TESIS DOCTORAL

Programa de Doctorado en Ciencias Sociales

Doctorado Internacional

Carga mental y complejidad en control de tráfico aéreo: simulación de situaciones operativas en estudios con registro de movimientos oculares

Cognitive workload and complexity in air traffic control: simulation of operational tasks in eye tracking studies

Mauro Marchitto

Director:
Prof. José Juan Cañas Delgado

Department of Experimental Psychology
Mind, Brain, and Behavior Research Center (CIMCYC)

Universidad de Granada
Junio 2015

Editorial: Universidad de Granada. Tesis Doctorales
Autor: Mauro Marchitto
ISBN: 978-84-9125-220-7
URI: http://hdl.handle.net/10481/40696

DOCTORAL CANDIDATE:

Mauro Marchitto

THESIS DIRECTOR:

Prof. José Juan Cañas Delgado (Universidad de Granada)

EXTERNAL REVIEWERS:

Marco Pedrotti, Ph.D. (Université de Neuchâtel)

Yasmina Okan Gil, Ph.D. (University of Leeds)

THESIS EXAMINATION COMMITTEE:

Prof. Andrés Catena Martínez (Universidad de Granada)

Prof. Rocío García-Retamero Imedio (Universidad de Granada)

Prof. Roberto Montanari (Universitá degli studi Suor Orsola Benincasa, Napoli)

Prof. Paulo Noriega (Universidade de Lisboa)

Prof. Javier Roca Ruiz (Universidad de Valencia)

El doctorando Mauro Marchitto y el director de la tesis Prof. José Juan Cañas Delgado Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 10 Junio 2015

Director/es de la Tesis	Doctorando
Fdo.:	Fdo.:

To my family

PUBLICACIONES EN LA TESIS

Di Stasi, L. L., Marchitto, M., Antolí, A., Rodriguez, E., & Cañas, J. J. (2010). From subjective questionnaires to saccadic peak velocity: A neuroergonomics index for online assessment of mental workload. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in understanding human performance: Neuroergonomics, human factors design, and special populations* (pp. 11–20). Boca Raton, FL: CRC Press.

Marchitto, M., Di Stasi, L. L., & Cañas, J. J. (2012). Ocular movements under taskload manipulations: Influence of geometry on saccades in air traffic control simulated tasks. *Human Factors and Ergonomics in Manufacturing & Service Industries, 22*(5), 407-419. doi: 10.1002/hfm.20320

Marchitto, M., Benedetto, S., Baccino, T., & Cañas, J. J. (manuscript submitted for publication). Air traffic control: ocular metrics reflect cognitive complexity. *International Journal of Industrial Ergonomics.*

Marchitto, M., Benedetto, S., Cañas, J. J., & Baccino, T. (manuscript submitted for publication). Workload and cognitive strategy in conflict detection: analysis of gaze transitions and ocular metrics. *Journal of Cognitive Engineering and Decision Making*.

La presente tesis doctoral se ha desarrollado dentro de los proyectos de investigación SEJ 2007-63850/PSIC financiado por el Ministerio de Educación y Ciencia (actual Ministerio de Educación, Cultura y Deporte), y PSI2012-39246 financiado por el Ministerio de Economía y Competitividad. La estancia de investigación en el LUTIN Lab (Université Paris VIII) ha sido financiada a través del programa de movilidad internacional de estudiantes de Doctorado (Vicerrectorado de Relaciones Internacionales y Cooperación al Desarrollo, Vicerrectorado de Política Científica e Investigación, y Escuela Internacional de Posgrado de la Universidad de Granada; CEI BioTic Granada).

CONTENTS

PREFACIO

(EN ESPAÑOL)

PREFACIO

El sistema de transporte aeronáutico ha conocido un crecimiento impresionante en la segunda mitad del siglo pasado, y la cantidad global de tráfico aéreo se ha doblado cada 15 años desde 1977. Esta tendencia es la misma al día de hoy, y continuará siendo en el futuro, como ha aclarado la Organización de Aviación Civil Internacional en sus previsiones de tráfico (ICAO, 2013). El tráfico de pasajeros ha crecido a una tasa del 6% anual aproximadamente en el periodo 2013-2015.

Como se ha señalado en la literatura sobre control de tráfico aéreo (ATC, del inglés *air traffic control*) "la carga mental del controlador será probablemente la única y mayor limitación a la capacidad del sistema de gestión del tráfico" (traducido desde Majumdar & Polak, 2001). Consecuentemente, la necesidad de medir de forma fiable la carga de trabajo experimentada por los controladores es un tema central y de actualidad en la gestión de la seguridad dentro del sistema de ATC.

Esta necesidad surgió a la vez que otro concepto fundamental, tanto para el diseño como para la gestión de las operaciones de ATC: el concepto de complejidad. El dominio del ATC ha dedicado esfuerzos considerables desde su nacimiento a definir y operacionalizar la complejidad (Arad, 1964). El interés hacia índices medibles de complejidad en el dominio del ATC fue estimulado a través de los años por el supuesto comúnmente aceptado de que la complejidad percibida de la situación de tráfico es un determinante de la carga de trabajo percibida.

Históricamente, los primeros factores de complejidad identificados estaban relacionados con la geometría estática del espacio aéreo (Arad, 1964), y con la cantidad de tráfico (número de aviones gestionados, Hurst and Rose, 1978; Stein, 1985) o su densidad. Sin embargo, surgió pronto la necesidad de incluir la dinámica de tráfico entre los determinantes de la complejidad, es decir, la evolución de los vuelos a lo largo del tiempo (por ejemplo los cambios de altitud, velocidad, y tasa de ascenso), junto a las características estructurales estáticas.

Estaba claro que diferentes fuentes de complejidad afectaban a la complejidad percibida a nivel individual, indicada como complejidad cognitiva. La complejidad cognitiva se consideraba como el resultado del impacto de las características del tráfico sobre un sistema cognitivo (el controlador), que debe aplicar recursos de procesamiento y estrategias cognitivas para cumplir la tarea de forma segura y con el mínimo coste.

El enfoque sobre los procesos cognitivos de los controladores evidenció progresivamente las principales debilidades de las aproximaciones anteriores al concepto de complejidad, como por ejemplo: una dependencia excesiva de la conducta observable para inferir los niveles de carga de trabajo experimentada en lugar de intentar medir los antecedentes perceptuales y cognitivos de la ejecución de una acción (detectar, evaluar, planificar, y monitorizar); la frecuente falta de consideración de los mecanismos cognitivos de autorregulación del controlador, como por ejemplo la adopción y el ajuste de estrategias compensatorias en respuesta a la carga de trabajo y al estrés (Hockey, 1997; Crévits, Debernard, & Denecker, 2002); la excesiva simplificación al asumir una transformación lineal de las demandas de la tarea en carga de trabajo percibida totalmente describible en términos de factores de complejidad geométrica.

Emergió progresivamente un marco teórico para la complejidad del ATC más centrado en el controlador, en el cual se asumía fundamentalmente que la actividad de los controladores es esencialmente cognitiva, es decir, de procesamiento de la información en presencia de recursos de procesamiento limitados en tiempo y esfuerzo (Wickens, Mavor, & McGee, 1997). Considerando la naturaleza cognitiva de las tareas de ATC y las múltiples estrategias disponibles para ordenar el tráfico de una forma segura, las acciones observables ejecutadas por parte de los controladores podrían expresar solo una información parcial respecto de la complejidad percibida, ya que ésta estaría en gran medida determinada por los procesos cognitivos como monitorizar, evaluar, y planificar (Pawlak et al., 1996). La complejidad cognitiva es, entonces, el resultado del efecto de mediación de los procesos cognitivos sobre las demandas de las tareas, y determina la carga de trabajo percibida.

Resumiendo, históricamente ha habido dos aproximaciones a la definición de la complejidad del ATC. La aproximación “abajo-arriba” (del inglés bottom-up) fue típica en los estudios de ingeniería y diseño, y consistió en identificar factores de complejidad geométrica y de tráfico como determinantes de la carga de trabajo (sobre todo por medio de entrevistas y técnicas de observación), intentando combinarlos en una ecuación de complejidad y compilando listas de factores de complejidad. La aproximación "arriba-abajo" (top-down) fue consecuencia de la aplicación de la psicología cognitiva a los estudios ergonómicos, y operó con un modelo opuesto. Básicamente, en estos estudios se usó un criterio para determinar la carga de trabajo de los controladores basado en su actividad cognitiva (como por ejemplo, juicio perceptual, toma de decisiones o asignación de recursos atencionales a lo largo del tiempo) para identificar qué factores de complejidad tendrían más impacto en dicha respuesta. Como se indica en la literatura (Hilburn, 2004), la opción más prometedora para la modelación de la complejidad en el ATC es una combinación de las dos aproximaciones, siendo las dos útiles para identificar factores de complejidad, manipularlos para evaluar la respuesta cognitiva asociada, e identificar las estrategias cognitivas adoptadas para gestionar la complejidad percibida.

Los estudios, entre otros, de Chatterji et al. (2001), Averty et al. (2002), Histon et al. (2002), y Boag et al. (2006) consideraron procesos cognitivos cómo las proyecciones de movimiento/trayectoria, o la presencia de conflictos como los verdaderos factores responsables de la carga de trabajo, la cual es esencialmente mental.

Casi todas las definiciones de carga de trabajo mental expresan el concepto como un coste experimentado para la aplicación de un esfuerzo mental para ejecutar una tarea de naturaleza más mental que física (Kahneman, 1973; Wickens; 2002; Parasuraman & Caggiano, 2002; Cañas, 2004). Aunque la relación funcional entre complejidad y carga mental no se ha comprendido todavía del todo (Athènes, 2002), la carga de trabajo experimentada se entiende normalmente en la literatura del ATC como una consecuencia de la complejidad cognitiva percibida (Mogford et al., 1995; Loft, Sanderson, Neal, & Mooij, 2007; Djokic, Lorenz, & Fricke, 2010). Por ejemplo, se ha demostrado como la carga de trabajo mental percibida está afectada por la presencia (predicha) de trayectorias conflictivas (Neal & Kwantes, 2009; Averty, 2005).

La detección de conflicto es un componente clave (Kallus, Van Damme, & Dittman, 1999) y una de las tareas más estudiada en el ATC (Remington, Johnston, Ruthruff, Gold, & Romera, 2000; Galster, Duley, Masalonis, & Parasuraman, 2001; Loft, Humphreys, & Neal, 2004; Eyferth, Niessen, & Spaeth, 2003). Se puede llevar a cabo por medio de diferentes estrategias (Rantanen & Nunes, 2005; Xu &

Rantanen, 2003; Stankovic, Raufaste, & Averty, 2008; Loft, Bolland, Humphreys, & Neal, 2009), las cuales generan diferentes niveles de carga mental de trabajo. Básicamente, consiste en la predicción de una pérdida simultánea de separación entre dos aviones tanto en la dimensión lateral como vertical.

La carga de trabajo mental está relacionada con la asignación de recursos mentales para cumplir con las demandas de la tarea, y no es directamente observable. No obstante, la carga de trabajo mental se puede inferir desde varias fuentes, es decir, es un constructo multidimensional. Los tres principales grupos de métricas de carga de trabajo mental son: las estimaciones subjetivas, los índices de desempeño, y las respuestas psicofisiológicas.

En el ATC, la conducta manifestada ha sido reconocida como imprecisa a la hora de evaluar la carga de trabajo percibida. Las medidas psicofisiológicas ofrecen la posibilidad de registrar respuestas objetivas en el desempeño de una tarea. Los métodos de registro de movimientos oculares permiten el registro de la conducta visual asociada al procesamiento cognitivo, interfiriendo mínimamente con la tarea en ejecución. El control del tráfico aéreo tiene un componente visual fundamental. Los movimientos oculares y las fijaciones representan los procesos primarios para la obtención de información desde los interfaces visuales. El estudio de la conducta visual de los controladores por medio del registro de métricas oculares podría beneficiar en un futuro próximo la modelación de carga de trabajo mental en el ATC. De hecho, el tiempo invertido por los controladores para procesar elementos específicos en la pantalla del radar, podría asumirse como indicativo de los procesos atencionales del operador y la representación mental del tráfico. Además, las estrategias de escaneo de los controladores pueden informar acerca de la asignación de recursos perceptuales y cognitivos (Kang, Bass, & Lee, 2014).

Las aplicaciones del registro de movimientos oculares en el dominio del ATC son relativamente recientes. La tasa y duración de los parpadeos han sido las primeras métricas oculares usadas en el ATC para la inferencia de la carga de trabajo mental experimentada (Brookings, Wilson, & Swain, 1996; Wilson & Russell, 2003). Normalmente, los resultados han mostrado una relación negativa entre las métricas de parpadeo y la carga mental. En paralelo, las métricas de fijación y mirada (donde el tiempo de mirada total para una cierta área de interés es dado por la suma de la duración de las fijaciones dentro de esa misma área), han sido también indicadas como positivamente relacionadas con la carga de trabajo mental en tareas de ATC (Ahlstrom & Friedman-Berg, 2006).

Por otro lado, el estudio de los patrones de miradas y el análisis de las métricas oculares durante la detección de conflicto en el ATC son relativamente recientes, pero prometedores (Hunter & Parush, 2009; Martin, Cegarra, & Averty, 2011; Kang & Landry, 2010; 2014).

Como ya se ha anticipado, la detección de conflicto en el ATC es fundamental dentro de las operaciones normales de control de tráfico. Muchas investigaciones han tratado el estudio de la detección de conflicto, con el objetivo de modelar los factores de complejidad relacionados con la misma, así como también las estrategias, y la carga relacionada. Ulteriores evidencias empíricas mediante el registro de movimientos oculares acerca de la detección de conflicto pueden contribuir a la modelación de estos procesos de toma de decisión tan importantes en el dominio del ATC.

RESUMEN DE LA TESIS

En esta Tesis Doctoral se presentan cuatro estudios experimentales en los cuales se midieron diferentes métricas oculares como índices de carga de trabajo mental en tareas simuladas de ATC. Más precisamente, las tareas experimentales estaban relacionadas con la detección de conflicto, es decir con la previsión de pérdida de separación mínima (lateral y vertical) entre dos aviones con trayectorias convergentes. El registro de los movimientos oculares para evaluar la carga de trabajo mental en tareas de detección de conflicto en el dominio del ATC es relativamente reciente. En todos los experimentos presentados, se variaron las demandas de la tarea manipulando unos reconocidos factores de complejidad, con el objetivo de generar diferentes niveles de complejidad percibida. Se asumió que esta manipulación de la complexidad iba a contribuir a la complejidad cognitiva percibida de la situación del tráfico, y consecuentemente, a la carga mental experimentada. La carga mental fue evaluada de forma multidimensional, ya que se midieron también el tiempo de respuesta y unas estimaciones subjetivas.

En el primer estudio (**Capítulo 2)** se usó una simple tarea de decisión perceptual sobre la posición de un objetivo como condición experimental fácil, mientras se añadieron unas tareas adicionales (una tarea de papel y lápiz, y una simple operación matemática) para crear una condición difícil. La carga de trabajo mental y la fatiga fueron medidas en función de la complejidad de la tarea y de la duración de la misma por medio de dos parámetros relacionados con los movimientos sacádicos (es decir velocidad máxima y duración, en función de la amplitud), del tiempo de respuesta, y de medidas subjetivas. Como se esperaba, la complejidad de la tarea tuvo su efecto tanto a nivel subjetivo como de ejecución. Por otro lado y de forma inesperada, hubo una mejora en la ejecución de la tarea (tiempos de respuestas más cortos) a través del tiempo, y solamente en la condición de mayor dificultad. Este resultado fue interpretado en términos de estrategia adoptada para el cómputo perceptual y, alternativamente, en términos de carga mental óptima y subcarga. La condición de mayor complejidad requería leer y memorizar el identificativo del avión (un numero de 3 dígitos), además de la tarea de decisión perceptual. Por otro lado, en la condición de menor complejidad, solo la posición del avión era relevante de cara al objetivo de la tarea, sin necesidad de procesar otra información. En esta última condición las demandas de la tarea fueron mínimas, dejando probablemente pocas posibilidades para acortar aún más los tiempos de respuestas a lo largo del tiempo. Diferentemente, el efecto de aprendizaje y mejor ejecución en la condición de mayor dificultad pudo deberse a un mejor procesamiento de las informaciones espacial y numérica. Una interpretación complementaria a la anterior fue apoyada por el análisis de los parámetros sacádicos. La velocidad máxima de las sácadas amplias (>11°) mostró un decremento significativo en la condición de menor dificultad, sobre todo en la primera parte de las sesiones experimentales. Este resultado sugirió que probablemente la condición de menor complejidad fue excesivamente simple, mientras que la de mayor complejidad generó en realidad un nivel de carga mental optimo, permitiendo un efecto de aprendizaje debido a una gestión más activa de los recursos de procesamiento. Las situaciones de subcarga están caracterizadas por una gran disponibilidad de tiempo para ejecutar una cantidad mínima de tareas, y de baja complejidad. La velocidad máxima de los movimientos sacádicos amplios mostró entonces una sensibilidad a las situaciones de subcarga, igual que en las situaciones de sobrecarga mental (Di Stasi, Marchitto, Antolí, Baccino, &

Cañas, 2010), es decir, cuando las demandas impuestas por la tarea y los recursos disponibles no se presenten en proporción óptima.

En el **Capítulo 3** se manipularon dos factores de complejidad: el ángulo de convergencia y la distancia con el centro de convergencia. La tarea experimental consistió en un juicio relativo, más precisamente, de la estimación de la llegada del primer avión al punto de convergencia. Los escenarios de tráfico, aunque estáticos (imágenes), fueron construidos con un simulador de tráfico aéreo. La velocidad máxima de las sácadas largas (>15º) mostró un decremento con ángulos de convergencias más amplios y con distancias diferentes (con respecto a igual) desde el centro de convergencia. La duración de las sácadas mostró el mismo patrón. Consecuentemente, es probable que los movimientos sacádicos largos fueran más precisos en presencia de una geometría más compleja. Las sácadas largas son las más demandantes para el sistema nervioso oculomotor, y podrían reducir su poder de disparo y su error posicional con geometrías más complejas, optimizando la conducta visual. La velocidad máxima de los movimientos sacádicos confirmó su sensibilidad a las variaciones de la carga mental, y un balance entre velocidad y precisión en su ejecución.

En el estudio presentado en el **Capítulo 4** se usaron escenarios dinámicos para una tarea de detección de conflicto. El ángulo de convergencia fue manipulado como factor de complejidad geométrica y la distancia mínima en el momento de máximo acercamiento como factor de complejidad de tráfico. Se construyeron situaciones de conflicto y de no conflicto, dependiendo de si la distancia mínima hubiese sido menor o mayor, respectivamente, del estándar de separación mínima (5 millas náuticas). En comparación con las tareas experimentales anteriores, en este estudio, el juicio relativo era solo una parte preliminar de la tarea de detección de conflicto. Una detección correcta fue predicha como más difícil (peor ejecución y mayor carga mental subjetiva) con una distancia mínima de acercamiento entre aviones muy parecida al estándar de separación mínima (por ejemplo, 6 millas náuticas). Efectivamente, la tendencia al error fue mayor en estas condiciones. No obstante, las situaciones de conflicto se mostraron como las más demandantes, como mostraron los resultados convergentes en las medidas de carga mental. Un esfuerzo de procesamiento adicional fue encontrado en el caso de rutas de convergencia perpendiculares (en posición vertical y horizontal). Los efectos del tiempo de respuesta (toma de decisión) fueron controlados en el análisis de las métricas oculares (análisis de covarianza). Las sácadas largas confirmaron la sensibilidad a la carga mental de la velocidad máxima y de la duración, mostrando valores menores para los conflictos, especialmente con rutas perpendiculares. Los movimientos oculares amplios se adaptaron al aumento de la carga de trabajo perceptual y cognitiva. En relación a las fijaciones, no se encontraron efectos de los factores de complejidad una vez controlada la influencia del tiempo de respuesta, así que frecuencia y duración de las fijaciones aumentaron con el tiempo de respuesta, y podrían ser predichas por éste. En entornos operacionales, donde el tiempo de respuesta es de difícil, si no imposible, computación, las métricas de fijación son entonces informativas del esfuerzo de procesamiento. Los datos sugirieron que las diferencias considerables que emergieron entre los conflictos y los no conflictos podrían ser cualitativas, más que exclusivamente cuantitativas, suponiendo la adopción de dos diferentes estrategias de detección de conflicto en los dos tipos de escenario de tráfico. Una estrategia basada en proporciones (por ejemplo entre distancia con el centro de convergencia y velocidad, para estimar el tiempo de llegada al punto de convergencia) fue probablemente adoptada en los no conflictos, donde una decisión correcta fue tomada tras realizar pocas fijaciones y movimientos sacádicos, y con valores bajos de carga mental subjetiva. Probablemente, los tiempos de llegada al punto de convergencia para los dos aviones fueron estimados como suficientemente diferentes, sugiriendo un mantenimiento de la separación lateral. Por otro lado, las situaciones de conflicto fueron más demandantes, y una estrategia basada en una proporción pudo resultar inviable. Los participantes, probablemente, tuvieron que basarse en el procesamiento recurrente de la posición actualizada de los aviones, para poder comparar el desplazamiento de los mismos. Una estrategia basada en el movimiento fue muy probablemente preferida en estas situaciones. Las diferencias en velocidad y distancia desde el punto de convergencia a principio de cada escenario fueron propuestas como factores determinantes para la adopción de una u otra estrategia. Por ejemplo, diferencias pequeñas de velocidad combinadas con diferencias considerables de distancia desde el centro de convergencia probablemente sugirieron una separación lateral segura durante el tiempo. Por otro lado, diferencias más pequeñas habrían podido añadir incertidumbre a la estimación de las futuras posiciones relativas, y por ende de la separación.

En el estudio presentado en el **Capítulo 5** se investigaron los valores diferenciales de los parámetros de vuelo como factores de complejidad. Se asumió que varias combinaciones de diferencias de velocidad, altitud, y distancia con el centro de convergencia habrían determinado la adopción de una u otra estrategia para la detección de conflicto, y consecuentemente afectado a la carga mental percibida. Como índice ocular de carga mental se midió esta vez la variación del tamaño de la pupila. En paralelo, se propuso un análisis cuantitativo de las transiciones de mirada entre áreas de interés relevantes para la tarea, con el objetivo de poder identificar la estrategia adoptada para la resolución de conflicto desde el análisis del patrón de mirada. Un aumento de frecuencia de una transición específica dentro del patrón de mirada total fue asumido como indicativo del papel jugado por dicha transición de mirada en la estrategia de detección de conflicto. Al contrario, un decremento significativo de frecuencia fue asumido como indicativo de escasa relevancia dentro de la estrategia adoptada. Varios resultados fueron encontrados tras el análisis de las transiciones de mirada. Por ejemplo, las transiciones entre las etiquetas de los aviones (en las cuales están contenidos los datos de altitud y velocidad) fueron prácticamente ausentes cuando la estimación de la separación lateral fue más demandante. Es más, se observó un aumento de la frecuencia de transiciones entre aviones y punto de convergencia en las situaciones en las que se requirió un procesamiento más preciso de la distancia (es decir, con las diferencias de distancias más pequeñas). Por último, la frecuencia de transiciones cruzadas entre elementos diferentes de los aviones (posición y etiqueta) aumentó en los escenarios más demandantes, en los cuales se tuvo que estimar la separación mínima tanto en la dimensión lateral como en la vertical.

El tamaño de la pupila se mostró afectado por la complejidad, ya que se observaron dilataciones significativas en las condiciones indicadas como demandantes, y con tiempo de respuesta más largos. Sin embargo, las dilataciones más grandes se registraron para los escenarios más fáciles, en los cuales no se cometieron errores, y se invirtieron menos recursos de procesamiento. Estos resultados se interpretaron en términos de reacción emocional positiva relacionada con la baja (o nula) incertidumbre respecto de la ausencia de conflicto.

En síntesis, la carga mental mostró una relación negativa con la diferencia de distancia y de altitud, como se esperaba. Por otro lado, el efecto de la diferencia de velocidad dependió de la combinación de ésta con la diferencia de distancia. Es más se observó una dificultad intrínseca en la proyección de las diferencias de velocidad.

En el **Capítulo 6** se presentan algunas conclusiones, junto a unas sugerencias de aplicación práctica de los estudios presentados. Los análisis de índices oculares y transiciones de mirada propuestos representan una contribución, aunque preliminar, al estudio de la complejidad y de la carga mental en tareas de detección de conflicto. La aplicación de los métodos de registro ocular en el dominio del ATC podría beneficiar en el próximo futuro tanto el diseño de sistemas (por ejemplo de alerta automática de conflicto) como al entrenamiento de nuevos controladores, por ejemplo identificando diferencias en los patrones de mirada en función de la experiencia, o los correlatos visuales de las estrategias de escaneo y asignación de recursos atencionales. Los métodos de registro de movimientos oculares podrían ser incorporados como una importante fuente de información añadida en los futuros programas de entrenamiento.

REFERENCIAS

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623-636.

Arad, B. A. (1964). The controller load and sector design. *Journal of Air Traffic Control, May,* 12-31.

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D. & Collet, C. (2002). Complexity and Controller Workload: Trying to Bridge the Gap. In *Proceedings of the 2002 International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero 2002)*. Cambridge, MA, USA: Massachusetts Institute of Technology.

Averty, P. (2005). Conflict perception by ATCs admits doubt but not inconsistency. In *Proceedings of the 6th Air Traffic Management Research & Development Seminar*.

Averty, P., Athènes, S., Collet, C., & Dittmar, A. (2002). Evaluating a new index of mental workload in real ATC situation using psychophysiological measures. In *Digital Avionics Systems Conference, 2002. Proceedings. The 21st* (Vol. 2, pp. 7A4-1). IEEE.

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, *49*(14), 1508-1526.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology*, *42*(3), 361-377.

Cañas, J. J. (2004). *Personas y máquinas. El diseño de su interacción desde la ergonomía cognitiva.* Madrid: Ediciones Pirámide.

Chatterji, G. B., & Sridhar, B. (2001, October). Measures for air traffic controller workload prediction. In *Proceedings of the First AIAA Aircraft Technology, Integration, and Operations Forum*. Los Angeles, California.

Crévits, I., Debernard, S., & Denecker, P. (2002). Model building for air-traffic controllers' workload regulation. *European Journal of Operational Research*, *136*(2), 324-332.

Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, *16*(6), 330-333.

Djokic, J., Lorenz, B., & Fricke, H. (2010). Air traffic control complexity as workload driver. *Transportation research part C: emerging technologies*, *18*(6), 930-936.

Eyferth, K., Niessen, C., & Spaeth, O. (2003). A model of air traffic controllers' conflict detection and conflict resolution. *Aerospace science and technology*, *7*(6), 409-416.

Galster, S. M., Duley, J. A., Masalonis, A. J., & Parasuraman, R. (2001). Air traffic controller performance and workload under mature free flight: Conflict detection and resolution of aircraft self-separation. *The International Journal of Aviation Psychology*, *11*(1), 71-93.

Hilburn, B. (2004). Cognitive complexity in air traffic control: A literature review. *EEC Note*, *4*(04). Bretigny-sur-Orge: Eurocontrol Experimental Research Center.

Histon, J. M., Hansman, R. J., Gottlieb, B., Kleinwaks, H., Yenson, S., Delahaye, D., & Puechmorel, S. (2002). Structural considerations and cognitive complexity in air traffic control. In *Digital Avionics Systems Conference, 2002. Proceedings. The 21st* (Vol. 1, pp. 1C2-1). IEEE.

Hockey, G. R. (1997). Compensatory control in the regulation of human performance under stress and high workload; a cognitive-energetical framework. *Biological psychology*, *45*(1-3), 73-93.

Hunter, A. C., & Parush, A. (2009, October). Using eye movements to uncover conflict detection strategies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 22, pp. 1729-1733). Sage Publications.

Hurst, M. W., & Rose, R. M. (1978). Objective Job Difficulty, Behavioural Response, and Sector Characteristics in Air Route Traffic Control Centres∗. *Ergonomics*, *21*(9), 697-708.

ICAO, International Civil Aviation Organization, (2013). *Global air navigation plan 2013-2028*. Doc 9750-AN/963, Fourth Edition, 2013.

Kahneman, D. (1973). *Attention and effort* (p. 246). Englewood Cliffs, NJ: Prentice-Hall.

Kallus, K. W., Van Damme, D., & Dittmann, A. (1999). Integrated task and job analysis of air traffic controllers–phase 2: task analysis of en-route controllers. *Deliverable HUM.ET1.ST01.1000-DEL04.* Eurocontrol: Bruxelles.

Kang, Z., Bass, E. J., & Lee, D. W. (2014, September). Air traffic controllers' visual scanning, aircraft selection, and comparison strategies in support of conflict detection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 77-81). SAGE Publications.

Kang, Z., & Landry, S. J. (2010). Capturing and analyzing visual groupings of multiple moving targets in an aircraft conflict detection task using eye movements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 23, pp. 1906-1910). SAGE Publications.

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors*, *56*(6), 1150-1162.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, *15*(2), 106.

Loft, S., Humphreys, M., & Neal, A. (2004). The influence of memory for prior instances on performance in a conflict detection task. *Journal of Experimental Psychology: Applied*, *10*(3), 173-187.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(3), 376-399.

Majumdar, A., & Polak, J. (2001). Estimating the capacity of Europe's airspace using a simulation model of air traffic controller workload. Paper No. 01-3250. In *80th Annual Meeting of the Transportation Research Board.* Washington, DC: Transportation Research Board.

Martin, C., Cegarra, J., & Averty, P. (2011). Analysis of mental workload during en-route air traffic control task execution based on eye-tracking technique. In *Engineering Psychology and Cognitive Ergonomics* (pp. 592-597). Springer Berlin Heidelberg.

Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). Complexity construct in air traffic control: a review and synthesis of the literature (DOT/FAA/-CT TN95/22). Atlantic City, NJ: FAA.

Neal, A., & Kwantes, P. J. (2009). An evidence accumulation model for conflict detection performance in a simulated air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(2), 164-180.

Parasuraman, R., & Caggiano, D. (2002). Mental Workload. In V. S. Ramachandran (Ed.) *Encyclopedia of the human brain (pp. 17-27).* San Diego: Academic Press.

Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996, July). A framework for the evaluation of air traffic control complexity.

In *Proceedings of the AIAA Guidance Navigation and Control Conference, San Diego, CA*.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *The International Journal of Aviation Psychology*, *15*(4), 339-362.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *42*(3), 349-366.

Stankovic, S., Raufaste, É., & Averty, P. (2008). Determinants of conflict detection: A model of risk judgments in air traffic control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(1), 121-134.

Stein, E.S. (1985) *Air traffic controller workload: An examination of workload probe* (Report FAA/CTTN90/60). Atlantic City, New Jersey: FAA.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wickens, C.D., Mavor, A.S. & McGee, J.P. (Eds.) (1997). *Flight to the future: human factors in air traffic control*. Washington, DC: National Academy of Sciences, National Research Council.

Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 381-389.

Xu, X., & Rantanen, E. M. (2003). Conflict detection in air traffic control: A task analysis, a literature review, and a need for further research. In *12th International Symposium on Aviation Psychology, 2003*.

CHAPTER 1:

INTRODUCTION

CHAPTER 1

1. INTRODUCTION

The aviation transportation system has seen an impressive growth in the second half of last century, and the global quantity of air traffic has doubled once every 15 years from 1977. This trend is still ongoing, and will continue in the next future, as made clear by the International Civil Aviation Organization (ICAO) in forecasts for future air traffic (ICAO, 2013). Passenger traffic in civil aviation has grown at a rate of approximately 6% per year in the last two years.

As reported in air traffic control (ATC) scientific literature, "Controller workload is likely to remain the single greatest functional limitation on the capacity of the ATM system" (Majumdar & Polak, 2001). Therefore, the need for reliably measuring controllers' experienced workload is a central aspect of ATC safety management still nowadays, after decades of investigations, for at least two reasons. First, as mentioned, traffic volumes are predicted to increase dramatically in the next future, generating additional workload in controllers' performance. Second, new operational paradigms have been proposed in the last years, which would partially (or totally) transfer aircraft route selection and separation assurance from ground (ATC operators) to air (cockpit), with the aim of optimizing airspace capacity. This paradigm has been named "free flight", as to differentiate it from traditional ground-controlled managed flight, and it is assumed to efficiently increase airspace capacity without threatening air traffic system safety (Hollnagel, 2007). However, the reorganization of control activities under new free flight paradigms would sensibly impact controller performance, as control activity would result more passive compared to the higher authority controllers have in managed flight (Metzger & Parasuraman, 2001; 2005). The 21st century started with a worldwide plan for renewing ATC system safely and efficiently, reducing costs and maximizing sustainability. Free flight programs have been investigated and tested in both the European Union and the US, called Sesar (Single European Sky ATM research) and Nextgen (Next Generation Air Transportation System), respectively (Brooker, 2008).

The need for reliable measurement of workload in ATC emerged in relation to a fundamental concept in the design and management of ATC operations, i.e. air traffic complexity. Complexity measures provide useful descriptions and quantifications of man-machine interactions, and are used for evaluating productivity and cost effectiveness of air traffic control centers (Flynn, Leleu, & Zerrouki, 2003), guiding optimal design and safe management of ATC system.

2. COMPLEXITY IN ATC

Complexity has been notably difficult to define (Hollnagel, 2012). Researchers have operationalized the concept in several ways, in different domains. Under a general perspective, some specific characteristics of complex systems have been reported, like: a large number of interrelated parts, whose connections are not completely observable (i.e. system functioning is opaque), and which interact dynamically in a non-linear fashion (Cilliers, 1998; Johnson, 2007). Consequently, complex systems functioning cannot be fully predicted or even understood, as system elements present partial (or even total) autonomy in functioning, and they often lack of feedback. Sociotechnical systems are complex systems, as they show opacity, autonomy, non-linear interactions of constituent parts, and partial predictability. The aviation transportation in general and the ATC domain in particular, are clear examples of complex sociotechnical systems.

The ATC domain has dedicated considerable efforts since its birth to define and operationalize complexity (Arad, 1964) with the double objective of assessing and predicting complexity on the basis of specific traffic features, thus serving management and design purposes. A considerable number of researches operationalized the concept by means of specific traffic features. The constant interest for measurable indexes of complexity in ATC domain was encouraged throughout the years by the commonly agreed belief that perceived complexity of ATC traffic situation is a determinant of workload.

Traditionally, the engineering approach to complexity contributed to the identification of traffic complexity factors and assessment of determined workload. Complexity factors were mainly related to airspace static geometry (Arad, 1964) and traffic count (number of managed aircraft, Hurst and Rose, 1978; Stein, 1985), or density. However, it soon emerged the need for including traffic dynamics in the definition of complexity. Assuming that complexity was determined by traffic count *and* traffic temporal evolution, complexity modeling had to consider flight progress through time (e.g. changes of speed, altitude, and climb rate), besides static structural features (e.g. sector size or routes shape) (Mogford, Guttman, Morrow, & Kopardekar, 1995; Laudeman, Shelden, Branstrom, & Brasil, 1998; Majumdar & Ochieng, 2002; Hilburn, 2004).

Traffic count and airspace structure were progressively considered insufficient to fully capture complexity experienced by controllers. For instance, a measure of traffic dynamic density was proposed and partially validated as complexity index (Wyndemere, 1996; Laudeman et al., 1998; Kopardekar, 2000; Masalonis, Callaham, Figueroa, & Wanke, 2003). Dynamic density considered the number of aircraft as well as flight progress-related metrics. In this sense, several aircrafts ordered in a standard flow (i.e. flying at the same constant speed on the same air route) would have generated lower complexity than fewer aircrafts flying in a less ordered manner (e.g. with crossing paths, ascending or descending attitudes, different climb rates), or with suspected conflicting trajectories. Consequently, the contributing factors to dynamic density were related to changes in heading, speed, and altitude, as well as estimations of minimum distance between aircraft and conflicts. Importantly, these factors accounted for traffic dynamics and included cognitive processes performed *on* traffic situation (e.g. conflict prediction) as complexity drivers. In fact, due to suspected conflict presence, two aircrafts were assumed to be perceived as more demanding, so that more attentional resources would have been allocated to them. Research on dynamic density represented an important milestone in the modeling of complexity as it first included cognitive activity recurrently performed by controllers (e.g. conflict prediction) among complexity factors that drive workload.

The focus on cognitive processes of ATC controllers progressively evidenced the main weaknesses of early approaches to complexity modeling, i.e.: overreliance on overt performance to infer workload levels instead of trying to measure perceptual and cognitive antecedents of action performance (detecting, evaluating, planning, and monitoring, Pawlak, Brinton, Crouch, & Lancaster, 1996); the frequent overlook to controller cognitive self-regulating mechanisms, like adoption and adjustment of compensatory strategies' as a response to workload and stress (Hockey, 1997; Crévits, Debernard, & Denecker, 2002); the excessive simplification by assuming a linear transformation of taskload into workload fully describable in terms of traffic complexity factors.

It is nowadays commonly accepted that different sources of ATC complexity affect the perceived complexity at the individual level, indicated as cognitive complexity. Cognitive complexity was considered the result of traffic characteristics' impact on a cognitive system (controller), who applies cognitive resources and strategies to accomplish the task safely and with minimal cost. Besides geometry complexity (e.g. sector shapes, number of intersecting airways, convergence angles, or presence of special use airspace), further forms of complexity were described, as traffic complexity (e.g. number of aircraft, dynamic density, flight profiles, presence of to-be-avoided convective weather), display complexity (e.g. visual interface design), and organizational complexity (e.g. special procedures, work conditions). All these different forms of complexity were assumed to contribute to cognitive complexity, intended as the result of internal cognitive processing of traffic situation (Cummings & Tsonis, 2005; 2006).

Besides the analytical approach to task complexity, a more human-centered theoretical framework to ATC complexity emerged in human factors literature, which substantially defined controller's cognitive activity as information processing with limited (in time and effort) cognitive processing resources (Wickens, Mavor, & McGee, 1997). It was implied that cognitive workload and cognitive complexity in ATC could not be fully captured in a formal model and predicted on the basis of objective task features, since it was already known that cognitive strategies can reorganize and restructure traffic information to accommodate taskload variations (Sperandio, 1971; 1980).

Considering the cognitive nature of ATC tasks and the multiple strategies available to order air traffic safely, observable actions performed by controllers might not convey the most appropriate information about perceived complexity, as this is mostly determined by mental activities as monitoring, evaluating, and planning. The adoption of a human information processing framework served as useful input for cognitive complexity modeling (Niessen, Eyferth, & Bierwagen, 1999; Histon, Hansman, Gottlieb, Kleinwaks, Yenson, Delahaye, & Puechmorel, 2002; Inoue, Furuta, Nakata, Kanno, Aoyama, & Brown, 2012), operational errors understanding (Shorrock, 2007), and sector design improvement by means of operator-based principles (Majumdar & Ochieng, 2007).

Chatterji and Sridhar (2001) related complexity factors and workload in a non-linear way, using parallel processing systems (neural networks) to classify traffic situations and predict workload on the basis of selected complexity factors and online (each 120s) subjective ratings. Importantly, the best classification performance was reached when factors relating to conflict presence and resolution difficulty were included in the training set, besides traffic geometry and dynamics. For instance, they included among complexity factors minimum separations (horizontal and vertical) and time-to-conflict, two metrics not directly observable which must be estimated (e.g. by means of perceptual matching, time and space estimations, motion projection), and that considerably impact experienced workload. Therefore, perceived complexity and workload were affected by *projected* traffic, i.e. by mental simulations performed in order to evaluate conflict presence and to build a timely and ordered "picture" of traffic situation (Mogford, 1997; Niessen & Eyferth, 2001; Durso & Manning, 2008).

The concept of relational complexity (Boag, Neal, Loft, & Halford, 2006) measured cognitive complexity for conflict detection of two convergent aircrafts as the number of transitions (above and below vertical and horizontal separation standard) predicted throughout flight progress. For instance, two aircraft flying at the same level with convergent trajectories presented one transition into loss of lateral separation state at some point; differently, if one of these two aircraft was climbing then two transitions were present, as both vertical and lateral separation would have been violated during flight progress. Conflict detection difficulty increased with aircraft transitions.

Athènes et al. (2002) considered cognitive activity performed by ATC to define complexity levels and build a workload index. In particular, they focused on self-regulation processes for cognitive resources and workload management, assuming that controllers perform along a speed-accuracy trade-off based on time pressure and uncertainty. These two metrics were related with the cognitive activity dedicated to each aircraft in a given traffic situation, and were modeled as cognitive complexity factors. Basically, nearly all controller observable actions are aimed at avoiding loss of separation between aircraft. Therefore, the time interval (called maturing time, MT) between first notice of a possible conflict until intervention (observable action) was assumed to express the mechanism of workload regulation: immediate actions (speed) can lower workload but could reveal as imprecise (accuracy), while delayed interventions add certainty to traffic situation but are more demanding and they shorten available time for resolution. Athènes et al. (2002) computed MT and added it in a measure of cognitive complexity (labeled task load index, TLI), assumed to be positively correlated with workload, validating it against subjective and psychophysiological arousal metrics.

In summary, two approaches emerged historically in the definition of ATC complexity. The bottom-up approach was typical of engineering and design research, and identified traffic and geometry complexity factors as workload drivers (mainly by means of interviews and observation techniques), trying to combine them in a complexity equation and compiling classification lists. The top-down approach emerged essentially from the application of cognitive psychology to ergonomic studies and operated on a reversed paradigm. Basically, such studies used a criterion for controllers' workload based on cognitive activity (as perceptual judgment, decision making process, allocation of attentional resources through time) to identify which complexity factors had an impact on such response. As pointed out in the literature (Hilburn, 2004), the most promising option for modeling ATC complexity is a combination of the two approaches, as they are useful for identification and manipulation of complexity factors to assess related cognitive response, identifying cognitive strategies adopted to manage perceived complexity. The work, among others, by Chatterji et al. (2001), Averty et al. (2002), Histon et al. (2002), and Boag et al. (2006), showed cognitive processes like perception and decision making as the actual drivers of controllers' workload, in particular in relation to conflict detection.

Conflict detection is a key component of ATC (Kallus, Van Damme, & Dittman, 1999) and one of the most investigated ATC tasks (Remington, Johnston, Ruthruff, Gold, & Romera, 2000; Galster, Duley, Masalonis, & Parasuraman, 2001; Loft, Humphreys, & Neal, 2004; Eyferth, Niessen, & Spaeth, 2003). It can be accomplished by means of multiple strategies (Rantanen & Nunes, 2005; Xu & Rantanen, 2003; Stankovic, Raufaste, & Averty, 2008; Loft, Bolland, Humphreys, & Neal, 2009), which generate different levels of workload. Basically, it consists of predicting the contemporary lateral and vertical loss of separation between two aircrafts. Cognitive workload has been demonstrated to be affected by the presence (predicted) of conflicting trajectories (Neal & Kwantes, 2009; Averty, 2005). Conflict detection is the result of cognitive projection of trajectories. Controllers need to project aircraft motion in time and space, and estimate aircraft future relative positions. The modeling of cognitive projection process for conflict detection is of primary importance in the ATC domain (Davison, 2006).

The need to find reliable workload measures in ATC complexity research that account for the cognitive activity in which controllers are engaged in has progressively received increasing attention, considering cognitive workload as reliable criterion for cognitive complexity, especially in relation to conflict detection.

3. COGNITIVE WORKLOAD

The concepts of workload and complexity are tightly coupled. The interest in cognitive workload measurement increased with the application of human information processing models to real work operations (Wickens, 2002). Moreover, the emergence of functional models of attention, which assumed that humans have limited resources for information processing (Kahneman, 1973; Wickens, 1984) and they manage them dynamically (Norman & Bobrow, 1975), further encouraged the investigation of cognitive workload in applied settings.

Similarly to cognitive complexity, cognitive workload has been heterogeneously defined since its conceptualization in ergonomics studies (Welford, 1977; Leplat & Welford, 1978; Moray, 1979). Despite this, nearly all cognitive workload definitions express it as the cost for applying mental effort while performing a task that is more cognitive than physical in nature (Kahneman, 1973; Wickens; 2002; Parasuraman & Caggiano, 2002; Cañas, 2004). Although the functional relationship between complexity and workload is still to be fully understood (Athènes et al., 2002), there is common agreement in the ATC scientific community on the position that considers experienced workload as a consequence of perceived cognitive complexity (Mogford et al., 1995; Loft, Sanderson, Neal, & Mooij, 2007; Djokic, Lorenz, & Fricke, 2010). Cognitive complexity results from the mediating effect of cognitive processes (e.g. perception, decision making, and memory) on taskload, and it determines experienced workload. Furthermore, individual characteristics (e.g. experience, familiarity, psychophysical state), and context (e.g. time pressure, noise, presence of distractors, etc.) affect cognitive processing, hence experienced workload.

Due to the mediating effect of cognitive processes, a linear transformation of taskload into experienced workload would be implausible. Cognitive strategies are to be intended as dynamic, subjective, and idiosyncratic (Davison, Histon, Ragnarsdottir, Major, Hansman, 2003). Despite this, there is agreement on the fact that controllers' cognitive strategies (e.g. for conflict detection) can be made explicit by investigation and, at least partially, formalized into a cognitive complexity model. For instance, heuristic processes (e.g. experienced-based pattern recognition) or higher level cognitive strategies like structured-based abstractions have been identified (see Histon, 2008). When higher workload is experienced, a strategy shift might occur, in order to manage experienced workload dynamically.

Workload deals with allocation of mental resources in order to accomplish task demands and, like complexity, is not directly observable. Despite this, workload can be inferred from several sources, i.e. is a multidimensional construct. Subjective estimations, performance indices, and psychophysiological responses are the three main groups of workload metrics. In ATC, overt behavior has been acknowledged as being a too elusive workload measure, due to the cognitive nature of control tasks. Subjective estimations of experienced workload have been widely used, and preferred to overt performance (Langan-Fox, Sankey, & Canty, 2009). However, due to possible subjective response biases, off-line administration, or intrusiveness of online ratings, the need for more objective and unobtrusive cognitive workload measures is still today of capital importance for ATC system safety. Psychophysiological measures offer the possibility of recording objective response of a person performing a task. Among them, eye tracking methods allows recording and processing of cognitively-related visual behavior data with little interference on the task currently performed. Since both bottom-up and top-down cognitive processes have been acknowledged to play a central role in the in ATC taskload-workload transformation, the study of controllers' visual behavior by means of eye metric recording might further benefit the modeling of ATC workload.

4. EYE METRICS IN ATC

ATC has a fundamental visual component. Ocular movements and fixations represent the primary process for obtaining relevant information from visual interfaces. Considering that retina is an extension of the central nervous system, the point of gaze is a reliable index of visual attended information. The eye-mind relationship assumes that observed region in the visual field are cognitively relevant for task accomplishment. In ATC tasks, time spent looking at specific elements on the radar display might therefore be assumed as representing operator's attention, thus being in relation with cognitive resources investment and mental representation of traffic. Furthermore, controllers' scanning strategy inform about allocation of perceptual and cognitive resources (Kang, Bass, & Lee, 2014).

The recording of ocular metrics in the aviation system can be traced back at the origin of the commercial aviation, when they started to be used mainly for pilots' performance assessment (Fitts, 1950; Morris & Miller, 1996; Veltman & Gaillard, 1998; Wilson, 2002), showing sensitivity to both workload and fatigue.

Differently, applications of ocular metrics recording in ATC domain are more recent. Blink duration and frequency are among the first ocular metrics used to infer ATC experienced workload (Brookings, Wilson, & Swain, 1996; Wilson & Russell, 2003): usually, results showed a negative relationship between blink metrics and workload. In parallel, fixation and dwell metrics (where total dwell time for a specific area of interest (AOI) is the sum of fixations' durations), have been indicated as positively related to workload. Fixation duration was observed to increase with higher processing complexity, together with the frequency of longer fixations (Stein, 1992; Metzger & Parasuraman, 2006; Rognin, Grimaud, Hoffman, & Zeghal, 2004; Zeghal, Grimaud, Hoffman, Rognin, Pellegrin, & Rodet, 2002; Hauland, 2008). Furthermore, pupil diameter (or pupil size variation) has been also related positively to cognitive workload of ATC tasks (Ahlstrom & Friedman-Berg, 2006).

In general, blinks and pupil size were used as workload metrics in most of researches. The diagnostic limitations of these metrics as *pure* workload indicators have been progressively evidenced. In fact, pupil size is also affected by general arousal and emotional reactions, while blink metrics do not provide spatial information about actual cognitive processing. Recent researches have shown alternative ocular metrics as workload indices. For instance, saccadic parameters (amplitude, duration, and peak velocity) have been shown to be sensitive to taskload and workload variations in several simulated settings (Di Stasi, Antolí, & Cañas, 2011; Di Stasi, Catena, Cañas, Macknik, & Martinez-Conde, 2013; Di Stasi, Marchitto, Antolí, Baccino, & Cañas, 2010).

As already mentioned, conflict detection in ATC is a fundamental component of controllers' operation, so that many researchers have focused on conflict detection study in order to model conflict detection complexity factors, strategies, and related workload. Differently, the study of ATC visual scanpaths and the analysis of ocular metrics during ATC conflict detection are relatively recent, but promising (Hunter & Parush, 2009; Martin, Cegarra, & Averty, 2011; Kang & Landry, 2010; 2014). Further empirical evidences from eye tracking data on conflict detection tasks can contribute to model such fundamental ATC decisional processes.

5. OVERVIEW OF THE THESIS

This Doctoral Thesis presents four experimental studies in which several ocular metrics were measured as workload indexes in simulated air traffic control (ATC) tasks. In particular, experimental tasks related to conflict detection. The use of eye tracking methods for evaluating workload related to ATC conflict detection task is relatively recent. In all experiments, task demands varied in order to generate different complexity levels, by manipulating acknowledged complexity factors. The complexity manipulation was assumed to contribute to perceived cognitive complexity of the traffic situation, hence experienced workload. Workload was assessed multidimensionally, as performance metrics and subjective ratings were also recorded.

In the first experiment (**Chapter 2**), a simple perceptual task about target position was used as easy experimental condition, and additional tasks (a paper-and-pencil task, and a simple mathematical operation) were added to create a difficult condition. Workload and fatigue were measured as a function of task complexity (TC) and time on task (TOT), by means of saccadic main sequence parameters (amplitude, peak velocity, and duration), response time, and subjective scales. As expected, TC affected both performance and subjective workload, and TOT affected fatigue scores. Differently, there was a performance improvement (decreased response times) through time for the higher complexity condition, exclusively. Such results were unexpected and were explained in terms of cognitive strategy and of optimal workload and underload. In order to accomplish additional tasks in higher complexity condition, aircraft callsign (3 digits) had to be memorized, besides the perceptual task. Differently, in lower complexity condition only aircraft position was relevant, with no need for processing further information. A better processing strategy could have been implemented through time in the higher complexity condition, while being unnecessary with lower complexity, as performance improvements were less possible. A second, complementary interpretation was supported by ocular data. Saccadic peak velocity of large saccades (>11°) presented a slow down effect for the low complexity condition exclusively, with bigger decrease in the first half of experiment. Such result suggested that lower complexity condition was most probably excessively simple, while the higher complexity condition probably determined optimal engagement, even allowing active management of cognitive resources investment to improve task performance. Underload situations are characterized by great time availability to perform a reduced number of tasks, most often of very low complexity. In this study, saccadic peak velocity showed to be sensitive to an underload condition, showing a slow down effect when workload is out of the range for optimal performance.

In **Chapter 3,** we modeled two complexity factors on the basis of specific ATC literature to create different complexity conditions and determine workload variations. Convergence angle and distance to convergence point were chosen for a conflict detection task experiment. More precisely, the task required to estimate arrival order at convergence point of two aircraft (i.e. a relative judgment, RJ). Although traffic scenarios were static (screenshots), they were built using an ATC flight traffic simulator. Results confirmed a slow down effect of saccadic peak velocity of large reaching saccades (>15°) with wider convergence angles and different (compared to same) distance to convergence point. Saccadic durations showed the same pattern. Therefore, there was probably an increase of precision in large saccades' performance with higher geometry complexity. Large saccades are the most demanding for ocular neuromotor system, and they might reduce their burst power and positional error with more complex geometries, thus optimizing visual scanning. Peak saccadic velocity confirmed its sensitivity to workload fluctuations, and the existence of a speed/accuracy trade-off.

Chapter 4 presents a study which used dynamic scenarios for conflict detection task. Convergence angle (CA) was manipulated as geometry complexity factor, while minimum distance at closest approach (MD) as traffic complexity factor. Conflicts and no conflicts scenarios were built on the basis of whether MD was below or above safe separation standard (5nm), respectively. Compared to previous experimental tasks, RJ was only a preliminary process within conflict detection task. Correct conflict detection was more difficult for scenarios with MD close to separation standard (e.g. 6nm), which showed higher error-proneness. However, conflicts were the most demanding scenarios, as showed by convergent workload measures. A further processing effort in presence of perpendicular routes was also observed. Effects of response time were controlled in the analysis of ocular metrics (analysis of covariance). Large saccades confirmed the sensitivity of peak velocity and duration to cognitive workload, as they showed lower values for conflicts, especially with right angles of convergence. Large ocular movements adapted to increased perceptual and cognitive workload. As for fixation data, they showed no effects once controlled for the effect of time needed to respond, so that fixation frequency and duration could be predicted by response time. In operational settings, where response time is difficult to compute (if not impossible), fixation metrics can inform about processing effort. Data suggested that the considerable differences emerged between conflicts and no conflicts might have been qualitative, rather than exclusively quantitative, meaning that two different conflict detection strategies were probably adopted in the two types of traffic scenarios. It was hypothesized that a ratio-based (e.g. distance-to-velocity ratio) strategy was adopted in no conflicts, in which a correct decision was rapidly made after performing few fixations and saccades, providing low subjective workload ratings. Probably, aircraft times to convergence point were estimated as considerably different, suggesting safe lateral separation. Differently, conflict trials were more demanding, so that a ratio-based estimation was deemed as unviable. Participants probably had to rely on recurrent processing of aircraft position updates in order to infer and compare aircraft motion equations.

The differences in speed and distance to convergence point at the beginning of each scenario might have been responsible for the successful application of a ratio-based strategy in no conflicts, or for a strategy shift in favor of a prediction motion strategy in conflicts. For instance, little speed differences combined with considerable differences of distance to convergent point probably suggested safe lateral separation throughout flight progress. Differently, moderated differences in both speed and distance might have added uncertainty about aircraft future relative position and separation. In summary, the study suggested that complexity of conflict judgment depends on relative judgments and estimations of speed, distance, and altitude differential values.

In **Chapter 5** differential values of flight metrics were therefore investigated as complexity factors. Several combinations of differential speed, altitude, and distance to convergence point were assumed to drive conflict detection strategy and to affect indirectly experienced workload. Pupil size was measured as workload index. In parallel, we proposed a quantitative analysis of gaze transitions between task relevant areas of interest (AOI), in order to identify adopted strategy from scanpath analysis. A frequency increase of a specific transition within the whole scanpath was assumed to indicate a more central role of such gaze transition type within the whole strategy for task resolution. Conversely, a significant decrement was assumed to indicate scarce importance within the strategy adopted. Increased workload would have limited the performance of low relevance gaze transitions. Important results emerged from gaze transition analysis. For example, transitions between aircraft data tags (altitude and speed information) were almost absent when lateral separation estimation was more demanding. Furthermore, it was observed an increase of transitions between aircrafts and convergence center when a more accurate processing of differential distance was required (i.e. with the littlest differential distance). Lastly, frequency of crossed transitions between aircraft elements (position and data tags) increased when both lateral and vertical separation estimation were required. Significant pupil dilations were observed in most demanding situations. However, the greatest pupil dilations were recorded for the easiest scenarios, in which no errors were made in conflict detection task, and minor resource investment was observed. These pupil dilations were interpreted as an emotional positive reaction due to the very low (or even null) uncertainty about conflict absence.

In summary, it was observed the expected negative relationship between differential distance to convergence point and workload, and between differential altitude and workload. Differently, the effect of differential speed highly depended on its combination with differential distance. However, it was observed an inherent difficulty for integrating speed differences into a projection model correctly.

In **Chapter 6** some conclusions are presented, together with some suggestions for practical applications. Proposed analysis of ocular metrics and of visual scanpaths by means of gaze transitions represent a relatively novel contribution to the study of complexity and workload in ATC conflict detection, and shall be considered as preliminary. The application of eye tracking methods in ATC domain could benefit in the future both the system design (e.g. conflict alert tools) and ATC training of new controllers, for instance by identifying experience-related differences in visual behavior, or the visual correlates of scanning strategies. Eye tracking methods might be successfully included as additional source of information in future programs for workload assessment.

REFERENCES

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623-636.

Arad, B. A. (1964). The controller load and sector design. *Journal of Air Traffic Control, May,* 12-31.

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D. & Collet, C. (2002). Complexity and Controller Workload: Trying to Bridge the Gap. In *Proceedings of the 2002 International Conference on Human-Computer Interaction in*

Aeronautics (HCI-Aero 2002). Cambridge, MA, USA: Massachusetts Institute of Technology.

Averty, P., Athènes, S., Collet, C., & Dittmar, A. (2002). Evaluating a new index of mental workload in real ATC situation using psychophysiological measures. In *Digital Avionics Systems Conference, 2002. Proceedings. The 21st* (Vol. 2, pp. 7A4-1). IEEE.

Averty, P. (2005). Conflict perception by ATCs admits doubt but not inconsistency. In *Proceedings of the 6th Air Traffic Management Research & Development Seminar*.

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, *49*(14), 1508-1526.

Brooker, P. (2008). SESAR and NextGen: Investing In New Paradigms. *Journal of navigation*, *61*(2), 195-208.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology*, *42*(3), 361-377.

Chatterji, G. B., & Sridhar, B. (2001, October). Measures for air traffic controller workload prediction. In *Proceedings of the First AIAA Aircraft Technology, Integration, and Operations Forum*. Los Angeles, California.

Cañas, J. J. (2004). *Personas y máquinas. El diseño de su interacción desde la ergonomía cognitiva.* Madrid: Ediciones Pirámide.

Cilliers, P. (1998). *Complexity and postmodernism: understanding complex systems*. London: Routledge.

Crévits, I., Debernard, S., & Denecker, P. (2002). Model building for air-traffic controllers' workload regulation. *European Journal of Operational Research*, *136*(2), 324-332.

Cummings, M. L., & Tsonis, C. G. (2005, September 26-30). Deconstructing complexity in air traffic control. In *49th Human Factors and Ergonomics Society Annual Meeting, Orlando, FL.*

Cummings, M. L., & Tsonis, C. G. (2006). Partitioning complexity in air traffic management tasks. *The International Journal of Aviation Psychology*, *16*(3), 277-295.

Davison, H. J. (2006). *Modeling the air traffic controller's cognitive projection process.* Report No. ICAT-2006-1. MIT: Massachusetts Institute of Technology.

Davison, H. J., Histon, J. M., Ragnarsdottir, M. D., Major, L. M., & Hansman, R. J. (2003, June). Impact of operating context on the use of structure in air traffic controller cognitive processes. Paper presented at the *5th Eurocontrol/FAA ATM R&D Seminar.* Budapest, Hungary.

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2011). Main sequence: an index for detecting mental workload variation in complex tasks. *Applied ergonomics*, *42*(6), 807-813.

Di Stasi, L. L., Catena, A., Cañas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews*, *37*(5), 968-975.

Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, *16*(6), 330-333.

Djokic, J., Lorenz, B., & Fricke, H. (2010). Air traffic control complexity as workload driver. *Transportation research part C: emerging technologies*, *18*(6), 930-936.

Durso, F. T., & Manning, C. A. (2008). Air traffic control. *Reviews of human factors and ergonomics*, *4*(1), 195-244.

Eyferth, K., Niessen, C., & Spaeth, O. (2003). A model of air traffic controllers' conflict detection and conflict resolution. *Aerospace science and technology*, *7*(6), 409-416.

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*.

Flynn, G. M., Leleu, C., & Zerrouki, L. (2003). Traffic complexity indicators and sector typology analysis of US and European centres. *EEC Note*, *20/03*. Bretigny-sur-Orge: Eurocontrol Experimental Research Center.

Galster, S. M., Duley, J. A., Masalonis, A. J., & Parasuraman, R. (2001). Air traffic controller performance and workload under mature free flight: Conflict detection and resolution of aircraft self-separation. *The International Journal of Aviation Psychology*, *11*(1), 71-93.

Hilburn, B. (2004). Cognitive complexity in air traffic control: A literature review. *EEC Note*, *4*(04). Bretigny-sur-Orge: Eurocontrol Experimental Research Center.

Histon, J. M., Hansman, R. J., Gottlieb, B., Kleinwaks, H., Yenson, S., Delahaye, D., & Puechmorel, S. (2002). Structural considerations and cognitive complexity in air traffic control. In *Digital Avionics Systems Conference, 2002. Proceedings. The 21st* (Vol. 1, pp. 1C2-1). IEEE.

Histon, J. M. (2008). *Mitigating complexity in air traffic control: the role of structure-based abstractions* (Doctoral dissertation, Massachusetts Institute of Technology Cambridge, MA 02139 USA).

Hockey, G. R. (1997). Compensatory control in the regulation of human performance under stress and high workload; a cognitive-energetical framework. *Biological psychology*, *45*(1-3), 73-93.

Hollnagel, E. (2007). Flight decks and free flight: Where are the system boundaries? *Applied ergonomics*, *38*(4), 409-416.

Hollnagel, E. (2012). Coping with complexity: past, present and future. *Cognition, Technology & Work*, *14*(3), 199-205.

Hunter, A. C., & Parush, A. (2009, October). Using eye movements to uncover conflict detection strategies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 22, pp. 1729-1733). Sage Publications.

Hurst, M. W., & Rose, R. M. (1978). Objective Job Difficulty, Behavioural Response, and Sector Characteristics in Air Route Traffic Control Centres. *Ergonomics*, *21*(9), 697-708.

ICAO, International Civil Aviation Organization, (2013). *Global air navigation plan 2013-2028.* Doc 9750-AN/963, Fourth Edition, 2013.

Johnson, N. F. (2007). *Two's company, three is complexity: a simple guide to the science of all sciences*. Oneworld Pubns Ltd.

Kahneman, D. (1973). *Attention and effort* (p. 246). Englewood Cliffs, NJ: Prentice-Hall.

Kallus, K. W., Van Damme, D., & Dittmann, A. (1999). Integrated task and job analysis of air traffic controllers–phase 2: task analysis of en-route controllers. *Deliverable HUM.ET1.ST01.1000-DEL04.* Eurocontrol: Bruxelles.

Kang, Z., Bass, E. J., & Lee, D. W. (2014, September). Air traffic controllers' visual scanning, aircraft selection, and comparison strategies in support of conflict detection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 77-81). SAGE Publications.

Kang, Z., & Landry, S. J. (2010). Capturing and analyzing visual groupings of multiple moving targets in an aircraft conflict detection task using eye movements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 23, pp. 1906-1910). SAGE Publications.

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors*, *56*(6), 1150-1162.

Kopardekar, P. (2000). *Dynamic Density – A Review of Proposed Variables*. FAA NAS Advanced Concepts Branch ACT-540: FAA.

Langan-Fox, J., Sankey, M. J., & Canty, J. M. (2009). Human factors measurement for future air traffic control systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(5), 595-637.

Laudeman, I. V., Shelden, S. G., Branstrom, R., & Brasil, C. L. (1998). *Dynamic density: an air traffic management metric.* Report NAS 1.15: 112226. Ames Research Center, Moffett Field, California: FAA.

Leplat, J., & Welford, A. T. (1978). Symposium on mental workload. *Ergonomics, 21,* 141-233.

Loft, S., Humphreys, M., & Neal, A. (2004). The influence of memory for prior instances on performance in a conflict detection task. *Journal of Experimental Psychology: Applied*, *10*(3), 173-187.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(3), 376-399.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, *15*(2), 106.

Majumdar, A., & Ochieng, W. Y. (2002). Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload. *Transportation Research Record: Journal of the Transportation Research Board*, *1788*(1), 58-69.

Majumdar, A., & Ochieng, W. Y. (2007). Air traffic control complexity and safety: Framework for sector design based on controller interviews of complexity factors. *Transportation Research Record: Journal of the Transportation Research Board*, *2007*, 70-80.

Majumdar, A., & Polak, J. (2001). Estimating the capacity of Europe's airspace using a simulation model of air traffic controller workload. Paper No. 01-3250. In *80th Annual Meeting of the Transportation Research Board.* Washington, DC: Transportation Research Board.

Martin, C., Cegarra, J., & Averty, P. (2011). Analysis of mental workload during en-route air traffic control task execution based on eye-tracking technique. In *Engineering Psychology and Cognitive Ergonomics* (pp. 592-597). Springer Berlin Heidelberg.

Masalonis, A. J., Callaham, M. B., & Wanke, C. R. (2003, June). Dynamic density and complexity metrics for realtime traffic flow management. In *Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar*.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 519-528.

Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(1), 35-49.

Metzger, U., & Parasuraman, R. (2006). Effects of automated conflict cuing and traffic density on air traffic controller performance and visual attention in a datalink environment. *The International Journal of Aviation Psychology*, *16*(4), 343-362.

Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). *Complexity construct in air traffic control: a review and synthesis of the literature* (DOT/FAA/-CT TN95/22). Atlantic City, NJ: FAA.

Mogford, R. H. (1997). Mental models and situation awareness in air traffic control. *The International Journal of Aviation Psychology*, *7*(4), 331-341.

Moray, N. (1979). *Mental Workload. Its Theory and measurement.* New York: Plenum Press.

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological psychology*, *42*(3), 343-360.

Neal, A., & Kwantes, P. J. (2009). An evidence accumulation model for conflict detection performance in a simulated air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(2), 164-180.

Niessen, C., Eyferth, K., & Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. *Ergonomics*, *42*(11), 1507-1520.

Niessen, C., & Eyferth, K. (2001). A model of the air traffic controller's picture. *Safety Science*, *37*(2), 187-202.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive psychology*, *7*(1), 44-64.

Parasuraman, R., & Caggiano, D. (2002). Mental Workload. In V. S. Ramachandran (Ed.) *Encyclopedia of the human brain (pp. 17-27).* San Diego: Academic Press.

Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996, July). A framework for the evaluation of air traffic control complexity. In *Proceedings of the AIAA Guidance Navigation and Control Conference, San Diego, CA*.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *The International Journal of Aviation Psychology*, *15*(4), 339-362.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *42*(3), 349-366.

Rognin, L., Grimaud, I., Hoffman, E., Zeghal, K., 2004. Assessing the impact of a new instruction on air traffic controller monitoring tasks. In: *Proceedings of the International Conference on Human–Computer Interaction in Aeronautics*. AAAI Press, Toulouse, France.

Shorrock, S. T. (2007). Errors of perception in air traffic control. *Safety science*, *45*(8), 890-904.

Sperandio, J. C. (1971). Variation of operator's strategies and regulating effects on workload. *Ergonomics*, *14*(5), 571-577.

Sperandio, J. C. (1980). *La psychologie en ergonomie* (Vol. 80). Paris: Presses universitaires de France.

Stankovic, S., Raufaste, É., & Averty, P. (2008). Determinants of conflict detection: A model of risk judgments in air traffic control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(1), 121-134.

Stein, E.S. (1985) *Air traffic controller workload: An examination of workload probe* (Report FAA/CTTN90/60). Atlantic City, New Jersey: FAA.

Stein, E.S. (1992). *Air traffic control visual scanning.* (DOT/FAA/CT-TN92/16). Atlantic City, NJ: Department of Transportation, Federal Aviation Administration Technical Center.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*(5), 656-669.

Welford, A. T. (1977). La charge mentale de travail comme fonction des exigences de la capacité de la stratégie et de la habilité. Rapport de synthèse. *Le travail humain*, *40*, 283-304.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–101). New York: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wickens, C. D., Mavor, A.S. & McGee, J.P. (Eds.) (1997). *Flight to the future: human factors in air traffic control*. Washington, DC: National Academy of Sciences, National Research Council.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, *12*(1), 3-18.

Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human*

Factors: The Journal of the Human Factors and Ergonomics Society, *45*(3), 381-389.

Xu, X., & Rantanen, E. M. (2003). Conflict detection in air traffic control: A task analysis, a literature review, and a need for further research. In *12th International Symposium on Aviation Psychology, 2003*.

Zeghal, K., Grimaud, I., Hoffman, E., Rognin, L., Pellegrin, A., & Rodet, L. (2002). Delegation of spacing tasks from controllers to flight crew: Impact on controller monitoring tasks. In *Digital Avionics Systems Conference, 2002. Proceedings. The 21st* (Vol. 1, pp. 2B3-1). IEEE.

CHAPTER 2

SACCADIC PEAK VELOCITY AS WORKLOAD INDEX

CHAPTER 2

FROM SUBJECTIVE QUESTIONNAIRES TO SACCADIC PEAK VELOCITY: A NEUROERGONOMICS INDEX FOR ONLINE ASSESSMENT OF MENTAL WORKLOAD[1]

ABSTRACT

Experts in human factors engineering and applied cognitive research are in search of reliable measures for the online monitoring of mental workload (MW) during active working. Due to the limitations associated with subjective and performance measures, researchers have turned their attention to oculomotor indices. We present data from an ongoing research project on the evaluation of MW from saccadic peak velocity (PV). Participants were tested in a complex experimental setting simulating typical air traffic control (ATC) tasks. Changes in MW were evaluated with a multidimensional methodology, using subjective ratings, behavioral indices, and saccadic dynamics data. Comparison of our results with the literature suggests that PV shows sensitivity in the real-time detection of differences in mental state and is a strong candidate for the online diagnosis of operator under- or over-load in the workplace.

---

[1] Published as:
Di Stasi, L. L., Marchitto, M., Antolí, A., Rodriguez, E., & Cañas, J. J. (2010). From subjective questionnaires to saccadic peak velocity: A neuroergonomics index for online assessment of mental workload. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in understanding human performance: Neuroergonomics, human factors design, and special populations* (pp. 11–20). Boca Raton, FL: CRC Press.

1. INTRODUCTION

Mental workload (MW) has been defined as "a composite brain state or set of states that mediates human performance of perceptual, cognitive, and motor tasks" (Parasuraman & Caggiano, 2002, p. 17) and has been used to explain how human operators face heightened cognitive demands associated with increased task complexity in job situations where cognitive skills are more important than physical ones (Cacciabue, 2004). In an information technology society, changes in MW could have significant impacts on operator performance, possibly causing delays in information processing or even cause operators to ignore or misinterpret incoming information (Ryu & Myung, 2005). Consequently, automation research has identified a need to monitor operator functional state in real-time in order to determine the most appropriate type and level of automated assistance for helping operators to complete tasks safely (e.g. Langan-Fox, Canty, & Sankey, 2009). The development of a method for monitoring operator attentional states in real-time during interactions with artifacts could be a good starting point for undertaking the investigation of this crucial issue. Our research is relevant to a variety of domains, from air traffic control (ATC) towers to call centers. For example, online monitoring of changes in an operator's attentional state (mental under/overload) could help in the design of adaptive systems that can allocate tasks in a dynamic way between the operator and the machine (e.g. Kaber, Perry, Segall, McClernon, & Prinzel, 2006).

2. PV AS A NEUROERGONOMICS INDEX FOR THE ONLINE ASSESSMENT OF MENTAL WORKLOAD

Due to the limitations associated with subjective and performance measures, researchers have turned their attention to oculomotor indices. Decades of investigations have focused primarily on pupil diameter and blink rate as the best indicators of workload dynamics. However, some considerable drawbacks limit the reliability of these indices and their applicability in the context of the workplace (Di Stasi, Marchitto, Antolí, Baccino, & Cañas, 2010). Both blink duration and rate increase with fatigue and time on task (e.g. Ryu & Myung, 2005), and the amount of not-processed information increases in parallel with these indices. It has been known for long time that there is a strong correlation between pupil amplitude variations and the amount of cognitive resources used to perform a task (e.g. Ahlstrom & Friedman-Berg, 2006) and these indicators continue to be used in workload assessments. However, pupil size is also influenced by several factors that can sometimes be difficult to control, including emotion and environmental lighting. The validity of pupil size as a MW index could be questioned if appropriate controls are not performed. Therefore, we are presently exploring saccadic peak velocity (PV) as an alternative to pupil diameter and blink rate to measure MW.

The relationship between saccadic amplitude, duration and PV has been called "main sequence", to indicate that PV and saccadic duration increase systematically with amplitude (Bahill, Clark, & Stark, 1975). Empirical results on the relationship between saccadic dynamics and activation state have shown that task complexity and other task variables can influence the PV response. Furthermore, in visual performance tasks, PV varies with the subject's state of mental activation, MW and fatigue (e.g. Di Stasi *et al.*, 2009; 2010a; 2010b).

Recently, Di Stasi *et al.* (2009) studied the influence of MW on saccadic dynamics in ecological and complex settings (fixed-base driving simulators) within the relationship between risky driving behavior and MW. The study included eye-activity parameters in the methodology used for driver assessment. It was found that the high-risk group had shorter saccade duration and a higher saccadic PV than the low-risk group. On the Mental Workload Test (MWT), the high-risk group scored significantly higher on both perceptual/central and answer demand. Furthermore, PV showed several significant correlations with MWT dimensions. The negative correlations of PV and subjective scales of MW suggested that, given a high level of risk proneness, lower PV was associated with a higher level of subjective workload. However, there are some caveats to this work. Due to the complexity of the simulated scenarios, the authors analyzed normalized (by saccade number) PV values, considering the difference between the test session and training session (baseline), but without applying the bin-analysis (analyzing PV as a function of saccade length; Di Stasi *et al.*, 2010a), necessary to control for the influence of amplitude on PV.

Similar results were obtained by Di Stasi *et al.* (2010a) in a more controlled experimental setting. In this study the authors demonstrated that PV was sensitive to variations in MW during ecological driving tasks, showing again an inverse relation between PV and task complexity. PV decreased by 7.2°/s as the MWT score increased by 15.2 and reaction time for a secondary task increased by 46msec. Saccade duration and average velocity were not affected by differences in task complexity. The design of this experimental investigation allowed the authors to differentiate between the effects of time-on-task (TOT) and changes in MW from the same dataset. In this experiment no effect of fatigue was found. Even if the analyses for influences of TOT on PV revealed no effects, the authors suggested that the relationship between fatigue and MW requires further investigations in a more controlled experimental setting.

In an experiment that simulated multitasking performance in ATC setting, Di Stasi *et al.* (2010b) studied the relation between the main sequence parameters and task load. The created tasks demanded different perceptual and central processing resources, as well as response resources. Results obtained from the subjective ratings (MWT) and behavioral measures (number of errors and delayed answers) confirmed that MW levels varied according to task demand. These different levels of MW were reflected in PV values. The authors found that there was a 6.3°/s reduction in PV when task complexity assessed by MWT increased by 10.6 and performance was also affected (6 delayed answers). However, there was one limitation to this work. The authors were unable to distinguish between the effects of task complexity and TOT, due to the nature of the experimental design. Indeed, to avoid any effect of task switching during the experimental session, the order of task complexity variable was not balanced across participants.

On the basis of these results we designed a well-controlled experiment to surmount the methodological problems encountered in the previous studies, and particularly the influence of TOT on the disruption of the main sequence rules. The experiment was conducted in the ATC domain.

3. THE EXPERIMENT

Modern complex systems such as nuclear power plants, air-flight control systems and weapon systems often impose heavy MW on their operators. The high rate of information flow, the complexity of the information, numerous difficult decisions and task-time stress could overwhelm the operators (Hwang *et al.*, 2008). In the aviation domain "controller workload is likely to remain the single greatest functional limitation on the capacity of the air traffic management system" (Hilburn, 2004, p. 1).

In this experiment participants were tested in a complex experimental setting simulating typical ATC tasks. To control for TOT effects the experiment was performed on two different days. Changes in MW were evaluated with a multidimensional methodology using subjective ratings (MWT), performance indices, and psychophysiological data.

3.1 PARTICIPANTS

Thirteen volunteers (4 males) took part in this experiment (mean age = 22.4 years; SD = 2 years). None of the participants had ATC experience. All subjects had normal or corrected-to-normal vision and signed a consent form that informed them of the risks of the study and the treatment of personal data. They received course credits for participating in the study. The study was conducted in conformity with the Declaration of Helsinki.

3.2 STIMULI AND INSTRUMENTS

The same equipment configuration and experimental setting of Di Stasi *et al.* (2010b) were used. Participants were tested on two different simulated ATC tasks. Tasks were created as a simplified version of some actual ATC operations, respecting the main artifacts and interaction sequences that the ATC operators have to deal with in their complex environments (Cox, Sharples, Stedmon, & Wilson, 2007).

The visible matrix of airspace consisted of 6 concentric green nodes presented on a black background. The radii of the six nodes were 1.5, 3, 4.5, 6, 7.5, and 9cm, respectively. Aircrafts (red dots with a concentric inner black dot) were always located on a visible node within the matrix and could appear on any of the five adjacent nodes (although never on the smallest). For each node, 8 positions were chosen in which aircraft could be shown (clockwise: up, 45°, 90°, 135°, 180°, 225°, 270°, 315°). A total of 40 different stimuli were constructed and stimuli were randomly presented four times per block (two blocks, 320 trials in total per experimental session). Aircraft position was updated every 1.5s, within which time the aircraft would be presented to one of the 5 adjacent nodes. Aircraft were represented visually with their call signs (3 digits). Forty such call signs were extracted from a random number table. Call signs were presented in a size 11, Calibri font. Aircraft color was constant and subtended 1° of visual angle.

3.3 DESIGN/PROCEDURE

The experimental design follows a 2 Task Complexity (TC: low and high) x 2 Time-on-Task (TOT: 1st block and 2nd block). TC was varied by manipulating the number of simultaneous tasks. Both the number of simultaneous tasks and TOT were assumed to lead to different attentional states (Wickens, 2002). To avoid any serial effect, the levels of TC were balanced across the two days. The levels of the TOT variable were obtained by dividing the session into two experimental blocks: the first part (first 20 min) and the second part (last 20 min of experiment).

We used a multiple-measures approach to evaluate the effectiveness of our manipulation. To evaluate the subjective ratings of mental state, we made use of three different questionnaires. First, the Stanford Sleepiness Scale (SSS) was used as a global measure of sleepiness (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). Second, the Borg rating of perceived exertion (Borg, 1998) was used to evaluate the perceived task fatigue. Third, the MWT was used to estimate subjective mental effort (for more details, see Di Stasi *et al.*, 2009). All tests were translated into the Spanish language.

To estimate possible effects of changes in activation state on eye movements indices, we analyzed saccadic main sequence (saccadic amplitude, duration, and PV). Eye-movement data were analyzed using medians, rather than means, to minimize the effects of outliers and noisy data.

Participants were tested in a quiet room, and sat approximately 60cm from a display screen. There were four experimental blocks, split over two days. At the beginning of each experimental day participants were required to complete the SSS questionnaires; after each experimental block subjects were required to complete the SSS, Borg scale, and MWT questionnaires. Subjects who scored higher than 3 on the SSS scale at the beginning of both experimental days were excluded from further testing.

The low-complexity task comprised a decision task performed using a computer mouse whose two buttons were the answer keys. Each subject was instructed to determine (and answer with the mouse) whether the position of each aircraft on the display screen was either “critical” or “non-critical”. An aircraft lying in one of the second and third nodes (3 and 4.5cm of radius) was defines as being in a critical position. The experimenter explained to the participants that the critical position reflected the supposed closeness of the aircraft to the airport and its priority in needing assistance. By contrast, if the aircraft was located in one of the three largest circles it was judged to be in a non-critical position. Participants were requested to perform the mouse task using the hand other than the one used habitually for writing. The high-complexity task introduced a concurrent paper-and-pencil task to be performed along with the decision task and added a simple mathematical operation to be carried out with the call sign written in the operator’s answer sheet (for more details see Di Stasi *et al*., 2010b).

Finally, the number of errors and reaction times for each trial at each complexity level were analyzed.

4. RESULTS

First, we examined the effectiveness of the TC and TOT manipulation by analyzing the subjective rating scores, number of errors, and reaction times (RT) on the detection task. Analyses were run on data obtained from 8 participants. Mean scores in the MWT, SS, and Borg scale were submitted to a 2 (TC: low and high) x 2 (TOT: 1st block and 2nd block) repeated measures analysis of variance.

For MWT analysis significant main effect was obtained for TC [$F(1, 7) = 7.0$, $p = .03$, $\eta_p^2 = .50$] (Table 1). No significant effects were found for TOT or for the interaction of both factors ($F < 1$).

Analysis of the SSS and Borg scale mean scores demonstrated a significant effect only for the TOT factor [$F(1, 7) = 5.6$, $p = .05$, $\eta_p^2 = .44$ and $F(1, 7) = 6.1$, $p = .04$, $\eta_p^2 = .47$] respectively for the SSS and Borg scores. No significant effects were found for TC or for the interaction of both factors ($F < 1.7$) (Table 1).

Mean RTs in the decision task for each participant were analyzed using repeated measures ANOVA, with TC and TOT as repeated factors. As expected, the main effect was significant for both TC and TOT [$F(1, 7) = 57.4$, $p < .001$, $\eta_p^2 = .89$, and $F(1, 7) = 16.6$, $p = .005$, $\eta_p^2 = .70$, respectively], confirming that our manipulation was accurate. The interaction was also significant [$F(1, 7) = 7.2$, $p = .03$, $\eta_p^2 = .51$]. Simple effects analysis showed significant differences for each comparison (all $p < .05$; Table 1), except for 1st vs. 2nd block in the low complexity condition ($p = .08$). The effect of TOT on RT was smaller in the low-complexity task than in the high-complexity task. In the high-complexity task (but not in the low-complexity task) it was observed that participants could reduce RT by implementing a more efficient strategy for information processing: learning effects could therefore play an active role in decreasing RT in the high-complexity condition. More studies will be required to evaluate the relationship between TOT and on-task learning.

Errors in the decision task were always less than 8% of the total number of trials (for all subjects irrespective of task complexity).

In the next step we analyzed the sensitivity of the saccadic main sequence parameters to detect variation in MW and fatigue across the TC and TOT manipulation. The amplitudes of the saccades were categorized into 9 bins (henceforth Saccade Length, Di Stasi *et al.*, 2010a), ranging from 3° to 12° (with 1° increments). The medians of the saccadic duration and PV were then submitted to two separate 2 (TC) x 2 (TOT) x 9 (Saccade Length) repeated measures ANOVA.

Regarding saccade duration (SD), there was only a main effect of Saccade Length [$F(8, 56) = 138.6$, $p < .001$, $\eta_p^2 = .95$]. TC, TOT, and the interactions of both factors with Saccade Length were not significant ($F < 1$; Table 1). As expected, for SD and PV, higher values were found for larger saccades (main sequences rules; Becker, 1989).

Regarding peak velocity (PV), ANOVA only revealed significant main effects for Saccade Length [$F(8, 56) = 563$, $p < .001$, $\eta_p^2 = .99$]. TOT and TC main effects were not significant ($F < 4.1$; Table 1). However, as expected, interactions between TC and Saccade Length, and between TC, TOT, and Saccade Length, were reliable [$F(8, 56) = 2.3$, $p = .03$, $\eta_p^2 = .25$, and $F(8, 56) = 2.58$, $p = .02$, $\eta_p^2 = .27$, respectively]. We next analyzed this interaction by separating the first versus the second block

(Figure 1). Simple-effects analyses revealed that the low-complexity task had lower PV values (M = 400.8, SE = 13.3 vs. M = 450.3, SE = 11.9) in the last bin (t = -2.63, p = .03). No significant effects were observed in the second block.

	TASK			
	1st block (first 20 min)		2nd block (last 20 min)	
	Low complexity	High complexity	Low complexity	High complexity
MWT score	47.0 (7.1)	58.3 (3.4)	47.4 (7.0)	60.3 (4.0)
SSS score	2.5 (.33)	2.63 (.26)	2.88 (.52)	3.25 (.37)
Borg Score	10.5 (.82)	10.25 (.75)	11.63 (1.33)	12.5 (1.0)
RT (msec)	807.1 (72.6)	1338.8 (127.2)	746.3 (62.5)	1071.4 (92.2)
PV (°/sec)	314.1 (6.7)	326.0 (11.8)	321.4 (8.0)	327.9 (11.6)
SD (msec)	41.20 (2.23)	41.23 (2.34)	41.44 (2.44)	41.5 (2.41)

Table 1. Overview of the experimental results. Mean values and SE (in parentheses) recorded on several dependent variables for 8 participants.

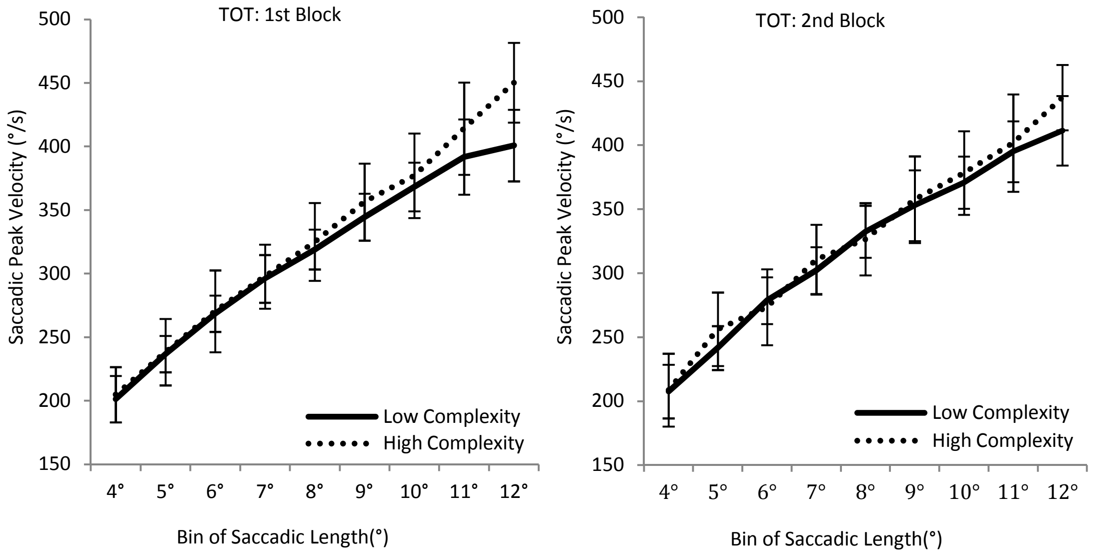

Figure 1. Illustration of TC x TOT x Saccade Length interaction effect on PV. Vertical bars denote 0.95 confidence intervals.

5. CONCLUSION

The detection of potential operator overload situations is a first step towards the avoidance of incorrect decisions brought about by increased MW and/or fatigue. Comparison of the results on PV reported here with those of other literature studies suggests that PV can afford a sensitive and real-time index of changes in mental state, and PV is therefore a candidate for under/overload diagnosis in complex and operational environments.

Because TC was related to the subjective rating of MW, the effect of TC on PV appears to be due to variation of MW rather than with scores on the SSS and Borg scales that assess fatigue and sleepiness. These two scales were however affected by changes in TOT that had no effects on PV.

The effect of TC on PV (Figure 1) could be explained by considering the nature of this parameter. When a saccadic movement starts, it has an initial velocity and then accelerates. The PV is the point at which acceleration becomes negative. Unlike velocity, PV is independent of saccadic duration because it is not *a priori* linked to it by a mathematical definition. Furthermore, PV is independent of the distance at which saccades terminate, even though the apparent duration of saccades depends on distance (Becker, 1989). PV therefore appears to afford a good index of saccadic programming, and can reflect MW effects on saccadic programming independently of distance and duration. The fact that we found a significant effect of PV only in the largest distance could be explained by the relatively short saccadic magnitude (from 3° to 12°); it is possible that the mathematical relationship between these parameters could mask the effect of our main manipulation. Stronger effects of PV are found when saccade amplitudes are larger, for example in driving simulation tasks (Di Stasi *et al*., 2010a).

Our research is relevant to a variety of domains ranging from ATC towers to call centers. For example, using real-time PV measures, neuroergonomists could better evaluate when an operator's attentional state is changing (mental under/overload), helping in the design of systems able to allocate tasks in a dynamic way between the operator and the machine.

ACKNOWLEDGEMENTS

This study was in part supported by Spanish national project SEJ2007-63850/PSIC (J. J. Cañas), and by national grant FPU AP2006-03664, awarded to the first author.

REFERENCES

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623-636.

Becker, W. (1989). Saccadic eye movements as a control system: Metrics. In: R. H. Wurtz, and M. E., Goldberg (Eds.), *Reviews of Oculomotor Research, vol. 3: The Neurobiology of saccadic Eye Movements,* 13-67*.* Hillsdale: Elsevier.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3), 191-204.

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Champaign, IL: Human kinetics.

Cacciabue, P. C. (2004). Guide to applying human factors methods: Human error and accident management in safety-critical systems. London: Springer Science & Business Media.

Cox, G., Sharples, S., Stedmon, A., & Wilson, J. (2007). An observation tool to study air traffic control and flightdeck collaboration. *Applied ergonomics*, *38*(4), 425-435.

Di Stasi, L. L., Álvarez-Valbuena, V., Cañas, J. J., Maldonado, A., Catena, A., Antolí, A., & Candido, A. (2009). Risk behaviour and mental workload: multimodal assessment techniques applied to motorbike riding simulation. *Transportation research part F: traffic psychology and behaviour*, *12*(5), 361-370.

Di Stasi, L. L., Renner, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Cañas, J. J., ... & Pannasch, S. (2010a). Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, space, and environmental medicine*, *81*(4), 413-417.

Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010b). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, *16*(6), 330-333.

Hilburn, B. (2004). Cognitive complexity in air traffic control - A literature review. *EEC Note, 4*(*04*). Brétigny-sur-Orge: Eurocontrol Experimental Research Center.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R. & Dement, W. C. (1973). Quantification of Sleepiness: A New Approach. *Psychophysiology*, *10*, 431–436.

Hwang, S. L., Yau, Y. J., Lin, Y. T., Chen, J. H., Huang, T. H., Yenn, T. C., & Hsu, C. C. (2008). Predicting work performance in nuclear power plants. *Safety science*, *46*(7), 1115-1124.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, *36*(5), 447-462.

Langan-Fox, J., Canty, J. M., & Sankey, M. J. (2009). Human–automation teams and adaptable control for future air traffic management. *International Journal of Industrial Ergonomics*, *39*(5), 894-903.

Parasuraman, R., & Caggiano, D. (2002). Mental workload. In: V. S. Ramachandran (Ed.), *Encyclopedia of the human brain*, *3*, 17-27. San Diego: Academic Press.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, *35*(11), 991-1009.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

CHAPTER 3

SACCADES AND COGNITIVE COMPLEXITY

CHAPTER 3

OCULAR MOVEMENTS UNDER TASKLOAD MANIPULATIONS: INFLUENCE OF GEOMETRY ON SACCADES IN AIR TRAFFIC CONTROL SIMULATED TASKS[2]

ABSTRACT

Traffic geometry is a factor that contributes to cognitive complexity in air traffic control. In conflict detection tasks, geometry can affect the attentional effort necessary to correctly perceive and interpret the situation; online measures of situational workload are therefore highly desirable. In this study, we explored whether saccadic movements vary with changes in geometry. We created simple scenarios with two aircraft and simulated a conflict detection task. Independent variables were the conflict angle and the distance to convergence point. We hypothesized lower saccadic peak velocity (and longer duration) for increasing complexity, that is, for increasing conflict angles and for different distances to convergence point. Response times varied accordingly with task complexity. Concerning saccades, there was a decrease of peak velocity (and a related increase of duration) for increased geometry complexity for large saccades (>15°). The data therefore suggest that geometry is able to influence "reaching" saccades and not "fixation" saccades.

---

[2] Published as:
Marchitto, M., Di Stasi, L. L., & Cañas, J. J. (2012). Ocular movements under taskload manipulations: Influence of geometry on saccades in air traffic control simulated tasks. *Human Factors and Ergonomics in Manufacturing & Service Industries, 22*(5), 407-419. doi: 10.1002/hfm.20320

1. INTRODUCTION

1.1. COMPLEXITY IN AIR TRAFFIC CONTROL

Complexity has been a largely debated matter in the air traffic control (ATC) domain since the origin of commercial aviation. A lot of effort has been spent addressing the many facets and levels of analysis of this concept in the operative domain (Hilburn, 2004). Cognitive complexity is defined as the perceived level of complexity at the individual cognitive level (Cummings & Tsonis, 2005). Many factors belonging to different levels (organizational procedures, traffic environment, and display complexity) have been indicated as influencing the ATC Operators' (ATCOs') cognitive complexity. One mission of design and assessment methodologies in ergonomics is to find measures that efficiently express complexity mitigation processes. These measures could help in identifying paradoxical situations in which perceived complexity is increased by the design and interaction with the technology. In this study, we do not deal with the assessment of specific interface solutions, but with the possibility of observing changes in psychophysiological activity that mirrors environmental – traffic geometry – manipulations. Traffic geometry is manipulated in simple scenarios, with two aircraft that originate different configurations due to their specific position in the control display. Traffic geometry is retained as an environmental factor that can modify perceived cognitive complexity. Therefore, we look for sensitive ocular measures influenced by traffic geometry for possible future online workload assessment.

Workload evaluation is intimately related to new ATC paradigms that are currently being investigated for future implementation, such as so-called "free flight" (Hollnagel, 2007). It is a matter of increasing importance how relaxing constraints on routes and latitudes will impact the controllers' ability to maintain a correct and constantly actualized mental model of the traffic situation (Remington, Johnston, Ruthruff, Gold, & Romera, 2000). If a reduction in traffic regularity would undermine information organization on the control display, it is reasonable to think that getting all relevant information from displays will take longer. If the role of air traffic controllers becomes more focused on passive monitoring than on active control (Metzger & Parasuraman, 2001), several issues that determine perceived cognitive complexity and therefore influence performance on primary control activities will have to be investigated. The ability to develop and maintain situational awareness (SA) depends on situational requirements specified in task analyses and on operators' understanding of the task (Hauland, 2008). The degree of predictability that the traffic situation permits will influence the effort in building a reliable representation of the situation, the so-called "picture" (Nunes & Mogford, 2003). Traffic geometry can influence these representations by increasing visual search demands. Moreover, in future scenarios some framing support elements (e.g., routes) will no longer appear, so evaluation of human factors issues such as vigilance, SA, and workload become of primary importance (Rantanen & Nunes, 2005). Understanding current possible measures and trying to figure out new reliable measures are part of future challenges (Langan-Fox, Sankey, & Canty, 2009).

1.2. TASKLOAD AND WORKLOAD RELATIONSHIP

Many methods have been proposed to address workload dynamics in operators' activities, focusing on different data sources. Performance measures have the benefit of overcoming the most common problems encountered when adopting subjective techniques, such as response biases or off-line administration. Physiological measures (e.g., heart rate variability and ocular metrics) may be more suitable than subjective scales for measuring workload in ATC dynamic environments because they are able to provide more continuous measures of workload (Brookings, Wilson, & Swain, 1996). Performance measures imply the intrinsic difficulty of score assignation. Moreover, in real work context, operators will invest cognitive resources to keep their performance safe and constant. Nowadays, technology allows portable and unobtrusive tools to be applied in operational contexts without jeopardizing normal job activities. In this sense stands our effort in investigating saccadic dynamics as a possible candidate for online workload assessment. Of course, both for validation purposes and for the sake of completeness, a multidimensional approach to workload evaluation proves to be the most fruitful solution. Trends in one dimension, if confirmed by other data sources, become more reliable, by relying on "external" criteria. A multidimensional approach (including the use of multiple psychophysiological measures) is common practice in modeling studies and classification processes by artificial intelligence solutions (Wilson & Russell, 2003). Several studies have manipulated task demands (and difficulty) with the aim of making the participants experience different levels of workload. In these cases, combining psychophysiological data with performance and subjective ratings will provide important confirmation of workload states and therefore of the construct validity of the study. Previous studies have used eye movements and blinks (Veltman & Gaillard, 1996) as measures of fatigue or of visual channel overload. In this study, we explore saccadic movement dynamics as possible workload measures, by observing whether variations in these parameters occur in parallel with changes in geometry and/or complexity. Saccades are ballistic movements of the eyes, carried out to bring situation-relevant information to the fovea (the part of the retina with the highest concentration of cone cells and with the best visual acuity). Cortical and subcortical neural structures are responsible for amplitude specification, from which saccadic peak velocity (PV) and saccadic duration (SD) are derived. By means of saccadic movements, attentional requirements are accomplished. Indeed, if there are effects of different task factors on saccadic velocity, they can be interpreted from an increased attentional demands perspective. In this sense, ocular movements can express perceived workload levels and also inform on the visual strategy that operators adopt.

The relationship between taskload (the objective demands needed to carry out a task evidenced from task analysis methods) and workload (the functional state resulting after resources are invested to accomplish the task) is matter of debate (Loft, Sanderson, Neal, & Mooij, 2007). Nevertheless, it is quite intuitive and logical to think that workload cannot depend only on task demands, but also on the subjects' capacity to address and satisfy task requirements. Cognitive strategies adopted for information processing as well as for optimizing actions are therefore important factors that make a linear relationship between taskload and workload quite improbable. Sperandio (1971) underlined the importance of considering operators' strategies when modeling and looking for workload assessment. Top-down influences related to familiarity, experience, and current objectives are key elements in successful strategy implementation and shift, and they cannot be ignored for reliable workload assessment. The present study on eye movements focuses on specific taskload situations, and aims in the future at discovering workable information on cognitive strategies, at least those related to relevant information search in ATC tasks. From this perspective, neuroergonomic measures are more promising than performance ones in mirroring taskload situations. Indeed, in all resource-limited tasks (Norman & Bobrow, 1975), performance could result satisfactory even if task demands and consequent workload augment.

In summary, we agree that the traffic picture accounts only partially for the workload levels generated in an operational situation because strategies also play a key role in resource management and in task demand optimization. Many studies inspired by open loop models (models that almost totally explain workload as a function of task demands, such as traffic factors, density metrics, and geometry) often tend to underestimate resource management processes and feedback that participants use to analyze traffic situations and to accomplish their objectives (Loft et al., 2007). Without denying the causal power of cognitive strategies on workload, however, we focus once more on task demands to observe whether their manipulations are captured by saccadic dynamics. These neuroergonomic measures could be fruitful for modeling purposes and/or adaptive technology prototyping.

1.3. GEOMETRY: COMMENTS ON CONFLICT DETECTION TASK ANALYSIS

Conflict detection is one subtask of ATC activities. ATCOs check for possible lateral, vertical, and longitudinal separation violations by estimating the trajectories of pairs of aircraft using symbols and graphical representations of flight path, together with flight information in data blocks (e.g., altitude and speed). Indeed, conflict detection makes intense use of the traffic display. Visual information acquisition is therefore the basic cognitive process that allows high-level cognitive elaboration and decision making. Spatial organization of information (i.e., traffic geometry and information location) affects the ATCO's ocular movements. Conflict geometry refers to several characteristics of the potential conflicting aircraft pair, including angle of convergence (conflict angle) and time before a conflict occurrence (conflict time). Increases in conflict angle have been shown to increase the time required for trajectory extrapolation judgments (Remington et al., 2000). Time to conflict (TTC) has not been systematically investigated, but a longer TTC will generally produce a longer detection time. The convergence angle between aircraft also influences the accuracy of trajectory extrapolation. In general, increased conflict angles reduce conflict detection accuracy. Anyhow, the influence of conflict angle on detection ability is also dependent on the concurrent speed values (as well as conflict times and traffic load; Remington et al., 2000). In general, if the speed difference is small, the comparison will be based on the distance to convergence point. Conversely, if speed difference is substantial, times to the convergence point will help in checking for possible conflicts. It has often been postulated that the physical distance between conflicting aircraft symbols on the display is the mediating variable in conflict detection ability (in terms of accuracy and rapidity). Small conflict times and angles would place aircraft closer to one another on the display (Remington et al., 2000). Nevertheless, more recent studies have questioned the influence of distance between aircraft on detection ability, for instance, by building scenarios in which traffic load is increased together with the distance between aircraft (Ahlstrom & Friedman-Berg, 2006). In conflict detection tasks, an information-based hierarchical strategy for aircraft comparisons has been proven (Rantanen & Nunes, 2005), demonstrating that altitude is the primary source for conflict discrimination. Aircraft that are sufficiently separated vertically will never enter into conflict and do not need any other comparison. Conversely, in the case of same altitudes, speed evaluation and trajectory computations are needed. In our experiment, altitude was not manipulated, and no information was provided about it, so it could be considered constant. Participants had to compare aircraft on the basis of distance and speed values alone, such as ATCOs may do after detecting two aircraft flying at same altitude. Such a decision is similar to those made in conflict detection situations, because the same steps are involved: localization and comparison of aircraft position, observation of speed values, and extrapolation of future position. An algorithm for the latter was given during introductory paper-and-pencil tasks (see *Procedure* section). Of two aircraft, the one that has to fly twice the distance must have a speed value bigger than twice the other to arrive at the convergence point first. Vice versa, aircraft that must cover half the distance compared to the other cannot arrive first if its speed value is smaller than half of the other.

1.4. MOTIVATION OF THE STUDY

Because ocular movements are intimately related to attentional demands and represent the primary process of obtaining relevant information for situation representation, specific display and traffic geometries could determine visual dynamics and, to a certain extent, strategy implementation. As described by Hauland (2008), operators must rely on visually represented information from the interface. Therefore, because of the eye–mind relationship, it should be possible to use the direction of the eyes as a measure of information acquisition, considering accessed information as part of cognitive processing. Information must be accessed through the fovea. When the scenario is demanding, there should be an increased probability that dwell times (the time spent looking near each target) represent the operator's attention (Wu, Kwon, & Kowler, 2010), and saccades his/her neuromotor primary process to accomplish these demands.

We start our exploration by considering simple scenarios and tasks in which strategy shifts are improbable. According to our methodological choices (screenshots of traffic situations with only two aircrafts), the unit of analysis was saccades in different geometry condition trials, measuring PV and SD controlled for amplitude. Indeed, we assume that observed regions of visual field represent cognitive meaning for accomplishing task objectives. We raise the question of whether there are observable differences in ocular movement data due to taskload manipulations under these simple conditions of "constant" cognitive strategy. The existence of these differences would support the validity of ocular data as an online workload measure. Other studies have used ocular data for online assessments; a review of scientific literature that applied ocular data analysis for workload assessment is beyond the scope of the present work. Examples of the measures used include blink rate and duration (Brookings, Wilson, & Swain, 1996; Veltman & Gaillard, 1998), pupil diameter (PD), saccadic extent, fixation frequency, and dwell time (Ahlstrom & Friedman-Berg, 2006). In general, the data show declines in blink duration and rate, and increases in PD as a function of increasing workload. To the best of our knowledge, few studies have focused on saccadic PV variations and their relation to task demands and workload (Di Stasi et al., 2009; Di Stasi, Marchitto, Antolí, Rodriguez, & Cañas, 2010a; Di Stasi, Marchitto, Antolí, Baccino, & Cañas, 2010b; Di Stasi et al., 2010c; Di Stasi, Antolí, & Cañas, 2011; Di Stasi et al., 2011).

In this study, we investigate the influence of traffic geometry (in terms of angle and distance to convergence point) while comparing two aircraft in different locations in the display. Distance was manipulated at two levels. In half of the cases, aircraft had the same distance, and thus a simple speed comparison was sufficient for making the decision. In the other half of the cases, participants had to calculate the TTC to decide which aircraft would have arrived at the convergence point first, because aircraft had different distances to it. These trials were more difficult because a further calculation was needed before deciding which aircraft would arrive at the convergence point first.

On the basis of results cited earlier in text, we expected geometry manipulations to create different workload levels and observable differences in performance as well as in saccadic dynamics. More precisely, we had foreseen decisions on time to convergence point to be quicker with the same distance trials, as a simple speed comparison would have been sufficient. Conversely, for aircraft at different distances, subjects had to half one speed value or double the other before comparing them for the final decision. In relation to the angle of convergence, we expected increasing angles to raise difficulty, in terms of speed and accuracy.

In summary, higher response times (RTs) were expected from different distance trials and for increasing angles. We also assumed that by manipulating angle and distances in a simplified version of conflict detection, geometry would have affected ocular movements. In particular, on the basis of previous results encountered in relation to PV (Di Stasi et al., 2010b), we expected a slowdown of maximum velocity in presence of increasing geometrical complexity (different distance and increasing angles of convergence). In our experiments (see the *Results* section), we checked for the potential influence of distance between aircraft in ocular and performance results.

2. METHODS

2.1. TASK AND STIMULI

Because the subject pool consisted of college students, a simplified task was created that contained realistic situations and data, yet did not require years of expertise to be carried out. It consisted of a decision about which of two aircraft (per trial) would arrive first to the convergence point. Decisions should have been made by means of a PC mouse after processing and comparing speed and distance to the convergence point. In all trials, aircraft were an Airbus and a Boeing, and they were associated with the left button (marked as A) and the right button (marked as B), respectively.

The stimuli used were screenshots of initial scenarios built using an ATC simulator, namely ATC-lab^{Advanced} (Fothergill, Loft, & Neal, 2009). This simulator presents a high level of realism and control; it is freely available on the Internet and allows a great flexibility in defining and programming scenarios because of its XML open code. In this manner, it is possible to design ad hoc scenarios for isolated ATC tasks.

The airspace was the same in all the trials, and it consisted of a central circle that subtended approximately 3.5° of visual angle from a viewing distance of approximately 60 cm. An aircraft was represented by a small circle placed on one of the four trajectories shown on the screenshots (vertical, horizontal, and oblique), with its text box containing the flight data. All trajectories passed through the center of the screen, which always coincided with the aircrafts' convergence point. An example of scenario is presented in Figure 1.

In total, eight different conditions resulted from the combination of two distance values and four angle levels (see *Experimental Design* section). For every condition, we built 8 screenshots, in order to cover the 4 airspace quadrants and 2 possible answers range. In this way, a total of 64 screenshots was built. Participants were trained on the meaning of the flight data in the text boxes before recording sessions. A particular stress was placed on the speed value to be sure that they would be able to locate this task-relevant information.

2.2. PARTICIPANTS

A total of 18 students (mean age = 21.4 years, standard deviation = 2.58, three males) of the Psychology Faculty at Granada University took part in the study for course credits. All subjects had normal or corrected-to-normal vision, and none of them had ATC experience. They signed a consent form that informed them of the risks of the study and the treatment of personal data.

2.3. APPARATUS

Eye movements were sampled monocularly at 500 Hz using an Eyelink II head-mounted eye tracking system (SR Research, Ontario, Canada). A nine-point calibration process and eye-position validation was carried out before each recording session. Saccade detection required a deflection in eye position greater than 0.1°, with a minimum velocity of 30°/s, and a minimum acceleration of $8000°/s^2$, maintained for at least 4 ms. Saccades around blinks, as well as those of less than 10 ms, were not considered in the analysis.

2.4. PROCEDURE

Before starting the recording of ocular data, participants went through a pretest phase, which consisted of three simple paper problems about two cars traveling toward the same point at different speeds (and different distances in half of the cases). Participants had to adopt the same logic for answering these questions that they would have to use for decisions on aircraft in the subsequent experimental task. This brief training phase was needed to be sure that subjects understood how to perform the task. If distances were the same, a speed match was sufficient. Conversely, if distances to convergence point were different, a speed transformation was needed first. After this phase, an explanation about how to solve the paper problems and the subsequent task on aircraft was provided, even if participants (as it was the case) correctly resolved the former ones. Successively, calibration and eye-data collection started. Altogether, 128 trials were presented (64 screenshots presented twice), for a total registration time of approximately 50 minutes. A chin rest was used to limit head movements.

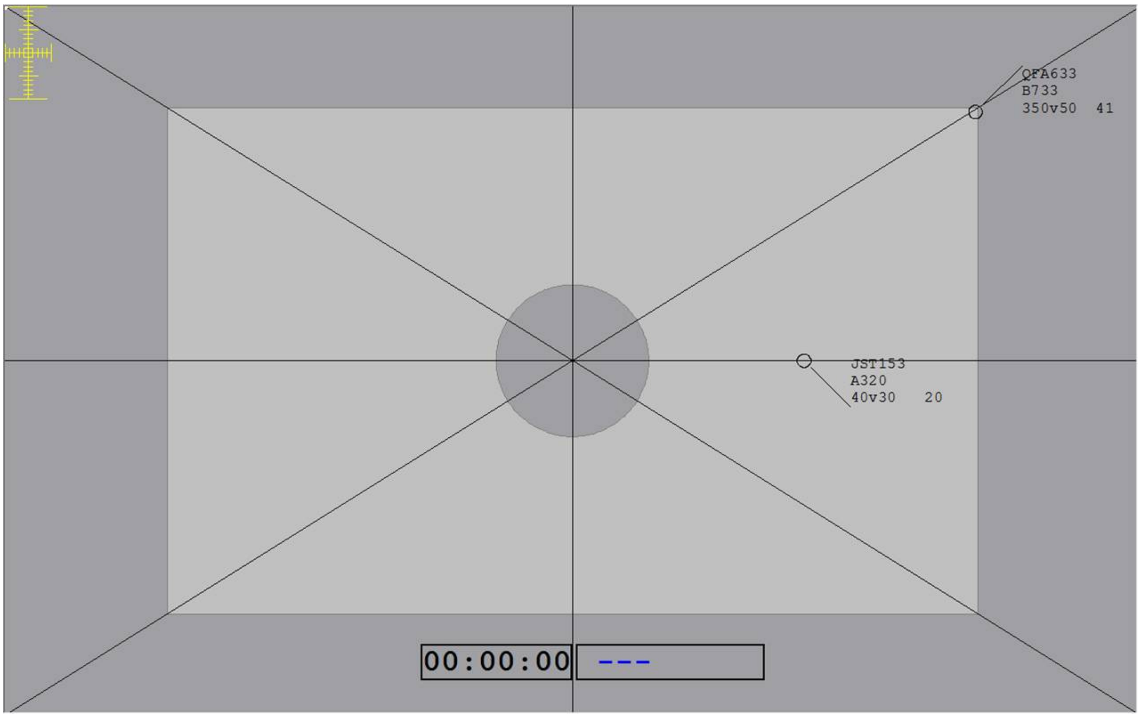

Figure 1. An example of the stimuli used in the study. In this case, aircrafts have different distances from the central convergence point, with a 30° angle of convergence.

2.5. EXPERIMENTAL DESIGN

The two main factors were the aircraft's convergence angle and its distance to the convergence point. The convergence angle was manipulated at four levels; special cases (e.g. 0° or 180°) were intentionally not included. Therefore, angles of 30°, 60°, 90°, and 120° were used. We assumed that increasing angles would have increased task difficulty, at attentional and perceptual level. Distance is a spatial expression of what in ATC is usually expressed in time units and indicated as TTC. We manipulated distance at two levels: same and different (i.e., one aircraft at double the distance to convergence point compared to the other). The two factors were manipulated within subjects, originating a fully factorial 2 × 4 repeated-measures experimental design. Saccadic amplitude must be considered as a control variable when studying eye movements. The velocity and duration of a saccade depends on the amplitude required, which is computed at the neural level before the start of each saccade. For these reasons, bins (amplitude intervals) were used to group and analyze the saccades. At a behavioral-performance level, the dependent variable was RT (i.e., the time needed to push the mouse button). Incorrect answers were excluded from the RT analysis. This is standard practice in RT studies. At a psychophysiological level, ocular data were analyzed using PV and SD median values.

3. RESULTS

3.1. RT

The means of the median RTs from correct trials for each participant in all eight experimental conditions were entered into a repeated-measures analysis of variance (ANOVA) with angle and distance as factors, manipulated at four (30°, 60°, 90°, 120°) and two levels (same and different), respectively.

The main effect of the angle factor was significant [$F(3, 51) = 4.51$, $p = .007$, $\eta_p^2 = .21$] and so was the distance factor [$F(1, 17) = 7.72$, $p = .013$, $\eta_p^2 = .31$]. Their interaction was not significant. Indeed, when aircraft distance was different, RTs were significantly higher, as expected, because different distance trials required a further mental operation (i.e., a speed transformation) compared to same-distance trials. For the angle factor, there was a positive trend with increasing RTs for increasing angle levels. This result is in accord with the complexity operationalization explained earlier in text and is probably due to the fact that increasing angles of convergence require longer exploration times. Nevertheless, post hoc comparisons (after Bonferroni correction) revealed a significant difference only between the 30° – 90° and 30° – 120° levels. The main effect was therefore due to 30° angles, which were explored more quickly than the others.

In summary, results from the RTs substantially confirmed our hypotheses made on the environmental complexity in terms of convergence angle and distance to convergence point.

3.2. RAPIDITY–ACCURACY TRADE-OFF EVALUATION

To evaluate any possible trade-off that could have taken place in the decision task, we checked whether RTs and error rates were correlated. There was no significant correlation between error rates and RTs in any experimental condition. Thus, no trade-off took place in the decision task. Participants did not commit fewer errors for increasing RTs. We will not attempt to interpret the behavioral results more, except to note that there is no evidence of a speed–accuracy trade-off. In summary, differences found in the behavioral data seem due to our complexity manipulation; therefore, they work as objective criteria for the subsequent ocular data analysis.

3.3. OCULAR MOVEMENTS

3.3.1. SACCADIC PV

Our main focus was to check for significant differences in PV as a function of experimental manipulation carried out on environmental complexity. Ocular data were appropriately filtered using the saccade definition criteria (see *Apparatus* section). Besides PV, we considered SD in the analysis. Separate ANOVAs were conducted for each one, following the same 2 × 4 experimental design of performance data. We present all the statistical results in these subsections.

When including saccadic velocity dynamics in experimental analysis, it is necessary to control for saccadic amplitude, as the so-called main sequence establishes (Bahill, Clark, & Stark, 1975); that is, PV and SD increase with amplitude in a relation that is almost linear for small amplitudes. For this reason, all saccades were grouped in bins (amplitude intervals) included as a further factor in the PV analysis. The bins schema and distribution across experimental conditions are shown in Figure 2.

We excluded saccades with amplitude smaller than 1° from our analysis and built nine bins in total. Only the first seven were considered for analysis, however, because saccades in bins 8 and 9 were present only in the 120° angle level. Therefore, a 4 × 2 × 7 fully factorial design was followed with angle, distance, and bin (saccadic amplitude) as repeated factors.

There was an obvious effect of bin [$F(6, 102) = 1128.3$, $p < .001$, $\eta_p^2 = .99$], together with a significant main effect of angle [$F(3, 51) = 8.3$, $p < .001$, $\eta_p^2 = .34$]. The distance factor was marginally significant [$F(1, 17) = 4.4$, $p = .051$, $\eta_p^2 = .21$]. Post hoc comparisons for angle revealed significant differences between 30° and all other levels (all $p < .01$). The two first-order interactions bin*angle [$F(18, 306)$

= 5.5, $p < .001$, $\eta_p^2 = .24$] and distance*angle [$F(3, 51) = 4.5$, $p = .007$, $\eta_p^2 = .21$] were also significant, as well as the three-way interaction [$F(18, 306) = 1.7$, $p = .03$, $\eta_p^2 = .09$]. These last results are presented in Figure 3. When analyzing the higher order interaction, significant differences were found only in bin 7. The angle factor generated two different patterns in the two levels of distance. When distance was different, there were significant differences between the 30° angle and the remaining levels (all $p < .001$). Conversely, when distance was equal, there were significant differences only for the 30° – 120° pair ($p < .001$). In relation to large saccades (>15°), the slow-down of saccades due to the angle factor was even stronger with different distances to convergence point compared to equal.

3.3.2. DISTANCE BETWEEN AIRCRAFT

One objection that could be raised against the effective influence of geometry (manipulated by angle and distance variations) on saccades is that observed differences could be due to the distance between aircraft, a variable that changes when angles between aircraft and distances from the center change. The encountered effects could therefore be simply mediated by this variable. It is quite probable that, by augmenting angles or distances to convergence point, distances between aircraft will increase. In any case, scenario exploration is surely carried out by performing saccades from one aircraft to the other and by observing aircraft distances to the center. We carried out a simple analysis to check for possible mediating influences of distance between aircraft on the effect found on PV and on RTs. Figure 4 orders conditions in the horizontal axis in relation to distance between aircrafts values (a). Corresponding PV and RTs values are then plot (b and c) maintaining the same order of conditions. It can be observed how, for instance, two relatively different conditions on the basis of geometry manipulation adopted, like SD-90° and DD-60°, present an almost equal distance between aircraft.

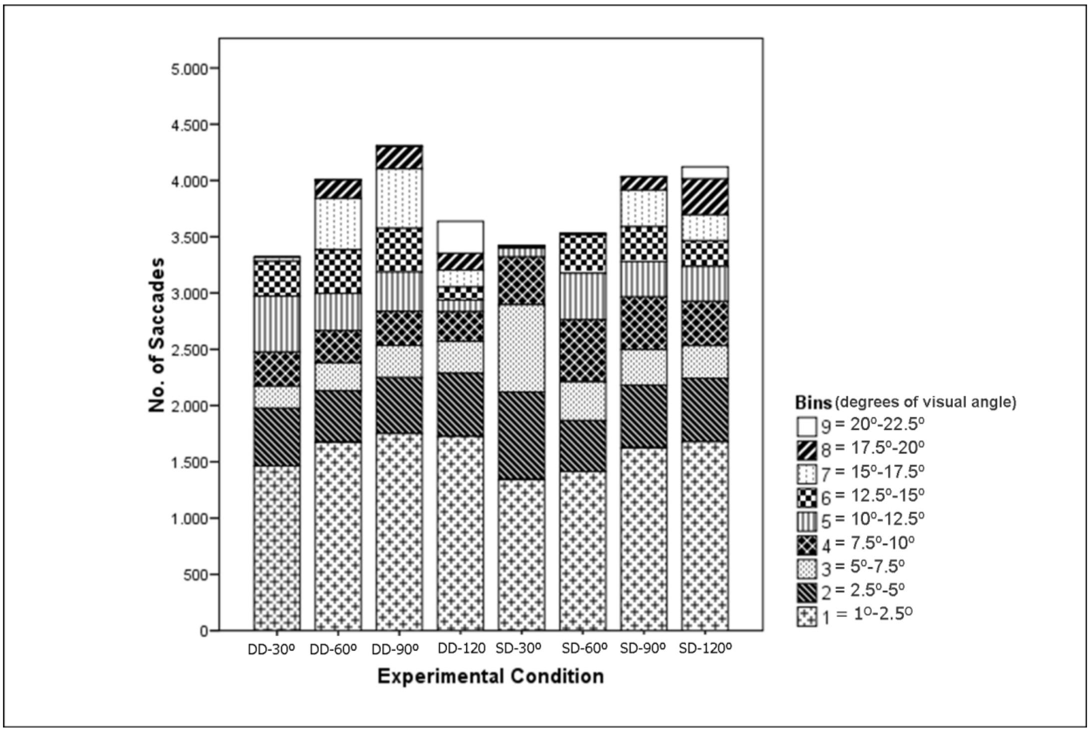

Figure 2. Saccadic amplitude interval subdivision (right part) and frequency plot as a function of the experimental condition (left part). DD means different distance, while SD means same distance.

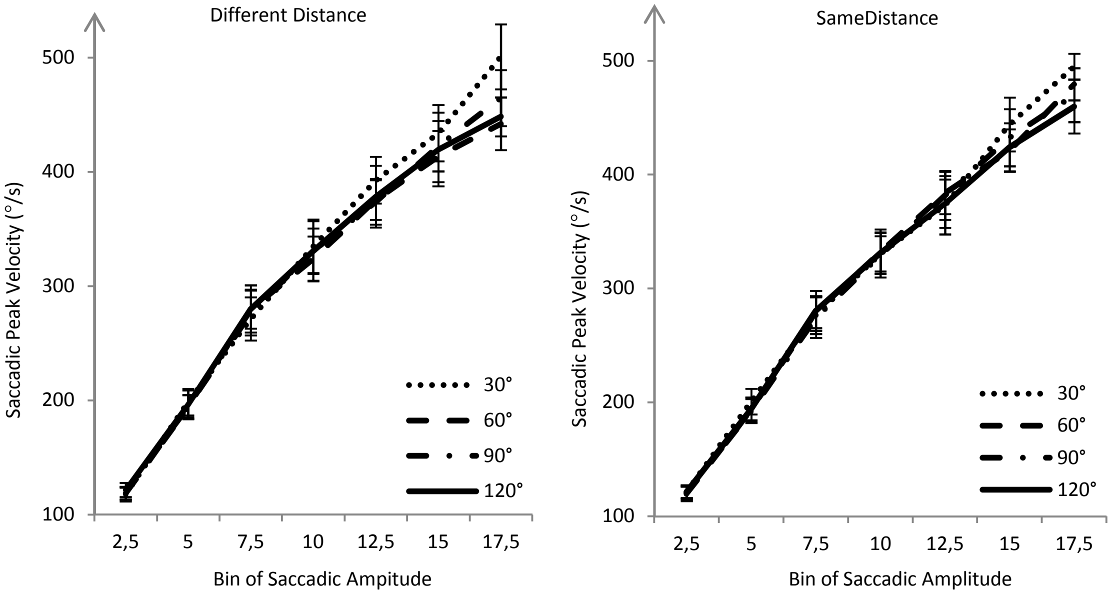

Figure 3. Three-way significant interaction for PV. Vertical bars denote 95% confidence intervals.

When conditions are arranged on the basis of the distance between aircraft (a), PV values do not clearly have the same linear trend (b). If this geometrical factor plays a central role, lower values of PV with increasing distance between aircraft would not be expected. Because the higher the distance between aircraft, the higher the proportion of large saccades in a trial, on the basis of the main sequence rule, PV means should increase for increasing distance between aircraft. The situation found in the PV data, however, seems different in relation to RTs. In this case, there seems to be a positive trend (c). Actually, in the experimental stimuli, larger angles tend to place aircraft farther from each other. The fact that this association is observable in RTs confirms how display factors influence behavioral data. We think however that, concerning PV values, our geometry manipulation remains valid.

This simple analysis suggests that the distance between aircraft cannot be considered the real mediating variable for the differences found in ocular data. Actually, the cognitive demands in this task did not require simple saccades toward a target point, but the implementation of more complex visual patterns, in which the relevant fixation points are not only those indicating aircraft position. Moreover, display exploration was obviously free, without any restriction on visual behavior.

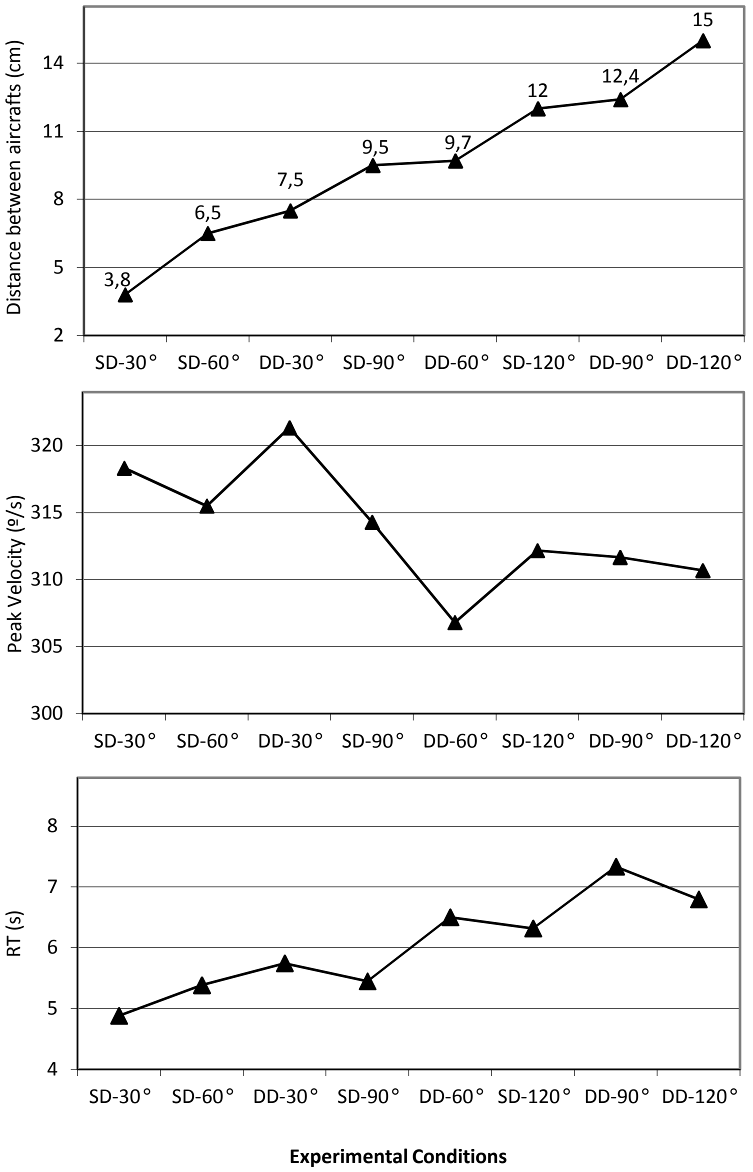

Figure 4. On the X axis, experimental conditions are ranked on the basis of the distance between the aircrafts (a). In (b), the PV is shown for each condition on the Y axis. In (c) RTs are shown on the Y axis.

3.3.3. SD

As for PV, the first seven bins were considered for repeated-measures ANOVA, following a 7 × 2 × 4 design with bin, distance, and angle as repeated factors. In the global ANOVA, apart from the main significant effect of bin and angle [$F(3, 51) = 4.97$, $p = .004$, $\eta_p^2 = .22$] factors, angle*bin was the only interaction that showed significance [$F(18, 306) = 2.6$, $p = .001$, $\eta_p^2 = .13$]. This interaction confirms the effect of the angle factor only on some amplitude intervals. Simple effects analysis (angle effects in every bin level) revealed significant differences only in bin 2 and bin 7. As for bin 2, the results are not immediately clear, because angles of 30° saccades had significantly longer durations than 90° and 120° ones. We are not able to explain this difference, which may be a random effect. The pattern found in bin 7 is more interesting [$F(3, 51) = 8.32$, $p < .001$, $\eta_p^2 = .06$], with higher values of SD for increasing angles. This last trend is opposite to that found in the PV values for the same amplitude interval and further confirms our hypotheses.

4. DISCUSSION

This experiment was designed to explore saccadic dynamics with increasing task requirements induced by geometry manipulations. If geometry is not only a complexity factor at an environmental level but also at an attentional one, eye movements could show important information on situational mental function. On the basis of previous results that showed main sequence sensitivity to task difficulty variations (Di Stasi et al., 2010b; 2010c), we hypothesized a decrease in PV and an increase in SD with increased attentional requirements induced by geometric manipulations. This basic psychophysiological process would account for situational workload variations during visual searches for task-relevant information on a display. The experiment represents a first attempt to model a conflict detection task in ATC while simultaneously observing eye movement behavior.

Our attempts to manipulate task difficulty by varying scenario geometry were highly successful. RTs are an external justification to the ocular movement analysis in taskload variation studies; RTs changed consistently with expectations. We can assume therefore that different attentional demands were experienced as a consequence of our geometric manipulations. The speed and type of aircraft were in the form of alphanumeric text, and their processing required eye fixations along with spatial attention allocation to specific display locations. Participants had to fixate and attend to the fine detail of the data block and perform large saccades to process visual information about position and distance to convergence point. Visual behavior in our task requires large-reaching saccades and small amplitude saccades around task-relevant text, to allow textual information processing. Because saccadic targets in the display often involve textual information, target positional error of saccades may be reduced by slower saccades (Harris & Wolpert, 2006). It is reasonable to think that visual strategy and saccadic trade-offs would adapt to the specific task requirements (Gancarz & Grossberg, 1999).

Concerning behavioral data, the correlation analysis between RTs and errors did not give any significant value for all the conditions. This result means that RTs and accuracy do not depend on each other and that there was no trade-off between the two variables.

Results on ocular data show that only saccades of certain amplitude (>15°) change with geometric influences. For saccades of small amplitudes, there is no effect. Actually, the relationship between amplitude, PV, and SD for small amplitude saccades is difficult to be modified by environmental factors (App & Debus, 1998).

In summary, geometry manipulations showed significant effects only for large-amplitude saccades. These effects were more evident on PV than SD. Saccades of more than 15° (bin 7) had large PV and small SD when the angle between aircraft was small and had smaller PV and a larger SD when the angle increased. More precisely, this slowdown effect on large saccades did not involve saccades in the 30° angle condition. Probably, in this last situation, geometric influences were not at work. Moreover, the effects of angle on PV values were different depending on the distance level.

Amplitude controls have proven fundamental for figuring out how some effects involve only saccades of a certain size. Some works have shown how saccades of different amplitude are differently influenced by environmental characteristics (Tatler, Baddeley, & Vincent, 2006): Small saccades, for instance, are influenced more by high frequency information (contrast, luminance, and edge information) than are large saccades (>8°). This finding seems to be based on overlearned mechanisms rather than on salience-guided ones, as they are less influenced by spatial features of destination. In this sense, large saccades can be considered more as the realization of an attentional intention (after a loose processing of destination location), rather than an exogenous bottom-up process. Saccades have been proposed (Harris & Wolpert, 2006) as the result of a speed–accuracy trade-off, as an optimal strategy that balances velocity (and duration) of the eye movement and endpoint variability of the saccade (positional error).

Maybe geometric factors have an influence on this neural compromise and large saccades are slower to benefit saccadic accuracy. If it was possible to get online measurement of such oculomotor parameters during the operational session, there would be the possibility of monitoring speed–accuracy trade-off variations throughout a session. If, as a consequence of increased visual workload, saccades become slower to increase endpoint accuracy, then it would prove fruitful for online workload monitoring an index that expresses these changes.

Frost and Poppel (1976) suggested that saccades to targets farther than 10°–15° away are executed using a different neural pathway than that for smaller saccades. The fact that we find significant reductions in velocity due to angle and distance only for large saccades (the saccades used to move gaze from an object to another more distant) suggests that geometry may influence “reaching” saccades rather than “fixation” saccades (the little saccades around the same object). Large saccades are the realization of endogenous saccadic goal specifications (Wu et al., 2010). This basic process is probably affected by the geometric features that account for traffic complexity.

In our opinion, attentional requirements represent a cost that contributes to experienced situational workload, a construct that is surely impossible to be captured by a subjective tool. To focus consciously on these environmental factors off-line remains considerably difficult. The results found on PV and SD seem to indicate that, in such simple scenarios, the angle of convergence can reduce saccadic PV. To a certain extent, the distance factor generated the same pattern, with lower PV when distance was different.

We are aware that the distance manipulations carried out in this study are quite unusual. In real control settings, distance is directly checkable by using support tools and vector aids. It is not so obvious that a situation with two aircraft at different distances from the convergence point is intended as more difficult compared to when the aircraft are at the same distance. Actually, the last case is less frequent. When dealing with a sample of non-experts, it could be the case that different-distance cases are immediately recognizable, whereas for same-distance trials further graphical checks are implemented to verify that the distance is actually the same. This could also be the reason that explains why there were more errors in same-distance trials than in different distance ones.

There are some drawbacks in this experimental work. A series of simplifications were necessary that may partly mask, or at least limit, the effects found in the ocular data. Surely, a series of independent trials of short duration is uncharacteristic of real ATC work. By using screenshots, all the information is crystallized in an image, and traffic evolution is nonexistent. Another caveat of this study is that we did not recruit actual ATCOs. We dealt with untrained university students: Consequently, particular importance was given on paper introductory tasks and pretest briefing. Nevertheless, what distinguishes expert controllers with respect to the general population may be more related to planning, strategy implementation, and procedural and declarative knowledge than basic perceptual and cognitive processing abilities.

As a result, caution is recommended when interpreting these results. It is important to note how the study of ocular movements could be useful to understand visual search behavior and relative workload as a function of traffic situations. Modeling ATCOs behavior could benefit from studies on geometric influences on ocular movements.

Other geometric factors that influence attentional requirements and display complexity have to be included in future studies to more reliably model conflict detection tasks. Traffic load, as a factor that can increase attentional task requirements, may also have an influence on saccadic velocity, by imposing new velocity–accuracy trade-offs on the oculomotor system. Maybe display density could shift neural priority toward accuracy, causing saccades with longer duration and smaller velocity. Another factor to be included in future work is the difference in speeds of the aircraft. For aircraft flying at similar speeds, controllers need only check the distances to the convergence point to find conflicts. Conversely, when speed differential increases, projections of future position have to be based both on speed and distance, i.e. a more difficult situation. Finally, aircraft altitude should be included as a complexity factor. When aircrafts change altitude (in dynamic scenarios), the difficulty of conflict detection is significantly increased. Many studies have shown that altitude is often the first step in hierarchical conflict detection strategy (Rantanen & Nunes, 2005).

5. CONCLUSION

The experiment conducted in this study represents a first attempt to model ATC conflict-detection tasks to observe whether geometric factors related to environmental complexity can influence saccadic dynamics. The results encountered indicate that only large saccades are influenced by geometric manipulation. Saccades of small amplitude do not show differences in speed for different complexities. They have speed characteristics that in general cannot be affected by contextual factors. Conversely, large-amplitude saccades had decreased PV when geometric factors increased environmental complexity.

In future work, other factors will be included and new experimental conditions tested with the aim of better modeling the conflict detection task and individual traffic factor influences on visual search patterns.

We first thought that promising results should have been found in simple situations, before addressing more complex ones in which many factors interact. In the future, units of analysis will have to change in accordance with the tasks investigated and the time frame of reference. Online reliable measures of visual task demands are therefore highly desired. Ocular data represent a direct link to the visual strategies implemented by ATCOs, and their measuring could prove useful in an operational context. This approach may be more useful in understanding ATCO workload than attempting to model the relationship between isolated traffic factors and ATCO speed or accuracy of responding. Moreover, with respect to future research, it will be necessary to correlate the saccadic parameters measured in this task with more objective criteria as, for instance, PD (by implementing all the necessary controls related to this index, as to keep constant environmental lighting). In addition, subjective data collected by means of a rating scale that could be administered online would strengthen the overall study. The major issue when dealing with randomized traffic scenarios in succession is that any subjective measure should be collected throughout the session, more than afterward, to better appreciate situational changes and to get meaningful correlations. Such a multidimensional methodology would provide additional validation of saccadic dynamics parameters, here proposed as indices for situational workload evaluation.

Overall, these results show that a great deal of information concerning operator functional state is contained in an individual's psychophysiological data. Although the causal relationship between psychophysiological measures and cognitive activity is not known, the former can be advantageously applied to solve problems involving human–technology interactions. Computational modeling of ocular movements may be useful (e.g., in classifying functional states). The benefits from different forms of (adaptive) automation depend on classification reliability and the sensitivity of measures. Ocular movement data have to be investigated for improving automation solutions that support cognitive processes, such as information acquisition.

ACKNOWLEDGMENTS

This study was conducted within Proyecto Rayuela, SEJ 2007-63850/PSIC, funded by the Spanish Government.

REFERENCES

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*, 623–636.

App, E., & Debus, G. (1998). Saccadic velocity and activation: Development of a diagnostic tool for assessing energy regulation. *Ergonomics*, *41*(5), 689–697.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *191*, 204–224.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*, 361–377.

Cummings, M. L., & Tsonis, C. G. (2005, September 26–30). *Deconstructing complexity in air traffic control*. Paper presented at the 49th Human Factors and Ergonomics Society Annual Meeting, Orlando, FL.

Di Stasi, L. L., Alvarez, V., Cañas, J. J., Maldonado, A., Catena, A., Antolí, A., et al. (2009). Risk behaviour and mental workload: Multimodal assessment techniques applied to motorbike riding simulation. *Transportation Research, Part F*, *12*, 361–370.

Di Stasi, L. L., Marchitto, M., Antolí, A., Rodriguez, E., & Cañas, J. J. (2010a). *From subjective questionnaires to saccadic peak velocity: A neuroergonomics index for online assessment of mental workload.* In T. Marek, W. Karwowski, & V. Rice (Eds.), Advances in understanding human performance: Neuroergonomics, human factors design, and special populations (pp. 11–20). Boca Raton, FL: CRC Press.

Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010b). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management, 16*, 330–333.

Di Stasi, L. L., Renner, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Cañas, J. J., et al. (2010c). Saccadic peak velocity is sensitive to variations in mental workload while driving in a virtual environment. *Aviation, Space, and Environmental Medicine*, *81*, 413–417.

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2011). Main sequence: An index for detecting mental workload variation in complex tasks. *Applied Ergonomics*, *42*(6), 807-813.

Di Stasi, L. L., Renner, R., Cañas, J. J., Catena, A., Velichkovsky, B. M., & Pannasch, S. (2011). Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data. *Transportation-research: Part C: emerging technologies, 21*(1), 122-133.

Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab$^{\text{Advanced}}$: An air traffic control simulator with realism and control. *Behavior Research Methods*, *41*(1), 118–127.

Frost, D., & Pöppel, E. (1976). Different programming modes of human saccadic eye-movements as a function of stimulus eccentricity—indications of a functional subdivision of visual-field. *Biological Cybernetics*, *23*(1), 39–48.

Gancarz, G., & Grossberg, S. (1999). A neural model of saccadic eye movement control explains task-specific adaptation. *Vision Research*, *39*, 3123–3143.

Harris, C. M., & Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off. *Biological Cybernetics*, *95*, 21–29.

Hauland, G. (2008). Measuring individual and team situation awareness during planning tasks in training of en route air traffic control. *International Journal of Aviation Psychology, 18,* 290–304.

Hilburn, B. (2004). Cognitive complexity in air traffic control - A literature review. *EEC Note,. 4*(04). Brétigny-sur-Orge: Eurocontrol Experimental Research Center.

Hollnagel, E. (2007). Flight decks and free flight: Where are the system boundaries? *Applied Ergonomics*, *38*(4), 409–416.

Langan-Fox, J., Sankey, M. J., & Canty, J. M. (2009). Human factors measurement for future air traffic control systems. *Human Factors*, *51*, 595–637.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modelling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors*, *49*, 376–399.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control vs. passive monitoring. *Human Factors*, *43*, 529–528.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*, 44–64.

Nunes, A., & Mogford, R. H. (2003, October, 13–17). *Identifying controller strategies that support the "picture"*. Paper presented at the 47th Annual Meeting of the Human Factors and Ergonomics Society, Denver, CO.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *International Journal of Aviation Psychology*, *15*(4), 339–362.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human Factors*, *42*, 349–366.

Sperandio, J. C. (1971). Variation of operator's strategies and regulating effects on workload. *Ergonomics*, *14*, 571–577.

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, *46*, 1857–1862.

Veltman, J. A., & Gaillard, A. W. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*, 323–342.

Veltman, J. A., & Gaillard, A. W. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*, 656–669.

Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors*, *45*, 381–389.

Wu, C. C., Kwon, O. S., & Kowler, E. (2010). Fitt's law and speed/accuracy trade-offs during sequences of saccades: Implications for strategies of saccadic planning. *Vision Research*, *50*, 2142–2157.

APPENDIX A: DESCRIPTIVE STATISTICS OF MEASURED VARIABLES

	Different distance				Same distance			
	30°	60°	90°	120°	30°	60°	90°	120°
M	5.8	6.5	7.3	6.8	4.9	5.4	5.4	6.3
SE	0.8	0.6	0.8	0.7	0.4	0.5	0.5	0.6

Table 1: Means (*M*) and standard errors (*SE*) of Response Times (RT) in the eight experimental conditions.

BIN	Different distance				Same distance			
	30°	60°	90°	120°	30°	60°	90°	120°
1	117.9	118.6	118.7	121.6	121.3	119.7	120.4	121.4
	3	*2.4*	*2.8*	*2.9*	*2.8*	*3*	*2.5*	*2.6*
2	198.4	194.9	194.3	196.2	200.6	193.5	192.3	194.2
	5.5	*4.6*	*4.9*	*6*	*5.3*	*5.1*	*5*	*4.8*
3	271.3	279.4	276.9	280.1	276.7	279	274.3	280.3
	9	*7.9*	*9.5*	*9.8*	*7.9*	*6.6*	*8.4*	*8.3*
4	334.5	323.8	331	330.8	330.4	332.2	329.2	331.1
	11.3	*9.3*	*9.2*	*12.4*	*7.3*	*9.3*	*9.3*	*8.5*
5	392.7	375.5	373.8	378.3	373.1	383.8	381.7	374.4
	9.7	*8.2*	*9.5*	*12.8*	*12.2*	*8.8*	*10.2*	*10*
6	433.9	413.5	422.5	419.6	443.9	421	432.4	424
	11.7	*10.6*	*10.5*	*15.3*	*11.2*	*8.8*	*11.9*	*9.9*
7	500.7	442	464.5	448.2	494.8	479.3	469.8	459.8
	13.5	*10.9*	*11.6*	*8.1*	*5.4*	*6.7*	*11.2*	*11.2*

Table 2: Means and standard errors (*italics*) of saccadic Peak Velocity (PV) in the eight experimental conditions, controlled for saccadic amplitude (Bin).

BIN	Different distance				Same distance			
	30°	60°	90°	120°	30°	60°	90°	120°
1	21,3	21,6	21,6	22,2	22,1	21,8	21,8	22
	0,6	*0,7*	*0,7*	*0,7*	*0,7*	*0,7*	*0,7*	*0,6*
2	32,2	31,3	31	30,8	33,2	31,4	31	31,2
	1	*0,7*	*0,7*	*0,7*	*0,8*	*0,7*	*0,8*	*0,9*
3	41,6	42	42,8	42	42,1	42,3	42,4	42,3
	0,8	*0,7*	*0,7*	*0,8*	*0,7*	*0,8*	*0,8*	*0,7*
4	48,9	47,1	47,3	47,4	46,4	47,9	48,4	48,2
	1	*0,8*	*1,1*	*0,9*	*0,7*	*0,8*	*0,8*	*0,8*
5	53,2	51,9	52,9	52,6	53,9	51,6	53,3	52,6
	0,9	*0,8*	*0,8*	*1,2*	*3*	*0,8*	*1,1*	*0,9*
6	57,5	56,1	57,6	59,7	56	55,6	57,7	57,4
	0,9	*0,8*	*1,1*	*3*	*1,2*	*0,7*	*1,2*	*1*
7	60,4	62	62,4	63,1	58	58,5	62,8	62,4
	1,2	*1,2*	*1,2*	*1,4*	*0,5*	*0,6*	*1,1*	*1,3*

Table 3: Means and standard errors (*italics*) of Saccadic Duration (SD) in the eight experimental conditions, controlled for saccadic amplitude (Bin).

CHAPTER 4

OCULAR METRICS AND CONFLICT DETECTION COMPLEXITY

CHAPTER 4

AIR TRAFFIC CONTROL: OCULAR METRICS REFLECT COGNITIVE COMPLEXITY[3]

ABSTRACT

The objective of the study was to evaluate effects of complexity on cognitive workload in a simulated air traffic control conflict detection task by means of eye movements recording. We manipulated two complexity factors, convergence angle and aircrafts minimum distance at closest approach, in a multidimensional workload assessment method based on psychophysiological, performance, and subjective measures. Conflict trials resulted more complex and time-consuming than no conflicts, requiring more frequent fixations and saccades. Moreover, large saccades showed reduced burst power with higher task complexity. A motion-based and a ratio-based strategy were suggested for conflicts and no conflicts on the basis of ocular metrics analysis: aircrafts differential speed and distance to convergence point at trial start were considered determinant for strategy adoption.

Relevance to Industry. Eye metrics measurement for online workload assessment enhances better identification of workload-inducing scenarios and adopted strategy for traffic management. System design, as well as air traffic control operators training programs, might benefit from online workload measurement.

---

[3] Submitted as:
Marchitto, M., Benedetto, S., Baccino, T., & Cañas, J. J. (under review). Air traffic control: ocular metrics reflect cognitive complexity. *International Journal of Industrial Ergonomics.*

1. INTRODUCTION

According to the International Civil Aviation Organization (ICAO) medium-term passenger traffic is expected to increase by 6.3% in 2015 (ICAO, 2012) and air traffic control (ATC) is going to be one of the most affected activities of the whole air transport system. ATC is a service offered by ground-based operators (ATCO), which aims at preventing conflicts between aircrafts, by managing the resulting complexity. It mainly consists of organizing the traffic flow, and providing information and support for pilots. In order to prevent collisions, ATCOs employ traffic separation rules, which constantly ensure the maintaining of a minimum amount of empty space around airplanes. The concept of complexity in the ATC domain has received considerable attention along the years, and a large amount of studies classify complexity at several levels of analysis: environmental, organizational, traffic, and display (e.g. Cummings & Tsonis, 2005; Mogford, Guttman, Morrow, & Kopardekar, 1995). These elements influence controllers' cognitive complexity, i.e. the perceived complexity at the individual level. Many studies have been carried out for identifying factors that affect cognitive complexity in air traffic scenarios (Hilburn, 2004; Histon et al., 2002), and to devise evaluation frameworks to be applied both in simulation and real work contexts (Majumdar & Ochieng, 2007; Pawlak, Brinton, Crouch, & Lancaster, 1996). Other studies have focused on the effects of complexity on ATCO, showing that it affects mental workload, the allocation of mental resources for accomplishing task demand in a safe and efficient manner (Athènes, Averty, Puechmorel, Delahaye, & Collet, 2002; Li, Cho, Hansman, & Palacios, 2010).

According to Leplat (1978) mental workload (also referred to as cognitive workload) is a multidimensional construct, rather than a unidimensional one. In this sense the multidimensional assessment of mental workload by triangulation from physiology, performance, and subjective assessments is a fruitful approach (Parasuraman, Sheridan, & Wickens, 2008; Wierwille & Eggemeier, 1993). In ATC, high performance standards are usually maintained, independently of task complexity. Therefore, the monitoring of controller's performance cannot always convey the real cognitive demand imposed by the task. For this reason the need of online measurement of mental workload is an opportunity to seek. When dealing with safety-critical work contexts, the best option for workload assessment relies on unobtrusive techniques (Langan-Fox, Sankey, & Canty, 2009) and psychophysiology seems to be one of the most promising fields for the online measurement of the operator's state (Wilson & Russell, 2003). The main advantage of psychophysiological responses is that they do not require an overt response by the operator, and can be collected continuously with relatively low obtrusiveness. In this framework, workload has been also investigated by means of real time research techniques such as electroencephalography (Weiland, Roberts, Fine, & Caywood, 2013), and optical brain imaging (Ayaz et al., 2010), but the level of intrusiveness of such techniques still represents a limitation. In contrast, the increased sophistication and accessibility of eye tracking technologies have generated a great deal of interest around eye measures (Ahlstrom & Friedman-Berg, 2006). While in the past they required invasive equipment, unsuitable for most applied settings, nowadays large advances in technology have made the equipment much more portable and capable.

In the framework of conflict detection in ATC, we explored the effects of complexity on mental workload by means of a multidimensional assessment method based on psychophysiological (eye movements), performance (response time and accuracy), and subjective measures. With respect to previous investigations, the main contribution of this study lies on the employment of eye movements for the online assessment of workload during a simulated ATC task.

Conflict detection is one of the main activities performed by ATCOs (Kallus, Van Damme, & Dittman, 1999). It consists of comparing trajectories of converging aircrafts and estimating the probability of a future simultaneous violation of vertical and lateral separation standards, which are commonly set to 1000ft (feet) and 5nm (nautical miles, the unit of distance used in aviation, which correspond to 1852 meters), respectively (Loft, Bolland, Humphreys, & Neal, 2009). Conflict detection is carried out on radar displays and involves several subtasks such as information location, change detection, and short- and long-term predictions in a complex and dynamic environment (Li et al. 2010). To the best of our knowledge, studies in ATC domain employing psychophysiological data for workload assessment are quite limited and those dealing with conflict detection by means of ocular behaviour recording are even sparser (for a review see Langan-Fox et al., 2009). Eye tracking has been used to explore attention allocation in control tasks

(Martin, Cegarra, & Averty, 2011; Lokhande & Reynolds, 2012), task discrimination (Imants & de Greef, 2011), system usability (Jacob & Karn, 2003), and it has been employed as head-free input device (Alonso et al., 2013).

According to the literature dealing with conflict detection, convergence angle (CA) and minimum distance at closest approach (MD) are key factors in determining cognitive complexity. CA is a geometry factor that influences visual information acquisition, the basic cognitive process that enables successive high-level cognitive elaboration and decision making. CA directly influences distance between converging aircrafts. While wider angles increase the separation between aircraft, smaller ones reduce it. For example, two aircrafts flying with the same speed that converge with CA = 135° will keep lateral separation for a longer period with respect to aircrafts converging with CA = 60°. As a result, the eyes must perform wider saccadic movements to transit from one aircraft to the other with wider CA. In this respect, Marchitto, Di Stasi, and Cañas (2012) showed that an increase of CA affects both conflict detection times and ocular movements, reducing the peak velocity of large reaching saccades. According to Remington, Johnston, Ruthruff, Gold, and Romera (2000), trajectory comparison is faster and more accurate for smaller angles, which are usually associated to higher probability of intervention by ATCO (Loft et al., 2009).

MD is a traffic complexity factor that affects the cognitive simulation process, which involves the projection of future aircrafts positions, the estimation of the distance between them, and the comparison of such distance to the separation standards (i.e. vertical and lateral). Controllers can apply different perceptual and cognitive methods (i.e. time- or space-based strategies) for estimating future relative positions of aircrafts on the basis of relevant flight information such as speed, distance to convergence point, heading, and movement observation (Xu & Rantanen, 2003). In conflict detection, MD has been shown to predict response time, probability of intervention (Loft et al., 2009; Stankovic, Raufaste, & Averty, 2008), and subjective ratings of difficulty and complexity (Boag, Neal, Loft, & Halford, 2006).

This paper is organized as follows: information about participants, stimuli, apparatus, and a detailed description of the dependent variables with relative hypotheses is provided in Section 2. Results are presented and discussed in Section

3 with a correlational analysis followed by a factorial ANCOVA with CA and MD as predictors, and response time as covariate. Finally, conclusions are drawn in Section 4, together with practical applications and some perspectives for future work.

2. MATERIALS AND METHODS

2.1 PARTICIPANTS

Twenty-six students (22 women, mean age = 22 years, SD = 2 years) from the University of Granada – Faculty of Psychology, volunteered for course credits after signing an informed consent. They all had normal or corrected-to-normal vision (contact lenses were accepted, but not glasses). None of them had previous ATC experience. An internal committee board approved the study, which was performed in keeping with the Declaration of Helsinki.

2.2 STIMULI AND APPARATUS

The ATC-lab^{Advanced} software (Fothergill, Loft, & Neal, 2009) was employed for building the air traffic scenarios, which consisted of a central point, i.e. convergence point, and two aircrafts with related flight information moving on predefined routes, as represented in Figure 1. Four different types of routes were employed, i.e. vertical, horizontal, and two oblique. Aircrafts' position was updated every 5s. Around the convergence point, a circle with a radius equal to the lateral separation standard for conflict definition (5nm) was always presented. In this experiment the conflict detection task dealt with leveled aircrafts flying at the same altitude, thus only lateral problems were considered (Loft et al. 2009). Stimuli were presented on a CRT monitor (21 inches) with 160Hz refresh capability, vertically.

Eye movements were recorded with a 500 Hz infrared video-based eye tracker (Eyelink II - SR Research, Ontario, Canada). A nine-point calibration process was carried out before each recording session and a drift correction was performed before the beginning of each trial. In order to limit head movements, we used a chin rest during data acquisition.

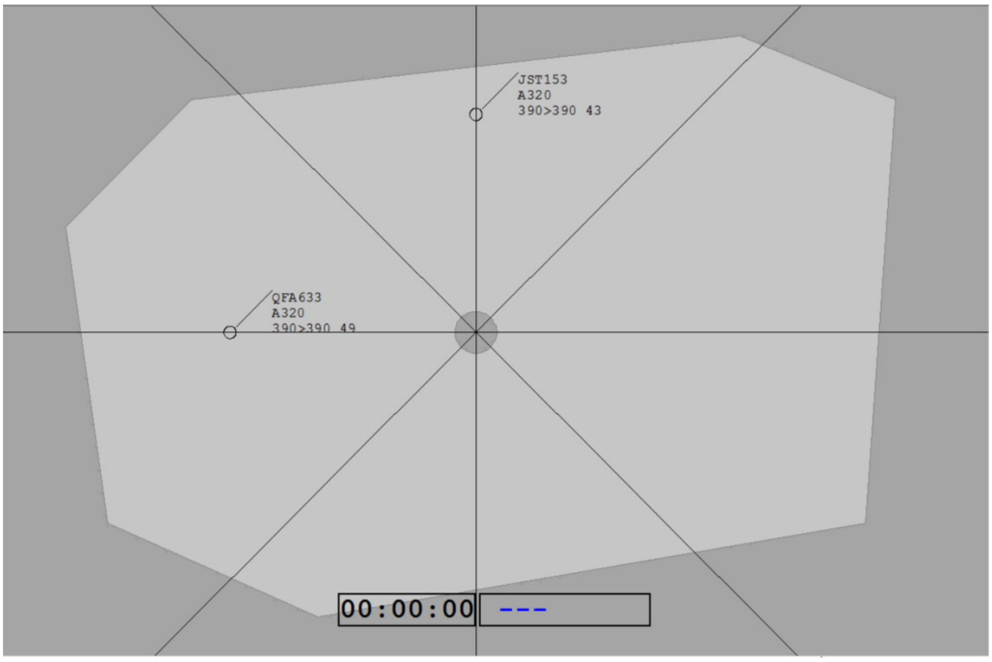

Figure 1: Airspace structure representation. The circle around the central convergence point has a radius of 5nm, and it subtends 1 ° of visual angle from a viewing distance of approximately 60 cm. Aircrafts are represented by a small circle, with relative tag containing flight data: call sign on the first line; type of aircraft on the second line; current altitude, assigned altitude, and speed on the third line.

2.3 EXPERIMENTAL DESIGN AND PROCEDURE

Two CA levels (90°; 135°) and five MD levels (0; 1; 6; 10; 12nm) were manipulated in a 2x5 within subject design. On the basis of lateral separation standard adopted in the study (5nm), two levels of MD indicated a conflict (0; 1nm), while the remaining three a no conflict (6; 10; 12nm). In total, 10 experimental conditions resulted from factors' combination. In order to cover both right and left halves of display with aircraft position at trial start, we built two scenarios for each condition. Therefore, a set of 20 trials was created and finally uploaded in Experiment Builder software (SR Research, Ontario, Canada). All the scenarios had the same time to minimum separation (i.e. 7 minutes, blind information). Screenshots from two examples are reported in Figure 2.

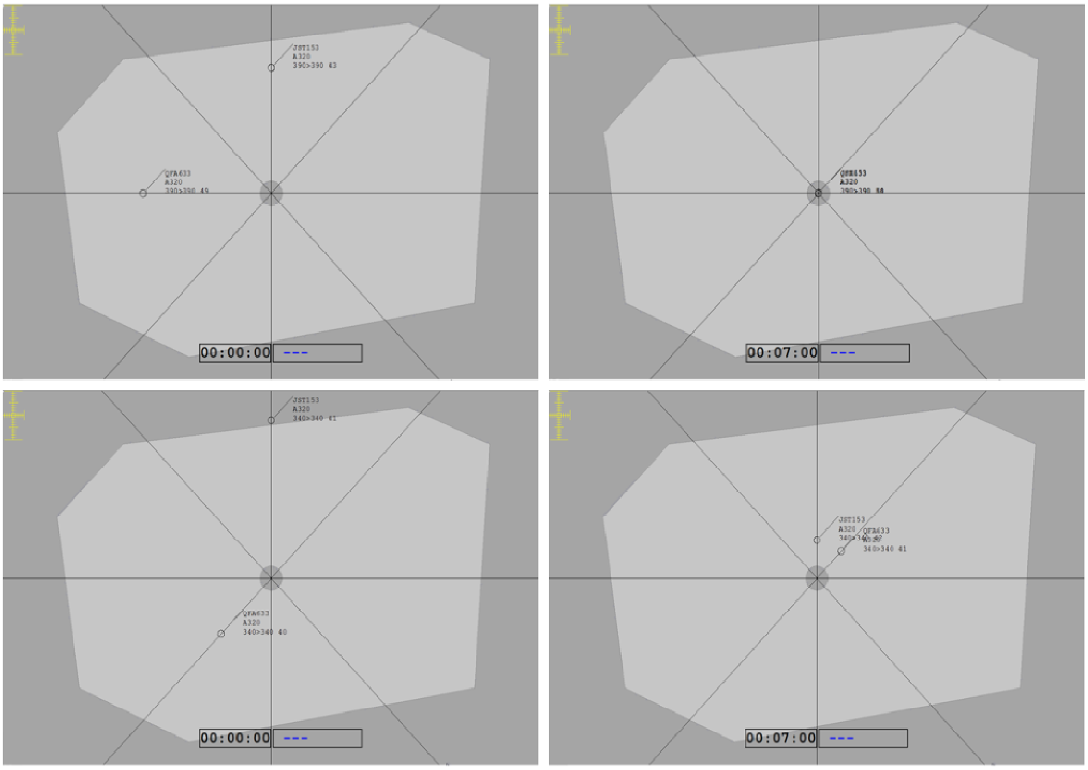

Figure 2: Examples of traffic scenarios. Left side represents trial start; right side represents aircrafts at closest approach, after 7 minutes. Top panel: conflict (MD = 0nm, CA = 90°). Bottom panel: no conflict (MD = 12nm, CA = 135°).

The conflict detection task consisted of judging whether two converging aircrafts would have lost lateral separation through time, i.e. whether the distance between them would have been below (conflict) or remained above (no conflict) the critical separation standard. After task instruction explanations, four training trials were performed (without recording eye tracking data), and accuracy feedbacks were given to improve learning. Experiment began with the randomized presentation of experimental trials. Participants performed the conflict detection task by means of a PC mouse whose buttons were colored in red (conflict, left button) and green (no conflict, right button), respectively. After each trial, participants were asked to verbally express their perceived mental workload. Experimental session lasted for approximately 1 hour. A schematic representation of the procedure is provided in Figure 3.

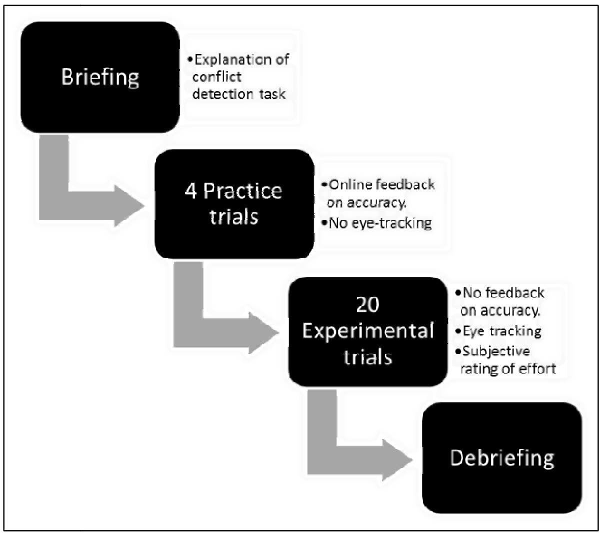

Figure 3: Experimental procedure breakdown. Participants were briefed on conflict detection task and had the opportunity of practicing it in four practice trials. After the experimental session, a final debriefing on study's hypotheses was provided.

2.4 DEPENDENT VARIABLES

Dependent variables are briefly described and relative predictions formulated in the following sections. A schematic representation of hypotheses is provided in Figure 4.

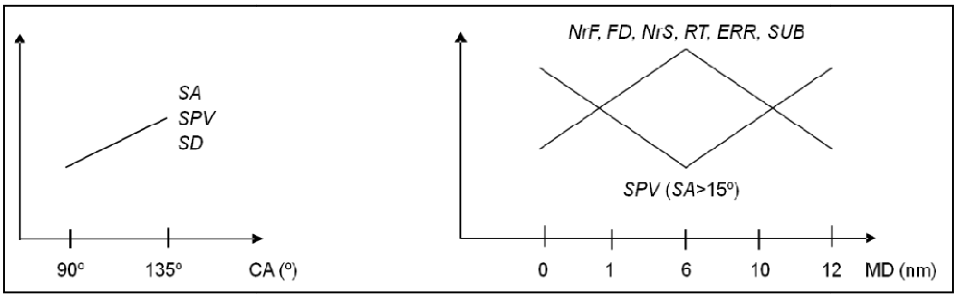

Figure 4: Graphical representation of hypotheses.

2.4.1 OCULAR MEASURES

On the basis of the so-called eye-mind relationship, eye tracking is a useful technique for monitoring operators' attention (Poole & Ball, 2005). Fixations and saccades constitute the basic core events of ocular behaviour. Fixations represent an overt visual attention process that directly inform about attentional resources allocation (Just & Carpenter, 1976; Henderson, 2003).

Concerning fixations, two metrics were selected: total number of fixations (*NrF*), and mean fixation duration (*FD*). Several studies have shown the positive correlation between *NrF* and workload (Goldberg & Kotval, 1998; Ha, Kim, Lee, & Seong, 2006; Lin, Zhang, & Watson, 2003). According to Henderson (2007), the comprehension of the visual world is strongly dependent on the fixated points, and, *FD* is a direct expression of the time spent during information processing. Longer *FD* has been associated to observer's difficulty in extracting information from a display (Fitts, Jones, & Milton, 1950; Goldberg & Kotval, 1998; Callan, 1998). In conflict detection task, MD has been showed to be strongly associated to mental workload (Loft et al., 2009; Xu & Rantanen, 2007; Xu, 2003). In this study we hypothesized that when MD value was close to the separation standard, conflict detection would have been more demanding if compared to MD values considerably above or below the separation criteria. Consequently, we expected that MD = 6nm would generate higher *NrF* and longer *FD* with respect to remaining MD levels.

Saccades are rapid eye movements occurring between fixations, and are performed to bring task relevant information to the fovea, the retinal area with greatest visual acuity. By means of saccades, attentional requirements are accomplished. No encoding takes place during saccades (saccadic suppression), but saccade dynamics are indicative of oculomotor system's effort during task completion (Goldberg & Kotval, 1999). In this experiment, four metrics were considered: total number of saccades (*NrS*), average saccadic amplitude (*SA*), average saccadic peak velocity (*SPV*), and average saccadic duration (*SD*). The last two parameters are related in a nonlinear manner to *SA*: according to the main sequence relationship, *SD* and *SPV* will increase as long as *SA* also increases (Bahill & Stark, 1975). Main sequence indicates the relationships between the size of saccades and their duration, as well as between size and peak velocity. It shows a linear trend for saccades with *SA* up to 15° or 20°, while reaching a soft saturation for larger saccades. This saturation effect is mainly due to ocular motoneurons firing at their maximum rates when realizing large saccadic movements, up to human *SPV* thresholds. Many studies evidenced how contextual factors such as fatigue (Schmidt, Abel, Dell'Osso, & Daroff, 1979; Cazzoli et al., 2014), arousal (Di Stasi, Catena, Cañas, Macknik, & Martinez-Conde, 2013; Benedetto, Carbone, Drai-Zerbib, Pedrotti, & Baccino, 2014), or task difficulty (Di Stasi, Marchitto, Antolí, Baccino, & Cañas, 2010) can affect saccadic velocities, in particular those of large saccades. On the basis of these evidences, we expected significant main effects of MD and CA on the selected saccade metrics. As to MD, we predicted that correct conflict detections would have required more time, and consequently produced more saccades, with higher complexity trials (MD = 6nm). Furthermore, on the basis of previous findings (Marchitto et al., 2012), we predicted that large saccades ($SA > 15°$) for high complexity trials would have triggered slower *SPV*, as to indicate the decreased average burst power of oculomotor system in such situations. As to CA, we expected that larger saccades would have been needed for higher CA, i.e. when aircrafts would have been farther each other in the display. According to this assumption, we predicted higher *SA* for wider convergence angles, together with higher *SD* and *SPV*.

2.4.2 PERFORMANCE MEASURES

As to performance measures, two metrics were selected: mean error rate (*ERR*) and response time (*RT*). Errors were produced both when no conflicts were gauged as conflict trials (false alarm) and when conflicts were assessed as no conflicts (miss). Although in real ATC situations misses are more dangerous than false alarms, in this experiment both type of errors were considered together, and a unique accuracy metric was computed. In line with previous statements on ocular measures, we expected that more complex trials (MD = 6 nm), would have increased both *ERR* and *RT*. With respect to *RT*, smaller values have been already found for MD = 0; 1; 10; 12nm in comparison with MD = 6; 8nm (Loft et al., 2009).

2.4.3 SUBJECTIVE MEASURES

We employed the Verbal Online Subjective Opinion (VOSO) for subjective workload estimation (*SUB*). The VOSO (Miller, 2001) is an eleven-point (from 0 to 10) unidimensional rating scale, which requires a verbal response of the perceived mental workload experienced during the task. It was administered at the end of each experimental trial. In line with previous predictions on ocular and performance measures, higher VOSO scores were expected for MD = 6nm, lower for conflicts (MD = 0; 1nm), and even lower for remaining no-conflict trials (MD = 10; 12nm), on the basis of MD difference with the lateral separation standard.

3. RESULTS AND DISCUSSION

Concerning data analysis, participants with performance accuracy below 70% (i.e. more than 6 errors out of 20 decision trials) were excluded. Such a threshold was employed mainly because in a binary decision task in which conflicts were present in 40% of trials, accuracy levels around 50% could have been reached by guessing. As a result, 9 participants over 26 were excluded from the analysis. Furthermore, only correct trials were included in the analysis of ocular, performance, and subjective data.

In relation to ocular measures, some further thresholds were applied in order to exclude abnormal values. As to fixations, *FD* ranged from 100ms to 1000ms. Fixations around blinks were dropped from the sample data. As to saccades, detection required a deflection in eye position greater than 0.1°, with a minimum velocity of 30°/s and a minimum acceleration of $8000°/s^2$, maintained for at least 4ms. Thresholds for duration (10–250ms) and peak velocity (35–900°/s) were also applied, and saccades around blink eliminated.

This present section is organized as follows. Results of correlational and inferential analysis are presented and discussed in sections 3.1 and 3.2, respectively. Finally, implications about adopted strategies are discussed in section 3.3.

3.1 CORRELATIONAL STUDY

The correlational analysis was carried out to investigate the occurrence of considerable associations between ocular, performance, and subjective measures. Due to the predictable high positive correlations between *RT* and ocular events frequencies (*NrF* and *NrS*) partial correlations were preferred: *RT* was set as continuous control variable and its effects controlled. Resultant correlations among these variables are reported in Table 1.

As to fixations, *NrF* showed a positive relationship with *NrS* (*r* = .95), while negative with *ERR* (*r* = -.38). The alternation of fixations and saccades explains this strong positive association. Moreover, the significant negative correlation with *ERR* suggested that larger frequencies of fixations and saccades were associated to higher accuracy. As to *FD,* a significant negative correlation was found with *NrS* (*r* = -.20). This association might be interpreted in terms of different visual behaviors that might be adopted in a visual task. Shorter *FD* and higher saccadic frequency might be typical of exploratory visual behaviour, while longer *FD* and lower saccadic frequency might be expression of focused ocular behaviour. In this sense *FD* confirms to be associated with attentional and cognitive effort in visual tasks.

In relation to saccades, *NrS* correlated negatively with *ERR* (*r* = -.39)*.* In this experiment, participants were asked to perform the conflict detection task as best they could. Therefore, the overall quantity of eye events (saccades and fixations) was associated with the accuracy of decision. In particular, it seemed that, independently of time needed to make a decision, conflict detection performance improved with higher frequency of fixations and saccades, i.e. more detailed exploration and frequent processing. In a dynamic decision task as conflict detection, the probability of making the correct decision increases with time, while available time for implementing a conflict solution decreases accordingly. Partial correlations controlled for this effect.

The remaining significant correlations between saccadic metrics were essentially due to main sequence relationship: *SA* highly correlated with *SPV* (*r* = .57), and *SD* (*r* = .40). Similarly, *SPV* and *SD* correlated positively (*r* = .41). Furthermore, *SD* was positively correlated with *SUB* (*r* = .24), as to indicate that less workload was estimated in presence of shorter saccades: short *SD* are associated to short *SA*, and thus to local exploration. In relation to *ERR*, there was a significant positive correlation with *SUB* (*r* = .35), meaning that lower accuracy was associated to higher subjective workload.

In summary, results of the correlation study confirmed that ocular metrics were associated to perceived complexity and resulting cognitive workload. In fact, subjective estimations correlated positively with saccadic parameters, suggesting that oculomotor system effort is taken into account when rating subjective workload. Performance errors (ERR) correlated negatively with the number of ocular events.

	NrF	*FD*	*NrS*	*SA*	*SD*	*SPV*	*SUB*
FD	-.135						
NrS	.954**	-.204**					
SA	-.013	.023	-.065				
SD	-.065	.046	-.092	.395**			
SPV	-.094	-.057	-.120	.573**	.411**		
SUB	-.091	.044	-.098	.091	.242**	.159*	
ERR	-.379**	.032	-.392**	-.094	.113	.083	.348**

Table 1. Partial correlations matrix. RT was set as control variable. Marked correlations (*) significant at $p < .05$, and marked correlation (**) significant at $p < .01$; N = 17.

3.2 INFERENTIAL STUDY

After computing partial correlations for dependent variables, we wanted to test the effects of CA and MD complexity factors while controlling for RT. A 2x5 repeated measures ANCOVA was performed for each of the dependent variables, with RT as covariate (M = 28s, SE = 4.8s). The significance level α was set at .05 for all statistical analyses. When the Mauchly sphericity test was significant, we applied the Greenhouse-Geisser correction. As to MD factor, we performed planned comparisons of MD = 6nm level with each other one, according to hypotheses. In relation to multiple post-hoc comparisons, we used Bonferroni correction. Means and standard deviations for each dependent variable are reported in Table 2.

DV	CA = 90°					CA = 135°				
	MD (nm)									
	0	1	6	10	12	0	1	6	10	12
NrF	165.3 (143.2)	138.2 (109.6)	100.4 (79.1)	121.1 (114.9)	84.0 (60.6)	122.8 (91.6)	119.6 (93.0)	110.9 (93.1)	86.3 (79.6)	53.3 (31.0)
FD (ms)	274.8 (47.8)	277.4 (45.7)	264.6 (28.9)	272.9 (35.9)	256.0 (45.0)	264.4 (38.9)	264.4 (35.1)	254.2 (38.9)	252.9 (43.8)	246.2 (35.3)
NrS	185.1 (154.0)	154.4 (129.0)	114.9 (94.8)	136.0 (133.4)	93.7 (75.0)	146.0 (119.8)	137.3 (109.5)	123.6 (110.0)	96.2 (98.2)	56.6 (32.0)
SA (°)	4.3 (0.8)	4.4 (1.1)	4.0 (0.6)	3.8 (0.8)	4.6 (0.9)	4.5 (0.7)	4.8 (1.1)	4.7 (1.0)	4.6 (1.0)	5.3 (1.1)
SD (ms)	42.7 (19.2)	44.1 (23.8)	47.6 (28.2)	46.6 (27.5)	48.5 (33.1)	44.1 (25.1)	43.5 (23.9)	42.7 (21.0)	43.0 (22.3)	44.8 (25.8)
SPV (°/s)	284.0 (152.6)	303.7 (181.7)	342.2 (212.1)	312.5 (188.5)	309.5 (196.0)	300.0 (181.6)	301.9 (175.5)	305.9 (183.2)	302.1 (179.9)	322.0 (202.6)
ERR	0.2 (0.3)	0.2 (0.3)	0.6 (0.4)	0.2 (0.4)	0 -	0.2 (0.3)	0.3 (0.3)	0.1 (0.2)	0 -	0 -
SUB	6.1 (1.6)	5.6 (1.5)	4.9 (1.4)	4.2 (1.8)	3.4 (1.2)	5.7 (1.7)	5.8 (1.4)	3.9 (2.2)	3.4 (1.6)	2.0 (0.9)
RT (s)	43.1 (36.4)	33.7 (28.4)	26.8 (19.0)	33.2 (31.0)	18.7 (19.5)	34.1 (28.4)	35.4 (25.5)	24.8 (25.1)	18.5 (23.7)	9.9 (6.6)

Table 2. Means and standard deviations (in parentheses) for each of the dependent variables and experimental conditions. Mean error rates (*ERR*) were calculated by dividing number of errors in each experimental condition by number of observations (2 trials in each condition x 17 participants = 34 observations). Subjective workload (*SUB*) was verbally rated on VOSO scale (from 0 to 10).

3.2.1 OCULAR MEASURES

Concerning *NrF* and *FD,* there were not significant main effects of CA and MD factors, as well as CAxMD interactions. Therefore, fixation metrics in the study could be predicted by *RT*: higher *RT* was assumed to indicate higher cognitive complexity of conflict detection and thus, increased frequency and duration of information collection episodes.

Concerning saccades, a further factor was introduced in the analysis, i.e. interval of saccadic amplitude (Bin), and consequently *NrS, SD,* and *SPV* were analyzed as a function of Bin, CA, and MD. Five amplitude intervals were selected: 0.1°-1° (microsaccades), 1°-5° (short saccades), 5°-10° (medium-short saccades), 10°-15° (medium-large saccades), and 15°-53° (large saccades). Wider saccades were expected with higher CA. On the basis of main sequence relationship, we assumed that *SD* and *SPV* should have varied similarly. However, we also hypothesized lower values of *SPV* in most difficult trials (MD = 6nm) for large saccades (Bin 5). Significant interactions for saccadic metrics are reported in Figure 5.

As to *NrS*, results replicated those of fixation metrics. Once the effect of *RT* has been controlled by covariance analysis, there was not any remaining effect of complexity factors CA and MD on saccadic frequency. As to *SA*, wider saccades were found for CA = 135°, as hypothesized [$F(1, 15) = 19.73$, $p < .001$, $\eta^2_p = .59$]. Consequently, the geometrical cost for oculomotor displacement with higher CA was observed in *SA* data: with wider CA aircrafts at trial start are placed further each other, and saccades with bigger amplitude need to be performed in order to locate task relevant information, like position and speed.

As to *SD*, bigger values were found for CA = 90° [$F(1, 15) = 16.6$, $p = .001$, $\eta^2_p = .52$]. Combined *SA* and *SD* results suggest that, on average, saccades were performed more slowly with right-angles. Main effect of MD was not significant, while there was a strong Bin effect [$F(4, 60) = 214.1$, $p < .001$, $\eta^2_p = .94$], due to main sequence. Several interactions were significant, namely CAxMD [$F(4, 60) = 3.18$, $p = .02$, $\eta^2_p = .18$], CAxBin [$F(4, 60) = 15.5$, $p = .001$, $\eta^2_p = .51$], and CAxMDxBin [$F(16, 240) = 3.76$, $p < .012$, $\eta^2_p = .20$]. We performed two separate 5x5 (MDxBin) ANOVA in each level of CA as simple effects analysis for the second order interaction. As to CA = 90°, there were significant differences for large saccades (Bin 5), which showed lower *SD* in MD = 0nm with respect to MD = 6; 12nm (both $p < .001$). The other conflict trials (MD = 1nm) presented the same trend (both $p = .06$). As to CA = 135°, significant differences were found for microsaccades (Bin 1), which showed higher *SD* in MD = 0nm with respect to MD = 12nm ($p = .01$) and, similarly to CA = 90°, for large saccades (Bin 5), although this effect was weaker ($p = .05$), with no significant post hoc comparisons. Therefore, duration of large saccades decreased and that of microsaccades increased with higher cognitive complexity, as in conflict trials. These results further support microsaccades as index of workload and experienced fatigue (Benedetto, Pedrotti, & Bridgeman, 2011; Di Stasi et al., 2013).

As to *SPV*, CA main effect was not significant, while MD main effect it was [$F(4, 60) = 6.82$, $p < .001$, $\eta^2_p = .31$]: planned comparisons in relation to MD = 6nm showed higher values with respect to MD = 0; 1; and 10nm (all $p < .001$). No other differences were found after post hoc comparisons. Therefore, there was a considerable difference in *SPV* data of conflicts with respect to no conflicts, the latter presenting higher values. Planned comparisons related to MD = 6nm in Bin 5 showed the same pattern as in MD main effect (all $p < .001$): this result was opposed to predictions, although it was a convergent evidence of the higher complexity and cognitive resources consumption of conflicts with respect to no conflicts. Bin effect also was significant [$F(4, 60) = 270.9$, $p < .001$, $\eta^2_p = .95$], as expected. Significant interactions were MDxBin [$F(16, 240) = 6.35$, $p < .001$, $\eta^2_p = .30$] and CAxMDxBin [$F(16, 240) = 2.67$, $p < .04$, $\eta^2_p = .15$]. As to the second order interaction, simple effect analysis revealed in CA = 90° trials significant lower *SPV* values of large saccades (Bin 5) for MD = 0nm with respect to no conflicts (MD = 6; 10; 12nm), as well as between MD = 1nm and MD = 6nm (all $p < .001$). As to CA = 135° trials, effects on *SPV* were weaker, and related to Bin 4: MD = 0nm presented lower values with respect to MD = 10nm ($p = .03$).

Therefore, *SPV* results resembled those of *SD*: large saccades in conflicts had lower values with respect to no conflicts, especially with vertical and horizontal routes. Hypotheses predicted CA effects on saccadic parameters, as geometrical complexity factor that, when increasing, requires oculomotor system to trigger movements of wider amplitude with congruent velocity and duration. This result was substantially confirmed for *SA*, but not for *SD* (which showed higher values with CA = 90°), neither *SPV* (CA main effect not significant). Effects on *SD* and *SPV* involved mainly large saccades with CA = 90° trials. We made a precise prediction on large saccades' *SPV* about a possible slow down effect for increased complexity, since large saccades already proved to be sensitive to task difficulty (Marchitto et al., 2012). Results about large saccades showed a substantial difference in perceived complexity between conflicts and no conflicts independently of *RT*, whose effects were controlled as covariate, while effects of complexity factors on remaining ocular metrics were strictly dependent on conflict detection times. Large saccades in conflicts showed reduced burst power with respect to no conflicts, especially when CA = 90°.

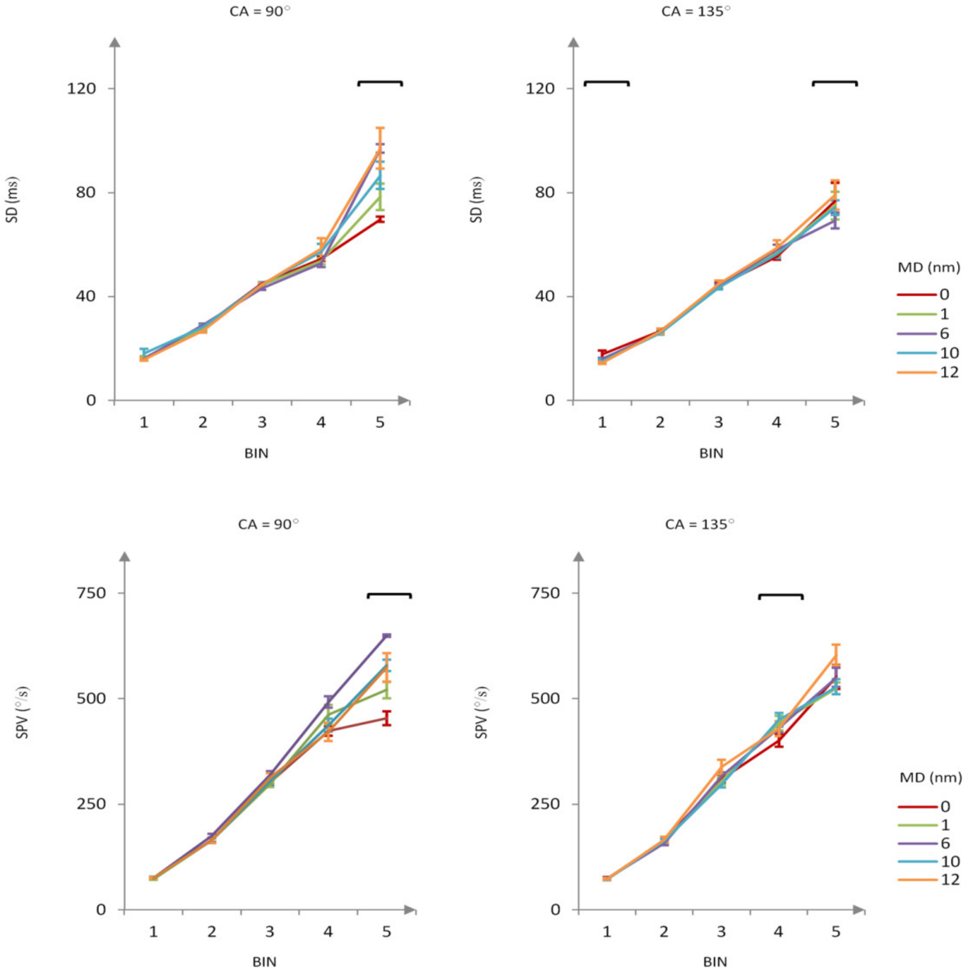

Figure 5: Saccadic metrics significant interactions. Top panel shows simple effects analyses for CAxMDxBin interaction on *SD*, bottom panel shows the same analysis in relation to *SPV*: CA = 90° on the left and CA = 135° on the right. Vertical bars denote standard errors (*SE*). Horizontal lines indicate Bins with significant MD effects.

These results constitute empirical evidence that *SPV* of large saccades decreased when cognitive complexity increased, as for example when dealing with horizontal and vertical routes. Since both large saccades *SD* and *SPV* decreased with higher complexity, they probably became more precise, with less burst power and reduced positional error, as a task adaptation process would suggest (Gancarz & Grossberg, 1999).

In summary, the geometry complexity factor CA affected as expected *SA*, while effects on *SD* presented a trend opposite to predictions: there was an intrinsic cognitive difficulty with CA = 90°, most probably due to horizontal and vertical flying routes. Perpendicular routes have been reported to be particularly difficult (Boag et al., 2006). The traffic complexity factor MD did not affected fixation parameters, once the effects of *RT* were removed. The expected effects of MD = 6nm were not encountered, but a clear difference between conflicts and no conflicts emerged after ocular metrics analysis. Trials with the biggest MD difference compared to lateral separation standard of 5nm (i.e. MD = 12nm) resulted to be the least complex for conflict detection task. Fixation parameters are further confirmed as reliable metrics for on line workload assessment, especially in more complex traffic scenarios (e.g. with higher traffic loads), in which conflict detection *RTs* for single pairs of aircraft would be hardly partitioned. It is reasonable to conclude that in conflict detection task the most fixated elements are those related to highest cognitive complexity. More importantly, manipulation of MD affected saccadic parameters: large saccades showed shorter *SD* and lower *SPV* in conflicts with respect to no conflicts (especially with MD = 12nm). Probably, as cognitive complexity augmented, large saccades became more precise, i.e. with shorter duration and weaker burst signal. Therefore, large saccades were affected by traffic and geometry complexity factors, increasing perceived cognitive complexity of the conflict detection task. Increased cognitive complexity augmented local exploration (more saccades with lower amplitude), information processing and thus, consequent workload. All these differences were highly confirmed by subjective and performance data. Next section reports on these findings.

3.2.2 PERFORMANCE MEASURES

We measured error rates (*ERR*) and response time (*RT*) as performance indices, and analyzed them separately in two repeated measures ANOVAs. As to *ERR*, no errors were made in three conditions, as reported in Table 2, namely when MD = 12nm in both CA levels, and when MD = 10nm with CA = 135°. These conditions had not variance and were not further considered in the analysis. As predicted, they were the easiest conditions for correct conflict detection. Consequently, a 2x3 ANOVA was performed.

There were higher *ERR* with CA = 90° [$F(1, 16) = 5.68$, $p = .03$, $\eta^2_p = .26$], and a significant CAxMD interaction [$F(2, 32) = 10.69$, $p < .001$, $\eta^2_p = .40$], due to higher *ERR* when MD = 6nm for CA = 90° with respect to CA = 135°. Therefore, MD = 6nm was effectively the most error-prone condition, with an intrinsic difficulty deriving from the little difference with lateral separation standard, as predicted. This effect was sensibly amplified in presence of vertical and horizontal routes. Results about *ERR* confirmed predictions.

As to *RT*, there were later responses for CA = 90° [$F(1, 16) = 18.05$, $p = .001$, $\eta^2_p = .53$]. Main effect of CA on *RT* was not explicitly predicted, but it demonstrated the inherent difficulty of CA = 90° situations, as ocular data already suggested. Moreover, MD main effect was significant [$F(4, 64) = 10.49$, $p < .001$, $\eta^2_p = .40$]. Planned comparisons revealed that *RT* in MD = 6nm was higher than in MD = 12nm ($p = .003$), but lower than in MD = 0nm ($p = .049$). Post hoc comparisons revealed higher *RT* values for MD = 0; 1nm vs. MD = 12nm ($p < .001$ and $p = .01$, respectively). There was not significant interaction. Main effects of MD on *RT* are reported in Figure 6. These results showed that there was a clear effect of MD as cognitive complexity factor that increased *RT* in conflict trials, requiring more cognitive effort and producing later responses, as opposed to no-conflict trials (especially MD = 12nm), which were easier conditions that required little effort and allowed a very quick response. Furthermore, the particular geometry configuration with CA = 90° determined a further cognitive cost observable in performance data, ocular metrics, and subjective ratings, as reported in next section.

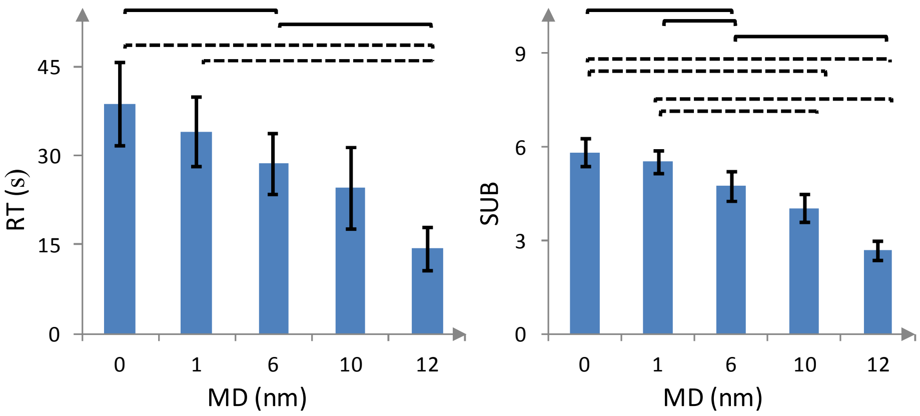

Figure 6: Main effects of MD on performance (*RT*) and subjective ratings of workload (*SUB*). Vertical bars denote *SE*. Horizontal lines indicate significant planned comparisons; dotted lines refer to post hoc comparisons.

3.2.3 SUBJECTIVE MEASURES

Subjective ratings of workload (*SUB)* were collected after each trial using VOSO scale (Miller, 2001), and were analyzed introducing *RT* as covariate (as for ocular metrics), in order to remove effects on workload ratings due to longer times for correct conflict detection. Trials with CA = 90° generated significant higher workload ratings [$F(1, 15) = 5.94$, $p = .028$, $\eta^2_p = .28$]. Main effect of MD was also significant [$F(4, 60) = 6.38$, $p < .001$, $\eta^2_p = .30$], as reported in Figure 6. Planned comparisons revealed that *SUB* in MD = 6nm was higher than in MD = 12nm ($p < .001$), but lower than in MD = 0; 1nm ($p < .008$ and $p = .024$, respectively). Post hoc comparisons further revealed significant differences between conflicts and no conflicts: MD = 0nm vs. MD = 10; 12nm ($p = .008$ and $p < .001$, respectively); MD = 1nm vs. MD = 10; 12nm ($p = .046$ and $p < .001$, respectively). Interaction was not significant. In accord with ocular and performance variables, *SUB* results showed a substantial difference between conflicts and no conflicts and, among no conflicts, between MD = 12nm and remaining levels.

3.3 GENERAL DISCUSSION

These results further support the interpretation of ocular metrics results in terms of cognitive workload generated by the conflict detection task. Fixation and saccadic parameters captured the higher cognitive complexity of conflicts with respect to no conflicts (MD = 10; 12nm), and of vertical and horizontal routes with respect to obtuse angle of convergence (CA = 135°). The most error-prone experimental condition was MD = 6nm, as predicted. Nevertheless, conflicts trials required longer visual exploration and information collection, and generated higher subjective workload. Ocular metrics confirmed these trends, especially peak velocity of large saccades.

These considerable quantitative differences might underlie qualitatively different cognitive processes for trajectory prediction and lateral separation estimation. As to conflicts, trajectory prediction was probably based on movement observation and aircraft position recurrent processing. Conversely, no conflicts presented fewer fixations (due to little *RT*), so that a strategy that allowed quick prediction after rapid computation of aircraft alphanumerical data and distance to convergence point was most probably adopted. Literature reports on several strategies for conflict detection task (Nunes & Mogford, 2003). It is normally assumed that the appropriate implementation of a strategy to a particular situation assures higher levels of performance.

According to Xu and Rantanen (2003), conflict detection task consists of two phases, performed not necessarily in a fixed order: a relative judgment task (RJ) to estimate which aircraft will reach the convergence point first, and a prediction-motion (PM) task to calculate when the aircraft will be at convergence point and to decide about their possible conflict situation after lateral separation estimation. In some special cases RJ is particularly easy, as when aircraft with the same distance from convergence point and same speed are rapidly estimated to collide at intersection point, i.e. to be in conflict. As to PM, some available strategies that simplify conflict detection are reported. For example, times to convergence point can be calculated for each aircraft in any instant by dividing distance to convergence point and speed, i.e. applying distance-to-speed ratio strategy. As a consequence, it is possible to make a prediction on lateral separation by projecting aircraft position once its motion has been included in a cognitive projection model, i.e. the cognitive simulation of flights future progress. This strategy is considerably helpful in conflict decision task when estimated times are very similar or considerably different: in fact, it would be relatively easy to estimate a conflict and a no conflict, respectively. Conversely, a medium difference in times to intersection point could make the estimation of lateral separation more difficult: in this case, distance-to-speed ratio strategy might be insufficient for correct conflict detection. Another strategy for PM is cognitive motion extrapolation (CME). This strategy does not imply time calculations, and it is based on motion representation in spatial dimension, leaving time dimension implicit. It assumes that the cognitive model of object's motion used for projection is explicitly spatial (DeLucia & Liddell, 1998). A recurrent processing of aircraft updated position is necessary for a correct spatial representation of motion.

It is normally assumed that the strategies that guarantee optimal result with minimal cognitive effort are applied whenever possible and that shifts towards more demanding strategies occur whether the former option is unfeasible. The difference between conflicts and no conflicts evidenced in data analysis could be a consequence of two different strategies that might have been adopted in no conflicts and conflicts. As to no conflicts, quick responses (after few fixations and saccades) might have been the result of adopting a ratio-based strategy after processing aircraft start position and flight data. As to conflict trials, a prediction-motion strategy based on spatial representation of motion was most probably adopted (e.g. CME): since flight data were kept constant through time, the higher number of fixations and saccades (predicted by late response time) were related to continuous aircraft position updates.

We suggested that there might have been a strategy shift from a ratio-based to a prediction-motion strategy in conflicts trials, due to the ratio-based strategy being unfeasible in such more uncertain situations. A quick decision on conflict presence might have been impeded by aircraft differences in speed and distance to convergence point, due to the difficulty of transforming the estimated difference in time of arrival into a reliable estimation of lateral separation. Conflicts were more difficult, demanding, and effortful, and this strategy shift might have contributed to final workload assessment. To support this conclusion, we computed speed and distance differences for each experimental condition at trial start, as reported in

Table 3. These differential values were the consequence of traffic dynamic calculations for the experimental scenarios, after setting an MD level, fixing a time-to-MD value (7 minutes in this study), and choosing a speed value for one of the two aircrafts.

CA	MD (nm)	Speed Difference (kn)	Distance Difference (nm)
90°	0	60*	10
	1	20	0
	6	10	10
	10	20	15
	12	20	20
135°	0	60*	10
	1	50*	12
	6	10*	23
	10	50	28
	12	10*	43

Table 3: Differential speeds in knots (kn) and differential distances to convergence point in nautical miles (nm) at trial start for each of the experimental conditions. Marked speed differences (*) mean that the aircraft closer to convergence point was slower than the other aircraft.

These differential values might have determined a strategy shift in conflict detection task. No conflicts (MD = 6; 12nm) with CA = 135° were easily answered probably because distance differences were considerable and speed differences little: consequently, RJ and PM were relatively easy. Similarly, when MD = 10nm with CA = 135°, a considerable speed difference simplified the task, also because the aircraft closer to intersection was faster: the distance difference with convergence point would have increased through time, further separating aircrafts in the lateral dimension. No conflicts with MD = 6nm and CA = 90° presented a moderate distance difference and a little speed difference. Participants tended to erroneously interpret this situation as a conflict, since difference in times of arrival was probably estimated as close to zero. Conflicts also presented little distance differences and might have made particularly difficult the application of a ratio-based strategy. In fact, in almost all conflicts there was a considerable speed difference at trial start, and the aircraft closer to intersection was slower. Consequently, the distance difference with convergence point would have decreased as effect of speed: calculations of time to convergence point should have been considerably precise. In this case, RJ resulted quite difficult, since it should have been decided whether the quicker aircraft would have had time to arrive before the closer one. When facing this situation, participants might have switched to a strategy based on position recurrent computation and projection, instead of relying on difficult and error-prone calculations.

It is acknowledged in ATC literature that differential (or relative) values of speed, distance and flight level are central information for correct conflict detection (Leplat & Bisseret, 1966; Ahlstrom, 2005). Task performance is better when differential values are considerable high or close to zero (Kimball, Hofmann, & Nossaman, 1973). Next section presents the conclusions drawn after quantitative analysis of ocular metrics discussed in this section.

4. CONCLUSION

This study explored ocular behavior during simulated ATC conflict detection task. Ocular metrics were registered together with performance and subjective ratings of workload, in order to capture workload variations as a function of conflict detection task complexity. We manipulated two complexity factors in order to create different levels of task difficulty, i.e. convergence angle (CA), and minimum distance at closest approach (MD). Triangulation method proved to be very fruitful for multimodal workload measurement, supporting the validity of ocular metrics as sensitive online workload indices.

We hypothesized that the most complex situations were those with MD close to separation standard (i.e. MD = 6nm). Hypotheses were confirmed only partially: MD = 6nm was the most error-prone condition, but conflict trials were more time-consuming and rated as the most demanding scenarios. On the contrary, the remaining no conflict trials (MD = 10; 12nm) resulted in lower subjective workload levels, more rapid responses, and little visual processing. Moreover, vertical and horizontal routes acted unexpectedly as further complexity factor. Ocular metrics were able to capture higher cognitive complexity. Effects on fixation metrics were absent once removed contemporary influence of response time. This substantially confirms the usefulness of fixations metrics recording for cognitive workload when it is difficult to isolate response time to a single aircraft pair, e.g. with higher traffic loads. As to saccadic metrics, we observed that large saccades (*SA* > 15°) were affected by MD factor, showing shorter duration and lower peak velocity in conflicts. It seemed that large saccades were more precise, i.e. more rapid and with weaker burst power for conflict trials, which were the most difficult situations. Velocity parameters of large saccades were sensitive to task difficulty, in accord with several recent findings (Di Stasi et al., 2010; 2011; 2013; Marchitto et al., 2012).

The substantial difference in cognitive complexity between conflicts and no conflicts was interpreted as the probable adoption of different strategies in the two conditions. In no conflicts (especially with MD = 12nm) it was suggested the application of a ratio-based strategy, at least at a rough level, that allowed quick estimations of differential times of arrival to convergence point. In conflict trials, a strategy shift might have occurred, by adopting a motion-prediction strategy. Fixation metrics and saccadic parameters of large saccades captured this cognitive cost. We concluded that participants first checked for differential speeds and distances to convergence point, and then made a quick decision if a considerable difference in times to convergence point was estimated, as in no conflicts. Conversely, the particular combination of differential speeds and distances to intersection in conflicts probably determined the adoption of a motion-based strategy, which required more cognitive effort with respect to a ratio-based strategy and increased decision times. As a consequence, conflicts were perceived as more complex and demanding.

Ocular metrics allowed for online workload monitoring and are a promising method for strategy characterization (Hunter & Parush, 2009). Workload modeling as well as its management would benefit from the refinement of temporal and spatial analysis of ocular indices, for example in ATCO training (Kang & Landry, 2014). The incorporation of quantitative and qualitative assessment of visual behaviour in novice ATCOs learning process might enhance the identification of high workload scenarios, task relevant information, and adopted strategy for traffic management. Ocular metrics measurement for online workload assessment could also be successfully applied in the field of adaptive technology, to identify overload episodes and to support controller by automating ATC subtasks. The manipulation of differential (or even relative) speed, distance to convergence point, and also flight level might help further in identifying the conditions for strategies' adoption, by means of scanpath visualization. For example, the use of specific Areas of Interests (AOI) could help us to better address cognitive complexity in conflict detection task (Kang & Landry, 2010).

ACKNOWLEDGEMENTS

This work was supported by Spanish Ministry of Economy and Competitiveness, PSI2012-39246. We are very grateful to all the researchers and colleagues who have reviewed and commented on this manuscript at various stages. The authors have no conflicts of interests to declare.

REFERENCES

Ahlstrom, U. (2005). Work domain analysis for air traffic controller weather displays. *Journal of safety research, 36*(2), 159-169.

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623-636.

Alonso, R., Causse, M., Vachon, F., Parise, R., Dehais, F., & Terrier, P. (2013). Evaluation of head-free eye tracking as an input device for air traffic control. *Ergonomics*, *56*(2), 246-255.

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). ATC complexity and controller workload: Trying to bridge the gap. In *Proceedings of the International Conference on HCI in Aeronautics* (pp. 56-60).

Ayaz, H., Willems, B., Bunce, B., Shewokis, P. A., Izzetoglu, K., Hah, S., Deshmukh, A., & Onaral, B. (2010). Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. In T. Marek, W. Karwowski & V. Rice (Eds.), *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations* (pp. 21-32): CRC Press, Taylor & Francis Group.

Benedetto, S., Carbone, A., Drai-Zerbib, V., Pedrotti, M., & Baccino, T. (2014). Effects of luminance and illuminance on visual fatigue and arousal during digital reading. *Computers in Human Behavior*, *41*, 112-119.

Benedetto, S., Pedrotti, M., & Bridgeman, B. (2011). Microsaccades and exploratory saccades in a naturalistic environment. *J. Eye Mov. Res*, *4*(2), 1-10.

Boag, C., Neal, A., Loft, S., & Halford, G. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics, 14,* 1508–1526.

Callan, D. (1998). Eye movement relationships to excessive performance error in aviation. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 1132-1136). Santa Monica, CA: Human Factors and Ergonomics Society.

Cummings, M. L., & Tsonis, C. (2005, September). Deconstructing complexity in air traffic control. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 1, pp. 25-29). Sage Publications.

Cazzoli, D., Antoniades, C. A., Kennard, C., Nyffeler, T., Bassetti, C. L., & Müri, R. M. (2014). Eye movements discriminate fatigue due to chronotypical factors and time spent on task – A double dissociation. *PloS one*, *9*(1), e87146.

DeLucia, P. R., & Liddell, G. W. (1998). Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 901-914.

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2011). Main sequence: an index for detecting mental workload variation in complex tasks. *Applied ergonomics*, *42*(6), 807-813.

Di Stasi, L. L., Catena, A., Cañas, J. J., Macknik, S. L., & Martinez-Conde, S. (2013). Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews*, *37*(5), 968-975.

Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, *16*(6), 330-333.

Di Stasi, L. L., McCamy, M. B., Catena, A., Macknik, S. L., Cañas, J. J., & Martinez-Conde, S. (2013), Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience, 38*, 2389–2398.

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review, 9*(2), 24-29.

Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab^{Advanced}: An air traffic control simulator with realism and control. *Behavior research methods*, *41*(1), 118-127.

Gancarz, G., & Grossberg, S. (1999). A neural model of saccadic eye movement control explains task-specific adaptation. *Vision research*, *39*(18), 3123-3143.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24,* 631-645.

Goldberg, J. H., & Kotval, X. P. (1998). Eye movement-based evaluation of the computer interface. In Kumar, S.K. (Ed.), *Advances in occupational ergonomics and safety*, 529-532. Amsterdam: IOS Press.

Ha, C. H., Kim, J. H., Lee, S. J., & Seong, P. H. (2006). Investigation on relationship between information flow rate and mental workload of accident diagnosis tasks in NPPs. *Nuclear Science, IEEE Transactions on*, *53*(3), 1450-1459.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, *7*(11), 498-504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219-222.

Hilburn, B. (2004). Cognitive complexity in air traffic control - A literature review. *EEC Note, 4*(04). Brétigny-sur-Orge: Eurocontrol Experimental Research Center.

Histon, J. M., Hansman, R. J., Gottlieb, B., Kleinwaks, H., Yenson, S., Delahaye, D., & Puechmorel, S. (2002). Structural considerations and cognitive complexity in air traffic control. In *Digital Avionics Systems Conference, 2002. Proceedings. The 21st* (Vol. 1, pp. 1C2-1). IEEE.

Hunter, A. C., & Parush, A. (2009). Using eye movements to uncover conflict detection strategies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 22, pp. 1729-1733). Sage Publications.

ICAO, International Civil Aviation Organization. *Forecasts of scheduled passenger traffic (2012)*. Retrieved on November, 6th, 2014, from [http://www.icao.int/sustainability/pages/eap_fp_forecastmed.aspx].

Imants, P., & de Greef, T. (2011, August). Using eye tracker data in air traffic control. In *Proceedings of the 29th Annual European Conference on Cognitive Ergonomics* (pp. 259-260). ACM.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Radach, R., Hyona, J., & Deubel, H. (Eds.), *The Mind's eye: Cognitive and Applied Aspects of Eye Movement Research*, 573 - 603. Amsterdam: Elsevier.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441-480.

Kallus, K. W., Van Damme, D., & Dittman, A. (1999). *Integrated job and task analysis of air traffic controllers*: *Phase 2. Task analysis of en-route controllers* (European Air Traffic Management Programme Rep. No. HUM. ET1. ST01.1000-REP-04). Brussels, Belgium: Eurocontrol.

Kang, Z., & Landry, S. J. (2010, September). Capturing and analyzing visual groupings of multiple moving targets in an aircraft conflict detection task using eye movements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 23, pp. 1906-1910). SAGE Publications.

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(6), 1150-1162.

Kimball, K. A., Hofmann, M. A., & Nossaman, R. O. (1973). Differential velocity and time prediction of motion. *Perceptual and motor skills*, *36*(3), 935-945.

Langan-Fox, J., Sankey, M. J., & Canty, J. M. (2009). Human factors measurement for future air traffic control systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(5), 595-637.

Leplat, J. (1978). The factors affecting workload. *Ergonomics, 21,* 143-149.

Leplat, J., & Bisseret, A. (1966). Analysis of the processes involved in the treatment of information by the air traffic controller. *The Controller, 5*(1), 13–22.

Li, L., Cho, H., Hansman, R. J., & Palacios, R. (2010). Aircraft-based complexity assessment for radar controllers in the multi-sector planner experiment. In *10th AIAA Aviation Technology, Integration, and Operations (ATIO) conference* (pp. 1-9).

Lin, Y., Zhang, W. J., & Watson, L. G. (2003). Using eye movement parameters for evaluating human–machine interface frameworks under normal control operation and fault detection situations. *International Journal of Human-Computer Studies*, *59*(6), 837-873.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, *15*(2), 106-124.

Lokhande, K., & Reynolds, H. J. D. (2012, September). Cognitive workload and visual attention analyses of the air traffic control tower flight data manager (TFDM) prototype demonstration. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 105-109). Sage Publications.

Majumdar, A., & Ochieng, W. Y. (2007). Air traffic control complexity and safety: framework for sector design based on controller interviews of complexity factors. *Transportation Research Record: Journal of the Transportation Research Board*, *2007*(1), 70-80.

Marchitto, M., Di Stasi, L. L. and Cañas, J. J. (2012), Ocular movements under taskload manipulations: Influence of geometry on saccades in air traffic control simulated tasks. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*(5), 407–419.

Martin, C., Cegarra, J., & Averty, P. (2011). Analysis of mental workload during en-route air traffic control task execution based on eye-tracking technique. In *Engineering Psychology and Cognitive Ergonomics* (pp. 592-597). Springer: Berlin Heidelberg.

Miller, S. (2001). Workload measures. *National Advanced Driving Simulator.* Oakland, IA: The University of Iowa.

Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). *The complexity construct in air traffic control: a review and synthesis of the literature*. CTA INC MCKEE CITY NJ.

Nunes, A., & Mogford, R. H. (2003, October). Identifying controller strategies that support the 'picture'. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 1, pp. 71-75). SAGE Publications.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140-160.

Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996, July). A framework for the evaluation of air traffic control complexity. In *Proceedings of the AIAA Guidance Navigation and Control Conference, San Diego, CA*.

Poole, A., & Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research: current status and future. In C. Ghaoui (Ed.), *Encyclopedia of Human-Computer Interaction* (211-219). Pennsylvania: Idea Group, Inc.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human Factors, 42,* 349–366.

Schmidt, D., Abel, L. A., Dell'Osso, L. F., & Daroff, R. B. (1979). Saccadic velocity characteristics: Intrinsic variability and fatigue. *Aviation, space, and environmental medicine*, *50*(4), 393-395.

Stankovic, S., Raufaste, E., & Averty, P. (2008). Determinants of conflict detection: A model of risk judgments in air traffic control. *Human Factors, 50,* 121–134.

Weiland, M. Z., Roberts, D. M., Fine, M. S., & Caywood, M. S. (2013, September). Real time research methods monitoring air traffic controller workload during simulation studies using electroencephalography (EEG). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1615-1619). SAGE Publications.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, *35*(2), 263-281.

Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 381-389.

Xu, X. (2003, October). Conflict detection with cockpit display of traffic information: what is it, what has been found, and what needs to be done? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 1, pp. 194-198). SAGE Publications.

Xu, X., & Rantanen, E. M. (2007). Effects of air traffic geometry on pilots' conflict detection with cockpit display of traffic information. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(3), 358-375.

Xu, X., & Rantanen, E. M. (2003). Conflict detection in air traffic control: a task analysis, a literature review, and a need for further research. In R. S. Jensen (Ed.), *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 1289–1295). Dayton, OH: Wright State University Press.

CHAPTER 5

GAZE TRANSITIONS, COGNITIVE STRATEGIES, AND WORKLOAD

CHAPTER 5

WORKLOAD AND COGNITIVE STRATEGY IN CONFLICT DETECTION: ANALYSIS OF GAZE TRANSITIONS AND OCULAR METRICS[4].

ABSTRACT

The monitoring of workload and cognitive strategy for conflict detection in the domain of air traffic control (ATC) is a key factor for improving system safety and capacity. In this study, conflict detection task demands were varied in a simulated ATC environment by manipulating aircraft differences in speed, distance to convergence point, and altitude. Gaze transitions were analyzed as a function of differential values, allowing the identification of strategies adopted for task completion. In addition, experienced workload was monitored by measuring performance (response time), subjective (workload rating), and psychophysiological (pupil size variation) indexes. Transition analysis enabled to observe the redistribution of attentional resources with increasing workload, the perceptual costs for finer processing of aircraft positions, and the cognitive costs of parallel separation estimation in both vertical and lateral dimensions. Workload was negatively related to differential altitude and distance to convergence point, as predicted. Differently, effects of differential speed on workload depended on the perceived probability of a lateral loss of separation. Air traffic control system safety might benefit from gaze transitions analysis for online strategy identification and workload monitoring in real or simulated operational contexts, as well as to support verbal protocols or post hoc performance analysis during training.

---

[4] Submitted as:
Marchitto, M., Benedetto, S., Cañas, J. J., & Baccino, T. (submitted). Workload and cognitive strategy in conflict detection: analysis of gaze transitions and ocular metrics. *Journal of Cognitive Engineering and Decision Making*.

1. INTRODUCTION

The problem of monitoring operator's cognitive workload is of major concern in complex socio-technical systems since the foundations of ergonomics (Fitts, Jones, & Milton, 1950). On line, unobtrusive, easy-to-administer methods for mental workload evaluation have been sought in many operational settings. In the domain of air traffic control (ATC), the monitoring of operator mental state during traffic management without jeopardizing operational safety is a major concern (Loft, Sanderson, Neal, & Mooij, 2007, Langan-Fox, Sankey, & Canty, 2009). Most of workload research aims at identifying overload situations and strategies adopted by ATC operators (ATCO) for releasing a safe and efficient service. However, performance detriments are rarely observed in ATC domain, as ATCO can invest more cognitive resources whenever task requirements increase. Moreover, it is acknowledged that ATCO can modify the cognitive strategy in order to reduce situational workload (Sperandio, 1978; Averty, Collet, Dittmar, Athènes, & Vernet-Maury, 2004). In practice, workload monitoring based on performance fluctuations recording is still difficult and often unviable in such operational settings.

Conflict detection is a fundamental activity performed by ATCO (Kallus, Van Damme, & Dittman, 1999; Rantanen & Nunes, 2005), which consists of recurrent pair comparisons of convergent aircrafts, in order to determine whether their lateral and vertical separation will be below prescribed minima, e.g. 5nm (nautical miles) in the lateral plane and 1000ft (feet) in the vertical plane. The early identification of conflicts is fundamental to leave available time for their resolution (Eyferth, Niessen, & Spaeth, 2003). Therefore, conflict detection is based on aircraft separation estimation (i.e. the perceptual and cognitive processes enabled by relative motion projection). In fact, conflict detection requires the construction of a cognitive model for motion projection through two steps. Relative judgment (RJ) allows to establish arrival order to convergence point, whereas prediction-motion (PM) allows to project aircrafts' relative movement in the future, and to estimate separation (Xu & Rantanen, 2003; Tresilian, 1995). The recurrent computations of speed, altitude and distance to convergence point values during flight progress are fundamental for performing RJ and PM correctly (Stankovic, Raufaste, & Averty, 2008; Rantanen, 2004) and to identify future loss of separation (LOS). Therefore, the development of methods and metrics sensitive to strategy shifts and variations in mental resources investment during conflict detection is fundamental for enabling ATC system safety (Nunes & Mogford, 2003).

Besides engineering approaches, which identified a considerable set of traffic and environmental complexity factors that moderate perceived complexity and consequent cognitive workload (for a review see Hilburn, 2004; Majumdar & Ochieng, 2002), the operator-centered approach, typical of psychological studies on cognitive complexity, has recently received growing attention (Boag, Neal, Loft, & Halford, 2006; Histon & Hansman, 2002). Such a human-centered cognitive approach to conflict detection aims at modeling controllers' decision processes through the application of knowledge about traffic complexity factors. Scenarios of different complexity are used to identify cognitive strategies adopted for task resolution, and to detect strategy shifts (Loft, Bolland, Humphreys, & Neal, 2009; Neal & Kwantes, 2009). Computational algorithms for conflict detection have been considerably fruitful for the design of automated detection tools, but less plausible for studying the management of operator mental resources. Indeed, the controller-centered approach to conflict detection focused mainly on ATCOs' heuristic judgments (Bisseret, 1981) and dynamic mental representation of traffic (Averty, Guittet, Lezaud, 2008).

Together with subjective and performance measures, psychophysiological metrics have been progressively applied in workload studies during the last decades for mapping cognitive processes related to conflict detection. Due to unobtrusiveness, portability, and relative ease of use, eye trackers might become the most promising technology in the ATC domain. Eye tracking provides fruitful information about visually attended elements in plenty of situations and tasks (Just & Carpenter, 1976; Viviani, 1989). The relation between visual attention and cognitive processing has been largely debated, but no definitive positions have been reached (Anderson, Bothell, & Douglass, 2004; Brandt & Stark, 1997). Nevertheless, in real operational settings as well as in human-in-the-loop simulations of operational tasks with a dominant visual component (e.g. computer display-supported process control), it is reasonable to assume that visually attended elements provide important clues about the resulting cognitive engagement (Just and Carpenter, 1985; Poole & Ball, 2005; Grant & Spivey, 2003; Thomas & Lleras, 2007). Therefore, eye tracking might be successfully applied for operator state monitoring. Although several ocular metrics have been employed to measure workload in ATC (Willems, Allen, & Stein, 1999; Ahlstrom & Friedman-Berg, 2006; Marchitto, Di Stasi, & Cañas, 2012; Di Stasi, Marchitto, Antolí, Baccino, & Cañas, 2010), to the authors' knowledge, research that address cognitive strategies during conflict detection on the basis of gaze transitions analysis is still modest (Hunter & Parush, 2009; Kang & Landry 2014). Eye movements are realized during visual exploration in order to maintain foveal vision on appropriate information. A transition is the movement of point of gaze that occurs between two fixations. The analysis of transitions between task-relevant areas of interest (AOI) allows to quantify ordered sequences of eye movements and to infer adopted cognitive strategies (Kang & Landry, 2010).

This study monitored workload and identified projection strategies in a simulated ATC conflict detection task by means of eye tracking methods, while varying task complexity by manipulating differential[5] speed (ΔS), distance to convergence point (ΔD), and altitude (ΔL). Information about altitude, heading, and speed has been reported to be fundamental for controller's ability to estimate future vertical and lateral separation between converging aircrafts (Rantanen, 2004). To this end, we assumed a positive relation between workload and ΔS, as higher values would have increased uncertainty about aircraft future relative position (i.e. with respect to the other aircraft) and lateral separation. Furthermore, we predicted a negative relation with ΔD and ΔL, as increasing values might have clued safe separation preservation (lateral and vertical, respectively). We considered that workload resulting from the combination of flight metrics differential values might have depended on cognitive strategy adopted for motion prediction. We assumed increasing values of pupil size variation (*PVar*), response time (*RT*), and subjective workload rating (*SUB*) as indicating increasing workload.

Our contribution is twofold. First, we proposed gaze transitions' analysis for inferring cognitive strategies adopted for task resolution as a function of ΔS, ΔD, and ΔL. Second, we assessed cognitive workload associated to traffic scenarios by

---

[5] The differential value is the difference in absolute value between two aircrafts for a specific flight metric at scenario onset (e.g., two aircrafts flying at 370kn — knots — and 340kn, respectively, present ΔS 30kn).

means of performance, subjective, and psychophysiological measures. Such approach was used to characterize strategy description and quantify related sustained effort. The paper is organized as follows. Section 2 details experimental method, information about participants, procedure, and stimuli. Results are presented and discussed separately for each scenario type in Section 3. Lastly, Section 4 presents the conclusions drawn, together with some reflections for practical applications.

2. METHODS

2.1 PARTICIPANTS

The sample consisted of 37 participants (29 female, mean age = 24 years, *SD* = 5). Most of them were students (*N* = 29) at Université Paris 8 (Psychology Department), who volunteered for course credits. The rest of the participants were employees at the Cité de Science et de l'Industrie (Paris), where the study was conducted. None of them had experience with ATC tasks, and they all had normal or corrected-to-normal vision (contact lenses were accepted, glasses were not). The experiment followed the tenets of the Declaration of Helsinki (World Medical Association, 2013).

2.2 TASK AND PROCEDURE

Participants were required to perform 36 conflict detection tasks, randomly presented. A conflict situation was defined as a contemporary absence of safe separation in both the lateral and the vertical plane, i.e. less than 5nm and 1000ft, respectively. The task consisted of indicating whether or not aircrafts would have entered in conflict during flight. Answers were provided by means of a PC mouse with buttons labeled "yes" (left button, in red), and "no" (right button, in green). After each response, participants were asked to rate verbally the perceived workload. Before the beginning of the experiment, participants went through a training session, which lasted approximately 30 min. The session consisted of two phases. During the first phase, participants were asked to read a manual in French concerning the conflict detection task they would have performed later. During the second phase, the experimenter had to ensure the proper understanding of the task by asking to the each of the participants a set of predetermined questions. In case of negative result, the procedure was repeated until the understanding was complete.

2.3 APPARATUS

Experimental setting consisted of a seat in front of stimulus screen with embedded eye-tracker, i.e. remote infrared SMI RED 5 (www.smivision.com), with a sampling rate of 250Hz. A 9-point calibration was performed for each participant at the beginning of experimental session. The average distance between participants and the 22'' LCD screen (Dell P2210; www.dell.com) employed for the task was 60 cm. The environmental light (provided by two fluorescent lamps TL84, 4000°K of luminosity intensity) was kept constant throughout the experiment: illuminance on participants' eyes was 140lx at a viewing distance of approximately 70cm, assessed by an Extech 403125 digital light meter (www.extech.com) pointed toward the screen.

2.4 STIMULI

We used the ATC-lab$^{\text{Advanced}}$ simulator to program experimental scenarios (Fothergill, Loft, & Neal, 2009). In total, 36 stimuli were built (see Table 1). Differential speed (ΔS), distance to convergence point (ΔD), and level (ΔL) were manipulated separately (three levels each) in scenario types A, B, and C, respectively. Scenario type D, E, and F, consisted of two metrics combination (3x3 levels), namely ΔS&ΔD, ΔS&ΔL, and ΔD&ΔL, respectively. Of note, we adopted a closer-is-slower paradigm in type D scenarios, i.e. presented the faster aircraft farther from convergence point, as we considered excessively simple the case with the faster aircraft closer to convergence point. We selected three differential values for each manipulated metric: ΔS (10kn, 30kn, 70kn), ΔD (10nm, 20nm, 30nm), and ΔL (1000ft, 2000ft, 3000ft). In each scenario type, aircrafts showed equal and constant values for not manipulated metrics (Δ = 0). We explored workload fluctuations as a function of differential values used within scenario types, considered separately. Figure 1 shows a schematic representation of the relevant elements for conflict detection.

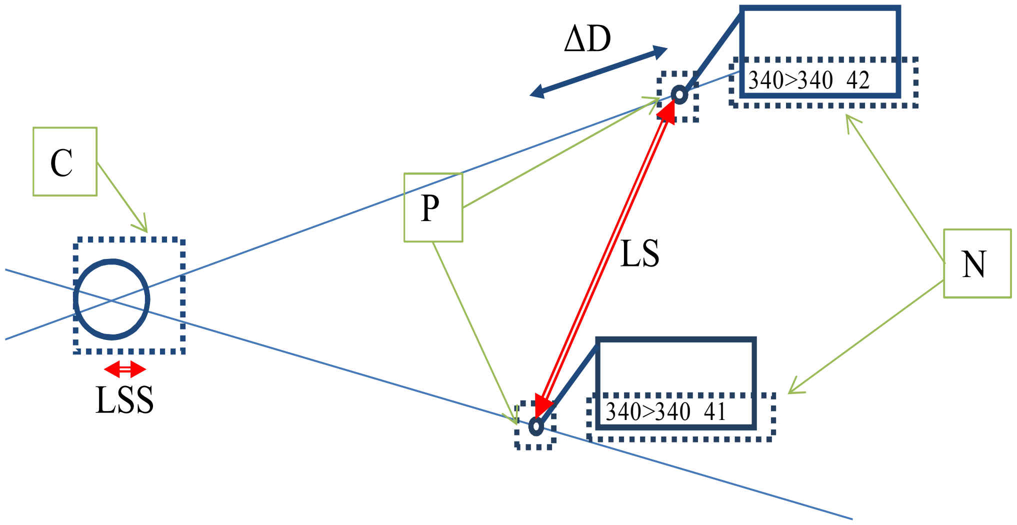

Figure 1. Schematic representation of a traffic scenario. Lateral separation standard (LSS), i.e. the radius of the circle around convergence center, and aircraft lateral separation (LS), were explicitly indicated to participants. In the example, ΔD is graphically indicated. Differently, ΔS and ΔL result from aircraft speeds and current altitudes subtractions. In the example, ΔS = 10kn (one "0" normally omitted in data tag) and ΔL = 0ft, with both aircrafts level at 34000ft (two "0" normally omitted). Dotted boxes indicate AOIs for gaze transitions analysis: convergence center (C); aircraft positions (P), and data tags with numbers (N).

Scenario	ΔS (kn)	ΔD (nm)	ΔL (ft)
A1*	10		
A2*	30		
A3	70		
B1		10	
B2		20	
B3		30	
C1*			1000
C2			2000
C3			3000
D1	10	10	
D2		20	
D3		30	
D4	30	10	
D5		20	
D6		30	
D7*	70	10	
D8		20	
D9		30	
E1*	10		1000
E2			2000
E3			3000
E4*	30		1000
E5			2000
E6			3000
E7	70		1000
E8			2000
E9			3000
F1		10	1000
F2			2000
F3			3000
F4		20	1000
F5			2000
F6			3000
F7		30	1000
F8			2000
F9			3000

Table 1. Differential values of ΔS, ΔD, and ΔL for experimental scenarios. Asterisk (*) denotes conflict scenarios; kn = knots; nm = nautical miles; ft = feet. A knot corresponds to 1nm/h (nautical mile per hour); 1nm correspond to 1852m (meters). No difference was present when not indicated (Δ = 0).

2.5 DEPENDENT VARIABLES

We defined 5 AOIs on the basis of the scenario elements considered fundamental for abstracting aircraft motion profiles and predict separation. Two AOIs consisted of aircraft positions (P), which were updated each 5s depending on speed and altitude. Following, the two bottom lines of numbers in aircraft data tags were considered (N), as they report altitude (current and assigned) and speed data. All these four AOIs were dynamic, i.e. changed their position accordingly to aircrafts moving forward. Lastly, convergence point (C) was the only static AOI. We computed frequencies of bidirectional transitions between AOIs and aggregated them in five types, as reported in Table 2. Cognitive strategies for conflict detection were inferred from transitions frequency variations as a function of differential values within each scenario type. For each subject, transitions frequencies were transformed in percentages (%) on the total number of transitions performed in that scenario, before aggregating data from different subjects. Repeated measures ANOVAs were performed for each transition within each type of scenario, with differential values as predictors.

Name	From – to
P&Ns	Position - Numbers: same aircraft
P&P	Position - Position
N&N	Numbers - Numbers
P&Nd	Position - Numbers: different aircraft
PN&C	Position or Number - Convergence point

Table 2. Name and definition of the 5 types of two-way transitions.

Complementary, we measured response time (*RT*), subjective workload ratings (*SUB*), and average pupil variation (*PVar*) as workload metrics. As to *RT*, we considered time from scenario onset until mouse click ("yes" or "no" buttons). As to *SUB*, we employed an eleven-point unidimensional rating scale (from 0 to 10) for collecting participants' workload rating immediately after response. Unidimensional scales are useful for obtaining *on line* estimations, rapidly and unobtrusively. Similar scales have been previously used in workload research (e.g. the Verbal Online Subjective Opinion, VOSO) (Miller, 2001), while in ATC domain they often required a manual response (e.g. Air Traffic Control Workload Technique, ATWIT) (Stein, 1985). As to *PVar,* we followed standard practice in pupillometry research and measured average pupil diameter variation (mm) with respect to a baseline (i.e. pupil size during the first fixation on traffic scenario). Therefore, positive variation indicated pupil dilation (Beatty & Lucero-Wagoner, 2000), independently of baseline pupil size. The positive relation between pupil dilation and workload has been largely documented both in relation to basic cognitive tasks (Chen & Epps, 2014) and to ATC simulated environments (Brookings, Wilson, & Swain, 1996; Ahlstrom & Friedman-Berg, 2006). We asked participants to indicate whether a conflict was present or absent, keeping as much constant as possible response-induced pupillary variations (Richer & Beatty, 1985).

3. RESULTS AND DISCUSSION

For each scenario, we performed transitions and workload analysis as a function of differential values. Descriptive statistics are reported in Table 3. In order to exclude performance influenced by guessing, we considered only participants with accuracy above sample mean (M = 72.2%, SD = 6.5, corresponding to 10 errors maximum out of 36 scenarios). The final sample size was N = 19. Furthermore, we excluded from analysis scenarios with error rate >.5. These error-prone scenarios were A3, B1, C1, D1, D4, and E4. As to type D scenarios, the final 3x2 repeated measures ANOVA had ΔS (10kn, 30kn, 70kn) and ΔD (20nm, 30nm) as predictors. As to type E scenarios, the final 3x2 repeated measures ANOVA had ΔS (10kn, 30kn, 70kn) and ΔL (2000ft, 3000ft) as predictors. Critical p value was set at .05 and partial eta-squared (η_p^2) was computed as effect size measure. When sphericity test was significant, Greenhouse-Geisser correction was used. We performed dependent *t*-tests for planned comparisons and used Bonferroni correction for multiple post hoc comparisons.

	Transition Type (%)					Workload metrics		
Scenario	P&Ns	P&P	N&N	P&Nd	PN&C	RT (s)	SUB (0/10)	PVar (mm)
A1	22.5 (4.7)	3.7 (2.2)	19.9 (5.8)	12.4 (3.5)*	8.2 (4.4)	24.1 (3.1)	4.8 (.41)*	0.18 (.034)
A2	16.6 (4.5)	8.8 (3.8)	30.9 (8.1)	2.1 (1.2)*	8.4 (5.2)	18.9 (2.5)	3.5 (.31)*	0.13 (.050)
B2	15.3 (3.5)	7.9 (3.6)	14.6 (4.9)*	7.2 (3.0)	16.1 (4.9)*	25.6 (5.0)*	4.3 (.58)*	0.26 (.071)
B3	37.5 (8.7)	6.1 (3.6)	35.6 (7.8)*	7.9 (3.8)	1.9 (1.3)*	11,1 (1.3)*	2,2 (.22)*	0,19 (.055)
C2	19.5 (6.0)	1.0 (0.6)	34.1 (7.8)	5.2 (1.7)	11.7 (5.7)	23,2 (2.6)	3,9 (.30)	0,15 (.055)
C3	17.0 (7.1)	0.4 (0.3)	41.6 (8.4)	17.6 (7.0)	5.8 (2.4)	21,5 (3.3)	3,6 (.44)	0,06 (.076)
D2	15.0 (4.0)	1.7 (1.0)	25.6 (5.6)	1.7 (1.0)	31.1 (7.1)*	24,0 (2.7)	4,8 (.44)	0,16 (.032)
D3	27.3 (7.4)	5.4 (2.6)	18.4 (5.8)	7.8 (3.1)	7.8 (3.6)*	21.5 (3.3)	3.9 (.45)	0.05 (.058)
D5	24.0 (5.8)	12.1 (3.5)	21.9 (4.9)	4.0 (1.8)	4.7 (2.0)	27.5 (2.8)*	4.8 (.50)	0.02 (.041)
D6	25.4 (7.4)	10.4 (5.6)	19.5 (6.3)	5.0 (2.6)	8.2 (4.4)	19.7 (2.6)*	4.0 (.43)	0.11 (.060)
D8	19.4 (4.6)	18.3 (4.5)	5.7 (1.4)*	5.5 (1.9)	14.7 (5.1)	55.0 (4.7)°	5.8 (.47)	0.24 (.052)
D9	16.6 (6.0)	5.9 (3.0)	24.3 (7.4)*	9.6 (5.5)	10.2 (3.9)	22.8 (4.0)°	3.8 (.46)	0.31 (.052)
E2	5.8 (3.0)	1.5 (1.1)	39.0 (8.7)	34.1 (8.1)*	3.0 (1.4)	32.1 (5.4)	4.2 (.41)	0.16 (.041)
E3	20.3 (6.6)	1.1 (0.7)	38.3 (9.4)	0.6 (0.5)*	6.4 (3.1)	25.0 (3.7)	3.9 (.39)	0.22 (.055)
E5	21.8 (6.0)	4.7 (2.7)	30.0 (7.3)	5.3 (2.4)	14.7 (7.1)	31.8 (4.6)	4.4 (.39)	0.09 (.041)
E6	20.9 (6.9)	4.0 (2.0)	45.0 (8.6)	4.1 (2.4)	14.8 (5.1)	23.2 (3.3)	3.6 (.44)	0.08 (.006)
E8	20.0 (6.3)	2.2 (1.3)	27.2 (7.1)	2.1 (1.0)°	11.0 (5.3)	25.7 (3.3)	3.8 (.39)	0.14 (.044)
E9	22.8 (4.9)	2.7 (1.4)	40.7 (6.0)	8.9 (2.0)°	2.8 (1.7)	23.5 (2.4)	3.9 (.44)	0.18 (.041)
F1	29.0 (5.0)	2.2 (1.1)	59.2 (5.8)	9.5 (2.8)	0.0 (0.0)	22.1 (1.6)	4.1 (.25)	0.07 (.028)
F2	14.5 (4.4)	4.2 (2.0)	37.8 (7.4)	10.0 (5.4)	7.2 (3.3)	29.4 (4.5)$^{\#}$	3.9 (.41)	0.11 (.055)
F3	28.0 (7.3)	4.4 (2.0)	31.5 (8.0)	2.0 (1.2)	7.9 (3.6)	17.1 (2.0) $^{\#}$	3.5 (.32)	0.22 (.050)
F4	20.4 (6.0)	1.3 (1.3)	20.8 (5.7)	7.9 (5.4)	18.0 (7.5)	25.6 (3.9)°*	4.4 (.50)°*	0.09 (.060)
F5	13.3 (5.0)	0.7 (0.7)	20.5 (7.4)	6.9 (3.3)	21.8 (7.9)	12.9 (1.6)°	2.7 (.21)°	0.10 (.064)
F6	15.3 (4.4)	3.1 (1.9)	39.3 (8.7)	3.1 (1.9)	12.9 (6.5)	15.2 (1.9)*	3.0 (.34)*	0.09 (.062)
F7	13.0 (4.9)	0.8 (0.8)	45.2 (10.0)	2.6 (1.5)	12.2 (6.2)	17.0 (2.1)	3.5 (.37)	0.20 (.057)
F8	15.8 (4.9)	1.8 (1.2)	37.7 (8.2)	5.6 (2.7)	12.8 (5.6)	16.9 (1.9)	3.1 (.30)	0.24 (.073)
F9	17.2 (6.3)	1.8 (1.3)	33.7 (9.4)	7.5 (5.4)	8.1 (4.0)	13.6 (1.5)	2.7 (.25)	0.30 (.062)

Table 3. Descriptive statistics for differential values effects: mean and *SE* (in parentheses) per single scenarios. Markers (*) in type A, B, and C data mean significant differences; markers (*, °, $^{\#}$) in type D, E, and F indicate significant simple effects (that explain significant interaction). Significant main effects are not indicated, as their relative means are calculated after collapsing levels of the ignored factor.

In types A, B, and D participants were requested to project in the lateral dimension, since aircraft had equal and constant altitude. Results for type B and D were very similar: two transition types showed significant effects as a function of ΔD and ΔS, namely N&N (altitude and speed data comparisons) and PN&C (gaze movements between aircrafts and convergence center).

In type B, ΔD = 20nm showed lower N&N transition frequency than 30nm [$F(1, 18) = 6.36$, $p = .004$, $\eta_p^2 = .26$]. The same effect was found in type D [significant ΔSxΔD interaction $F(2, 36) = 6.0$, $p = .006$, $\eta_p^2 = .25$], with lower frequency for ΔD = 20nm only when ΔS = 70kn [$F(1, 18) = 6.25$, $p = .022$, $\eta_p^2 = .26$], as reported in Figure 2. In practice, N&N transition frequency decreased dramatically when lateral separation estimation was more difficult, i.e. with bigger ΔS and smaller ΔD, as predicted.

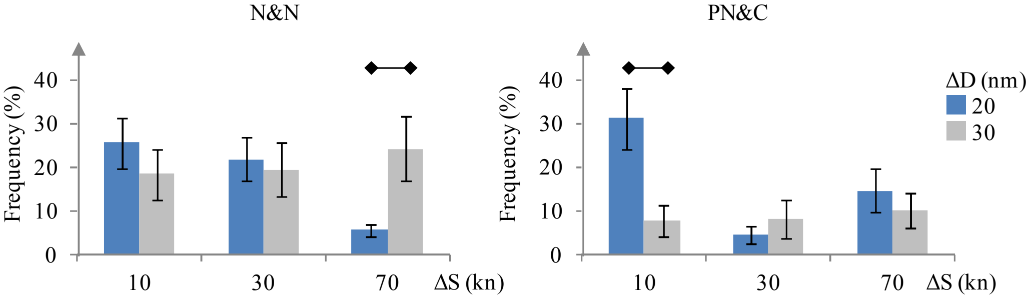

Figure 2. Type D scenarios. Effects of ΔS and ΔD on N&N and PN&C transition types. Vertical bars mean *SE*; horizontal lines indicate significant effects.

In fact, workload metrics confirmed these conditions to be the most demanding, while transition analysis evidenced a redistribution of attentional resources with higher workload, showing that attentional focus was retired from aircraft data tags, which were less relevant for conflict detection in presence of equal and constant speed and altitude. In type B, ΔD = 20nm showed higher *RT* and *SUB* [$F(1, 18) = 8.94$, $p = .008$, $\eta_p^2 = .33$, and $F(1, 18) = 10.25$, $p = .005$, $\eta_p^2 = .36$, respectively] compared to ΔD = 30nm. In type D, both main effects of ΔS and ΔD in *RT* data were significant [$F(2, 36) = 33.5$, $p < .001$, $\eta_p^2 = .65$ and $F(1, 18) = 32.8$, $p < .001$, $\eta_p^2 = .65$, respectively], qualified by a significant ΔSxΔD interaction [$F(2, 36) = 11.7$, $p < .001$, $\eta_p^2 = .40$]: ΔD = 20nm showed higher *RT* compared to 30nm, only when ΔS = 30kn and ΔS = 70kn [$F(1, 18) = 7.04$, $p = .016$, $\eta_p^2 = .28$ and $F(1, 18) = 28.24$, $p < .001$, $\eta_p^2 = .61$, respectively]. Furthermore, ΔS main effect was significant in *PVar* data [$F(2, 36) = 11.6$, $p < .001$, $\eta_p^2 = .39$], with larger pupil dilations with ΔS = 70kn compared to 10kn and 30kn ($p = .039$ and $p = .008$, respectively). Lastly, higher workload for ΔD = 20nm compared to 30nm was found also in *SUB* data [$F(1, 18) = 23.5$, $p < .001$, $\eta_p^2 = .57$].

As to PN&C, we found frequency increases in type B for ΔD = 20nm compared to ΔD = 30nm [$F(1, 18) = 10.13$, $p = .005$, $\eta_p^2 = .36$], an opposite trend with respect to N&N results. Such an effect was confirmed in type D [significant ΔSxΔD interaction: $F(2, 36) = 6.75$, $p = .007$, $\eta_p^2 = .27$], but only when ΔS = 10kn [$F(1, 18) = 14.5$, $p = .001$, $\eta_p^2 = .45$], i.e. in less demanding situations (Figure 2). In fact, ΔD modification throughout flight progress due to ΔS was minor, facilitating both TTC estimation and RJ (i.e. to abstract order of arrival at convergence center): all these flight metrics can be estimated on the basis of correct processing of distance to convergence center. We interpreted higher PN&C transitions frequency in presence of lower workload for projection task (i.e. little ΔS) as the perceptual cost for a more precise processing of ΔD. PN&C transitions increased when finer position differences had to be processed, even if final workload for conflict detection would have remained low. Larger ΔD differences were perceptually more evident and they could have been processed also by parafoveal or even peripheral vision, allowing to heuristically perform RJ and estimate lateral separation at closest approach: this was the reason why ΔD = 30nm was an easy situation in which most probably no motion projection occurred at all, but conflict was ruled out on the basis of perceptual processing of ΔD, without being affected by ΔS. Differently, with ΔD = 20nm, we found a perceptual cost for precise ΔD computation, especially in type D when ΔD is coupled to little (10kn) ΔS. In such a situation, ΔD changed minimally throughout flight progress as a consequence of ΔS, and TTC are abstracted more easily.

Importantly, *PVar* data showed considerable pupil dilation for increasing ΔS, as predicted. The highest *PVar* dilations were found for ΔS (70kn), for both D8 and D9 conditions (ΔD = 20nm and ΔD = 30nm, respectively). By contrast, *RT* and *SUB* data revealed that ΔD = 30nm was estimated as sufficiently large as not to produce any lateral loss-of-separation (LOS), independently of ΔS. Our explanation is that *PVar* variation in D9 was not related to cognitive projection difficulty *per se*, but most probably it was related to cognitive effort for initial aircraft motion extrapolation, before building a motion projection cognitive model that would have allowed separation estimation. Although participants were able to make a correct and quick decision about conflict presence once aircraft motion equations were inferred, the initial visual processing for their abstraction might have demanded a considerable perceptual-cognitive effort, in both D8 and D9. Consequently, abstraction of aircraft TTC was more time-demanding. This increment of cognitive load in the first period after scenario onset could not be captured by *RT* and *SUB* data. Therefore, *PVar* confirmed as a valid online metric for perceptual and cognitive workload monitoring (Chen & Epps, 2014), which can be used for investigating workload related to aircraft motion abstraction in ATC conflict detection task.

To summarize, we expected higher workload for bigger ΔS as well as for smaller ΔD. Such predictions were confirmed in scenario type B and D by *RT*, *SUB*, and *PVar* data. Altitude and speed data comparisons (N&N) were almost absent when workload for lateral projection increased. Furthermore, gaze transitions between aircrafts and convergence center (PN&C) could be considered an index of cognitive load for aircraft motion abstraction, the basic processes for performing lateral estimation. The latter could have been resulted as easy, independently of the difficulty of motion abstraction. Of note, with ΔD = 20nm participants actually performed motion prediction revealing cognitive workload as positively related with ΔS. Differently, with ΔD = 30nm lateral LOS was easily excluded, most probably on the basis of a heuristic judgment (i.e. without performing motion prediction) and deciding after quick processing of aircraft data after scenario onset.

Lateral separation estimation was also performed in type A scenarios, in which aircraft had equal distance to convergence point. Modifications of N&N and PN&C transition frequencies were not encountered, neither lower workload for minor ΔS: among workload metrics, *SUB* showed higher values for Δ*S* = 10kn compared to 30kn [$F(1, 18) = 7.52$, $p = .013$, $\eta_p^2 = .30$], without further converging data about workload differences in type A levels. Differently, P&Nd transitions (gaze movements between position or data tags of different aircrafts) showed higher values with ΔS = 10kn compared to 30kn [$F(1, 18) = 9.76$, $p = .006$, $\eta_p^2 = .35$]. Similar results were found in type E scenarios, which allowed projecting in the two dimensions. Nevertheless, lateral separation estimation was preferred to vertical one. In type E scenarios, P&Nd transition showed significant main effects of both ΔS and ΔL [$F(2, 36) = 6.93$, $p = .009$, $\eta_p^2 = .28$, and $F(1, 18) = 7.78$, $p = .012$, $\eta_p^2 = .30$, respectively], qualified by a significant ΔSxΔL interaction [$F(2, 36) = 17.4$,

$p < .001$, $\eta_p^2 = .49$], as shown in Figure 3. Higher frequency of P&Nd transitions was found with ΔS = 10kn (as in type A), but only with ΔL = 2000ft [$F(1, 18) = 16.5$, $p = .001$, $\eta_p^2 = .48$]. The trend was opposite with ΔS = 70kn [$F(1, 18) = 11.4$, $p = .003$, $\eta_p^2 = .39$]. P&Nd frequency for ΔL 3000ft remained minimal, independently of ΔS. Similarly to result encountered for large ΔD values in type D scenarios, large ΔL values (3000ft) in type E scenarios enabled a quick decision without needing to perform further separation estimation. By contrast, vertical LOS had major occurrence probability with minor ΔL values (1000ft and 2000ft). In parallel, similar aircraft TTC could be inferred with minor ΔS values (10kn). In such cases, lateral LOS was considered as more probable, and determined higher cognitive resources investment for conflict detection.

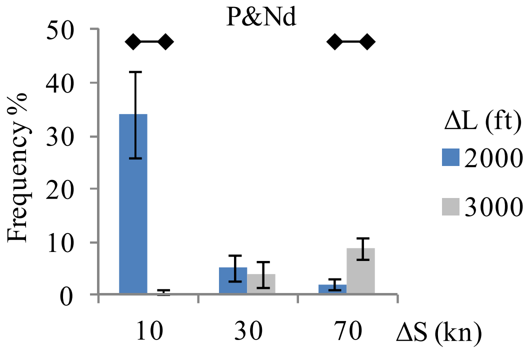

Figure 3. Type E scenarios. Effects of ΔS and ΔL on P&Nd transition frequency. Vertical bars mean *SE*; horizontal lines indicate significant effects.

We observed an increase of P&Nd frequencies in type A and E scenarios when ΔS = 10kn (and distances to convergence center were equal), i.e. when changes in distance to convergence point were very subtle and required more time (with respect to higher ΔS values) to be visible and appreciated by the perceptual-cognitive system. Lateral LOS was deemed as more possible with minor values of ΔS, as they made separation projection more difficult, and the perceptual/cognitive effort for aircraft relative motion abstraction augmented. Such higher frequency of crossed transitions between aircrafts' position and data tags might express the cognitive cost for lateral separation estimation with little ΔS values, plus the further cost for parallel checking of vertical separation preservation (with ΔL =

2000ft). Workload metrics supported these interpretations. *RT* was longer for ΔL = 2000ft than ΔL = 3000ft [$F(1, 18) = 7.0$, $p = .016$, $\eta_p^2 = .28$], while *PVar* showed significant ΔS main effect [$F(2, 36) = 4.0$, $p = .027$, $\eta_p^2 = .18$], with larger *PVar* for ΔS = 10kn compared to ΔS = 30kn [$t(18) = 2.81$, $p = .012$, $r = .64$]. No ΔS or ΔL effects were found for *SUB*.

Therefore, the negative relationship between workload and ΔL was substantially confirmed. However, we did not find a positive relationship between workload and ΔS. Similarly to type A data, little (10kn) ΔS was perceived as a cognitive cost for lateral separation estimation compared to bigger values, determining higher visual processing demands for lateral separation estimation and, consequently, of workload. With ΔS 30kn or 70kn, participants felt more confident about lateral separation preservation. Similarly, participants performed lateral separation estimation when vertical LOS was deemed as more probable, i.e. with ΔL = 2000ft compared to ΔL = 3000ft. Such visuocognitive costs for separation estimation were observable in wider pupil dilation, as well as in the dramatic increment of P&Nd transition frequency. When conflict was deemed more probable pupil enlarged in preparation of finer visual information processing for aircraft motion abstraction.

In practice, ΔS = 10kn conditions in presence of equal distance to convergence center imposed a harder projection task and determined higher visual and cognitive activation, which was impossible to be captured by performance and subjective workload metrics, being off-line measurements. *PVar* data were coherent with gaze transitions data in showing differences in visual processing load. Pupil dilation in cognitive research has been intended as the need for visual exploration and scanning and it has been connected with mental effort and perceived cognitive complexity (Beatty, 1982). In this type of scenarios, higher pupil dilation denoted cognitive workload that arose after early RJ was made (detecting similar TTC and estimating both lateral and vertical LOS as highly probable). Whenever vertical LOS could be excluded after early processing of differential altitude (i.e. in ΔL = 3000ft scenarios), no further motion projection occurred.

In type F scenarios both lateral and vertical separation estimation was possible. N&N transition frequency changed as a function of ΔD and ΔL: ΔDxΔL interaction was significant [$F(4, 72) = 3.7$, $p = .008$, $\eta_p^2 = .17$], as represented in Figure 4. Simple effects analysis revealed a ΔL effect only with ΔD = 10nm [$F(2, 36) = 6.95$, $p = .003$, $\eta_p^2 = .28$], with higher N&N frequency when ΔL = 1000ft compared to 2000ft and 3000ft ($p = .028$ and $p = .003$, respectively). In such a scenario type, N&N transitions were fundamental for climb rate extrapolation and vertical separation estimation. Data showed significantly increased frequency for ΔL = 1000ft when ΔD = 10nm, i.e. when a lateral LOS was considered more probable, and conflict presence had to be ruled out also on the basis of vertical separation estimation. Differently, with ΔD = 20nm or 30nm participants easily excluded lateral LOS, they avoided to assess vertical separation and, consequently, no ΔL effect was found. Therefore, it seemed that participants had to project vertical separation variation in order to rule out conflict presence whenever a lateral LOS was considered as highly probable (as with ΔD = 10nm). Of note, no lateral LOS would have occurred in type F scenarios.

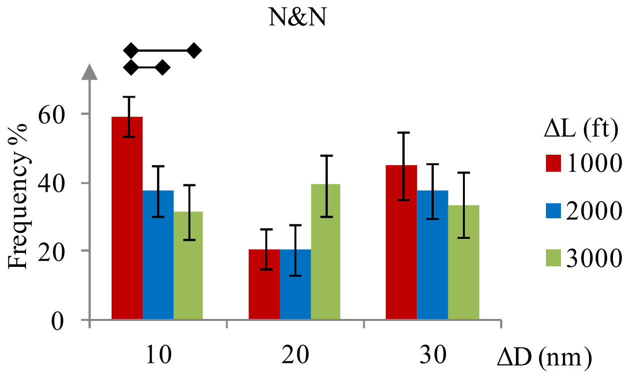

Figure 4. Effects of ΔD and ΔL on N&N transition frequency. Vertical bars indicate *SE;* horizontal lines indicate significant effects.

The increase of N&N frequency indicated vertical separation estimation performance: otherwise, minimal frequency would have been found, as in type D and B. Participants performed motion prediction with small ΔD or ΔL values (10nm and 1000ft, respectively) while with higher values conflict was easily ruled out on the basis of RJ exclusively, or a shallow separation estimation, without performing any motion prediction. Analysis of workload metrics for type F scenarios supported this interpretation, confirming negative relationship of workload with ΔD and ΔL.

As to *RT*, both main effects of ΔD and ΔL were significant [$F(2, 36) = 10.8$, $p < .001$, $\eta_p^2 = .38$, and $F(2, 36) = 6.71$, $p = .003$, $\eta_p^2 = .27$, respectively], qualified by a significant ΔDxΔL interaction [$F(4, 72) = 6.3$, $p = .003$, $\eta_p^2 = .26$], as reported in Figure 5. There was a ΔL effect with ΔD = 10nm [$F(2, 36) = 4.91$, $p = .025$, $\eta_p^2 = .21$] and ΔD = 20nm [$F(2, 36) = 11.7$, $p < .001$, $\eta_p^2 = .39$]. In the former, *RT* was longer for ΔL = 2000ft compared to ΔL = 3000ft ($p = .011$). In the latter, *RT* was longer for ΔL = 1000ft compared to ΔL = 2000ft and 3000ft ($p < .001$ and $p = .002$, respectively). This last result was confirmed in *SUB* data: both ΔD and ΔL main effects were significant [$F(2, 36) = 5.46$, $p = .008$, $\eta_p^2 = .23$, and $F(2, 36) = 7.74$, $p = .004$, $\eta_p^2 = .30$, respectively], qualified by a significant ΔDxΔL interaction [$F(4, 72) = 2.78$, $p = .034$, $\eta_p^2 = .13$]. There was a ΔL effect only with ΔD = 20nm [$F(2, 36) = 9.32$, $p = .001$, $\eta_p^2 = .34$], and higher *SUB* were found for ΔL = 1000ft compared to ΔL = 2000ft and 3000ft ($p < .001$ and $p = .006$, respectively).

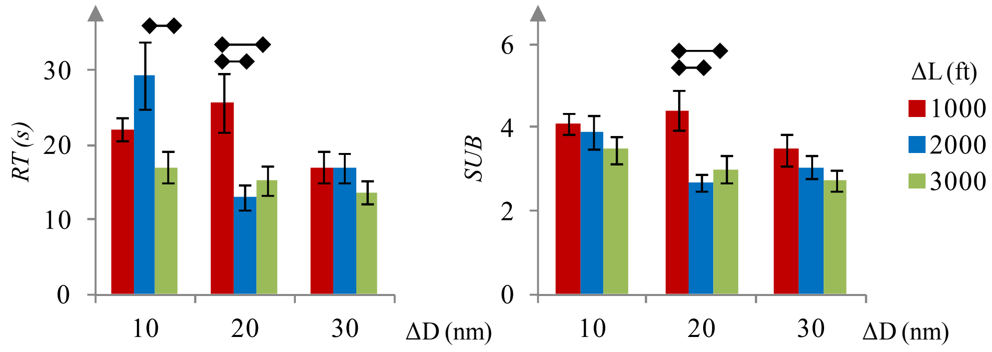

Figure 5. ΔDxΔL interaction in *RT* and *SUB* data for type F scenarios. Vertical bars mean *SE*; horizontal lines indicate significant effects.

Interestingly, *PVar* data showed an almost opposite trend. Only ΔD main effect was significant [$F(2, 36) = 5.77$, $p = .007$, $\eta_p^2 = .24$], due to larger dilations in ΔD = 30nm with respect to ΔD = 20nm [$t(18) = -2.97$, $p = .008$, $r = .57$]. Therefore, the largest pupil dilation occurred in easy situations (ΔD = 30nm and ΔL = 3000ft), in which conflicts were rapidly ruled out on the basis of early RJ, instead of performing subtle PM. We interpreted pupil size dilations in these situations as expression of arousal, or emotional activation, probably due to a feeling of confidence about decision once facilitators for heuristic conflict detection were detected, i.e. a considerable vertical separation (3000ft) that is preserved during most part of flight progress, or a considerable difference in distance to convergence point (e.g. 30nm), that strongly cues future lateral separation preservation. Such result is in part in opposition to workload-induced pupil dilations encountered for lateral projection strategy (type D and E). Participants might have raised activation in such easy conditions, so that emotion-driven pupil dilations were observed, which were even larger than cognitive-related dilations. Several researches evidenced how pupil dilation might be related to a number of psychological processes and emotional responses, including emotional processing (Lemercier, Guillot, Courcoux, Garrel, Baccino, & Schlich, 2014), drowsiness (Van Orden, Jung, & Makeig, 2000), target detection (Privitera, Renninger, Carney, Klein, & Aguilar, 2010), or general arousal (Bradshaw, 1967; Bradley, Miccoli, Escrig, & Lang, 2008).

In practice, in type F participants detected a future lateral LOS when ΔD = 10nm and, consequently, had to estimate vertical separation for conflict detection, experiencing higher workload and showing a considerable increase of transitions between aircraft data tags, in order to abstract climb rate and monitor vertical separation. Differently, with ΔD = 20nm, lateral LOS was less probable but deemed as still possible: participants experienced higher workload for projection in the lateral dimension only when ΔL = 1000ft, since probably higher uncertainty about vertical LOS was detected. Finally, altitude effect was absent with ΔD = 30nm as the combined effects of ΔD and ΔL on workload (negative relationship) allowed for quick answer about conflict absence. In summary, participants checked for vertical separation when ΔL = 1000ft or 2000ft, after having detected a highly probable lateral LOS, i.e. when ΔD = 10nm and, to a lesser extent, ΔD = 20nm. With ΔD = 30nm or ΔL = 3000ft separation estimation resulted unnecessary, after having heuristically predicted lateral and vertical separation preservation, respectively.

Lastly, in type C scenarios, participants had to necessarily perform vertical separation estimation. No transition types, nor workload indices, showed differences as a function of ΔL. Nevertheless, the negative relationship between workload and ΔL was supported by the high error rate when ΔL= 1000ft, due to the difficulty in projecting vertical separation beyond the moment in which one aircraft would have reached assigned altitude. Such a result might be in relation with prospective memory difficulties in ATC tasks (Loft, 2014; Loft, Smith, & Remington, 2013).

4. CONCLUSION

In this paper we manipulated conflict detection task demands in a simulated ATC environment and recorded ocular metrics: subsequent analysis of gaze transitions between task relevant areas of interests (AOI) was conducted in order to identify the strategies adopted during task completion. Furthermore, we measured workload by means of performance (response time), subjective (workload rating), and psychophysiological (pupil size variation) indexes. We manipulated aircraft differential values of speed, distance to convergence point, and altitude at scenario onset, determining different levels of traffic complexity and, consequently, of cognitive workload. The study main contribution is a preliminary and novel application to conflict detection task of gaze transition analysis for cognitive strategy identification and monitoring, together with a multidimensional assessment of workload fluctuations as a function of speed, distance, and altitude differential values.

When performing lateral separation estimation, it was observed a distribution of attentional resources on task relevant information for higher workload situations. Consecutive fixations on aircraft data tags almost disappeared, as they were not crucial for conflict detection in such situations. Similarly, it was observed how recurrent fixations between aircrafts and convergence center increased when finer differential distance and position processing were requested, independently of global workload for conflict detection, i.e. an effect on perceptual system. Differently, when checking for safe separation in both vertical and lateral dimension (as in type E), it was observed how crossed transitions between aircraft elements increased in frequency as long as workload increased. Such visual scanpath feature was connected with the cognitive cost for aircraft motion equations abstraction. Transition analyses allowed identification of recurrent gaze movements and the quantification of their relative weight within the whole scanpath. As a support, workload metrics allowed to couple transition data to workload levels. A more precise distinction between perceptual and cognitive load might be possible in future application of such a methodology.

We correctly predicted a negative relationship between workload and differential values of altitude and distance. Differently, the predicted positive relationship between workload and differential speed was found only in part. In fact, big differential speed was considered a workload-inducing factor only when aircraft did not presented an equal distance to convergence point, for which lateral LOS was more probable for little speed differences (10kn). In such situations, little differential speed was related to higher workload. We interpreted these results as the increased urgency for correct prediction when the difference in speed was not too big, depending on the estimated difference in TTC (time to convergence point). When differential TTC was perceived as little, higher workload was experienced, as a consequence of major cognitive resources investment. In contrast, situations with higher differences in TTC due to higher differential speed were probably interpreted as less urgent and resulted less demanding.

Differential distance facilitated projection as long as speeds were equal, or presented little differences. RJ in such situations was in fact based on perceptual discrimination skills, instead of on the computation of alphanumerical data. Discriminative power of visuocognitive system is higher for bigger differential distances, and diminishes (increasing visual workload) for little differential distances. In type D scenarios, where differential distance was counterbalanced by differential speed effects in a closer-is-slower paradigm, higher speed differences were related to major workload, and the higher was speed difference, the weaker was the facilitating effect of differential distance.

When differential distance was zero, TTC estimation had to be based on aircraft speed. Researchers have reported intrinsic difficulties with motion projection with different speed (Davison Reynolds, 2006). Our study confirmed such a cognitive limitation. In fact, lateral separation estimation with differential speed was particularly difficult. For instance, the considerable error rate in scenario A3, in which equal distance to convergence point was combined with a difference in speed of 70kn, demonstrated how difficult was to consider speed effects on a distance scale. In type E scenarios the same finding was encountered.

The cognitive costs of such difficulties and the consequent resources investment were observed in performance, subjective, and pupil variation metrics.

Interestingly, the biggest pupil dilations were observed in low workload conditions (i.e. easy scenarios, types B and F). Such results might be linked with arousal, i.e. with sympathetic activation, which would cause sensible pupil dilations. Such mydriatic reaction might have been limited by a parasympathetic coactivation in higher workload situations, which might have originated after a first activation (sympathetic) phase, once aware of the difficulty of conflict detection and the need for further attentional resources allocation. Sympathetic-parasympathetic coactivation has been related to emotional reactions, e.g. sadness (Kreibig, 2010) and it might have played a crucial role in high workload scenarios, whereas more time was needed for correct performance. In such situations, pupil dilations were smaller but significant (e.g. in type D scenarios). Differently, sympathetic reactions (pupil dilations) might have been dominant in easy trials (e.g. type F trials with considerable differential distance), in which a rapid, heuristic decision might have been made instead of motion projection, justifying a positive emotional reaction for the certainty about decision correctness.

Lastly, we observed almost systematic incorrect lateral separation estimation for big differential speed (70kn, type A) and small differential distance to convergence point (10nm, type B and D): lateral LOS was incorrectly predicted in such conditions, demonstrating the constraints of cognitive projection model. Similarly, omission errors (i.e. to consider as no conflict a future LOS) were observed in presence of little differential level (1000ft, type C and E): such errors were related to the difficulty for correct vertical projection once an aircraft would have changed altitude status (i.e. from "on climb" to "level"), and directly affected projection model quality.

In summary, transitions frequency analysis prompted fruitful insights about adopted strategies for correct motion projection and spatial regions of major investment of cognitive resources. Such insights were supported by a multidimensional assessment of workload. Moreover, we observed a large, arousal-related pupil dilation (in presence of low workload), distinct from a less prominent, workload-related pupil dilation. Pupil size variation is a useful online index of cognitive resources investment, able to capture situational workload flows in absence of visible effects on performance. Since pupil reactions might be due to a number of cognitive, emotional, and environmental factors, important distinctions might be made when analyzing pupil signals between workload responses and other reactions.

The approach presented in this study might be useful to workload researchers in different ways. Analysis of transitions might considerably benefit ATC training, for instance by further contributing to experts' mapping of attentional resources allocation (Kang & Landry, 2014), or to support retrospective verbal protocols in post hoc analysis of trainees' performance. The mapping of attentional and cognitive resources allocation by means of eye tracking might also positively impact the design of error-preventing technologies tailored on operator's psychophysical state.

ACKNOWLEDGEMENTS

This work was supported by Spanish Ministry of Economy and Competitiveness, PSI2012-39246. We thank all the volunteers for their time and effort. The authors have no conflicts of interests to declare.

REFERENCES

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623-636.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye Movements Do Not Reflect Retrieval Processes Limits of the Eye-Mind Hypothesis. *Psychological Science*, *15*(4), 225-231.

Averty, P., Collet, C., Dittmar, A., Athènes, S., & Vernet-Maury, E. (2004). Mental workload in air traffic control: an index constructed from field tests. *Aviation, space, and environmental medicine*, *75*(4), 333-341.

Averty, P., Guittet, K., & Lezaud, P. (2008). An ordered logit model of air traffic controllers conflict risk judgment. *Air Traffic Control Quarterly*, *16*(2), 101-132.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276–292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge, UK: Cambridge University Press.

Bisseret, A. (1981). Application of signal detection theory to decision making in supervisory control. The effect of the operator's experience. *Ergonomics*, *24*(2), 81-94.

Boag, C., Neal, A., Loft, S., & Halford, G. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics, 49*(14)*,* 1508–1526.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602-607.

Bradshaw, J. L. (1967). Pupil size as a measure of arousal during information processing. *Nature*, *216*, 515–516.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of cognitive neuroscience*, *9*(1), 27-38.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology*, *42*(3), 361-377.

Chen, S., & Epps, J. (2014). Using Task-Induced Pupil Diameter and Blink Rate to Infer Cognitive Load. *Human–Computer Interaction*, *29*(4), 390-413.

Davison Reynolds, H. (2006, February). *Modeling the air traffic controller's cognitive projection process* (Report No. ICAT-2006-1). Cambridge, MA: MIT International Center for Air Transportation.

Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, *16*(6), 330-333.

Eyferth, K., Niessen, C., & Spaeth, O. (2003). A model of air traffic controllers' conflict detection and conflict resolution. *Aerospace science and technology*, *7*(6), 409-416.

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, *9*(2), 1-6.

Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab$^{\text{Advanced}}$: An air traffic control simulator with realism and control. *Behavior research methods*, *41*(1), 118-127.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: guiding attention guides thought. *Psychological Science*, *14*(5), 462-466.

Hilburn, B. (2004). Cognitive complexity in air traffic control: A literature review. *EEC note*, *4*(04). Bretigny-sur-Orge: Eurocontrol Experimental Research Center.

Histon, J. M., & Hansman, R. J. (2002). *The impact of structure on cognitive complexity in air traffic control* (Rep. No. ICAT-2002-4). Cambridge, MA: MIT International Center for Air Transportation.

Hunter, A. C., & Parush, A. (2009). Using eye movements to uncover conflict detection strategies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 22, pp. 1729-1733). Sage Publications.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441-480.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*, 137–172.

Kallus, K. W., Van Damme, D., & Dittman, A. (1999). *Integrated job and task analysis of air traffic controllers*: *Phase 2. Task analysis of en-route controllers* (European Air Traffic Management Programme Rep. No. HUM. ET1. ST01.1000-REP-04). Brussels, Belgium: Eurocontrol.

Kang, Z., & Landry, S. J. (2010). Capturing and analyzing visual groupings of multiple moving targets in an aircraft conflict detection task using eye movements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 23, pp. 1906-1910). SAGE Publications.

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors*, *56*(6), 1150-1162.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, *84*(3), 394-421.

Langan-Fox, J., Sankey, M. J., & Canty, J. M. (2009). Human factors measurement for future air traffic control systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(5), 595-637.

Lemercier, A., Guillot, G., Courcoux, P., Garrel, C., Baccino , P., & Schlich, P. (2014). Pupillometry of taste: Methodological guide – from acquisition to data processing - and toolbox for MATLAB. *The Quantitative Methods for Psychology*, *10*(2), 179-195.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modelling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors*, *49*, 376–399.

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air traffic control: incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, *15*(2), 106-124.

Loft, S. (2014). Applying Psychological Science to Examine Prospective Memory in Simulated Air Traffic Control. *Current Directions in Psychological Science*, *23*(5), 326-331.

Loft, S., Smith, R. E., & Remington, R. W. (2013). Minimizing the disruptive effects of prospective memory in simulated air traffic control. *Journal of Experimental Psychology: Applied*, *19*(3), 254–265.

Majumdar, A., & Ochieng, W. Y. (2002). Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload. *Transportation Research Record: Journal of the Transportation Research Board*, *1788*(1), 58-69.

Marchitto, M., Di Stasi, L. L., & Cañas, J. J. (2012). Ocular movements under taskload manipulations: Influence of geometry on saccades in air traffic control simulated tasks. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*(5), 407–419.

Neal, A., & Kwantes, P. J. (2009). An evidence accumulation model for conflict detection performance in a simulated air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(2), 164-180.

Nunes, A., & Mogford, R. H. (2003, October). Identifying controller strategies that support the ‘picture’. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 1, pp. 71-75). SAGE Publications.

Poole, A., & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future. In C. Ghaoui (Ed.), *Encyclopedia of Human-Computer Interaction* (211-219). Pennsylvania: Idea Group, Inc.

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision*, *10*(10):3, 1-14.

Rantanen, E. (2004). *Development and validation of objective performance and workload measures in air traffic control*. (Report No. AHFD-04-19/FAA-04-7). Oklahoma City, OK: Federal Aviation Administration Civil Aeromedical Institute.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical conflict detection in air traffic control. *The International Journal of Aviation Psychology*, *15*(4), 339-362.

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, *22*, 204–207.

Sperandio, J. C. (1978). The regulation of working methods as a function of work-load among air traffic controllers. *Ergonomics*, *21*(3), 195-202.

Stankovic, S., Raufaste, É., & Averty, P. (2008). Determinants of conflict detection: A model of risk judgments in air traffic control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(1), 121-134.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: U. S. Department of Transportation, Federal Aviation Administration Technical Center.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic bulletin & review*, *14*(4), 663-668.

Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation: analysis of prediction-motion and relative judgment tasks. *Perception & Psychophysics*, *57*(2), 231-245.

Van Orden, K. F., Jung, T. P., & Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological psychology*, *52*(3), 221-240.

Viviani, P. (1989). Eye movements in visual search: cognitive, perceptual and motor control aspects. *Reviews of oculomotor research*, *4*, 353-393.

Willems, B., Allen, R. C., Stein, E. S. (1999). Air traffic control specialist visual scanning II: task load, visual noise, and intrusions into controlled airspace, Technical Note (DOT/FAA/CT-TN99/23), FAA.

World Medical Association. (2013). Declaration of Helsinki, ethical principles for medical research involving human subjects. *64th WMA General Assembly, Fortaleza, Brazil.*

Xu, X., & Rantanen, E. M. (2003). Conflict detection in air traffic control: A task analysis, a literature review, and a need for further research. In R. S. Jensen (Ed.), *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 1289–1295). Dayton, OH: Wright State University Press.

CHAPTER 6

CONCLUSION, PRACTICAL IMPLICATIONS, AND FUTURE RESEARCH

CHAPTER 6

CONCLUSION, PRACTICAL IMPLICATIONS AND FUTURE REASEARCH

1. INTRODUCTION

Cognitive complexity refers to the perceived level of complexity at the individual level (Cummings & Tsonis, 2005) while accomplishing specific task requirements. Cognitive complexity is determined by objective task requirements (taskload), context characteristics (environmental and organizational), and by personal characteristics (psychophysiological state) (Hart & Staveland, 1988). In ATC domain task requirements can vary due to numerous complexity factors, which refer to geometry, traffic dynamics, organizational flight rules, or display's interface. Cognitive processes (attention, perception, memory, recognition) mitigate the influence of complexity factors on perceived cognitive complexity. For instance, several cognitive strategies might be available for conflict detection, all aimed at reducing perceived cognitive complexity and reducing experienced workload. Workload is the effect of using less (limited) mental resources than those demanded by the complexity of the task to accomplish it, and could be suboptimal (underload), optimal, or even become excessive (overload).

A human-centered approach to complexity focuses on the interaction between task characteristics and cognitive activity, in order to disentangle cognitive strategies, strategy shifts, and etiology of experienced workload. Online methods for workload assessment are highly requested, as they could identify high workload situations in absence of performance decrements, which are most difficult to observe in applied settings. Psychophysiological measures, in particular eye metrics, are promising methods for future online cognitive workload assessment, as they are reaching considerable levels of portability and unobtrusiveness in recent years.

This thesis presents a series of experiments that showed sensitivity of specific eye metrics to variations in ATC conflict detection task complexity, and that used gaze analysis to identify adopted cognitive strategies. **Chapter 6** resumes and discusses experimental results, provides some practical implications of presented studies, and indicates some directions for future research.

2. ATC CONFLICT DETECTION TASK, WORKLOAD, AND OCULAR MEASURES

The research line of this thesis started (Chapter 2) with testing ocular measures sensitivity to task complexity (TC) and time on task (TOT). We created a simple decision task about target position, and added further tasks to create a difficult condition (a paper-and-pencil task and a simple math operation). Expected positive effects of TC on workload ratings and response times, and of TOT on fatigue/sleepiness scores were found. Differently, we encountered shorter response times in the second half of the experiment, especially for the difficult task.

We provided two complementary explanations for such performance optimization effect in the difficult task. First, a better visual strategy might have been adopted in difficult task, in which besides perceptual decision main task, aircraft callsign (3 digits) codification was requested to accomplish additional tasks. Increased visual task requirements might have been optimized through time. Differently, short response times in easy condition were maintained throughout experiment, without further possible improvements. Second, structural aspects of experimental tasks were considered. Low complexity condition only consisted of visual task, with no information about remaining trials before session's end. Differently, in high complexity condition, writing and math tasks were performed on a printed sheet, informing participants about number of remaining trials. Consequently, an activation effect in the difficult task might have generated performance optimization. Participants were probably more activated towards the end of the session, speeding up response times, and finally providing higher fatigue ratings.

Ocular data supported the interpretation performance results. Saccadic peak velocity and duration were expected to present lower values for increased task complexity. We encountered a slow down effect of saccadic peak velocity for low complexity task, compared to high complexity task. Such an effect was stronger in the first part of the experimental session, and involved exclusively saccades wider than 11°. Workload is optimal when accomplishing an optimal number of tasks (quantitative load) or tasks of optimal complexity (qualitative load). If time to perform the tasks becomes unavailable, task requirements exceed operators' cognitive resources availability, and overload conditions (quantitative or quantitative) might be experienced (Díaz Cabrera, Hernández Fernaud, & Rolo González, 2012). Conversely, underload situations are characterized by great time availability to perform a reduced number of tasks, or very low complexity tasks (or both). Lack of feedback, or poor number of alternatives for decision making are examples of underload drivers (Ferrer & Dalmau, 2004). In operational contexts underload effects are safety-critical as well as overload effects. Operators might be "out-of-the-loop" and last more time to take control of the situation in underload situations (Bainbridge, 1983).

It is acknowledged since more than a century that medium arousal (as a consequence of medium task demands) are related to optimal workload and performance, while excessive or excessively low arousal levels relate to performance decrements (Yerkes & Dodson, 1908). Therefore, a visual task in an underload condition might reveal extremely boring, with minimal employment of visual cognitive resources. Low complexity task in this study consisted of only a simple visual decision task, with sufficient available time for performance and lack of temporal feedback on session duration. Such features are typical of quantitative and qualitative underload situations. The slow down effect in PV of large saccades was found for an underload situation, while optimal workload did not produce any significant modification of main sequence parameters. The effect was present only in the first block, in which underload situations might have been more affecting, due to great resources availability (the same trend was present in the second block without reaching significance). Results also suggested that longer experimental blocks were needed to observe fatigue effects: fatigue onset might take more time when dealing with simple tasks and sufficient time to accomplish them. Previous works suggested that fatigue effects on saccadic peak velocity were observable after at least 60 minutes, although such time interval was related to tasks very different from simulated ATC, like driving (Di Stasi, Renner, Staehr, Helmert, Velichkovsky, Cañas, Catena, & Pannasch, 2010), or problem solving in microworld simulations (Di Stasi, Antolí, & Cañas, 2011).

All these results obliged to reconsider task complexity manipulation. Acknowledged reference literature on ATC complexity factors was studied to manipulate taskload (i.e. task demands) more effectively, determining different levels of experienced workload and testing ocular metrics sensitivity to workload variations. Ecological validity, replication, and application of results would benefit from more plausible simulations of real operational tasks. Therefore, a series of logical and methodological decisions were made for a better approach to workload assessment as a consequence of complexity variations in the ATC tasks.

First, we used a flight traffic simulator with a sufficient degree of realism in order to generate appropriate cognitive engagement and realistic traffic scenarios. Second, we manipulated acknowledged complexity factors in scientific literature about ATC domain in order to vary task complexity in a more reliable fashion. Third, we adopted conflict detection task, since it is a key component of ATC released service. Controllers often described their job as "to separate aircraft" (Averty, 2005), to indicate conflict detection (and resolution) resulting from aircraft trajectories as the fundamental part of safe traffic control. Last, dealing with naïve subjects, appropriate training sessions and briefing materials should have been arranged and tested before running data collection. Therefore, in successive works we focused on complexity of conflict detection task and multidimensional assessment of workload with specific attention on ocular metrics.

The study presented in Chapter 3 dealt with perceived cognitive complexity as a function of traffic geometry. We used a basic subtask of conflict detection, which consisted of inferring which aircraft would have arrived first to convergence point (RJ, relative judgment), presenting static scenarios (i.e. a screenshot of the traffic scene) with two converging aircrafts. Convergence angle and aircrafts' distance to convergence point were the manipulated geometry complexity factors. Correct RJ is fundamental to assess future aircraft relative positions and separation, i.e. to detect conflict. Workload was assessed in relation to RJ performance with lateral problems (vertical dimension was not included at this stage). Saccadic peak velocity (PV) was adopted as psychophysiological measure for assessment of experienced workload. Geometry complexity of traffic scenarios might increase perceptual load, as elements' spatial disposition could make distances perception and comparison more difficult.

In parallel, the mediating influence of cognitive strategies was limited. Participants were trained on a specific strategy to perform decision making. A simple speed comparison for aircrafts was sufficient for RJ with equal distance to convergence point. Differently, in case of different distances to convergence point, participants had to transform one speed value (calculating a proportion) or compute times to convergence point (by means of distance-to-speed ratios) before comparing them.

Taskload variations (cognitive demands) successfully produced workload variations. We therefore observed changes in psychophysiological activity due to traffic geometry manipulations. Geometrical complexity affected attentional and perceptual requirements, determining workload variations detected by saccadic metrics of reaching saccades (>15°), which showed a slow down effect. We explained these differences as effect of a speed/accuracy tradeoff in favor of accuracy when complexity increased, with lower saccadic peak velocity and reduced positional error. Furthermore, we found longer response times when further processing was requested and for increasing convergence angles, confirming workload effects of geometrical complexity. Recent investigations disentangled effects of perceptual load from those of cognitive load, on the basis of performance and psychophysiological measures (Jesse, 2010). The distinction of these two constructs on the basis of empirical evidences based on ocular metrics could positively impact interface design as well as functions' allocation for adaptive assistance technology design (e.g. automatic conflict alert tools).

The study in Chapter 4 continued the investigation about cognitive complexity and workload experienced during conflict detection, by manipulating again a geometry complexity factor (CA, convergence angle), together with a traffic complexity factor (MD, minimum distance at closest approach). In fact, dynamic traffic scenarios (videos instead of screenshots) were adopted, allowing manipulation of traffic dynamics on a spatial and temporal basis: we set desired separation and fixed a time to closest approach, held constant in all scenarios. MD could be below (conflict) or above (no conflict) safe separation standard. Importantly, participants needed to be trained on basic traffic dynamics (e.g. the relation between speed and altitude, and the consequences on climb rate), simulation display, and safe aircraft separation. A graphical reference of separation standard was also included in simulation display (the radius of the circle around convergence point).

Relative judgment (RJ) performance was only a preliminary step to conflict detection. Participants had to effectively abstract aircraft motion equations on the basis of position updates (5s period), and combine them into an integrated cognitive model of motion projections to estimate future lateral separation. Wide CA were chosen in order to maintain a considerable level of geometry complexity. We assumed that correct conflict detection would have been easier with higher differences between MD and safe separation standard (5nm). Conversely, conflict detection was assumed to be more difficult when MD was very close to safety standard (e.g. 6nm). In practice, we assumed PM (prediction motion) as constant strategy for conflict detection.

On the basis of previous works, conditions that were assumed as more complex (MD = 6nm) were expected to show higher workload, i.e. higher error-rates and subjective workload ratings, longer response times, slower peak velocities of large saccades, and increased fixation duration. Increasing CA (geometry complexity) was expected to increase saccadic amplitude, duration, and velocity. Since late responses in conflict detection task were related to higher cognitive effort and experienced workload, complexity effects on number of ocular events (saccades and fixations), subjective ratings, fixation duration, saccadic metrics, and pupil size were analyzed controlling for response time.

Correlational study confirmed the association between ocular, performance, and subjective metrics. In particular, results suggested two ocular behaviors for task solution: detailed aircrafts position observation and frequent comparison (with longer fixations, many large saccades, lower saccadic and fixation frequency, higher subjective workload, and lower accuracy), vs. rapid aircraft localization and flight metrics comparison (with shorter fixations, few large saccades, higher fixation and saccadic frequency, lower subjective workload, and higher accuracy).

Performance, subjective, and ocular measures consistently showed conflicts as more demanding, and further experienced workload with perpendicular routes, which increased geometry complexity. In fact, large saccades confirmed to be sensitive to workload induced by cognitive complexity. Results showed lower saccadic duration and peak velocity of large saccades (>15°) for conflicts, especially with perpendicular routes. Therefore, large ocular movements adapted to increased perceptual and cognitive task requirements, being slower and, most probably, more precise. Reduced positional error might be the trade-off for such a reduction of saccadic peak velocity. Saccades are ballistic movements: their amplitude is coded before performance, while duration and peak velocity are determined by motor neurons activity. Saccadic duration is composed of two phases, an acceleration phase that lasts up to reaching saccadic peak velocity, and a deceleration phase for completing the movement. Firing activity of motor neurons is observed in the acceleration phase (Bahill & Stark, 1975). Therefore, saccadic peak velocity is reached approximately at half the duration of the saccade, at least for small saccades which present a symmetric velocity profile (Leigh & Zee, 1999). Large saccades present a skewed velocity profile, with acceleration phase that lasts about 1/3 of saccadic duration (Lin, Chen, Chen, Tsai, & Huang, 2004). In fact, in order to realize large saccades, burst neurons in the pontine reticular formation have to increase considerably their firing rate and duration of firing signal up to reaching asymptotic performance, beyond which no further improvements in neural signal transmission might occur (Zils, Sprenger, Heide, Born, & Gais, 2005). When realizing saccades of maximum amplitude, nearly all motor neurons are firing in synchrony at the top of their capacity. Muñoz & Everling (2004) suggested that variations in attentional processing affect excitatory signals from frontal cortical areas to pontine reticular formation, which contains burst neurons. Large saccades therefore showed reduced burst power, i.e. lower velocities, when perceived complexity increased.

There were no effects of CA and MD on fixation parameters (number and duration) and saccadic number, once response time effects were removed. Fixation metrics could be predicted by response times, substantially confirming eye-mind relationship for such a cognitive task. Therefore, in operational settings, whereas no response times can be measured, most fixated elements are most probably those which cognition is most engaged with. Fixation durations progressively increased as more attention was allocated to such elements. Therefore, fixation parameters showed as valid workload metrics. Lastly, saccadic amplitude increased with CA, a predicted geometry complexity effect (almost 60% of variance in saccadic amplitude was explained by CA).

In summary, the slow down effect of large saccades was consistent, as found in previous works. Saccadic peak velocity was lower in presence of higher cognitive complexity, i.e. in conflicts, especially with perpendicular routes. Saccadic duration showed lower values in the same conditions. Differences in duration and velocity were connected to cognitive complexity of conflict detection, and involved large saccades, exclusively. Therefore, duration and peak velocity of large saccades were confirmed as oculomotor indexes sensitive to applied cognitive effort.

On the basis of such strong distinctions between conflicts and no conflicts showed by performance, fixation, and saccadic metrics, qualitative differences in visual and cognitive processing were hypothesized, as expression of different strategies adopted for conflict detection task. In no conflicts, few fixations and saccades, short decision times, and low subjective ratings suggested the adoption of a strategy for rapid assessment of flight data. We hypothesized the adoption of a ratio-based strategy (e.g. distance-to-velocity ratio, D-to-V) in no conflict trials, as the estimated times to convergence point were considerably different, suggesting safe lateral separation.

Differently, we thought that trajectory prediction in conflict trials was performed on the basis of aircraft updating position processing and of motion projection. Cognitive motion projection was more time consuming and demanding with respect to D-to-V strategy, which might have been unviable in conflict trials, producing a strategy shift in favor of a projection motion-based strategy. The difference in estimated times to convergence point might have been perceived as insufficient for safe lateral separation. Recurrent processing of aircraft position was necessary for correct motion projection. Therefore, all further fixations and saccades that were performed in conflict trials with respect to no conflicts were most probably related to position processing.

Differences between aircraft speed and distance to convergence point at trial start were identified as possible drivers for such strategy shift. In conflicts little differential distances were combined with considerable differential speeds in a closer-is-slower fashion, resulting in increased workload for lateral separation estimation. Differently, in no conflicts, differential speeds were limited and differential distances were considerable, suggesting different times to convergence point, i.e. a higher probability of safe lateral separation.

Conflict detection is fundamentally a matter of relation and of aircraft flight progress comparisons. Operators recurrently compute aircraft position and flight metrics differences to estimate times or distances. Strategy adoption and eventual strategy shifts for motion projection and separation estimation could be assumed to be a consequence of relative judgments and the feeling of uncertainty associated to them. A relative judgment consists of aircraft comparisons in relation to a specific flight metric. Examples of basic relative judgments are, for instance, to detect the aircraft closer to convergence point, or the slowest one. Such basic relative judgments were considered as the basis for more complex relative judgments, such estimation of order and times of arrival to convergence point, aircraft relative position at closest approach, minimum lateral (or vertical) separation, all leading to final conflict judgment. In substance, we assumed that complexity of conflict judgment depends on relative judgments and estimation (or computation) of speed, distance, and altitude differential values.

In Chapter 5 study we manipulated differential values of flight metrics as cognitive complexity drivers. Differential speed, distance, and altitude were manipulated alone (scenario types A, B, and C in the study, respectively), or in combination (speed and distance in type D, speed and altitude in type E, and distance and altitude in type F scenarios). In addition, we analyzed gaze transitions between task relevant areas of interest (AOI) and tested pupil size as psychophysiological workload index in conflict detection, together with performance metrics and subjective workload ratings.

The application of gaze transition analysis to conflict detection task was an innovative contribution. The study of visual scanpath in ATC conflict detection task is a relatively new application of ocular metrics recording in workload research. Results available in scientific literature (Hunter & Parush, 2009; Kang & Landry, 2014) inferred cognitive strategy on the basis of observational analysis of scanpaths data, in the form of experts' judgments, or as off line verbal protocol about recorded data. Therefore, we proposed a quantitative analysis of gaze transitions between task relevant AOIs as a more objective method to analyze visual scanpaths in ATC conflict detection task and to identify adopted strategies.

An increase in frequency of a specific transition within general scanpath was assumed to indicate an increased importance of such visual (micro)behavior for task resolution. Cognitive strategy adopted for task resolution will be based on, or strongly characterized by, this transition type. Vice versa, a significant decrement, or even a null frequency, was assumed to indicate low importance within the strategy adopted. In presence of increased workload, focused visual behavior could limit the performance of low relevance gaze transitions.

Effectively, for lateral separation estimation (in scenario types B and D) we observed an almost null frequency of transitions between aircraft data tags (labeled "N&N" in the study) in high workload conditions. Data tags were showing static information (speed and altitude values) in those conditions. Attentional focus and gaze were removed from low relevance areas when perceived complexity increased, as revealed by the highest response times, subjective ratings, and pupil dilations. In parallel, we found an increment of transitions between aircraft and convergence center (labeled "PN&C") in situations that requested finer processing of aircraft (little) differential distance, even if conflict detection workload was low (for instance, when little differential speed will change differential distance minimally throughout flight progress, easing separation estimation).

Importantly, gaze transitions showed an increase of crossed transitions between aircrafts' different elements (labeled "P&Nd"), in presence of little differential speed (in type A scenarios). The same occurred when checking for vertical separation also (combined differential speed and altitude manipulation, type E scenarios). Little speed differences produced subtler changes in differential distance, which were harder to be assessed by the perceptual-cognitive system. In parallel, when vertical separation was also checked, such transition type resulted even more frequent. Higher frequency of P&Nd transitions was interpreted as the cost for lateral motion projection and separation estimation while, in parallel, checking for vertical separation maintenance. Response times were longer and pupil dilation larger when crossed transitions between aircrafts' elements increased in frequency. Pupil enlarged for finer aircraft position processing and motion abstraction when conflict was estimated as more probable.

As predicted, workload for lateral separation estimation increased when big differential speeds and distances were combined (type D scenarios): changes in differential distance throughout flight progress were bigger with big differential speed, making relative position estimation more difficult. With same distance to convergence point (as in scenario types A and E, which manipulated speed alone, and speed and distance, respectively), big differential speed conditions were particularly error-prone, denoting the difficulty for relative motion projection and lateral estimation. However, analysis of correct answers showed higher resources investment (larger pupil dilations) for little differential speed conditions. These situations were probably perceived as more urgent, as lateral loss of separation (LOS) was deemed as more probable. The aforementioned increase of crossed transitions between aircraft elements was expression of such perceived difficulty.

The expected negative relation between differential distance to convergence point and workload was also confirmed (except when combined with big differential speeds, as said). Larger differential distances were perceptually more evident, leading to confident no conflict judgments. Actually, no motion projection was probably performed for conflict detection with large differential distance (i.e. 30nm). A loose estimation of difference in times to convergence point was sufficient for a confident no conflict judgment. Similarly, no motion prediction was performed with large differential altitudes (i.e. 3000ft), e.g. in type F. By contrast, when conflict presence could not be rapidly excluded, we observed a dramatic increment of gaze transitions between aircraft data tags ("N&N"), as indicating vertical separation estimation.

Importantly, we found the biggest pupil variations for type F scenarios, especially for very easy situations. Increased arousal and positive emotional state in a situation with low perceived cognitive complexity (low response times, error rates, and subjective ratings) were probably responsible for such pupil dilations. Besides sustained workload-induced pupil dilations (accompanied by increase of N&N transitions frequency), we found an even more pronounced emotionally-induced pupil dilation in low workload scenarios, in which conflict presence was easily ruled out after rapid exploration. High confidence about the answer probably triggered positive emotional state and big pupil dilations.

Pupil size variation showed workload effects that were not captured by performance and subjective metrics. For instance, big differential speed was a workload factor, except when combined with large differential distances (type D scenarios), which would have suggested lateral separation preservation without performing demanding motion projections. However, big pupil dilations with high differential speed occurred also with large differential distances. This effect was interpreted as the cognitive cost for comparing aircrafts' flight progress at scenario onset. Comparison of aircraft moving forward is necessary to obtain a measure of relative displacement, to be projected for separation estimation. Big differential speed maximized the difference between aircrafts' flight progresses, requiring a finer motion processing, at least at scenario onset, observed in by pupil dilations.

3. PRACTICAL IMPLICATIONS

Plenty of investigations have accumulated empirical evidences that relate ocular indices to perceived complexity, applied cognitive effort, and also stress experienced during performance of several experimental tasks (Di Stasi, Marchitto, Antolí, & Cañas, 2013; Pedrotti, et al., 2014) and, more rarely, real work operations. Therefore, ocular metrics measurement in ATC domain could deserve more attention in the future, as its practical contributions might be several.

Importantly, ocular metrics benefit workload research by both allowing quantitative assessment (analyzing indexes of mental resources investment), and qualitative assessment (inferring cognitive strategies from the analysis of visual scanpaths). For instance, recent investigations used visual scanpaths to describe and classify cognitive strategies adopted during conflict detection on the basis of observational techniques (visual inspection), by means of both experts' judgments and self-reported written notes on performed visual behavior (Kang & Landry, 2014; Kang, Bass, & Lee, 2014). Such material proved very useful for training purposes, as novices benefited from experts' scanpaths exposure. Benefits obtained during training from visual scanpaths could be amplified by gaze transitions analysis, as we proposed, contributing both to experts and trainees to identify ocular patterns within the whole scanpath that characterize optimal performance, or strategy shifts. The minimization of resources needed to reach a satisfactory level of knowledge and practice is a fundamental practical aspect of aviation training (Wise, Hopkin, & Garland, 2009). Eye metrics and gaze transitions analysis might help to reduce training cost and improve benefits, i.e. training effectiveness.

Ocular metrics confirmed as valid indexes for monitoring workload flows, as well as detecting under- and overload situations. Velocity of large saccades, pupil size, and fixation parameters confirmed to vary with mental resources investment. Considering that ocular metrics are on line measures, resource consumption and attended information could be analyzed *while* cognitively processing traffic information, tracing the whole decision making process that ends with conflict judgment. The emergence of conflict judgments can be studied as a function of time by manipulating timespan for required conflict judgment and recording uncertainty associated to judgment (Averty, 2009). The application of eye tracking recording while performing visual behavior for conflict judgment could be applied to the study of judgment evolution, by informing about attended regions and current strategy.

The aviation domain is in an expansion phase in recent years, and traffic is predicted to increase considerably in the next future (ICAO, 2012). ATC controllers' workload has been reported as the most important constrain to such increment of airspace capacity. New paradigms (often called "free flight") for ATC are being investigated in the last years, which would transfer active control and separation processes from controllers to pilots. ATC will be transformed from active control and commands issuing, into more passive monitoring. Reliable multidimensional workload assessment for detecting both under- and overload situations is needed in research dealing with new ATC paradigms.

Lastly, measurement of ocular metrics for workload and cognitive strategy assessment could contribute enormously to the design of conflict detection support tools. Conflict alert systems exist that can alert controllers about conflict presence by means of projecting aircraft trajectories (Nolan, 2010). Nevertheless, the only information available to technology for trajectories prediction is aircraft flight past history. Since alert devices are unaware of controller's intentions, they might results in routinely false alerts, e.g. when a controller has planned to issue a turn command to one of two converging aircraft in the next future. Ocular metrics measurement relates to cognitive strategy being implemented, and could help to pave the way for designing strategy-aware alert technologies, that use situational cognitive information to provide better and timely assistance.

4. CONCLUSIONS

The assessment of cognitive workload experienced while performing highly demanding tasks in a dynamic environment is a safety-critical aspect of complex sociotechnical systems, like transportation, heavy industry, or healthcare systems. Undoubtedly, in the aviation domain, piloting and air traffic control (ATC) are the sectors that most needed to reliably model operator's cognitive performance and to monitor experienced workload for training, operational, and design purposes. Workload measurement is performed by means of performance, subjective, or psychophysiological metrics. Among the latter, ocular metrics might receive more attention in the future, as eye tracking technologies have reached in the last years considerable portability, unobtrusiveness, and ease of use.

The studies presented in this thesis aimed to provide further validity to ocular metrics as workload indexes in simulated air traffic control (ATC) tasks. We modeled conflict detection task and varied task complexity through manipulation of acknowledged traffic and geometry complexity factors, which constitute workload drivers for ATC operators. We measured main sequence parameters (i.e. saccadic amplitude, peak velocity, and duration), fixation parameters, pupil size variation, and gaze transitions between task-relevant elements.

Results showed that saccadic dynamics were related to mental activity and experienced workload, therefore contained information about subject visual-cognitive effort in conditions of different complexity. Saccadic peak velocity confirmed to be sensitive to geometry and cognitive complexity (**Chapter 3**), and to be a reliable workload index, showing a consistent slow down effect for increased workload, as well as in underload situations (**Chapter 2**).

Furthermore, we could find important insights from eye metrics analysis in relation to adopted strategy for conflict judgments (**Chapter 4** and **Chapter 5**). Both ratio-based and motion prediction-based strategies were identified. The latter were associated to higher workload, as they were adopted in more complex situations, when the former resulted insufficient for a confident decision. Ocular metrics revealed how quick and confident comparison of flight metrics in low workload situations was replaced by recurrent aircraft position processing and motion projections of relative movement for separation estimation.

Finally, we presented a preliminary application of gaze transition analysis from visual scanpaths (**Chapter 5**). We collected gaze data in relation to dynamic areas of interests (AOI) that evolved with flight progress. Analysis of gaze transitions between task-relevant AOIs allowed evidencing variations of specific ocular patterns as a function of task complexity and workload and supporting strategies' inference. Furthermore, the use of flight metrics differential values as complexity factors allowed identification of ideal thresholds for strategy shifts during conflict detection resolution task. In fact, motion prediction was not performed for large differential distance and altitude (30nm and 3000ft, respectively).

In summary, our results supported the application of eye tracking methods for workload assessment in ATC tasks, although representing partial and preliminary applications. Future work will be needed to further support these outcomes, addressing some critical issues, both practical and methodological. First, ocular metrics should be measured in scenarios with higher traffic loads, computing gaze metrics (total gaze time, transitions, and gaze revisits) in relation to attended aircrafts. Second, different temporal bases (e.g. hours, shifts, or consecutive sessions) should be considered for ocular metrics analysis, so that reliable fatigue effects due to time on task could be distinguished. Third, a deeper investigation of saccades should focus also on microsaccades (<1°), as recent investigations have observed a workload-induced slow down effect on microsaccades (Di Stasi, McCamy, Catena, Macknik, Cañas, & Martinez-Conde, 2013). In real work settings, visual scanning on visual display might be limited to limited areas (within 20°), generating smaller magnitude saccades. Last, strategy identification based on gaze transitions analysis will need powerful interpretation frameworks when applied to more complex traffic scenarios.

REFERENCES

Averty, P. (2005). Conflict perception by ATCS admits doubt but not inconsistency. In *Proceedings of the 6th Air Traffic Management Research & Development Seminar*. Baltimore, MD.

Bahill, A. T., & Stark, L. (1975). The high-frequency burst of motoneuronal activity lasts about half the duration of saccadic eye movements. *Mathematical Biosciences*, *26*(3), 319-323.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3), 191-204.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775-779.

Cummings, M. L., & Tsonis, C. (2005, September). Deconstructing complexity in air traffic control. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 1, pp. 25-29). SAGE Publications.

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2011). Main sequence: an index for detecting mental workload variation in complex tasks. *Applied ergonomics*, *42*(6), 807-813.

Di Stasi, L. L., Marchitto, M., Antolí, A., & Cañas, J. J. (2013). Saccadic peak velocity as an alternative index of operator attention: A short review. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *63*(6), 335-343.

Di Stasi, L. L., McCamy, M. B., Catena, A., Macknik, S. L., Cañas, J. J., & Martinez-Conde, S. (2013). Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience*, *38*(3), 2389-2398.

Di Stasi, L. L., Renner, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Cañas, J. J., ... & Pannasch, S. (2010). Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, space, and environmental medicine*, *81*(4), 413-417.

Díaz Cabrera, D., Hernández Fernaud, E., Rolo González, G. (2012). *Carga Mental de Trabajo.* Madrid: Editorial Síntesis.

Ferrer, R., & Dalmau, I. (2004). Revisión del concepto de carga mental: evaluación, consecuencias y proceso de normalización. *Anuario de psicología*, *35*(4), 521-545.

Galley, N. (1998). An enquiry into the relationship between activation and performance using saccadic eye movement parameters. *Ergonomics*, *41*(5), 698-720.

Galley, N., & Andres, G. (1996). Saccadic eye movements and blinks during long-term driving on the autobahn with minimal alcohol ingestion. *Vision in Vehicles*, *5*, 381-388.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. A. M. Hancock (Ed.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.

Hunter, A. C., & Parush, A. (2009, October). Using eye movements to uncover conflict detection strategies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 22, pp. 1729-1733). Sage Publications.

ICAO, International Civil Aviation Organization. Forecasts of Scheduled Passenger Traffic (2012). Retrieved on May, 4th, 2015, from [http://www.icao.int/sustainability/pages/eap_fp_forecastmed.aspx].

Jessee, M. S. (2010, September). Ocular activity as a measure of mental and visual workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 18, pp. 1350-1354). SAGE Publications.

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors*, *56*(6), 1150-1162.

Kang, Z., Bass, E. J., & Lee, D. W. (2014, September). Air traffic controllers' visual scanning, aircraft selection, and comparison strategies in support of conflict detection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 77-81). SAGE Publications.

Leigh, R. J., & Zee, D. S. (1999). *The neurology of eye movements* (Vol. 90). New York: Oxford university press.

Lin, H. H., Chen, Y. F., Chen, T., Tsai, T. T., & Huang, K. H. (2004). Temporal analysis of the acceleration and deceleration phases for visual saccades. *Biomedical Engineering: Applications, Basis and Communications*, *16*(06), 355-362.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218-228.

Nolan, M. (2010). *Fundamentals of air traffic control*. Cengage Learning.

Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J. R., Mérienne, F., Benedetto, S., & Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction*, *30*(3), 220-236.

Schmidt, D., Abel, L. A., DellOsso, L. F., & Daroff, R. B. (1979). Saccadic velocity characteristics- Intrinsic variability and fatigue. *Aviation, space, and environmental medicine*, *50*(4), 393-395.

Wise, J. A., Hopkin, V. D., & Garland, D. J. (Eds.). (2009). *Handbook of aviation human factors*. CRC Press.

Yerkes, R. M., Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482.

Zils, E., Sprenger, A., Heide, W., Born, J., & Gais, S. (2005). Differential effects of sleep deprivation on saccadic eye movements. *Sleep*, *28*(9), 1109-1115.

ACKNOWLEDGMENTS

It is my pleasure to thank the many people who have been important for the realization of this work. I wish I could write an exhaustive list, but I know it is nearly impossible.

I express my deep gratitude to my Thesis Director, Prof. José Cañas, of the Department of Psychology, and director of the Cognitive Ergonomics Group. He offered me sound advice, proper guidelines, stimulating research perspectives, and remarkable approval. Gracias, Pepe.

Thanks to my Group colleagues, for their important support since the beginning of my Spanish adventure: Leandro, Dori, Sonia, Ana.

I am also grateful to all the members of the Learning, Emotion, and Decision Making Group, lead by Prof. Andrés Catena, Prof. Antonio Maldonado, and Prof. Antonio Cándido. Thank you for your academic advice, your encouragement, and for the friendly climate within the group.

Many thanks go to Prof. Thierry Baccino (Université Paris VIII), for giving me the opportunity of a research internship at LUTIN Lab (Paris), and for his precious suggestions. I thank all the crew of LUTIN Lab, in particular Dr. Simone Benedetto and Dr. Andrea Carbone, for their precious help and sincere friendship.

Special thanks to Dr. Yasmina Okan (University of Leeds), and Dr. Marco Pedrotti (Université de Neuchâtel), who acted as external reviewers, for their time and effort.

I thank all the wonderful friends I met in these years in Granada. There are no words to describe all the moments we lived together. Thanks to Ana and Fabiano for all the time, conversations, dinners, holidays, games, etc., we shared. Thanks to Alberto, for his unlimited availability, his positive attitude, his sincere collaboration. Thanks to Paola, Maryem, Elisa, Tommaso, Francesco, Joaquín, Cristi, Gloria, Jose, Francesca, Dafina, Francesca. You all have been fundamental for me during these years.

Finally, I would like to thank my parents, Francesco and Angela, my sister Paola, my brother-in-law Claudio. For their unconditional love, trust, and help demonstrated throughout my whole life. Vi vorró sempre bene!

The last, special thanks, is for a special person, Francesca. Your support has been, and is still today, emotional, physical, moral, and social. Simply, human.

FURTHER PUBLICATIONS

Casucci, M., Marchitto, M., & Cacciabue, P. C. (2010). A numerical tool for reproducing driver behaviour: Experiments and predictive simulations. *Applied ergonomics*, *41*(2), 198-210.

Di Stasi, L. L., Marchitto, M., Antolí, A., & Cañas, J. J. (2013). Saccadic peak velocity as an alternative index of operator attention: A short review. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *63*(6), 335-343.

Di Stasi, L. L., Marchitto, M., Antolí, A., Baccino, T., & Cañas, J. J. (2010). Approximation of on-line mental workload index in ATC simulated multitasks. *Journal of Air Transport Management*, *16*(6), 330-333.

Marchitto, M. (2005, September). Safety culture assessment as crucial phase for tailoring human factors training: the case of an Italian refinery. In *Proceedings of the 2005 annual conference on European association of cognitive ergonomics* (pp. 35-42). University of Athens.

Marchitto, M. (2011). El error humano y la gestión de seguridad: la perspectiva sistémica en las obras de James Reason. *Laboreal*, *7*(2), 56-64.

Marchitto, M., & Cañas, J. J. (2011). User Experience As A Challenge For Cognitive Psychology And Ergonomics. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *7*(3), 268-280.

Odom, W., Zimmerman, J., Forlizzi, J., López Higuera, A., Marchitto, M., Cañas, J., ... & Moore, H. (2013, April). Fragmentation and transition: understanding perceptions of virtual possessions among young adults in Spain, South Korea and the United States. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1833-1842). ACM.

RESUME

Mauro Marchitto graduated in Psychology at the University of Padua, Italy, in 2004. The final thesis was prepared at the Joint Research Center of the European Commission, (Ispra, Italy) during a one-year stage at the Human Factor Sector, under the supervision of Prof. Pietro C. Cacciabue. He then worked in an Italian SME as Human Factors trainer in aviation maintenance and as a researcher on several European Projects (e.g. Aide, Sensation, Rankers, Virthualis) about industrial and transport safety, road design, sensor technologies for sleep and stress detection. He moved to Granada in 2008 and obtained his M.Sc. in Cognitive and Behavioral Neuroscience from the University of Granada in 2009. Successively, he started to work in the Cognitive Ergonomics Group (same institution), on projects related to workload modeling, mobile technology innovation, usability, and user-experience evaluation. Currently, he is a PhD candidate in Social Sciences at the University of Granada, under the supervision of Prof. José J. Cañas Delgado. His research topic deals with eye tracking and workload in Air Traffic Control (ATC). During his doctoral studies, he spent three months at the LUTIN Lab (University of Paris VIII), under the supervision of Prof. Thierry Baccino.

NOTES