ICONO 14

REVISTA ICONO 14, 2009, Nº 13, pp. 310-333. ISSN 1697-8293. Madrid (España)
María Teresa Soto Sanfiel: *Interfaces comunicativas*
Recibido: 12/07/2009 – Aceptado: 15/10/2009

Nº 13 – REVISTA DE COMUNICACION Y NUEVAS TECNOLOGIAS – ISSN: 1697 - 8293

INTERFACES HABLADAS
Caracterización, usos y diseño

María Teresa Soto Sanfiel

Profesora titular

Departament de Comunicació Audiovisual I Publicitat. Universitat Autònoma de Barcelona. Campus UAB, Edif. I. Bellaterra (08193). Barcelona, Spain. Tel: + 34 93 581 16 43 / Fax: + 34 93 581 20 05. Email: MariaTeresa.Soto@uab.cat

Palabras clave

Comunicación humano-máquina, interfaces habladas, reconocimiento y síntesis del habla, voz, habla, lenguaje, formación de impresiones, recepción

Key Words

Human-machine communication, speech interfaces, speech recognition and synthesis, voice, speech, impression formation, language, reception

Resumen

Este artículo describe y piensa al fenómeno de las Interfaces habladas (IH) desde variados puntos de vista y niveles de análisis. El texto se ha concebido con los objetivos específicos de: 1.- procurar una visión panorámica de aspectos de la producción y consumo comunicativo de las IH; 2.- ofrecer recomendaciones para su creación y uso eficaz, y 3.- llamar la atención sobre su proliferación e inspirar su estudio desde la comunicación. A pesar de la creciente presencia de las IF en nuestras vidas cotidianas, hay ausencia de textos que las caractericen y analicen por sus aspectos comunicativos. El trabajo es pertinente porque el fenómeno significa un cambio respecto a estadios comunicativos precedentes con consecuencias en las concepciones intelectuales y emocionales de los usuarios. La proliferación de IH nos abre a nuevas realidades comunicativas: hablamos con máquinas.

Abstract

This article describes the phenomena of Speech interfaces (SI) from different perspectives and levels of analysis. It has three specific purposes: 1. to offer a panoramic overview of the communicative aspects related to SI production and consumption; 2. to give recommendations for the optimal and efficient design of this interfaces, and 3. to inspire SI studies from audiovisual communication researchers. In spite of the increasingly presence of SI in our daily lives, there is an absence of texts that characterize and analyze them from its communicative aspects. Also, this work is necessary as its proliferation opens new communicative realities: we're speaking to machines, indeed.

Introducción

Somos protagonistas de un notable incremento en la disponibilidad y el uso de sistemas de comunicación humano-máquina basados en el lenguaje hablado. Podemos hablar, en variedad de contextos, con interlocutores a los que conscientemente identificamos como irreales y que, sin embargo, son capaces de iniciar, mantener o proseguir conversaciones (aunque sea de manera limitada). Podemos hablar con un sistema que reserva nuestros billetes de avión, transcribe nuestras palabras, toma nota de nuestro consumo del gas, o nos dirige hacia el agente de servicio indicado para responder consultas específicas. Podemos hablar con una máquina que nos da información sobre el estado del tiempo y de la carretera… Los avances en el diseño de tecnologías del habla han producido máquinas aptas tanto para reconocer voces y discursos humanos, como para generar los propios. Aún queda mucho trabajo por delante para los especialistas en el diseño de estas aplicaciones porque, en algunos aspectos, son aún rudimentarias. Sin embargo, tanto su perfeccionamiento como su ubicuidad prosiguen vigorosamente.

El presente texto tiene como protagonistas a la *Interfaces habladas* (IH), los dispositivos técnicos que permiten la comunicación oral entre un humano y una máquina, que se han desarrollado, generalmente, como sistemas de recuperación de información o servicios asistenciales/transaccionales y a los que se puede acceder a través de variados medios (teléfonos, instalaciones u ordenadores). El uso de IH tiene varias ventajas: se organiza sobre el vehículo de comunicación más natural entre los humanos (el habla), posibilita que se puedan realizar otras acciones de manera simultánea y permite acceder a mucha información en poco tiempo, por ejemplo (Llisterri, 1988). Se espera que, a corto plazo, las IH proliferen en numerosos ámbitos -institucionales, creativos, educativos o sanitarios- porque contribuyen a reducir costos operativos e incrementan la eficacia de las gestiones y porque se ha demostrado que son un buen vehículo de marketing para empresas e instituciones ya que contribuyen a potenciar la imagen de marca (Kotelly, 2003). Pero, además, se espera que, en un plazo mayor, las IH se logren desarrollar en formas más ambiciosas. De hecho, ya se trabaja para que brinden asistencia terapéutica a personas de la tercera edad, con disminuciones sensoriales, enfermedades psicológicas o que padezcan de privaciones sociales. Asimismo, en la actualidad se busca adaptarlas tanto a situaciones educativas o formativas como asociadas a nuevas propuestas de entretenimiento. Las IH tienen una presencia creciente en nuestra cotidianidad.

Este artículo describe y piensa al fenómeno de las IH desde variados puntos de vista y niveles de profundidad. En esencia, extien-

de a los lectores el resultado de un esfuerzo de reflexión en torno al fenómeno y a sus diferentes aristas. De manera particular, el texto se ha concebido con los objetivos específicos de: 1.- procurar una visión panorámica de aspectos de la producción y consumo comunicativo de las IH; 2.- ofrecer recomendaciones para su creación y uso eficaz, y 3.- llamar la atención sobre este nuevo reto a ser explicado desde el área de la comunicación. Esta ambiciosa concepción se justifica por la ausencia de referencias en español sobre el tema, a pesar de su creciente presencia en nuestras vidas cotidianas. Creemos que este trabajo es pertinente porque el fenómeno significa un cambio respecto a estadios comunicativos precedentes con consecuencias en las concepciones intelectuales y emocionales de los usuarios (Barnes, 2001, 2003; Barnes y Strate, 1996; Levinson, 1997, 1999; Meyrowtiz, 1985; Postman, 1995; Strate, 1999). Creemos que la proliferación de IH, abre numerosas interrogantes a la investigación en comunicación: ahora ya podemos hablar con máquinas[1].

La información del artículo se dispone en tres partes. La primera, explicativa, incluye una categorización de las IH y la definición de sus características técnicas. Su objetivo es centrar al lector en el objeto de estudio ofreciendo pinceladas tanto de la oferta comercial de estas aplicaciones, como de sus aspectos técnicos y componentes básicos. La segunda parte, muy diferente a la anterior, pretende describir la relación comunicativa que se produce entre las IH y sus usuarios durante el consumo. Este enfoque es primordial no solo porque el diseño de estas aplicaciones se enriquece en la medida en que se conocen los comportamientos y actitudes de las audiencias, sino porque de él se desprenden preguntas, sugerencias y reflexiones sobre las nuevas realidades de la comunicación mediada. Por ejemplo, esta parte se refiere a la formación de impresiones a partir del habla de las IH, a las relaciones *parasociales* que generan o a las actitudes hacia la personalidad que manifiestan en sus voces. Se recomienda tomar en cuenta que esta es la parte central del trabajo; aquella que, desde una perspectiva de las ciencias de la comunicación, contiene la información más académica y científica. La tercera y última parte, propone tanto métodos como técnicas específicas para la creación y diseño de IH. También, ofrece una propuesta de modelo de validación de la calidad de las IH. Finalmente, las conclusiones, cortas, tras la profusión de información, resumen los retos actuales en la investigación y desarrollo de estas aplicaciones.

Objetivos

A partir de la observación de aspectos relevantes de la producción y consumo de las Interfaces habladas, el texto propone recomendaciones de diseño y uso eficaz de dichas aplicaciones desde una óptica comunicativa. El artículo considera que la proliferación de máquinas que hablan con humanos implica nuevas realidades comunicativas que merecen la atención de los investigadores.

Metodología

La propuesta se produce tras el análisis de literatura precedente proveniente de distintos campos: tecnologías del habla, psicología de la recepción, comunicación interactiva, y producción audiovisual de mensajes.

1. ¿De qué hablamos al referirnos a las IH?

1.1. De las aplicaciones: ¿Cómo y dónde se hallan disponibles?

Sea como parte del equipamiento del hogar, de la oficina o en dispositivos móviles, las IH se pueden encontrar, en mayor medida: 1.- aplicadas a la telefonía (las más frecuentes); 2.- en el control o comando de ordenadores y equipos; 3.- como sistemas de dictados, documentación y transcripción; 4.- en aplicaciones para la identificación del hablante; 5.- en automóviles; 6.- en aplicaciones manos libres y 7.- como personajes animados y bustos parlantes[2]. Considérese que la presente no es una clasificación excluyente y que se ofrece con el propósito principal de definir a nuestro objeto de estudio. A continuación, se describe brevemente a estas categorías de aplicaciones.

De las aplicaciones de telefonía (1), las IH más frecuentes y expandidas, son los sistemas de "respuesta de voz interactiva" (IVR)[3], un tipo de interfaz que permite que, en el transcurso de una llamada telefónica, se pueda interactuar con un ser humano mediante la captura (reconocimiento) de vocabularios reducidos del habla y la producción (intercambio) de información oral (en la mayoría de los casos humana porque tiene mejor calidad), previamente grabada y automatizada. Típicamente, estos sistemas aparecen asociados a servicios de atención de las empresas o instituciones y guían al interlocutor a través de menús u opciones preestablecidas (a

modo de centralita telefónica). Sin embargo, también se utilizan para la búsqueda de información en todo tipo de teléfonos, para gestiones bancarias (manejo de cuentas, obtención de saldo, realización de pagos), para concertar visitas o citas, para comprar entradas o billetes de transporte, para seguir el estado de un envío o servicio, para reportar equipos estropeados o saldos de contadores, y/o para activar/desactivar prestaciones telefónicas. Con el uso de estas aplicaciones, las empresas reducen tanto el tiempo de espera de los clientes, como el de gestión de las llamadas entrantes. Buena parte de la eficacia de estas aplicaciones de gestión se encuentra en la redacción de los mensajes (directa, sencilla y económica en palabras).

Las IH de control o comando de ordenadores y equipos (2) son también muy frecuentes. Se utilizan, en general, en sistemas de respuestas a preguntas, de acceso a grandes listas, cuando no se puede/quiere utilizar las manos y/o para humanizar ordenadores4. Su diseño concreto depende de la aplicación que la contiene. Así, estas IH se encuentran en programas de entretenimiento (en conversaciones con personajes o interfaces de juegos o simuladores electrónicos) o de edición de documentos (p.ej.: para cortar textos, seleccionar tipos de letras o moverse entre menús). Normalmente, ofrecen el listado de su oferta de operaciones posibles y el usuario escoge entre ellas (Huang, Acero y Hon, 2001: 922-923).

Las IH de dictado o transcripción (3) son, esencialmente, sistemas de reconocimiento de voz usados para convertir a texto grandes cantidades de palabras (largos discursos) o cuando se trabaja con manos ocupadas. Son muy útiles para determinadas profesiones (abogados, radiólogos o periodistas) y, en la actualidad, se comercializan adaptados a las jergas profesionales[5].

Por su parte, las IH de identificación del hablante (4) pretenden verificar y detectar identidades mediante la autentificación del habla y son usadas en dispositivos de seguridad y control de acceso. Se basan en la noción de "huella sonora" (de que existe, y se pueden aislar en parámetros, una identidad del habla sonora única e intransferible, determinada por la configuración del tracto vocal y sus hábitos de habla). Los hablantes que deben ser identificados producen una serie de textos sonoros que son tratados y almacenados en un ordenador para cotejar con una muestra, llegado el momento de la verificación. Como se deduce, estas IH se encuentran aplicadas en servicios de banca y compras telefónicas a distancia, en servicios forenses o sanitarios y en dispositivos de acceso o control de entradas. Se muestran útiles para las situaciones en que un usuario debe entrar repetidamente a un sistema o cuando existen grandes cantidades de personas que deben ser identificadas.

En las IH diseñadas para automóviles (5), también cada vez más difundidas, además de las aplicaciones de control, se incluyen

sistemas de navegación GPS[6] manejadas por voz y/o programas que permiten leer mensajes de correo o de texto. Otra categoría, la de las IH adaptadas a dispositivos manos libres (6), se usa con frecuencia para vehicular el acceso a información que no aparece disponible visualmente en, sobre todo, agendas electrónicas, reproductores musicales o teléfonos móviles. Finalmente, por lo que respecta a este panorama, las IH (7) se encuentran, también, presentadas en forma de bustos parlantes o animaciones de personajes que forman parte de aplicaciones multimedia, de aprendizaje de lenguas o de gestión de aplicaciones.

1.2. De la arquitectura de las IH (componentes principales)

Es preciso tener, al menos, nociones básicas sobre su configuración: ¿qué incluyen? ¿Cómo es su estructura? Técnicamente, las IH se componen de una aplicación de *síntesis del habla*, de otra de *reconocimiento del habla*, y de un dispositivo técnico que permite la interacción entre los componentes del sistema, el llamado API[7]: 1.- Las aplicaciones de *síntesis del habla* (TTS)[8] son las llamadas a reproducir/construir los fragmentos del lenguaje (el habla de la máquina). Para ello, traducen, recuperan o transforman textos escritos o representaciones fonológicas a un código que puede ser interpretado por una máquina (preferiblemente una forma de onda) para posteriormente ser transmitido a los oyentes; 2.- Las aplicaciones de *reconocimiento automático del habla* extraen/convierten las señales vocales en parámetros acústicos y fonéticos (a veces también léxicos) según modelos que emulan la percepción o la articulación del habla, y 3.- Las API[9] que facilita la comunicación/interconexión entre los diferentes componentes, *softwares* o utilidades que forman parte del sistema de una IH. Se acepta que las IH son más eficaces, potentes y consistentes en la medida en que: 1) identifican e interpretan el habla de un mayor número usuarios del sistema y, 2) producen habla artificial, que parece humana, para preguntar o responder a los requerimientos de sus usuarios[10]

En relación a las aplicaciones de *síntesis*, existen tres métodos (o formas) de generar lenguaje artificial (TTS) hoy día. Todos estos modelos conciben a la sustancia sintética del habla a partir de la caracterización de una sustancia real. Sin embargo, los métodos se diferencian en el origen y parámetros que usan para modelar el habla: 1.- el de *síntesis articulatoria* emula el proceso físico de producción del habla humana para lo que imita las dinámicas de los órganos articuladores durante el paso del aire a través de ellos; 2.- el de *síntesis de formantes* describe matemáticamente el proceso acústico de producción del habla, considera al tracto vocal como un grupo de resonadores que producen ondas sonoras y lo representa como una función matemática que varía con el tiempo, y 3.- el de *síntesis analítica* genera una forma de onda específica a

modo de representación visual (Nusbaum y Shintel, 2006: 20).

Las aplicaciones de *reconocimiento* del habla son también de varias clases. Las más sencillas permiten identificar fragmentos cortos del habla y tienen un vocabulario limitado (menos de 50 palabras) para comandar o controlar la aplicación; las más complicadas pueden reconocer habla continúa, y espontánea, de cualquier hablante, en gran número de situaciones (más de 20.000 palabras) (Huang, Acero y Hon, 2001: 936). De ello se puede extraer que la robustez de un sistema de reconocimiento de voz depende de un conjunto de factores: a.- del modo de elocución del hablante (palabras sueltas, encadenadas o habla continua); b.- del número de locutores que es capaz de comprender (desde uno, el que la ha entrenado a comprender su propia voz, hasta varios locutores que nunca han utilizado el sistema); c.- del tipo (y dificultad) del vocabulario; d.- del modo en que adquiere la palabra que reconoce (en local o a distancia); e.- de la forma y lugar en que se capta el sonido del habla (micrófono cerca o lejos de la boca y cantidad de ruido ambiente), y f.- de la naturaleza y complejidad del habla a decodificar (Haton et al., 2006: 285-286).

Concerniente a la API, la herramienta que conecta a los componentes del sistema, en la actualidad existen una serie de aplicaciones estándares que se aplican a la creación de IH: 1.- SALT, una extensión XML de HTML para la construcción de aplicaciones multimodales y que permite añadir interfaces vocales a la *web*[11]; 2.- VoiceXML, desarrollado por el consorcio W3C para desarrollar aplicaciones sonoras semejantes a las visuales con base en HTML[12]; 3.- Microsoft SAPI, la interfaz natural de Windows que, en su última versión incorpora a un personaje virtual femenino[13]; 4.- JSAPI[14], que permite integrar las tecnologías del tratamiento de la palabra (síntesis y reconocimiento) a aplicaciones JAVA; 5.- MMIL[15], un lenguaje común para integrar informaciones multimodales que estandariza el comité ISO/TC 37/ SC4, y 6.- MPEG7 que más allá de su aplicación para interfaces habladas, se creó para gestionar contenidos audiovisuales multimedias[16].

2. Cuando los humanos hablamos con máquinas

En este apartado se concentra la información que consideramos más relevante de este artículo. Incluye datos científicos sobre el fenómeno de las IH analizado desde una perspectiva comunicativa. Busca dar cuenta de las relaciones que se producen en el intercambio humano-máquina y justifican las recomendaciones de creación que se incluyen en la tercera parte. Su propósito es, también, explicativo de los factores de las IH.

Tal y como se ha mencionado, algunas investigaciones en comunicación sustentan la teoría de que los seres humanos respondemos automática e inconscientemente a las máquinas como si fuesen seres humanos (Brennen, 1998; Lee y Nass, 2004; Moon y Nass, 1996; Nass y Lee, 2001; Sundar y Nass, 2000); que entre los individuos y los medios, en general, se producen patrones de relación propios de la comunicación interpersonal[17](Reeves y Nass, 1996)[18] y que les consideramos actores sociales. Por ejemplo, en nuestro trato con ellas, hacemos uso de los estereotipos (Lee, Nass y Brave, 2000; Nass, Moon y Green, 1997) o tenemos obligaciones morales (Fogg y Nass, 1997).

Lo dicho lleva indefectiblemente a asumir que los actuales conocimientos sobre algunos de los fenómenos presentes en los intercambios interpersonales, y en la formación de impresiones sobre seres humanos, sean aplicados consciente y voluntariamente al diseño de IH con el fin de garantizar su eficacia comunicativa. Parte del objetivo de este apartado es aportar evidencia a favor de estas ideas. Tómese en cuenta, sin embargo, que en la creación de IH, los especialistas han volcado sus experiencias: no puede ser no-humano, aquello que se deriva del hombre[19].

La primera pregunta que nos surge es: ¿existen (y cuáles son) determinadas características en la configuración de una máquina que provocan que reaccionemos de dicha manera? Y, encontramos respuestas para estas preguntas en una experiencia de la primera parte del siglo pasado: Turing (1950) demostró que solo era preciso un mínimo de comportamientos específicamente humanos para engañar a la percepción. El investigador creó una máquina, y un test, que mostraron empíricamente que su interlocutor la consideraría humana si en su comportamiento comunicativo exhibía características propias de una persona. Luego, las máquinas son consideradas humanas si se comunican como humanos. Y ello nos conduce al más natural y genuinamente humano de los comportamientos comunicativos: el habla (del que, además, se sirven las IH). Tan solo con la presencia de esta propiedad en una máquina, un ser humano puede considerarla su par.

El discurso anterior lleva implícito dos aspectos que están, en este caso, el de las IH, intrínsecamente relacionados: 1.- los procesos de formación de impresiones, y 2.- el habla como fuente de información en las relaciones interpersonales. En la intersección de ambos terrenos, en el de la formación de impresiones sobre los hablantes situamos esta reflexión: los seres humanos atribuimos cualidades y características a las máquinas que hablan, en función de nuestras experiencias con la percepción del habla, y en especial de nuestras experiencias con el habla mediada. Las características del habla y del discurso de las IH, por tanto, tienen que adecuarse no solo a las propias del lenguaje natural, sino a las del comunicador prototípico (el que debe determinarse según el caso). Abundamos en estas ideas a continuación.

2.1. La formación de impresiones en los humanos a través del habla de las IH (prototipos cognitivos)

La formación de impresiones es un estadio del proceso perceptivo que se desata (según los conocimientos actuales de la psicología de la percepción) de forma increíblemente rápida, automática e inconsciente con el propósito de que podamos relacionarnos con la información novedosa. Debido a que los seres humanos tenemos una capacidad limitada de procesamiento, debemos categorizar los datos en base a nuestros conocimientos precedentes. De esta manera, compensamos nuestras limitaciones perceptivas (Abele y Petzold, 1988; Lang, 2000; Lang et al., 2004). El mecanismo que subyace a la atribución de cualidades humanas a una máquina que nos habla, a una IH, se podría explicar por la siguiente frase: ya que nuestras experiencias previas con el habla se han realizado siempre entre humanos, quien habla es humano. Y es por eso, porque la mente funciona por condicionamientos previos (al menos hasta que incorpora un nuevo aprendizaje), es lógico usar su *background* de manera operativa en la creación de IH eficaces.

La primera noción a tomar en cuenta en la construcción del personaje de la IH concreta es que los seres humanos tenemos imágenes del *comunicador ideal* para cada situación y que esas imágenes funcionan como categorías en el proceso de formación de impresiones. En el procesamiento de los hablantes usamos prototipos cognitivos sobre estilos de comunicación que funcionan de manera pragmática y orientada por la tarea que desempeña el hablante. Buena parte de dicho prototipo se determina a partir de la noción de competencia comunicativa (Pavitt y Haight, 1985; 1986). Es decir, la IH será una comunicadora ideal en tanto que competente en la producción de determinado mensaje[20].

La segunda noción tiene que ver con los aspectos del habla que proveen de información sobre los hablantes y provocan nuestra formación de impresiones. En este sentido, se sabe que los seres humanos obtenemos

dos tipos de datos del habla. Nos fijamos en aspectos del contenido (las palabras o las frases) y, en las cualidades funcionales de la voz: en los parámetros acústicos (p.ej.; tono, timbre, velocidad del habla, uso de las pausas e intensidad) (Scherer, 1979). De hecho, en todos los actos del lenguaje en los que participamos, evaluamos a los hablantes por la comparación con el prototipo cognitivo de la cualidad verbal de su lenguaje y de la no-verbal (paralingüística). Una virtud de la creación de las IH es que ambos aspectos se pueden concebir y manipular a voluntad (la segunda en mayor o menor grado dependiendo del dispositivo tecnológico con el que se cuenta).

2.2. De las relaciones parasociales con la IH

Durante la interacción con IH se pueden producir relaciones *parasociales*[21]. Al igual que ocurre en el consumo de los medios audiovisuales (Barnes, 2003; Giles, 2002) frente a ellas: los seres humanos nos comportamos como si nos relacionásemos con una fuente, cuando verdaderamente nos estamos relacionamos con el medio (Nass y Sundar, 1994). Por ejemplo, también aplicamos estereotipos de género en la percepción de las voces –masculinas o femeninas- de los ordenadores (Green, 1993), o las consideramos nuestros compañeros de trabajo (Nass, Fogg y Moon, 1994) o las tratamos siguiendo normas de cortesía (Finkel, Guterbock y Borg, 1991; Jones, 1964; Kane Macaulay, 1993).

Por otra parte, como sucede con los personajes o medios convencionales, la producción y consumo de las IH es semejante al del proceso de producción audiovisual (se preproduce, graba y se emite). Igualmente, es común también que nuestras reacciones a estos productos enlatados sean espontáneas (y que sigan un repertorio de cogniciones aprendido). Luego, es plausible considerar que parte de la credibilidad del personaje o del sistema de IH dependa de factores como: 1.- la completitud del escritor/diseñador del guión en la comprensión de los personajes mediáticos; 2.- las necesidades o valores de las audiencias; 3.- la representación-actuación del personaje, y 4.- la gestión de los turnos de palabra y aspectos de cooperación para alcanzar metas sociales.

2.3. Sobre la actitud de la IH y nuestra sensación de no-mediación: coherencia del personaje

Todo lo dicho no significa, en absoluto, que los seres humanos seamos incapaces de detectar que nuestro interlocutor es una IH. En realidad, eso no importa tanto; lo relevante es nuestra voluntaria renuncia a la percepción de mediación (la suspensión del descreimiento sobre la artificialidad de la máquina en la acción). El proceso es, de alguna manera, semejante al que ocurre cuando consumimos una ficción: vemos los fenómenos a los que nos sometemos como si fuesen reales, a pesar de que sabemos que son irreales. En nuestra interacción

con una IH podemos renunciar a nuestra noción de realidad y aceptar la irrealidad de la máquina. No obstante, ello ocurre únicamente con la condición de que las acciones del personaje-máquina-hablante parezcan adecuadas a nuestros esquemas perceptivos y cognitivos previos. Cualquier duda sobre la consistencia del comportamiento comunicativo de la IH inmediatamente reduce la validez de la experiencia de no-mediación (Vorderer, Klimmt y Ritterfeld, 2004). Luego, aceptamos, y nos envolvemos de buen grado en la irrealidad de la situación, cuando los comportamientos del habla exhibidos por la IH se ajustan a nuestras experiencias (y expectativas) sobre el habla mediada. Convenimos en ser persuadidos de su realidad, si nos creemos lo que la IH dice. Esto es, aceptamos la no-mediación, ser transportados, estar presentes, ser involucrados, estar inmersos o fluir con la experiencia (Biocca, 2001; Lee, 2004; Lombard y Ditton, 1997), si el personaje es creíble. Es preciso, por lo tanto, construir el "personaje" de la IH con coherencia.

2.4. Rasgos del habla de las IH y personalidad: parámetros acústicos (el tono)

Asociamos voces con formas de ser. La sola presencia de habla provoca juicios inminentes sobre la personalidad, independientemente de que las voces sean naturales o sintéticas, como en las aplicaciones TTS. Sin embargo, y aunque se ha sugerido que es muy urgente avanzar en el estudio de las claves vocales que manifiestan personalidad en las máquinas -especialmente en relación con sus variaciones en tiempo real -(Nass y Lee, 2001), todavía no se dispone de un cuerpo de referentes específicos sobre el tema. Ahora bien, sobre los rasgos vocales asociados a la personalidad de los hablantes en otras situaciones comunicativas si se dispone datos. A continuación, nos basaremos en esos estudios para ofrecer los rasgos considerados más importantes en la generación de voces para las IH; hablaremos de los parámetros acústicos (en particular el tono)[22] que deben incorporarse al diseño de las interfaces.

Uno de los aspectos más relevantes en la formación de impresiones sobre la personalidad de los hablantes es la percepción de atractivo vocal. De hecho, buena parte de las percepciones sobre la personalidad de los humanos se producen a partir de la evaluación, inconsciente e inmediata, de dicho factor, como demostraron Zuckerman y Miyake (1993). Específicamente, estos investigadores encontraron que: a.- los seres humanos tendemos a asociar las voces atractivas con impresiones globales de la personalidad más positivas; b.- las voces atractivas estimulan nuestro deseo de afiliación; c.- aumentan el índice de similitud que asumimos con respecto al hablante, y d.- producen empatía (creemos que el hablante tiene un estatus semejante al nuestro).

Además de lo anterior, y específicamente relacionado con los parámetros acústicos, los investigadores demostraron que el atractivo vocal se relacionaba con el tono de voz (frecuencia fundamental): las voces con tonos más bajos son consideradas más atractivas. En sintonía con ello, Collins (2000) encontró que las voces masculinas más graves son consideradas más atractivas por las mujeres.

Otro de los aspectos relevantes en la formación de impresiones sobre la personalidad y atractivo vocal de las voces de las IH es la Credibilidad (Nass y Lee, 2001). De hecho, se trata del factor más relevante en la percepción de los hablantes que se dirigen a las audiencias para transmitir mensajes prediseñados y complejos porque orienta y dirige las evaluaciones de los receptores durante los fenómenos de percepción sonora (Soto, 2000; 2008; 2008ª). Las investigaciones han demostrado que los comunicadores atractivos son más creíbles que los no atractivos (Eagly and Chaiken, 1975). Nass y Lee (2001) explican este fenómeno por el principio de similitud y atracción de la personalidad: nos gusta parecernos, encontrarnos y rodearnos de gente con cualidades positivas (creíbles y atractivos). Desde el punto de vista acústico, además de más atractivas -como se dijo- también son consideradas más creíbles, sean las voces graves masculinas o femeninas (Soto, 2000; 2008).

Concerniente en particular con el sexo del hablante, existen indicios acerca de las diferencias en los juicios que provocan las voces masculinas y femeninas. Las primeras son percibidas como más amigables y objetivas cuando elogian o alaban (Ashmore, 1981; Basow y Silberg, 1987; Eagly y Wood, 1982), mientras que las segundas son consideradas mejores para la enseñanza y la transmisión de mensajes cálidos o afectivos. Como se puede deducir, estas características son coherentes con los estereotipos de género.

Por otra parte, se ha demostrado que existen relaciones entre la personalidad de la voz y el texto que transmite una IH (Nass y Lee, 2001; Soto, 2008a). Así, un texto es considerado extrovertido si lo escribe una persona extrovertida (Nass y Lee, 2001). También a propósito del texto, otra investigación ha probado que la transmisión de textos, formales y no formales, por voces profesionales (las entrenadas para la producción de variedades controladas de la voz) es más creíble que la transmitida por los no-profesionales. En dicha investigación se comprobó que los hablantes profesionales son considerados más creíbles que los no profesionales, independientemente del texto que interpreten (Soto, 2008ª).

Por lo dicho, se justifica la utilización de voces atractivas y creíbles: graves, profesionales (con capacidad para producir pausas y variaciones melódicas controladas), claras, inteligibles y de un género adecuado al mensaje para conseguir una percepción más favorable de la IH y de la interacción con ella (al menos en las IH de servicio).

Sin embargo, no solo serán el atractivo de la voz y su credibilidad, los rasgos imprescindibles. Como resumen Nass y Lee (2001) se necesitará también un "texto atractivo y creíble" (las intervenciones de la IH), hecho también por un "escritor atractivo y creíble", y capaz de crear una personalidad de la máquina "atractiva y creíble"; una fórmula que, parafraseando a esos investigadores, es vieja conocida de la industria audiovisual.

3. Eficacia comunicativa de las IH: recomendaciones

Al presentar datos sobre los factores acústicos relacionados con la percepción de personalidad en las IH, introdujimos evidencias acerca de la cercana relación existente entre la voz y el texto. Ahora, en este apartado, nos centraremos en enumerar las condiciones esenciales que el texto de las IH debe cumplir para lograr una comunicación eficaz. Cuando hablamos de textos, nos referimos a los comandos[23], las instrucciones que la máquina da al usuario y que deben ser acordes con la personalidad de la IH, (y esta con el servicio para el que se diseña o el posicionamiento de la marca a la que pertenece). Considerar dichos aspectos del personaje-máquina significará tomar decisiones no sólo acerca del sonido de la voz, sino del tipo de contenido en micro y macro niveles. No obstante, fundamentalmente, es preciso tomar en cuenta que la percepción del mensaje sonoro es secuencial, inmediata y fugaz y que ello condiciona la concepción de los comandos. Daremos más detalles a continuación.

Primero, desde una aproximación macro, recomendamos mantener en mente las reglas pragmáticas del lenguaje conversacional, conocidas como Máximas de Grice[24]. Específicamente, es conveniente que: 1) cada intervención de la IH sea lo suficientemente informativa para el intercambio concreto y nunca de más información de la necesaria; 2) que la IH reproduzca solo información de calidad, lo que pueda ser comprobado y sea creíble; 3) que únicamente diga lo importante o relevante para la tarea o la relación con el usuario, y 4) que sea directa, clara, escueta y ordenada. El seguimiento de estos principios ofrece la posibilidad de controlar y salvaguardar la imagen positiva de la IH (o del servicio, empresa, institución o persona a la que representan).

Desde una aproximación micro, es aconsejable: a.- redactar comandos muy cortos; b.- utilizar estructuras del lenguaje semejante a las del habla cotidiana: estructura simple, (sujeto + verbo + complementos) sin frases subordinadas; c.- lenguaje llano,

convencional, ausente de jergas profesionales o palabras en otros idiomas, y d.- una única idea, completa, en cada frase (con principio y final).

Además de lo dicho, es ideal que, en el caso de los textos que incluyen preguntas de la IH a los usuarios, se tomen en cuenta las siguientes ideas: 1.- escribir textos directos que guíen a los usuarios acerca de cómo responder naturalmente (esto es muy importante dado que pueden usar un repertorio de combinaciones del lenguaje muy amplio e ilimitado); 2.- tomar en cuenta los turnos de interacción (las posibilidades de que sea el usuario quien controle la comunicación); 3.- considerar los errores de captación/interpretación del sistema (retroceder para comprobar o remitir la conversación a un humano cuando sea necesario).

Una vez redactado el texto, se recomienda normalizarlo (adecuado para la traducción fonológica de la máquina). Por ello, puede considerarse los siguientes aspectos: a.- recordar que algunos símbolos de puntuación (las comas y los puntos) se utilizan como información prosódica y definen tanto las pausas, como la entonación (pausa corta y punto final/descenso de la entonación); 2.- evitar determinados signos de puntuación (las comillas, o los puntos suspensivos); 3.- huir de los acrónimos, a menos que se unan para convertirse en palabras (o separar sus letras para que sean correctamente pronunciadas); 4.- convertir números a letras (palabras); 5.- revisar la coherencia sintáctica (sujeto + verbo + predicado); 6.- repetir el texto en voz alta para definir los acentos de las palabras (que no necesariamente deben coincidir con la gramática formal, sino con la pragmática) y el sonido global de la frase, y 7.- comprobar la adecuación de la pronunciación de las palabras en relación con el lugar en que se produce la comunicación (las variedades dialectales ocasionan que existan palabras que son pronunciadas de distintas maneras por las variedades dialectales de las lenguas o que la duración de los segmentos fónicos sea distinto) (Nusbaum y Shintel, 2006).

3.1. Del método para el diseño adecuado de las IH

Lo inmediatamente apuntado son datos para momentos y aspectos concretos (las más importantes) de la creación de las IH. No obstante, antes de llegar al nivel de desarrollo que la selección de la voz y la construcción específica del texto significan, se deberían seguir una serie de pasos. En este apartado se presenta un método que, formalizado desde la experiencia, busca garantizar que las IH reflejen tanto las necesidades e intenciones de su promotor, como la actitud favorable de sus receptores (Lai y Yankelovich, 2003)[25]. Sus autores advierten que el proceso de diseño de una interfaz es iterativo (requiere verificación y validación en el tiempo, necesitar estar orientado a satisfacer a los usuarios y debe admitir correcciones sucesivas).

El método consta de seis fases: 1.- modelado de la tarea de la interfaz; 2.- escucha de los hablantes ejecutando la tarea para la que se diseña la interfaz; 3.- redefinición del modelo; 4.- selección de la tecnología, 5.- diseño del diálogo, y 6.- experimentación en ambientes naturales y artificiales. Nos detendremos brevemente en cada fase. No obstante, antes agregaremos que el éxito de una IH depende de un buen número de factores. Quizás, el más importante de todos es que el sistema responda, verdaderamente, a una necesidad real y que, por ello, aporte beneficios a quien las utiliza. También, se requiere que la IH cumpla con otras condiciones, al menos, para que funcione adecuadamente. Por ejemplo, debe ser consistente tecnológicamente y con un modo de empleo fácil de memorizar (debido a que, como se ha apuntado antes, los seres humanos tenemos capacidades perceptivas limitadas, nos cuesta recordar grandes frases y nos es difícil comprender el habla sintética)[26]. Además, debe ser eficaz en el control de los tiempos de la interacción (tanto en el reconocimiento del discurso, como el de su respuesta a las peticiones) para que el usuario no se impaciente (Haton et al., 2006). Finalmente, las IH de servicio deben poder remitir la comunicación a un operador humano en los casos en que no pueda, por si misma, dar respuesta a las demandas del usuario (Haton et al., 2006).

Volviendo al modelo, y concerniente a la primera fase -la de modelado de la tarea (1)-, el método de Lai y Yankelovich aconseja que se avance en la sistematización de los pasos lógicos de la realización de la tarea que llevará a cabo la interfaz, lo que incluye los turnos de palabra (se refiere a una unidad de interacción completa entre IH y usuario) y la información precisa para cada paso. Los autores recomiendan tomar en cuenta que los usuarios pueden aproximarse a realizar las tareas para la que se usa la IH de distinta manera. En este primer paso, debe considerarse aspectos del emisor (objetivos, metas, posicionamiento, capacidad tecnológica, inversión o plataforma) y del receptor (edad, relaciones con la tecnología, idiomas, variedades dialectales, usos y costumbres culturales).

En la segunda fase, la de escuchar (2), los creadores de la interfaz deben documentarse sobre cómo las personas hablan durante el ejercicio de la tarea en su cotidianidad. Según los autores, esto incluye escuchar a los usuarios en prácticas semejantes a las de la que se está modelando. De esta manera, obtendrán información sobre el vocabulario-conceptos, la estructura de las oraciones, el tono de voz adecuado para la tarea, los modelos de interacción y los métodos de *feedback* que se usan en el diálogo natural.

A continuación, re-definirán el modelo de la tarea (3) volviendo al esquema inicial y modificándolo en base a lo obtenido en la fase previa (2). En esta fase se hacen ajustes de vocabulario, del ritmo y toma de turnos. Además, y porque ya los creadores

tendrán información más detalladas sobre características de los usuarios ý la plataforma de consumo de la aplicación (teléfono, PDA, móvil), se podrá definir el tono de la aplicación. En esta fase, los autores recomiendan que se utilice de referencia la peor de las condiciones ambientales de consumo (el móvil, por ejemplo) para prevenir y desarrollar soluciones a los inconvenientes de recepción.

Después, los creadores deberán decidir la tecnología (4); seleccionarán el sistema de reconocimiento de habla y el de TTS. Para el primero tomarán en cuenta su oferta acústica, el vocabulario y la gramática. Además, decidirán si la IH permitirá la interrupción por el hablante. Para el segundo tendrán que decidir entre habla natural y grabada. La primera otorga más naturalidad pero presenta dificultades técnicas si no se graba de antemano. La segunda es mejor para habla dinámica pero pierde calidad acústica.

Seguido, diseñarán el diálogo (5). En este paso es donde se incluye el desarrollo de los comandos, de los que se habló específicamente antes y la selección del talento vocal. Para ello, se revisará nuevamente los datos recogidos, en la primera fase. También, además de los comandos, los creadores deberán diseñar el feedback de la IH a cada intervención del usuario para hacerle saber si ha reconocido su habla correctamente. Es más popular, según los autores, preferir una combinación de *feedbacks* y comandos en los diálogos (turnos) para que el intercambio prosiga lo más rápidamente posible. En esta fase, se diseñan los mecanismos de manejo del error. La corrección requiere de feedback para que los usuarios sepan que el error ha sido conocido y corregido.

La fase final del modelo incluye probar la aplicación, y su usabilidad, en situaciones de laboratorio y ambiente natural. En la primera, se pide a los participantes que ejecuten varias tareas que deba cumplir la aplicación. En la segunda, se deja al libre uso. Por supuesto, esta fase incluye el ajuste y sintonizado final de la IH.

3.2. Sobre como evaluar su calidad: Factores envueltos en la percepción de IH

En este apartado se ofrecen herramientas teóricas para, una vez implantado el sistema de IH, evaluar la percepción de los usuarios acerca de su calidad. El modelo que se presenta, creado por Polkolski (2005[27]) a partir del *meta-análisis* de la literatura precedente, puede implementarse en forma de cuestionario a los usuarios del sistema o de otros métodos cualitativos (entrevistas), si se desea profundizar en las respuestas.

Según Polkolski, los estudios de IH relacionan cuatro factores con la percepción de calidad en los sistemas de IH: 1) metas alcanzadas por el usuario con el uso del sistema; 2) características del habla; 3) expresividad y uso del verbo, y 4) comportamiento del sistema con el cliente. Cual-

quier IH puede ser evaluada en estos términos (que incluyen las propiedades que se describen a continuación).

La investigadora englobó en *metas alcanzadas* (1) al grado en que la IH es capaz de captar las necesidades del usuario y promover su sentido de afiliación. Según la investigadora, una buena IH promueve la sensación de control ofreciendo facilidades para la ejecución de las opciones/tareas, promoviendo una actitud activa y productiva y trabajando correctamente. De esta manera, concluye, el usuario deseará volver a interactuar con el sistema. Por otra parte, halló que las *características del habla* (2) adecuadas eran la selección de una voz natural y agradable para el sistema, que tuviera parecido a la de la gente que aparece en los medios (radio y televisión) y que fuese entusiasta y enérgica. Asimismo, encontró que, en relación a la *expresividad y uso del verbo* (3), era necesario no reiterar excesivamente los mensajes, dar los detalles justos para la realización de la tarea y ofrecer rapidez en la escucha del usuario. Finalmente, por *comportamiento con el cliente* (4), definió al grado en que el sistema parecía similar a las expectativas del usuario. Para ello, recomienda usar términos familiares, palabras actuales, tener un discurso organizado y lógico, hablar a un ritmo comprensible, ser educado, cortés, amistoso y profesional. Según Polkolski, todos estos aspectos están correlacionados con la satisfacción del usuario de las IH.

Conclusiones: retos en el diseño de las IH

Además del avance en el desarrollo de sistemas capaces de comprender habla con alto grado de confianza y variabilidad o de generar voces sintéticas naturales, el mayor reto con el que se enfrenta el diseño de las "máquinas que hablan" es reconocer eficazmente la emoción en la voz humana e imitarla en las voces sintéticas que producen. Aunque se han hecho grandes avances en ambos terrenos, todavía se carece de modelos capaces de adaptarse satisfactoriamente con la variabilidad sintáctica o semántica que las emociones humanas introducen en la codificación y decodificación del lenguaje. En este esfuerzo convergen, necesariamente, los aportes de variadas disciplinas. Por ejemplo: es preciso conocer cómo procesamos los humanos las emociones (tarea de la psicología), o cómo afectan a los ritmos, pausas y sonidos de nuestro habla (de la fonética), y qué señales del habla emocionada escuchamos (acústica).

El interés por dominar la producción y captación de emociones habladas, además, ha estimulado la imaginación de los científicos quienes ya trabajan en el diseño de técnicas que minimicen la angustia y el

estrés de los receptores mediante la interacción hablada (por ejemplo, comunicando sus emociones a las interfaces). Para ello, se prosigue en la creación de los llamados *ordenadores emocionales*, los que parecen inteligentes y dignos de confianza (Picard, 1997)[28]. En el contexto mediático, asimismo, existe interés porque se sabe que la aceptación y el uso de la tecnología se vincula a las motivaciones. Se cree entonces que detectándolas y creándolas se podrá manipular la satisfacción de las audiencias. Y ello porque, tal y como se ha demostrado ampliamente en los trabajos asociados al paradigma de los usos y gratificaciones, se ha contrastado en abundancia que las emociones justifican y explican la satisfacción del consumo mediático de las audiencias (Blumer, 1979; Katz, Blumler y Gurevitch, 1974; Katz y Lazarsfeld, 1955; Perse y Courtright, 1993; Swanson, 1977).

El presente trabajo ha tenido como ambicioso objetivo dar cuenta de las diferentes aristas del fenómeno de las IH de una manera sencilla en la forma (comprensible y divulgativa), pero justificada académica y científicamente. Por ello, ha aportado abundantes evidencias, reflexiones y datos. Dicha revisión nos permite predecir que tanto por la vertiginosa evolución tecnológica como por el interés que despierta el tema, los aspectos aportados en las últimas líneas serán superados muy rápidamente, en un futuro muy próximo. Además, creemos, que la industria del entretenimiento tendrá un papel central porque será el vehículo por el que las nuevas generaciones se entrenen en la interacción fluida, normalizada y popularizada con las máquinas. Hoy día convivimos naturalmente con la interacción y aceptamos su irrealidad cuando la máquina hablante es coherente en la manifestación de comportamientos comunicativos. Lograr similitud, empatía e identificación con ellas es sólo una cuestión de hábito (de superar las barreras, prejuicios y desconocimientos iniciales). Creemos que muy pronto importará poco hablar indistintamente con hombres o máquinas si las segundas se ajustan a nuestras necesidades de compañía, información, entretenimiento o afecto. Tiempo al tiempo. Poco.

Referencias

ABELE, A. E., PETZOLD, P. (1998). Pragmatic use of categorical information in impression formation. *Journal of personality and social psychology, 75*(2) 347-358.

ASHMORE, R. D. (1981). Sex stereotypes and implicit personality theory. En D. L. Hamilton (ed.), *Cognitive processes in stereotyping and intergroup behavior* (p. 37-82). Hillsdale, NJ: Lawrence Erlbaum Associates.

BARNES, S.B. (2001). *Online connections: Internet interpersonal relationships*. Cresskill, NJ: Hampton press.

BARNES, S.B. (2003). *Computer-mediated communication: Human-to-human communication across the Internet*. Boston, MA: Allyn & Bacon.

BARNES, S.B., STRATE, L. (1996). The educational implications of the computer: a media ecology critique. *New Jersey journal of communication, 4*(2), 180-208.

BASOW, S. A., SILBERG, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of educational psychology, 79* (3), 308-314.

BLUMLER, J.G. (1979). The role of theory in uses and gratifications studies. *Communication research, 6,* 9-36.

BRENNEN, S. (1998). The grounding problem in conversations with and through computers. En S. Fusell y R. Kreuz (eds.), *The new handbook of language and social psychology* (pp. 601-623). New York, NY: John Wiley & Sons.

BIOCCA, F. (2001). Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleoperators And Virtual Environments,10*, 247–265.

BURGOON, M. (1990). Language and social influence. En H. Giles y W.P. Robinson (eds.), *Handbook of language and social psychology* (pp. 51-72). Chichester, UK: John Wiley & Sons.

COLLINS, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*(6).

COSMIDES, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187-276.

COWIE, R., DOUGLAS-COWIE, E., TSAPATSOULIS, N., VOTSIS, G., KOLLIAS, S., FELLENZ, W. TAYLOR, J.G. (2001). *Emotion recognition in human-computer interaction*. IEE Signal processing magazine.

COOPER, A. (1995). *About face: the Essentials of user interface design*. Boston: Adison-Wesley.

DENNETT, D. C. (1988). Précis of the intentional stance. *Behavioral and brain sciences, 11*, 495-546.

EAGLY, A., CHAIKEN, SH. (1975). An attribution analysis of the effect of communicator characteristics on opinion change: The case of communicator Attractiveness. *Journal of personality and social psychology, 32*(1), 136-144.

EAGLY, A. H., WOOD, W. (1982). Inferred sex differences in status as determinant of gender stereotypes about social influence. *Journal of personality and social psychology, 43*(5), 915-928.

FINKEL, S. E., GUTERBOCK, T. M., Borg, M. J. (1991). Race-of interviewer effects in a preelection poll: Virginia 1989. *Public opinion quarterly, 55*(3), 313-330.

FOGG , B.J., NASS, C. (1997). How users reciprocate to computers: an experiment that demonstrates behavior change. *CHI '97 extended abstracts on Human factors in computing systems: looking to the future*. March 22-27, Atlanta, Georgia.

FOX, J., BRADLEY, S.S., WANG, Z. (2006). Emotional context and typicality in encoding and reality assessment of television scenarios. *Artículo presentado en el encuentro anual de la Asociación Internacional de Comunicación* (Internacional Communication Association), Dresden, Alemania.

GILES, D. (2002). Parasocial interaction: A review of the literature and a model for future research. *Media psychology, 4,* 279-305.

GREEN, N. (1993). Can computers have genders? Artículo presentado en el encuentro anual de la *Asociación Internacional de Comunicación (Internacional Communication Association)*. Washington, DC.

HATON, J-P., CERISARA, CH., FOHR, D., LAPRIE, Y., SMAÏLI, K. (2006). *Reconnaissance automatique de la parole. Du signal à son interprétation*. Paris: Dunod.

HEIDEGGER, M. (1977). *The question concerning technology, and other essays*. New York, NY: Harper & Row.

HORTON, D., WOHL, R. R. (1956). Mass communication and para-social interaction: Observation on intimacy at a distance, *Psychiatry, 19*, 215-229.

HOULBERG, R. (1984). Local television news audience and the para-social interaction. *Journal of broadcasting, 28*, 423-429.

HUANG, X., ACERO, A., HON, H-S. (2001). Spoken language processing. Upper Saddle River, NJ: Prentice Hall.

JONES, E. E. (1964). *Ingratiation: A social psychological analysis*. New York, NY: Meredith Publishing Company.

KANE, E. W., MACAULAY, L. J. (1993). Interviewer gender and gender attitudes. *Public opinion quarterly, 57*(1), 1-28.

KATZ, E., BLUMLER, J.G.,GUREVITCH, M. (1974). Utilitization of mass communication by the individual. En J.G. Blumler y E. Katz (eds.), *The use of mass communication* (pp.19-32). London, UK: Sage.

KATZ, E., LAZARSFELD, P. F. (1955). *Personal influence: The part played by people in the flow of mass communication*. Glencoe, IL: Free Press.

KOTELLY, B. (2003). *The art and business of speech recognition: creating the noble voice*. Boston, MA: Addison Wesley.

LAI, J., YANKELOVICH, H. (2003). Conversational speech interfaces. En J. Jacko y A. Sears (eds.), *The human-computer interaction handbook: Fundamentals, evolving Technologies and emerging applications* (pp. 698-713). Mahwah, NJ: Lawrence Erlbaum.

LANG, A. (2000). The limited capacity model of mediated meesage processing. *Journal of Communication, 50*, 46-70.

LANG, A., SPARKS, J.V., BRADLEYN S.D., LEE, S., WANG, Z. (2004). Processing arousing information: Psychophysiological predictors of motivated attention. *Psychophysiology, 41*(1), S61.

LASSWELL, H.D. (1948). The structure and function of communication in society. En L. Bryson (ed.), *The communication of ideas* (pp. 37-52). Nueva York, NY: Harper and Brothers.

LEE, K. M. (2004). Presence, explicated. *Communication Theory, 14, 27–50.*

LEE, K., NASS, C. (2004). The multiple source effect and synthesized speech. *Human communication research, 30*(2), 182-207.

LEE, E., NASS, C., BRAVE, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotype. *Extended Abstracts of CHI'00: Conference on Human Factors in Computing Systems*, 289-290.

LEVINSON, P. (1997). *The soft edge*. New York, NY: Routledge.

LEVINSON, P. (1999). *Digital McLuhan*. New York, NY: Routledge.

LLISTERRI, L. (1988). La síntesis del habla: estado de la cuestión. *Procesamiento del lenguaje natural, 6*, 17-41. Disponible en http://liceu.uab.es/~joaquim/publicacions/Llisterri_88_Sintesis.pdf

LOMBARD, M., DITTON, T. (1997). At the heart of it all: The concept of presence. *Journal of computer mediated communication, 3*. Obtenido el 26 de noviembre de 2008 desde: http://209.130.1.169/jcmc/vol3/issue2/lombard.html

MCQUAIL, D., BLUMLER, J., BROWN, R. (1972). The television audience: a revised perspective. En: D. McQuail (ed.), *Sociology of mass communication*. London, UK: Longman.

MEYROWITZ, J. (1985). *No sense of place*. New York: Oxford University Press.

MOON, Y., NASS, C. (1996). How real are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication research, 23,* 651-674.

NASS, C., LEE, K.M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology: Applied, 7*(3), 171-181.

NASS, C., MOON, Y., GREEN, N. (1997). Are computers gender neutral? Gender stereotypic responses to computers. *Journal of applied social psychology, 27*, 864-876.

NASS, C., SUNDAR, SH. (1994). Is human-computer interaction social or parasocial? Obtenido el 25 de noviembre de 2008, desde http://www.stanford.edu/group/commdept/oldstuff/srct_pages/Social-Parasocial.html

NIELSEN, J. (1993). *Usability engineering*. Boston, MA: Academic Press.

NUSBAUM, H.C., SHINTEL, H. (2006). Speech synthesis. En K. Brown (ed.), *Encyclopedia of language & linguistics* (pp. 19-31). Amsterdam: Elsevier.

PAVITT, CH., HAIGHT, L. (1985). The 'competent communicator' as a cognitive prototype. *Human communication research, 12,* 225-241.

PAVITT, CH., HAIGHT, L (1986). Implicit theories of communicative competence: situational and competence lever differences in judgements of prototype and target. *Communication monographs, 53*(4), 221-236.

PERSE, E.M., COURTRIGHT, J.A. (1993). Normative images of communication media: Mass and interpersonal channels in the new media environment. *Human communication research, 19*(4), 485-503.

PICARD, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.

POLKOLSI, M.D. (2007). Machine as mediators. En E. Konijn, S. Utz, M. Tanis y S. B. Barnes (eds.), *Mediated interpersonal Communications* (pp. 34-57). New York, NY: Routledge.

POSTMAN, N. (1995). *The end of education*. New York, NY: Alfred. A. Knopf.

REEVES, B., NASS, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York, NY: Cambridge University Press.

RUBIN, A.M. (1985). Uses of daytime television soap operas by college students. Journal of Broadcasting & Electronic Media, 29, 241-258.

RUBIN, A. M., PERSE, E. M. (1987). Audience activity and television news gratifications. *Communication research, 14*, 58-84.

RUBIN, A. M., RUBIN, R. B. (1985). Interface of personal and mediated communication: A research agenda. *Critical studies in mass communication, 2*, 36-53.

SCHERER, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech communication, 40*(1-2), 227-256.

SEARLE, J. R. (1981). Minds, brains, and programs. En D. R. Hofstadter y D. C. Dennett (eds.), *The mind's I* (pp. 353372). Toronto, CAN: Bantam.

STRATE, L. (1999). The varieties of cyberspace: problems in definition and delimitation. *Western journal of communication, 63*(3), 382-412.

SOTO, M.T. (2000). *Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz*. Tesis doctoral. Bellaterra: Departamento de Comunicación audiovisual y Publicidad. Universidad Autónoma de Barcelona.

SOTO, M.T. (2008). Efecto del tono de voz y de la percepción del rostro en la formación de impresiones sobre los hablantes mediáticos. *Comunicación y sociedad, 10,* 129161.

SOTO, M.T. (2008a). Impresiones sobre los hablantes mediáticos a partir de la profesionalidad en su elocución y el contenido de su discurso. *Signo y pensamiento, 53* (27), 246-266.

SUNDAR, S., NASS, C. (2000). Source orientation in human-computer interaction – Programmer, networker, or independent social actor? *Communication research, 27*(6), 683-703.

SWANSON, D. (1977). The uses and misuses of uses and gratifications. *Human communication research, 3*, 214-221.

TOMKINS, S. (1962). *Affect imagery consciousness: Vol 1. Positive affects.* New York, NY: Springer.

TOMKINS, S. (1963). *Affect imagery consciousness: Vol II. Negative affects.* New York, NY: Springer.

TOMKINS, S. (1982). Affect theory. En P. Ekman (ed.), *Emotions in human face*. New York, NY. Cambridge university press.

TURING, A. (1950). Computing machinery and intelligence. *Mind, 59,* 433-460.

VORDERER, P., KLIMMT, G., RITTERFELD, U. (2004). Enjoyment: At the heart of media entertainment. *Communication theory, 4,* 388-408.

WRIGHT, CH. (1959). *Mass communication. A sociological perspective*. Nueva York, NY: Random house.

YEGIYAN, N., BRADLEY, S.D., LANG, A. (2005). Approach or avoid: How motivation type affects the processing of risky information. *Artículo presentado en el encuentro anual de la Asociación internacional de comunicación* (Internacional communication association). New York, USA.

ZUCKERMAN, M., MIYAKE, K. (1993). The attractive voice: What makes it so?, *Journal of nonverbal behavior, 27*(2), 119-135.

Cita de este artículo
Soto Sanfiel, M. T. (2009) Interfaces habladas. Caracterización, usos y diseño. *Revista Icono14 [en línea] 30 de 11 de 2009, Nº 13.* pp. 310-333. Recuperado (Fecha de acceso), de http://www.icono14.net

[1] Por ejemplo: ¿cómo afecta la posibilidad de relacionarse con máquinas a la comunicación entre humanos?, ¿qué nuevas pautas de comunicación surgen por la existencia de esos nuevos interlocutores?, ¿qué influencia tiene el desarrollo de máquinas sobre la comunicación interpersonal?, ¿qué buscamos los seres humanos creando máquinas con las que hablar?, o ¿hacia dónde nos lleva esta carrera tecnológica de emulación humana mediante máquinas? También, surgen otras preguntas de índole más práctico: ¿cuáles características humanas se escoge atribuir a las máquinas en su diseño y cuáles se desestiman?, ¿qué herramientas se utilizan?, o ¿hasta dónde podemos llegar en esos intercambios con el conocimiento disponible? Sin duda, estos son campos para abonar con la imaginación (y la investigación).

[2] También existen aplicaciones destinadas a personas con minusvalías o discapacidades visuales y que leen textos escritos en voz sintética, traducen o permiten controlar ordenadores o equipos. Para mayor información visite http://www.lighthouse.org/accessibility/accessible-technology/speech-solutions/ y asímismo http://www.cforat.org/main_page/assisitve_technology_links.htm

[3] Nos referimos a la traducción del inglés del término (y sus siglas): *Interactive Voice Response.*

[4] Véase, por ejemplo, el conjunto de aplicaciones *Microsoft Agent*, que permite crear interfaces de diálogo humanizadas (para lo que, en buena parte, usan personajes sintéticos), en http://www.microsoft.com/msagent/prodinfo/

[5] Para el dictado general, existe una amplia oferta de programario. Algunos de los comerciales más populares son *Dragon's Naturally Speaking* (http://www.nuance.com/naturallyspeaking), *Lernout & Hauspier* (http://www.voicerecognition.com/1998/products/lernout_hauspie/voicexpressplus.html) y *Via Voice* de IBM (http://www.dif.gob.mx/cta/soluciones/viavoice.htm)

[6] Acrónimo de *Global Postioning System*.

[7] Nos referimos a los progresos adquiridos por dos áreas de investigación en tecnologías de la voz cuya preocupación es la captación y reproducción sintética de la voz humana. Por una parte, a los avances sobre *reconocimiento automático del habla,* una rama de la inteligencia artificial que pretende proponer sistemas de

captación e interpretación de los mensajes hablados. Por otra, a los desarrollos de la *síntesis del habla*, un campo que persigue producir habla humana artificial inteligible con apariencia de naturalidad.

[8] Más conocidos por su nombre en inglés: *Text to Speech* (TTS).

[9] Acrónimo del nombre inglés: *Application Programming Interface*.

[10] No obstante, el progreso en ambas actividades se debe a la incorporación de las evidencias aportadas por un gran número de disciplinas científicas y técnicas que exploran el habla en sus distintas facetas (vg. lingüística, acústica, psicología, informática, ingenierías…)

[11] Acrónimo del nombre inglés: *Speech Aplication Language Tags*. Consúltese: http://msdn.microsoft.com/en-us/library/ms994629.aspx

[12] Véase: http://www.voicexml.org

[13] Consúltese: http://msdn.microsoft.com/en-us/library/ms723627(VS.85).aspx

[14] Acrónimo del nombre inglés: *Java Speech Application Programming Interface*. Véase http://java.sun.com/products/java-media/speech

[15] Acrónimo del inglés: *MultiModial Interface Language*. Visítese: http://ralyx.inria.fr/2003/Raweb/led/id2643443.html

[16] Consúltese: http://www.chiariglione.org/mpeg

[17] Polkolski (2007) nos recuerda que, en general, los estudios de comunicación interpersonal han partido de la base de que: 1.- los dos intervinientes son personas; 2.- de que es una forma de comunicación con especificidades, lo que la hace distinta a otras formas de comunicación (mediada, intrapersonal, impersonal), y 3.- que el principal reto de la comunicación es la creación o mantenimiento de las relaciones. Según el investigador, estas consideraciones han impedido que se cruzaran y estimularan las perspectivas de esos ámbitos.

[18] Teoría de la ecuación de los medios (*Media Equation Theory*).

[19] La idea de que la tecnología es el reflejo de sus creadores es una idea defendida por algunos cognoscitivistas (Dennett, 1988; Heidegger, 1977). Específicamente, algunos investigadores defienden que es una representación de sus creadores (Cosmides, 1989; Dennett, 1988, 1991; Searle, 1981).

[20] Lo inmediatamente apuntado ha sido explorado con profusión en el marco de los estudios del lenguaje. Como todo sistema simbólico, el habla tiene reglas y expectativas de uso, además de demostrar el vínculo de los individuos participantes en los actos comunicativos. Esto es: los seres humanos esperan percibir un tipo de lenguaje adecuado (generado en praxis comunicativas previas) para cada situación y reaccionan de distintas maneras según sea la violación de sus prototipos cognitivos. Burgoon (1990) en el marco de la teoría de la expectativa del lenguaje, aisló las reacciones de los receptores a la actitud o comportamiento del comunicador: 1.- una violación positiva de las expectativas del habla, por un comportamiento mejor del esperado, provocan una actitud a favor de la fuente; 2.- una violación positiva por una fuente previamente negativa conduce a una actitud positiva hacia la fuente, y 3.- una violación negativa, por un comportamiento negativo, no afecta la actitud de la fuente.

[21] El término fue acuñado por Horton y Wohl (1956) para describir el fenómeno de la interacción en el que una de las partes sabe mucho de otra y esta nada de la primera. El fenómeno se genera, típicamente, entre los personajes mediáticos y los televidentes. Las audiencias desarrollan una relación de cercanía con dichos personajes, hasta el punto de que pueden considerarse sus amigos. La identificación del fenómeno por Horton y Wohl estimuló la aparición de otros estudios, los que, a su vez, han propuesto otras definiciones (Houlberg, 1984; Rafaeli, 1990; Rubin y Pere, 1987; Rubin y Rubin, 1985). No obstante, todos coinciden la idea de que las

relaciones parasociales ocurren cuando los individuos interactúan con una representación mediática de una persona como si la persona estuviera realmente presente (Nass y Sundar, 1994).

[22] El tono o frecuencia fundamental es el parámetro acústico de la voz más estudiado y el que se considera que mejor predice las actitudes de los perceptores respecto a una fuente. Por eso, y por razones de espacio, nos referimos únicamente a él en este trabajo. No obstante, existen otros parámetros también estudiados e influyentes: la velocidad de la elocución, la intensidad o amplitud de la voz, el timbre, el uso de pausas, etc. Véase a Scherer (2003).

[23] El nombre que, en inglés, tienen estos textos o instrucciones es *prompt*.

[24] Nos referimos al principio cooperativo del filósofo Paul Grice quien estableció estas reglas de pragmática conversacional basadas en la cortesía.

[25] El método está inspirado en la filosofía del *diseño centrado en el usuario*. Véase a Nielsen (1993) o Cooper (1995).

[26] Huang, Acero y Hon (2001).

[27] Citada en Polkolsi, M.D. (2007). Ver referencias.

[28] En inglés ya existe un nombre para esta área de investigación: *affective computing* ("computación afectiva").