



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

MAESTRÍA EN ARTES VISUALES (COMUNICACIÓN Y DISEÑO GRÁFICO)

POSGRADO DE ARTES Y DISEÑO

FACULTAD DE ARTES Y DISEÑO

LA INTERFAZ GESTUAL.

ANÁLISIS FORMAL DEL GESTO COMO SISTEMA DE COMUNICACIÓN

EN LOS PROGRAMAS DE COMPUTADORA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN ARTES VISUALES

PRESENTA:

GONZALO LEÓN GONZÁLEZ-PACHECO

TUTORA:

DRA. ELIA DEL CARMEN MORALES GONZÁLEZ

FACULTAD DE ARTES Y DISEÑO

CIUDAD DE MÉXICO, A 25 DE ENERO DE 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas

Tesis Digitales

Restricciones de uso

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

La interfaz gestual.

Análisis formal del gesto como sistema de comunicación en los programas de computadora

Contenido

Introducción	1
Capítulo 1 – La interfaz – conceptos generales	9
1.1 La interfaz como medio de comunicación.....	9
1.2 La interfaz gráfica.....	14
1.3 El paradigma WIMP (<i>Windows, Icons, Menus, Pointing Device</i>).....	18
1.4 Problemáticas de la interfaz gráfica WIMP	20
1.5 Puntos guía para el diseño de aplicaciones con interfaz gráfica.....	29
Capítulo 2.- Aspectos de la interfaz gestual	33
2.1 Características de la interfaz gestual.....	33
2.2 El gesto como signo.....	38
2.3 El gesto como elemento lingüístico.....	45
2.4 El gesto y su inclusión en el diseño de la interfaz gráfica.....	48
2.5 Consideraciones de usabilidad, accesibilidad e inclusividad de las interfaces gestuales.....	54
2.6 Posibilidades expresivas de la gestualidad.....	55
2.7 Puntos guía para el diseño de aplicaciones con interfaz gestual.....	58

Capítulo 3.- Análisis formal de la interfaz Gestual	62
3.1 Elementos comunes de las tipologías del gesto y propuesta de tipología.....	62
3.2 Elementos formales de la interfaz WIMP que incluye gestos.....	73
3.3 Elementos formales de la interfaz gestual que incluye WIMP.....	76
3.4 Elementos formales de la interfaz gestual con lenguaje de gestos estricto.....	81
3.5 Elementos formales de la interfaz gestual con sintaxis flexible.....	83
3.6 Elementos formales de la interfaz gestual de ubicación libre.....	86
3.7 Elementos formales de la interfaz gestual con control de voz.....	89
3.8 Elementos formales de la interfaz gestual especializada en la inclusión de personas con discapacidad.....	91
3.9 Casos de estudio.....	95
3.10 Diseño de interfaz gestual a través del análisis formal propuesto.....	112
Conclusiones	114
Índice de ilustraciones	117
Índice de Tablas	119
Bibliografía	120

Agradecimientos

Un agradecimiento especial por lo que aportaron a esta tesis a mis profesores del programa de maestría en los años de 1998 a 2000, principalmente a la Maestra Luz del Carmen Vilchis con quien hicimos el esquema teórico inicial y a la Maestra Elia del Carmen Morales, cuya acertada dirección ha sido invaluable.

Gracias a todos los integrantes del Laboratorio de Multimedia del CNA, especialmente a Isaías Ortega y Rebeca Sánchez, con quienes exploré de manera práctica los problemas con que esta tesis tomó forma, y otro agradecimiento en el mismo sentido a los compañeros de trabajo en el proyecto Enciclomedia, en el ILCE especialmente a Leonardo Aguilar, Ángel González, Mónica Prieto, Alex Guzmán y Jesus Morales, con quienes compartí el asombro y la curiosidad del control gestual de la consola Wii y los problemas y triunfos de hacer multimedia para pizarrones táctiles.

Y por su apoyo incondicional, su aliento en los momentos difíciles, su acompañamiento constante, todo derivado de un amor enorme, le agradezco a mi familia: mis padres Leticia y Gonzalo, mis hermanas Leti, Vero y Ale, a mi esposa Priscila y a mi hijo Juan G., y por lo mismo gracias a mi familia extendida en Ecuador que no se queda atrás en el apoyo y cariño.

Gracias a mis amigos y colegas con quienes discutí y revisé todas las ideas de estas páginas: Yesica del Moral y Peter Aronson, Marco Sandoval, Gerardo Gómez, Omar Lezama, Edgar León, Ricardo Sánchez B., Andrés Sánchez M., Toño Malpica, Marcela García, y el PCM-Foro General, Arnulfo Zepeda, Citlalli Rojón, Emmanuel Rojón, Felipe González C. y a los integrantes activistas del capítulo profesional Siggraph-CDMX.

Dedicado a Gonzalo Maximiliano León Dorantes



“Después de más de un siglo de tecnología eléctrica, hemos extendido nuestro mismísimo sistema nervioso central haciendo que abarque el globo, aboliendo el espacio y el tiempo al menos en lo que concierne a nuestro planeta. Nos acercamos velozmente a la fase final de las extensiones del hombre -La simulación de la conciencia mediante la tecnología, cuando el proceso creativo del saber será extendido colectivamente y corporalmente a la totalidad de la sociedad humana, extendiéndose tanto como hasta donde ahora hemos extendido nuestros sentidos y nuestros nervios a través de los medios de comunicación”

Marshall McLuhan

Introducción

Las computadoras se volvieron objetos de uso masivo desde que floreció el mercado de las computadoras personales en los primeros años de la década de los ochentas y desde entonces ha crecido de manera exponencial el número de computadoras que una persona usa cotidianamente, ya que no sólo se tienen en el entorno doméstico o del trabajo, sino que se las encuentra embebidas en aparatos electrónicos, electrodomésticos, autos y hasta en accesorios de vestimenta, sólo que el usuario no las identifica como computadoras. Los profesionales en el campo del diseño y la comunicación visual han sido requeridos en este mercado (dejando aparte el papel del diseño industrial para definir forma de los objetos) ya que se trata de los que comprenden mejor el fenómeno de interacción cuando se realiza en forma distinta a la verbal. La interacción con las computadoras utiliza una forma de comunicación gráfica presentada en un medio que es llamado interfaz, como se explicará en el primer capítulo de esta tesis.

En la elaboración de la interfaz es necesario considerar tanto los aspectos relativos a los procesos de comunicación, incluyendo las características físicas y ergonómicas del usuario, como los relativos a la configuración del mensaje gráfico, en este último se articulan los códigos tipográficos, morfológicos y cromáticos que

darán sentido al mensaje, de aquí se deriva el diseño de interfaz como un campo pertinente a la comunicación visual.

Las interfaces gráficas de usuario (GUI, por sus siglas del inglés *Graphical User Interface*) se han vuelto la forma común de interactuar con la computadora. Los estudios y análisis de la GUI no solo han derivado en el establecimiento del campo de diseño de interfaz dentro de la disciplina de diseño, también han llamado la atención sobre las limitaciones de la GUI en cuanto a su eficiencia o eficacia para comunicar a la máquina con el ser humano, y han señalado el camino de la posibilidad de interfaces alternativas, donde se maneje la comunicación en niveles complementarios o superiores a la comunicación visual.

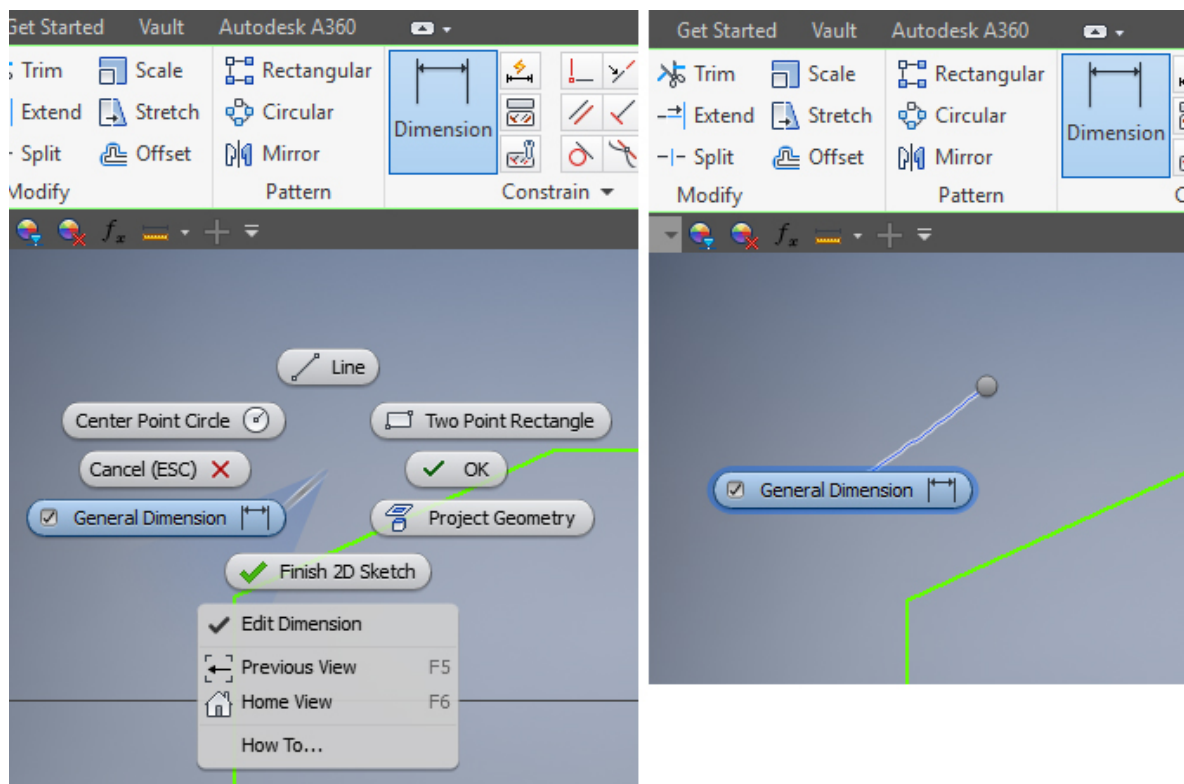
Una de esas posibilidades alternas es la que se usará para el marco teórico y hasta filosófico de ésta tesis. Proviene de las ideas expuestas por Bill Buxton¹ a partir de 1991. En esa fecha Buxton propone, en una ponencia para el Simposio Internacional sobre la siguiente generación de Interfaz Humana (FRIEND21 – Tokio, Japón, noviembre 22 a 25 de 1991) que la GUI no refleja de manera correcta la completa capacidad sensorial del ser humano, y ese reflejo del humano en la interfaz puede definirse como 3 “espejos”: El espejo de lo físico (refleja como somos los humanos, nuestras habilidades motoras y sensoriales) el espejo de lo cognitivo (refleja las habilidades para conocer, aprender, resolver problemas, del ser humano) y el espejo de lo social (refleja las habilidades del humano para relacionarse en el ámbito social).

Buxton es partidario de incluir la interacción física y expresiva en la interacción hombre-máquina, y propone calificar la eficacia de la interfaz para comunicar hombre y máquina de acuerdo a que tan bien logra incluir el ámbito físico, cognitivo y

¹William Buxton es un destacado investigador de los sistemas de interacción hombre-máquina. De nacionalidad canadiense, llegó al campo de diseño de interfaz por un camino indirecto. Su formación profesional es en la carrera de música y artes escénicas, realizó una maestría en ciencias de computadora por su interés en diseñar y construir instrumentos musicales digitales. Fue parte de los laboratorios Xerox PARC (Palo Alto Research Center) de investigación donde, entre tantos otros avances, a finales de los 70s se desarrolló la interfaz gráfica. Desde 1982 y hasta 1994 su labor principal se dividía como investigador del Computer Systems Research Institute, University of Toronto, en Canadá y como investigador en el Xerox PARC. Su biografía y publicaciones hasta la fecha se pueden consultar en <https://www.billbuxton.com/> (URL visitado en enero 2019)

social. El trabajo académico de Buxton ha sido muy influyente entre los ingenieros de *software* y científicos de computadoras, especialmente los que desarrollan trabajo en el campo de la realidad virtual. El mismo Buxton dirigió al equipo encargado de crear la interfaz del *software* Alias|Wavefront entre 1984 y 1994, donde aplicó bastantes de sus ideas sobre el uso de gestos. El resultado, en uso hasta el día de hoy, es el sistema de “*marking menus*”, que se puede ver en uso en todos los programas de diseño 3D de la compañía AutoDesk. Estos “*marking menus*” fueron una de las piezas de tecnología que adquirieron al adquirir Alias|Wavefront en 1994.

En el siguiente ejemplo, a la izquierda: “*marking menus*” en el programa de



CAD AutoDesk Inventor. Al mantener presionado el botón derecho del mouse aparecen menús alrededor del punto donde se hizo click. Las opciones que aparecen son las más relevantes, según el modo en el que está el programa, y el movimiento de selección es más corto que el de ir hasta el menú en la parte superior de la pantalla. El tiempo en que se demora en leer la etiqueta del elemento también es más corto, ya que cada elemento está bien separado y delimitado de los demás. A

la derecha: el “*marking menu*” se convierte en un gesto, que es definido como un movimiento de una parte del cuerpo que expresa una idea o significado, a diferencia de los movimientos para elegir una opción es un movimiento que se ejecuta rápidamente, al mismo tiempo que se hace click derecho se mueve el mouse en la dirección donde aparecería el botón del “*marking menu*”, se hace la elección de la acción sin necesidad de esperar a que se despliegue el menú, es decir, un gesto diagonal en un sentido significa una acción, en otro sentido significa otra, etc. El gesto se vuelve parte de la memoria muscular, puede hacerse hasta con los ojos cerrados, es más rápido y eficiente que esperar al despliegue del menú y leer las etiquetas de los elementos.

Una interfaz gestual consiste, de acuerdo a esta idea, en una forma de comunicar órdenes a un programa de computadora donde el usuario utiliza movimientos y ubicación en el espacio y en el tiempo. Este conjunto de movimientos y cambios en el tiempo pueden ser por el uso del cuerpo o de una parte del cuerpo, o por el uso de un aparato que se manipula, como una batuta o un aparato apuntador, y que hace de extensión corporal. La Interfaz gestual se presenta como un extensión sensorial y expresiva. Hay un elemento de naturalidad en el gesto como forma de controlar una computadora, especialmente cuando se usa para interactuar con los ambientes de realidad virtual tridimensional, estos ambientes ya desde su aparición han marcado la tendencia como el tipo de interfaz con más probabilidad de substituir las GUI en el futuro.

Desde la introducción de la GUI en los laboratorios de la Xerox PARC no se han hecho muchos cambios en el diseño de la metáfora inicial, conocida como metáfora WIMP, que significa *(W)indows, (I)cons, (M)enus and (P)ointing device* (en español, Ventanas, íconos, Menús y aparato apuntador). Esta metáfora está basada en elementos de interacción que dividen la pantalla en áreas similares a ventanas, en menús y cajas de diálogo que despliegan una lista de acciones cuando son seleccionados, y en algún aparato especializado para señalar.

Esta interfaz fue revolucionaria en su tiempo, y su efectividad se comprueba por el hecho de que ya ha cumplido 30 años prácticamente sin variaciones. Sus ventajas más reconocidas son: que hace más natural la interacción con las computadoras, que acorta la curva de aprendizaje de los nuevos programas y que se percibe de una forma más agradable. Sin embargo, estas ventajas vienen por la comparación con la interfaz de línea de comando, o “modo de texto”; y hay que tener en mente que dicha comparación se dio en la fecha de la introducción masiva de la interfaz WIMP, en el año de 1984. Hoy día la interfaz de línea de comando no ha desaparecido del todo, se mantiene como la preferida por una minoría de usuarios avanzados (o nostálgicos), y la gran mayoría de los usuarios está usando GUI. La masificación de la experiencia con GUI en las computadoras personales hace que estas ventajas se vuelvan algo que se da por sentado, lo que en inglés se llama “*commodity*” y que se podría traducir al español como característica básica, la masificación ha puesto en evidencia vicios y fallas de la metáfora WIMP, que van desde falta de velocidad en la ejecución de operaciones repetitivas (es ineficiente) hasta la discriminación de usuarios con discapacidad. Aunque se conocen estos vicios y fallas casi desde la primera vez que se implementó la GUI de forma experimental, no se ha visto con el paso de los años que se haya desarrollado una interfaz alternativa que sustituya exitosamente la GUI, y después de tantos años, lo sorprendente es que la metáfora WIMP siga en uso cuando pudiera considerarse agotada e incluso obsoleta. Cabe preguntarse si no se debe a que aunque las interfaces alternativas muestran avances rápidos y continuos en la parte tecnológica, a nivel de medio de comunicación no han sido igualmente investigadas y desarrolladas.

La interfaz gestual de Buxton que mostramos como ejemplo anteriormente fue una de las primeras alternativas exitosamente implantadas, derivada de investigaciones académicas cuyo énfasis es la ergonomía y la eficiencia del trabajo a la hora de usar una interfaz, pero que en lo más profundo es sólo un primer acercamiento a una sustitución completa, de acuerdo a lo postulado en el ensayo de los 3 espejos de la interacción [BUXTON 1994] aún no alcanza esa inclusión total de las

posibles interacciones del humano con la máquina, sin embargo fue el punto de inspiración para múltiples interfaces gestuales que exploran y amplían la capacidad de comunicarse con las computadoras, yendo progresivamente de interfaces donde el gesto no es tan relevante a interfaces donde el gesto es el medio principal y más relevante de la interacción hombre máquina.

La hipótesis de este trabajo de investigación postula: dado que las interfaces gestuales constituyen medios de comunicación visual, su diseño requiere del conocimiento de la taxonomía y los elementos formales que constituyen estos medios.

Para entender e impulsar el diseño de las interfaces gestuales es un punto de partida clasificar las formas que toma el gesto en la interacción con las computadoras y proponer una tipología. La clasificación de las formas de comunicación con las computadoras a través de gestos se basará en el nivel de importancia del gesto dentro de toda la interacción, considerando al conjunto de gestos como un lenguaje con el que se da la comunicación humano-máquina.

El alcance de esta tesis es el de conocer los límites dentro de los cuales se puede diseñar una interfaz gestual, de aquí que se utiliza un método deductivo, en el que se inicia a partir de lo general, el estudio de la interfaz, a lo particular.

La investigación se conseguirá por una metodología de análisis formal cualitativo, revisando los análisis formales de las implementaciones técnicas del gesto en forma histórica en el campo de diseño de interfaz desde el punto de vista de desarrollo de sistemas de computación, y uniéndolo al análisis del gesto realizado en la disciplina de la psicología en el campo de la proxémica y de los estudios de psicología del aprendizaje lingüístico, que tiene un fuerte componente de análisis de significado lingüístico y semiótico. La validación provendrá de aplicar la taxonomía resultante en el análisis de ejemplos de interfaces existentes, identificando sus elementos formales y su correspondencia a los tipos propuestos. La tipología provendrá de un análisis cualitativo de los elementos que pueden corresponder a los siguientes conjuntos:

Elementos gráficos: las representaciones visuales de las interacciones de los usuarios con los programas de computadora.

Elementos mecánicos de interacción: controles, pantallas táctiles, etc.

Elementos de comunicación social: que hacen de la interfaz un facilitador de la comunicación para socializar entre grupos humanos.

Elementos de inmersión tridimensional que son necesarios cuando la interfaz abarca más allá de un solo plano de representación

Elementos específicos derivados del reconocimiento de gestos, ya que la metáfora gestual tiene elementos exclusivos que no son necesarios en otras metáforas de interfaz.

Se plantea como objetivo general: valorar las características del gesto como sistema de comunicación hombre-máquina, mediante una tipología de interfaces gestuales. Para ello se proponen los siguientes objetivos particulares:

Identificar los elementos formales de las interfaces basadas en el reconocimiento de gestos.

Comparar las interfaces basadas en la verificación de gestos con las interfaces gráficas de metáfora WIMP

Determinar una tipología de las interfaces gestuales, tomando como criterio de diferenciación de los tipos de interfaz el concepto de Bill Buxton de los 3 espejos de la interacción: reflejo del ser humano en lo físico, lo cognitivo y lo social

Contribuir a la discusión que se está llevando a cabo sobre la validez y obsolescencia de las interfaces gráficas de usuario.

De acuerdo a lo anterior, la investigación está estructurada en tres capítulos, el primero revisa los conceptos generales de la interfaz; las características de la interfaz gestual y particularidades del gesto en el diseño de la misma se revisan en el segundo capítulo. Finalmente, el tercer capítulo aborda los elementos formales de la interfaz gestual y la propuesta taxonómica de la misma, aspectos que permiten valorar su importancia con la interacción hombre-máquina. Los estudios de caso incluidos en ese capítulo ejemplifican algunas de las posibilidades y aplicaciones de la interfaz gestual.

Capítulo 1 – La interfaz – conceptos generales.

1.1 La interfaz como medio de comunicación.

Una interfaz consiste en todos los elementos que representan un control. Esta definición se toma directamente de la terminología de ingeniería de procesos [OPPENHEIM, et al 1998] y nunca debe ser confundida con la interfase²

En la ingeniería de procesos, que es la disciplina que se encarga de estudiar las acciones con las cuales se desarrolla un trabajo, es donde aparece por primera vez el concepto de interfaz. Un trabajo, en su enunciación más simple, es la aplicación de energía en un periodo de tiempo. En forma práctica, el trabajo es el resultado de la aplicación de energía para ejecutar transformaciones. Estas transformaciones son genéricamente llamadas procesos, y dichos procesos siempre se emparejan a un sistema de control. [OGATA 2003]

² Hay cierta confusión sobre la grafía de la palabra interfaz y su plural interfaces. La confusión procede de que el término interfase, y su plural interfases, también se utilizan en el ámbito de computación. Las interfases, para los programadores, son aquellas piezas de código que se ejecutan al trasladar datos de un estado (o fase) a otro. La interfaz, sin embargo, se refiere a todo aquello que sirve de intermediario entre el usuario y el programa que controla. De latín Inter-Entre y Facia-Cara y su plural es interfaces, como el modelo haz-haces, paz-paces etc.

De manera estricta, una interfaz es la parte de un sistema de control que presenta la información relevante al estado de un proceso y en la interfaz se puede interferir, modificar o aprobar los procesos en desarrollo. [MAYBURY 1993] Es decir, una interfaz consta tanto de una manera de representar información como de instrucciones o guías de acciones a tomar. La interfaz se puede abstraer, por lo tanto, como un medio de comunicación. En ella se dan fenómenos de comunicación que pueden ser estudiados de la misma manera que los fenómenos del lenguaje y la comunicación en otros niveles de la interacción humana. A continuación se compara el diagrama clásico de Jakobson con un diagrama simplificado de comunicación entre hombre y máquina.



Diagrama de Jakobson

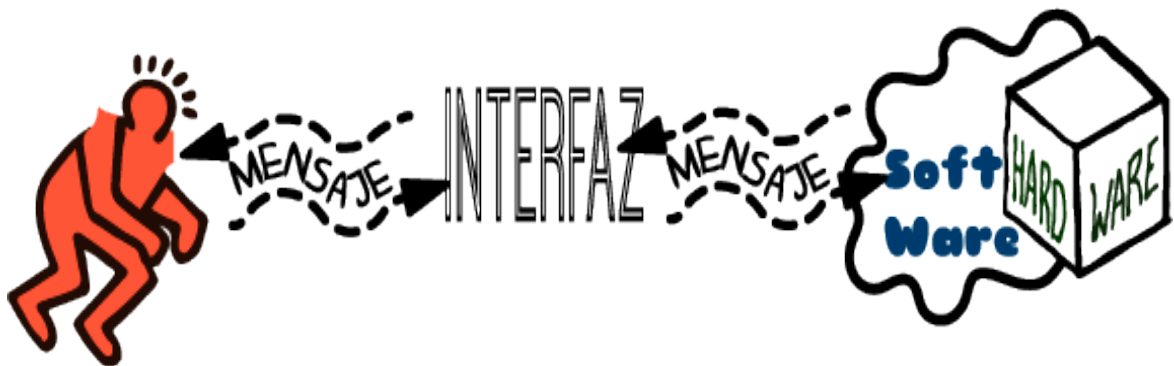


Diagrama de Jakobson adaptado a la comunicación entre humano y máquina

Ilustración 1: Comparación del Diagrama de Jakobson con los procesos de comunicación a través de una interfaz

El esquema clásico de un proceso de manufactura industrial (y que por analogía se extenderá a los procesos de computación) implica la existencia de un circuito de operación, un circuito de comando, un circuito de retroalimentación y un circuito de transporte. La interfaz (o las interfaces, porque puede haber más de una) de control permite comprender el estado de los múltiples circuitos para informar al operario (o al supervisor, en caso de un sistema automático) sobre las acciones que se requieren, los cambios de estado y los puntos álgidos del proceso. [OGATA 2003]

El siguiente esquema describe este concepto

Ilustración 2: Esquema de operación y control de un proceso industrial

Los elementos que constituyen una interfaz, en este esquema son:

- elementos informativos, por ejemplo, medidores, luces y etiquetas
- elementos de control, por ejemplo, reguladores de velocidad, de temperatura, etc.
- elementos de operación, por ejemplo, interruptores, llave de inicio/paro, etc.

En el concepto de interfaz industrial, los tres tipos de elementos que derivan de los tres tipos de circuito en los que se realizan los procesos son específicos del proceso. Esta cualidad unívoca, específica del tipo de mensajes que viajen por la interfaz es una cualidad que no existe en el mundo del *hardware* y *software*.

En un principio el concepto de interfaz de las computadoras se derivó de manera natural desde este concepto de interfaz para control de procesos industriales. [MARCUS 1992] La interfaz consiste, pues, en los elementos por los cuales un usuario recibe información de lo que pasa con un programa de computadora, y los elementos con los cuales él es capaz de darle instrucciones a la computadora. Pero la interfaz de computadora tiene que contemplar características que no existen en los sistemas industriales, específicamente por la naturaleza no específica de los procesos que realiza, naturaleza a la que muchos autores se refieren como "*sistema abierto*". [MANDEL 1994]

Al volver a revisar el esquema de la comunicación de Jakobson, y aclarando un poco más el esquema de Ogata de la ilustración 2, se puede entender que el mensaje que se construye dentro del *software* está dividido en elementos que informan el estado de la máquina, elementos que comunican el comando de control de la máquina, y elementos que comunican cuál es el siguiente paso de operación de la máquina. En la comprensión de este mensaje, el usuario manipula la interfaz para a su vez construir un nuevo mensaje de entrada, indicando a la máquina que reconoce en qué estado se encuentra, escogiendo la nueva operación que requiere y afinando el estado de control. Lo ideal es que este circuito de comunicación tenga un equivalente (a manera de un reflejo en un espejo) de las capacidades sensoriales y cognitivas humanas. [BUXTON 1990A]

A diferencia de un sistema industrial, donde la interfaz que controla, por ejemplo, un proceso industrial de enlatado de alimentos no puede usarse para controlar un proceso industrial de manufactura de televisiones, la interfaz de control de una computadora debe tomar en cuenta que el "*sistema abierto*" significa que la interfaz debe servir igual si la máquina está ejecutando un programa de

conteo estadístico en un momento y en un momento posterior un programa de cálculo astronómico. La interfaz de computadora se aplica en sistemas cuya meta no está dirigida a un solo tipo de acción ni determina el tipo de resultados de manera unívoca.

La computadora es una máquina dedicada a procesar datos, y considerada como máquina ejecuta un trabajo que debe ser controlado. La interfaz posibilita al programador y al usuario de los programas establecer un canal de comunicación a través del cual se mantiene el control y se obtienen los resultados del trabajo solicitado. En una máquina industrial, la interfaz es dependiente del tipo de proceso y no puede aplicarse a un proceso distinto. Sin embargo, en la computadora se espera que la interfaz permanezca aunque cambie el tipo de proceso que se está ejecutando. Por ello se convierte también en un ambiente, un lugar general en el cual se posibilita la comprensión de los procesos sin especificidad. [MANDEL 1994]

La interfaz, para resumirlo con otros términos, representa información sobre el control de los procesos y las actividades a tomar por parte del operador de la máquina (es decir, el usuario) y además provee el ambiente donde se pueden programar nuevos procesos.

Las interfaces de los programas de computadora mantuvieron la constante de elementos de información, de control y de operación, lo que se puede atribuir a una inercia heredada de las interfaces industriales; por mucho tiempo cada programa proponía una interfaz propia. Este es un fenómeno que continua sucediendo, aunque cada vez con menos frecuencia.

La interfaz, en la computadora, ha dependido de los avances tecnológicos. Durante mucho tiempo, la interfaz más común no tenía nada que ver con interfaces visuales. Estas primeras interfaces consistían en interfaces mecánicas, como los tableros de switches, las tarjetas perforadas y el teletipo. Aún después de la adopción de las pantallas como sistema de despliegue de la interfaz, estas seguían manteniendo apariencias y mecanismos que se desprendían de la concep-

ción original de interfaz. No era extraño que cada programa tuviera una interfaz propia, y que los métodos de representación de los procesos fueran muy dispares [BOLTER, GRUSIN 2000]. Esta situación en términos lingüísticos, se puede describir como que cada interfaz es un dialecto, y las acciones comunes entre los programas recibían asignaciones semánticas distintas. La interfaz gráfica no es otra cosa que uno entre varios tipos posibles de asignación semántica [BUXTON 1983].

1.2 La interfaz gráfica

Como no deja de tratarse de interfaces, una interfaz gráfica contiene los elementos que contiene cualquier interfaz, esto es, elementos de información, elementos de operación y elementos de control de los procesos que representa. Lo que cambia es el nivel de representación: se utiliza una representación gráfica. Esto quiere decir que puede haber otros tipos de representación. Por ejemplo, una interfaz visual puede representar los elementos por indicadores luminosos; una interfaz mecánica consiste en botones, palancas y diales que tienen un accionar mecánico; y así sucesivamente.

Para recibir el apelativo de interfaz gráfica, la interfaz debe presentarse con características, en primer lugar, visuales, y en segundo lugar mediante una representación interpretativa del significado, es decir, usando una codificación propia del lenguaje gráfico.

Muchos programas usaron desde hace mucho tiempo interfaz gráfica, pero la adopción generalizada de este tipo de interfaces se produce cuando se consigue la confluencia de dos factores: primero cuando los avances tecnológicos hacen posible la existencia de pantallas con capacidad de desplegar información gráfica, y segundo cuando el sistema operativo establece un modelo estandarizado de representación gráfica.

Gestación de la interfaz gráfica del sistema operativo.

Ya que no es el objetivo primordial de esta tesis, pero es necesario para entender las interfaces gestuales como la evolución de las GUI, se mencionará una breve historia de como se gestó la interfaz gráfica.



Ilustración 3: Izquierda - Ivan Sutherland con SketchPad, 1962. Derecha - Douglas Engelbart con On-Line, 1968

En primer lugar, la cuna de las interfaces gráficas es universalmente reconocida en el PARC (*Palo Alto Research Center*) de la compañía Xerox. PARC desarrolló una interfaz a partir del trabajo de precursores, concretamente del *On-line System* de la Universidad Stanford (para el cual se inventó el primer *mouse*) y el *SketchPad*, sistema de Diseño Asistido por computadora de la compañía de ingeniería Sutherland que utilizaba un lápiz de luz como aparato indicador para realizar dibujos técnicos. *On-line* proveyó de la idea de hipertexto y cursor, mientras que *SketchPad* propuso la idea de elementos gráficos para información y operación.

Otro elemento que *On-line* hizo patente es que las interfaces gráficas requerían de aparatos especializados en apuntar. El equipo de Stanford era dirigido por Douglas Engelbart, a quien se da crédito como el padre del aparato que se necesitaba para la manipulación de elementos gráficos: el primer “*ratón*” o

“mouse”³. La presentación del sistema de Engelbart causó un cierto revuelo en su demostración ante un público reducido en el año de 1968. El sistema englobaba la idea de hipertexto, la equivalencia gráfica de elementos que hasta ese momento existían sólo como *hardware* (botones, barras de desplazamiento), el emparejamiento de un procesador de datos con un procesador de imágenes y por primera vez hizo su aparición un cursor controlado por un *mouse*. [ENGELBART 1968]

Lo realmente revolucionario de la propuesta GUI de PARC estaba en que se planteaba como una capa de interfaz a nivel de sistema operativo, esto es, como un marco de trabajo que establecía un lenguaje único para el uso de todos los programas que se fueran a ejecutar en la computadora. La primera computadora que usó este avance fue modelo Alto, en 1973, y el sistema de interfaz gráfico se llamaba Star. Este sistema contiene todos los elementos básicos de las GUI: reglas específicas de representación estandarizada de las acciones de las que es capaz el sistema operativo, señales de control y de información.

Una de las características que se derivaron de este acercamiento a la interfaz fue la necesidad de demostrar "modo", es decir, de establecer claramente para el usuario cuando la interfaz está en modo informativo, cuando está en modo de control y cuando está en modo de operación. Es por ello que a estas interfaces también se les conoce con el adjetivo de "multimodales".

La estandarización de la representación gráfica es un paso muy importante, dado que fue la condición sin la cual no se conseguiría el predominio comercial en pocos años, fue con lo que se alcanzó la base estable necesaria para que la adopción de la GUI de forma comercial fuera factible incluso en su implementación en múltiples plataformas tecnológicas.

³ Engelbart y su equipo aceptan que nunca se les ocurrió otro nombre para el singular aparato. En castellano, algunas personas tienden a favorecer el uso de la palabra en inglés precisamente para distinguir de cuando se habla del mamífero roedor. Sin embargo, el término traducido directamente al español es igual de común.

Fortalezas de la GUI de sistema operativo

Varios programas aparecen en el lapso de 1973 a 1984 que proponen una GUI como base de la estandarización de la representación de procesos. Por mencionar algunos, NeWS, que usaba el lenguaje postscript en computadoras SUN en 1980, DeskMate, en computadoras TRS-80 de Radio Shack en 1982, GEM de Digital Research, desarrollado desde 1979, que funcionaba en computadoras CP/M y IBM PC, y que fue seleccionado como el GUI de la Atari ST en 1985 y el GEOS en las computadoras C64 de Commodore Business Machines en 1986. Pero es el MacOS de Apple, en 1984 el que tuvo el primer éxito comercial indiscutible. [BOLTER, GRUSIN 2000][LISTER 2003]

El éxito del MacOS como GUI de la nueva computadora Macintosh de la compañía Apple puso en relieve las ventajas de la GUI, que son las siguientes:

Se establece una manera estándar de programación de modos, esto es, de tener representaciones de información, control y operación para los diferentes programas. Por ejemplo, reloj de arena que indica modo de espera, cambio de cursor que indica modo de operación, pantallas de alerta que indican modo de información, etc.

La estandarización libera a los desarrolladores de *software* de la necesidad de programar interfaces.

La estandarización fomenta una experiencia de usuario única, lo que reduce el tiempo de aprendizaje de nuevos programas. A este fenómeno se le conoce como "nivelación de la curva de aprendizaje".

Se depende menos de la mnemotecnia.

Se establecen las características ergonómicas del mouse y el cursor como aparatos de interacción hombre-máquina [APPLE, 1995]

1.3 El paradigma WIMP (*Windows, Icons, Menus, Pointing Device*)

Las interfaces gráficas derivadas del Star de Xerox han dado pie a que se pueda reconocer un lenguaje visual cuyos elementos comunes le han dado nombre al paradigma como interfaz WIMP

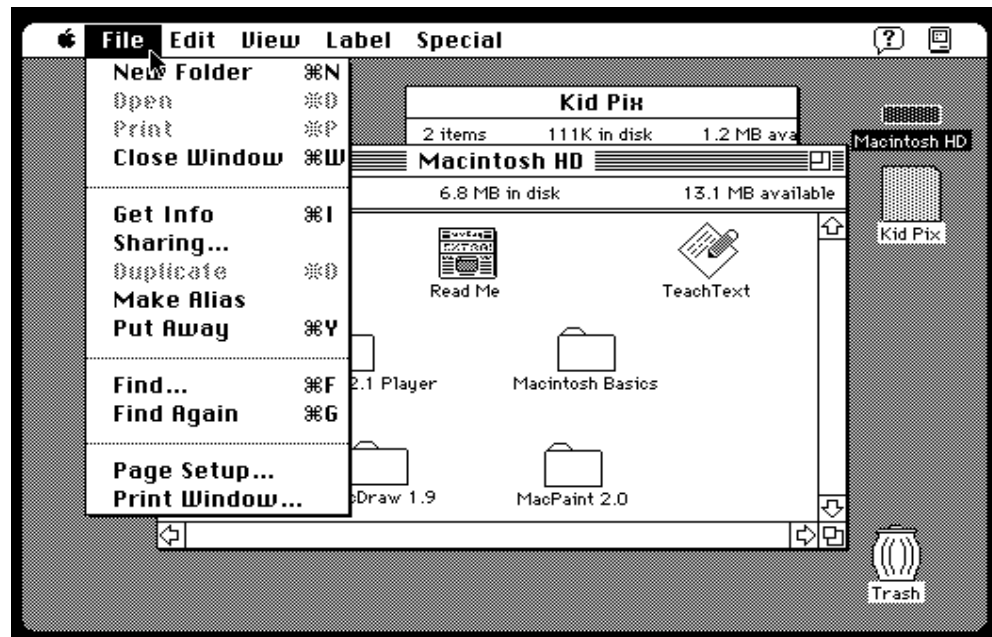


Ilustración 4: WIMP: *Windows, Icons, Menus and Pointers*

Los elementos son:

a) Los procesos que requieren comunicación con el usuario generan una representación visual en la pantalla, un espacio que se llama Ventana (*window*). Este espacio codifica los demás elementos de la interfaz. Los procesos simultáneos pueden verse de manera simultánea, y los cambios visuales establecen los cambios de modo (activo/inactivo, alerta/reposo, etc.) Se establece además un sistema jerárquico, donde los subprocesos son ventanas que están dentro de la ventana que representa el proceso superior del que derivan (por ejemplo: la ventana del programa contiene las ventanas de los documentos). Esta jerarquía incluso implica que todas las ventanas residen en la ventana del Sistema Operativo, que normalmente es llamada el escritorio, el área de trabajo, o algún término similar.

b) Representar acciones mediante íconos, y ejecutar esas acciones al elegir con el *mouse*. La representación icónica debe interpretar la acción a realizar y por lo tanto tener su propio campo semántico asociado. Así, un ícono que se ve como una impresora debe conducir al programa que maneja la impresora, un ícono que representa una brocha conduce al programa de pintura, etc. Este campo semántico tiene elementos claramente establecidos, a los que se llama "íconos del sistema" y cuya representación visual está delimitada por un manual de estilo y que normalmente se consideran invariables, constantes, y los "íconos de usuario" cuyo único campo semántico es que llevan a acciones, pero cuya representación visual puede ser modificada a voluntad del usuario. Uno de los efectos interesantes de este acercamiento es que los íconos, por un lado, tienen una carga de significado aprendida: el usuario sabe que todo ícono es una interacción con la computadora y todos los usuarios del mismo sistema reconocen los íconos y las acciones que producen en cualquier computadora en la que los encuentren y por el otro lado una carga de significado personal que se presta a la expresividad individual: un usuario puede escoger ciertos íconos por su carga emocional o por su cierre semántico, que no necesariamente son los mismos para alguien que no es usuario de ese sistema en particular

c) Los menús describen todas las operaciones del programa. Los menús pertenecen a la ventana en la que se encuentran asociados. El manual de estilo establece una jerarquía estricta, por ejemplo, para que el primer menú de cualquier ventana siempre sea el mismo y contenga los mismos elementos, con lo que se garantiza una experiencia consistente de ventana a ventana, y se elimina la dependencia de la memoria para dar con las instrucciones precisas para establecer las operaciones a realizar. Los menús pueden flotar libremente para establecer otros modos de operación, por ejemplo, cajas de alerta para el modo informativo y cajas de opciones para el modo de control de procesos. Estos otros menús se conocen como cajas de diálogo.

d) Aparato para indicar. En el más alto nivel metafórico, el cursor transmite la información con la que el usuario quiere que el programa haga el proceso de datos. El cursor es controlado por un aparato que puede señalar un punto preciso en la pantalla y así el sistema operativo puede interpretar, por el momento y la posición de dicha señal, qué acción ejecutar, ya sea de operación o de control. El más conocido de los aparatos que realizan esta función es desde luego el *mouse*, pero existen otros aparatos que se usan de igual forma: el *track ball*, las tabletas digitalizadoras, los *joysticks*, etc.

Se le ha llamado paradigma WIMP a este modo de interactuar por las iniciales de los elementos: *Windows*, *Icons*, *Menus*, *Pointer*. Los nombres de estos elementos son indicadores del referente visual o metafórico con el que se construyen, y efectivamente esta interfaz logra integrar toda la gama de acciones y mensajes entre usuario y máquina con esos cuatro elementos y sus variaciones. [LAUREL 1995][FISHER 1994]

1.4 Problemáticas de la interfaz gráfica WIMP

Por su persistencia en el tiempo desde su adopción generalizada el paradigma WIMP se considera como un gran éxito, pero no ha estado exento de críticas en estudios que señalan problemáticas diversas en cuanto a su ergonomía y eficiencia, y también se ha hecho patente que no es adecuado a nuevos modos de uso de las computadoras.

Estas son las problemáticas más relevantes para el objetivo de comparar la interfaz GUI con las interfaces gestuales y entender la racionalización de cómo las interfaces gestuales pueden ser un sustituto viable de las GUI en un futuro cercano:

- I) Pobre desempeño en la presentación visual de información multidimensional.
- II) Ineficiente en la automatización de tareas repetitivas.

III) Ineficiente como extensión sensorial.

IV) Ineficiente como sistema de comunicación hombre-máquina

V) Ineficiente como sistema de socialización.

I) Pobre desempeño en la presentación visual de información multidimensional.

Edward Tufte ha estudiado el problema de la presentación visual de información cuantitativa y ha llegado a conclusiones interesantes, una de ellas es sobre la multidimensionalidad de las representaciones gráficas de información. [TUFTE 1983][TUFTE 1990]

Una gráfica simple, por ejemplo una típica gráfica de “pastel”, presenta en un solo vistazo dos dimensiones: la dimensión comparativa de cada pedazo del pastel con el todo y la dimensión comparativa entre cada pedazo del pastel. Mientras más información queremos mostrar, el problema de cómo mostrar cada dimensión y su relación con las demás dimensiones se complica. Por ejemplo, si en un mapa del metro es necesario informar, adicionalmente a las estaciones que conforman cada una de las 14 líneas que existen, las distancias a recorrer en kilómetros entre cada estación, los tiempos de llegada y salida del primer y del último tren, y las posibles conexiones con otros sistemas de transporte públicos, como líneas de autobuses o de tranvías.

Para Tufte la solución de gráficas multidimensionales es un problema exclusivo de diseño gráfico. Para solucionarlo, propone seis principios: Escapar del plano (usar representaciones tridimensionales, como maquetas, realidad virtual, etc.), establecer niveles de lectura de la información a nivel micro y macro, utilizar capas y separaciones, utilización de pequeños múltiplos, ligar color e información, y establecer narrativas de espacio y tiempo.

Aunque Tufte establece estos principios con el enfoque de las representaciones gráficas impresas, donde se entiende como gráficas las que son el re-

sultado o la conclusión de una serie de operaciones, y con las que se presenta dicho resultado como mensaje final y total, sus consideraciones se pueden adaptar (y de hecho se han adaptado) a las interfaces dentro de los sistemas de cómputo, puesto que se trata de manifestaciones gráficas que muestran resultados y que tienen que incluir múltiples dimensiones informativas.

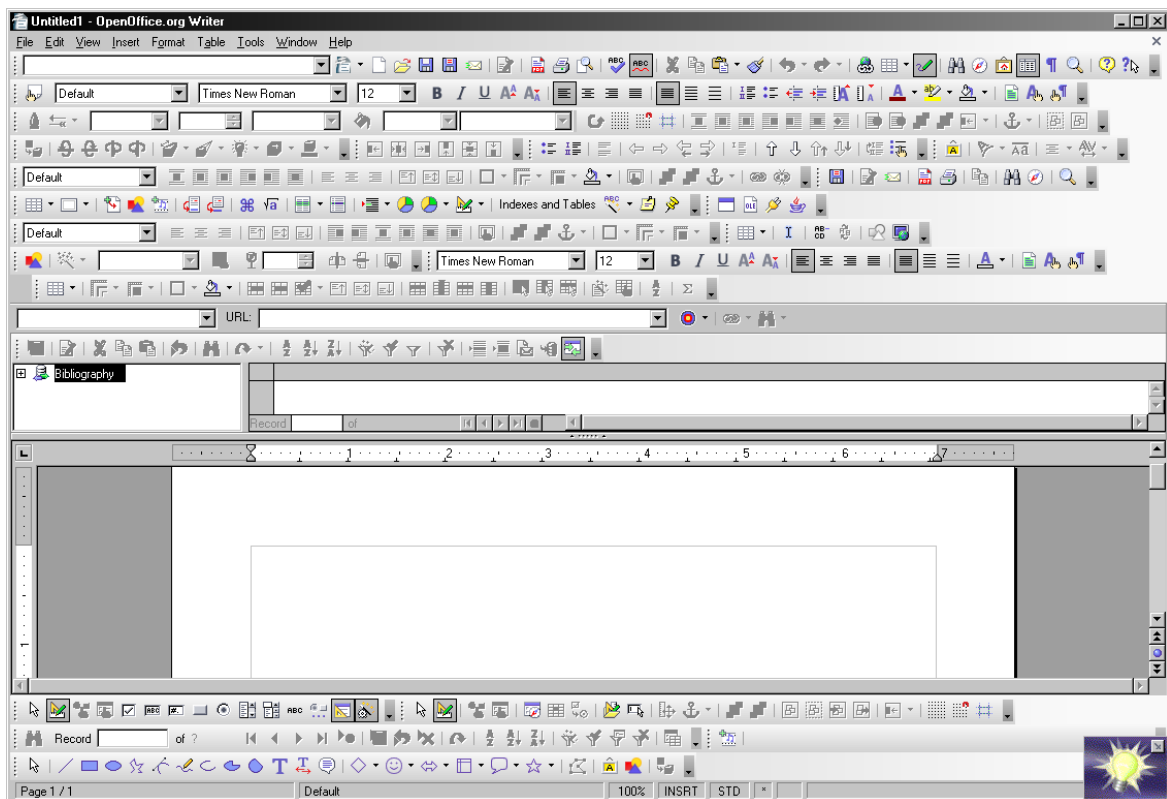


Ilustración 5: Procesador de textos con todas las herramientas abiertas. Lo inadecuado de WIMP para comunicar múltiples dimensiones se hace evidente

Para ejemplificar esta adaptación se puede postular una interfaz que requiere transmitir las capacidades de un programa medianamente complejo, por ejemplo, un procesador de textos (Ilustración 5). Cada ícono muestra una actividad posible, pero todos los íconos compiten por espacio (que es un recurso escaso) en una pantalla y también requieren ser lo suficientemente diferenciados de manera que el usuario

recuerde el significado de cada ícono o pueda deducirlo por interpretación del gráfico usado.

Además, la interfaz tiene que mostrar información sobre el modo de proceso de forma continua, el usuario sabe si está en modo de ingresar texto, de ejecutar operaciones sobre archivos, de esperar resultados. La interfaz tiene que mostrar las acciones obvias y cancelar las acciones inapropiadas, tiene que sugerir las acciones futuras posibles y enviar todos estos mensajes de manera clara y sin equívocos.

Al aplicar el sistema de análisis de Tufte, esta interfaz solo tiene una dimensión, un sólo nivel de comunicación en cada elemento. En un momento dado, hay mas de cien íconos distintos a la vista al mismo tiempo, además de que duplican funciones en el menú superior. Los íconos que representan lo que se puede realizar con el programa no descubren su función hasta que el usuario hace aparecer una explicación mientras deja el cursor sobre ellos para leer una línea de explicación (el "*tool tip*" o "guiño sobre la herramienta"). Aunque se espera que el usuario tenga cada vez que ocupar menos tiempo en recordar la función de cada ícono, la realidad es que la interpretación de las imágenes icónicas es complicada e inconsistente. [NIELSEN 1993]

II) Ineficiente en la automatización de tareas repetitivas.

Bill Buxton encontró, en investigaciones que realizó entre 1987 y 1989, que una de las deficiencias más sobresalientes de la interfaz WIMP era la lentitud con la que las tareas repetidas se llevaban a cabo [BUXTON 1990B]. Esto se debe principalmente a que la interfaz WIMP no toma en cuenta el uso de otras capacidades humanas más allá de sus habilidades cognitivas. Como Buxton mismo hace notar, el ser humano tiene memoria muscular, que le permite realizar tareas de manera automática y optimizar su desempeño en tareas para las que ya conocemos el resultado. Al manejar un coche, una vez pasada la etapa inicial de familiarización con los movimientos para cambiar de velocidad acelerando o

desacelerando, se realizan dichos movimientos de manera inconsciente. Aunque lleva mucho tiempo tener una imagen mental del un teclado QWERTY de memoria, finalmente se logra escribir rápidamente porque los dedos saben como moverse para producir la letra que se necesita, sin tener que voltear a ver el teclado. En contraste con esta forma de aprender la interfaz, es claramente inadecuado tratar de escoger cualquier elemento de un menú de una ventana "sin ver". Con esto se hace patente cómo la interfaz WIMP no ha explotado la cualidad de memoria muscular que se describió en los otros casos. A la larga esto lleva a los usuarios avanzados o "*power users*" a demostrar que es más eficiente el uso del teclado que el del *mouse* cuando lo que se pretende es tener un desempeño rápido. La definición de eficiencia es la de la función inversa del trabajo y el tiempo, esto es, la mejor eficiencia es la que realiza mayor trabajo en menor tiempo. Además, según la teoría de Buxton sobre la correspondencia de la interfaz con el modo de aprender del humano, se está omitiendo la posibilidad de hacer espejo del control de la computadora mediante control físico y control muscular.

III) Ineficiente como extensión sensorial.

La interfaz WIMP favorece el sentido de la vista sobre otros sentidos humanos. Buxton y otros han realizado estudios que demuestran que hay ventajas en la inclusión de otros sentidos en el manejo de computadoras.[WACHS et al, 2011] Claramente el uso de señales auditivas, por ejemplo, puede agregar dimensiones al conjunto de información que se maneja para indicar errores en los procesos, en los controles o en la entrada de datos [NIELSEN 1993]. Aún no se ha generalizado el uso de comandos de voz para controlar una computadora, aunque el *hardware* existente ya tiene la capacidad para reconocer voz, el tener que integrar esta capacidad al paradigma WIMP es sumamente complejo.

Bill Buxton plantea que las computadoras, como cualquier tecnología que ha usado el ser humano, son una extensión de las capacidades humanas [BUXTON

1994], en una reflexión que es similar a la que realiza McLuhan sobre los medios de comunicación. Mientras que McLuhan considera que los medios de comunicación son la extensión de nuestro sistema nervioso [MCLUHAN 1997], Buxton va más allá al proponer que las computadoras son extensiones de nuestros procesos cognitivos, y que las computadoras se proyectan con capacidades cada vez mayores (acercándose y asemejándose a las capacidades humanas) para extraer datos con múltiples "sentidos virtuales", por ejemplo, dándole a la computadora capacidad de procesar video, sonido, comunicarse por redes, conocer su geoposición etc.; sin embargo esas capacidades cognitivas de la máquina se ven pobremente reflejadas en los niveles de interacción con los seres humanos.

El autor manifiesta que se debe diseñar la interfaz con las máquinas con las mismas capacidades que ya tiene la máquina para establecer rutas cognitivas. En [BUXTON 1994] El autor declara que hay que considerar la tecnología en términos de la fidelidad con que refleja las capacidades humanas, y que dichas capacidades se especifican en tres niveles: Nivel físico: cómo está construido el ser humano y que habilidades motoras y sensoriales posee. Nivel cognitivo: Cómo piensa, aprende, resuelve problemas y transmite conocimientos el ser humano y Nivel social: Cómo se estructuran los grupos, cómo se estructura la dinámica entre individuos, la dinámica del poder, de la política y en general la estructura de la comunicación en el ámbito social.

Se considera entonces que una interfaz es mas efectiva según qué tan bien refleja las capacidades humanas en las interacciones con las máquinas en la medida de estos tres niveles. Al revisar bajo este criterio la interfaz de metáfora WIMP se llega a la conclusión de que como extensión del ser humano es mejor que la la interfaz de texto, pero que no refleja en alto grado todas las capacidades del ser humano. Con el paso del tiempo el progreso sostenido de la tecnología que está llevando a las computadoras un nivel más humano en sus capacidades de percepción, haciendo más evidente el rezago de la interfaz WIMP para ofrecer interacciones equivalentes a esa percepción.

IV) Ineficiente como sistema de comunicación hombre-máquina.

Toda interfaz de paradigma WIMP puede analizarse como un campo semántico, en el cual existen un conjunto de reglas específicas que rigen y facilitan la comunicación de los objetivos del programador de un *software* específico hacia el usuario de dicho *software*. El campo semántico se establece de antemano, es la normativa estandarizada del sistema operativo, y deja muy poco espacio a las manifestaciones comunicativas expresivas, sensoriales y hasta poéticas. De hecho, cuando las aplicaciones de *software* requieren de un mayor nivel de comunicación, es decir que requieren ampliar su campo semántico, casi siempre lo hacen proponiendo una interfaz que se sale de los límites de la norma de interfaz del sistema operativo, aún dentro de los postulados de una interfaz WIMP, pero proponiendo una separación conceptual de la interfaz estándar, usando casi siempre una diversificación del diseño gráfico. Es decir, se considera que se mejora la comunicación si se rompen las reglas del campo semántico estandarizado.

Un ejemplo directo es cómo prácticamente todos los juegos proponen una interfaz no estándar, ya que su campo semántico está muy alejado del que enmarca la interacción del sistema operativo.

Es interesante verificar como las consolas de juego, que son computadoras de gran potencia en si mismas, pueden proponer gran cantidad de las acciones que una interfaz de GUI propone, sin utilizar un sistema tan cerrado y reglamentado [ANDERSEN, et al, 1993].

Otro ejemplo que hace evidente esta insuficiencia es el de la propuesta de interactividad vía Internet, que ha hecho hablar a algunos teóricos sobre la emergencia de un "sistema operativo de Web" [BIOCCA, LEVY 1995][BOYLE 1997]. La interactividad por internet impacta en los sistemas operativos tradicionales al integrar la flexibilidad y ubicuidad comunicativa, sus primeras manifes-

taciones se ven en los cambios de diseño de las páginas web para acomodar interacción en computadoras y aparatos móviles por igual (lo que recibe el nombre de “diseño responsivo”) sin seguir estrictamente las reglas de estilo de la GUI del sistema operativo y en algunos casos incluso supliendo las carencias de éste en, por ejemplo, problemas de accesibilidad para personas con discapacidades. En manifestaciones más avanzadas podemos comprobar la falta de adaptación de la interfaz GUI para el control por voz, la interacción social mediante video, y el control remoto de aparatos domésticos por mencionar solo algunas de las funciones que en materia de comunicación se han dado en fechas recientes.



Ilustración 6: Un Teléfono “inteligente” en realidad es una computadora con muchas extensiones de los sentidos humanos, pero con poco espacio para una interfaz WIMP. Las pantallas son físicamente pequeñas ya que se espera que quepan en una mano y sean muy portátiles

La ubicuidad de computadoras conectadas en redes no necesariamente implica que en todos lados hayan pantallas en las cuales se puedan abrir ventanas para manipular íconos y menús. El ámbito lógico para las interfaces de voz es cla-

ramente el de los teléfonos "inteligentes", así como las interfaces de respuesta auditiva son las más lógicas en las computadoras que son el núcleo funcional de un aparato para oír música. La ubicuidad de las computadoras que no necesitan interfaz visual tiene como efecto por un lado de hacer evidentes las limitantes conceptuales de la interfaz WIMP y por otro lado dan impulso a la búsqueda de interfaces nuevas que correspondan a los nuevos ámbitos de interacción.

V) Ineficiente como sistema de socialización.

Otro de los aspectos en los que los GUI WIMP no se han podido mantener a nivel de la evolución de los sistemas de cómputo es al tratar de aplicarlos en las interacciones sociales que la computación móvil y las redes informáticas ubicuas hacen posibles. La emergencia de tecnologías como bluetooth y WiFi hablan de la posibilidad de mantener el flujo de datos en cualquier ubicación y el avance inexorable del poder computacional enunciado en la ley de Moore⁴ hacen que hasta los juguetes tengan hoy día computadoras cien veces más potentes que la famosa UNIVAC.

El fenómeno de la comunicación de persona a persona y de persona a grupo social se ha potenciado al ritmo de este progreso tecnológico. Se puede observar una expansión de la interacción social humana. Es sintomático de lo inadecuado de la interfaz de metáfora WIMP para estos nuevos ámbitos de interacción que la GUI sea lo primero que se oculta cuando la computadora es un mediador entre el ser humano y su grupo social, la GUI se trata más como un obstáculo que como una facilitadora de la interacción.

⁴Aunque es llamada "ley", se trata más bien de una observación que el ingeniero Gordon Moore, cofundador de Intel, hiciera en un artículo sobre miniaturización de componentes electrónicos en el año de 1965. En ella señalaba que la tendencia entre 1958 y 1965 de los circuitos impresos a contener cada vez más transistores continuaría, haciendo que los circuitos típicos contuvieran el doble de transistores cada año y que dicha tendencia continuaría por lo menos 10 años más. Esta observación fue revisada por su autor en varias ocasiones, siendo la versión más conocida (y usada) la de que el número de transistores en un chip procesador se duplica cada 18 meses. La exactitud a través del tiempo de esta predicción se debe, entre otras causas, a que la industria la ha adoptado para establecer sus metas de producción e investigación y desarrollo.

1.5 Puntos guía para el diseño de aplicaciones con interfaz gráfica.

Para concluir este análisis de las interfaces GUI, se propone a continuación de manera sintetizada un conjunto de estrategias y puntos de guía para el diseño de interfaces gráficas y que son comúnmente reconocidos como lineamientos para el diseño de interfaces efectivas en esta metáfora de interacción, lo que se podría titular como mejores prácticas para el diseño.

- a) El diseñador debe conocer el manual de estilo de la interfaz gráfica de sistema en la que su diseño estará inmerso. La aplicación de diseño de interfaz estándar implica mucho más que la consideración de orden de menús y tamaño de íconos, ya que también especifica ciertos modos de interacción que el usuario conoce y de cierta manera no pueden ser contradichos. Si el usuario de una Mac espera cierto comportamiento al mantener presionado el botón del mouse sobre un elemento (aparición de un menú contextual) que a su vez es el tipo de interacción que el usuario de Windows espera del evento de "hacer click con el botón derecho", es un error relativamente grave proponer una interacción distinta.

- b) La separación conceptual del diseño de una interfaz particular con respecto al diseño de una interfaz estándar debe corresponder a la separación conceptual entre la funcionalidad de la aplicación individual y la funcionalidad del sistema operativo. Esto quiere decir que se pueden proponer interfaces más complejas y heterodoxas, cuando la aplicación tiene funciones más complejas y heterodoxas. Para proponer una interfaz de un programa que requiere funciones como "guardar un archivo" o "imprimir" la propuesta de diseño tanto visual como en modo de interacción no debe derivar muy lejos de lo normativo en el sistema

operativo, cuando la aplicación necesita, por dar un ejemplo, que el usuario pueda ocupar metafóricamente el rol de un pirata del siglo XVII en busca de aventuras, es obligatorio que el diseñador provea de una interfaz que refuerce las funciones de fantasía y hasta poéticas de dichas necesidades.

- c) El diseño de interfaz gráfica WIMP tiene un fuerte componente ergonómico, esto es, su posibilidad funcional siempre será más importante que su posibilidad estética. Atenerse a las reglas de la interfaz estándar normalmente garantiza la ergonomía de la interfaz, en tanto a que ésta ya fue probada con anterioridad, y el usuario puede predecir los resultados de sus interacciones. En el caso de las interfaces heterodoxas se debe tener una verificación que puede tener como punto de partida los estudios de usabilidad de interfaz como los constantemente actualizados por Jakob Nielsen y que hace públicos en su sitio web⁵. De sus estudios se enuncian los siguientes diez puntos prácticos para verificar heurísticamente la usabilidad de una interfaz:[NIELSEN, MOLICH 1990]

1.- Visibilidad del estado del sistema

El sistema siempre debe mantener al usuario informado de lo que está pasando con él, a través de una retroalimentación apropiada dentro de un tiempo razonable.

2.- Correspondencia entre el sistema y el mundo real

El sistema debe hablar con un lenguaje natural para el usuario, con palabras, frases y conceptos que le sean familiares, más que con terminología orientada al sistema. Se deben seguir las convenciones del mundo real, haciendo la información aparecer en orden lógico y natural.

3.- El usuario tendrá control y libertad

Esto es especialmente necesario cuando el usuario con cierta frecuencia escoge funciones del sistema por error y debe tener una "salida de emergencia" claramente marcada para poder salir del estado indeseado sin tener que pasar por un dialogo extendido. En resumen, debe haber soporte de "deshacer" y "rehacer".

⁵<https://www.nngroup.com/articles/> desde febrero de 2018, anteriormente useit.com. Los diez puntos heurísticos pueden consultarse en <https://www.nngroup.com/articles/ten-usability-heuristics/> (revisado en agosto de 2018)

4.- Consistencia y Estándares

Los usuarios no deben dudar si diferentes palabras, situaciones o acciones llevan a la misma cosa. Se deben mantener las convenciones del sistema operativo.

5.- Prevenir errores

Mejor que tener mensajes de error bien diseñado es un diseño cuidadoso en el que se previene que se cometan errores desde el principio. Esto se consigue ya sea eliminando las situaciones en las que se tiende a cometer un error o estableciendo puntos de revisión en los que se solicita al usuario confirmación de una opción antes de ejecutar la acción.

6.- Reconocer es mejor que recordar

Se debe minimizar la carga en la memoria del usuario haciendo los objetos, acciones y opciones visibles. El usuario no tiene que recordar la información mientras pasa de una parte de un diálogo a otra. Las instrucciones de cómo usar el sistema deben ser visibles o fáciles de invocar cada vez que sea apropiado.

7.- Flexibilidad y eficiencia de uso

Un usuario experto debe contar con aceleradores, que no serán visibles para el usuario novato. Un sistema debe proveer la forma de uso tanto de usuarios experimentados como usuarios novatos. Esto generalmente implica que cada usuario pueda ajustar su accionar a su gusto.

8.- Diseño minimalista y estético

No agregar información a los diálogos si es irrelevante o raramente necesaria. Cada unidad extra de información en un diálogo compite contra las unidades relevantes de información y disminuye su visibilidad relativa.

9.- Ayudar a los usuarios a reconocer, diagnosticar y recuperarse de los errores

Los mensajes de error deben mostrarse en lenguaje sencillo (no en códigos de error), indicar con precisión cual es el problema y sugerir una solución constructivamente.

10.- Ayuda y documentación

Aunque siempre es mejor un sistema si puede usarse sin documentación, puede ser necesario proveer ayuda y documentación. Debe ser fácil buscar dentro de la información de este tipo, incluso con el enfoque de cuál es la tarea que realiza en ese momento el usuario, y proveer de una lista no muy larga de pasos concretos a realizar.

- d) El diseño gráfico de interfaz sigue siendo diseño gráfico, esto es, no se cancelan las reglas de composición, balance, contraste etc. que caracterizan un buen diseño gráfico.

- e) El acto diseñístico es un acto de interpretación, y en este caso el diseñador de interfaces interpreta los aspectos funcionales del programa que subyace a la interfaz que está diseñando; interpreta los mensajes que la interacción entre usuario y máquina deben transmitir y por último utiliza su interfaz como vehículo de interpretación de la experiencia que el usuario tendrá con el programa.

Capítulo 2.- Aspectos de la interfaz gestual

2.1 Características de la interfaz gestual.

La principal diferencia de las interfaces con otros fenómenos de comunicación es que uno de los receptores/emisores no es un ser humano. Este fenómeno es por lo tanto descrito también como interacción Hombre-máquina o, refiriéndose específicamente a las computadoras (por su prevalencia desde hace algunas décadas), interacción humano-a-computadora, abreviado CHI (por sus siglas en inglés: *Computer to Human Interaction*).

Para la CHI se usan lenguajes de comunicación que no necesariamente son visuales, pero ciertamente el lenguaje visual se ha vuelto el más común y por lo tanto el más asociado con la CHI cuando se habla del diseño de interfaz para *software*. Como vimos en el capítulo anterior, existen ciertas justificaciones de orden práctico y ergonómico que han mantenido a las interfaces gráficas o GUI como la interfaz predominante en la comunicación hombre-a-computadora.

Es erróneo pensar que tal dominio es debido a lo perfecta que resulta la comunicación a través de estas interfaces. Un par de ejemplos generales que ponen en evidencia esta imperfecta comunicación: de manera continua aparecen métodos de personalización de la GUI estándar; y los diseñadores tratan de pro-

poner elementos no estándares todo el tiempo. Estos ejemplos nos indican una insatisfacción en las necesidades de comunicación que cubre la GUI, o al menos que no hay una correspondencia completa entre las funciones que se han establecido como estándar y las necesidades de los usuarios.

En la tabla que se presenta a continuación Jakob Nielsen resume los avances de las computadoras ligándolos a los avances de: tecnología de *hardware*, modo de operación, lenguajes de programación, tecnología de terminal (que hoy día tiene más sentido llamar tecnología de despliegue de la información), tipo de usuario, paradigma de imagen en la publicidad, y paradigma de interfaz de usuario [NIELSEN 1993].

Generación	Tecnología de Hardware	Modo de operación	Lenguajes de programación	Tecnología de terminal	Tipo de usuarios	Imagen publicitaria	Paradigma de interfaz de usuario
0 -1945 Prehistórica	Mecánica, electromecánica (Babbage, Zuse Z3)	Realmente no fue "usada" excepto para hacer cálculos.	Enchufar y desenchufar cables	Lectura de luces parpadeantes y tarjetas perforadas	Los inventores de las máquinas eran los mismos usuarios	Ninguna	Ninguno (El poder meter las manos directamente en el hardware era lo único que importaba)
1 1945-1955 Pionera	Bulbos, Máquinas enormes, necesidad de mucho enfriamiento, corto tiempo de operación entre fallas.	Un usuario a la vez se vuelve "dueño" de la máquina (pero sólo por un tiempo limitado)	Lenguaje de máquina 001100111101	TTY, teletipo electromecánico. Sólo usado en el centro de computación	Expertos, pioneros	La computadora es una calculadora	Programar
2 1955-1965 Histórica	Transistores; computadoras más estables se empiezan a usar fuera de los laboratorios.	Lotes ("La computadora como un templo" al cual se le hacen ofrendas y se recibe a cambio respuestas como de un oráculo)	Ensamblador ADD A,B	Terminales orientadas a línea de comandos ("TTY de vidrio")	Tecnócratas, profesionales comunicacionales	La computadora como un procesador de información	Lenguajes de línea de comandos
3 1965-1980 Tradicional	Circuitos integrados. Los negocios pueden justificar el costo porque ya se ve como una necesidad.	Tiempo compartido (sistemas de proceso de transacciones en línea)	Lenguajes de "Alto nivel". Fortran, Pascal	Terminales de pantalla completa, que despliegan sólo caracteres alfanuméricos. Es común el acceso remoto.	Grupos especializados sin conocimientos computacionales (por ejemplo cajeros bancarios)	Mecanización de labores administrativas	Menús de pantalla completa y estrictamente jerarquizados. Llenado de formularios.

Generación	Tecnología de Hardware	Modo de operación	Lenguajes de programación	Tecnología de terminal	Tipo de usuarios	Imagen publicitaria	Paradigma de interfaz de usuario
4 1980-1995 Moderna	VLSI. ⁶ Los individuos pueden comprar sistemas personales	Computadoras personales para un sólo usuario	Lenguajes orientados a problemas, hojas de cálculo	Pantallas gráficas con resolución adecuada. Estaciones de trabajo compactas y portátiles pesadas.	Profesionistas, hombres de negocios, aficionados	Productividad personal (la computadora como herramienta)	WIMP (Windows, Icons, Menus, and a Pointing device)
5 1996-? Futura	Integración a escala de microchip, la-computadora-en-un-chip. Los individuos pueden comprar muchas computadoras.	Sistemas de usuario único conectados en red, sistemas embebidos	No-imperativos, posiblemente gráficos.	Con I/O multimedia, fácilmente portátiles, conectados inalámbricamente.	Cualquier persona	La computadora como entretenimiento	Interfaces sin comandos

Tabla 1 Las generaciones son aproximadas, y no todas las dimensiones se desarrollaron y cambiaron sincrónicamente. La “imagen publicitaria” fue tomada de Aspray, W., and Beaver, D. *Marketing the monster: Advertising computer technology*. Annals of the History of Computing 8, 2 (Abril 1986) ACM Press, New York, E.U.A. pags. 127-143.

Esta tabla fue publicada en 1993, por lo que la predicción de Nielsen sobre la fecha de partida de la 5a generación y el tiempo que le asigna a cada generación es un tanto arbitraria, pero lo suficientemente acertada en cuanto a que las características que describe para la 5a generación de aparatos de cómputo es bastante efectiva para describir el estado actual (2018) de las computadoras. De hecho le ha tomado a las interfaces sin comando unos diez años más en materializarse con respecto al punto de partida del año de 1996, pero Nielsen estableció las fechas para darle un espacio temporal redondeado de 10 años a cada “generación” y de 15 años a las generaciones “tradicional” y “moderna”. Nielsen no ha vuelto a revisar su tabla de generaciones o a proclamar una 6a generación, pero su concepto de interfaz sin comando resulta muy útil para el análisis de las interfaces gestuales, además de qué con esta tabla se comprende muy bien el antecedente y la implantación en forma gradual de las tecnologías de interfaz en paralelo al avance tecnológico y a los cambios en el mercado y tipo de usuario.

⁶Very Large Scale Integration o integración a muy grande escala. Se refiere a la integración de mayor número de transistores en diseños de un sólo circuito integrado, permitiendo por lo tanto mayor cantidad de operaciones y funciones. Los microprocesadores son aparatos VLSI

Nielsen llama interfaz sin comando aquella en la que los diálogos precisos para especificar un comando desaparecen [NIELSEN 1993]. No desaparece el diálogo en sí, sino que el diálogo toma forma más concretas de interacción, donde incluso la computadora adivina hasta cierto punto las acciones. En vez de haber una certeza unívoca de qué es lo que el programa tiene que ejecutar (un comando) lo que hay es una alta probabilidad de ejecutar, en muchos casos aprendida (una intuición).

El ejemplo que Nielsen utiliza para entender estas interfaces es el de una interfaz basada en gestos, inmersa en una realidad virtual. Un juego de deportes en el cual el jugador realiza los movimientos del personaje que controla no mediante clicks y teclas, sino haciendo los gestos de los movimientos reales de un jugador en esa situación, o gestos equivalentes. Nielsen plantea este ejemplo en 1993, y la interfaz sin comando se implementó casi al pie de la letra de este ejemplo en algunos sistemas experimentales y otros comerciales (para galerías de videojuegos “*Arcade*”, principalmente), alcanzando un éxito masivo con el juego *Wii Sports* de la consola Nintendo Wii a partir de 2006.

Una interfaz gestual es, pues, primordialmente, una interfaz en la que los diálogos de hombre-a-computadora se realizan como no-comandos. Para determinar cuál es el curso de acción más probable, el programa controlado por una interfaz sin comandos necesita de un sistema de interpretación. Varios sistemas de este tipo se han realizado a través del tiempo, dedicados de alguna manera u otra a colaborar con interfaces WIMP. Los precursores de las interfaces sin comando son sistemas como el reconocimiento de voz o el reconocimiento de “escritura natural” y que generalmente se plantean como complementarios de la interfaz WIMP.

Las posibilidades que los aparatos de toque (*touchpad*, *touchscreens* y similares), las tabletas digitalizadoras, los aparatos de retroalimentación háptica⁷ y

⁷Los aparatos hápticos son aquellos que reciben y transmiten al usuarios sensaciones de posición, tamaño, resistencia y masa. Uno de los problemas de manipular objetos virtuales es que el manipulador debe recibir retroalimentación. Si en el mundo real yo levanto un objeto con la mano, puedo sentir sus límites físicos, su peso, y hasta cierto punto la resistencia que me implica realizar un mayor o menor esfuerzo para moverlo. Los aparatos hápticos simulan estas condiciones.

finalmente las cámaras web pareadas con *software* de reconocimiento de imágenes sugieren que la interfaz gestual es la que puede acoplarse más rápidamente a las computadoras que necesitan interfaces sin comandos. Uno de los impactos de la mayor capacidad de cómputo al paso del tiempo es de poder analizar en tiempo real los datos de los sensores del ambiente y realizar predicciones de probabilidad de los comandos que el usuario va a realizar. El sistema probabilístico comienza a usarse en el reconocimiento de video [WACHS et al, 2011] y la combinación de técnicas para interpretar datos del ambiente y convertirlos en un curso de acción probable recibe el nombre genérico de técnicas heurísticas⁸

La siguiente es una definición general de interfaz gestual que sintetiza los conceptos expuestos: es una interfaz sin comandos, adecuada específicamente a aparatos de cómputo personales, integrados y que incluyen un sistema para reconocer los gestos⁹ y determinar heurísticamente las acciones a seguir.

En los equipos de desarrollo *software*, necesariamente interdisciplinarios, se dedica al menos a una persona de manera específica al diseño de interfaz y al análisis de interacción, esta persona es la que toma las decisiones sobre los elementos que conforman la interfaz gestual como sistema de control, partiendo del conocimiento especializado de los procesos de comunicación entre usuarios y computadoras, y las características de *hardware* y *software* que diseña el resto del equipo; este diseñador cuando está diseñando interfaces gestuales puede reconocer las unidades semánticas del lenguaje de gestos y aplicarlos en la construcción de mensajes, y puede enunciar las metas de comunicación del sistema; para este diseñador poder prefigurar el tipo de diseño a conseguir es muy útil pues le permite poder dirigir sus esfuerzos de manera más efectiva.

⁸Del griego *εὕρισκειν* que significa «hallar, inventar» y que se refiere al predecir un resultado por el conocimiento previo

⁹La gestualidad puede comprender una amplia gama de actitudes, estáticas y dinámicas, de todo el cuerpo o solo parte de él, o incluso de elementos que no son parte del cuerpo humano, como el *mouse* u otros aparatos.

2.2 El gesto como signo.

El estudio de los signos dentro de la teoría de la comunicación, comúnmente llamado semiótica [ECO 1975], ha tenido corrientes que en diversos tiempos han intentado clasificaciones como parte de su metodología en el estudio del signo y otros fenómenos de comunicación asociados al signo. Los modelos más aceptados hablan de cómo el significado incluye percepciones externas al signo mismo, por ejemplo las nociones de Barthes sobre la moda [BARTHES 1967] y la influencia del contexto temporal en los significados .

El gesto, que se define como las actitudes corporales expresivas, voluntarias o involuntarias, que una persona realiza como parte de un discurso [HALL 1994], puede ser (y en algunos casos ha sido) estudiado como signo en la comunicación. La mayor parte de las teorías semióticas ven al gesto como subordinado a los significados expresados en un discurso principal, es decir, como un subtexto, un anclaje de un significado expresado por otra vía [BARTHES 2002].

El gesto, como elemento de comunicación, como un signo por sí mismo, fue estudiado principalmente por un grupo de psicólogos, formados en la corriente *conductista* (también llamada “*behaviorista*”) que tuvo su mayor auge en las escuelas de psicología estadounidenses en las primeras décadas del siglo XX. Son notorios los experimentos sobre la comunicación gestual (o corporal) del Dr. Birdwhistell y la teoría que explica los gestos como elementos lingüísticos, desarrollada por Edward T. Hall y a la que dio el nombre de *Proxémica*. [HALL 1994]

Generalizando, la Proxémica trata de entender los gestos como un código y analizarlo dentro de un discurso exclusivamente lingüístico, siempre en contexto de la palabra hablada. Birdwhistell, quien es uno de los principales investigadores de la proxémica, afirma que el lenguaje de los gestos es una representación de procesos mentales que el sujeto tiene pero que no llegan a expresarse en palabras . [BIRDWHISTELL 1994]

Movimientos de cabeza como marcadores



Voy a ir al centro luego iré a casa de Bill . . .



. . . Luego iré a casa ¿Qué vas a hacer?

Los párpados como marcadores



. . . Luego iré a casa ¿Qué vas a hacer?

Movimientos de mano como marcadores



. . . Luego iré a casa ¿Qué vas a hacer?

Ilustración 7: Ejemplo de identificación de gestos marcadores en los estudios de proxémica de Birdwhistell

La proxémica tiene como idea central que los gestos pueden estudiarse como un sistema simbólico, estudiando y describiendo este sistema simbólico de manera dependiente de la lingüística, sin dejar de considerarlo parte de un fenómeno de comunicación verbal. Los estudios y experimentos proxémicos demostraron que un interlocutor percibe cambios muy finos en la posición corporal, en los movimientos de las extremidades, en los movimientos musculares involuntarios, en la dirección de las miradas, etc. y que los discrimina con un significado unívoco y como modificadores del discurso hablado [BIRDWHISTELL 1994].

En semiótica, partiendo del modelo de Peirce podemos distinguir tres tipos de signo: Índice, ícono y símbolo. Aunque existen otros modelos y hay inves-

tigaciones en semiótica que amplían las categorías de Peirce, este modelo en tres partes es suficiente para revisarlo desde el análisis proxémico.

El índice es el signo que identifica el objeto referido por la conexión con el objeto mismo. Decimos que el humo nos indica que hay fuego, por lo que el humo es un signo índice del fuego. En los gestos, un gesto que funcione como signo índice es, prototípicamente, el gesto de señalar con el dedo. Por ejemplo, al estar en la mesa y decirle a otra persona “¿me pasas el salero? Y acompañarlo con un gesto apuntando con un dedo al objeto que es el salero, creando en el interlocutor una relación entre el dedo que apunta y el salero, una línea imaginaria que completa el significado de indicar la acción. El gesto es pues, un gesto signo índice, hace de índice del salero, identifica el salero entre otros objetos a los que potencialmente podría referirse, su equivalente gráfico podría ser el de una flecha que dirige la atención del sujeto en una dirección específica o a una acción concreta, en lenguaje podría ser similar a la expresión “el salero está aquí” siendo la palabra “aquí” el signo de la ubicación espacial, causalmente conectado a la ubicación actual del que habla.

El ícono es el signo que identifica el objeto por el parecido. El retrato de una persona es un ícono, no es la persona real, tiene suficiente parecido a la persona para permitirle al observador identificarla. En el caso del lenguaje oral, existen expresiones icónicas, por ejemplo las onomatopeyas, donde la palabra se parece al sonido del concepto que representa. Con frecuencia los gestos son de tipo icónico. Para representar “enorme” una persona puede hacer un gesto abriendo los brazos en la actitud de tratar de abarcar algo que es muy grande. Incluso muchas veces se usa la distancia entre manos para establecer el tamaño real de lo representado. “un hilo de este tamaño” “la pared tiene este espesor”, el espacio vacío entre las extremidades tiene una relación de parecido al tamaño del objeto al que se alude, de ahí la clasificación de este gesto como icónico.

Finalmente, completa esta clasificación el símbolo, que contienen un significado sin tener que indicar al objeto representado o sin mantener un parecido de

identidad. Exactamente por qué una forma visual o una unidad lingüística se cargan de significado hasta conseguir ser símbolos es el campo de estudio de la semiótica y hacer una exposición detallada de cómo adquieren significado los símbolos está fuera del alcance de esta tesis. Un ejemplo de símbolo es el de una bandera. El conjunto de formas y colores que son la bandera no están funcionando como índice ni tienen un parecido que pueda decirse icónico. La carga de significados, que se ha llamado el campo semántico, de la bandera puede ser conocida a quien la diseñó, que pudo haberse basado en campos semánticos para los que existe una taxonomía bien establecida y aceptada, como por ejemplo escoger los colores de acuerdo a los significados establecidos por la heráldica, y en otros casos por la asignación hermenéutica del símbolo utilizado, por ejemplo el águila devorando una serpiente sobre un nopal como alusión al mito fundacional del imperio azteca, en la bandera de México, o la hoz y el martillo en las banderas de países comunistas, en otro ejemplo. Un observador que no ha sido educado en el significado de estos campos semánticos reconoce el valor simbólico de la bandera, el país que representa, sin menoscabo de lo poco o mucho que reconoce de los otros campos semánticos.

En cuanto a gestos, lo que interesa es el fenómeno ya observado por Birdwhistell y los otros investigadores de la proxémica, donde ciertos gestos alcanzan un carácter simbólico al conseguir un nivel de reconocimiento que no depende de la cercanía con el objeto como en el caso del gesto índice o del parecido como en el caso del gesto icónico. El interés de los proxémicos era el de establecer una identidad entre procesos de pensamiento y gestos, que pudieran identificarse de manera común a grupos amplios de personas. En el caso de la comunicación hombre máquina, el potencial de establecer gestos como símbolo reside en la posibilidad de diseñar interfaces que puedan ser usadas con más efectividad eliminando la barrera del lenguaje, sea este escrito o visual.

En el caso de el signo simbólico, algunos investigadores como Herbert Clark [CLARK 1999] han identificado que los tres tipos de Peirce pueden coexistir en un

solo elemento de significación, un solo signo puede tener grados de índice, iconicidad y simbolismo. Este modelo mixto es bastante útil para la propuesta de taxonomía que es objeto de esta tesis, ya que las interfaces pueden usarse para comunicar información muy heterogénea, y los gestos que se usen como base de la interactividad tienen que adaptarse a esta cualidad. Así como para Herbert Clark los campos semánticos mixtos son “dinámicos”, implicando que pueden tener variaciones entre índice, ícono y símbolo según el uso o jerarquía en el discurso, la relación del gesto con las partes no gestuales de la interfaz se beneficia de tener una cualidad dinámica donde su campo semántico puede ampliarse o reducirse, variar en el tiempo y adaptarse por el uso y la individualidad del usuario.

El gesto como símbolo lo podemos ejemplificar en el uso de ciertos íconos en la interfaz gráfica. Los gráficos de la siguiente ilustración son reconocibles como símbolos de “internet”.



Ilustración 8: Cualquiera de estos tres gráficos está asociado como símbolo de “internet”, siendo que en su origen son símbolos de programas o compañías que hacen programas para usar internet. Izq a der: Logotipo de Mozilla Firefox (ver57), logotipo de Microsoft Edge (ver44), logotipo de Google Chrome (ver71)

Si vemos uno de estos símbolos en el aparador de un local, en una pantalla de un teléfono inteligente o en la barra de aplicaciones de la pantalla de un sistema operativo, identificamos que hay conexión a los servicios de red, sea porque se trata de un local de cabinas de computadoras con renta por tiempo, o la activación de un programa de navegación por red celular o red ethernet.

Trasladando esto al valor simbólico de los gestos, ya fue observado por Clark que el gesto comúnmente combina los valores icónicos e indiciantes, y que muchas veces el valor simbólico no está completamente desligado de la representación icónica. Un ejemplo prototípico es el del gesto que se muestra a continuación:



Ilustración 9: Gesto "pistola". Foto de Gonzalo León (2018)

Se puede reconocer, y por tratarse de un parecido es por tanto un significado icónico, el gesto como una pistola, pero hay muchas diferencias donde la mano y la pistola no tienen ningún parecido en absoluto, que van desde la forma misma de una pistola, color, textura, etc, hasta los posibles significados simbólicos, que van desde los casi icónicos como “esta persona es tan antipática que quisiera darle un balazo” a los más simbólicos como “yo disparo” que en México es una forma coloquial de decir “yo pago” o en inglés “*you call the shots*” “tu eres el que ordena cuando disparar”, que es una forma coloquial de decir “tu eres el jefe”. Esta forma resumida en la que se construye el gesto Clark la llama representación selectiva, ya que solo algunas partes del objeto son representadas.

Otra subcategoría que se desprende del estudio del gesto simbólico es la de emblema, que se refiere a un gesto que trasciende en significación los equivalentes lingüísticos o culturales. La investigadora de psicología y aprendizaje del lenguaje, Susan Goldin-Meadow ha diferenciado el emblema o gesto emblemático por contener un significado autónomo y único, incluso clasificando un gesto como emblema por el hecho de tener asociado un referente o etiqueta verbal llamado glosa. Por ejemplo, al referirse al “señal de amor y paz”, “gesto de OK”, “saludo de Hitler” la persona se imagina claramente cual es el gesto referido con esas glosas, y

aunque puede haber bastante laxitud en la manera de realizarlos, su existencia como gesto emblema es indiscutible y su significado invariable. A diferencia del gesto icónico que puede variar de significado al variar el discurso, por ejemplo el gesto que puede usarse para significar una pelota de basquetbol en un un contexto puede usarse para significar pelota de futbol o de golf en otro contexto, el gesto emblema no cambia de significado, y a diferencia del gesto simbólico que puede construirse con elementos del gesto icónico y del gesto índice, el gesto emblema no tiene una dependencia de parecido a un objeto o de orientar la atención en una dirección específica. [GOLDIN 2005]

Los emblemas son gestos simbólicos ya que basan su campo semántico en significados convencionales, esto es, por aceptación cultural. El campo semántico de un gesto simbólico no alcanza el nivel de signo universal, dado que siempre habrá diferencias culturales y temporales en los grupos humanos que lo usan. Algunos de estos gestos han conseguido un estado intercultural, un reconocimiento más amplio, gracia a que se difunden por los medios de comunicación globales. Una persona que no tiene referencias sobre los movimientos culturales que originaron el emblema de “paz y amor” probablemente vea ese gesto como símbolo icónico de orejas de un animal, o de unas tijeras.

Esta categorización del gesto basada en las tres categorías del signo de Peirce es útil en la comprensión de muchos tipos de discurso que incluyen gestos, pero no necesariamente es el único tipo de categoría que existe del gesto como signo. Especialmente cuando se considera como parte del discurso, cuando se construye un lenguaje a partir del gesto o incluyendo el gesto como elemento lingüístico, la categorización solo del gesto como signo es bastante incompleta en cuanto a la generación de significado mediante gestos y la comprensión del gesto como medio de comunicación.

2.3 El gesto como elemento lingüístico.

Los estudios de la proxémica analizan y organizan como código complementario al discurso verbal las posiciones en el espacio de objetos y personas, así como los gestos que acompañaban el lenguaje. Estas nociones fueron utilizadas también en estudios de ergonomía, y en aplicaciones de diseño industrial y gráfico, como refuerzos del discurso visual. Por ejemplo, el conocimiento de cuál es la distancia que las personas de una cultura reconocen como “privada” (es decir, la distancia que deben guardar dos personas para que ninguna de las dos se sienta invadida en su privacidad) ha sido utilizado para el diseño de espacios arquitectónicos. Otro ejemplo: el reconocimiento instintivo de ciertos gestos como gestos de aprobación o desaprobación se ha usado para la generación de gráficos de señalización, como la síntesis morfológica que representa un dedo cruzándose sobre unos labios para indicar que se debe guardar silencio.

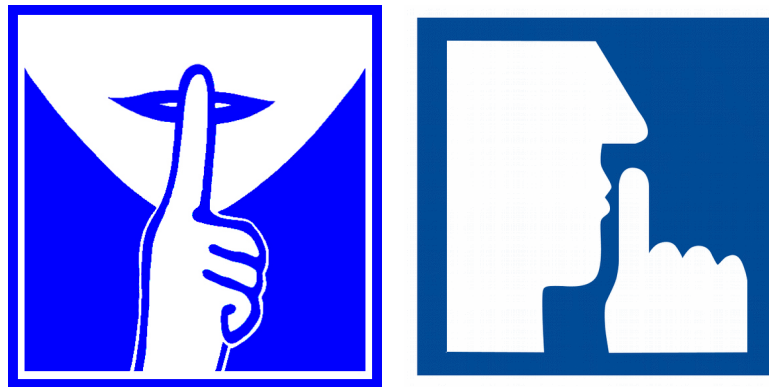


Ilustración 10: Arriba, izq: Cartel de silencio (G. León, 2017). Arriba, der: Señal de silencio, norma ISO 7010:2007 Abajo: secuencia de la frase "se guarda silencio" del Lenguaje de señas americano (ASL, por sus siglas en inglés). Sólo el dedo sobre los labios, en el ASL, se usa para ordenarle a alguien que esté sin hacer ruido. La secuencia completa es una frase descriptiva o imperativa según el contexto.

Un caso aparte es la utilización de gestos como un lenguaje por si mismo, como los lenguajes de sordos, los alfabetos marinos de banderas, o los gestos estandarizados con los que el personal de pista dirige el rodar de un avión en el aeropuerto. Estos lenguajes pueden ser estudiados por la lingüística de la misma manera que los lenguajes hablados, con significaciones y convenciones sintácticas y ortográficas equivalentes a las del lenguaje hablado o escrito [DAVIS 1999] .

Un enfoque muy interesante sobre la función del gesto como lenguaje proviene de la neurosemiótica, desde donde se produce la clasificación propuesta por Nespoulous y Lecours [NESPOULOUS et al, 1986] basada en la distancia (abstracción) del movimiento gestual y la interpretación deseada a dicho movimiento. Se puede representar la taxonomía del gesto como la figura siguiente:

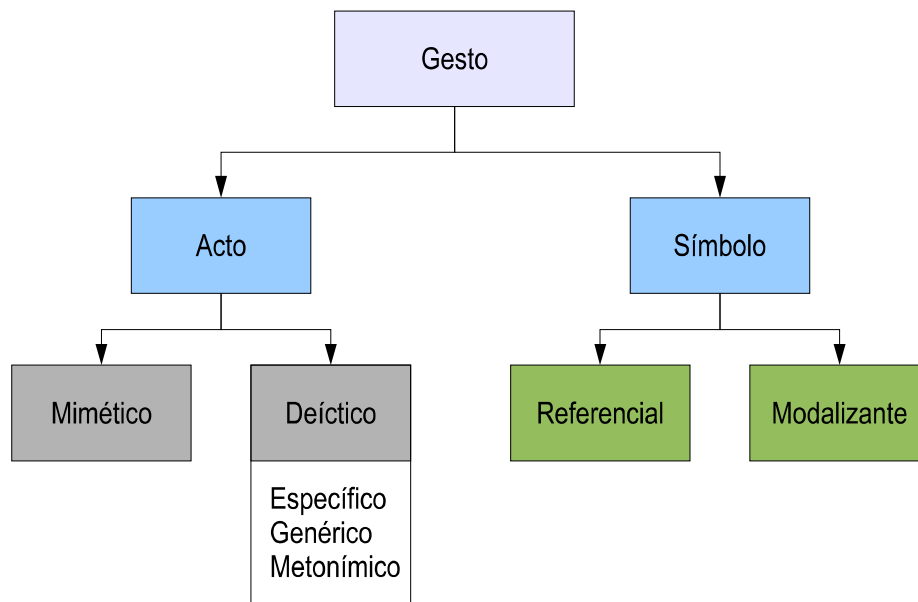


Ilustración 11: Taxonomía del gesto de Nespoulous y Lecours. Fuente:[QUEK 1994]

En los gestos como acto, el movimiento se relaciona específicamente con la interpretación esperada. Los gestos como símbolo son primordialmente arbitrarios, por naturaleza y la diferenciación se centra en entender el gesto como un conjunto de movimientos (gesto como acto) y como una unidad lingüística (gesto como símbolo).

El gesto como símbolo se entiende como referencial al discurso y como modalizante del discurso. El gesto es referencial cuando se usa para aludir a un objeto o un concepto, es decir, para hacer presente al *referente*. Por ejemplo, tallar el pulgar sobre el dedo índice y medio, en círculos, para hacer referencia al dinero.

El gesto es modalizante cuando acompaña a un mensaje, generalmente transmitido por otro medio de comunicación, y comunica la opinión del emisor. Por ejemplo, si una persona le dice a otra “¿ya viste a fulano?” y al mismo tiempo hace un gesto con los brazos para indicar obesidad, posiblemente provocando una risa en el receptor. Si posteriormente alguien escuchara una grabación del discurso hablado, la risa provocada por el comentario sería difícil de entender. Se dice que el gesto ha cambiado de modo la afirmación, de un modo neutro o declarativo a un modo irónico o exagerado. Esta cualidad modalizante puede hacer que un mismo gesto cambie de significado dependiendo del contexto en que se genera. Por ejemplo, levantar una mano y hacer girar el dedo índice señalando hacia arriba puede significar “Dame más ejemplos” cuando se le dirige a alguien que está exponiendo una explicación, o “acerca más el vehículo” si lo está usando un apoyo vial dirigiéndole a un conductor.

El gesto como acto se puede clasificar en *gesto de significado mimético* y en *gesto de significado deíctico*. Decimos que un gesto es mimético cuando el gesto tiene una alta identidad icónica con el acto representado; es decir que el gesto es una pantomima del acto en si. Por ejemplo, el gesto de un fumador que pide que le enciendan el cigarrillo haciendo el movimiento que usa para activar un encendedor. El gesto mimético depende de la familiaridad del receptor con el acto aludido para reconocer su significado. Si el receptor está poco familiarizado con el gesto, es posible que el emisor del gesto tenga que recurrir a la exageración hasta conseguir el grado de reconocimiento necesario.

El gesto deíctico es el que funciona como indicador o apuntador de un objeto o concepto. (Deixis = apuntar, indicar). Podemos clasificar más finamente los gestos deícticos si tomamos en cuenta el grado de separación del objeto al que se

apunta con respecto al contexto. Así, hablamos de gestos *específicos* cuando el gesto selecciona particularmente un objeto o lugar. Por ejemplo, hacer un gesto que selecciona un ícono y así activar un programa. El gesto *genérico* es el que se usa para elegir a un individuo para de ahí deducir como significado su categoría o clase superior. Finalmente, el gesto *metonímico*; siendo la metonimia la figura retórica en las que se hace referencia a las causas para hablar de los efectos y viceversa, un gesto metonímico es aquel en el que, por ejemplo, el usuario apunta a una grúa para ordenarle a un robot que haga la acción de levantar o cargar. Los tres tipos de gesto deíctico funcionan de manera similar a nivel mecánico, pero se diferencian en el mayor o menor nivel de interpretación de su significado.

2.4 El gesto y su inclusión en el diseño de la interfaz gráfica.

En las interfaces gráficas hay un factor de gestualidad que se integra en mayor o menor medida en los diferentes sistemas operativos según la decisión consciente o inconsciente del diseñador de la interfaz. Ya desde la aparición del mouse y el cursor como medios de apuntar en pantalla se conocía que podría haber otros métodos y aparatos de apuntador. Por ejemplo, el lápiz de luz que se mencionó entre los precursores de las GUI en el sistema SketchPad de Ivan Sutherland. Este sistema tenía un fuerte componente de gestualidad, en tanto que el lápiz se sostenía directamente sobre la pantalla para crear líneas y formas geométricas [SUTHERLAND 1965].

Sin embargo, el análisis de los movimientos para hacer entrada como un lenguaje gestual tardó algunos años en aparecer. El *software* de dibujo prefigura la interfaz de gestos para control de los elementos del *software*, los intentos de incluir dicha gestualidad en otros aspectos de la GUI son hasta cierto punto rudimentarios y en todo caso lentos. Es equiparable, por ejemplo, a la falta de avances en la interfaz de control de voz. [LISTER 2003].

Nielsen, hace un análisis de las razones por las que la GUI se estancó en su progreso [NIELSEN 2009] y analiza porqué las interfaces que tendría que ser su evolución no la sustituyeron, siendo el evento que alargó el dominio de la interfaz gráfica el advenimiento de las páginas web como principal método de despliegue de información en la pantalla de las computadoras; en dicho análisis Nielsen menciona que consideraba los elementos de la metáfora WIMP como obsoletos en 1994, cuando sólo habían pasado 10 años desde su popularización (1984 a 1994). y que le sorprendía estar haciendo una revisión y actualización de ese análisis cuando la metáfora WIMP estaba por cumplir 25 años de existencia y llega a la conclusión de que pese a la percepción de que el mundo de la tecnología computacional avanza rápidamente, los cambios de paradigma de interacción no lo hacen tan rápido. Después de todo, el mouse, inventado en 1963, fue presentado en un prototipo funcional en 1968 y le llevó 16 años más para aparecer en un sistema de aceptación comercial y popular (en la MacIntosh, en 1984), y aún así la interfaz GUI pasó por un periodo de implantación y popularización que duro algunos años. Esta paradoja puede extenderse a los intentos de evolucionar la GUI con la inserción de nuevos paradigmas de interfaz.

Por ejemplo, la tableta digitalizadora ha estado disponible desde la aparición de las computadoras (y aún antes, puesto que la primera tableta que podía capturar y transmitir caracteres, el Teleautograph, fue inventada por Elisha Gray en 1888) y ha tenido desarrollos notables sobre todo en los sistemas especializados de dibujo técnico (CAD, o dibujo asistido por computadora) y de dibujo artístico (inicialmente conocidos como *Computer Paintbox*, posteriormente abreviado simplemente a *Paint*).

La característica “relación posicional entre cursor y *stylus*” de la tableta digitalizadora sugirió el uso de gestos en los sistemas de CAD desde la mitad de los años ochentas. Las compañías fabricantes incluían típicamente algún tipo de *software* que permitía establecer áreas de la tableta con finalidades especí-

ficas, o "áreas de macros", ya que se podían asociar a algún pequeño *script*¹⁰ de computadoras. No tardó en aparecer algún emprendedor que escribiera *scripts* en los que gestos simples podían ejecutar conjuntos de comandos, y el gesto simple casi siempre tenía algún tipo de asociación semántica con la operación a realizar. Por ejemplo, una diagonal hacia abajo podía implicar el minimizar una ventana, un gesto de espiral en la dirección opuesta a las manecillas del reloj, deshacer.

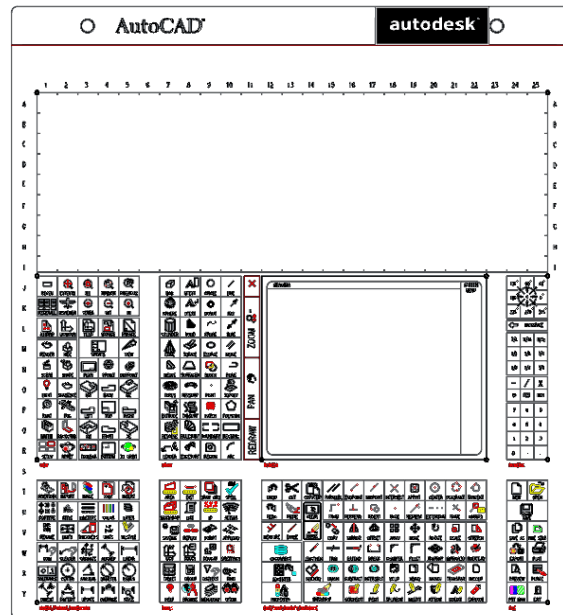


Ilustración 12: Plantilla de tableta digitalizadora de AutoCAD. En el área en blanco cerca del centro se tenía el área de dibujo, el rectángulo superior era el sitio donde algunos programadores empezaron a experimentar con el reconocimiento de gestos.

Este nuevo paradigma podía provocar entusiasmo en el reducido grupo de usuarios de CAD o *paint*, pero no se alcanzaba la masa crítica, la demanda de mercado que hiciera mella en el uso de GUI y que hiciera evolucionar la interfaz hasta un nuevo paradigma.

Otra idea que se desprende del uso de tabletas digitalizadoras es la posibilidad de realizar reconocimiento de texto manuscrito. El campo del reconocimiento óptico de caracteres impresos se ha desarrollado casi simultáneamente

¹⁰ Un *script* es un programa que no se compila para ejecutarse, sino que depende de un compilador integrado en el programa principal que lee el documento donde se aloja el *script*. La traducción literal es "guion" como en el guion o libreto de una obra dramática, pero en español se usa sin traducir en su contexto de informática.

con las computadoras, generando varios métodos convencionales para que las computadoras puedan convertir imágenes de letras en texto de computadora. Reconocer texto manuscrito tenía muchas aplicaciones interesantes, como por ejemplo establecer la autenticidad de firmas en documentos bancarios. Si se agrega la idea de una tableta digitalizadora, es natural que se piense en reconocer el acto de la escritura al tiempo que sucede, esto es, en reconocer los movimientos que hacen cada letra al ser dibujada. Movimientos que son, desde luego, gestos. Es una idea que surge en paralelo a la del control mediante comandos de voz, o a la de reconocimiento de voz para hacer una conversión en tiempo real a texto de computadora.

Otro aparato que pugna por llevar los gestos a la GUI es muy parecido a la tableta digitalizadora. Se trata de las superficies de toque, ya sea sobre la pantalla (*touchscreen*) o sobre áreas planas horizontales sin transparencia (*touchpad*), y en las que se puede interactuar usando estiletes (*stylus*), similar a las tabletas digitalizadoras, o directamente con los dedos o uñas.

Un sistema relativamente exitoso fue el comercializado como "*graffiti input system*". Su éxito se da cuando un nuevo tipo de aparatos, los Personal Digital Assistants, (PDAs) se popularizan. Graffiti estaba especialmente orientado a sustituir el uso de un teclado, proveyendo de un área de "ingreso" de caracteres en un área de touchpad o touchscreen, y acelerando el reconocimiento de caracteres mediante una convención de gestos, una codificación que no era propiamente reconocimiento de manuscrito sino reconocimiento de gestos. Para tener reconocimiento de letra manuscrita se requiere de un sistema que pueda aprender las diferencias en el trazo de letras de diferentes usuarios, incluso de diferentes alfabetos. Uno de los primeros PDAs, comercializado con la marca Newton por la compañía Apple, tenía un sistema de reconocimiento de manuscrito que no funcionaba muy bien debido a que estaba muy orientado a reconocer palabras en inglés y sus límites de variaciones hacían que solo a los usuarios con "buena letra" le hiciera un reconocimiento más o menos efectivo. Graffiti es más eficiente porque requiere que el

usuario aprenda los trazos y cada trazo está diseñado de manera que sea muy diferenciado de los demás. Aunque los trazos están inspirados en la forma de las letras, para promover un aprendizaje mnemotécnico o mimético, en realidad se trata de un conjunto de gestos arbitrario.

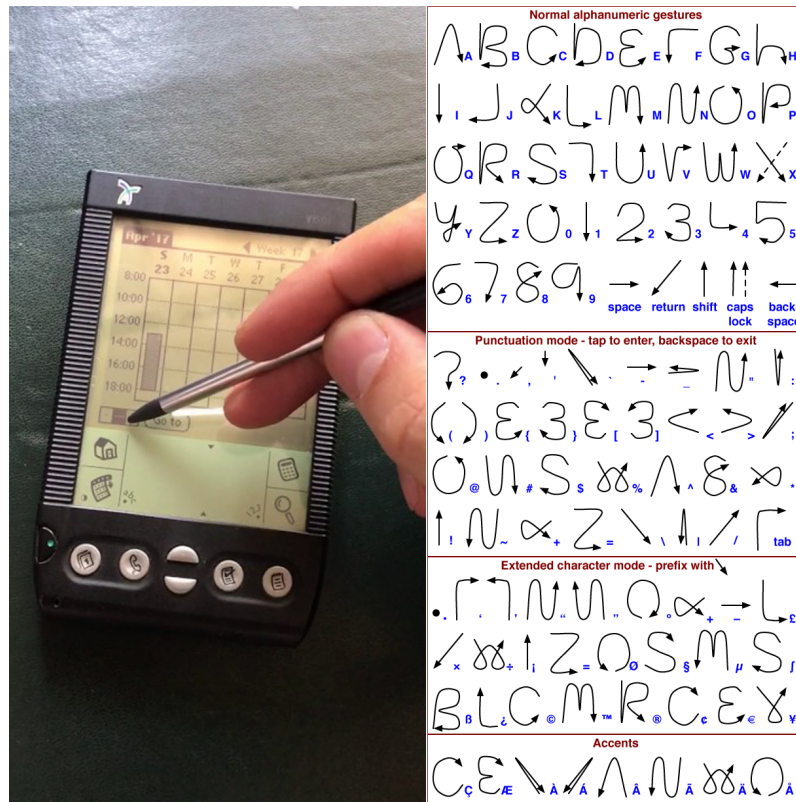


Ilustración 13: Izquierda: PDA marca Handspring. Derecha: los gestos del sistema Graffiti

El sistema graffiti prefigura varios de los aspectos que Nielsen propone como avances que sustituirán a la GUI. Nielsen llama a su propuesta la interfaz antiMac, pero en realidad es una interfaz que se contrapone a la interfaz gráfica WIMP, por lo que se presenta y se comprende mejor como la interfaz antiWIMP [GENTNER y NIELSEN 1996]. Aunque se trata de un comparativo que puede entenderse como una crítica de la GUI, y se pudiera interpretar como calificando a la interfaz WIMP como “mala”, más bien esta tabla pone en evidencia que las necesidades actuales

no coinciden con las necesidades que dieron origen a la GUI de paradigma WIMP. El cuadro fue traducido del inglés.

GUI WIMP	Interfaz antiWIMP
Los usuarios son "el resto de la gente" (no han tenido experiencia previa en el uso de computadoras)	Los usuarios son "La generación post Nintendo" (han crecido con computadoras)
Aplicaciones de "productividad" orientadas a las funciones de oficina	Trabajo, juego, interacción de grupos, aplicaciones embebidas, y computación ubicua
Computadora débil (128K RAM, 68000 CPU)	Computadora monstruosa (RAM multi-gigabyte, supercomputación en el bolsillo)
Ancho de banda para comunicación muy empobrecido (pantallas pequeñas, ingresar datos con teclado o mouse)	Comunicación enriquecida (la computadora puede verte, sabe tu ubicación en la tierra con su GPS, usa pantalla de alta resolución, ingresa datos con aparatos de entrada/salida novedosos)
El sistema está aislado y se mantiene estable hasta que el usuario decide hacer un cambio	El sistema está conectado y se mantiene cambiando sin intervención del usuario
Manipular íconos	Usar lenguaje
Débilmente orientado a objetos (usa un pequeño número de objetos muy grandes con pocos atributos)	Fuertemente orientado a objetos (un gran número de pequeños objetos, cada uno con muchos atributos)
El muy visible sistema manejador de archivos ("finder" o "explorer") es la base unificadora de inicio, y los archivos son el objeto básico de interacción	La recuperación de información personal es el principio de unificación con unidades de información atómicas como el objeto básico de interacción
Navegar por el disco duro	La información llega a ti
"El poder para ser lo mejor que puedes ser" (eslogan de la compañía Apple en 1984)	No siempre necesitas trabajar duro

Tabla 2. Comparación de las interfaces GUI WIMP y las interfaces antiWIMP. Fuente: [GENTNER y NIELSEN 1996]

2.5 Consideraciones de usabilidad, accesibilidad e inclusividad de las interfaces gestuales.

Uno de los puntos de inflexión en el que se da el avance de diseño de las interfaces gestuales es el de interfaces para la inclusión de las personas discapacitadas. Ya sea para el uso de la computadora como mediador del discapacitado, por ejemplo las computadoras con síntesis de voz para permitirle hablar a los que sufren parálisis de esclerosis múltiple; ya sea para permitirle el uso de la computadora misma y dar acceso inclusivo a grupos con alguna discapacidad, por ejemplo interfaces para ciegos y débiles visuales, para personas que perdieron el uso de un brazo, etc.

Algunos de los precursores del mouse de Engelbert fueron sistemas de apuntador montados en la cabeza o en el cuerpo. Aunque fueron desechados “por ser extremadamente incómodos”, dieron pie a la exploración de sistemas para los casos en los que el usuario estaba impedido del uso de las manos.

Desde luego que una propuesta obvia que se experimentó desde los inicios es el reconocimiento de los gestos del lenguaje de mudos. [KADOUS 1995] [STARNER y PENTLAND 1995] Varios de los métodos de reconocimiento de formas y patrones se verificaron usando el lenguaje de signos para mudos como objetivo o como punto de partida. Una de las ventajas de diseñar un sistema gestual que tome en cuenta personas con discapacidad es que muchas veces se trata de diseños que a las personas comunes les facilita el trabajo. En el caso de los sistemas de interfaz con el objetivo de recibir comandos a través de los gestos del lenguaje de mudos, se tiene el valor agregado de entrenar a gente que no es muda en un método de comunicación que a la larga beneficia a los mudos fuera del contexto de uso de la computadora; lo que es similar a lo que pasa con la gente que practica la lectoescritura en inglés por la disponibilidad de mayor cantidad de información en ese idioma en internet.

2.6 Posibilidades expresivas de la gestualidad.

A partir de la taxonomía de Nespoulous y Lecours, en la que el gesto es símbolo o acción dependiendo del ejercicio interpretativo que realiza el receptor y de la distancia de la interpretación del contexto en que se realiza el gesto, se desprende que una de las cualidades del gesto es que puede construirse de manera expresiva y estética, y que hay una dependencia del gesto de los contextos por lo que se puede establecer un sistema expresivo en el que para cambiar el significado de los gestos basta con hacer cambios en el contexto, dichos cambios están orientados a estimular o controlar la experiencia emocional del usuario.

La dicotomía que usan los semióticos en cuanto a la transparencia-opacidad de la interpretación puede ser útil para entender esta naturaleza del gesto como símbolo y sus posibilidades para agregar expresividad y emotividad a una interfaz. Una observación superficial lleva a la conclusión inmediata de que ciertos gestos son de una clase transparente que es inmediatamente evidente a todos los observadores sin tener que establecer una convención previa sobre el significado. Sin embargo, una observación más profunda lleva a la conclusión que dichas interpretaciones no son dominantes, y que incluso presentan problemas si quiere aplicarse en la práctica una interpretación como una clase de significados. Aunque algunos gestos tienen una extensión transcultural más amplia que la de sus equivalentes lingüísticos, las diferencias culturales subyacen y no pueden ser consideradas como insignificantes. Inclusive puede hablarse de diferencias en la interpretación individual. De hecho, la evolución diacrónica¹¹ de las culturas es la que lleva a la cercanía simbólica de ciertos gestos en cierto momento evolutivo, y la misma evolución diacrónica hace que ese significado social cambie y se vuelva opaco con el paso del tiempo. [QUEK 1994] Por ejemplo, el gesto con la mano que se asocia con la frase “OK” para indicar que algo está bien, tiene un tiempo de adopción distinto en cada cultura, y aunque en este momento se puede asumir una adopción casi universal de este gesto, puede haber culturas en las que el significado aun no esté consolidado, o sitios en las que el

¹¹ Diacronía es considerar todo el tiempo de existencia y la secuencia en la que algo existe

contexto de uso de ese gesto es más específico que en la cultura en que se originó (por ejemplo, que sea correcto usarlo en privado pero mal visto usarlo en público)

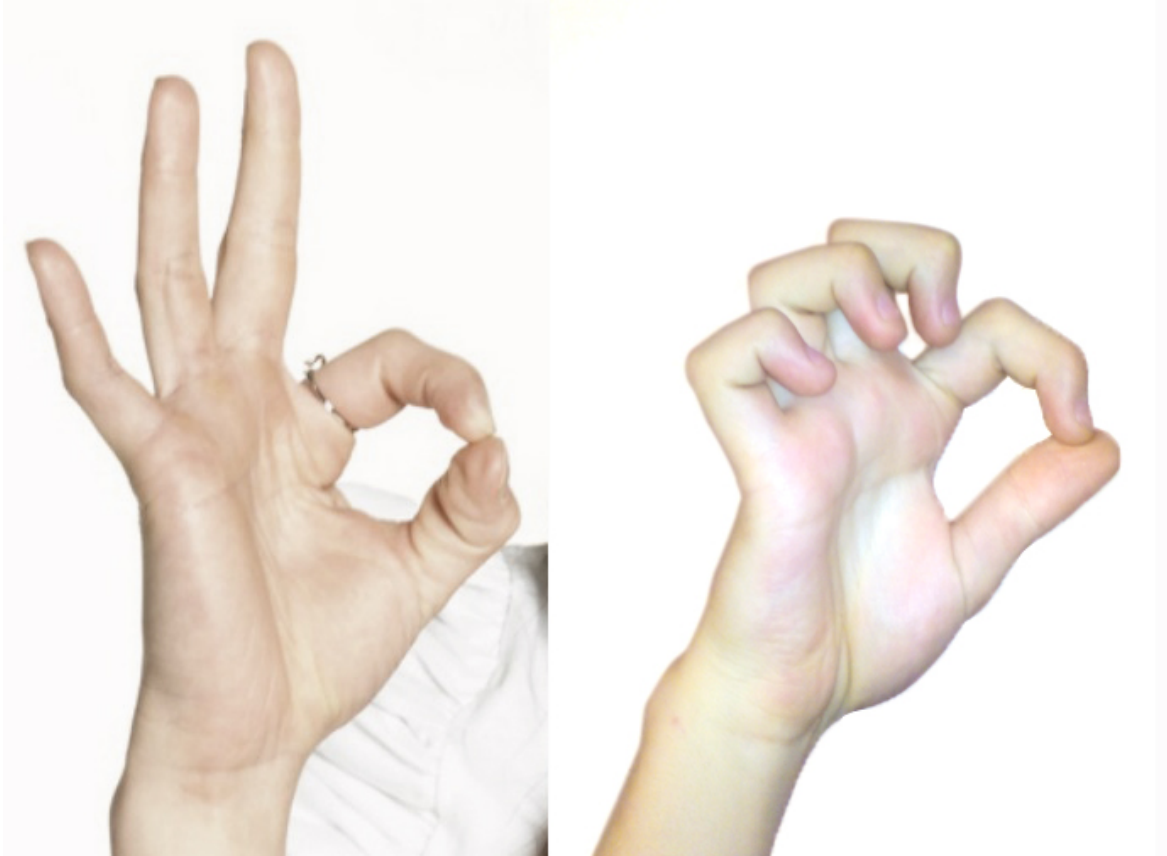


Ilustración 14: El signo aceptado para "OK" (lado izquierdo) puede ser fácilmente confundido con el gesto grosero de la derecha, pero solo en el contexto cultural de México.

Así como la aplicación de los gestos como lenguaje en el caso de los discapacitados orientó a los diseñadores de interfaz en las posibilidades del gesto como medio de control, los estudios neurolingüísticos han fundamentado el uso del gesto como medio de expresión e incluso como vehículo de un mensaje estético. Mientras que algunas interfaces toman el acto gestual en su aspecto deíctico y lo equiparan a las acciones que se realizan en una interfaz WIMP (apuntar, íconos que representan comandos, etc.) otros desarrolladores proyectaron las experiencias de la danza y la expresión corporal en la producción de sistemas interactivos.

La naturaleza opaca del gesto como símbolo se alinea muy bien con las pretensiones artísticas de dejar al espectador una interpretación abierta; especialmente cuando en una obra artística interactiva el espectador se convierte en usuario. Mencionaré como ejemplo la instalación interactiva de Camille Utterback y Romy Achituv “Text Rain”¹² o la instalación del colectivo japonés WOWlab “Light Rain”¹³

Ilustración 15: Text Rain, (1999)



Ilustración 16: Light Rain, (2008), WOWLab

Al analizar los ejemplos mencionados, se encuentra que el atractivo y el impacto emocional que estas instalaciones tienen sobre el espectador es pro-

¹²<http://camilleutterback.com/projects/text-rain/> (consultado en junio 2017)

¹³<http://wowlab.net/en/works/light-rain> (consultado en junio 2017)
<http://www.youtube.com/user/wowlab2008> (videos)

porcional a qué tanto se involucra al espectador con las reacciones visuales de los elementos proyectados como respuesta a los gestos. Al mismo tiempo que se usan los gestos como forma de control se está generando una interpretación del mensaje del artista. Sin los gestos, la interfaz de estas instalaciones no lograría construir la experiencia estética ni la inmersión del espectador en la obra artística. [LOPES 2009]

No son los únicos casos en los que la interfaz adquiere nivel de expresión artística, pero si son casos modelo de cómo el gesto posibilita llevar la interfaz a un grado de expresividad artística que no logran interfaces de otros tipos.

2.7 Puntos guía para el diseño de aplicaciones con interfaz gestual.

La decisión de usar una interfaz gestual puede partir de las más diversas razones, pero siempre es conveniente tomar en cuenta algunos lineamientos que se enumeran a continuación.

a) Imponer límites

Delimitar el conjunto de gestos a utilizar, de preferencia usando un muestreo poblacional, para lo cual se puede seguir el procedimiento propuesto por el grupo de investigación del Laboratorio de Visión de computadora y tecnología de medios de la Universidad de Aalborg, Dinamarca [NIELSEN et al 2003] que consta de tres pasos:

1.) Delimitación taxonómica.

Establecer las palabras y adjetivos con las cuales se va a describir el vocabulario de gestos que se va a extraer. Considerar dos campos de aplicación:

Etiquetas de descripción – describen el gesto en sí, ya sea estático (postura) o dinámico (movimiento)

Etiquetas de tipo semántico – describen lo que el gesto quiere comunicar. Esta comunicación puede identificarse como consciente, espontánea, interactiva o proposicional.

Al establecer la delimitación taxonómica es importante conocer el impacto cultural. Recordar que los gestos son manifestaciones de la educación y cultura de una población y que comunican con una fuerte carga de la cultura en que se originan. Es importante saber desde el principio si la dependencia cultural puede ser un problema en el tipo de aplicación al que se va a diseñar una interfaz gestual.

2.) Escoger el vocabulario de gestos.

Para esto una serie de ensayos y pruebas estadísticas con prototipos de papel o en situaciones simuladas pueden ser la mejor manera de delimitar el

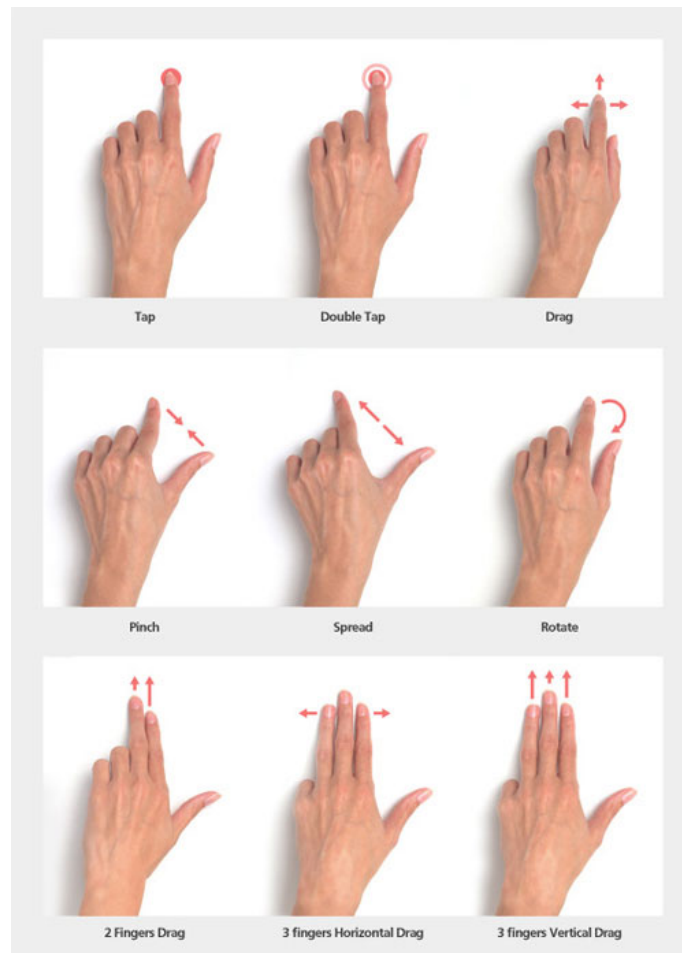


Ilustración 17: Gestos multitouch, que son el estándar de facto desde su introducción en 2007 en los productos de Apple

conjunto de gestos a usar. Este es quizás el paso más problemático ya que

existe el riesgo de establecer un vocabulario que sea difícil de recordar para el usuario o que no se ajuste a las necesidades de comunicación del proyecto.

3.) Filtrar el vocabulario con un criterio técnico.

Los gestos deben ser factibles técnicamente. Si la tecnología de reconocimiento de gestos no permite la captura más que de los gestos de las manos, no debe haber gestos faciales en el vocabulario, si la tecnología solo va reconocer posturas, la interfaz no debe sugerir que el usuario deba moverse mucho o muy rápido, etc.

b) Tamaño

Establecer el rango de escala del gesto. Esto consiste en saber si la sintaxis del gesto es la misma en diferentes grupos antropológicos, por ejemplo si se reconoce de igual manera cuando lo ejecuta un adulto de tercera edad o un niño pequeño, o si se va a limitar el rango de edades para no ampliar demasiado la necesidad de reconocimiento del gesto. Otro factor de escala importante es el de la escala temporal que debe definir cuanto tiempo se espera que el usuario pase ejecutando los gestos de la interfaz y si en estos tiempos puede causarse fatiga, ansiedad o fastidio, o por el contrario es posible que el usuario pueda mantenerse largo tiempo repitiendo los gestos. Qué se considera un tiempo largo o corto es muy relativo del tipo de gesto y el tipo de digitalización que se van a usar. Otro rango de escala es el de contraste, especialmente en los sistemas de reconocimiento visual o auditivo donde la legibilidad del gesto puede depender de factores externos como el nivel de iluminación, la distancia relativa entre el sujeto y el sistema digitalizador del gesto, etc.

c) Medir la eficiencia.

La interfaz gestual debe ser más conveniente y eficaz que la interfaz WIMP, o no será usada. La escala de eficiencia debe establecerse desde antes de empezar el diseño, aunque sea con medidas meta originadas en sistemas WIMP actuales.

d) Medir la curva de aprendizaje de la interfaz gestual.

Si la interfaz no se explica por si misma, debe haber una guía rápida y no obtrusiva que le ayude al usuario a familiarizarse con la interfaz.

e) Reconocimiento modal

Permitir el pausado o la inhabilitación del reconocimiento de gestos, y la reanudación del servicio a la conveniencia del usuario. Proveer alternativas no gestuales especialmente en las aplicaciones móviles.

f) Retroalimentación

Proveer un sistema que afirme al usuario que el gesto ha sido reconocido, especialmente en sistema donde la latencia del reconocimiento puede causar confusión e inseguridad en el usuario.

Esta guía de diseño establece los criterios con que se diseña el conjunto de gestos. La integración de estos gestos al resto del sistema y con los elementos adicionales que conforman una interfaz propiamente dicha se beneficiará mucho si en la planeación de este diseño de interfaz se tiene un objetivo, un tipo preestablecido al que apuntar como meta. La propuesta de tipología se plantea en el siguiente capítulo.

Capítulo 3.- Análisis formal de la interfaz Gestual

De lo expuesto hasta este punto se tiene una comprensión, aunque de manera un tanto inconexa, de los elementos que conforman una interfaz y de la forma que toma el gesto cuando se le estudia como elemento de comunicación. La unión de estos dos conceptos es posible mediante un análisis formal partiendo de las taxonomías ya existentes, produciendo una taxonomía que usa la diferenciación por la jerarquía del gesto en la capacidad de comunicación de la interfaz, capacidad que será determinada en gran medida por la forma en que incluye la semántica gestual. Se ejemplificará esta taxonomía con el análisis de algunos casos de estudio, validando las decisiones de diseño en lo que atañe a la conformación de la interfaz gestual, y con una visión crítica usando como criterio el acercamiento a la experiencia humana según el planteamiento de los 3 espejos de la interacción propuesto por B. Buxton.

3.1 Elementos comunes de las tipologías del gesto y propuesta de tipología.

Para entender el gesto como elemento formal de una interfaz, se presenta a continuación un análisis que enumera los tipos de interfaz usando como criterio

de identificación cual es la forma de inclusión del gesto y su efectividad en comparación con la interfaz gráfica. Esta comparación toma en cuenta que la interfaz gestual se produce en algunos casos por evolución empírica (lo que Nielsen y otros han definido como diseño iterativo)[NIELSEN 2000] y en otros casos por diseños basados en estudios con más solidez cuantitativa, como los estudios enfocados al diseño de interfaces específicas de programas de realidad virtual [VRST'94] o las interfaces experimentales de los Human Computer Interaction courses de la Universidad de Toronto¹⁴.

Las tipologías actuales se centran en aspectos del *hardware* usado, ya sea para la digitalización del gesto, la programación de la interactividad o para el despliegue de los contenidos. Estas tipologías están enfocadas a los aspectos más relevantes para los desarrolladores e ingenieros, enfatizan los elementos necesarios del sistema en cuanto a *hardware* y *software* en caso de incluir el gesto como medio de control. La tipología que se propone con esta tesis tiene como objetivo facilitar a los diseñadores gráficos entender el impacto del gesto como elemento del discurso de comunicación y aprovecharlo para diseñar un mejor mensaje visual que complete la experiencia del usuario. El énfasis de esta tipología está en el aspecto de comunicación del gesto y en cual es la manera de diseñar la interfaz entendida como medio de comunicación.

De la revisión de las tipologías existentes se desprende que hay elementos que vamos a integrar en la tipología que se propone, ya que se trata de elementos comunes en las diferentes tipologías, pero que en nuestra propuesta van a quedar enmarcadas de manera secundaria a la función de comunicación que realiza el gesto. Los elementos a usar son los siguientes:

a) Aparato digitalizador del gesto.

Todas las interfaces que utilizan gestos utilizan algún tipo de conversión analógica a digital en la cual el gesto se convierte en datos. Los métodos de conversión están muy bien estudiados y categorizados y por lo mismo es fácil

¹⁴<http://www.dgp.toronto.edu/> revisado el 10 diciembre de 2018

pensar que una tipología basada en el método de digitalización del gesto es la mejor manera de hacer una tipología de las interfaces gestuales.

Pero lo cierto es que los problemas técnicos de conversión de gestos en datos digitales no son los mismos que los problemas expresivos y estéticos que se desprenden del gesto como elemento de un mensaje. Estas tipologías dan relevancia al “*input device*” y consideran que la descripción de dicha pieza del sistema es la que da la definición más importante de las características formales de este. En realidad, es un poco a la inversa.

Si sabemos de antemano que existen diferentes tipos de gesto y cuales son los posibles discursos que se pueden construir con ellos, podemos elegir entre los múltiples métodos de digitalización al que se acopla con mejor conveniencia al mensaje que se construye. Sin embargo, no es posible hablar de una interfaz gestual sin contemplar que existe al menos un método de digitalización del gesto. Los más comunes hasta la fecha actual son:

Captura de video – Estos aparatos pueden capturar una secuencia de imágenes. El digitalizador de video típicamente tiene al menos una cámara, pero existen sistemas con múltiples cámaras que dan mayor libertad al usuario en cuanto a su posición para digitalizar. Usualmente los datos capturados se procesan mediante algoritmos de reconocimiento visual.

Captura de movimiento – Se usan aparatos para capturar una secuencia de cambios de velocidad y dirección. Típicamente el aparato de captura es una combinación de acelerómetros y giroscopios. Los datos capturados se procesan mediante algoritmos que interpretan ritmo, dirección y duración de los cambios en velocidad y ángulo.

Captura de cambios espaciales – Se usan aparatos para establecer cuál es el cuerpo que produce los gestos y su ubicación espacial. Se procesan los cambios no sólo de velocidad y ángulo sino también las variaciones de posición en un espacio tridimensional. Los aparatos de captura suelen ser apa-

ratos de digitalización 3d y de captura de movimiento, algunos usando combinaciones de captura de imagen, acelerómetros, giroscopios y otros aparatos especiales. Los procesos de interpretación casi siempre usan algoritmos predictivos para determinar no sólo un gesto sino preparar de antemano las consecuencias de la acción en un espacio virtual o real.

Captura táctil y háptica – Se usan aparatos para digitalizar puntos de contacto táctil o de fuerza de compresión o expansión. Los aparatos son típicamente superficies que el usuario puede tocar, ya sea directamente o con mediación de algún stylus y que pueden o no ser sensibles a la presión ejercida sobre la superficie. Algunos aparatos digitalizadores pueden incluso oponer resistencia a la fuerza ejercida sobre ellos, lo que normalmente se conoce como aparatos hápticos. Los procesos de interpretación asociados generalmente a este tipo de captura consisten en algoritmos de reconocimiento de formas mediante clasificadores o entrenamiento neuronal.

b) Algoritmo de interpretación del gesto.

Otro elemento común a todos los sistemas con interfaz gestual es el sistema que interpreta el gesto para asignarle una unidad semántica una vez que se ha digitalizado la información que lo forma. Mientras que para establecer un método de digitalización el enfoque se orienta a definir cuáles son las variables a digitalizar, los sistemas de interpretación toman los datos digitalizados y los procesan de acuerdo a algoritmos con los que se efectúa la separación de unidades significativas y en ocasiones se busca la organización de estas unidades en un discurso construido con gestos; la decisión de hacer la interpretación a nivel de cada gesto o de profundizar en interpretación de oraciones o discursos más completos depende de muchos factores, queda a discreción de los programadores el establecer los rangos de muestreo y la granularidad mínima del sistema, esto es, la fineza con la que se puede distinguir un gesto de otro. Otro factor que incide en el nivel de interpretación es el de la velocidad de respuesta que se espera en la interacción, y lo cual en muchas ocasiones implica programas que puedan filtrar las condiciones externas a los signos a reconocer propiamente.

Esto quiere decir que los valores digitalizados siempre tendrán una cantidad de datos no significativos, a los que se caracteriza como “ruido” entre los datos que realmente forman parte del gesto a reconocer, y que la labor de reconocimiento del gesto es tanto una labor de entender un gesto del siguiente, cómo la de separar los datos que no son parte del gesto de los que si lo son.

El programa de reconocimiento de los gestos es un elemento que existe en una capa abstracta en el sistema operativo, esto quiere decir que su funcionamiento no requiere una intervención del usuario, también se puede decir que funciona en forma pura lógica o matemática, de ahí su adjetivo de “abstracto”. Las capacidades del programa de reconocimiento de gestos van a establecer los parámetros de la capacidad de la interfaz, y esto si le interesa al diseñador de la interfaz gestual pues implica un límite en la factibilidad de los gestos de entrada o salida del sistema. Por el tipo de algoritmo que se utiliza, el sistema de reconocimiento de gestos se clasifica de la siguiente manera:

Sistema de sintaxis predeterminada y rango fijo de tolerancia a variaciones – El sistema reconoce gestos a partir de un diccionario preestablecido, y el usuario conoce el conjunto de gestos “permitidos”. El usuario se entrena y recuerda los gestos como recuerda un lenguaje formal. Estos gestos no pueden personalizarse o hacerse fuera del orden que el sistema espera, lo único que acepta el sistema de interpretación es cierta variación en la escala o la exactitud de cada gesto. Por ejemplo, el sistema que hace el reconocimiento de los gestos de desplazamiento de los dedos en una pantalla táctil o en un *touchpad*

Sistema entrenable asociado a una red neuronal – En la capa de abstracción se tiene programado un algoritmo de aprendizaje modelado a partir de la imitación de las conexiones neuronales de un cerebro. Estos sistemas aprenden por evaluación de ejemplos, y han sido usado tradicionalmente como base del reconocimiento de características o conjunto de cualidades que no es fácil programar en un set de reglas cerrado. Por

ejemplo, los sistemas biométricos que reconocen una persona por su forma de caminar, o que pueden reconocer formas (personas, animales, etc.) dentro de una fotografía.

El sistema de aprendizaje puede limitar el estilo o tipo de reconocimiento de gestos, es decir, establece por sí mismo un nivel de discriminación, o umbral, en el cual separa lo que no va a reconocer como gestos de lo que sí va a reconocer como tales. El aprendizaje del sistema no es sólo por reconocimiento visual, sino que usa también ritmos, repeticiones o combinación de entradas de sensores diversos pero en secuencias reconocibles, y es común que estos sistemas aprendan usando el modelo de Markov¹⁵ aplicado como método de aprendizaje de Inteligencias Artificiales. El modelo de Markov es un modelo de la teoría de probabilidades que se usa para encontrar patrones en conjuntos que cambian de manera aleatoria o estocástica. Lo que facilita su uso en sistemas programados es que el conjunto de datos con el que se entrena un aprendizaje puede ser replicado, lo que significa que se puede entrenar un sistema en reconocimiento de gestos y luego copiar el resultado de ese entrenamiento para no tener que entrenar cada sistema individualmente. La base de datos de lo aprendido puede ampliarse si se quiere refinar o perfeccionar el sistema, pero por lo general se puede detener el aprendizaje de la máquina una vez que se alcanza un nivel de precisión aceptable. En el caso del reconocimiento de gestos, este nivel aceptable ya incluye las variaciones más comunes en los gestos a reconocer, es decir, que se garantiza un reconocimiento positivo en la mayoría de la población de usuarios, a partir de un muestreo estadísticamente significativo.

Sistemas predictivos o heurísticos – En estos sistemas se construye un servicio de cálculo probabilístico que adelanta o predice los resultados más probables aunque el reconocimiento realizado del gesto sea parcial. A diferencia del aprendizaje basado en el modelo de Markov, estos sistemas heu-

¹⁵El modelo lleva el nombre del matemático ruso Andrei Markov, quien describió las propiedades que hacen posible este método en las últimas décadas del siglo XIX. El modelo consiste en describir cuál es la secuencia de eventos posibles donde la probabilidad de cada evento depende solamente del estado que se obtuvo del evento precedente.

rísticos son más veloces, pues no esperan a tener una confirmación dentro de la secuencia de eventos probables, y el conjunto de datos de entrenamiento para reconocer el gesto puede quedar “abierto”, es decir que puede afinarse o terminarse de entrenar de manera continua, siendo más preciso conforme más uso se le da al sistema. Esta cualidad puede ser una desventaja ya que el sistema se va refinando de acuerdo al uso individual, lo que impide replicar un sistema altamente entrenado para otro usuario diferente.

El sistema heurístico se usa principalmente cuando es crítico reducir al mínimo la latencia del reconocimiento gestual, en aplicaciones en las que se espera una interacción veloz y fluida; un ejemplo típico es en los juegos de video controlados por gestos, en los que el conjunto de acciones que se pueden realizar con, por ejemplo, la mano izquierda, son un conjunto de acciones bastante reducido y además no pueden ser realizados por, digamos, el pie o la cabeza, por lo que se puede depender de una predicción antes de que se de un reconocimiento completo. El sistema puede corregir sus predicciones aprovechando la velocidad de respuesta del sistema antes de tener que efectuar una retroalimentación visual.

Sistema adaptable o evolutivo con aprendizaje continuo – Este es un sistema mixto donde se combinan aprendizaje mediante un sistema neuronal tipo Markov y algún sistema heurístico abierto. Los algoritmos con que se combinan estos dos sistemas que le dan a las aplicaciones la capacidad de determinar cuando el refinamiento en el aprendizaje implica una diferenciación o mejora, y por lo tanto el sistema mismo puede cambiar y derivar sistemas más adaptados a situaciones específicas o más eficientes en la situación general inicial, es decir, “descendientes más evolucionados”. Estos sistemas adaptables serían capaces de reconocer gestos nuevos y las situaciones en que se aplican, y autogenerar nuevas versiones del sistema de abstracción del gesto. Esta capacidad es deseable cuando se quiere tener un sistema que sea cada vez más preciso y refinado para un usuario específico, esto es, un sistema personalizado para no tener errores en el reconocimiento

de gestos de al menos de un usuario principal. Este sistema puede ser un componente más de un sistema de Inteligencia Artificial con otros objetivos más amplios que la comunicación gestual.

c) Uso del gesto en el sistema de despliegue

Otro elemento común a las interfaces gestuales y que se usa para diferenciarlos es el tipo de despliegue que utilizan para enviar mensajes al usuario. En este caso se enfatiza al gesto como sistema de control de la interfaz, como queda expuesto en el primer capítulo la interfaz es un medio de control en la interacción hombre máquina. El elemento de despliegue es el que se encarga de la retroalimentación al usuario sobre los actos realizados, y cuando la interfaz es basada en mayor o menor medida en gestos también se puede hacer que el elemento de despliegue sea basado en gestos. Los algoritmos de reconocimiento de gestos pueden funcionar también como generadores de gestos, proveyendo de un sistema abstracto al que puede conectarse un sistema de despliegue. Este despliegue puede tener integrado los gestos en mayor o menor grado, por lo que según el grado en que se usan dichos gestos, podemos hablar de:

Despliegue sin respuesta gestual – La retroalimentación ocurre en el despliegue tradicional, como la pantalla, el sistema de audio, etc. La respuesta gestual puede derivarse en simulaciones de gestos, pero no hay un subsistema específico para general respuesta gestual. Por ejemplo, el reconocimiento de un gesto en un juego de video ejecuta las acciones esperadas, pero se afirma al usuario por cambios en las gráficas o por generación de sonidos. Incluso ciertas simulaciones, como el hacer que el personaje que se controla en el juego se alegre o entristezca, pueden considerarse como retroalimentaciones visuales gráficas, no gestuales.

Despliegue gestual con subsistema dedicado – la retroalimentación ocurre en un subsistema específicamente establecido para dar notificaciones. Este subsistema generalmente consta de mecanismos activación de servomotores, sistemas táctiles, elementos biónicos y similares, que son parte del

mundo real, implican una estimulación sensorial al usuario que no es simulada. Por ejemplo, los controles con capacidad de vibración, puntos metálicos que pueden producir presión en alguna parte del cuerpo para hacer saber al usuario el estado del programa que está usando, exoesqueletos que pueden transmitir temperatura o cambiar la rigidez de la articulación, etc.

Despliegue gestual con la misma sintaxis del sistema de digitalización – Como se mencionó anteriormente el algoritmo de reconocimiento de gestos puede usarse también como método de conversión de datos a gestos. Al usar un *hardware* especializado en la digitalización del gesto, es también posible usar ese mismo *hardware* para dar la retroalimentación en el mismo nivel de comunicación gestual. Este tipo de despliegue se está desarrollando especialmente en los robots androides. Por ejemplo, si el reconocimiento de la interacción del usuario se da por gestos faciales, la respuesta se produce en expresiones faciales en un robot dotado de algún tipo de cara. El sistema está equipado para responder sin tener que traducir la respuesta a otro código gestual.

Propuesta de Tipología.

En las taxonomías actuales con estos tres elementos comunes se construye una interfaz gestual: Aparato digitalizador, algoritmo de reconocimiento y sistema de despliegue. En todos estos casos se considera el gesto como un elemento subordinado a una de las funciones del sistema, mientras que todas las interfaces se tipifican como gestuales.

Es posible analizar cuál es el papel del gesto y separarlo como un cuarto elemento, y de ahí establecer una clasificación de las interfaces gestuales de acuerdo a la jerarquía que adquiere el elemento gestual. De este análisis se desprende cuáles son las variaciones en la creación de un campo semántico en el gesto, lo que es importante a la hora de proponer un nuevo diseño de interfaz o de realizar mejoras en un diseño existente.

Para la tipología que se propone a continuación se considera el tipo de comunicación que se desarrolla y cómo substituye a la interfaz gráfica WIMP. Por lo expuesto en los capítulos anteriores, se ve la necesidad de que el gesto se tome como factor de diferenciación, de considerar al gesto como un signo, base de un sistema de comunicación y comprender las necesidades que va a satisfacer el diseño de interfaz según el papel mayor o menor del gesto en lo comunicado.

La tipología propuesta ha determinado siete tipos de interfaz, la diferenciación de un tipo a otro se basa en la exclusividad del gesto en la transmisión de significado.

Tipo WIMP que incluye gestos

El gesto es un método adicional de entrada de datos, redundante a los métodos tradicionales de la interfaz WIMP que son típicamente mouse y teclado, el gesto tiene menor o misma jerarquía que los otros métodos de interfaz.

Tipo gestual que incluye WIMP

La interfaz está diseñada para que el principal método de interfaz sea el gesto, pero se incluye en el diseño (muchas veces como alternativa oculta o complemento) una interfaz WIMP que es de menor jerarquía que la interfaz gestual.

Tipo gestual con lenguaje de gestos estricto

La interfaz es controlada únicamente con gestos, se diseña un lenguaje de gestos preestablecidos lo que implica que el usuario debe entrenarse en la sintaxis y morfología del gesto.

Tipo gestual de sintaxis flexible

La interfaz es controlada únicamente con gestos, pero la morfología de los gestos está parcialmente definida de manera que el usuario puede

establecer variaciones, que el sistema aprende a reconocer, es decir que el método de comunicación se adapta y entrena para cada usuario a partir de una sintaxis incompleta que debe ser entrenada para el reconocimiento de gestos.

Tipo gestual de ubicación libre

El sistema de reconocimiento de gestos no está restringido por el espacio donde se efectúa la entrada de datos, sino que puede reconocer usuarios y gestos en diferentes posiciones espaciales. El gesto tiene una cualidad tridimensional que facilita el uso del cuerpo entero en la generación de la comunicación gestual

Tipo gestual con control por voz

El diseño de la comunicación con el sistema es a través de gestos pero se agrega un sistema de reconocimiento de voz para control y activación de la interfaz. El control de voz está limitado a la activación de la interfaz gestual, en este sentido no es una interfaz del sistema, sino más bien una parte de la interfaz gestual.

Tipo especializado en inclusión de personas con discapacidad

El diseño de interfaz es dirigido por las necesidades especiales de un grupo de usuarios con capacidades diferentes. Esto determina características de los gestos que probablemente sean más naturales a estos grupos que la comunicación oral o escrita, y el diseño puede alimentarse de un sistema de comunicación gestual elaborado para el mundo real y que ya ha sido usado y refinado por largo tiempo, similar a como se ha migrado el uso de el texto y el lenguaje a la interfaz de computadora no gestual. Este tipo de interfaz ha sido impulsada desde la política de inclusividad y se ha comprobado que los mejores diseños trascienden el uso en el grupo para el que son diseñados, es decir, pueden ser adoptados exitosamente por usuarios que no tienen la discapacidad para la que estas interfaces están dirigidas.

3.2 Elementos formales de la interfaz WIMP que incluye gestos.

Como esta interfaz se ajusta al paradigma WIMP, los elementos siguen la guía de uso del sistema y están subordinados a los elementos gráficos. Las capacidades gestuales que se agregan no toman un lugar jerárquico superior al de la interfaz gráfica. Se espera que cada elemento gestual tenga una equivalencia en los elementos WIMP. Por ejemplo, si formalmente se tiene que un manipulador de ventanas requiere arrastrar un botón de izquierda a derecha para aumentar el tamaño de lo que se visualiza, el gesto equivalente respetará formalmente que la acción se realice de izquierda a derecha para aumentar. La jerarquía del gesto está al mismo nivel que el teclado o el mouse, la interfaz gestual es redundante a otros métodos de entrada de datos y el usuario aprende la equivalencia de los símbolos gestuales a las acciones de la interfaz gráfica WIMP. Ha sido el tipo de interfaz más rápidamente implementado.

Antecedentes:

Casi todos los ejemplos mencionados en los capítulos anteriores son implementaciones de este tipo de Interfaz. Es interesante notar que en muchos casos estas interfaces son implementaciones de algún tipo de *hardware* de entrada de datos, no necesariamente de un digitalizador de gestos, pero que agrega una nueva capacidad al sistema operativo y por lo tanto el *software* de controlador incluye el reconocimiento de gestos.

La adopción generalizada de las pantallas táctiles y de los *TouchPad* como alternativa al *mouse*, especialmente en los aparatos móviles como *tablets* y teléfonos, ha sido la puerta de entrada de la interfaz gestual más difundida, la de los gestos de control propuestos por Apple en su touchpad “*multitouch*”.

Las implementaciones más populares están basadas en el código de la investigación realizada en la Universidad de Washington con inclusión de inves-

tigaciones de desarrolladores de Microsoft [WOBBROCK, WILSON y LI 2007] Aunque la investigación se publicó en 2007, el código con el que se podían realizar prototipos funcionales ya existía, estuvo en proceso de desarrollo y disponible públicamente desde 2001, con kit de desarrollo en versión 1 liberado públicamente desde el 2005, por lo que fue ampliamente adoptado por múltiples emprendedores públicos y privados, lo que dio pie a un conjunto de gestos que, por el uso y la amplia difusión pueden considerarse estándar *de facto*. El conjunto de gestos al que nos referimos puede verse en la Ilustración 17 en el capítulo precedente.

Características formales

El sistema de digitalización de gestos está implementado en el mismo *hardware* común que se usa también para el control de la interfaz gráfica WIMP; las formas gestuales gestual se diseñan aprovechando la capacidad gestual de un sistema de input que no fue diseñado específicamente para gestos. La sintaxis gestual es establecida de antemano y tiene un equivalente en la sintaxis general del sistema.

Los elementos gestuales son modales, esto es, pueden activarse o desactivarse, individualmente o en conjunto.

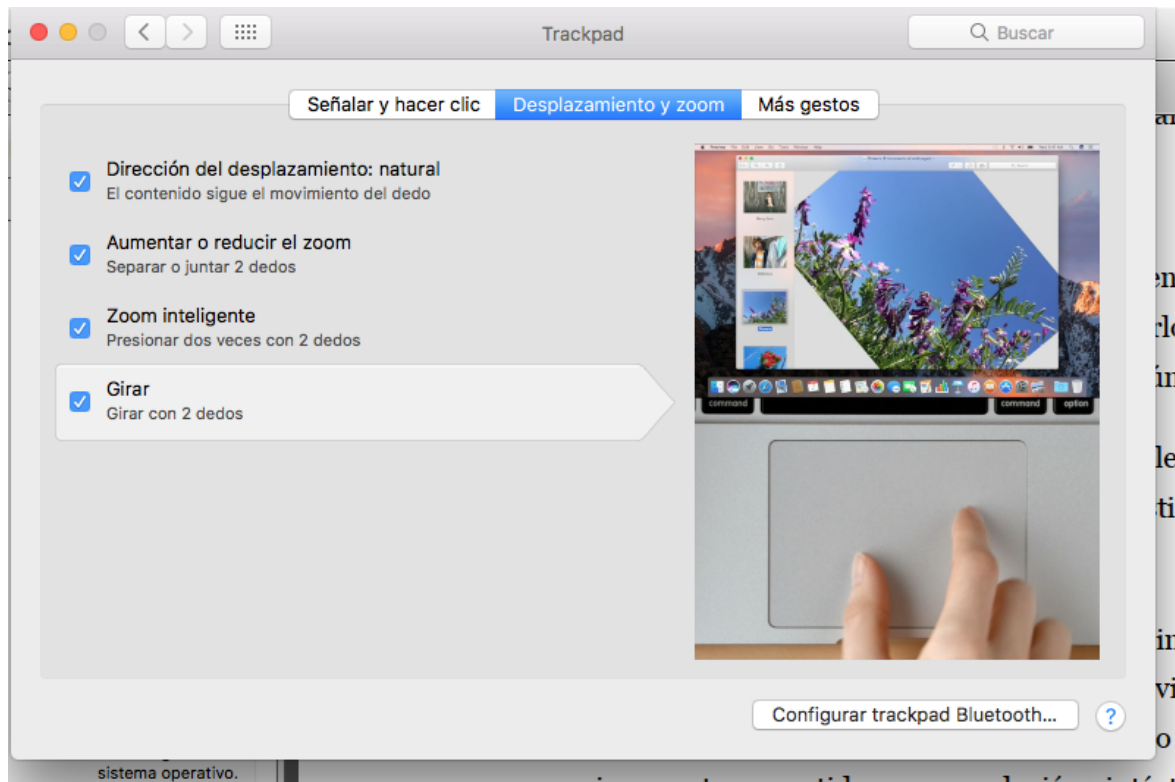


Ilustración 18: Configuración de gestos en un sistema con interfaz Wimp, en este caso MacOS X

Los gestos se reconocen en una capa de abstracción que no requiere respuesta gestual. La respuesta del sistema es manejada en el nivel estándar del sistema (por ejemplo, despliegue en el monitor de alertas de la interfaz gráfica) una vez que la comunicación de los comandos es confirmada. Como el despliegue está dentro del estándar del sistema, el input gestual es considerado sólo como un complemento de las acciones de la interfaz gráfica WIMP de selección, arrastrar, ejecutar, etc.

En este tipo de interfaz Los gestos están fijos en su forma de ejecución, sin embargo el usuario puede configurar respuestas personalizadas al gesto reconocido; el panel de control del sistema operativo permite sobreponerse a las acciones predeterminadas, es decir, redirigir el significado de un gesto a un comando distinto. Esto no es una característica exclusiva de la interfaz gestual, sino una característica para todos los métodos de input, por lo que es posible que se tome la deci-

sión de restringir el diseño de la interfaz gestual a unos pocos elementos sabiendo que pueden ser personalizables con facilidad.

Todo elemento gestual tiene un elemento equivalente en la interfaz gráfica WIMP. En la jerarquía del sistema la interfaz gráfica tendrá prioridad, la interfaz gestual será de una jerarquía secundaria o menor.

Ventajas y limitaciones:

Es una interfaz muy eficiente en el uso de recursos del sistema, al estar confinado a un conjunto limitado de gestos y a sistemas de digitalización estándar. Se han desarrollado por varios años suficientes sistemas como para que al día de hoy sean comunes y populares, y con suficiente experiencia se ha podido refinar su diseño corrigiendo los problemas y verificando los aciertos.

En lo que se refiere a la curva de aprendizaje, el usuario debe entrenarse en la sintaxis de los gestos antes de poder usarlos intuitivamente. El que el conjunto de gestos se haya implementado de manera casi idéntica en diferentes plataformas hace que este aprendizaje no dependa de una plataforma específica e incluso le facilita al usuario el uso de gestos en diferentes aparatos domésticos que los implementan.

En cuanto al acercamiento a la experiencia humana integral, este tipo de interfaz gestual no hace una aportación importante a la intención de reflejar las capacidades humanas. La supeditación a la interfaz gráfica siempre limitará su potencial expresivo y sensorial.

3.3 Elementos formales de la interfaz gestual que incluye WIMP.

Esta interfaz puede considerarse como un modelo de transición hacia una interfaz gestual más pura o estricta. Es un modelo mixto en el que se conserva la interfaz gráfica WIMP, pero en un nivel más bajo de la jerarquía de métodos de

input del sistema. Si la comparamos con el tipo de interfaz propuesto en el punto 3.2 podemos decir que se invierte la relación jerárquica entre la interfaz gestual y la interfaz gráfica.

Antecedentes

Se puede fijar el inicio de este tipo de interfaz en el año de 1987, cuando se hacen disponibles de forma comercial los sistemas de realidad virtual¹⁶, lo que requería de diseño de interfaces que dieran solución a los problemas de los sistemas inmersivos para manipulación de mundos virtuales ya sea para objetivos funcionales o recreativos. En ese año compañías como SUN o Silicon Graphics estaban proveyendo de sistemas especializados de *hardware*; y emprendimientos privados y públicos estaban creando aplicaciones en múltiples campos. Se hizo patente que al avanzar la representación virtual de la realidad, se necesitaba avanzar también en el diseño de una interacción más similar a la interacción del ser humano con el mundo real. La interfaz gráfica tradicional para estas interacciones claramente no era suficiente. Las exploraciones académicas de la digitalización de gestos se aprovecharon para propuestas de interfaz que suplieran esta insuficiencia.

La primeras implementaciones comerciales dirigidas al público en general que incluyen interfaces gestuales fueron en el campo de los juegos de video, y aparecieron alrededor del año 1992. Implementaban la interfaz con reconocimiento de gestos como sistema primario de interacción, las más interesantes de este tipo de interfaces derivan sus elementos gestuales de posturas y actividades que los usuarios usan en situaciones del mundo real, por lo que no tienen que pasar por un entrenamiento en una sintaxis especial para el control del sistema. Entre 1992 y 2006 el *hardware* y el *software* que utilizaban interfaz gestual fueron avanzando, saliendo de los laboratorios universitarios y de los video arcades comerciales hasta llegar a la consola Wii de la compañía Nintendo, que estuvo dispo-

¹⁶A finales de 1985 Jaron Lanier y otros fundaron la compañía VPL Research, que introduce tecnología de realidad virtual con las marcas Dataglove, eyePhone y AudioSphere. Mattel, la compañía fabricante de juguetes, hizo disponible el PowerGlove en 1987, marcando la primera vez que un accesorio de realidad virtual se comercializó al público en general.

nible a partir de noviembre de 2006 y cuyo gran éxito comercial significó que millones de personas conocieran y usaran de manera doméstica una interfaz controlada por gestos.

Características formales:

Los elementos gestuales son dependientes de un sistema de reconocimiento de gestos específico, es decir, su función primaria es para el input de gestos. Los elementos de interfaz gráfica WIMP no son parte de la interacción principal del sistema y la comunicación del sistema de gestos no tiene un equivalente en la interfaz gráfica WIMP. La interfaz gestual se diseña específicamente para el sistema digitalizador, y la parte de interfaz gráfica se separa porque es para realizar funciones que no son las de la interacción principal, como el “*setup*” inicial del sistema de gestos, o indicar el inicio y el fin de la interacción gestual.

Otro elemento de este tipo de interfaz es que la interfaz gráfica es el canal por el que se presenta la información de retroalimentación al input gestual. La solución de diseño más común para esto es una interfaz mínimamente visible, en muchos casos inspirada en el HUD (*Heads-Up Display*), despliegue sobrepuesto a la cara, nombre que se deriva de la proyección dentro del casco de piloto de aviones jet de la información relevante de todos los sistemas de control.

Los elementos gestuales serán un conjunto bien delimitado con una sintaxis clara y sin ambigüedades, puesto que el punto de partida de su diseño está en las características y limitaciones del aparato digitalizador de gestos.

Si el aparato de digitalización lo permite, la comunicación gestual se ejecutará en oraciones, es decir instrucciones complejas producto de la concatenación de gestos que signifiquen acciones y sujetos de esas acciones. Por ejemplo, un simulador de tenis puede combinar velocidad, dirección y posición del control para que el gesto de “golpear la pelota” se interprete, primero, como si hubo contacto o no

hubo contacto con la pelota y, segundo, cómo en qué dirección y con qué velocidad va a cambiar la trayectoria de dicha pelota.

Si el aparato de digitalización no permite comunicación gestual en oraciones complejas, o se toma la decisión de limitar el método de comunicación gestual a elementos simplificados, la comunicación gestual se diseña para hacer una equivalencia de cada gesto con una instrucción o comando del sistema. Por ejemplo, en un *software* de presentación de diapositivas, con un gesto se da la instrucción de avanzar a la siguiente, otro gesto es regresar, otro gesto es ampliar, otro es terminar la presentación, y así sucesivamente.

Estos gestos, sean complejos o sencillos, son un conjunto finito y determinado de antemano, forman un vocabulario fijo, con poca o ninguna tolerancia a las variaciones que puede introducir el usuario, y son necesarias dos condiciones: un periodo de entrenamiento del usuario en el uso del vocabulario; y homogeneidad de la experiencia en el uso de la interfaz gestual para todos los usuarios.

Ventajas y limitaciones

La curva de aprendizaje puede ser muy prolongada, ya que requiere que el usuario conozca y se entrene en una sintaxis gestual preexistente. Se puede acelerar esta curva de aprendizaje optando por una sintaxis de pocos elementos, en algunos casos, o por una sintaxis derivada de gestos ya bien establecidos en el mundo real. A esta última sintaxis también se le llama de simulación de gestos. La simulación de gestos tiene limitantes culturales, la naturalidad de ciertos gestos depende del grupo humano que los inventó. Sin embargo, construir un vocabulario de gestos que no dependa de la simulación de gestos reales puede balancear la dificultad de aprendizaje con la trascendencia entre grupos humanos. El conjunto de gestos es independiente de las condiciones individuales del usuario, lo que favorece una adopción universal y favorece asimismo la inclusión de grupos

heterogéneos de edad, de grupos limitados por cualidades físicas o simplemente la inclusión a través de barreras culturales, geográficas o históricas.

Mantener la interfaz gráfica WIMP es ventajoso ya que se puede separar la interacción principal de las interacciones secundarias; asignar las interacciones secundarias a la interfaz gráfica simplifica y hace muy eficiente la interacción principal.

Una desventaja es que la recuperación de la interacción cuando hay una falla crítica del sistema cuando se está en la interfaz gestual no se diseña de la misma manera que cuando se hace la recuperación del sistema con interfaz gráfica. Es decir, es deseable que si se está en la interfaz gestual no se tenga que salir de ella para detener, cancelar, reiniciar y en general recuperarse de una situación inesperada o errónea. Es muy fácil omitir el diseño de esta parte de la interfaz por estar acostumbrado a que la interacción normal en estos casos procede de la interfaz gráfica.

Un vocabulario limitado de gestos, aunque ventajoso como ya se explicó antes, también puede implicar que no sea una interfaz que se pueda trasladar fuera del sistema de digitalización del gesto y del contexto preciso para el que se diseñó. Por ejemplo, los gestos para una simulación de tenis no tienen el mismo significado en un juego de baile, aunque la interfaz gestual de ambos se ejecute en la misma consola y con los mismos controles.

En este tipo de interfaz gestual la interacción lograda refleja mejor las capacidades humanas. Especialmente cuando la interfaz deriva su vocabulario de gestos de una simulación de los gestos usados en el mundo real se refleja mejor la expresividad y emociones humanas. Mantener la interfaz gráfica WIMP para aspectos del manejo del sistema puede considerarse un lastre pero necesario por las limitaciones tecnológicas existentes.



Ilustración 19: Imagen publicitaria del juego de entrenamiento y cultura física "active" de la empresa Electronic Arts. A la interacción del wiimote se agregaba una banda sujeta en la pierna con monitor de ritmo cardíaco, temperatura corporal y presión arterial, con el objetivo de verificar si efectivamente se estaba ejercitando con el programa.

3.4 Elementos formales de la interfaz gestual con lenguaje de gestos estricto.

Esta interfaz se origina en el sistema digitalizador del gesto y construye las posibilidades de interacción de manera exclusiva alrededor de las posibilidades de este sistema.

Antecedentes:

Estas interfaces son las mas utilizadas en las aplicaciones de tipo artístico y expresivo. Casi siempre tienen que ver con el desarrollo de algún tipo de *hardware* y con alguna aplicación de demostración de las capacidades de este *hardware*. Las pantallas montadas en la cabeza (*head mounted display*), los guantes hápticos (*dataglove*, 1987) las “varitas mágicas” y otros dispositivos apuntadores con *tracking* (registro de la posición relativa) en espacio tridimensional, y otros dispositivos que incluían de forma implícita la capacidad de digitalizar gestos estuvieron en desarrollo académico entre 1960 y 1990, pero es la dis-

ponibilidad comercial de estos sistemas, a partir de 1985, la que da pie al desarrollo de este tipo de interfaz. Hoy día está migrando a aplicaciones en teléfonos inteligentes y sistemas domésticos de entretenimiento, principalmente por el crecimiento exponencial de la capacidad de cómputo y despliegue gráfico.

Características formales:

Un solo método de interacción basado en un dispositivo de digitalización del gesto.

Sin interacción WIMP, o el sistema operativo WIMP es un sistema anfitrión de una aplicación que en primer plano no tiene interacción WIMP.

El usuario se entrena en la sintaxis de la interfaz, que es de tipo estricto pero contiene un conjunto más amplio de gestos reconocibles, sobrepasando con facilidad los 50 gestos.

Ventajas y limitaciones:

No hay personalización de gestos, pero el sistema tiene tolerancia a las variaciones sintácticas.

Para facilitar la curva de aprendizaje, suele utilizar gestualidad natural o mimética de situaciones reales.

Tiene que tomar en cuenta los problemas de latencia entre el reconocimiento del gesto y el resultado de la interacción.

Puede presentar problemas de ergonomía al imitar situación del mundo real pero sin proveer de los equivalentes virtuales de estructura, confort, espacio y retroalimentación háptica. Problemas típicos de estas interfaces son, por ejemplo: gesto fuera de escala (gestos que requieren más espacio del disponible en el puesto de interacción), cansancio por posturas incómodas, latencia inaceptable.

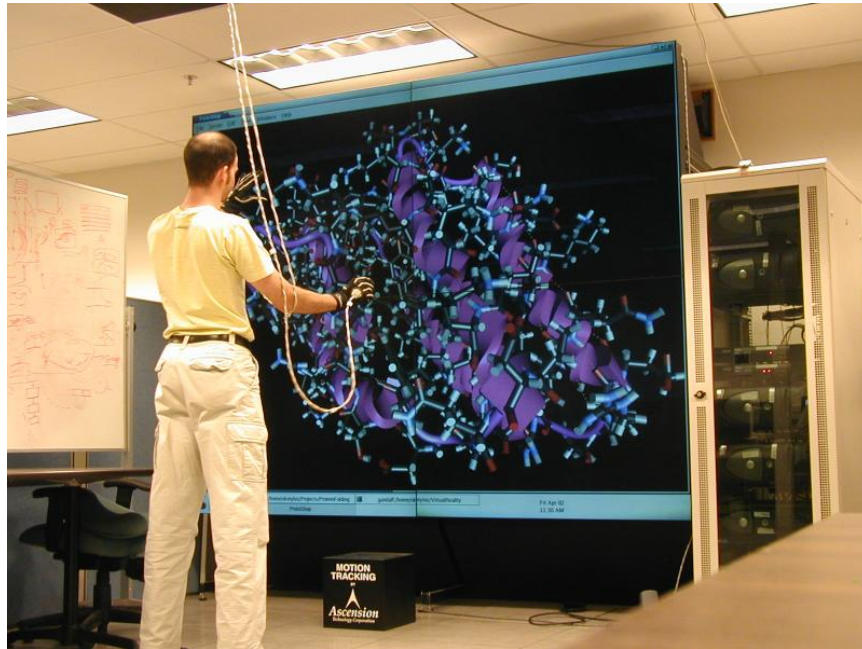


Ilustración 20: Manipulación de modelos de proteínas mediante el *software* ProteinShop y dos guantes datagloves. ProteinShop tiene una sintaxis estricta de comandos mediante gestos, desarrollo de la University of California, Davis- Computer Science department. La persona en la foto es autor de ProteinShop , Oliver Kreylos

3.5 Elementos formales de la interfaz gestual con sintaxis flexible.

Estas interfaces se diseñan para sistemas que necesitan reconocer los gestos específicos de un usuario por razones de privacidad, para limitar el acceso a la aplicación a un sólo propietario; o por razones de comodidad, cuando la interfaz aprende las cualidades únicas de un usuario para eliminar obstrucciones ocasionadas por las desviaciones de la normal en la ejecución de un gesto y hacer más eficiente el sistema de reconocimiento de gestos. La sintaxis con la que inicia la aplicación se va refinado y ajustado con el uso continuo, hasta que se vuelve exclusiva de una personalidad de usuario.

Antecedentes:

Este tipo de interfaz se ha desarrollado en estrecha relación al desarrollo de las máquinas capaces de aprender. Para estas máquinas es necesario un proceso de entrenamiento, y dado que en la mayoría de los casos estos sistemas no tienen

un despliegue visual de pantallas o dispositivos gráficos, puesto que se incluyen en aparatos robóticos, es deseable que el proceso de aprendizaje se realice por el reconocimiento de gestos. Los programas que pueden aprender por si mismos se empezaron a desarrollar desde mediados de la década de 1960 pero los estudios con diseño de interfaces gestuales para el proceso de entrenamiento no ocurrieron hasta después del año 1992 posiblemente porque el desarrollo y eficiencia del *hardware* no lo hizo factible antes [WU y HUANG 1999].

Otra necesidad para la que este tipo de interfaz gestual es muy apropiada es la de proveer control mediante factores biométricos. Se tienen estudios que indican que la forma de caminar, los manierismos, el ritmo de las articulaciones y otros gestos conscientes o inconscientes pueden usarse para identificar individuos de manera tan exacta como las huellas dactilares.[LIANG HUANG et al 2003]

Características formales:

El sistema de digitalización de gestos tiene cualidades de inteligencia artificial para realizar el reconocimiento del gesto e irlo refinando en diferentes iteraciones hasta conseguir distinguirlo de una manera única

Los gestos que utiliza la interfaz se establecen de forma abierta y se refinan con el uso de la aplicación, con la intención de distinguir las sutilezas más mínimas en la ejecución de un gesto.

El diseñador sólo establece el conjunto inicial de datos de los gestos, con una sintaxis que organiza la relación con las actividades a realizar una vez que el gesto se reconoce a nivel aceptable. Las acciones no varían a partir de ese punto, el gesto sí va a estar sometido a una evolución sintáctica, al grado de poder identificar de manera individual a un usuario de otro.

Ventajas y limitaciones:

Estos sistemas permiten la personalización completa. Otorgar acceso y control a un solo usuario, con las ventajas de privacidad y comodidad que eso implica.

Otra ventaja es la de la posibilidad de entrenar el sistema de manera expresiva reflejando la individualidad del usuario. Esto puede ser un atractivo que impulse aplicaciones de carácter emocional y artístico.

Estos sistemas son los más adecuados para incluirse en robots autónomos y autómatas pues el uso de gestos para controlarlos tiene su contraparte en la posibilidad de recibir retroalimentación gestual, y la personalización de los gestos de entrada puede producir una personalidad de reflejo en los gestos de retroalimentación, lo cual puede ser base de una comunicación más empática con las máquinas. Esto puede ser importante en, por ejemplo, robots de apoyo médico.

Entre las desventajas de este tipo de interfaz gestual es que el *software* aprende mediante algoritmos de inteligencia artificial, que consumen muchos recursos de cómputo y es un campo en el que los avances han ido a un ritmo más lento que en el desarrollo de otros tipos de algoritmo.

Otra desventaja a tomar en cuenta es que hay una curva de aprendizaje por parte del *software*, siempre se requiere un periodo de tiempo para que con la interacción del usuario el sistema aprenda las variaciones individuales; el diseñador de la interfaz debe proveer de un conjunto inicial de gestos lo suficientemente completos para que el sistema sea útil durante el periodo de entrenamiento y prever cual es el tiempo normal de entrenamiento y como puede continuar en funcionamiento un sistema si por alguna razón no puede ser entrenado en el tiempo asignado.

Una última desventaja es que en ciertos casos factores externos o temporales pueden modificar la capacidad gestual del usuario, y si el sistema está suficientemente avanzado en el reconocimiento de un gesto de manera exclusiva, esto puede efectivamente negarle al usuario la comunicación con la aplicación al no

existir una alternativa a la interfaz. Por ejemplo, si un sistema está entrenado a partir del uso de gestos con la mano derecha, y el usuario sufre un accidente que le impide el uso de esa mano, el usuario queda efectivamente bloqueado del uso de los programas.

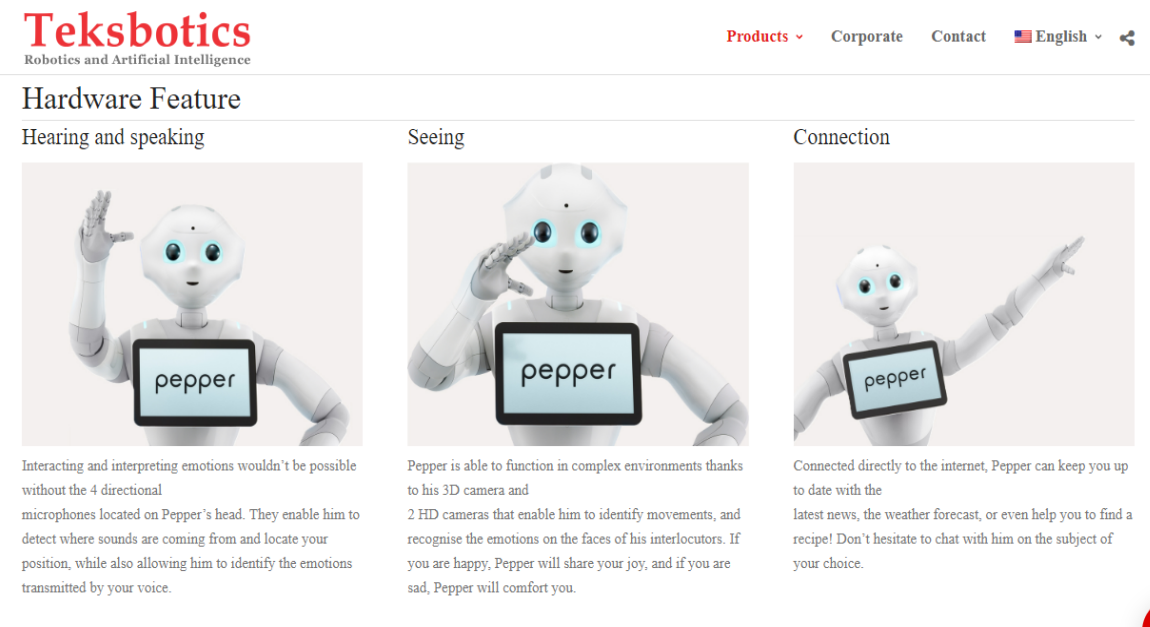


Ilustración 21: Robot Pepper, puede reconocer gestos y emociones del interlocutor y aprender gestos para recibir órdenes

3.6 Elementos formales de la interfaz gestual de ubicación libre.

Para digitalizar el gesto esta interfaz tiene sensores dirigidos directamente al cuerpo del usuario. Los sensores pueden transmitir datos muy heterogéneos (posición espacial, velocidad, dirección, orientación, calor, humedad, ritmo cardíaco etc.) y el *software* puede utilizar reconocimiento de patrones para establecer un gesto, aunque este gesto no sea consciente por parte del usuario.

Antecedentes:

Estas interfaces se derivan de los sistemas de monitoreo de salud, por una parte, de los sistemas hápticos para entrenamiento en mundos de realidad virtual, por otra parte, y del desarrollo de protocolos de comunicación de bajo

consumo de energía o de activación pasiva (como el RFID - *Radio frequency ID* – identificación por radiofrecuencia).

Está también fuertemente asociado a lo que a partir de 1994 se conoce como “*Weareable computers*” (computadoras que se usan como vestimenta) y que en los primeros años de los 2000s empezó a ser llamado como “*Internet of things*” (internet en las cosas). En estos casos se trata de la integración de computadoras en objetos que les permita la cercanía al cuerpo humano de manera que se tenga una captura de los datos de postura, ubicación, velocidad, etc. datos que pueden identificarse como un gesto. No se requiere de un algoritmo de inteligencia artificial para el reconocimiento del gesto, solo un sistema relacional de los diferentes datos de los sensores para establecer un rango de reconocimiento positivo.

Estas computadoras han dado pie a la aparición, en tiempos más recientes, de múltiples controladores domésticos que pueden ser manejados con gestos capturados por los acelerómetros de un teléfono. Estos controladores pueden accionar interruptores eléctricos, servir de control remoto a la televisión, abrir cerraduras de puertas, etc. La disponibilidad comercial de estos sistemas se ha popularizado bajo la etiqueta de “*Smart Home*” o casa inteligente.

Características formales:

El reconocimiento del gesto se realiza por la deducción de un conjunto heterogéneo de datos, en los cuales se reconoce un patrón. Los sistemas de reconocimiento heurístico suelen ser los más comunes para esta tarea.

Los gestos se diseñan para ser reconocidos independientemente de la posición relativa del usuario con respecto al aparato digitalizador. El reconocimiento del gesto puede suceder en ambientes heterogéneos, y pueden ser resultado de la activación precisa de más de un tipo de digitalizador del gesto.

El usuario tiene que dar inicio de manera inequívoca al reconocimiento de gesto, por lo que se debe diseñar el cambio de modo.

El diseño de la notificación al usuario debe tomar en cuenta que así como el gesto no tiene una posición fija en el espacio, el modo de retroalimentación debe considerar que el usuario se encuentra en un espacio indeterminado, incluso en movimiento o transición entre espacios y ambientes.

Las formas de los gestos no requieren de tener un equivalente en un gesto del cuerpo humano, es decir que puede diseñarse una sintaxis a partir de patrones de conducta, y puede identificarse como gesto un patrón de modificación de las condiciones externas al usuario. Por ejemplo, una combinación de datos biométricos más datos sobre la ubicación geográfica, la hora y el día de la semana pueden considerarse como un gesto que signifique “el usuario se ejercita haciendo ciclismo” y este gesto puede ser el gatillo para ejecutar un programa de seguimiento de calorías consumidas, notificación de sugerencias de mejora, etc.

Esta interfaz suele tener un componente de aprendizaje de máquina continuo, aunque no necesariamente orientado a identificar a un usuario, sino a identificar patrones de conducta.

Lo que es considerado gesto para esta interfaz puede estar oculto al usuario como lenguaje, por lo que el usuario no tiene que aprender el uso de la interfaz, sino que la interfaz es la que adapta, aprende del usuario las variaciones que tiene que aplicar al conjunto de gestos con el que fue diseñada.

Ventajas y limitaciones:

El tiempo de aprendizaje de uso de interfaz es corto, el tipo de gestos y sus variaciones son naturales al usuario. El sistema es el que realiza el aprendizaje, más que el usuario.

Se puede usar la aplicación de la interfaz sin tener que situarse en un espacio fijo, o manipular un sistema digitalizador adicional. La computadora puede quedar oculta y liberar al usuario en el uso de sus extremidades o en la ejecución de otras tareas físicas.

Una de las principales desventajas de esta interfaz es que es la más afectada por el exceso de datos que no son parte del gesto, y el sistema de digitalización del gesto requiere un servicio de filtrado de ruido, que puede ser de alto gasto de recursos de cómputo.

Otra desventaja es que la activación del modo de reconocimiento y las notificaciones de retroalimentación tienen que tener algún tipo de conectividad con el usuario, sin importar como cambia de posición en el tiempo y espacio, por lo va a existir una dependencia entre este tipo de interfaces y alguna o varias redes de comunicación. Esta dependencia puede implicar riesgos en la privacidad del usuario, latencia en la aplicación de las instrucciones a la computadora, y errores en la transmisión de los mensajes entre hombre y máquina.

3.7 Elementos formales de la interfaz gestual con control de voz.

Antecedentes:

Las interfaces de voz han tenido un desarrollo y evolución paralelo al de las interfaces gestuales. Las mismas condiciones que han ido haciendo factible comercialmente las interfaces gestuales han hecho factibles las interfaces de voz. Y se complementan muy bien el caso específico de sobreponerse a las limitaciones de la interfaz WIMP.

Características formales:

La interfaz de voz activa o desactiva una aplicación gestual de sintaxis estricta.

La aplicación gestual generalmente va a tener una interacción más eficiente que el uso de comandos de voz o WIMP.

Los comandos de voz pueden ser modificadores del gesto, abriendo las posibilidades de que un mismo gesto pueda usarse de múltiples maneras, similar a lo que sucede con la aplicación de la combinación de teclas normales con las teclas ctrl o Alt en las interfaces WIMP

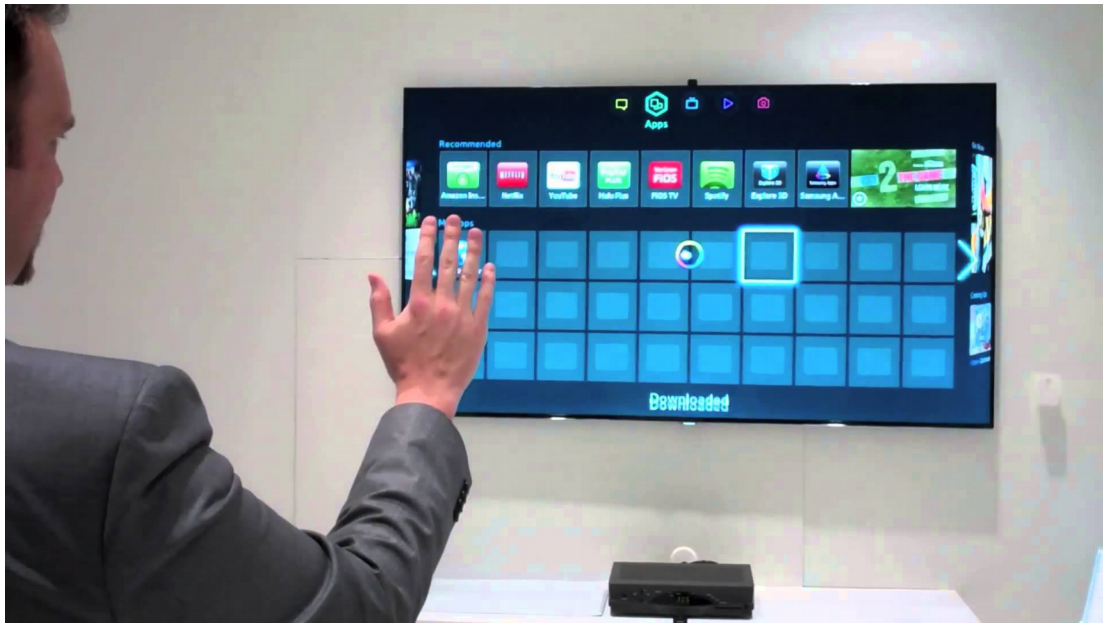


Ilustración 22: Television Samsung en venta desde 2013 tiene control de voz, reconocimiento de gestos, control remoto con TouchPad e identificación biométrica (por rasgos faciales) de los usuarios

Ventajas y desventajas:

Uno de los problemas de la interfaz gestual es delimitar el inicio y el final de la interacción por gestos. Aunque no es raro que la interacción gestual se mantenga de manera continua, si es problemático especialmente cuando se tiene interfaces que minimizan el componente visual o lo eliminan del todo.

La combinación de la interfaz gestual con una interfaz de voz se ha dado sobre todo facilitar el aspecto modal de la interfaz gestual. Mediante comandos de voz se establece la condición inicial en la que se quiere iniciar la interacción gestual, y asimismo se da por terminada la interacción gestual, asegurando que no se van a dar interacciones gestuales por casualidad.

Por último, es de notar que las inflexiones de la voz pueden ser usadas para identificar estados emocionales, información que puede modificar el reconocimiento del gesto de manera significativa, por ejemplo, el mismo gesto acompañado de voz enojada producirá un efecto distinto del gesto acompañado de voz en calma.

3.8 Elementos formales de la interfaz gestual especializada en la inclusión de personas con discapacidad.

Interfaces diseñadas para satisfacer las necesidades de comunicación con la computadora de personas con un tipo específico de discapacidad física que hace difícil o imposible el uso de interfaz gráfica con paradigma WIMP y para facilitar la inclusión mediada por computadoras o mediada por otros sistemas mecánicos controlados por computadora.

Antecedentes:

La interfaz gestual ha tenido un alto potencial de uso para integración de personas con discapacidades, ya que la interfaz de paradigma WIMP ha tenido un desarrollo muy sesgado al uso visual y manual, y al no haber sido diseñada desde el origen para la inclusión de personas con dificultades visuales o de manipulación de interfaz gráfica ha impulsado la exploración de alternativas ya sea como agregados para paliar las carencias de la la interfaz gráfica, ya sea como sustitutos completos diseñados como interfaces especializadas.

De ahí el atractivo de la interfaz gestual , ya que la mayoría de las discapacidades no impiden la generación de gestos con alguna parte del cuerpo.

Históricamente casi todos los sistemas de este tipo se han desarrollado orientados a un tipo específico de discapacidad, con *hardware* y *software* adaptado a la medida. El caso más famoso es, quizás, el del sintetizador de voz del Dr. Stephen

Hawking, que es controlado con gestos de los ojos y los párpados para convertirlos en discurso hablado.



Ilustración 23: El aparato digitalizador del gesto montado en el marco de los anteojos permite manipular una interfaz especial con gestos de las pestañas, movimiento de los ojos y algunos movimientos de la mejilla. Foto de Associated Press /Elizabeth Dalziel (2015)

Estas funciones de control de comunicación para personas con algún tipo de parálisis fueron desarrolladas desde los años setentas. En las décadas siguientes estas interfaces para control y comunicación siguieron siendo las más relevantes, y con la diversificación de los programas a las plataformas móviles se han ido dando en mayor medida las interfaces controladas por gestos como apoyo en el aprendizaje y la integración a la vida diaria de las personas de diferentes discapacidades.

Características formales:

Los elementos gestuales con los que se diseña esta interfaz requieren de una investigación especializada en el tipo de discapacidad. Los ejemplos más exitosos se derivan de la experiencia directa del diseñador con el tipo de discapacidad al que se dirigen.

Aunque a nivel experimental se han dado casi siempre soluciones a la medida de cada usuario individual, es cada vez más común encontrar interfaces que están dirigidas a un nicho, es decir que agrupan discapacidades de manera más amplia y ofrecen ajustes dentro del programa para que la interfaz se adapte a un rango de discapacidad.

Otro elemento a tomar en cuenta es el del uso de la interfaz por parte de personas que no tienen la discapacidad objetivo del diseño, pero que por apoyo al usuario discapacitado tienen que tener acceso a los programas, sea a través de la misma interfaz que el usuario discapacitado o a través de elementos redundantes con una interfaz gráfica WIMP. Una buena interfaz gestual, sin embargo, puede ser igual de accesible para el discapacitado como para el usuario sin discapacidad.

Por último, el diseño de estas interfaces puede resolver problemas de inclusión de grupos de personas con algún problema físico que aunque no se califica como discapacidad, sí puede caracterizarse como causante de problemas de accesibilidad al uso de programas y sistemas de cómputo. Grupos de tercera edad, de gente con lesiones en corto y mediano plazo, personas que pertenecen a los percentiles extremos de medidas corporales, debilidad visual moderada, etc. pueden beneficiarse de una interfaz diseñada para la inclusión de un grupo de discapacidad, al momento de hacer el diseño es posible hacer ajustes mínimos que beneficien a un grupo de personas más amplio que el exclusivo de la discapacidad estudiada.

Ventajas y desventajas

Es una interfaz que promueve la inclusión y cierra la brecha digital en grupos humanos que no siempre representan un aliciente comercial.

Es una Interfaz que puede ser útil a más personas que la población con el tipo de discapacidad al que está dirigido.

No siempre puede omitirse algún tipo de conversión a una equivalencia de control de interfaz gráfica WIMP, ya que en muchos casos es una interfaz cuya finalidad es dar acceso a los programas de computadora diseñados para la interfaz gráfica WIMP.

En algunos casos es una interfaz que requiere de sistemas expertos para ejecución de tareas muy específicas, como sintetizar discurso o interactuar con aparatos domésticos, lo que requiere de *hardware* y *software* a la medida que puede ser costoso o de complicada factibilidad técnica.

Para resumir esta taxonomía, se ha expuesto como se integran los elementos necesarios cuando se usa el gesto como medio de comunicación hombre-máquina, estableciendo la base de esta clasificación en el nivel del gesto en la jerarquía de la interacción y el nivel jerárquico de los otros elementos de la interfaz que tienen una función enfocada en la parte tecnológica: la parte de *hardware* que convierte el gesto en datos, la parte de *software* que convierte los datos en unidades semánticas y las interpreta, y el elemento de retroalimentación que puede a su vez estar basado o no en gestos. Son cuatro elementos que el diseñador de interfaz maneja para establecer la relación de unos con otros en el diseño de una nueva interfaz, para establecer las especificaciones de los elementos que no son gestuales o para indicar cuales son los objetivos a alcanzar a nivel de experiencia de comunicación humana.

Esta tipología también es aplicable en el caso de análisis para evaluación de una interfaz existente, y enfocar los esfuerzos de mejora o rediseño de interfaz para por ejemplo, la actualización para aprovechar avances tecnológicos. Es importante reconocer que este tipo de interfaces agregan una mejor interacción, con ventajas sobre las interfaces puramente gráficas, pero que son un escalón más en el avance hacia una interfaz que refleje fielmente las capacidades humanas.

3.9 Casos de estudio.

Para ejemplificar la utilidad de esta taxonomía para el diseñador de interfaces, se presentaran algunos casos de estudio que parten de ejercicios profesionales y académicos y se hará notar cómo se usó el conocimiento analítico derivado de la tipología para, según sea el caso, iniciar el diseño de la interfaz, modificar interfaces existentes, o analizar las fortalezas y debilidades de una interfaz propuesta por otro diseñador.

3.9.1 Reconocimiento de caracteres “Graffiti”

Este es un ejemplo de interfaz WIMP que incluye interfaz gestual. Es un sistema de reconocimiento de caracteres que se comercializó con la marca “*Graffiti*”. Este sistema fue muy exitoso ya que era parte de los asistentes digitales personales, PDA (por sus siglas en inglés; *Personal Digital Assistant*), que representan el primer tipo de computadora personal móvil, antecedente de los teléfonos inteligentes, y cuyo éxito comercial hizo que *Graffiti* pueda considerarse cómo el primer sistema de gestos usado de forma masiva; este sistema se integra sin conflictos con cualquier sistema WIMP existente, su función equivalente en la interfaz WIMP es el de ingreso de texto. La ventaja sobre el equivalente WIMP de que proveía un sistema más ergonómico que un teclado en aparatos que no tenían el espacio físico para un teclado completo, este es uno de los problemas recurrentes de las interfaces que se diseñan en el paradigma WIMP para aparatos móviles con limitados recursos de despliegue, como los susodichos PDAs y otros que se vislumbraban a futuro en ese momento, por ejemplo lectores de libros electrónicos, visores multimedia portátiles o de bolsillo, navegadores web integrados a la vestimenta, consolas de juego portátiles conectadas a Internet, etc.

Graffiti es un sistema de reconocimiento de gestos muy específico, funciona sólo como entrada de texto, es decir, substituye las funciones del teclado, y requiere de un sistema de digitalización de textos que fuera capaz de reconocer “grafías”. Patentado por Jeff Hawkins, fundador de las compañías Palm Inc. y

Handspring Inc. dedicadas al desarrollo de PDAs. En la patente US 6295372 B1 se puede leer que un PDA es un sistema de cómputo para sostenerse en la mano y con entrada de texto basada en una pluma, y que el *software* de reconocimiento del gesto *Graffiti* requiere que se separe un área exclusiva para la actividad de la pluma. Esto puede traducirse en varias configuraciones, en una de ellas el área dedicada es un digitalizador que no es parte de la pantalla, sino que es un elemento auxiliar a un display de tipo panel plano. El usuario ingresa grafías en esta área, imitando la escritura a mano. Otra configuración usa un digitalizador virtual, generado por *software* en un área de la pantalla, de la misma forma a como se genera un teclado virtual en un teléfono inteligente.

La integración de *Graffiti* es, por lo tanto, dependiente de un digitalizador dedicado pero este digitalizador puede ser emulado en una caja de diálogo o ventana emergente del sistema, de ahí su clasificación dentro de la tipología propuesta como un agregado de interfaz gestual a un sistema WIMP.

El sistema funciona con reconocimiento de cada carácter, mas algunos gestos especiales para modificar la entrada, estos gestos especiales son el equivalente de las teclas alt-ctrl-Fn en un teclado normal.

Se derivó de intentos de desarrollar un sistema de reconocimiento de letra manuscrita, pues los desarrolladores del equipo de Hawkins provenían del equipo que estaba haciendo el reconocimiento de letra manuscrita en el PDA *Newton* de Apple. Este equipo sabia que uno de los problemas principales del *Newton* es el de tener un área muy limitada de ingreso de texto, y por lo tanto no se puede ingresar mas de dos o tres palabras antes de llegar al margen de la pantalla. Una solución prometedora fue la de reconocer una letra a la vez, en un área pequeña en la que cada letra se escribe sobrepuesta a la anterior. El sistema se completó cuando vieron que se podía establecer un sistema de gestos preestablecidos, de manera que las variaciones en el gesto fueran fácilmente discriminadas de la letra que representaban. Este alfabeto de grafías es muy similar a los sistemas de taquigrafía desarrollados y popularizados desde el siglo 18. Hawkin dice, en el libro “Piloting

Palm” [BUTTER y POGUE, 2002] que el punto de separación de *Graffiti* de los sistemas de reconocimiento de texto que se habían desarrollado antes vino cuando se dieron cuenta de que en vez de desarrollar un sistema de inteligencia artificial que aprendiera a reconocer la letra manuscrita de un usuario, podían hacer un sistema algorítmico basado en un solo tipo de gestos, y hacer que el usuario aprendiera a hacer dichos gestos. Después de todo, dice Hawkin en el libro citado, el usuario también tiene que aprender a escribir en un teclado, y aprender un conjunto de gestos no debe llevarle más tiempo. De la misma manera que un usuario con más práctica puede usar todos los dedos de las dos manos y concentrarse en la pantalla, el usuario de *Graffiti* puede, con la práctica, ingresar texto de manera natural, incluyendo números y caracteres especiales, sin necesidad de mantener la vista en el área de digitalización.

El algoritmo de *Graffiti* está protegido por patentes, por lo que su uso requiere pago de licencia. Sin embargo sirvió de inspiración a otros sistemas gestuales que entendieron que la capacidad de las pantallas de tener entrada táctil iba a ser más generalizada a partir de la aparición de los teléfonos inteligentes y las computadoras ultra-portátiles genéricamente conocidas como “tablet”.

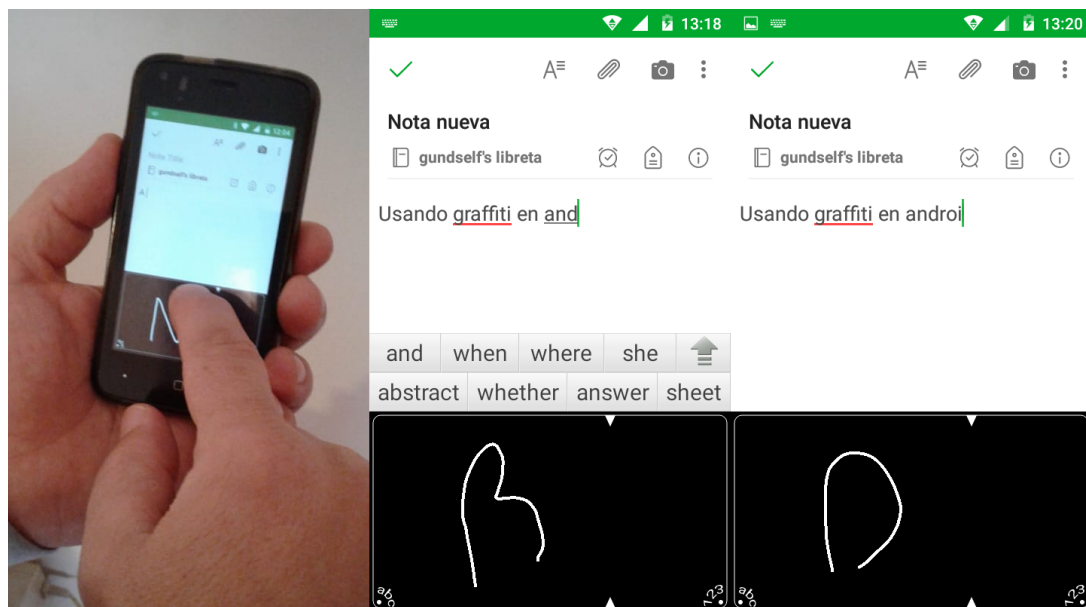


Ilustración 16: Izq: Graffiti como app del SO Android, impulsado por la compañía Access Co a partir del año 2010

Graffiti estuvo muy cercanamente ligado a la suerte de la plataforma Palm de *hardware*, por lo que cuando ésta quedó relegada del mercado lo mismo pasó con este sistema. Los usuarios estaban bastante cómodos con los teclados en pantalla y las mejoras exponenciales en la velocidad y capacidad de los procesadores en los aparatos portátiles, así como la posibilidad de la computación en la nube han hecho que el reconocimiento de voz como alternativa al teclado sea más factible que el ingreso textual, e incluso han puesto cada vez más al alcance el reconocimiento de letra manuscrita basado en inteligencia artificial y aprendizaje de máquina. Sin embargo *Graffiti* sigue activo, como lo demuestra la existencia de la aplicación para sistema Android y sistema iOS desarrollada por la misma compañía que hizo el sistema original.

3.9.2 Menú de restaurante proyectado sobre mesa



Ilustración 24: Demostración del sistema de menú virtual. El proyector está completamente cenital a la mesa

Este es un ejemplo de de interfaz gestual que incluye WIMP. La interfaz propuesta fue diseñada como asesoría académica de un proyecto de tesis para la Maestría de diseño multimedia de la Universidad del Azuay. El postulante fue el Diseñador Mathias Villacís y el tema a desarrollar era el de control mediante gestos de

aplicaciones para el sector de servicios turísticos, como facilitador de la comunicación con turistas extranjeros no hispanohablantes. Se dirigió este proyecto especialmente para conseguir que la interfaz gestual sustituyera a la interfaz WIMP por sus ventajas de comunicación intercultural.

Otro objetivo de ésta tesis era diseñar gestos que fueran de significado neutro a la cultura de origen del usuario, ya que el grupo de usuarios al que iba dirigido era el de turistas de lengua distinta al español. Los gestos que se diseñaron se basaron en el supuesto de que ciertos gestos como apuntar para elegir, girar para examinar, hacer a un lado para descartar, etc son comunes las experiencias humanas sin importar la cultura de origen.

Las conclusiones de esa tesis se apoyaron con un demo de una aplicación para substituir un el menú de un restaurante por un sistema gestual, que se proyecta sobre la mesa y los comensales pueden revisar lo que se ofrece e ir armando su orden. La comanda para la cocina se genera automáticamente una vez que los comensales terminan de elegir. Se hizo una programación básica de reconocimiento de gestos, por limitaciones de tiempo y presupuesto se hizo el reconocimiento usando un lápiz electrónico que era más exacto en el reconocimiento de movimientos y posición sobre la mesa, y que tenía un botón en la punta. Las mismas limitaciones de tiempo y presupuesto hicieron que no se rompiera totalmente con la interfaz WIMP, los pasos finales de elección de menú requerían confirmación en cajas de diálogo. Sin embargo el sistema era capaz de reconocer 9 gestos con una tolerancia amplia que acomodaba a una gran mayoría de dimensiones humanas.

Otro aspecto interesante de este demo es el que los gestos se programaron con la punta del lápiz electrónico como punto de reconocimiento relativo, lo que permitía usarlo desde cualquier orientación alrededor de la mesa, y que el modo de “abrir el menú” era en realidad un instructivo de uso, diseñado para que el mesero demostrara los gestos básicos y el modo de uso de manera práctica.

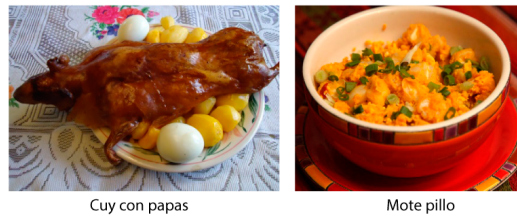
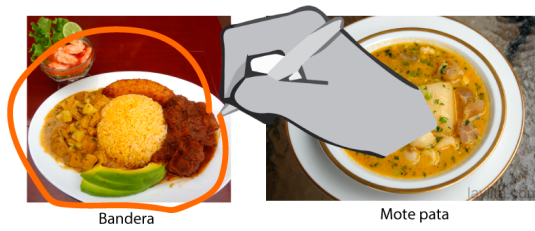


Ilustración 25: Tres de los gestos principales en la elección del menú. De arriba a abajo: ampliar foto, selección de platillo (el movimiento inverso quita el platillo seleccionado de la lista de comanda), quitar foto y volver a categoría.



El diseño de interfaz se benefició del análisis previo y el enmarcado en las características de un tipo de menú gestual, pues los recursos limitados para el desarrollo pudieron aprovecharse más efectivamente. Este enfoque puede usarse hoy día con las aplicaciones de realidad aumentada, ya que seleccionar un conjunto de gestos ahorra espacio en pantalla y hace más eficiente el proceso de reconocimiento de ubicación del entorno en el que se va a insertar el objeto virtual.

Una característica del *software* desarrollado es que tenía bastante tolerancia a las variaciones de los gestos diseñados, y un pequeño subsistema para apoyar, en caso de una desviación muy grande de la norma, antes de que la frustración pudiera anular el avance. Estas dos características se apoyaban fuertemente en interfaz WIMP, que se activaba una vez detectado que había un impedimento en el sistema de reconocimiento del gesto, ya sea por tiempos de respuesta largos o por demasiado ruido en el sistema (es decir, el sistema sabía que había entrada de gestos pero no lograba empatar ninguno con un comando preciso). La familiaridad de WIMP en un momento así prevenía que el comensal desistiera de su intento de elegir en el menú por frustración.

El demo se probó de manera limitada con un universo de 30 voluntarios, principalmente angloparlantes, y hubo la satisfacción de que no se activó el sistema de respaldo y ayuda, y que la comprensión y asimilación de los gestos de control era muy acelerada, sin mucha diferencia por la edad o procedencia del usuario.

En esta aplicación se hicieron patentes algunos de los problemas de ergonomía que son exclusivos de una interfaz gestual: Oclusión de la sombra de la mano que manipulaba el lápiz, oclusiones accidentales producidas por el entorno, la necesidad de miniaturizar los elementos de proyección, (solucionar esa necesidad quedaba fuera del alcance de esa tesis) y la imposibilidad de competir con costo/beneficio con la interacción directa con el mesero.

Pudimos también comprobar que la barrera de comunicación cultural puede ser menos alta cuando la interacción es a través de gestos, puesto que agregaba un elemento lúdico y socializante con resultados más inmediatos al problema de comunicarse con alguien que no habla la lengua propia. Sin embargo, el riesgo de inadvertidamente incluir gestos de un significado erróneo, agresivo o descortés para alguna cultura quedó latente. Este demo fue muy positivo al permitir vislumbrar, al menos en el sector de los turistas en una cultura extraña, una cierta universalidad de experiencias, una cierta homogeneidad en el ámbito de la actividad “viaje de turismo” y de los servicios turísticos, es decir que este tipo de experiencias cada vez son más universales, menos dependientes de una región culturalmente, y se pudo verificar que el gesto participa y evoluciona con esa universalidad. [VILLACÍS, M. 2013]

3.9.3 Actuación en escenario con proyecciones interactivas controladas por el actor

Este es un ejemplo de interfaz gestual con lenguaje de gestos estricto. El diseño de esta interfaz se derivó de asesorías para algunas actividades de danza y teatro en la Universidad de Cuenca (Universidad pública del Ecuador). Estas asesorías se derivaron del diseño de animaciones interactivas para ser proyectadas como escenografía de la Opereta “El Elixir de amor” (*L’elisir d’amore* – G. Donizetti, (1832)). La proyección de las animaciones en la opereta era controlada de manera tradicional, por activación de un operador, similar a como se controlan las luces de teatro, pero a algunos actores participantes les interesó la posibilidad de realizar un performance teatral o de danza donde dichas animaciones pudieran ser controladas por el actor en el escenario. La posibilidad de que el encargado de sincronizar los eventos del proyector fuera el actor mismo en el escenario puede ampliarse a la de tener cierto número de videos o animaciones cuyos ritmos y mensajes pudieran combinarse de manera rítmica y armónica, incluso fomentando un poco de improvisación a manera de los riffs de jazz y rock.

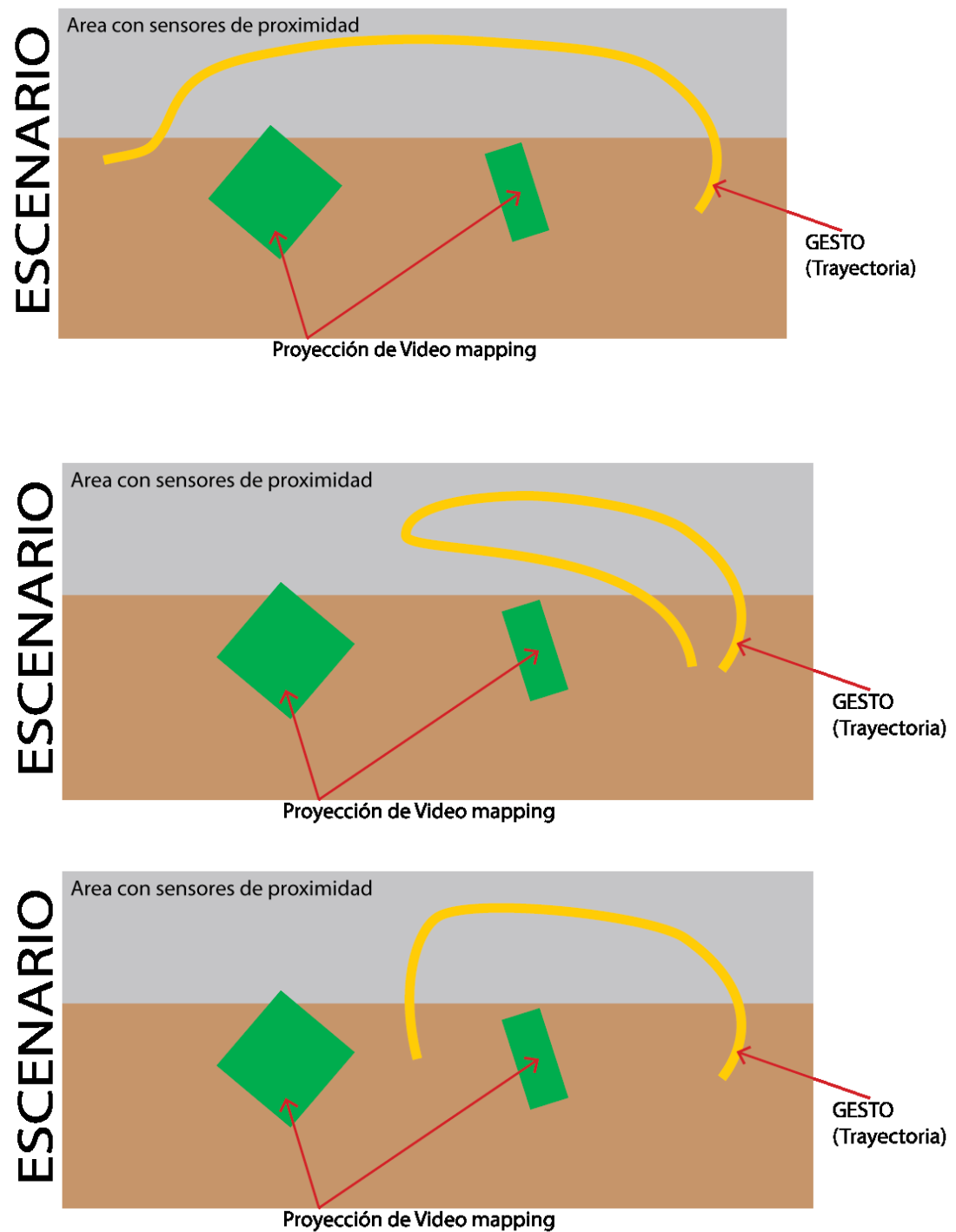


Ilustración 26: Sintaxis de los gestos de "Cuerpo-viento". Arriba: ingresar y salir por puntos opuestos del escenario: cambiar a segmento de video siguiente. Enmedio: Ingresar y salir por el mismo punto lateral: generar audio en ritmo más rápido siguiente. Abajo: Ingresar por punto lateral, salir por el centro: generar audio en ritmo mas lento siguiente.

Una de las interesadas fué Rita G. Rodríguez, estudiante de la carrera de danza que en su tesis de licenciatura realizó una integración de un sistema de reconocimiento de gestos con coreografía y música generativa. [RODRÍGUEZ 2013] Los profesores de la facultad Ernesto Santos y Daniel López proveyeron de la programación del *software* interactivo y la música, respectivamente. La asesoría provista a Santos y López consistió en la definición de cuales elementos de la coreografía se iban a considerar como gesto, es decir, en la definición del lenguaje estricto que efectivamente es un subconjunto de un lenguaje expresivo más amplio cuyos elementos son los pasos de baile y la partitura musical.

En la ejecución de la pieza, la bailarina va alterando la ejecución de la música con ciertos gestos, que pueden cambiar el algoritmo generativo del ritmo (acelerándolo, atrasándolo, reiniciándolo) y la armonía (cambiando la semilla de generación de los sonidos) y al mismo tiempo esos controles desencadenan cambios en el sistema de luces escénicas. El reconocimiento de gestos se simplificó al máximo con 2 tipos de entrada: una videocámara que podía reconocer 6 gestos, y un sensor de proximidad que podía reconocer si el cuerpo de la bailarina estaba en el frente o en la parte trasera del escenario.

El uso de gestos con un código estricto se apoyaba en el análisis algunas características del proyecto. Al tratarse de una bailarina profesional, se podía contar con una exactitud en la forma del gesto. Asimismo, el sistema iba a estar entrenado por un solo usuario. El diseño coreográfico se vuelve a su vez diseño de interfaz. Se eliminó la instalación de actuadores, típicamente cableados en el escenario, apoyando un concepto de escenografía minimalista cuya variación e impacto procedían del video mapping proyectado sobre los elementos. La propuesta de escenografía, video mapping y posibilidad de improvisación iban de acuerdo a la técnica propuesta en la tesis, conocida como Flying Low, por su enfoque hacia el suelo y a la fluidez en relación con el suelo, y Passing-Through, que es un sistema de improvisación que requiere que el bailarín esté muy pendiente del entorno.



Ilustración 27: Durante la ejecución de Cuerpo Viento. Foto: A. Jaramillo

Un componente interesante en esta interfaz es el del tiempo, marcado por compases y ritmos, que tuvieron un papel más preponderante que en otros diseños de interfaz gestual. Mientras que otros diseños se piensan con tolerancias amplias para el tiempo de ejecución del gesto, y tratando de reconocer con el mínimo tiempo posible para así poder construir un lenguaje de gestos veloz, en este caso se apoyó el sistema de reconocimiento con el hecho de que se podía consultar el sistema de control musical sobre el compás y ritmo, y de esa manera establecer el reconocimiento de gesto en ese marco de tiempo justo.

Por último, el elemento improvisatorio representó un problema para el sistema de ejecución y una limitante para el diseño de los gestos a reconocer, ya que había que diferenciar cuando los gestos se estaban usando como parte de la coreografía pero no con la intención de transmitir una orden al sistema. No era válido li-

mitar la ejecución de los gestos que eran parte de la interfaz por separado de los movimientos de la coreografía, de ahí que se haya tomado la decisión de establecer una “zona de actuación” separada de una “zona de control”. A eso hubo que agregar el problema técnico de latencia en el tiempo entre el reconocimiento del gesto y la activación del efecto deseado. De ahí que se tomara el sensor de proximidad como gatillo del sistema gestual. La artista podía realizar todo tipo de movimiento en la parte delantera del escenario, pero al momento que entraba en la parte trasera el sistema iniciaba la tarea de reconocer comandos y se estableció que una vez que reconociera un gesto se desactivaba la interfaz gestual hasta que se hiciera la secuencia de ir al frente del escenario y regresar al área trasera, lo que puede considerarse como un gesto por si mismo. De esta manera se podía tener una confirmación positiva de que los cambios iban a realizarse, y la bailarina podía realizar su improvisación sin preocuparse de que se efectuaran cambios en la música y el escenario que no fueran los deseados.

3.9.4. Robot que aprende comandos de control

Este es un ejemplo de Interfaz gestual con sintaxis flexible. La interfaz está integrada a un juguete robótico desarrollado por la compañía Sphero Ltd. (China) en la marca WowWee, con el nombre comercial de Roboreptile. Está a la venta en todo el mundo desde el año 2008. Aunque se trata de un juguete, es un buen ejemplo de como se diseña una interfaz gestual para interactuar con un *software* que puede incluir un aprendizaje con inteligencia artificial, y donde una interfaz gráfica no es aplicable. Es de esperar que los robots y los aparatos con capacidad robótica sean cada vez más comunes, producto de la miniaturización electrónica y el abaratamiento de los componentes de cómputo más poderosos y componentes mecánicos más ligeros y resistentes. Las aplicaciones de estos robots cubren un amplio rango, van de juguetes a verdaderos robots antropomorfos (androides) para el cuidado de personas de tercera edad o con discapacidad.



Ilustración 28: Roboreptile. Los sensores infrarrojos están situados en la cabeza, parecen fosas nasales, y en la cabeza también está situado el sensor auditivo.

El desarrollo robótico ha impulsado muchos de los avances en el campo de la inteligencia artificial que es necesaria no sólo para el reconocimiento de los gestos de control sino también para el reconocimiento del entorno y de las situaciones en las que se requiere activar un comportamiento específico. En ellos la interfaz es como una plantilla en blanco en la cual el sistema de aprendizaje va grabando y modificando las asociaciones del gesto reconocido con el tipo de acciones que el *software* puede realizar.

En el ejemplo del roboreptile, que está modelado a manera de un reptil bípedo, su conjunto de sensores y su programación básica le permiten aprender a partir del sonido y refinar su comportamiento para que el usuario pueda refinar el aprendizaje del robot y enseñarle “trucos”, que son respuestas gestuales que simulan “bailar”, “sentarse”, y “hablar”.

La interfaz esta diseñada para que el roboreptile identifique cercanía o lejanía del usuario usando dos sensores infrarrojos, activación de ciertos modos por un sensor de toque en el lomo, y señales auditivas por vía de un micrófono ambiental.

Una vez que se enciende por primera vez el juguete, el usuario ejecuta una secuencia de calibración que también le va a servir para mnemotecnia de los

gestos básicos. Cuando el roboreptile esta “activo”, se le llama la atención con un ruido fuerte (chiflido o palmada), el robot se dirige hacia la fuente del sonido y si detecta una fuente de calor cercana a su cabeza entra en modo de espera de un comando. Si no logra identificar el gesto después de un periodo de tiempo, regresa a modo de exploración libre, se dedica a caminar usando los sensores infrarrojos para evitar los obstáculos en el espacio.

Si logra identificar el gesto, se desencadenará una secuencia de respuesta gestual: sentarse, comer, bailar, etc. Si el usuario toca el lomo del robot una vez que termina la secuencia, es un refuerzo de que el reconocimiento del gesto fue correcto. El usuario también puede tocar el lomo del robot antes de que termine la secuencia para “castigar” o indicar que el reconocimiento del gesto fue incorrecto.

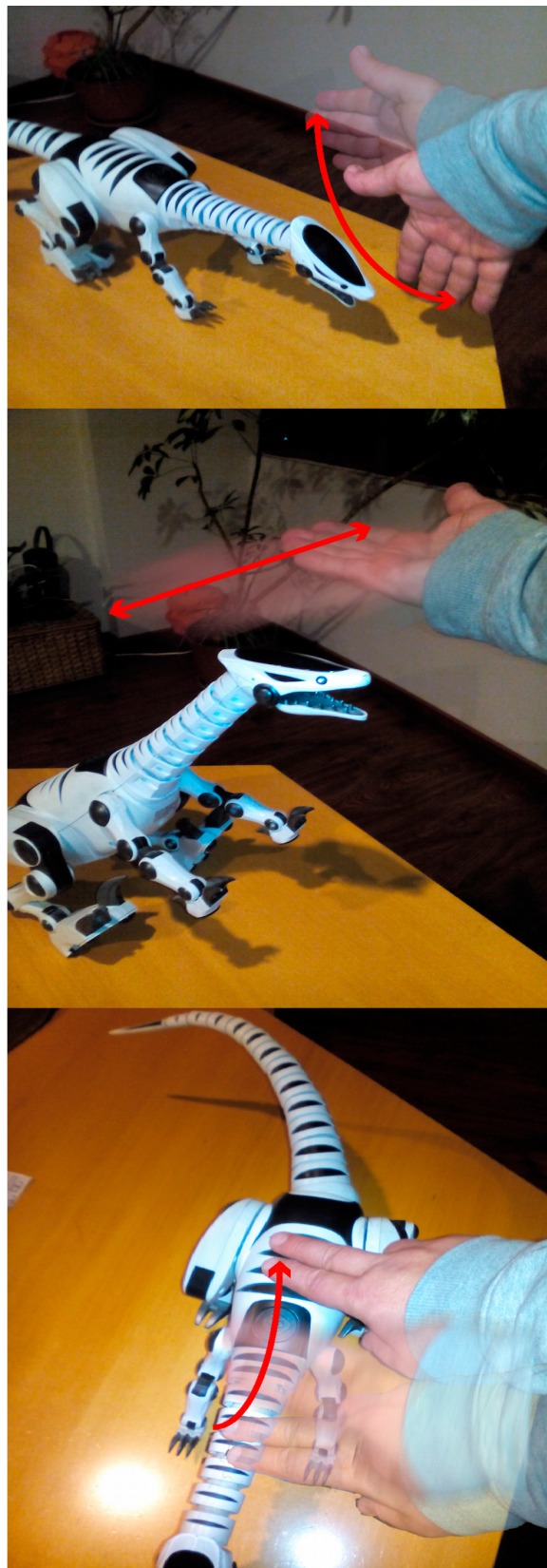
Una de las limitantes de esta interfaz es que tiene que filtrar las condiciones ambientales para determinar si está recibiendo o no un comando, y es un proceso que puede tomarle de 3 a 10 segundos antes de ejecutar el comando asociado. Aunque el manual hace notar este tiempo de espera para el reconocimiento de un gesto va mejorando por efecto del entrenamiento y la adaptación de los algoritmos de inteligencia artificial que usa para el reconocimiento de gestos, 3 a 10 segundos es una pausa de tiempo demasiado larga y llega a producir incertidumbre en el usuario, que tiene tiempo suficiente como para producir nuevos gestos es espera de que hay una respuesta.

Cuando el roboreptile ha fallado en reconocer gestos un numero de veces, se activa una secuencia de “enojo” donde el roboreptile se aleja y actua de manera “salvaje”, autónoma sin atender llamadas por un tiempo corto (3 minutos) .

La falta de sensibilidad o dificultad de filtrado es reconocida por el fabricante en cuanto a que provee de un control remoto donde los gestos pueden ejecutarse como secuencias de teclas y de un “capuchón” que permite cubrir los sensores, con lo cual el robot entra en modo de “dormir”, que es la manera de tranquilizar en caso de que se haya activado la secuencia “salvaje” .

Ilustración 29:

Vocabulario de gestos del roboreptile. De arriba a abajo: Gesto lateral. Gesto superior. Gesto de recompensa (tocando el sensor en el lomo)



3.9.4 **Software de enseñanza del lenguaje de señas ecuatoriano**

Este es un ejemplo de Interfaz gestual especializada en inclusión de personas con discapacidad. Esta interfaz se diseñó en el año 2013 como parte de la asesoría para la tesis de Maestría en Diseño Multimedia, del Diseñador Diego Larriva. [LARRIVA 2013] El objetivo de la tesis era el de verificar como las nuevas tecnología podían mejorar la inclusión de personas con discapacidad. Como caso particular se escogió el aprendizaje de la lengua de señas para las personas sordomudas o para las personas que necesitan comunicarse con las personas sordomudas.

El programa final contó con la asesoría adicional de expertos en el lenguaje de señas y con maestros especializados en los problemas de inclusión en el campo de la sociología y la educación.

La parte de interfaz gestual se diseñó para darle al programa la posibilidad de evaluar, calificar y corregir los ejercicios de aprendizaje, usando la cámara y el reconocimiento de formas. Esta manera es similar a la de los sistemas de aprendizaje de lenguas extranjeras con capacidad de reconocer voz y evaluar una correcta pronunciación.

Adicionalmente se estableció una parte de la interfaz gestual para el control de ingreso al programa, para la navegación dentro de las secciones y lecciones, y para el proceso de finalizar la sesión. El diseño visual tomaba en cuenta estos gestos para enfatizar la comodidad y efectividad de uso de esta ayuda educativa.

Un valor adicional que se tomó en cuenta es que este programa era igualmente efectivo para el aprendizaje de lengua de señas en usuarios sin discapacidad vocal. Un objetivo del diseño gráfico y de la interfaz gestual era que el sistema gestual fuera más atractivo que el uso de la interfaz WIMP precisamente para atraer a las personas que no tienen la necesidad absoluta de usar lengua de señas. La interfaz gestual debería ser más atractiva y cómoda que la interfaz usada con el

mouse y así dar un incentivo extra a las personas con todas las capacidades para que no sientan el proceso de aprendizaje como engorroso o difícil.



Ilustración 30: Algunas pantallas del programa de Aprendizaje del lenguaje de gestos. La niña hacía frases relativas al entorno y el alumno repetía esas frases frente a la cámara de la computadora para ser evaluado.

El prototipo de esta interfaz no llegó a estar 100% funcional, por limitaciones técnicas, de tiempo y presupuesto. El programa final que se realizó implementó los gestos de navegación que debían ser reconocidos por la cámara como gestos de TouchPad y no alcanzó un prototipo funcional del sistema de evaluación basado en el reconocimiento visual de los ejercicios de los estudiantes. Quedó documentado, sin embargo, los beneficios que dicho sistema autónomo de evaluación puede tener en la efectividad de aprendizaje mediante computadora.

En la investigación de esta tesis una de las especialistas en inclusión y educación de sordomudos, la Profesora Matilde Renjfo, señalaba que los lenguajes de señas del mundo son mucho más parecidos entre sí que los lenguajes hablados, así que aprender a hablar, por ejemplo, japonés de señas, es más fácil si ya se sabe español de señas. Estimular el aprendizaje del lenguaje de señas en personas que no son discapacitadas puede ser, finalmente, un facilitador de inclusión entre gente de diferentes culturas. Fue sorprendente que en este caso la herramienta para facilitar inclusión de los mudos tenga más que ver con la inclusión de las personas sin discapacidad en el mundo de las personas con discapacidad, que a la inversa.

3.10 Diseño de interfaz gestual a través del análisis formal propuesto.

La aplicación de esta taxonomía tiene una visión pragmática en cuanto a la utilización del análisis formal para la generación de nuevos diseños. La elección de un vocabulario gestual es equiparable a la elección de un sistema simbólico visual o gráfico, y si en el caso de la interfaz gráfica los aspectos subyacentes de tecnología como cual es la capacidad de despliegue de color, cual es el área real de interacción en píxeles, resolución, etc en el caso de la interfaz gestual estos aspectos subyacentes tienen una gama aun en su infancia, aun en proceso de evolución y mejora, pero que el diseñador de interfaz ya puede echar mano e influir en la elección. En este caso se habla de elementos como la velocidad de respuesta del algoritmo de interpretación de gestos, resolución de la digitalización en volumen de datos generado por ciclo al digitalizar el gesto, etc.

Las elecciones que realice el diseñador de interfaz van a determinar la amplitud del vocabulario de gestos, y la libertad del usuario final para modificar e individualizar ese vocabulario. Aspectos tecnológicos que pueden parecer muy rudimentarios en este momento pueden determinar una interfaz gestual efectiva precisamente por la manera creativa como el diseñador resuelve esas limitaciones. El trabajo de composición de los elementos ya no sólo implica el uso de elementos visuales, sino de elementos de comunicación, y con ello la interfaz empieza a trascender el aspecto de elemento de control para acercarse más a un elemento emocional y social.

Conclusiones

Al considerar las interfaces gestuales desde el punto de vista de la comunicación, y entender el gesto como unidad comunicativa, se puede comprender y delimitar el proceso de diseño orientado a este tipo de interfaz.

El análisis realizado permite visualizar las posibilidades del gesto en la interfaz, entre ellas destaca que el gesto puede utilizarse de manera fragmentada, diseñando una equivalencia de un gesto a un comando, para comunicarse con la computadora de una manera muy similar a la realizada por la interfaz gráfica, y también se puede diseñar un conjunto de acciones que al ser ejecutadas y reconocidas como construcciones sintácticas más completas puedan completar una comunicación más significativa, más parecida a un lenguaje oral.

El conjunto de elementos que conforman una interfaz gráfica se amplía y modifica al considerar el gesto. La inclusión de elementos para digitalización y reconocimiento de gestos abre la posibilidad de usar el mismo circuito de comunicación hombre-máquina para la comunicación de retroalimentación máquina-hombre. No se ha explotado este camino en las interfaces gestuales más comunes, pero es una posibilidad a futuro, y en el ejercicio de diseñar una interfaz gestual no puede descartarse esta posibilidad, más bien debería reforzarse en vistas de conseguir un mejor reflejo en la máquina de las capacidades del ser humano. Las

interfaces más experimentales a la vanguardia de esta tendencia son las que usan cara con emociones, actitudes corporales, y reacciones gestuales en las máquinas, con lo que se comprueba que la creación de lazos emocionales entre usuario-programa puede hacer más efectivas algunas aplicaciones. En este campo las aplicaciones que es posible aparezcan más pronto en el mercado son las que cumplen la función de monitor médico, educación de infantes, y acceso a la tecnología de grupos con discapacidades físicas o cognitivas.

La revisión de las interfaces existentes y de los elementos comunes que se presentan en ellas posibilitó el planteamiento de una tipología que describe cada tipo de interfaz según la importancia del gesto como método de comunicación. Se reconocieron dos tipos de interfaz gestual en los que se da una integración mixta con interfaces gráficas. Dos tipos de interfaz que usan exclusivamente gestos como método de comunicación, un tipo de interfaz que se complementa con la interfaz de reconocimiento de voz, y un tipo de interfaz gestual que suple las carencias de la interfaz gráfica cuando se refiere a usuarios con algún tipo de discapacidad para los cuales la interacción gestual es más adecuada que la interacción gráfica.

Esta tipología se propone a los diseñadores de interfaz o de experiencias interactivas como de punto de partida para nuevos diseños o para analizar los sistemas existentes y entender las fortalezas y debilidades con lo cual se pueda ejecutar mejoras rediseñando la interfaz.

La comprensión de cuáles son las características de una interfaz lleva a la más efectiva propuesta de diseño, incluso a la realización de menos prototipos y a acortar los tiempos de desarrollo. Esto puede ser especialmente útil para proyectos en que el mayor trabajo se realiza en el espacio multidisciplinario fuera del diseño de interfaz, en áreas como el desarrollo de *software* o la ingeniería de sistemas. Saber cómo se va a diseñar la interfaz o dentro de qué tipo de interfaz se va a realizar el diseño de la experiencia con el usuario es una retroalimentación valiosa al principio del desarrollo del proyecto.

Las interfaces gestuales están alcanzando un punto de madurez, no tan rápido como algunos de los investigadores de experiencia de usuario predecían, pero definitivamente apoyada por la popularización de aparatos como los video juegos y los teléfonos celulares inteligentes. Esta tesis ha hecho evidente cuales son las posibilidades del diseñador en su aportación al futuro del campo, y llama la atención hacia la posibilidad de usar el gesto y el discurso que se construye con él como un sistema que además de utilidad práctica puede tener calidad expresiva y estética.

Las interfaces donde la interacción gestual recibe también retroalimentación gestual tienden a producir empatía y reacciones emocionales. Esta cualidad es, de acuerdo a la propuesta de los tres espejos de interacción de B. Buxton, más humana y más efectiva en cuanto a tener interfaces que sean extensiones del ser humano en lo físico, cognitivo y social.

Las interfaces gestuales no implican el uso de todos los sentidos como sería el ideal holístico planteado por Buxton, pero si son un progreso con respecto a las interfaces gráficas en cuanto a que involucran mayores capacidades humanas y dirigen las interacciones de un modo más sensorial. Una interfaz que aporta una experiencia más humanizada que las que se tienen hasta la fecha.

Esta tesis termina con una invitación a continuar la crítica y discusión de este tipo de interfaces desde el punto de vista de los diseñadores, los especialistas en interacción y comunicación hombre máquina, en los que la disciplina del diseño se ha visto involucrada por la necesidad de convertir esta comunicación en una experiencia más emocional y expresiva, y a los que aporta desde una especial sensibilidad y enfoque.

Cuenca, Ecuador – Diciembre de 2018

Índice de ilustraciones

Ilustración 1: Comparación del Diagrama de Jakobson con los procesos de comunicación a través de una interfaz.....	10
Ilustración 2: Esquema de operación y control de un proceso industrial.....	11
Ilustración 3: Izquierda - Ivan Sutherland con SketchPad,1962. Derecha - Douglas Engelbart con <i>On-Line</i> , 1968.....	15
Ilustración 4: WIMP: <i>Windows, Icons, Menus and Pointers</i>	18
Ilustración 5: Procesador de textos con todas las herramientas abiertas. Lo inadecuado de WIMP para comunicar múltiples dimensiones se hace evidente.....	22
Ilustración 6: Un Teléfono “inteligente” en realidad es una computadora con muchas extensiones de los sentidos humanos, pero con poco espacio para una interfaz WIMP. Las pantallas son físicamente pequeñas ya que se espera que quepan en una mano y sean muy portátiles.....	27
Ilustración 7: Ejemplo de identificación de gestos marcadores en los estudios de proxémica de Birdwhistell.....	39
Ilustración 8: Cualquiera de estos tres gráficos está asociado como símbolo de "internet", siendo que en su origen son símbolos de programas o compañías que hacen programas para usar internet. Izq a der: Logotipo de Mozilla Firefox (ver57), logotipo de Microsoft Edge (ver44), logotipo de Google Chrome (ver71).....	42
Ilustración 9: Gesto "pistola". Foto de Gonzalo León (2018).....	43
Ilustración 10: Arriba, izq: Cartel de silencio (G. León, 2017). Arriba, der: Señal de silencio, norma ISO 7010:2007 Abajo: secuencia de la frase "se guarda silencio" del Lenguaje de señas americano (ASL, por sus siglas en inglés). Sólo el dedo sobre los labios, en el ASL, se usa para ordenarle a alguien que esté sin hacer ruido. La secuencia completa es una frase descriptiva o imperativa según el contexto.....	45
Ilustración 11: Taxonomía del gesto de Nespoulous y Lecours. Fuente:[QUEK 1994]	46

Ilustración 12: Plantilla de tableta digitalizadora de AutoCAD. En el área en blanco cerca del centro se tenía el área de dibujo, el rectángulo superior era el sitio donde algunos programadores empezaron a experimentar con el reconocimiento de gestos.....	50
Ilustración 13: Izquierda: PDA marca Handspring. Derecha: los gestos del sistema Graffiti.....	52
Ilustración 14: El signo aceptado para "OK" (lado izquierdo) puede ser fácilmente confundido con el gesto grosero de la derecha, pero solo en el contexto cultural de México.....	56
Ilustración 15: Text Rain, (1999).....	57
Ilustración 16: Light Rain, (2008), WOWLab.....	57
Ilustración 17: Gestos multitouch, que son el estándar de facto desde su introducción en 2007 en los productos de Apple.....	59
Ilustración 18: Configuración de gestos en un sistema con interfaz Wimp, en este caso MacOS X.....	75
Ilustración 19: Imagen publicitaria del juego de entrenamiento y cultura física "active" de la empresa Electronic Arts. A la interacción del wiimote se agregaba una banda sujeta en la pierna con monitor de ritmo cardíaco, temperatura corporal y presión arterial, con el objetivo de verificar si efectivamente se estaba ejercitando con el programa.....	81
Ilustración 20: Manipulación de modelos de proteínas mediante el <i>software</i> ProteinShop y dos guantes datagloves. ProteinShop tiene una sintaxis estricta de comandos mediante gestos, desarrollo de la University of California, Davis- Computer Science department. La persona en la foto es autor de ProteinShop , Oliver Kreylos.....	83
Ilustración 21: Robot Pepper, puede reconocer gestos y emociones del interlocutor y aprender gestos para recibir órdenes.....	86
Ilustración 22: Television Samsung en venta desde 2013 tiene control de voz, reconocimiento de gestos, control remoto con TouchPad e identificación biométrica (por rasgos faciales) de los usuarios.....	90
Ilustración 23: El aparato digitalizador del gesto montado en el marco de los anteojos permite manipular una interfaz especial con gestos de las pestañas, movimiento de los ojos y algunos movimientos de la mejilla. Foto de Associated Press /Elizabeth Dalziel (2015).....	92

Ilustración 24: Demostración del sistema de menú virtual. El proyector está completamente cenital a la mesa.....	98
Ilustración 25: Tres de los gestos principales en la elección del menú. De arriba a abajo: ampliar foto, selección de platillo (el movimiento inverso quita el platillo seleccionado de la lista de comanda), quitar foto y volver a categoría.....	100
Ilustración 26: Sintaxis de los gestos de "Cuerpo-viento". Arriba: ingresar y salir por puntos opuestos del escenario: cambiar a segmento de video siguiente. Enmedio: Ingresar y salir por el mismo punto lateral: generar audio en ritmo más rápido siguiente. Abajo: Ingresar por punto lateral, salir por el centro: generar audio en ritmo mas lento siguiente.....	103
Ilustración 27: Durante la ejecución de Cuerpo Viento. Foto: A. Jaramillo.....	105
Ilustración 28: Roboreptile. Los sensores infrarrojos están situados en la cabeza, parecen fosas nasales, y en la cabeza también está situado el sensor auditivo.....	107
Ilustración 29: Vocabulario de gestos del roboreptile. De arriba a abajo: Gesto lateral. Gesto superior. Gesto de recompensa (tocando el sensor en el lomo).....	109
Ilustración 30: Algunas pantallas del programa de Aprendizaje del lenguaje de gestos. La niña hacía frases relativas al entorno y el alumno repetía esas frases frente a la cámara de la computadora para ser evaluado.....	111

Índice de Tablas

Tabla 1 - Resumen de los avances de la tecnología de computadora, organizándolos en 5 eras o generaciones, según la tecnología de hardware, modo de operación, Lenguaje de programación, Tecnología de terminal, tipo de usuarios, imagen publicitaria y paradigma de interfaz de usuario.....	34
Tabla 2 - Comparación de las interfaces GUI WIMP y las interfaces antiWIMP. tabla de Gentner y Nielsen, 1996, Traducción de Gonzalo León.....	53
Tabla 3 – Vocabulario de gestos multitouch. Apple, 2007.....	59

Bibliografía

- [ANDERSEN, et al, 1993] *Andersen, P., Holmqvist, B. y Jensen, J., **The computer as medium***, Cambridge University Press., Nueva York, E.U.A., 1993. (494 págs.)
- [APPLE, 1995] *Apple computer Inc., **MacIntosh human interface guidelines***, Addison-Wesley, Reading, Ma. E.U.A., 1995. (86 págs.)
- [BARTHES 2002] *Barthes, Roland, **Lo Obvio y lo Obtuso: Imágenes, gestos y voces.***, Editorial Paidós., Barcelona España., 2002. (380 págs.)
- [BARTHES 1967] *Barthes, Roland, **El sistema de la moda***, Editoriál Paidós, Barcelona, España, 1967. (448 págs.)
- [BIOCCA, LEVY 1995] *Biocca, Frank y Levy, Mark R. (compiladores), **Communication in the age of virtual reality***, Lawrence Earlbaum Assoc, Hillsdale, N.J. E.U.A., 1995. (401 págs.)
- [BIRDWHISTELL 1994] *Birdwhistell, Ray L., **Un ejercicio de kinésica y de lingüística: la escena del cigarrillo***, Incluido en el libro La nueva comunicación ,Editorial Kairós, Barcelona, España, 1994 pags. 166-197
- [BOLTER, GRUSIN 2000] *Bolter, Jay David ; Grusin, Richard , **Remediation: understanding new media***, MIT Press, E.U.A., 2000. (295 págs.)
- [BOYLE 1997] *Boyle, Tim, **Design for multimedia learning***, Prentice Hall, Londres, R.U., 1997. (240 págs.)
- [BUTTER y POGUE, 2002] *Butter, Andrea; Pogue, David, **Piloting Palm: The Inside Story of Palm, Handspring, and the Birth of the Billion-Dollar Handheld Industry***, John Wiley & Sons, Inc., Nueva York, EEUU, 2002. (353 págs.)
- [BUXTON 1994] *Buxton, William, **The three mirrors of interaction: a holistic approach to user interfaces.***, Capítulo del libro Interacting with virtual environments ,John Wiley and Company Publishers, New York, E.U.A., 1994 pags. 122-165

- [BUXTON 1990B] *Buxton, William, A three state model of graphical input., en INTERACT '90 Proceedings of the IFIP TC13 Third International Conference on Human-Computer Interaction, vol. 13, num. 3, mes de Agosto, 1990, IFIP, North-Holland Publishing Co., Amsterdam, Holanda, pages 449-456*
- [BUXTON 1990A] *Buxton, William, The natural language of interaction: A perspective on non-verbal dialogues., Capítulo del libro The Art of Human-Computer Interface Design ,Addison-Wesley, Reading, MA. E.U.A., 1990 pags. 405-416*
- [BUXTON 1983] *Buxton, William, Lexical and Pragmatic Considerations of Input Structures., en ACM SIGGRAPH Computer Graphics Newsletter, vol. 17, num. 1, mes de enero, 1983, ACM SIGGRAPH, ACM Press, Nueva York, E.U.A., pags 31-37*
- [CLARK 1999] *Clark, Herbert H., Using language, en Journal of Linguistics, vol. 35, num. 1, mes de , 1996, , Cambridge University Press, Cambridge, pags 167-222*
- [DAVIS 1999] *Davis, Flora, El lenguaje de los Gestos, Emecé editores, Argentina, 1999. (257 págs.)*
- [ENGELBART 1968] *Engelbart, Douglas C., Augmenting Human Intellect, Capítulo del libro Multimedia: From Wagner to Virtual Reality , Stanford Research Institute, Stanford, E.U.A, 1968 pags. 64-90*
- [FISHER 1994] *Fisher, Scott, Multimedia authoring: building and developing documents, AP Professional / Harcourt Brace, Boston, MA, E.U.A., 1994. (286 págs.)*
- [GENTNER y NIELSEN 1996] *Gentner, D.; Nielsen, J., The Anti-Mac interface, en Communication of the ACM, vol. 39, num. 8, mes de Agosto, 1996, ACM, ACM Press, New York, E.U.A, pags 70-82*
- [GOLDIN 2005] *Goldin-Meadow, Susan, Hearing gesture: How our hands help us think, Harvard University Press, Cambridge, Mass, EEUU, 2005. (280 págs.)*
- [HALL 1994] *Hall, Edward T., Proxémica, Capítulo 2 del libro La nueva comunicación ,Editorial Kairós, Barcelona, España, 1994 pags. 198-230*

- [KADOUS 1995] *Kadous, Waleed, **GRASP: Recognition of Australian Sign Language Using Instrumented Gloves***, The University of New South Wales, Australia, 1995. (85 págs.)
- [LARRIVA 2013] *Larriva Calle, Diego F., **Diseño de Software Educativo incluyente***, Tesis para obtener el grado de Maestría en Diseño Multimedia, Universidad del Azuay, , 2011 (69 pags)
- [LAUREL 1995] *Laurel, Brenda, **The Art of Human-Computer Interface Design***, Addison-Wesley, Reading, MA. E.U.A., 1995. (523 págs.)
- [LIANG HUANG et al 2003] *Liang Wang, Tieniu Tan, Huazhong Ning and Weiming Hu,, **Silhouette analysis-based gait recognition for human identification**, en IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, num. 12, mes de Diciembre, 2003, IEEE, IEEE press, , pags 1505-1518*
- [LISTER 2003] *Lister, Martin, **New media: a critical introduction***, Routledge, E.U.A., 2003. (404 págs.)
- [LOPES 2009] *Lopes, Dominic, **A Philosophy of Computer Art***, Taylor & Francis, E.U.A., 2009. (170-210 págs.)
- [MANDEL 1994] *Mandel, Theo, **The elements of user interface design***, John Wiley and Co. Publishers, Nueva York. E.U.A., 1994. (434 págs.)
- [MARCUS 1992] *Marcus, A., **Graphic design for electronic documents and user interfaces***, ACM press / Addison-Wesley, Nueva York, E.U.A., 1992. (386 págs.)
- [MAYBURY 1993] *Maybury, Mark T., **Intelligent multimedia interfaces***, MIT Press, Menlo Park, E.U.A., 1993. (405 págs.)
- [MCLUHAN 1997] *McLuhan, M. ; Fiore, Q., **El medio es el masaje. Un inventario de efectos***, Editorial Paidós, Buenos Aires, Argentina, 1997. (160 págs.)
- [NESPOULOUS et al, 1986] *Nespoulous, J. L., Perron, P. y Lecours, A. R. (editores)., **The Biological Foundations of Gestures: Motor and Semiotic Aspects.***, Lawrence Erlbaum Associates, Inc. - Psychology press Hillsdale, New Jersey, USA, E.U.A., 1986. (336 págs.)
- [NIELSEN, MOLICH 1990] *Nielsen, J., and Molich, R., **Heuristic evaluation of user interfaces***, Ponencia incluida en Proceedings of the ACM CHI'90 Conference ,ACM Press, E.U.A., 1990 pags. 249-256

- [NIELSEN 2009] *Nielsen, Jakob, **Macintosh: 25 Years***, Publicado en el web en <http://www.useit.com/alertbox/macintosh.html> ,Useit.com, , 2009 pags. 10, consultado en enero 2010
- [NIELSEN 2000] *Nielsen, Jakob, **Why You Only Need to Test with 5 Users***, Artículo disponible en <http://www.useit.com/alertbox/20000319.html> ,Useit.com, , 2000 pags. 12, Consultado en junio 2017
- [NIELSEN 1993] *Nielsen, Jakob, **Noncommand user interfaces***, en Communications of the ACM, vol. 4, num. 36, mes de Abril, 1993, ACM, ACM Press, E.U.A., pags 83-99
- [NIELSEN et al 2003] *Nielsen, M., Störring, M., Moeslund, T. B., & Granum, E., **A procedure for developing intuitive and ergonomic gesture interfaces for HCI*** , Ponencia, incluida en International Gesture Workshop ,Springer Berlin Heidelberg, Alemania, 2003 pags. 409-420
- [OGATA 2003] *Ogata, Katsuhiko, **Ingeniería de Control Moderna***, Pearson / Prentice Hall, Madrid, España, 2003. (965 págs.)
- [OPPENHEIM, et al 1998] *Oppenheim,Alan V., Willsky, Alan S., y Nawab, S Hamid , **Señales y sistemas***, Pearson Educación, Madrid, España, 1998. (956 págs.)
- [QUEK 1994] *Quek, Francis K. H., **Toward a Vision Based Hand Gesture Interface***, Incluido en el libro Visual Reality Software and technology: proceedings of the VRST'94 conference ,World Scientific - The University of Illinois at Chicago, Singapore, 1994 pags. 17-25
- [RODRÍGUEZ 2013] *Rodríguez V, Rita G., **Uso de desplazamientos con la técnica de Danza Contemporánea Flying Low dentro de un espacio creado con proyección multimedia sobre estructuras tridimensionales, para generar un dialogo entre los componentes que serán parte del proceso creativo de la obra Cuerpo-viento***, en Tesis para obtener el grado de Licenciado en Danza, 2013, , Universidad de Cuenca, Ecuador, (100 pags)
- [VRST'94] *Singh, G.; Feiner, S. K.; y Thalmann, D., **Virtual Reality Software and technology.***, Conferencia. Ponencias publicadas como Proceedings of the VRST '94 Conference. Singapore, 23 – 26 August 1994 ,World Scientific, Singapore, 1994 pags. 356

- [STARNER y PENTLAND 1995] *Starner, Thad E.; Pentland, Alex*, **Visual Recognition of American Sign Language Using Hidden Markov Models**, MIT Press, Massachusetts, USA, 1995. (págs.)
- [SUTHERLAND 1965] *Sutherland, Ivan*, **The ultimate display**, conferencia, incluida en Proceedings of the International Federation of Information Congress, Spartan Books, Washington DC. E.U.A., 1965 pags. 506-508
- [TUFTE 1990] *Tufte, Edward R.*, **Envisioning Information**, Graphics Press, Connecticut, E.U.A., 1990. (130 págs.)
- [TUFTE 1983] *Tufte, Edward R.*, **The visual display of quantitative information**, Graphics Press, Connecticut, E.U.A., 1983. (200 págs.)
- [ECO 1975] *Umberto Eco*, **Tratado de semiótica general**, Lumen, Barcelona, España, 1975. (520 págs.)
- [VILLACÍS, M. 2013] *Villacís Luzuriaga, M.*, **Proyecto multimedia como facilitador de la comunicación humana de grupos heterogéneos culturalmente**, Tesis para obtener el grado de Maestría en Diseño Multimedia, Universidad del Azuay, Cuenca, Ecuador, 2013 (85 pags)
- [WACHS et al, 2011] *Wachs, J.P.; Kölsch, M.; Stern, H.; Edan, Y.*, **Vision-based hand-gesture applications**, en Communications of the ACM, vol. 54, num. 2, mes de Febrero, 2011, ACM, ACM Press, New York, E.U.A., pags 60-71
- [WOBBROCK, WILSON y LI 2007] *Wobbrock, Jacob O.; Wilson, Andrew D.; Li, Yang*, **Gestures without Libraries, Toolkits or Training: A \$1 Recognizer for User Interface Prototypes**, Conferencia incluida en Proceedings of the 20th annual ACM symposium on User interface software and technology, ACM Press, E.U.A., 2007 pags. 159-168
- [WU y HUANG 1999] *Wu, Y.; Huang, T.H.*, **Vision-Based Gesture Recognition: A Review**, Ponencia para la conferencia Gesture-Based Communication in Human-Computer Interaction, Springer, Berlin, 1999 pags. 103-115