Diseñando la Participación del Humano en los Sistemas Autónomos

Vicente Pelechano, Miriam Gil, Joan Fons, Manoli Albert
Centro de Investigación PROS, Universitat Politècnica de València
Camí de Vera s/n, 46022 Valencia, Spain
{pele, mgil, jjfons, malbert}@pros.upv.es

Abstract. Estamos entrando gradualmente en la era de los sistemas que pretenden dotar de capacidades de computación autónoma a servicios cotidianos. La búsqueda de la autonomía completa es un reto que se está persiguiendo en diversos ámbitos de aplicación y sectores industriales. Sin embargo, la realidad es que la variedad de sistemas, dominios, entornos y contextos de ejecución, restricciones legales y sociales, hace vislumbrar un mundo donde esta autonomía completa será una utopía a corto y medio plazo. En los escenarios en que el sistema autónomo no pueda automatizar completamente sus tareas, se requerirá pues de la participación humana. Desde un punto de vista ingenieril la colaboración entre el humano y estos sistemas (*Human in the Loop*) introduce un considerable número de retos y problemas a resolver. En este trabajo se identifican los retos tecnológicos que introduce esta colaboración humano-sistema, y se define un marco conceptual que identifica los aspectos a considerar desde un punto de vista abstracto e ingenieril.

Keywords. Sistemas Autónomos, "Human in the Loop", Interacción Hombre-Máquina, Marco Conceptual, Sensibilidad al Contexto

1 Introducción

El 'mundo inteligente' del futuro se está diseñando como complejos ecosistemas conectados compuestos por una amplia variedad de dispositivos y servicios distribuidos, que interactúan entre sí y que están controlados por un gran número de elementos de computación y humanos. Este tipo de sistemas se caracterizan por requerir adaptarse en tiempo de ejecución a nuevas condiciones del entorno, situaciones impredecibles, necesidades cambiantes de los usuarios, nuevos tipos de dispositivos, nuevas tecnologías con que interactuar o nuevos servicios que consumir.

En este contexto tecnológico es un hecho palpable que estamos entrando en la era de los *Sistemas Autónomos* (SA) [13]. Estos sistemas son capaces de auto-gestionarse por ellos mismos mediante la toma de decisiones. En este ámbito los futuros sistemas software introducirán, de manera gradual y de forma generalizada, capacidades autónomas [13]. La búsqueda de la autonomía completa es un reto que se está persiguiendo en diversos ámbitos de aplicación y sectores industriales (como los vehículos autónomos, los robots o los drones). Sin embargo, la realidad es que la variedad de sistemas, dominios, entornos y contextos de ejecución, restricciones legales y sociales, hace vislumbrar un mundo donde la autonomía completa será una utopía a corto y medio plazo [8, 10, 23]. En este escenario podemos constatar que, como ingenieros de software, nos enfrentamos al reto de desarrollar SA que puedan soportar un determinado Nivel de Autonomía (NA) [13] o distintos niveles ajustables de forma contextual y dinámica. Los NA surgen al reconocer que un SA no se puede automatizar completamente y requiere de la participación humana para llevar a cabo ciertas tareas.

Asumiendo como ciertas las premisas de que 1) conseguir la **autonomía completa** para todos los sistemas en cualquier ámbito de aplicación es una **utopía alcanzable a muy largo plazo** solo por un subconjunto de sistemas, 2) los SA ofrecerán **distintos grados** o **niveles de autonomía** y 3) la **autonomía parcial** implica la **necesidad** de que los **humanos ayuden o complementen al SA** tomando ciertas decisiones o realizando las tareas que el sistema no pueda ejecutar, podemos afirmar que estamos en un escenario en el cual el humano debe participar ayudando en la ejecución de ciertas tareas del SA.

Bajo estas premisas, humanos, máquinas y software están obligados a entenderse e interactuar para poder trabajar conjuntamente de la forma más efectiva y robusta posible [20]. Incluso en escenarios de autonomía total (vehículos autónomos nivel 5, robots en fábricas del futuro, etc.), los SA tienen la oportunidad, si no la obligación de explicar o proporcionar realimentación sobre su comportamiento a los humanos. Ambos deben entenderse y construir relaciones de confianza para poder trabajar conjuntamente y promover la aceptación de estos sistemas por parte de los humanos.

Desde un punto de vista ingenieril, la participación del humano en los SA (*Human in the Loop*) introduce un considerable número de retos y problemas a resolver para conseguir una participación correcta y sin fisuras, bien integrada en el SA, incluso en situaciones donde los usuarios tengan limitados los recursos atencionales, cognitivos o físicos para llevar a cabo la interacción [2]. Algunos trabajos han identificado el problema de cómo introducir al humano en el bucle de adaptación de los SA pero todavía no se ha propuesto una solución ingenieril que ayude a los desarrolladores de software a diseñar esta participación [2, 14].

En este trabajo pretendemos identificar aquellos retos tecnológicos que introduce la participación del humano en los SA y abrir caminos para diseñar soluciones en el ámbito de la **Ingeniería del Software** y la **Interacción Hombre-Máquina.** En particular, en este trabajo se define un marco conceptual que permitirá caracterizar la cooperación humano y SA. De esta manera, se abre el camino hacia el diseño soluciones ingenieriles (técnicas y métodos adaptados al dominio de aplicación) para construir SA que involucren adecuadamente a los humanos en los procesos automáticos para los que fueron diseñados.

El artículo está estructurado de la siguiente manera: en la Sección 2 se estudia y argumenta la necesidad de participación del humano y se identifican las tareas que debe realizar. La Sección 3 identifica un conjunto de retos y objetivos para abordar esta participación. La Sección 4 presenta los conceptos básicos sobre los cuales construir una propuesta metodológica. Para ello, se introduce el modo de funcionamiento de los SA con participación del humano. A partir de esta especificación del comportamiento de los SA, se presentan los aspectos que deben tratarse desde el punto de vista de la Ingeniería del SW y la HCI tanto para el diseño de las tareas que el humano debe llevar a cabo para complementar la funcionalidad del SA, como para el diseño de la interacción humano-sistema que requieren las tareas. Finalmente, la Sección 5 presenta cuál es nuestra intención de futuro con la propuesta y las conclusiones.

2 El Humano en los Sistemas Autónomos. ¿Es necesaria su participación?

Durante mucho tiempo se han realizado estudios para definir qué tipo de interacciones deberían existir entre humanos y sistemas que automatizan tareas [6]. Esta automatización desplaza el rol que juega el humano sobre el control/ejecución de una tarea desde un rol de responsabilidad de planificación, ejecución y monitorización, a un rol relacionado con la supervisión de la tarea [16].

Tal y como se plantea en [19], los SA pueden ofrecer mejoras sustanciales, como incrementar la seguridad, reducir el error humano y disminuir los tiempos de respuesta, entre otros. Sin embargo, antes que estos sistemas sean una realidad, hay que resolver otros retos relacionados con factores humanos como el efecto negativo causado por adaptaciones provocadas por confusiones o malos usos, una dependencia excesiva sobre los procesos automáticos o incluso en cambios en el nivel de atención (distracciones) de los humanos.

La autonomía de un sistema representa hasta qué grado el sistema es capaz de tomar decisiones, actuar o entender el contexto en el que se encuentra 'por sí solo', sin intervención humana [15]. En [13] se argumenta que vamos a entrar gradualmente en el mundo de los SA en el que tendremos sistemas con diferentes grados o niveles de autonomía (NA). La autonomía completa se define como la capacidad de los sistemas de actuar, bajo cualquier circunstancia y para todas sus capacidades, de manera autónoma. Según [13], este escenario no llegará a todos los niveles en muchos años. Mientras tanto los SA requerirán soporte humano para poder garantizar un funcionamiento correcto en todas las situaciones.

2.1 Niveles de Autonomía y la Necesidad del Humano

A mediados del siglo pasado, en [6] se indicaba cuáles eran los procesos, tareas y funciones, que podía desempeñar de una manera más efectiva un humano o una máquina. En este trabajo se identifican aspectos que podían ser mejor desempeñados por humanos (detección, percepción, juicio, inducción, improvisación, memoria a largo plazo), frente a otros que podían ser mejor resueltos por máquinas (velocidad, potencia de cómputo, capacidad de replicación, simultaneidad de operaciones y memoria a corto plazo). A pesar de las limitaciones tecnológicas de la época, y del cambio paradigmático en la computación que ponen en duda los resultados que se obtuvieron, hoy en día sigue siendo un trabajo relevante [21, 22].

El camino hacia la autonomía completa de los sistemas se traduce en la progresiva traslación del control y la supervisión entre humanos y máquinas [16]. Este cambio en la responsabilidad sobre la planificación, ejecución y monitorización activa a un papel de supervisión permite una descarga del proceso cognitivo del operador humano, una reducción del estrés sufrido por estas tareas, a la vez que ofrece la posibilidad de mejorar tiempos de respuesta, niveles de seguridad, y soporte a múltiples tareas en paralelo [7].

A raíz de estos trabajos se propusieron nuevas aproximaciones y taxonomías que trataban de identificar escenarios colaborativos donde humanos y máquinas podían cooperar para completar (colaborativamente) estas tareas, procesos y funciones. Podemos observar que existen propuestas de ámbito general, y otras que se enfocan claramente hacia un dominio específico (vehículos no tripulados espaciales, aéreos o terrestres, maquinaria y robótico en procesos de fabricación, etc.). Cada aproximación presenta diferentes NA (desde aproximaciones con 4 niveles, a otras con más de 10) y asocia a cada nivel, en función de la complejidad de las tareas a realizar, diferentes acciones que debe realizar el humano para asistir al sistema.

Una de las taxonomías generales ampliamente usada fue presentada por [17] (ver Tabla 1), quienes definieron 10 niveles de automatización que representan el foco de control (humano o sistema) y cómo se presenta la información al humano. A menor NA (1), mayor control por parte del humano y menor asistencia recibida por el sistema. A NA (10), mayor control por el sistema y menor supervisión o acciones a realizar por el humano.

Nivel	**Descripción**
1	El humano realiza la tarea sin apoyo del sistema
2	El sistema ayuda identificando las opciones
3	El sistema ayuda identificando las opciones y sugiere una, que el humano no tiene porqué seguir
4	El sistema selecciona la acción y el humano puede o no hacerla
5	El sistema selecciona la acción y la realiza si el humano la aprueba
6	El sistema selecciona la acción para realizarla, e informa al humano con tiempo suficiente para que la pueda cancelar
7	El sistema realiza la tarea y avisa siempre al humano de este hecho
8	El sistema realiza la tarea y si éste lo solicita explícitamente, se le informa
9	El sistema realiza la tarea y decide si debe o no informar al humano
10	El sistema realiza la tarea si decide que debe hacerse. En tal caso, decide además si informar o no al humano

Tabla 1. Niveles de Automatización de Sheridan y Verplank (1978)

El nivel 1 implica el control total del humano sobre la tarea, sin apoyo del sistema. Nos referiremos a este nivel como “grado de autonomía 0”. Este nivel no es de interés en este trabajo debido a que no hay autonomía subyacente. En los niveles 2 y 3, la realización de una tarea la lleva a cabo el humano, pero es asistido por el sistema por medio del cálculo de una serie de opciones, entre las que el humano debe seleccionar y aplicar. En los niveles 4 al 6, es el sistema el que decide la mejor opción e informa al humano para que éste realice la acción. En los niveles 7 al 10, el sistema realiza la acción y notifica al humano de la tarea realizada (en diferente grado según el nivel).

Esta taxonomía fue modificada más tarde por [15] para incluir las cuatro fases del procesado (adquisición de datos, percepción y análisis, toma de decisiones, y ejecución) para cada nivel. En este trabajo se define el rol que juega el humano dentro de la tarea autónoma/automática a realizar, según la fase en la que participe, y se le asignan tareas de monitorización, de verificación o cancelación, de toma de decisiones o de ejecución (control), además de informarle sobre procesos automáticos.

Paralelamente, en el ámbito de las naves o vehículos autónomos se han propuesto otras taxonomías específicas basadas en identificar el control del piloto/conductor versus del vehículo no manejado (UMV) sobre las funciones primarias del vehículo. En estas taxonomías, a medida que se incrementa el NA el rol del conductor se va desplazando desde el control primario, pasando por el control supervisado hasta la completa autonomía. En la propuesta [19] se definen 5 niveles (del 0 al 4). En los niveles inferiores, el humano debe realizar tareas de control y decisión, mientras que en los niveles superiores es el vehículo el que tiene el control y sólo ante ciertas condiciones solicita al humano que tome decisiones (con unas restricciones temporales pre-establecidas) o incluso que tome el control del vehículo temporalmente. En la propuesta [4] se definen 6 niveles (del 0 al 5) muy similares a la anterior, pero introduce explícitamente la capacidad o responsabilidad de monitorizar el entorno de conducción al conductor (niveles 0 al 2) o al vehículo (niveles 3 al 5).

Otros trabajos relacionados han intentado identificar problemas relacionados con las interacciones humanas en los SA. Por ejemplo, en [12, 18] abordan el problema de la desconfianza que puede generar al humano el hecho de delegar responsabilidades de control en los sistemas. Promueven el desarrollo de SA que mantengan informado, den retroalimentación continua sobre los procesos subyacentes, descarguen carga mental y estrés, a la vez que mantienen en el piloto/conductor cierta 'consciencia situacional' (representación mental) sobre estos procesos. Argumentan la necesidad de diseñar soluciones que satisfagan estos requisitos para aumentar las posibilidades de que los humanos acepten estos sistemas.

Es decir, la participación colaborativa del humano en los SA es un hecho que habrá que tratar durante mucho tiempo para desarrollar correctamente este tipo de sistemas. En la siguiente sección se analiza esta responsabilidad del humano en los SA identificando las acciones que el humano debería realizar.

2.2 Tareas del Humano en los Sistemas Autónomos

Tras analizar las diferentes aproximaciones, hemos llevado a cabo un proceso de abstracción para clasificar las acciones que estos NA requieren sobre el humano. Teniendo en mente las capacidades propuestas por [6, 15], proponemos los siguientes grupos de tareas:

1. Tareas de **monitorización** e introducción de información: el humano es responsable de realizar acciones para identificar patrones o situaciones en el contexto de la tarea en ejecución. También lo es de proveer información al SA para que éste pueda tomar decisiones en base a esta información.
2. Tareas de **decisión**: el humano debe decidir entre un conjunto de opciones predefinidas por el sistema, o completamente por su cuenta. En función de la situación, el sistema deberá proveer suficiente información o explicaciones para que el humano pueda tomar la mejor decisión posible.

3. Tareas de **control** o de ejecución: el humano es el responsable de manejar, ejecutar o dirigir la tarea (asistido o no por el sistema). También puede tener la opción de ceder el control (parcial o temporalmente) sobre la tarea al sistema.

Adicionalmente, para dar soporte a los aspectos de ***confianza*** y ***consciencia situacional*** identificados en [12, 18], necesitamos un tipo de tarea más que, si bien no es una acción del humano, es necesaria para la interacción adecuada humano-SA:

4. **Acciones de realimentación** (feedback) o **explicaciones** que realiza el SA con respecto al humano con el objetivo de: informar sobre procesos subyacentes (por ejemplo, hacer consciente sobre alguna acción automática en curso); explicar consecuencias sobre acciones tomadas (por ejemplo, informar sobre el recálculo automático de una ruta debido a una incidencia en el camino). En este trabajo no vamos a considerar este tipo de tareas, que aunque pueda requerir la intervención humana, abarcarían todo el espectro funcional del SA y no requieren acciones directas por parte del humano.

Uno de nuestros objetivos ha sido identificar las acciones que debía realizar el humano con independencia del NA en el que se ubica. Es decir, cada tarea puede estar ubicada en un NA diferente (en cualquiera), y su gestión puede requerir, en función de dicho nivel, una interacción con el humano diferente. En la siguiente sección se identifican los objetivos para involucrar al humano en la realización de estas tareas.

3 Objetivos para Involucrar al Humano

Una vez argumentada la necesidad que el humano participe en los SA, tomamos como ciertas las siguientes premisas: 1) conseguir la *autonomía completa para todos los sistemas en cualquier ámbito de aplicación es una utopía difícilmente alcanzable a medio plazo*, 2) los *SA soportarán distintos grados o NA* y 3) la *autonomía parcial implica la necesidad de que los humanos colaboren con el SA* tomando las decisiones o realizando las tareas que el sistema no pueda llevar a cabo por su NA.

Tal y como hemos comentado en la Sección 1, la participación del humano en los SA introduce un considerable número de retos a resolver para conseguir una participación correcta y sin fisuras, bien integrada en el SA incluso en situaciones donde los usuarios tengan limitados los recursos atencionales, cognitivos o físicos para llevar a cabo la interacción. Simplificando el problema y concentrándonos en un subconjunto de aspectos esenciales, pensamos que, en base a estos retos, una posible solución debería permitirnos alcanzar una serie de objetivos básicos que garanticen la calidad del sistema desarrollado. Estos objetivos son:

1. **Complementar la funcionalidad del SA** con la colaboración/participación del humano. El SA, de forma colaborativa con el humano, debe conseguir llevar a cabo "toda" la funcionalidad para la cual fue concebido.
2. **Facilitar que el humano se implique/involucre** de forma adecuada para llevar a cabo o ejecutar las tareas que el SA no puede llevar a cabo por sí solo. El SA debe ofrecer al humano los mecanismos de interacción y la información adecuada para poder ejecutar la funcionalidad requerida.

3. **Incorporar los mecanismos necesarios para ofrecer feedback** de forma que el humano comprenda el funcionamiento del SA y deposite su confianza en él. Los humanos puede que no confíen en los SA que actúan a sus espaldas. Si no saben qué va a suceder en el sistema o cómo van a reaccionar ante una situación el comportamiento autónomo puede provocar desconfianza.

La colaboración del humano servirá como herramienta para conseguir la ***completitud*** y la ***corrección*** del SA. Con su ayuda, el SA realizará aquello para lo que fue concebido. Estos objetivos básicos, a su vez, se deben llevar a cabo bajo una **gestión específica de los recursos de atención del humano**: El SA debe ser capaz de **captar la atención** del humano y evitar molestarle en exceso (**reducir la molestia**). El humano puede que esté distraído o realizando otras tareas en el momento que se requiera su participación. El SA deberá captar su atención dependiendo del grado de implicación que sea necesario y lo implicará en la ejecución de las tareas, evitando abrumarlo con interacciones innecesarias y acciones que demanden excesiva atención y destreza. A pesar de que sería deseable conseguir un alto nivel de autonomía en los SA, es importante establecer un equilibrio entre el NA y la intervención humana para evitar llegar a resultados no deseados o una mala experiencia de usuario. Por tanto, la **gestión de la participación humana** es un aspecto crucial para **evitar el rechazo** a estos sistemas.

El marco conceptual que planteamos tiene como objetivo proporcionar las bases para diseñar técnicas y métodos que nos permitan alcanzar los objetivos marcados: complementar la funcionalidad del SA, implicar adecuadamente al humano, captar su atención reduciendo la molestia y generar confianza en el sistema desarrollado.

4 Caracterizando la Participación del Humano

En esta sección se define un marco conceptual que identifica los aspectos a considerar en el diseño de la participación del humano en los SA desde un punto de vista abstracto e ingenieril. Para ello en primer lugar se determina el modo de funcionamiento (o modelo de ejecución) de los SA que involucran colaborativamente a los humanos (Sección 4.1), y a partir de éste se introducen los conceptos esenciales que componen el marco conceptual (Sección 4.2).

4.1 Modelo de Ejecución

Un SA está formado, entre otros, por un conjunto de tareas, procesos y actividades automáticas que permiten desempeñar la función para la que está concebido. El sistema es responsable en todo momento de identificar qué tareas debe realizar (puede implicar ejecutar varias tareas en paralelo). Para ello, estos sistemas deben monitorizar activa y continuamente el entorno en el que operan para activar estos procesos automáticos. En este modo autónomo, el sistema es también responsable de identificar situaciones en las que no se pueda solucionar de manera completa, correcta o con garantías una determinada tarea (generando logs/avisos, o tomando acciones para protegerse contra una potencial incidencia). Es en este caso cuando el sistema debería hacer partícipe al humano para que, de manera colaborativa, resuelvan la tarea. Para que esta colaboración sea factible, se debería exigir que exista al menos un humano a una distancia y en una situación en la cual es viable la comunicación e interacción entre ambos. Esto es, el humano debe estar físicamente cerca de una interfaz del SA (consola de interacción, mandos, volante, interfaces móviles, etc.). El proceso de involucrar al humano se puede abstraer con los siguientes pasos (ver diagrama de flujo en Figura 1):

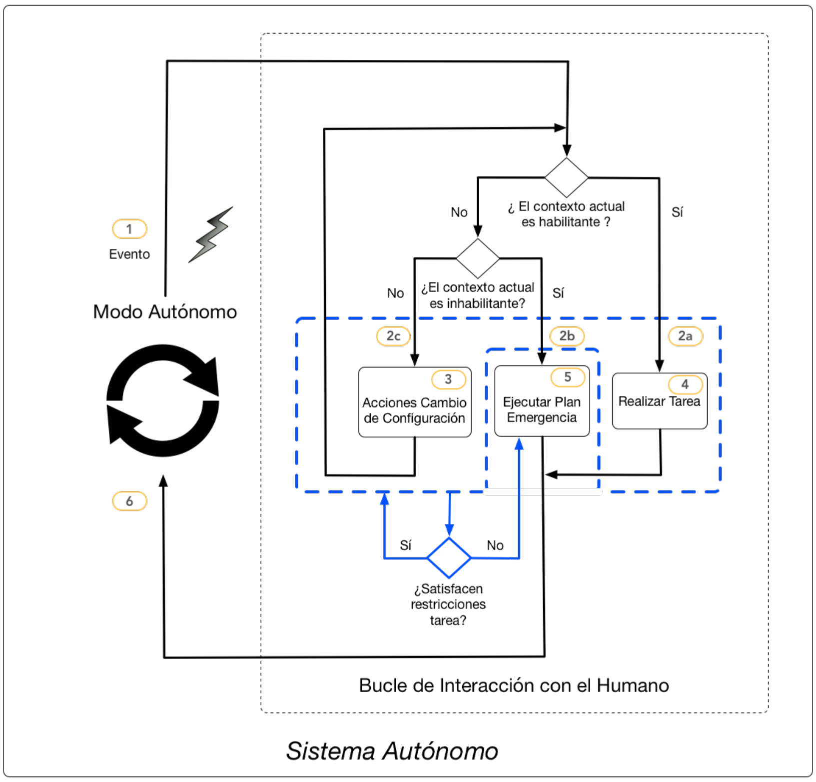

Figura 1. Modelo de ejecución.

1) El SA está continuamente monitorizando su entorno para identificar situaciones en las que se requiera la participación humana. En tal caso, eleva un **evento** que dispara la necesidad de que el humano ayude o lleve a cabo una **tarea**.

2) Se comprobarán **condiciones** sobre el **contexto**.
   2.a) Si las condiciones son adecuadas para realizar la tarea (**contexto habilitante**) pasaremos al **paso 4**.
   2.b) Si las condiciones no son adecuadas y no permiten la ejecución de la tarea (**contexto inhabilitante**) pasaremos al **paso 5**.
   2.c) Si las condiciones de contexto no son *habilitantes* ni *inhabilitantes* pasaremos al **paso 3**.

3) El sistema, mediante **mecanismos de interacción** realizará acciones para cambiar el **contexto** con el objetivo de alcanzar el contexto habilitante (por

ejemplo, requiriendo la atención del humano y/o preparando el entorno de forma que le habilite para la realización de la tarea). Volver al **paso 2**.

4) Realización de la tarea: el SA facilitará que el humano lleve a cabo la tarea mediante los **mecanismos de interacción** y **control** necesarios. Ir al **paso 6**.
5) Ejecutar **acciones de emergencia** o plan alternativo que dejen al sistema en un estado seguro. Ir al **paso 6**.
6) Una vez finalizada la tarea el sistema pasará automáticamente a **modo autónomo** (hasta detectar nuevas situaciones que requieran la participación del humano).

Toda tarea con intervención humana se realizará cumpliendo estrictamente una serie de restricciones (*temporales, contextuales, etc*.) previamente definidas. El sistema comprobará en todo momento (**pasos 2**, **3** y **4**) que las restricciones se cumplen hasta la finalización de la tarea (representado con líneas azules discontinuas en la Figura 1). Si las restricciones no se cumplen se ejecutará el **paso 5.**

4.2 Un Marco Conceptual

En esta sección se presenta el marco conceptual para el diseño de la participación humana en los SA partiendo de la caracterización previa del comportamiento y teniendo en cuenta los retos identificados en la Sección 3. El objetivo del marco conceptual es identificar los términos de interés en este dominio para construir modelos y técnicas que permitan: 1) especificar las tareas que el humano debe realizar para complementar la funcionalidad del SA mediante mecanismos de interacción, y 2) definir una interacción adecuada para conseguir que el humano pueda llevar a cabo una tarea, que no le resulte molesta en exceso y que maximice la confianza de éste con el SA (objetivos identificados en la Sección 3).

4.2.1 Definición de Tareas

La funcionalidad del SA en la que el humano puede involucrarse viene determinada por un conjunto de tareas. Las **tareas** describen acciones que debe llevar a cabo el humano interactuando con el SA. Las tareas se clasifican en 4 tipos (ver Sección 2), atendiendo al tipo de participación requerida:

- De **control**: el humano ejecuta una acción.
- De **decisión**: el humano toma una decisión.
- De **monitorización**: el humano proporciona al SA datos que éste no puede obtener u observar sin su ayuda.

Las tareas tienen asociado un **perfil de humano.** El perfil describe una categoría de humanos que comparte un conjunto de características y que juega un determinado rol en el SA. Los humanos que cumplen el perfil asociado a la tarea son los únicos que pueden ejecutarla.

Las tareas se ejecutan bajo ciertas condiciones de contexto. Es importante definir en primer lugar qué entendemos por **contexto**. Dey [3] define el contexto como "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user* *and an application, including the user and applications themselves*." De esta definición se deriva que son **propiedades** importantes del **contexto** el **estado del entorno físico** (como puede ser la temperatura, ruidos, etc.), el **estado de los recursos próximos** al sistema, el **estado del propio sistema** o la **situación del usuario** (como su localización, actividad, etc.). En este trabajo cobra especial importancia la *situación* o *contexto del humano* para determinar si el humano está "preparado" o "puede" participar en la tarea. El modelo "*opportunity-willingness-capability*" (OWC) [5] nos sirve para precisar el contexto del humano identificando una serie de factores que condicionan la ejecución de una tarea:

- **Oportunidad**: Identifica el conjunto de variables relacionadas con la situación del humano, como puede ser la localización del usuario, la actividad que está llevando a cabo, etc.
- **Disposición**: Indica la predisposición del humano a realizar la tarea. Este factor está relacionado con su nivel de atención o de estrés, o con lo ocupado que esté, por ejemplo. El usuario no podrá realizar la tarea si no tiene los recursos atencionales necesarios disponibles que requiere esa tarea.
- **Capacidad**: Define las habilidades y destrezas necesarias para realizar la tarea. Tiene que ver con sus conocimientos, nivel de experiencia o entrenamiento, con estados emocionales que puedan mermar su capacidad, etc.

La **ejecución** de la tarea se habilita bajo una casuística no trivial, puesto que nos encontramos ante tres posibles situaciones que condicionarán que se lleve a cabo la tarea (como se ha visto en el modelo de ejecución):

- **Contexto habilitante.** Representa la situación idónea para realizar la tarea. Es decir, la situación en la que el sistema, su entorno y el humano están **preparados** para ejecutar la tarea en unas condiciones adecuadas. Por ejemplo, en el caso de un vehículo autónomo, para pasar de un control autónomo a un control manual debería existir un pasajero con capacidad de conducción ubicado en el asiento del conductor y con un nivel de atención alto.
- **Contexto inhabilitante**. Este contexto definirá una situación bajo la cual no es posible realizar la tarea. En este *contexto* no se permitirá la realización de la tarea (por lo consiguiente el usuario no podrá participar en ella). Por ejemplo, en el caso anterior, para pasar a un control manual un contexto que inhabilitaría la ejecución de la tarea sería la no existencia de pasajeros con capacidad de conducir.

Pero, ¿qué ocurre con aquellas situaciones de contexto en las que no se cumplen ni las condiciones de contexto habilitante ni las del inhabilitante? En ellas, el contexto, el sistema, su entorno y/o el humano no están en la situación idónea para realizar la tarea, pero tampoco sería inviable ejecutarla. A este caso lo llamamos **contexto favorable**, en el que el SA ejecutará acciones con el objetivo de cambiar la configuración de contexto para tratar de alcanzar el contexto habilitante, y así poder realizar la tarea.

En la Figura 2 se presenta el metamodelo de este Marco Conceptual donde se pueden ver los conceptos introducidos y sus relaciones. El núcleo del metamodelo lo constituyen las clases coloreadas en amarillo que representan las tareas, sus tipos, el perfil del humano que se requiere y las restricciones que deben satisfacerse durante la ejecución de la tarea. Coloreadas en verde están las clases relacionadas con el concepto de contexto del SA. El contexto se define a través de propiedades de contexto, entre las cuales se distinguen las del humano, con los factores introducidos por el modelo OWC (*Oportunidad*, *Disposición* y *Capacidad*). Las tareas definen unas condiciones de contexto habilitantes e inhabilitantes, representado con sendas relaciones de asociación entre tarea y condición de contexto. Las condiciones de contexto se formularán mediante expresiones definidas sobre propiedades del contexto.

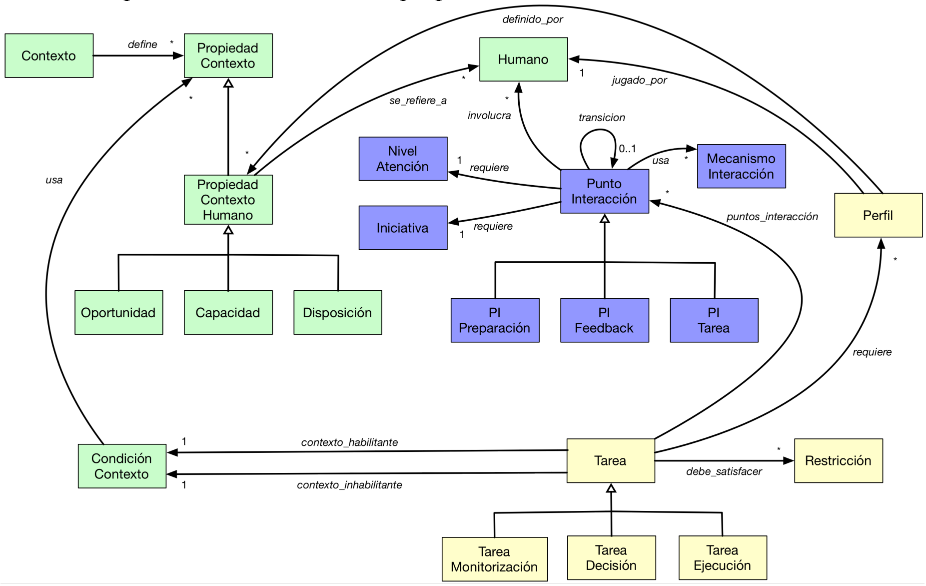

Figura 2. Conceptos y relaciones esenciales de la participación del humano.

4.2.2 Definición de la Interacción con el Humano

Llegados a este punto, debe definirse cómo interactuará el SA con el humano para poder llevar a cabo las tareas previamente identificadas. Las interacciones del humano con el SA se deben diseñar de forma que se permita:

- **Llevar a cabo la tarea** por el humano (complementar la funcionalidad del SA).
- **Preparar al humano** y **captar su atención** (facilitar su implicación en la tarea).
- **Proporcionar feedback** al humano (comprender acciones del SA y confiar en él).

De forma abstracta, las interacciones con el humano se llevarán a cabo a través de los *puntos de interacción.* Un ***punto de interacción (PI)*** se define como un intercambio básico entre dos entidades [11]. Por la naturaleza de este trabajo, asumimos que una entidad será el SA y la otra el humano, y se producirá un intercambio de información en esa interacción. Según esta definición, una tarea que involucre al humano (control, decisión, monitorización y feedback) se llevará a cabo mediante *uno o varios PI*, dependiendo de la complejidad de la tarea y del contexto. Se ha definido un tipo específico de PI para satisfacer cada uno de estos objetivos:

- *PI de la Tarea:* permiten que el humano realice una tarea concreta (control, decisión, monitorización). Están asociados al contexto habilitante de una tarea.

- *PI de Preparación*: garantizan que el humano realiza la tarea en el contexto idóneo. Cuando la tarea a realizar se debe ejecutar en un momento en el que el contexto no es el adecuado (habilitante), por medio de estos PI de preparación se tratará de 'convertir' este contexto en uno habilitante.
- *PI de Feedback:* ofrecen información al usuario sobre el comportamiento y contexto del sistema y así maximizar la confianza del humano en el sistema.

Los PIs asociados a una tarea tendrán un orden que determina la secuencia de ejecución de la interacción con el humano. El orden se definirá mediante **transiciones** entre PIs. Las transiciones se podrán etiquetar con las **restricciones** asociadas a la tarea; el incumplimiento de la restricción implica la finalización de ejecución de la tarea.

Los PIs se diseñarán **gestionando los recursos de atención** del humano (subobjetivo identificado en la Sección 3) de forma que se consiga **captar la atención** del humano pero a su vez se evite molestarle en exceso (**reducir la molestia**). La HCI ha propuesto trabajos para la gestión de la atención basados en niveles de automatización y carga de trabajo [1, 9]. Para este trabajo vamos a analizar y diseñar los puntos de interacción desde dos puntos de vista: 1) quién **inicia la interacción** en dichos puntos (si el sistema o el humano) y 2) los **recursos de atención** necesarios para atender la interacción (demanda de atención impuesta sobre el humano). Esta especificación de la iniciativa y la atención la vemos adecuada puesto que estos son factores que varían independientemente. Por ejemplo, un punto de interacción de feedback (iniciado por el sistema) puede requerir más atención o menos dependiendo de la importancia/criticalidad de la información a ofrecer. Al igual que un punto de interacción de preparación podría ser iniciado por el sistema (si intenta captar la atención del usuario) o por el humano (si esperamos contestación por su parte de que está preparado). Este diseño de los puntos de interacción por medio de la especificación de la iniciativa y la atención que requieren guiará a los diseñadores para la posterior selección de los **mecanismos de interacción** mas adecuados.

La Figura 2 muestra en color violeta las clases del metamodelo que representan la parte del marco conceptual relacionada con la interacción humano-sistema. Los puntos de interacción asociados a las tareas se clasifican en tres tipos (de tarea, de preparación y de feedback). Un punto de interacción requiere de un nivel de atención y de una iniciativa, y se concretará en unos mecanismos de interacción.

5 Conclusiones y Trabajos Futuros

Este trabajo constituye un primer paso hacia una solución ingenieril para construir SA que involucren adecuadamente a los humanos. Para ello, hemos analizado cuál debería ser la participación del humano en el ámbito de los SA y cuáles son los aspectos a tener en cuenta para su diseño. En primer lugar introducimos una serie de retos que plantean estas cuestiones para posteriormente definir un marco conceptual en el que se identifican qué aspectos se deben considerar para diseñar la participación del humano teniendo en cuenta los retos marcados. Este es el primer ladrillo, todavía falta analizar y dar solución a muchos aspectos tanto conceptuales como técnicos para conseguir una propuesta completa, rigurosa y sólida. Este trabajo abre el camino a muchos trabajos futuros que aborden aspectos que todavía quedan por tratar. A continuación se presentan potenciales trabajos futuros que se derivan de este trabajo:

- *A nivel de requisitos* sería adecuado (1) identificar qué es necesario especificar a este nivel para describir la participación humana en el SA, (2) proponer técnicas para especificar los requisitos, y (3) definir las correspondencias entre los conceptos que aparecen en los requisitos y los elementos de modelado.
- *A nivel de modelado* es necesario definir técnicas (1) para capturar los conceptos identificados en el Marco Conceptual, y (2) para diseñar los mecanismos de interacción concretos idóneos para cada nivel de iniciativa/atención. El uso de Lenguajes Específicos de Dominio (DSL) sería muy adecuado para este menester.
- *A nivel de infraestructura software* será necesario diseñar una solución arquitectónica y un framework de ejecución que contenga las piezas software necesarias para implementar prototipos funcionales de estos SA.
- *A nivel de validación* es necesario prototipar la propuesta y aplicarla de manera sistemática a suficientes casos de estudio en diferentes dominios de aplicación, para generalizar, si fuera necesario, el modelo de ejecución, o para extender y/o refinar los conceptos del marco conceptual.
- *A nivel de verificación* sería recomendable evaluar hasta qué punto la participación del humano en estos SA le genera la confianza esperada sobre el sistema. Para ello, deberían conducirse experimentos con usuarios finales para medir el grado de satisfacción con estas interacciones a partir de su feedback.

Agradecimientos. Este trabajo ha sido financiado por la Generalitat Valenciana bajo la ayuda post-doctoral APOSTD/2016/042.

Referencias

1. Buxton, B.: Integrating the periphery and context: A new model of telematics. In Proceedings of Graphics Interface, pp. 239–246 (1995).
2. Cámara, J., Moreno, G., Garlan, D.: Reasoning about human participation in self-adaptive systems. In: SEAMS 2015, pp. 146–156 (2015).
3. Dey, A. K.: Understanding and Using Context. Personal and Ubiquitous Computing 5:4–7 (2001).
4. ERTRAC. Automated Driving Roadmap. ERTRAC Task Force. Connectivity and Automated Driving, (2015).
5. Eskins, D. And Sanders, W. H.: The Multiple-Asymmetric-Utility System Model: A Framework for Modeling Cyber-Human Systems. QEST '11 Proceedings of the 2011 Eighth International Conference on Quantitative Evaluation of SysTems, 233-242 (2011).
6. Fitts, P. M.: Human engineering for an effective air-navigation and traffic-control system. Washington, DC: National Research Council, (1951).
7. Frost, C.R.: Challenges and Opportunities for Autonomous Systems in Space. National Academy of Engineering's U.S. Frontiers of Engineering Symposium, Armonk, New York, September 23-24, (2010).

8. Gómez, M. Coches eléctricos y autónomos, el cambio llega sobre ruedas. El Pais.http://tecnologia.elpais.com/tecnologia/2017/01/26/actualidad/1485435443_182055.html
9. Horvitz, E., Kadie, C., Paek, T., and Hovel, D.: Models of attention in computing and communication: from principles to applications, Commun. ACM, vol. 46, no. 3, pp. 52–59, (2003).
10. Jaynes, N. Timeline: The future of driverless cars, from Audi to Volvo. Mashable. http://mashable.com/2016/08/26/autonomous-car-timeline-and-tech/#eBMDuSgNhEqF
11. Ju, W., Leifer, L.: The design of implicit interactions: making interactive systems less obnoxious. Des. Issues 24(3), 72–84 (2008).
12. Lee, J. D., & See, K. A.: Trust in automation: Designing for appropriate reliance. Human Factors, 46, 50–80, (2004).
13. Moore, A., O'Reilly, T., Nielsen, P. D., and Fall, K.: Four Thought Leaders on Where the Industry Is Headed. IEEE Softw. 33, 1, 36-39 (2016).
14. Nunes, D. S., Zhang, P., Sá Silva, J.: A Survey on Human-in-the-Loop Applications Towards an Internet of All, in IEEE Communications Surveys & Tutorials, vol. 17, no. 2, pp. 944-965, (2015).
15. Parasuraman, R., Sheridan, T. B., and Wickens, C. D. 2000.: A model for types and levels of human interaction with automation. Trans. Sys. Man Cyber. Part A 30, 3, 286-297, (2000),
16. Sheridan, T.: On how often the supervisor should sample. IEEE Transactions on Systems Science and Cybernetics, SSC-6, 140-145, (*1970*).
17. Sheridan, T. B., and Verplank, W. L.: Human and Computer Control of Undersea Teleoperators," MIT Man-Machine Systems Laboratory, MA, Tech. Rep., (1978).
18. Stanton, N. A., & Young, M. S.: Vehicle automation and driving performance. Ergonomics, 41(7), 1014-1028, (2010).
19. Trimble, T. E., Bishop, R., Morgan, J. F., & Blanco, M.: Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts. (Rep. No. DOT HS 812 043), (2014, July).
20. Venneti, S., and Rosenthal, S.: Humans and Machines: Working Better Together. Sei Webinar Series (2016).
21. Winter, J.C.F., Dodou, D.: Why the Fitts list has persisted throughout the history of function allocation. Cogn Tech Work, 16:1–11 (2014).
22. Winter, J.C.F., Hancock, P.A.: Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.
23. World Economic Forum White Paper. Digital Transformation of Industries. Automotive Industry. http://reports.weforum.org/digital-transformation/wp-content/blogs.dir/94/mp/files/pages/files/wef-dti-automotivewhitepaper-final-january-2016-200116a.pdf.