

## **Medida de la sensibilidad en experimentos de vigilancia: Consecuencias estadísticas del uso de índices basados en la teoría de detección de señales**

Manuel J. Blanco\* y David Soto

Universidad de Santiago de Compostela

Este artículo examina si el decremento de sensibilidad que frecuentemente se observa en tareas de vigilancia puede estar afectado por el índice o estadístico utilizado para medir la sensibilidad. Informamos de los resultados de un estudio de simulación de dicho decremento utilizando dos medidas de sensibilidad:  $d'$  y  $A'$ . Los resultados de estas simulaciones indican que el decremento es algo mayor con  $d'$  que con  $A'$  y que la probabilidad de que dicho decremento resulte significativo en un análisis de varianza de los datos es también mayor para  $d'$  que para  $A'$ .

**Palabras clave:** Atención sostenida, Teoría de detección de señales, Vigilancia.

La atención sostenida es la habilidad para mantener la atención en alguna fuente de información con objeto de informar de la ocurrencia de un estímulo señal durante un período de tiempo prolongado. En el laboratorio, esta habilidad se investiga generalmente utilizando las llamadas “tareas de vigilancia”. Estas tareas implican la observación prolongada de una pantalla y requieren informar de un estímulo señal infrecuente e impredecible que aparece mezclado con estímulos ruido frecuentes. El hallazgo básico en la investigación con estas tareas es el “decremento de vigilancia” o declive en la proporción de señales correctamente detectadas durante el período de observación.

La aplicación de la teoría de detección de señales (TDS) al análisis del rendimiento en tareas de vigilancia (Broadbent y Gregory, 1963), lo que ocurrió muy pronto después de que la teoría fuese introducida a los psicólogos (Swets, Tanner y Birdsall, 1961), ayudó a establecer las condiciones bajo las cuales el decremento de vigilancia puede atribuirse a un proceso sensorial o a otros procesos de nivel superior. Esta teoría deja bien

---

\* Manuel J. Blanco. Facultad de Psicología, Departamento de Psicología Social y Básica, 15706 Santiago de Compostela, España. E-mail: psmlanc@usc.es Esta investigación fue subvencionada por la Xunta de Galicia con la beca XUGA21107B98.

claro que el declive en la proporción de aciertos (proporción de señales correctamente detectadas o identificadas) a lo largo del tiempo de trabajo puede resultar de una pérdida de sensibilidad o de cambios en el criterio de decisión o respuesta. Puesto que el decremento en la proporción de señales detectadas está normalmente acompañado de un declive en la proporción de falsas alarmas (proporción de respuestas “sí” a estímulos ruido), el decremento de vigilancia puede deberse a un deterioro en la habilidad del observador para discriminar los estímulos señal de los estímulos ruido, o bien a un proceso relacionado con la selección de respuesta que hace que el observador sea más cauto para informar de la presencia del estímulo señal.

Desde un punto de vista metodológico, pueden distinguirse tres etapas en la investigación sobre vigilancia. La primera abarca desde los primeros estudios de Mackworth (1948), al que se considera que fue el primero en investigar en el laboratorio el problema del decremento de vigilancia, hasta los años sesenta. Estos estudios se caracterizan, siempre desde un punto de vista metodológico, porque el decremento en la proporción de detecciones o aciertos se consideraba que era resultado de un declive en el nivel de sensibilidad y no se contemplaba la posibilidad de que parte de este decremento se debiese a cambios en el criterio de decisión.

La segunda etapa abarca desde los años 60 hasta 1977, que es el año en el que Swets publicó una revisión de la TDS aplicada a vigilancia. Estos estudios se caracterizan por la aplicación de los métodos técnicos basados en la TDS al análisis de los datos de vigilancia. En la mayoría de los experimentos de vigilancia realizados durante este período, los investigadores empezaron a reconocer que lo que había sido definido como “decremento de vigilancia” (un declive en la proporción de aciertos a lo largo de la tarea) iba generalmente acompañado de un decremento en la proporción de falsas alarmas. Si esto ocurriese, entonces el llamado “decremento de vigilancia” podría resultar de un cambio en el criterio de decisión adoptado por el observador, de un cambio en la sensibilidad o de ambos. Si esto se aceptaba, como así ocurrió, los resultados experimentales de la etapa previa se hacían ahora difíciles de interpretar. Durante este período, en la inmensa mayoría de investigaciones se utilizó el conocido índice  $d'$  como medida de sensibilidad. Como es bien sabido,  $d'$  exige asumir supuestos matemáticos específicos (véase más adelante) respecto a las distribuciones de probabilidad asociadas a los efectos sensoriales que el estímulo (señal o ruido) produce en el observador. Por consiguiente, el uso generalizado de  $d'$  durante esta etapa de la investigación sobre vigilancia descansaba sobre estos supuestos teóricos. Fue Swets (1977) el primero en reconocer la necesidad

de que los investigadores probasen dichos supuestos en el caso de que optasen por emplear  $d'$  como medida de sensibilidad.

El tercer y último período abarca los estudios experimentales realizados durante los últimos 25 años, aproximadamente. La característica metodológica de estos últimos estudios es que, inmediatamente después de la sugerencia hecha por Swets (1977), el índice paramétrico de la TDS,  $d'$ , ha ido sustituyéndose progresivamente por  $A'$ , que es un índice no paramétrico o libre de supuestos referentes a las distribuciones de probabilidad asociadas a los efectos sensoriales del estímulo. Puesto que la definición de  $d'$  es en términos de las distribuciones de probabilidad subyacentes, este índice parece más apropiado que  $A'$  cuando estas distribuciones han sido probadas, pero cuando estas distribuciones no pueden examinarse, entonces  $A'$  parece el índice más apropiado.

See, Howe, Warm y Dember (1995) hicieron un meta-análisis con datos tomados de 42 experimentos realizados entre 1980 y 1992. El 62% de estos experimentos utilizaron  $A'$  como índice de sensibilidad y el 38% restante  $d'$ . See y colaboradores (1995) calcularon el tamaño de efecto como indicador del decremento de sensibilidad, definido como la diferencia entre los niveles de sensibilidad en el primer y último período de la prueba dividida por la raíz cuadrada del error cuadrático medio para el factor Tiempo observado en el análisis de varianza de los datos experimentales. Entre otras cosas, See y colaboradores (1995) compararon los tamaños de efecto medios para experimentos en los que la variable dependiente había sido  $d'$  con aquellos correspondientes a los experimentos en los que la variable dependiente había sido  $A'$ . El resultado de esta comparación indicó que la magnitud del decremento no estaba relacionada con el índice de sensibilidad. Sin embargo, el problema de esta comparación es que los dos índices de sensibilidad se computaron a partir de datos experimentales diferentes, puesto que es usual que los autores sólo informen de resultados tomando un único índice, aunque existe alguna excepción.

Por ejemplo, Parasuraman y Mouloua (1987) examinaron los efectos del tipo de discriminación señal/ruido (sucesiva o simultánea) y de la razón señal/ruido (baja, moderada y alta) sobre el decremento de sensibilidad en tareas de vigilancia. Calcularon los dos índices de sensibilidad usuales,  $d'$  y  $A'$ , para cada condición y, para cada uno, realizaron un análisis de varianza (ANOVA). El ANOVA de los valores  $d'$  mostró efectos principales significativos del tipo de discriminación, razón señal/ruido y tiempo en la tarea, mientras que el ANOVA de los valores  $A'$  sólo mostró efectos significativos de la razón señal/ruido y tiempo, pero no del tipo de discriminación.

El propósito de este artículo es comprobar cómo el decremento de sensibilidad, indicado por el tamaño del efecto, puede variar en función de la medida de sensibilidad que utilice el investigador. En primer lugar, haremos una breve descripción de las principales medidas de sensibilidad utilizadas en la investigación de la vigilancia. Más adelante, ilustraremos cómo el índice de sensibilidad puede afectar al decremento de sensibilidad con los resultados de simulaciones por ordenador de datos de experimentos de vigilancia.

Asumiremos una tarea de vigilancia típica en la que existen sólo dos tipos de ensayos: 1) ensayos en los que se presenta un estímulo señal y 2) aquellos en los que se presenta un estímulo ruido. En cada ensayo, la tarea del observador es decidir si el estímulo es o no señal. Un acierto se define como una respuesta “sí” a un estímulo señal, mientras que una respuesta “sí” cuando el estímulo es ruido constituye una falsa alarma. La proporción de aciertos es la probabilidad condicional de una respuesta “sí” dado que un estímulo señal ha sido presentado, y la proporción de falsas alarmas es la probabilidad condicional de una respuesta “sí” dado que el estímulo era ruido. Estas dos probabilidades especifican conjuntamente la sensibilidad o grado en que el observador puede discriminar los dos tipos de estímulos.

En la TDS, la representación sensorial del estímulo,  $X$ , es una variable aleatoria continua. Esta variable tiene asociada una función de densidad de probabilidad,  $S$  o  $f(x/\text{señal})$ , cuando el estímulo es señal y otra,  $R$  o  $f(x/\text{ruido})$ , cuando el estímulo es ruido. El observador establece un valor específico de  $X$ , o cualquier transformación monotónica de este valor (p.e. la razón de verosimilitud, o probabilidad de que la respuesta del sistema sensorial provenga de  $S$  en relación a la probabilidad de que dicha respuesta provenga de  $R$ ) como criterio o punto de corte para responder. Si él/ella considera que la observación es mayor que el criterio responde “sí” y, en caso contrario, responde “no”.

La primera medida de sensibilidad basada en la TDS utilizada en la investigación de la vigilancia es también el primer índice definido en la aplicación psicológica de esta teoría, el índice de detectabilidad  $d'$  (Tanner y Swets, 1954). Este índice descansa sobre dos supuestos referentes a las distribuciones de densidad de probabilidad asociadas a los dos posibles estímulos: 1) ambas distribuciones son normales (gaussianas) y 2) ambas distribuciones tienen la misma varianza.  $d'$  se define en términos de integrales de estas dos distribuciones, o puntuaciones  $z$ , y toma un valor igual o mayor a 0, con 0 representando incapacidad para discriminar el estímulo señal del estímulo ruido.

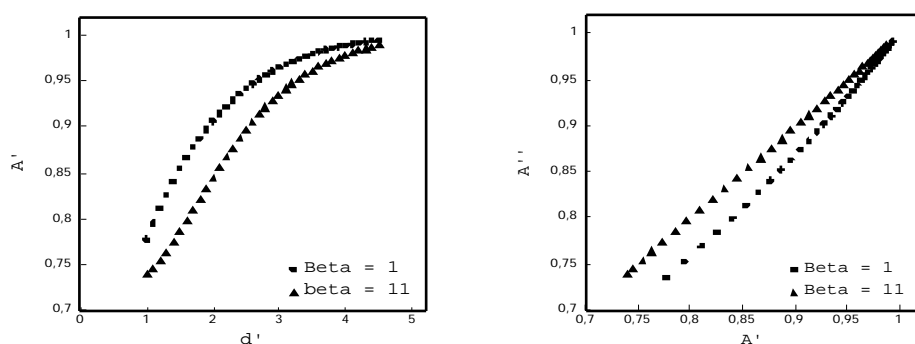
El uso de  $d'$  está sólo justificado cuando los supuestos matemáticos acerca de las funciones de densidad de probabilidad subyacentes han sido confirmados. Esta prueba requiere obtener la llamada curva de operación característica del receptor (ROC, abreviaturas en inglés) para cada observador. La ROC es una función que relaciona las probabilidades de falsas alarmas y aciertos conforme el criterio de decisión varía mientras los factores que supuestamente afectan a la sensibilidad permanecen constantes.

Dado que estas probabilidades están afectadas por las funciones de densidad de probabilidad subyacentes, la forma de la curva ROC también está determinada por las características de estas distribuciones. La ROC predicha por el modelo de distribuciones normales e idéntica varianza, dentro del cual se define  $d'$ , es lineal y paralela a la diagonal positiva del espacio ROC cuando ambas probabilidades condicionales se transforman a puntuaciones  $z$ .

Las primeras aplicaciones de la TDS en investigaciones de vigilancia utilizaron  $d'$  como medida de sensibilidad (véase Broadbent, 1971, para una revisión de estos primeros trabajos), pero los autores trataban de justificar su utilización calculando las funciones ROC y probando que estas funciones se ajustasen bien a las predichas por el modelo de la TDS que asume distribuciones normales y de la misma varianza. Sin embargo, la práctica común en la mayoría de los experimentos de vigilancia hechos con posterioridad ha sido emplear  $d'$  como índice de sensibilidad pero con ningún estudio previo de la función ROC que pudiera justificar ese uso. La situación comenzó a cambiar en los años 70 cuando el índice  $A'$  sustituyó a  $d'$  como medida de sensibilidad, sobre todo después de la recomendación hecha en este sentido por Swets (1977).

$A'$  se basa en una idea original de Norman (1964) y de Pollack y Norman (1964), pero su uso en psicología no apareció hasta que Grier (1971) proporcionó las fórmulas computacionales.  $A'$  es una estimación del área bajo la curva ROC cuando sólo un único punto ROC es conocido. Más específicamente,  $A'$  es la media de las áreas ROC máxima y mínima bajo todas las curvas ROC posibles que atraviesan un punto ROC dado. En realidad,  $A'$  no es exactamente la media algebraica de estas áreas, como hasta hace poco se pensaba. Más bien,  $A'$  debe interpretarse como una media ponderada de estas áreas, con un peso mayor asociado al límite superior y variando este peso con la posición del punto. Smith (1995) ha llamado recientemente la atención sobre este aspecto de  $A'$  y ha propuesto otro índice para sustituirlo,  $A''$ , que sí tendría la interpretación original de  $A'$  (media algebraica no ponderada de las áreas mínima y máxima correspondientes a las curvas ROC que atraviesan un punto ROC dado). Al

igual que  $A'$ ,  $A''$  toma valores entre 0.5, que significa una discriminación señal/ruido nula, y 1, que representa una discriminación perfecta. Ambos índices,  $A'$  y  $A''$ , pertenecen a la categoría de índices denominada habitualmente “índices libres de distribución”, porque no dependen de supuestos matemáticos acerca de las distribuciones  $S$  y  $R$  descritas anteriormente. Sin embargo, ambos índices descansan sobre un supuesto



**Figura 1.** Relaciones entre tres índices diferentes de sensibilidad:  $d'$ ,  $A'$  y  $A''$ .

importante acerca de la función ROC (véase el artículo de Smith, 1995, para una discusión de este supuesto). El supuesto es que la curva ROC subyacente debe ser “propia”. Se denominan ROCs propias (véase Egan, 1975) a aquellas que tienen una pendiente que es no creciente.

La Figura 1 representa las relaciones entre los tres índices descritos ( $d'$ ,  $A'$  y  $A''$ ) para dos niveles del criterio de decisión,  $\beta = 1$  y  $\beta = 11$ . Estas relaciones se calcularon a partir de pares de aciertos y falsas alarmas correspondientes a valores  $d'$  que variaban desde 1 a 4.5 en incrementos de 0.10 unidades. Un valor  $\beta = 1$  corresponde a un criterio de decisión simétrico y un valor  $\beta = 11$  a un criterio estricto y es un valor próximo a los que usualmente se observan en experimentos de vigilancia. La relación entre  $A'$  y  $A''$  es claramente lineal para ambos niveles del criterio de decisión, con un intercepto y una pendiente que varían ligeramente con el criterio. Para el criterio de decisión simétrico, el intercepto y la pendiente eran  $-0.218$  y  $1.211$  ( $r^2 = 0.997$ ), mientras que para el criterio estricto eran  $0$  y  $1$  ( $r^2 = 0.999$ ), respectivamente. Como puede verse en la Figura 1, la función que relaciona  $d'$  y  $A'$  es asintótica conforme se alcanza el límite superior de  $A'$ , debido al hecho de que  $A'$  tiene un límite superior (1) y  $d'$  no (en teoría,

infinito). Además, y esto puede ser importante en investigaciones sobre vigilancia, la diferencia entre la función que relaciona  $d'$  y  $A'$  cuando el criterio es flexible ( $\beta = 1$ ) y la que relaciona ambos valores cuando el criterio es estricto ( $\beta = 1/1$ ), se hace mayor para valores medios de  $A'$ , que son precisamente los que se observan en experimentos de vigilancia. Este hecho puede ser importante en la investigación sobre vigilancia y puede ser otro argumento para justificar la necesidad de controlar, mediante ensayos de práctica, que el observador no modifique su criterio de decisión durante la tarea.

## MÉTODO

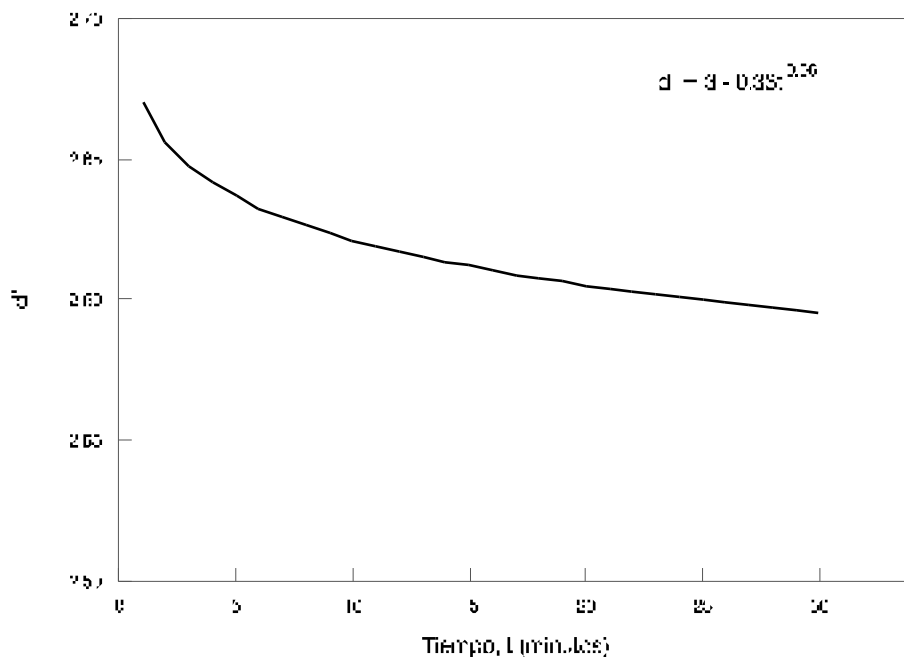
Hicimos un estudio de simulación por ordenador para estimar el efecto del índice de sensibilidad utilizado ( $d'$  o  $A'$ ) sobre el decremento en una tarea de vigilancia típica. Supusimos una única condición experimental y que el propósito del experimento fuese evaluar la magnitud del decremento de sensibilidad en dicha condición. La tarea consistía en tres bloques de 10 minutos cada uno, la probabilidad de señal era 0.10 y el criterio de decisión óptimo  $\beta = 9$ . Supusimos, además, que el número de participantes fuese 20. Los datos se generaron asumiendo una función potencial para el decremento de sensibilidad<sup>1</sup>. La función era la siguiente:

$$d' = a - bt^s$$

$d'$  es el nivel de sensibilidad en cada período o bloque, y  $a$ ,  $b$  y  $s$  son parámetros. El parámetro  $a$  representa el nivel de sensibilidad al comienzo de la tarea.  $s$  es el parámetro que representa la pendiente o magnitud del decremento a lo largo de la prueba; un valor positivo de  $s$  representa un decremento y un valor negativo un incremento. Finalmente,  $b$  es otro parámetro que refleja también la magnitud del decremento.  $s$  podría ser un parámetro relacionado con características individuales del observador y  $b$  un parámetro relacionado con características de la tarea. En todas las simulaciones que condujimos, supusimos que  $b$  era igual a  $1/a$ . La Figura 2 representa esta función para  $a=3$  y  $s= 0.06$ . En general, esta función satisface dos características importantes de las funciones decremento empíricas. En primer lugar, presenta una pendiente decreciente. Se sabe bien que si, por ejemplo, la tarea comprende tres períodos continuos de 10 minutos cada uno, al menos la mitad del decremento total ocurre durante el primer período (See y colaboradores, 1995). En segundo lugar, el parámetro  $b$  ( $b= 1/a$ ) se incluyó en la ecuación para hacer que el decremento fuese

<sup>1</sup> Examinamos también otras funciones matemáticas (p.e. exponencial, logarítmica), pero los resultados fueron similares a los obtenidos con esta función potencial.

inversamente proporcional al nivel inicial de sensibilidad. Esta relación entre el decremento de sensibilidad y el nivel de sensibilidad al comienzo de la tarea es una predicción de la hipótesis propuesta por Parasuraman (1979, 1985), que sugiere que el decremento de sensibilidad está relacionado con las demandas de procesamiento de la información impuestas por la tarea.



**Figura 2.** Representación gráfica de la función decremento potencial examinada en el presente estudio. Los parámetros  $a$  y  $s$  fueron 3.00 y 0.06, respectivamente. La función era:  $d' = 3.00 - 0.33t^{0.06}$

La hipótesis se basa en los tres supuestos típicos de los modelos de capacidad atencional (Kahneman, 1973): 1) existe un conjunto limitado de recursos de procesamiento que pueden ser destinados a una tarea de procesamiento; 2) el nivel de discriminación es una función de la cantidad de recursos destinados a la tarea; 3) cuando las demandas de la tarea exceden la capacidad disponible, la eficiencia del rendimiento disminuye y esto será más probable que ocurra conforme aumenta la dificultad de la prueba. Para explicar el decremento de sensibilidad en tareas de vigilancia, Parasuraman



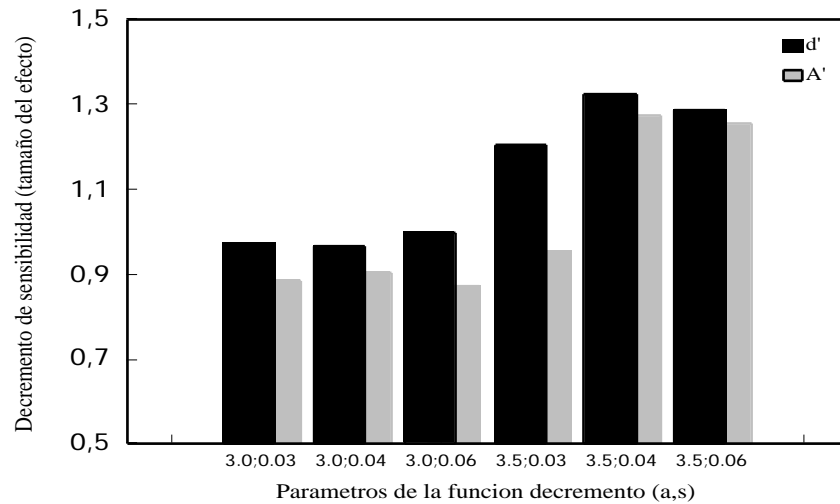
(1979, 1985) añade un supuesto adicional: la cantidad de recursos atencionales decrece progresivamente a lo largo de la tarea. En general, la hipótesis propone que, cuando las demandas de recursos atencionales impuestas al observador por la tarea de vigilancia son bajas, la sensibilidad será alta y el decremento pequeño, mientras que cuando las demandas son altas, la sensibilidad será baja y el decremento grande. La evidencia que apoya esta hipótesis proviene principalmente del hecho de que el decremento de sensibilidad ocurre con mayor probabilidad en tareas que supuestamente demandan una cantidad alta de recursos atencionales. En general, el decremento ocurre con más frecuencia y es mayor cuando los estímulos se presentan a un ritmo rápido en tareas de discriminación sucesiva, que demandan el uso de la memoria de trabajo (Parasuraman y Davies, 1977; Parasuraman, 1979), o cuando la discriminación señal/ruido es simultánea pero la razón señal/ruido es alta (Parasuraman y Mouloua, 1987).

En nuestra simulación, hicimos dos supuestos respecto a la variabilidad en la población de los parámetros de la función decremento. En primer lugar, puesto que  $d'$  puede variar desde 0 hasta infinito (aunque la TDS permite que  $d'$  tome cualquier valor real), asumimos que la distribución de  $a$ , el parámetro que representa la sensibilidad al comienzo de la prueba, era gamma. Este parámetro se varió entre simulaciones: su media se situó en 3 ó 3.50 y su desviación típica se fijó en  $1/5$  de esta media. Estos estadísticos producían valores que están dentro del rango de los observados habitualmente en experimentos de vigilancia. En segundo lugar, asumimos que la distribución de la pendiente de la función,  $s$ , el parámetro que representa la magnitud del decremento de vigilancia, era normal. La media de esta distribución varió también entre simulaciones: 0.03, 0.04 ó 0.06. En cada simulación, la desviación típica de esta distribución fue igual a tres veces la media. Finalmente, hicimos otro supuesto referente a la distribución de  $d'$  en la población. En todas las simulaciones asumimos una distribución gamma del criterio con media igual a 9 y desviación típica 0.9. Todas las variables aleatorias se generaron utilizando algoritmos adaptados de Press y colaboradores (1986)<sup>2</sup>. Cada experimento simulado se definió por una combinación de los dos parámetros manipulados ( $a$  y  $s$ ) y se repitió 1000 veces. Con los datos de cada experimento hicimos un ANOVA, con Tiempo en la tarea como único factor (10, 20 y 30 minutos). El nivel de significación se situó en 0.01. Con los datos de cada experimento estimamos el tamaño del efecto, empleando la misma fórmula que See y colaboradores (1995).

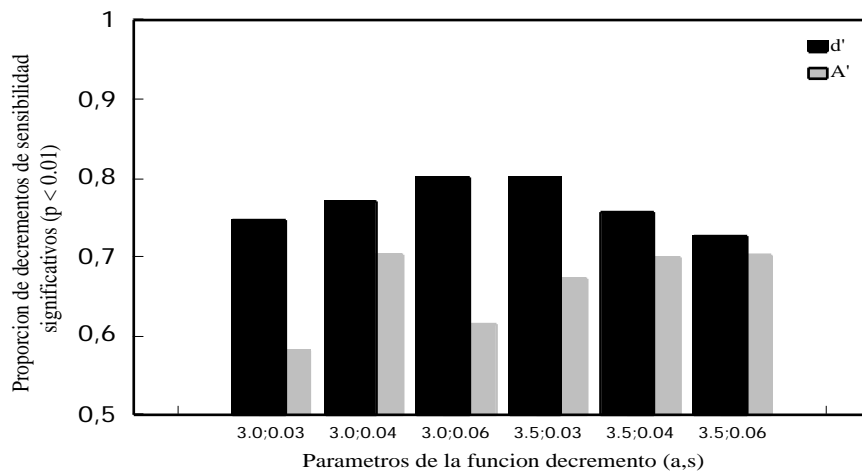
---

<sup>2</sup> El lector puede consultar el programa empleado para computar los índices TDS y el tamaño de efecto solicitando una copia a los autores

## RESULTADOS



**Figura 3.** Decremento de sensibilidad medio medido por el tamaño del efecto para grupos diferentes de simulaciones de experimentos de vigilancia. El tamaño de efecto se definió como la diferencia entre los niveles de sensibilidad en el primer y último período de la prueba dividida por la raíz cuadrada del error cuadrático medio para el factor Tiempo observado en el ANOVA de los datos experimentales simulados.



**Figura 4.** Proporción de decrementos de sensibilidad significativos ( $p < .01$ ) observados en los ANOVAs de experimentos de vigilancia simulados.

La Figura 3 representa el tamaño de efecto medio para las simulaciones hechas con cada par de valores  $a, s$ . Como puede verse en esta figura, todas los grupos de simulaciones dieron resultados similares. Los análisis del rendimiento basados en  $d'$  dieron tamaños de efecto mayores que los análisis de  $A'$ . El tamaño de efecto medio para  $d'$  fue aproximadamente un 8.05% mayor que el tamaño de efecto medio para  $A'$ . La Figura 4 representa la proporción de decrementos de sensibilidad significativos para cada grupo de simulaciones ( $p < 0.01$ ). Como muestra esta figura, el número de decrementos significativos fue mayor cuando la sensibilidad se medía con  $d'$  que cuando se medía con  $A'$ .

## CONCLUSIÓN

Nuestro estudio sugiere que la probabilidad de obtener un resultado significativo en un experimento de vigilancia es mayor cuando el investigador toma  $d'$  como medida del rendimiento que cuando toma  $A'$ . Este efecto del índice de sensibilidad sobre la significación de los resultados en un análisis de varianza de los datos de vigilancia puede plantear un serio problema cuando se trata de comparar y contrastar conclusiones derivadas de estudios que han utilizado diferentes índices. Esto ocurre con frecuencia cuando el investigador compara estudios recientes con otros realizados en los años sesenta y setenta, ya que en esa época los investigadores optaban mayoritariamente por  $d'$  mientras que ahora se trabaja sobre todo con  $A'$ .

El problema que puede plantear el índice de sensibilidad sobre la significación de resultados estadísticos en experimentos de vigilancia podría darse también en otras áreas de investigación en las que se han aplicado los métodos derivados de la TDS, pero este es un asunto que habría que investigar. En principio, a la luz de los resultados de nuestro estudio de simulación, podríamos pensar que hubiera problemas similares en la investigación sobre memoria de reconocimiento, evaluación de sistemas de diagnóstico médico, evaluación de sistemas de predicción meteorológica, control de calidad, etc (Swets, 1986), ya que también en estas áreas se han utilizado  $d'$  y  $A'$  en diferentes estudios y con distinta frecuencia en etapas distintas.

Otra cuestión muy diferente a la tratada aquí es la que se refiere al índice de sensibilidad más recomendable. Está claro que si los supuestos de normalidad y homogeneidad de varianzas se cumplen, el índice más apropiado es  $d'$ ; en cualquier otro caso, el investigador debería optar por  $A'$ . Sin embargo, conviene recordar dos cosas respecto a la confirmación de estos

supuestos en experimentos de vigilancia. En primer lugar, se sabe que en muchos experimentos de vigilancia este supuesto no siempre se cumple. Craig (1977) revisó los análisis ROC de varios experimentos de vigilancia y encontró que una parte importante (30%) de las curvas ROC no satisfacían los supuestos. En segundo lugar, en muchos experimentos de vigilancia no es posible conducir un análisis ROC previo a la elección del índice de sensibilidad. Como se sabe, la curva ROC se construye registrando el rendimiento del observador, bajo una determinada condición experimental, mientras él/ella emplea diferentes criterios de decisión. El criterio del observador puede manipularse por medio de instrucciones o una matriz de costes/beneficios. Sin embargo, ambos procedimientos son difíciles de aplicar en experimentos de vigilancia. Puesto que en un estudio ROC el criterio se manipula entre bloques de ensayos, el experimento de vigilancia requeriría repetir  $n$  veces la tarea de vigilancia para cada observador para así obtener  $n-1$  puntos ROC. Un procedimiento que reduce significativamente el tiempo necesario para obtener datos ROC para cada observador implica el uso de escalas de confianza. Este procedimiento ha sido el usual para generar curvas ROC en experimentos de vigilancia (Broadbent, 1971; Swets, 1977). En este procedimiento, el observador debe utilizar  $n$  respuestas diferentes para indicar su confianza en que el estímulo señal ha sido presentado. Se asume que el observador adopta, dentro de un bloque de ensayos,  $n-1$  criterios para separar cada par de respuestas adyacentes. Usualmente se emplean 4 ó 5 categorías de respuesta. Sin embargo, este rango de categorías crea un problema estadístico serio cuando se desea probar el ajuste de la curva ROC empírica a la predicha por algún modelo teórico. El problema es que el número de categorías de respuesta que el observador puede utilizar de modo eficiente y, consecuentemente, el número de puntos ROC con el que la curva puede construirse (número de categorías menos 1), es siempre muy bajo. Otro procedimiento similar que podría mitigar en alguna medida este problema es el empleado por Watson, Rilling y Bourbon (1964). En este, el observador expresa su confianza utilizando una escala continua. Sin embargo, este procedimiento, que es ampliamente utilizado hoy con muy buenos resultados en estudios ROC de técnicas médicas de diagnóstico (p.e. Rockette, Gur y Metz, 1992; King, Britton, Gur, Rockette y Davis, 1993), no parece apropiado para experimentos de vigilancia, ya que requiere un tiempo considerable, que afecta negativamente al decremento de sensibilidad.

En definitiva, el índice TDS más apropiado para investigar el decremento de sensibilidad en tareas de vigilancia es  $A'$  porque: 1) está libre de supuestos acerca de las distribuciones  $S$  y  $R$  y 2) no requiere realizar un

análisis ROC previo, un procedimiento difícil de aplicar en experimentos de vigilancia.

## ABSTRACT

**Vigilance and Signal Detection Theory.** This article assesses the extent to which the sensitivity decrement over time frequently observed in vigilance tasks can be affected by the index or statistic used to measure the sensitivity. The paper reports the results of a simulation of sensitivity decrement in vigilance tasks using two measures of sensitivity,  $d'$  and  $A'$ . The results of these simulations indicate that the sensitivity decrement is somewhat larger for  $d'$  than for  $A'$ , and that probability to obtain a significant decrement in an analysis of variance of the data is also higher for  $d'$  than for  $A'$ .

**Key words:** Sustained attention, Signal detection theory, Vigilance.

## REFERENCES

- Beasley, J. D. y Springer, S. G. (1977). Algorithm AS 111: The percentage points of the normal distribution. *Applied Statistics*, 26, 118-121.
- Broadbent, D. E. *Decision and stress*. London: Academic Press.
- Broadbent, D. E. y Gregory, M. (1963). Vigilance considered as a statistical decision. *British Journal of Psychology*, 54, 309-323.
- Brophy, A. L. (1985). Approximation of the inverse normal distribution function. *Behavior Research Methods, Instruments, & Computers*, 17, 415-417.
- Craig, A. (1977). Broadbent and Gregory revisited: Vigilance and statistical decision. *Human Factors*, 19, 25-36.
- Davies, D. R. y Tune, G. S. (1969). *Human Vigilance Performance*. New York: American Elsevier.
- Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*. New York: Academic Press.
- Green, D. M., y Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley & Sons.
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424-429.
- Hedges, L. V. y Olkin, I. (1985). *Statistical Methods for Meta-analysis*. San Diego, CA: Academic Press.
- Hunter, J. E. y Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage.
- Kahneman, D. (1973). *Attention and effort*. New York: Prentice-Hall.
- King, J. L., Britton, C. A., Gur, D., Rockette, H. E. y Davis, P. (1993). On the validity of the continuous and discrete confidence rating scales in receiver operating characteristic studies. *Statistics in Radiology*, 28, 962-963.
- Koelega, H. S., Brinkman, J., Hendriks, L. y Verbaten, M. N. (1989). Processing demands, effort, and individual differences in four different vigilance tasks. *Human Factors*, 31, 45-62.

- Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1, 6-21.
- Norman, D. A. (1964). A comparison of data obtained with different false-alarm rates. *Psychological Review*, 71, 243-246.
- Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205, 924-927.
- Parasuraman, R. y Davies, D. R. (1977). A taxonomic analysis of vigilance performance. En R. R. Mackie (Ed.), *Vigilance: Theory, Operational Performance, and Physiological Correlates*. New York: Plenum, 559-574.
- Parasuraman, R. y Mouloua, M. (1987). Interaction of signal discriminability and task type in vigilance decrement. *Perception and Psychophysics*, 41, 17-22.
- Pollack, I. y Norman, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science*, 1, 125-126.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. y Vetterling, W. T. (1986). *Numerical recipes*. Cambridge, England: Cambridge University Press.
- Rockette, H. E., Gur, D. y Metz, Ch. E. (1992). The use of continuous and discrete confidence judgments in receiver operating characteristic studies of diagnostic imaging techniques. *Statistics in Radiology*, 27, 169-172.
- See, J. E., Howe, S. R., Warm, J. S. y Dember, W. N. (1995). Meta-Analysis of sensitivity decrement in vigilance. *Psychological Bulletin*, 117, 230-249.
- Smith, W. D. (1995). Clarification of sensitivity measure A'. *Journal of Mathematical Psychology*, 39, 82-89.
- Swets, J. A. (1977). Signal detection theory applied to vigilance. En R. R. Mackie (Ed.), *Vigilance: Theory, Operational Performance, and Physiological Correlates*. New York: Plenum, 705-718.
- Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks. *Psychological Bulletin*, 99, 181-198.
- Swets, J. A., Tanner, W. P. y Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68, 301-340.
- Tanner, W. P. y Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401-409.
- Watson, Ch. S., Rilling, M. E. y Bourbon, W. T. (1964). Receiver-operating characteristics determined by a mechanical analog to the rating scale. *Journal of the Acoustical Society of America*, 36, 283-288.

(Recibido 5/9/00;; Aceptado: 11/1/01)