría tolerar tanto el desequilibrio en el muestreo como la no linealidad - pero esta tarea excede el alcance del presente proyecto.

Latencia mínima de respuesta. Los resultados de los efectos del parámetro MRL se muestran gráficamente, trazados para los 4 diferentes efectos a lo largo del vector de valores de los parámetros MRL para los datos EDA y CA de

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx

las muestras de laboratorio y de campo. Los efectos representados incluyeron la correlación biserial de punto de las relaciones log(R/C) con el estado del criterio de culpabilidad frente al de inocencia, la correlación biserial puntual de las puntuaciones ESS-M, junto con la proporción de puntuaciones ESS-M que coinciden con el estado del criterio del caso y la proporción de puntuaciones distintas de cero.

Figura 1. Análisis de latencia EDA (muestra de laboratorio, n = 40)

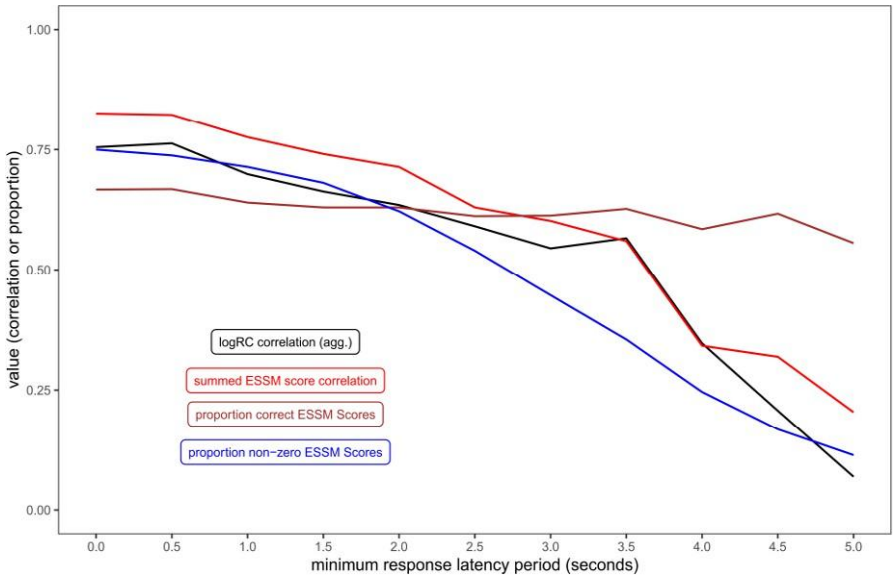


Figura 2. Análisis de latencia cardíaca (muestra de laboratorio, n = 40)

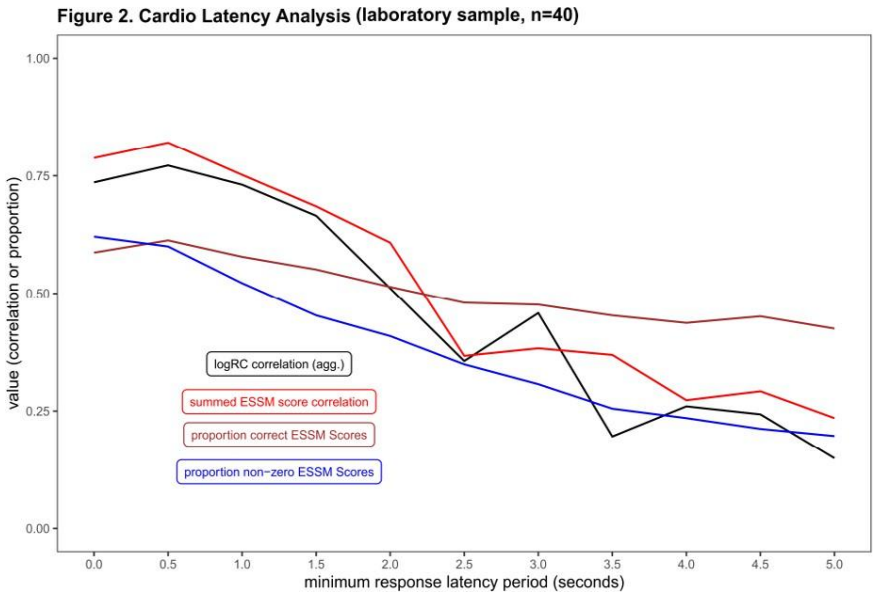


Figura 3. Análisis de latencia EDA (muestra de campo, n = 40)

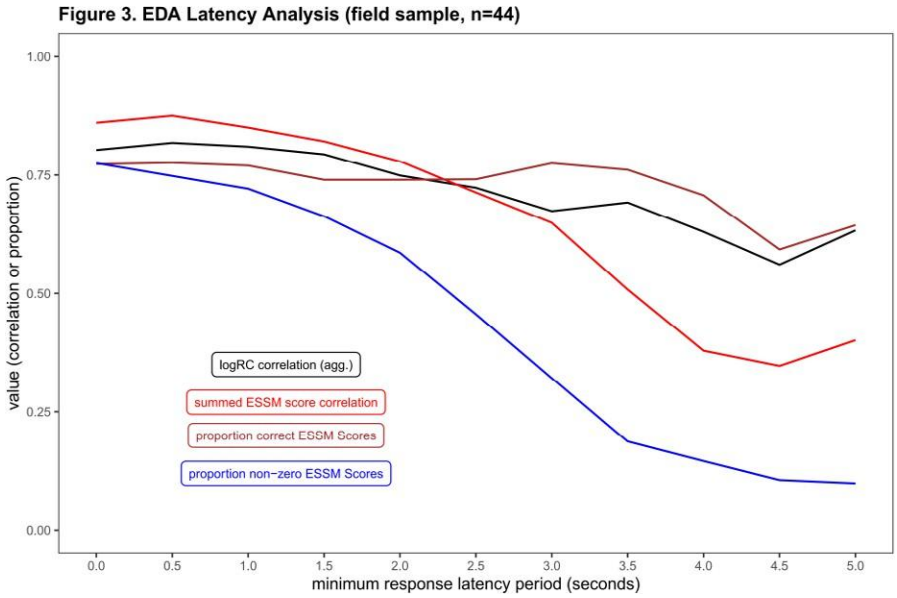
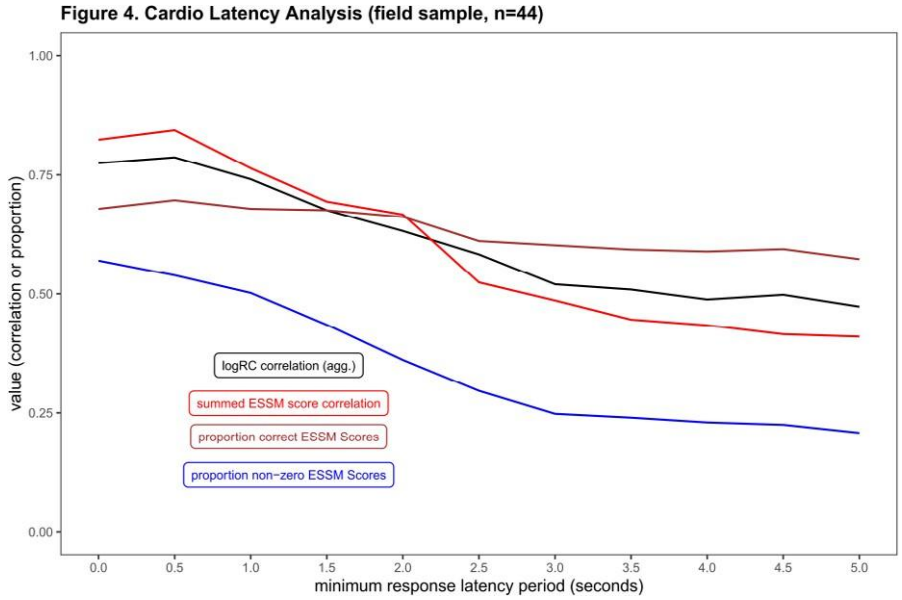


Figura 4. Análisis de latencia cardíaca (muestra de campo, n = 44)



La tabla 2 muestra las correlaciones biseriales de punto para las proporciones log(R/C) de los datos EDA y CA y la verdad de terreno para las muestras de campo y de laboratorio. Las correlaciones son fuertes sin una restricción MRL y aumentan ligeramente hasta alcanzar un pico de correlación utilizando un parámetro MRL de 0.5 segundos. Las correlaciones se debilitan progresivamente a medida que aumenta la restricción MRL.

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx

Tabla 2. Correlación biserial puntual de las relaciones log (R/C) con la verdad fundamental. Medias de bootstrap (SE) [95 %]

MRL value	EDA (lab)	CA (lab)	EDA (field)	CA (field)
.0s	.755 (.049) [.651, .840]	.736 (.054) [.620, .828]	.804 (.043) [.721, .883]	.775 (.043) [.687, .849]
.5s	.763 (.048) [.660, .845]	.772 (.047) [.676, .858]	.815 (.037) [.737, .881]	.785 (.045) [.687, .863]
1.0s	.699 (.066) [.561, .812]	.731 (.061) [.591, .833]	.811 (.046) [.709, .892]	.739 (.054) [.628, .833]
1.5s	.663 (.079) [.491, .800]	.665 (.074) [.509, .790]	.794 (.044) [.700, .871]	.681 (.065) [.544, .795]
2.0s	.635 (.078) [.477, .773]	.511 (.122) [.239, .716]	.753 (.064) [.612, .86]	.633 (.082) [.445, .772]
2.5s	.591 (.099) [.385, .759]	.356 (.131) [.092, .581]	.728 (.073) [.562, .849]	.588 (.087) [.405, .751]
3.0s	.545 (.093) [.335, .701]	.458 (.103) [.215, .639]	.674 (.096) [.455, .829]	.521 (.119) [.256, .718]
3.5s	.566 (.098) [.355, .726]	.196 (.178) [<.001, .568]	.691 (.099) [.485, .859]	.500 (.125) [.224, .715]
4.0s	.347 (.157) [.028, .632]	.26 (.159) [<.001, .547]	.629 (.132) [.342, .853]	.489 (.122) [.224, .706]
4.5s	.207 (.210) [<.001, .634]	.243 (.164) [<.001, .548]	.569 (.154) [.245, .825]	.497 (.122) [.233, .703]
5.0s	.069 (.191) [<.001, .488]	.149 (.167) [<.001, .452]	.637 (.129) [.359, .851]	.465 (.128) [.179, .681]

Discusión

Como era de esperar, la proporción de puntuaciones distintas de cero se maximizó sin el uso de la restricción MRL. La inspección de la tabla 1 y las figuras 1-4 muestran que una restricción de 3 segundos no haría posible la extracción de características (sin puntuación) para más del 50% de las presentaciones de estímulos, tanto para las muestras de campo como para las de laboratorio.

En contraste con la proporción de puntuaciones distintas de cero, la proporción de puntuaciones correctas fue menos vulnerable a los aumentos del parámetro MRL, con un ligero aumento en la proporción de puntuaciones correctas de EDA y CA utilizando una restricción de 0.5 segundos. Se observaron algunas diferencias en los casos de laboratorio y de campo para esta proporción. Los casos de laboratorio mostraron disminuciones progresivas en la proporción de puntuaciones EDA y CA correctas. Sin embargo, los casos de campo mostraron un ligero aumento en la proporción de puntuaciones correctas con una restricción MRL de EDA de 3.0 segundos. La razón de esto sigue siendo desconocida. Sin embargo, es posible que se deba a diferencias en el procesamiento de la señal EDA. Los datos EDA de los

casos de laboratorio se procesaron utilizando un método que cumple con los estándares de procesamiento de señales EDA publicado por la American Polygraph Association (2023b), mientras que los casos de campo se registraron con un sistema de centrado automático basado en hardware. No se utilizó ningún procesamiento de señal adicional para los datos EDA de los casos de campo.

Las correlaciones biseriales de punto para las puntuaciones log(R/C) y ESS-M alcanzaron su máximo nivel, utilizando una restricción de 0.5 segundos tanto para los datos EDA como para los CA de ambas muestras. Incluso pequeños aumentos en el umbral MRL provocan rápidas disminuciones en las relaciones log(R/C) y las correlaciones de las puntuaciones ESS-M, ya que la eliminación de las respuestas tempranas debilita de manera desproporcionada la asociación estadística entre los datos y el criterio conocido de engaño/veracidad tanto para los sensores EDA como para los CA.

Limitaciones

Este proyecto es el primer intento de publicar un análisis empírico de los cambios en los efectos de la prueba PDD para el parámetro MRL. Un objetivo secundario de este proyecto era presentar los resultados analíticos en forma gráfica con una carga estadística mínima para los lectores interesados. No se completó ninguna prueba estadística y se observaron algunas dificultades en la posible selección de un método estadístico (desequilibrio en el tamaño de la muestra y no linealidad). No se proporciona un análisis acerca de las diferencias del tamaño del efecto al inferir o introducir el inicio de una respuesta debido a un cambio en la inflexión cuando la pendiente de los datos EDA o CA ya son positivos antes del inicio del estímulo, sin ningún segmento de pendiente negativa dentro del ROW. Este proyecto se completó mientras se inferían algunas ubicaciones de inicio de la respuesta a partir de los cambios en la inflexión. Este parámetro debería ser objeto de futuras investigaciones.

Otras limitaciones de este proyecto incluyen el limitado alcance del objetivo. No se ofrece ningún análisis de optimización para los parámetros ROW o WOE. El parámetro ROW, cuyo punto final está anclado a la respuesta verbal, está sujeto a variaciones entre los profesionales de campo, que pueden introducir variaciones de tiempo al registrar el punto de la respuesta verbal. El parámetro ROW se ve influido además por la longitud y la duración de cada pregunta del estímulo.
Por último, no se realizó ningún análisis para la selección de la línea cardíaca óptima para el análisis (diastólica, sistólica o línea media). Los datos de CA para este proyecto se analizaron utilizando la línea media. Las investigaciones futuras deberían explorar el uso del modelado jerárquico bayesiano como método alternativo de selección de parámetros para estimar los umbrales de latencia óptimos, teniendo en cuenta la variabilidad de la respuesta

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx

fisiológica.

Conclusión

Los datos de las pruebas PDD se registran como datos de series temporales - en los que se muestrean y registran una serie de valores numéricos con una tasa de muestreo fija - y suelen mostrarse en forma gráfica en el que el eje x (dimensión horizontal) indica la escala de tiempo, mientras que el eje y (dimensión vertical) indica el delta o cambio en los valores de los datos. Las pruebas PDD se basan fundamentalmente en el estímulo y la respuesta. La información contenida en una respuesta puede contribuir a los efectos deseados de la prueba (es decir, la discriminación y clasificación del engaño y la veracidad) solo si los cambios observados en la actividad fisiológica pueden atribuirse razonablemente a los estímulos de la prueba. Se utilizan tres parámetros de extracción de características para determinar si un cambio observado en la actividad fisiológica es oportuno ante una pregunta de estímulo de la prueba. Estos son los parámetros WOE, ROW y MRL. Ya que los datos consisten en valores observados que se someten a análisis, estos parámetros sirven para estructurar el proceso de extracción de características.

Este proyecto, que consistió en dos análisis con datos confirmados de campo y de laboratorio, ofrece el primer análisis empírico del parámetro MRL para datos EDA y CA. Aunque se han proporcionado sugerencias para el uso de un período más largo, no se pudo encontrar ningún análisis publicado anteriormente. Además, no se ha podido encontrar ninguna especificación MRL publicada para datos CA. Este análisis mostró que un parámetro MRL de 0.5 segundos producía la correlación log(R/C) máxima para datos CA.

Este análisis respalda continuar con el uso del parámetro MRL tradicional de 0.5 segundos para la extracción de características EDA. Además, este análisis respalda el uso de un parámetro MRL de 0.5 segundos para los datos CA. Aumentar el MRL por encima de 0.5 segundos debilita los tamaños del efecto estadístico al eliminar respuestas fisiológicas tempranas pero válidas, lo que reduce el número de puntos de datos utilizables y disminuye la fuerza de la precisión de la clasificación. Esto no es sorprendente si se tiene en cuenta que las descripciones publicadas de los algoritmos informáticos en el pasado no incluían ninguna descripción del parámetro MRL y presumiblemente incluían todos los inicios de respuesta después del inicio del estímulo, al tiempo que proporcionaban resultados cuya precisión era igual o superior a la de la puntuación manual realizada por expertos humanos.

Los resultados de este estudio no respaldan la hipótesis de que una MRL extendida aumentará los efectos deseados en la precisión del PDD. Por el contrario, los datos indican que alargar el parámetro MRL puede contribuir a aumentar los resultados inconclusos. Aunque la sugerencia de alargar el parámetro MRL puede parecer intuitiva para algunos, la ciencia es un proceso de adhesión a la evidencia. Las pruebas disponibles en este momento sugieren que se puede extraer información útil de las respuestas EDA y CA

que comienzan después de 0.5 segundos tras el inicio del estímulo. Y aunque pueda parecer contrario a la intuición y difícil de entender basándose en el estado actual de los conocimientos psicofisiológicos, ignorar las pruebas en favor de la intuición acabaría por impedir la capacidad de aprender más en el futuro. Hay abundantes ejemplos en muchas áreas de la investigación científica y de práctica de campo basada en la evidencia, en las que el análisis de datos desafía tanto nuestra intuición como los límites de nuestra comprensión actual. Más importante aún, la identificación de inconsistencias entre la hipótesis intuitiva y la evidencia replicable proporciona una oportunidad potencial y una dirección para ampliar aún más nuestro conocimiento.

Esto no significa que las conjeturas anteriores carezcan por completo de fundamento. Más bien, sugiere que los debates retóricos son incompletos. Los efectos que se muestran en este análisis - que indican que un MRL extendido puede ser problemático - sugieren que es posible que los examinados, tras haber revisado cuidadosamente todas las preguntas de estímulo de las pruebas - antes de la colocación de los sensores y el registro de los datos de la prueba - son más conscientes de las preguntas, incluyendo los tipos de preguntas (RQ o CQ) que han decidido responder con sinceridad o de forma engañosa, de lo que se había supuesto anteriormente.

Este proyecto investigó el parámetro MRL y encontró evidencia empírica que respalda las prácticas existentes en el campo, aunque difieren de algunas opiniones profesionales. Las opiniones profesionales, a diferencia de las opiniones personales, se basan en el conocimiento, la experiencia y los principios establecidos, a menudo respaldados por la investigación y los estándares profesionales (Kruglanski y Gigerenzer, 2011). Por el contrario, las opiniones personales se ofrecen de forma subjetiva, sin pruebas ni análisis reproducibles, y a menudo se utilizan para expresar preferencias personales que pueden introducir fácilmente sesgos en la toma de decisiones profesionales. Es necesario continuar el debate sobre el parámetro MRL para armonizar las recomendaciones prácticas y las opiniones profesionales con las pruebas empíricas disponibles.

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx

References

American Polygraph Association. (2023a). *APA Standard for Polygraph Instrumentation.*

American Polygraph Association. [Retrieved online from https://www.polygraph.org/docs/APA_Standard_for_Polygraph_Instrumentation_approved_2023Aug25.pdf on September 25, 2024.]

American Polygraph Association. (2023b). *Standard for EDA Signal Processing.*

American Polygraph Association. [Retrieved online from https://www.polygraph.org/docs/Standard_for_EDA_Signal_Processing_Final.pdf on September 25, 2024.]

Bell, B. G., Raskin, D. C., Honts, C. R., & Kircher, J. C. (1999). The Utah numerical scoring system. *Polygraph, 28*(1), 1-9.

Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*(4), 647–658. Editorial Staff. (2019).

Introduction to the NCCA ASCII standard. *Polygraph & Forensic Credibility Assessment, 48* (2), 125-135.

Kircher, J. C. & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73*(2), 291-301.

Krapohl, D. J., Dutton, D. W. & Nix, K. A. (2019). A brief discussion of the lower latency limit of the electrodermal response in polygraph testing. *Polygraph & Forensic Credibility Assessment, 48* (2), 98-104.

Krapohl, D. J., & Nix K. (2019). Evidence and common sense: suggestions for scoring electrodermal responses. *APA Magazine, 52*(1), 60-76.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review, 118*(1), 97–109.

Nelson, R. (2020). Accuracy effects for ESS and three-position scores of Federal ZCT exams using the grand total rule with traditional/federal and multinomial cutscores. *Polygraph & Forensic Credibility Assessment, 49*(2), 157-175.

Nelson, R. (2024). Reference guide for the Empirical Scoring System – Multinomial (ESS-M). *Polygraph & Forensic Credibility Assessment, 53*(2), 103-145.

Nelson, R. & Handler. M. (2013). Pneumograph signal processing and feature extraction. *Polygraph, 42*(3), 163-172.

R Core Team. (2022). R: A Language and Environment for Statistical Computing.

Vienna, Austria: R Foundation for Statistical Computing. [https://www.R-project.org/]

Sjouwerman, R., & Lonsdorf, T. B. (2019). Latency of skin conductance responses across stimulus modalities. *Psychophysiology, 56*(4), e13307.

Timm, H. W. (1982) Analyzing deception from respiration patterns. *Journal of Police Science and Administration, 10*(1), 47-51

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor: rodolfo@poligrafia.com.mx