

# Usability Questionnaires to Evaluate Voice User Interfaces

A. L. Iñiguez, L.S. Gaytán, M.A. García-Ruiz and R. Maciel

**Abstract**— Voice user interfaces (VUI) have been increasingly used in everyday settings and they are growing in popularity. These interfaces have predominantly eyes-free and hands-free interactions. This kind of experiences continues to be an inceptive field compared to other input methods such as touch or using the keyboard/mouse. Thus, it is important to identify tools used to evaluate the usability of VUIs. This article presents a systematic review, in which we analyzed 57 articles and describes nine questionnaires used for evaluating the usability of VUIs, assessing the potential suitability of these questionnaires to measure different types of interactions and various usability dimensions. We found that these questionnaires were used to evaluate the usability of voice-only and voice-added VUIs: AttrakDiff, ICF-US, MOS-X, SUIQ-R, SUS, SASSI, UEQ, PARADISE and USE, where the SUS questionnaire is the most commonly used. However, its items do not directly assess voice quality, although it evaluates the general user interaction with a system. All the questionnaires include items related to three usability dimensions (effectiveness, efficiency, and satisfaction). The questionnaire with the most homogeneous coverage regarding the number of items in each aspect of usability is the SASSI questionnaire. It is a normal practice to use multiple questionnaires to obtain a more complete measurement of usability. We perceive the necessity to increase usability research about the differences between the voice interaction with diverse display types (voice-first, voice-only, voice-added) and the dialog types (command-based and conversational), and how usability affects the user expectations about the VUIs.

**Index Terms**— Questionnaires, Usability, Voice User Interfaces.

## I. INTRODUCCIÓN

Cada vez son más las personas que aprecian el potencial de una interfaz de usuario de voz (VUI, por sus siglas en inglés), utilizando reconocimiento automático del habla, síntesis de voz y procesamiento de lenguaje natural para interactuar a través del habla con un sistema computacional. Si bien los sistemas de respuesta de voz interactiva (IVR, por sus siglas en inglés) y los sistemas de dictado han existido durante mucho tiempo, la implementación de los asistentes de voz en un teléfono móvil (Siri, Cortana) o en un altavoz inteligente (Alexa, Google Assistant) han contribuido a que la interacción

con voz de un usuario y un sistema automatizado sea una tecnología cada vez más habitual. Existe una creciente adopción de las tecnologías basadas en voz, tan sólo en Estados Unidos de América casi 90 millones de adultos tiene un altavoz inteligente en su hogar y alrededor del 50% dice que usa los dispositivos todos los días [1].

La interacción con voz es extraordinaria cuando ofrece una forma más rápida, fácil y placentera de hacer las cosas. Una forma de verificarlo es aplicando evaluaciones de usabilidad. Sin embargo, las técnicas de usabilidad tradicionales pueden no ser aplicables en la forma en que fueron concebidas debido a las características particulares de utilizar el habla como medio de interacción, como la ausencia de retroalimentación visual, periféricos o interacción estilo WIMP (ventanas, íconos, menús y puntero). Utilizar cuestionarios es uno de los métodos más populares en la evaluación de la usabilidad, pues resultan sencillos de aplicar y son menos costosos [2]. Estos suelen utilizarse para recopilar las percepciones de los usuarios y para identificar fallas de usabilidad [3]. Se han publicado estudios sobre la evaluación de la usabilidad en VUIs, como el del Holmes et al. [4] donde se hace una revisión de los métodos convencionales para evaluar las interfaces de usuario conversacionales el cual está enfocado principalmente a chatbots. Kouroupetoglou & Spiliotopoulos [5] presentan un estudio sobre los métodos y enfoques de evaluación de la usabilidad para interfaces web basadas en voz. Sin embargo, no se detectan revisiones sistemáticas sobre el uso de cuestionarios para evaluar la usabilidad en VUIs que utilicen la voz como medio de interacción.

Por lo anterior, resulta importante identificar los cuestionarios que se han utilizado para evaluar la usabilidad en VUIs, entender sus diferencias, reconocer sus beneficios y deficiencias para seleccionar el más adecuado dependiendo las características del sistema a evaluar. Por medio de una revisión sistemática, se proporciona una síntesis de estudios científicos relevantes para identificar los cuestionarios que se utilizan para evaluar la usabilidad en interfaces que utilizan la voz como medio de interacción. Se hace una diferenciación entre los diferentes tipos de diálogo y los tipos de interacción con VUIs, se establecen las dimensiones de usabilidad que abordan estos cuestionarios, así como los retos que quedan por superar.

Este artículo está organizado de la siguiente manera. En la sección II y III se describen las interfaces de voz y la evaluación de usabilidad y cómo se han realizado en este tipo de interfaces. En la sección IV se describe la metodología utilizada y búsqueda de la literatura. En la sección V se presentan los resultados de la revisión, se muestran los cuestionarios que se

February 4, 2021.

A. L. Iñiguez is with Universidad de Guadalajara, Cd. Guzmán, JAL 49000 MEX (e-mail: adriana.iniguez@academicos.udg.mx).

L. S. Gaytán, is with Universidad de Colima, Coquimatlán, COL 28400 MEX. (e-mail: laura@uacol.mx).

M. A. García-Ruiz is with Algoma University, Sault Ste. Marie, ON P6A 2G4 CAN (e-mail: miguel.garcia@algonau.ca).

R. Maciel, is with Universidad de Guadalajara, Zapopan, JAL 45100 MEX. (e-mail: rmaciel@cucea.udg.mx).

utilizan para evaluar la usabilidad en VUIs y sus características. En la sección VI se hace una reflexión sobre los hallazgos y los desafíos por superar en la evaluación de la usabilidad en las VUIs. Finalmente, se enumeran las principales conclusiones en la sección VII.

II. INTERFACES DE VOZ

Una VUI es una interfaz que permite la interacción entre un humano y un sistema informático, en donde el diálogo hablado se utiliza para desarrollar procesos o tareas específicas [6]. Algunos de los beneficios de este tipo de interacción son que los usuarios pueden realizar otras actividades a la par que solicitan un proceso a la interfaz de voz, por ejemplo, estar manejando y solicitar que se realice una llamada telefónica sin necesidad de quitar las manos del volante. Asimismo, contribuye a la accesibilidad, ya que usuarios con alguna discapacidad visual o motriz pueden utilizar la voz para solicitar un proceso al sistema de voz. El uso de VUIs es conveniente ya que realizar una conversación en lenguaje natural, resulta más intuitivo y se requiere de menor capacitación que en otros tipos de sistemas [7]. Puede implementarse en diferentes dispositivos electrónicos, como teléfonos inteligentes, altavoz inteligente, reloj inteligente, rastreador de actividad física, computadora personal, sistema doméstico inteligente, o automóviles [8][9][10].

Las VUIs pueden clasificarse por su tipo de diálogo en basado en “comandos” y basado en “conversación” [6]. En los basados en comandos el sistema es capaz de reconocer y procesar el lenguaje, por ejemplo, indicar a un dron que suba, baje, gire a la izquierda o derecha. Mientras que uno basado en conversación puede establecer un diálogo, sostenerlo por varios turnos y guardar el contexto de comandos anteriores, resultando una interacción más natural, similar a la interacción humano-humano [11]. También se pueden clasificar por su tipo de interacción en sólo voz, primero voz y voz agregada [12]. En las interacciones con sólo voz no se proporciona una retroalimentación visual, por ejemplo, el uso altavoces inteligentes. En la interacción de primero voz, el habla es el principal medio de interacción, sin embargo, se da al usuario una retroalimentación visual, los dispositivos Echo Show y Google Home Hub son ejemplos de este tipo de interacción. Las interacciones con voz agregada, se refiere a teléfonos inteligentes, tabletas y pantallas con la adición de control por voz sistemas, sin ser la interacción de voz necesaria sino una opción.

III. EVALUACIÓN DE USABILIDAD

La usabilidad es un atributo de calidad, la norma ISO 9241-11 de la Organización de Estándares Internacionales (ISO) la define como "la medida en que un producto puede ser utilizado por usuarios específicos para lograr objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso específico" [13]. La efectividad es la capacidad de conseguir resultados esperados, por ejemplo, la tasa de finalización de tareas exitosas. La eficiencia es el esfuerzo requerido, por ejemplo, el tiempo y los turnos de conversación que se realizan

para completar un objetivo. La satisfacción se refiere a aspectos emocionales del usuario, donde se toma en cuenta la comodidad, el placer, la utilidad y la aceptabilidad del sistema. Esta definición de usabilidad aplica a cualquier sistema computacional, sin embargo, en las VUIs, existen desafíos que hay que tomar en cuenta para tener una buena usabilidad (Tabla I). La comunicación debe realizarse con diálogos claros y simples, de preferencia en pequeñas conversaciones para minimizar la carga cognitiva del usuario, ya que debe prestar atención a la conversación y recordar fragmentos de información. Así como proporcionar recuperación de errores, es decir, el sistema siempre debe de brindar al usuario una respuesta. Al no contar con una retroalimentación por una pantalla, es importante que el sistema proporcione confirmaciones auditivas (tono corto o bip) o visuales (alguna señal de luz en el dispositivo) que indique al usuario que está siendo escuchado o se está procesando una solicitud. Una interfaz de voz tiene como objetivo hacer una tarea más rápida y fácil de usar, por lo que el usuario debe tener claro lo que se puede realizar con la interfaz de voz de una manera que se sienta cómodo y sea fácil de aprender a usar. A su vez el sistema debe proporcionar al usuario una sensación de consistencia, control y confianza para que no se sienta perdido durante la interacción con el sistema [14]. Diversos sistemas que manejan VUIs cuentan con servicios de procesamiento en la nube, por lo que la rapidez de respuesta del sistema se puede ver afectada por una mala conectividad en la red de datos. No es lo mismo tener un retardo en una interfaz con retroalimentación visual, que en una VUI donde se espera una respuesta inmediata y mantener una conversación sin pausas. También es importante que el sistema sea capaz de interpretar diferentes maneras de decir una misma sentencia para interpretar de manera efectiva los diálogos de una conversación. Por ejemplo: quiero abrir mi correo, deseo abrir mi correo, abre mi correo, etc. Se considera que el campo de las técnicas y medidas de evaluación de este tipo de sistemas se encuentra en una fase inicial de desarrollo [15].

La evaluación de usabilidad puede ser formativa o sumativa dependiendo del momento en que se realiza la evaluación. La evaluación formativa se centra en los problemas de usabilidad

TABLA I  
DESAFÍOS PARA MEJORAR DE LA USABILIDAD EN VUIs

Desafío	Soluciones propuestas
Diálogos claros y simples	Minimizar la carga cognitiva.
Recuperación de errores	Brindar siempre una respuesta.
Retroalimentación	Dar confirmaciones de que el mensaje fue escuchado.
Tiempo de respuesta	Optimizar la resiliencia y respuesta del sistema.
Intuitivo	Fácil de aprender a usar, saber que palabras utilizar para interactuar con el sistema o iniciar una conversación.
Flexible	Soporta diferentes maneras de solicitar el inicio de un proceso.
Consistente	Evitar utilizar cambios en el tipo o tono de voz que sean innecesarios.

que deben resolverse durante la etapa de diseño del prototipo antes de que un diseño final pueda ser aceptado para su lanzamiento y una evaluación sumativa para evaluar la eficacia del diseño final [16]. También existen medidas objetivas y subjetivas, la evaluación objetiva se basa en la captura de datos analíticos sin interacción directa con los usuarios, los métodos de evaluación subjetiva se centran en capturar las actitudes y los juicios de los usuarios en la usabilidad percibida [3]. Uno de los métodos de evaluación subjetiva más utilizados son los cuestionarios, los cuales permiten recopilar datos cuantitativos y realizar análisis estadísticos [17]. Los cuestionarios de usabilidad estandarizados tienen mayor confiabilidad que los cuestionarios de usabilidad personalizados o ad hoc. Además, que son sencillos de aplicar, de replicar, económicos, incrementan la objetividad, y logran una comunicación más efectiva entre investigadores [18].

#### IV. MÉTODO

El objetivo de esta revisión sistemática es identificar de qué manera se evalúa la usabilidad utilizando cuestionarios. Se establecieron tres preguntas de investigación: ¿cuáles son los cuestionarios que se han utilizado para evaluar la usabilidad en interfaces de voz? (Q1), ¿en qué tipo de VUIs han sido utilizados? (Q2) y ¿cuáles son las dimensiones de la usabilidad que se incluyen en las categorías de los cuestionarios? (Q3).

Esta revisión se basa en la guía propuesta por Barbara Kitchenham [19], la cual cuenta con las etapas: identificación de la investigación, selección y evaluación de la calidad de los estudios, extracción de datos y síntesis.

##### A. Identificación de la Investigación

Se utilizó el instrumento PICO para la estrategia de búsqueda bibliográfica [20], el cual estructura la búsqueda identificando los elementos: *población (P)*, referido al grupo particular en el que se realiza la búsqueda, en este caso la población son las VUIs; *intervención (I)*, que es lo que se está haciendo para la población, en este caso la evaluación de usabilidad; *comparación (C)*, que representa el dominio de aplicación, en el cual son investigaciones aplicadas en interfaces con diálogo por comandos o conversacionales, así como por su tipo de interacción voz agregada, solo voz y primero voz, y *resultados (O, outcome)*, haciendo referencia al resultado esperado, que para esta revisión es encontrar cuestionarios que se utilizan para evaluar VUIs. La investigación se realizó en bases de datos digitales que contienen datos científicos y especializados, como IEEE Explore, ACM Digital, SPRINGER, SCOPUS y Web Of Science (WoS). Elegir los términos adecuados aplicados en el

TABLA II  
ELEMENTOS DEL INSTRUMENTO PICO

Acronimo	Encabezado	Elementos
P	Población	Interfaces de Voz
I	Intervención	Evaluación de Usabilidad
C	Comparación	Tipo de diálogo, Tipo de interacción
O	Salida	Cuestionarios

dominio de las interfaces de voz es un desafío, ya que se utilizan diferentes términos tales como bocina inteligente, asistente personal, asistente inteligente, sistema de habla, por mencionar algunos. Con los elementos del modelo PICO de la tabla II se crean los términos de búsqueda tomando en cuenta términos semejantes como se muestra en la tabla III. La búsqueda se realizó en inglés y en español, considerando que algún término de los mencionados en el elemento 1 se mencionara en el resumen del artículo y que los elementos 2 y 3 se mencionaran en el documento. Considerando artículos presentados en congresos o revistas científicas y académicas publicados durante el período enero de 2013 a julio 2020.

TABLA III  
TÉRMINOS DE BÚSQUEDA

Elementos	Equivalencias Español	Inglés
1 Interfaces de Voz / Voice User Interfaces	bocina inteligente, asistente personal, asistente inteligente, sistema de habla, interfaz conversacional	VUI, CUI, speech user interface, spoken system, dialogue systems, voice interactions, voice assistant, smart speaker, personal assistant, conversational interface
2 Evaluación de Usabilidad / Usability Evaluation	Test de usabilidad	Usability Assessment
3 Cuestionarios / Questionnaires	-	Test

##### B. Selección y Calidad de Estudios

Como resultado de las condiciones de búsqueda (Tabla IV) se encontraron 171 trabajos científicos. En un primer filtro se omitieron trabajos repetidos y se seleccionaron los trabajos presentados en revistas indexadas, revistas arbitradas y publicaciones de congresos. En un segundo filtro se revisaron los títulos, resúmenes y conclusiones. Se excluyeron los trabajos donde la interacción no se realiza por medio de la voz, como las interfaces conversacionales basadas solo en texto. Las interfaces conversacionales multimodales que tienen la opción de interactuar por medio de la voz sí fueron incluidas.

TABLA IV  
RESULTADOS DE LOS TÉRMINOS DE BÚSQUEDA

Base de datos	Filtro 1	Filtro 2	Filtro 3
ACM	89	48	17
EBSCO	6	4	2
IEEE Xplore	28	20	11
SCOPUS	6	5	4
SPRINGER	39	23	22
WOS	3	1	1
	171	101	57

Se incluyeron estudios que evalúan la usabilidad a través de cuestionarios, varios de los estudios incluidos utilizan a la par

de los cuestionarios otros métodos para evaluar la usabilidad como heurísticas, recorridos cognitivos, la técnica del Mago de Oz, por mencionar algunos. Finalmente, se seleccionaron 57 artículos, los cuales fueron leídos en su totalidad buscando respuestas a cada una de las preguntas de la investigación.

### C. Extracción y Síntesis de Datos

Se definió un formulario de extracción de datos para registrar con precisión la información que contribuye a dar respuesta a las preguntas de la investigación: nombre del trabajo, cuestionario utilizado, tipo de dialogo, tipo de interacción. Los estudios obtenidos se clasificaron según las preguntas de investigación: sus posibles respuestas, el número de estudios y el porcentaje de estudios que respondieron a cada una de las preguntas de investigación. Posteriormente, se realizó una clasificación de los cuestionarios, identificando sus características generales: número de ítems, escala psicométrica, categorías que evalúa y tipo de interfaz. Se identificó en qué tipo de interfaces se han utilizado los cuestionarios, diferenciando por el tipo de diálogo (comando o conversación) y por su tipo de interacción (sólo voz, primero voz o voz agregada). Finalmente, se realizó un análisis de las dimensiones de usabilidad que se toman en cuenta en los cuestionarios, con el objetivo de identificar la proporción de cada dimensión de la usabilidad en cada instrumento.

## V. RESULTADOS

### A. Cuestionarios Utilizados para Evaluar la Usabilidad en VUIs

Para dar respuesta a la pregunta Q1 ¿cuáles son los cuestionarios que se han utilizado para evaluar la usabilidad en interfaces de voz? Se revisaron los 57 artículos y se identificaron los cuestionarios utilizados para evaluar la usabilidad. De los trabajos analizados el 44% evalúan la usabilidad con cuestionarios personalizados y el 56% utiliza cuestionarios estandarizados. Los siguientes cuestionarios son los más utilizados en los trabajos examinados para evaluar la usabilidad o la satisfacción subjetiva en interfaces que utilizan la voz como medio de interacción:

#### 1) *AtrakDiff*

Se utiliza para evaluar la usabilidad, atractivo y estética de un sistema [21]. Es un cuestionario de 28 preguntas con 3 categorías: factores pragmáticos que miden la utilidad y usabilidad, factores hedónicos que evalúan el placer, emociones y estimulación percibida por el uso del sistema; y atractivo general que se basan en la combinación de los factores hedónicos y pragmáticos. Mide a través de una escala psicométrica diferencial semántica con la técnica de pares de palabras.

#### 2) *ICF-US Test*

Clasificación internacional de la escala de usabilidad basada en el funcionamiento (ICF-US Test, International Classification Of Functioning Based Usability Scale). Evalúa respecto a principios de usabilidad como: facilidad de uso, satisfacción, facilidad de aprendizaje, eficacia, coherencia, flexibilidad, familiaridad, capacidad de respuesta, claridad de la retroalimentación, visibilidad de la

retroalimentación. Tiene 10 ítems que se puntúan de -3 a 3, el valor 3 es el puntaje más alto y -3 es el más bajo [22].

#### 3) *MOS-X*

Escala de opinión media reducido (MOS-X, Mean Opinion Scale - Reduced) [23]. Mide la calidad del habla sintética, fue diseñado para sistemas de voz con su enfoque. Se utiliza como complemento para medir la usabilidad de un sistema, sin embargo, no se considera suficiente para evaluar la usabilidad [24]. Tiene un total de 15 ítems, cuatro para inteligibilidad, cuatro para naturalidad, tres para prosodia y cuatro para impresión social, con una escala psicométrica de 7 puntos.

#### 4) *PARADISE*

Paradigma para el sistema de evaluación del diálogo (PARADISE, Paradigm for Dialogue Evaluation System). Considera las calificaciones de satisfacción del usuario como un indicador de usabilidad [25]. Pondera medidas de éxito de tareas, eficiencia de dialogo, calidad de diálogo y satisfacción de usuario con preguntas sobre aspectos de usabilidad de interacción con un sistema.

#### 5) *SASSI*

Evaluación subjetiva de las interfaces del sistema de habla (SASSI, Subjective Assessment of Speech System Interfaces) [26]. Evalúa la usabilidad general del sistema de voz y la experiencia de usuario. Consiste en 34 ítems distribuidos en seis factores: Precisión de respuesta del sistema (9 ítems), Simplicidad (9 ítems), Demanda cognitiva (5 ítems), Molestia (5 ítems), Habitabilidad (5 ítems) y Velocidad (2 elementos).

#### 6) *SUISQ-R*

Calidad de servicio de interfaz de usuario de voz (SUISQ-R, Speech User Interface Service Quality - Reduced) [27]. Evalúa los atributos de usabilidad de las aplicaciones de Respuesta de voz interactiva (IVR), cuenta con cuatro factores: orientación de objetivos de usuario (4 elementos), comportamiento de servicio al cliente (4 elementos), características del habla (3 elementos) y verbosidad (3 elementos). Los elementos de Orientación del objetivo del usuario se relacionan con la eficiencia del sistema, la confianza del usuario, la confianza en el sistema y la claridad de la interfaz de voz. El comportamiento del servicio al cliente incluye elementos relacionados con la amabilidad y la cortesía del sistema, su ritmo de conversación y el uso de términos familiares. El factor de Características del habla se relaciona con la naturalidad y el entusiasmo de la voz del sistema. La verbosidad incluye elementos relacionados con la locuacidad y la repetitividad del sistema.

#### 7) *SUS*

Escala de usabilidad del sistema (SUS, System Usability Scale) [12]. Es un cuestionario de 10 preguntas que evalúa la Usabilidad de un sistema, su puntuación puede ser entre 0 y 100, donde cuanto mayor sea la puntuación, mejor será la percepción de la usabilidad del sistema. Ghosh *et al.* [28] sugieren que el SUS es una herramienta de evaluación válida para VUIs.

#### 8) *UEQ*

Cuestionario de experiencia del usuario (UEQ, User Experience Questionnaire) [29]. Evalúa la usabilidad y experiencia de usuario, se enfoca en las cualidades

TABLA V  
CUESTIONARIOS PARA EVALUAR LA USABILIDAD EN INTERFACES DE VOZ

Cuestionario	Mide	# ítems	Tipo de ítem	Escala Psicométrica	Categorías	Tipo de interfaz
AttrakDiff	Usabilidad, atracción y estética	28	Pares de palabras	Diferencial semántica 7 puntos	3	Propósito general
ICF-US	Usabilidad	10	Sentencia	Likert 6 puntos	10	Propósito general
MOS-X	Calidad de voz y habla sintética	15	Preguntas	Likert 7 puntos	15	Sistemas de voz
PARADISE	Satisfacción del usuario	9	Preguntas	Valoración 5 puntos	4	Agentes de diálogo
SASSI	Usabilidad	34	Sentencias	Likert 7 puntos	6	Sistemas de voz
SUISQ-R	Usabilidad	14	Sentencias	Likert 7 puntos	4	Sistemas de respuesta de voz interactiva
SUS	Usabilidad	10	Sentencias	Likert 5 puntos	5	Propósito general
UEQ	Usabilidad y experiencia de usuario	26	Pares de palabras	Diferencial semántica 7 puntos	6	Propósito general
USE	Usabilidad	30	Sentencias	Likert 7 puntos	4	Propósito general

hedónicas y pragmáticas con 26 ítems distribuidos en 6 categorías: atracción, visibilidad, eficiencia, confiabilidad, estimulación y novedad.

#### 9) USE

Cuestionario de utilidad, satisfacción y facilidad de uso (USE, Usefulness, Satisfaction, and Ease of Use Questionnaire). Mide la usabilidad utilizando 30 preguntas divididas en 4 categorías: utilidad, satisfacción, facilidad de uso y facilidad de aprendizaje. Se basa en que la utilidad y facilidad de uso están correlacionadas. Estos factores a su vez impulsan la satisfacción del usuario y la frecuencia de uso.

La cantidad de ítems con los que cuentan cada uno de los cuestionarios, los atributos que miden, el tipo de ítem, la escala psicométrica, las categorías con las que cuenta y el tipo de interfaz para la que fue diseñado se pueden observar en la Tabla V.

Los cuestionarios MOS-X, SUISQ-R, SASSI y PARADISE fueron diseñados para evaluar sistemas que utilizan la voz en la interacción. MOS-X, SUISQ-R y PARADISE se enfocan en evaluar la calidad de la salida voz y SASSI en la calidad de la entrada de voz. Los cuestionarios AttrakDiff, ICF-US, SUS, UEQ y USE fueron diseñados para evaluar la interacción general del sistema con el usuario, en sus ítems no hay preguntas o sentencias para evaluar la calidad de la voz. El uso de dos o más cuestionarios es una práctica común en la evaluación de sistemas, Cordasco *et al.* [30] utilizan el cuestionario SUS junto con el cuestionario AttrakDiff para evaluar la eficacia, eficiencia, disfrute, compromiso y atractivo del uso del sistema. Robb *et al.* [31] utilizan el cuestionario SUS a la par de un cuestionario estilo PARADISE, para comprender qué métricas objetivas contribuyen a la satisfacción del usuario. Pyae & Joelsson [9] y Verde *et al.* [32] utilizan el cuestionario SUS junto al cuestionario UEQ, en ambos trabajos los resultados de la usabilidad del SUS fueron corroborados con los

del cuestionario UEQ, que confirmaron que la aplicación evaluada logró buenos resultados para las variables analizadas. Grisol & Callejas [33] realizaron un cuestionario de evaluación que se basa en los cuestionarios AttrakDiff y SASSI, evalúan el grado en que el usuario percibe que el sistema lo entendió y el grado en el que el usuario entendió el sistema, la tasa de interacción percibida, el nivel de dificultad percibido de la interacción con el sistema, la presencia de errores, la certeza del usuario sobre qué hacer en cada momento y el nivel global de satisfacción con el sistema. Adicionalmente, existen casos donde los cuestionarios se complementan con ítems ad hoc; Biermann *et al.* [34] utilizan el cuestionario SASSI, al cual se añadieron ítems para evaluar la longitud de la respuesta, la naturalidad del diálogo, el nivel de humor y la probabilidad de usar el sistema, así como mencionar aspectos positivos y negativos del diálogo.

#### B. Uso de Cuestionarios de Acuerdo con el tipo de VUI

Respecto a qué tipo de VUIs se han utilizado en estos cuestionarios, dando respuesta a la pregunta de investigación Q2. Se realiza una revisión del tipo de VUI que utilizan para hacer la evaluación en cada uno de los trabajos analizados. De los trabajos analizados 37 utilizan interfaces de voz agregada, que se refiere a interfaces donde la voz no es el principal medio de interacción, sino que es una opción en un sistema multimodal. Veintidós de esos trabajos utilizaron cuestionarios como AttrakDiff, SASSI, ICF-US, PARADISE, SUISQ-R, SUS o USE, los restantes utilizaron cuestionarios personalizados. Veinte trabajos se clasificaron como sistemas con interfaces de solo voz, donde el habla es el único medio de interacción; de los cuales doce de ellos utilizaron cuestionarios como SASSI, SUISQ-R, SUS o UEQ. De los trabajos analizados ninguno ha utilizado VUIs del tipo primero la voz, donde la interacción principal sucede por voz, y algunos de ellos pueden tener una retroalimentación visual; se puede inferir

que la razón es que este tipo de dispositivos tienen poco tiempo en el mercado, tales como los dispositivos Amazon Echo Show y Google Home Hub. En la tabla VI se pueden observar la clasificación por su tipo de interacción.

Cabe destacar que en general los cuestionarios analizados no hacen referencia sobre el tipo interacción con voz que se está utilizando, PARADISE tiene un ítem que compara interfaces “En esta conversación, ¿cómo se compara la interfaz de voz del sistema con la interfaz de tono táctil para el correo de voz?” y el cuestionario IFC-US tiene un ítem sobre la flexibilidad del sistema al preguntar “La posibilidad de interactuar de varias maneras”. Los demás están enfocados a la percepción del usuario sobre la interacción en general, eficiencia de las tareas, facilidad de uso percibida, utilidad del sistema, inteligibilidad,

TABLA VI  
TIPOS DE INTERACCIÓN EN DONDE HAN SIDO UTILIZADOS LOS CUESTIONARIOS

	Comandos	Conversacional
<i>Solo voz</i> 12		
SASSI	1	-
SUISQ-R	1	-
SUS	3	5
UEQ	1	1
<i>Voz agregada</i> 22		
AttrakDiff	1	2
IFC-US	1	-
PARADISE	1	1
SASSI	-	1
SUISQ-R	1	1
SUS	4	7
USE	1	1
<i>Total</i>	15	19

naturalidad, prosodia, verbosidad, ritmo, tono, calidad hedónica, la calidad pragmática, etc. Sin embargo, se maneja de manera indistinta las características de la multimodalidad o como la voz es utilizada como principal medio de interacción o como sólo una opción.

Por su tipo de diálogo las VUIs se pueden clasificar en conversacionales o basado en comandos. Veinticinco de los trabajos se consideran sistemas conversacionales, de los cuales diecinueve utilizaron los cuestionarios como AttrakDiff, SASSI, PARADISE, SUS, UEQ o USE. Treinta y dos trabajos se consideran sistemas basados en comandos, de los cuales quince utilizaron cuestionarios como AttrakDiff, SASSI, IFC-US, PARADISE, SUISQ-R, SUS, UEQ y USE (Tabla VI).

El cuestionario PARADISE contiene ítems que indagan respecto a la experiencia de la conversación, por ejemplo: ¿Fue fácil de entender el sistema en esta conversación?, en esta conversación, ¿fue fácil encontrar el mensaje que deseabas?, en esta conversación, ¿sabía lo que podía decir en cada punto del diálogo? Los cuestionarios SASSI y MOS-X sus ítems se refieren a una interacción basada en comandos al utilizar el término “palabra”, por ejemplo: “algunas veces me preguntaba si estaba usando la palabra correcta”, “¿fueron las palabras simples difíciles de entender?” o “Se hizo hincapié en las palabras importantes?”. El atributo de naturalidad está presente en el cuestionario MOS-X y SUISQ-R, haciendo referencia sobre la entonación y ritmo; indagando si la voz se parece a la de un humano, si el sistema sonaba como una persona normal y no una voz sintética robotizada. Los demás cuestionarios no

cuentan con ítems que diferencien entre la interacción basado en comando o en conversación, por ejemplo, AttrakDiff ha sido utilizado para evaluar sistemas de diálogo basado en conversación [30] así como en sistemas que utilizan interacción con voz basado en comandos [35]. Teixeira *et al.* [36] utilizan el ICF-US test para evaluar un sistema enfocado a adultos mayores que utiliza una interfaz de voz por medio de comandos.

C. Dimensiones de la Usabilidad

Cada uno de los cuestionarios analizados evalúa diferentes categorías para los que fue diseñado. Para dar respuesta a la pregunta de investigación Q3, se realizó una comparativa con las dimensiones de usabilidad analizando cada uno de los ítems de los cuestionarios. Para clasificar en que porcentaje se presentan las dimensiones de la usabilidad: eficiencia, efectividad y satisfacción; dos expertos revisaron de manera independiente y posteriormente de manera conjunta cada uno de los ítems de los cuestionarios analizados clasificando cada ítem en la dimensión de la usabilidad que corresponden (Tabla VII). Por ejemplo, el cuestionario SASSI, en la categoría de *simpatía*, tiene el ítem: *Fui capaz de recuperarme fácilmente de los errores*, el cual se relacionó con la dimensión de eficiencia, ya que se refiere al sí fue apropiado el esfuerzo del usuario. Todos los cuestionarios analizados tienen ítems en las tres dimensiones de la usabilidad (Fig. 1), sin embargo, la cobertura más homogénea respecto a la cantidad ítems en cada aspecto de la usabilidad la tiene el cuestionario SASSI.

Los cuestionarios AttrakDiff, MOS-X, SUISQ-R y UEQ tienen una mayor cantidad de ítems relacionados a evaluar la satisfacción. Los cuestionarios IFC-US, PARADISE, SUS y USE tienen más ítems que evalúan la eficiencia. En proporción general la efectividad es la dimensión que tiene menos ítems. Tractinsky [37] considera que estos aspectos de la usabilidad pueden considerarse como componentes independientes, donde su independencia permite elegir entre ellos, dependiendo de su idoneidad para la tarea o el contexto estudiados.

Algo importante que destacar es que la evaluación de la usabilidad puede ser interpretada de forma diferente por un

TABLA VII  
NÚMERO DE ÍTEMAS ASOCIADOS CON LAS DIMENSIONES DE USABILIDAD

	Eficiencia	Efectividad	Satisfacción
AttrakDiff	5	2	13
IFC-US	6	2	1
MOS-X	3	1	8
PARADISE	4	2	2
SASSI	13	9	12
SUISQ-R	5	1	6
SUS	5	2	3
UEQ	8	3	10
USE	14	6	10

usuario que utiliza las VUIs de manera cotidiana que el que las utiliza con poca frecuencia. Los enunciados repetitivos pueden ser diferentes para un usuario que recién está utilizando una VUI y se está familiarizando con la interacción, a diferencia de un usuario que debido a su experiencia con este tipo de interfaces ciertas instrucciones le pueden parecer obvias y repetitivas, por lo que perciba una disminución de la efectividad. Se debe ser cautelosos para reducir los enunciados repetitivos y una solución posible es brindar opciones de rutas

de enunciados cuidando que la reducción excesiva de expresiones semánticamente similares no cause una falta de guía de conversación para los usuarios [38].

TABLA VIII  
DESAFÍOS DE USABILIDAD POR CUESTIONARIO

	1	2	3	4	5	6	7
AttrakDiff	✓						
ICF-US			✓		✓		
MOS-X	✓						
PARADISE				✓	✓		
SASSI	✓	✓				✓	✓
SUISQ-R	✓						
SUS	✓				✓		✓
UEQ	✓						
USE		✓			✓	✓	

- 1 Diálogos claros y simples
- 2 Recuperación de errores.
- 3 Retroalimentación
- 4 Tiempo de respuesta
- 5 Intuitivo
- 6 Flexible
- 7 Consistente

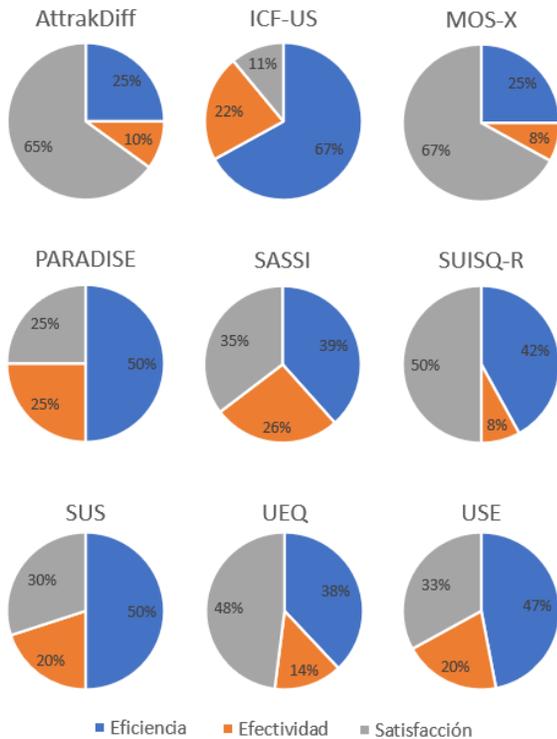


Fig. 1. Comparativa de las dimensiones de usabilidad.

## VI. DISCUSIÓN

En este estudio el 44% de los trabajos analizados evalúan la usabilidad con cuestionarios personalizados, un problema con esto es que los cuestionarios autoconstruidos generalmente no están debidamente validados [39], lo cual no garantiza que el instrumento realmente mide lo que pretende medir; el resto (56%) utiliza cuestionarios estandarizados que han sido sometidos a una calificación psicométrica. Estos cuestionarios

tienen mayor confiabilidad que los cuestionarios de usabilidad personalizados o ad hoc [40]. Una posible razón de que los investigadores personalicen los cuestionarios es que desean conocer percepciones específicas de los usuarios, puede ser que se enfoquen más en la facilidad de uso, otras ocasiones en la satisfacción; la mayoría utiliza preguntas tipo likert y promedia las respuestas de los usuarios.

Se puede observar que los cuestionarios no contienen ítems que tomen en cuenta todos los desafíos de usabilidad que se mencionan en la sección 3 de este documento. En la Tabla VIII se muestra que la mayoría tiene preguntas sobre si la interacción es sencilla o fácil de usar, pero no existe un cuestionario que evalúe los desafíos mencionados.

Una forma de obtener evaluaciones más completas es utilizar varios cuestionarios, por ejemplo: AttrakDiff evalúa en su mayoría la satisfacción del usuario y el atractivo del sistema, por lo que es recomendable aplicarlo en conjunto con otras pruebas de usabilidad, como por ejemplo con el cuestionario SASSI, con el que se completa la evaluación de la voz. ICF-US es sencillo de usar, sin embargo, solo tiene un ítem relacionado con la satisfacción por lo que se recomienda combinar con algún otro cuestionario con el que pueda evaluar esta dimensión de la usabilidad. MOS-X no se considera suficiente para evaluar la usabilidad, sino que es un cuestionario que ayuda a evaluar la calidad de la voz, por lo que se recomienda utilizarlo junto con otro cuestionario. SUIQ-R fue diseñado para evaluar sistemas IVR, tiene preguntas sobre servicio al cliente que pueden no ser aplicables a todos los contextos de VUIs; tiene más ítems para evaluar la eficiencia y la satisfacción, por lo que se podría complementar con otro cuestionario que evalúe la efectividad. SUS es un cuestionario muy utilizado y fácil de usar, se ha utilizado a la par de otros cuestionarios como el AttrakDiff, PARADISE y EUQ. SASSI fue creado para evaluar sistemas de voz, tiene un equilibrio sobre la cantidad de ítems que hay de cada dimensión de usabilidad, incluye categorías sobre demanda cognitiva, habitabilidad y precisión de la respuesta, sin embargo, es un cuestionario largo; con lo que se puede incrementar la no respuesta y la tasa de abandono. UEQ tiene pocos ítems sobre efectividad, fue diseñado para evaluar sistemas interactivos de propósito general, por lo que es recomendable utilizarlo a la par con un cuestionario que evalúe la calidad de la voz. PARADISE se enfoca más en evaluar aspectos de la dimensión satisfacción, es sencillo de implementar y fue diseñado para evaluar la calidad de la voz. USE es un cuestionario que en su mayoría de ítems están enfocados a evaluar la eficiencia y la facilidad de uso, es de propósito general, por lo que se recomienda combinarlo con un cuestionario que evalúe la calidad de la voz. Otra alternativa es crear un cuestionario combinando categorías de diferentes cuestionarios, el cual utilice el mismo tipo de escala y es indispensable comprobar la validez y fiabilidad de este nuevo cuestionario.

Los nueve cuestionarios se han utilizado para evaluar la usabilidad sin hacer una diferencia entre sistemas con interacción basada en comandos o basados en conversación. En ambos tipos de sistemas es importante evaluar la calidad de la voz, sin embargo, en un sistema basado en conversación la adecuada cantidad de información por ronda de conversación y la naturalidad de la voz pueden influir en la usabilidad

percibida. PARADISE contiene ítems respecto a experiencia de la conversación; SASSI y MOS-X respecto a comandos; SUIQ-R y MOS-X tienen ítems que indagan respecto a la naturalidad de la voz utilizada en el sistema.

Las interacciones donde la voz es el único medio de interacción requieren cierto esfuerzo cognitivo del usuario al estar prestando atención a las respuestas del sistema, algunas veces se utiliza este tipo de interacción al estar haciendo otra actividad, por ejemplo, cocinar o ir manejando un auto, por lo que es importante analizar la carga cognitiva y la atención que requiere el usuario, dado que podría mermar su percepción de usabilidad. En la interacción de primero voz, se solicitan procesos por medio de la voz, sin embargo, la respuesta del sistema es multimodal (auditiva, visual y/o táctil), por ejemplo, una pantalla puede dar retroalimentación a la par de la voz, la multimodalidad apoya a que la eficiencia se ve beneficiada, ya que hablamos más rápido de lo que escribimos y leemos más rápido de lo que hablamos. En la interacción con voz añadida, la voz no es sólo una opción para interactuar con un sistema, por lo que el usuario elige usar la voz cuando le resulta más conveniente que los otros medios de interacción. Detectar el tipo de interacción ideal para cada contexto y actividad puede mejorar las percepciones de las dimensiones de la usabilidad de un sistema.

## VII. CONCLUSIÓN

Este trabajo contribuye al conocimiento y difusión de cuestionarios que pueden ser utilizados para evaluar la usabilidad en VUIs. Se hace énfasis en la necesidad de incrementar los estudios en esta área. Las VUIs se están implementando en diversas tecnologías, a medida que se expande su uso es necesario contar con instrumentos de evaluación que ayuden a identificar procesos, conductas o acciones que reducen la usabilidad de un sistema, y con ello poder identificar mejoras en su diseño, uso y aplicación. Este trabajo presenta 9 cuestionarios que se han utilizado para evaluar la usabilidad en VUIs, sus características y en qué tipo de interfaces de voz pueden ser utilizados. Por medio de una revisión sistemática se analizaron 57 trabajos de investigación en base de datos científicas publicados de enero de 2013 a julio de 2020. Se puede observar que hay pocos trabajos de investigación que utilizan cuestionarios para evaluar la usabilidad en VUIs. Una de las principales razones es que las interacciones con voz han tenido un incremento de uso en los últimos años debido principalmente a los avances de la tecnología de reconocimiento de voz y de la inteligencia artificial. Estas razones han permitido que el uso de las interacciones de voz sea cada vez más frecuente y con ello la necesidad de incrementar los estudios orientados a la mejora de diseño y evaluación de VUIs que ayuden a consolidar la comprensión de las prácticas y la experiencia de uso [41][42].

Se establecieron 3 preguntas de investigación, ¿cuáles son los cuestionarios que se han utilizado para evaluar la usabilidad en interfaces de voz? Se identifican los cuestionarios MOS-X, SUIQ-R, SASSI, PARADISE, AttrakDiff, ICF-US, SUS, UEQ, y USE. Los primeros cuatro tienen ítems que evalúan la calidad de la voz, los demás fueron diseñados para sistemas interactivos de propósito general por lo que no evalúan la calidad voz. El cuestionario SUS es el más utilizado, Ghosh et

al. [28] sugieren que es una herramienta de evaluación válida para evaluar VUIs. Se suelen utilizar a la par diferentes cuestionarios para obtener una medición más completa de la usabilidad. Por ejemplo, se aplica el cuestionario SUS a la par de un cuestionario PARADISE. O bien, los cuestionarios AttrakDiff y SASSI. Otra pregunta fue, ¿en qué tipo de VUIs han sido utilizados? En general los cuestionarios analizados no hacen referencia sobre el tipo interacción con voz que se está utilizando (solo voz, primero voz o voz añadida) por lo que se considera es necesario incrementar las investigaciones sobre las diferencias en la interacción que tienen estas distintas modalidades para poder seleccionar un cuestionario que tome en cuenta sus particularidades. Existen diferencias en la experiencia cuando la voz es el único medio de interacción que cuando se tiene la opción de una interacción multimodal, donde la voz es opcional. Es importante hacer un análisis sobre las diferencias entre los tipos de interacción con voz, cómo aumentan o disminuyen la usabilidad; proponer instrumentos estandarizados y validados que tomen en cuenta estas características para conocer las expectativas de los usuarios y cómo afecta la usabilidad. La tercera pregunta es ¿Cuáles son las dimensiones de la usabilidad que se incluyen en las categorías de los cuestionarios? La efectividad, eficiencia y satisfacción se consideran componentes independientes entre sí, estos aparecen en los nueve cuestionarios en proporciones diferentes. Si bien los cuestionarios dan como resultado medidas subjetivas y cada uno mide las categorías para las que fue diseñado con esta comparativa se puede observar que todas las dimensiones de la usabilidad están presentes en los 9 cuestionarios.

El principal desafío para mejorar la usabilidad es que las VUIs deben adaptarse a las necesidades de los usuarios y no al revés. Es importante que el sistema permita mantener un lenguaje natural con el usuario, haciendo la interacción más agradable y cotidiana. Sobre todo, en los sistemas que sean conversacionales. Otro elemento importante es el tiempo de respuesta, lo ideal es mantener una conversación sin pausas prolongadas. Además, que el sistema debe ser capaz de interpretar diferentes formas de solicitar un proceso. Identificar estos desafíos y tomarlos en cuenta al momento de evaluar la usabilidad de una VUI, es fundamental para una amplia adopción de la tecnología de voz. Con ella se busca una interacción más natural con los usuarios, aumentar su satisfacción y colaborar para lograr el cumplimiento de sus objetivos de una manera sencilla. A su vez, incrementar la usabilidad en las VUI ayuda a posicionar la presencia de VUIs en una gama más amplia de actividades.

## REFERENCIAS

- [1] V. Research, "Smart Speaker Consumer Adoption Report April 2020," 2020.
- [2] F. Paz and J. A. Pow-Sang, "Usability Evaluation Methods for Software Development: A Systematic Mapping Review," Proc. - 8th Int. Conf. Adv. Softw. Eng. Its Appl. ASEA 2015, vol. 10, no. 1, pp. 1–4, 2016.
- [3] A. Assila, K. Oliveira, and H. Ezzedine, "Standardized Usability Questionnaires: Features and Quality focus.," Electron. J. Comput. Sci. Inf. Technol., vol. 6, no. 1, pp. 15–31, 2016.
- [4] S. Holmes, A. Moorhead, R. Bond, H. Zheng, V. Coates, and M. McTear, "Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?," ECCE

- 2019 - Proc. 31st Eur. Conf. Cogn. Ergon' 'Design Cogn., no. November, pp. 207–214, 2019.
- [5] G. Kouroupetroglou and D. Spiliotopoulos, "Usability methodologies for real-life voice user interfaces," *Int. J. Inf. Technol. Web Eng.*, vol. 4, no. 4, pp. 78–94, 2009.
- [6] C. Pearl, *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly Media, Inc., 2016.
- [7] A. Mhaidli, M. K. Venkatesh, Y. Zou, F. Schaub, and M. Kandadai, "Listen Only When Spoken To: Interpersonal Communication Cues as Smart Speaker Privacy Controls," *Proc. Priv. Enhancing Technol. ...* (vol. 2020, no. 2, pp. 1–20, 2020).
- [8] T. Uchiya, R. Nakano, D. Yamamoto, R. Nishimura, and I. Takumi, "Extension with intelligent agents for the spoken dialogue system for smartphones," 2015 IEEE 4th Glob. Conf. Consum. Electron. GCCE 2015, pp. 281–282, 2016.
- [9] A. Pyae and T. N. Joelsson, "Investigating the usability and user experiences of voice user interface: A case of Google home smart speaker," *MobileHCI 2018 - Beyond Mob. Next 20 Years - 20th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. Conf. Proc. Adjun.*, pp. 127–131, 2018.
- [10] M. Braun, A. Mainz, R. Chadowitz, B. Pflöging, and F. Alt, "At your service: Designing voice assistant personalities to improve automotive user interfaces a real world driving study," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1–11, 2019.
- [11] B. Axtell, C. Murad, B. R. Cowan, C. Munteanu, L. Clark, and P. Doyle, "Hey Computer, Can We Hit the Reset Button on Speech?," *Proceedings of the SIGCHI 2018 Workshop on Voice-based Conversational UX Studies and Design*, p. 6, 2018.
- [12] D. A. Coates, *Voice Applications for Alexa y Google Assistant*. Editorial Manning, 2019.
- [13] ISO, "Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts, ISO 9241-11:2018(en)," 2018. [Online]. Available: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>.
- [14] C. Murad, C. Munteanu, L. Clark, and B. R. Cowan, "Design guidelines for hands-free speech interaction," *Proc. 20th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. Adjun. - MobileHCI '18*, pp. 269–276, 2018.
- [15] J. F. Quesada Moreno, Z. Callejas Carrión, and D. Griol Barres, "Informe sobre sistemas conversacionales multimodales multilingües," *Tecnologías y arquitecturas para el desarrollo de asistentes virtuales, sistemas de dialogo y otros interfaces conversacionales. Reporte Técnico, Plan de Impulso de las Tecnologías del Lenguaje.*, 2019. [Online]. Available: <https://www.plantl.gob.es/tecnologias-lenguaje/actividades/Estudios%20tcnicos%20y%20de%20gobernanza/Estudio%20de%20sistemas%20conversacionales/estudio-sistemas-conversacionales.pdf>. [Accessed: 28-Mar-2020].
- [16] H. R. Hartson, T. S. Andre, and R. C. Williges, "Criteria For Evaluating Usability Evaluation Methods," *Int. J. Hum. Comput. Interact.*, vol. 15, no. 1, pp. 145–181, 2003.
- [17] B. Weiss, I. Wechsung, C. Kühnel, and S. Möller, "Evaluating embodied conversational agents in multimodal interfaces," *Comput. Cogn. Sci.*, vol. 1, no. 1, p. 6, 2015.
- [18] J. R. Lewis, "Standardized Questionnaires for Voice Interaction Design," *Voice Interact. Des.*, vol. 1, no. 1, pp. 1–16, 2016.
- [19] B. Kitchenham, "Procedures for performing systematic reviews," *Keele, UK, Keele Univ.*, vol. 33, no. TR/SE-0401, p. 28, 2004.
- [20] A. M. Methley, S. Campbell, C. Chew-Graham, R. McNally, and S. Cheraghi-Sohi, "PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews," *BMC Health Serv. Res.*, vol. 14, no. 1, p. 579, 2014.
- [21] M. Hassenzähl, "AttrakDiff." [Online]. Available: <http://attrakdiff.de/index-en.html>. [Accessed: 24-Mar-2020].
- [22] A. I. Martins, A. F. Rosa, A. Queirós, A. Silva, and N. P. Rocha, "Definition and Validation of the ICF - Usability Scale," *Procedia Comput. Sci.*, vol. 67, no. Dsai, pp. 132–139, 2015.
- [23] M. D. Polkosky and J. R. Lewis, "Expanding the MOS: Development and psychometric evaluation of the MOS-R and MOS-X," *Int. J. Speech Technol.*, vol. 6, no. 2, pp. 161–182, 2003.
- [24] A. B. Kocaballi and E. Coiera, "Measuring User Experience in Conversational Interfaces: A Comparison of Six Questionnaires," *Proc. Br. Comput. Soc. Hum. Comput. Interact. Conf. (BCS HCI '18)*, no. July, pp. 1–12, 2018.
- [25] M. WALKER, C. KAMM, and D. LITMAN, "Towards developing general models of usability with PARADISE," *Nat. Lang. Eng.*, vol. 6, no. 3&4, pp. 363–377, 2000.
- [26] K. Hone, "Usability measurement for speech systems: SASSI revisited," *Proc. CHI*, no. 1, 2014.
- [27] J. R. Lewis and M. L. Hardzinski, "Investigating the psychometric properties of the Speech User Interface Service Quality questionnaire," *Int. J. Speech Technol.*, vol. 18, no. 3, pp. 479–487, 2015.
- [28] D. Ghosh, P. S. Foong, S. Zhang, and S. Zhao, "Assessing the utility of the system usability scale for evaluating voice-based user interfaces," *ACM Int. Conf. Proceeding Ser.*, pp. 11–15, 2018.
- [29] M. Schrepp, "User Experience Questionnaire Handbook," 2019. [Online]. Available: <https://www.uq-online.org/Material/Handbook.pdf>. [Accessed: 24-Mar-2020].
- [30] G. Cordasco, "Assessing Voice User Interfaces: The vassist system prototype," 5th IEEE Int. Conf. Cogn. Infocommunications, CogInfoCom 2014 - Proc., pp. 91–96, 2014.
- [31] D. A. Robb, F. J. Chiyah Garcia, A. Laskov, X. Liu, P. Patron, and H. Hastie, "Keep me in the loop: Increasing operator situation awareness through a conversational multimodal interface," *ICMI 2018 - Proc. 2018 Int. Conf. Multimodal Interact.*, pp. 384–392, 2018.
- [32] L. Verde, G. De Pietro, and G. Sannino, "Vox4Health: Preliminary Results of a Pilot Study for the Evaluation of a Mobile Voice Screening Application," vol. 476, no. ISAmI 2016, pp. 131–140, 2016.
- [33] D. Griol and Z. Callejas, "Mobile Conversational Agents for Context-Aware Care Applications," *Cognit. Comput.*, vol. 8, no. 2, pp. 336–356, 2016.
- [34] M. Biermann, E. Schweiger, and M. Jentsch, "Talking to Stupid?!? Improving Voice User Interfaces," Fischer, H. & Hess, S., *Mensch und Computer 2019 - Usability Professionals*. Gesellschaft für Informatik e.V. Und German UPA e.V. no. September, pp. 53–61, 2019. DOI: 10.18420/muc2019-up-0253
- [35] P. L. Mateo Navarro, S. Hillmann, S. Möller, D. Sevilla Ruiz, and G. Martínez Pérez, "Run-time model based framework for automatic evaluation of multimodal interfaces," *J. Multimodal User Interfaces*, vol. 8, no. 4, pp. 399–427, 2014.
- [36] A. Teixeira, "Design and development of Medication Assistant: older adults centred design to go beyond simple medication reminders," *Univers. Access Inf. Soc.*, vol. 16, no. 3, pp. 545–560, 2017.
- [37] N. Tractinsky, "The Usability Construct: A Dead End?," *Human-Computer Interact.*, vol. 33, no. 2, pp. 131–177, 2018.
- [38] H. Jung, "Understanding Differences between Heavy Users and Light Users in Difficulties with Voice User Interfaces," In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI'20)*. Association for Computing Machinery, New York, NY, USA, Article 51, 1–4. 2020. DOI: <https://doi.org/10.1145/3405755.3406170>
- [39] L. B. Larsen, "Assessment of spoken dialogue system usability - What are we really measuring?," *EUROSPEECH 2003 - 8th Eur. Conf. Speech Commun. Technol.*, pp. 1945–1948, 2003.
- [40] J. Sauro and J. R. Lewis, "Correlations among prototypical usability metrics: evidence for the construct of usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1609–1618. 2009. DOI:<https://doi.org/10.1145/1518701.1518947>
- [41] S. Reeves, "Voice-based conversational ux studies and design," in *2018 CHI Conference on Human Factors in Computing Systems, CHI EA 2018, 2018*, vol. 2018-April.
- [42] B. Cowan, J. Fischer, and L. Clark, "CUI @ CHI: Mapping Grand Challenges for the Conversational User Interface Community," pp. 1–8, 2020.



**Adriana L. Iñiguez** es profesora en la Universidad de Guadalajara en México. Es ingeniera en sistemas computacionales, tiene una maestría en computación y actualmente está realizando un doctorado en Tecnologías de Información. Colabora en el Centro de Innovación de Ciudades Inteligentes de CUCEA de la Universidad de Guadalajara. Tiene experiencia profesional como coordinadora de programas universitarios, en el desempeño de análisis de

calidad y relevancia de los planes de estudio en el área de informática. Entre sus principales intereses de investigación está la Interacción Humano-Computadora, las interacciones con voz e inteligencia artificial.



**Laura S. Gaytán** es profesora Investigadora en la Universidad de Colima en México, en la Facultad de Ingeniería Mecánica y Eléctrica. Se enfoca en la Interacción Humano-Computadora. Sus intereses de investigación incluyen juegos serios para la educación, experiencia de usuario, computación social y accesibilidad. Tiene un doctorado en

Tecnologías de Información de la Universidad de Guadalajara y es miembro del Sistema Nacional de Investigadores de México, así como de la Academia Mexicana de Computación. Se desempeña como presidenta de la Comunidad Latina de Interacción Humano-Computadora de la ACM SIGCHI. Así como parte del Comité Ejecutivo de la ACM-W Norteamérica.



**Miguel A. García-Ruiz** se graduó en ingeniería de sistemas informáticos y obtuvo su maestría en ciencias de la computación de la Universidad de Colima, México. Miguel recibió su doctorado en Ciencias de la Computación e Inteligencia Artificial del Centro de Investigación en Ciencias Cognitivas (COGS), Universidad de Sussex, Reino Unido. Realizó un taller de realidad virtual en la Universidad de

Salford, Inglaterra, y una pasantía en técnicas de gráficos por computadora en la Universidad Politécnica de Madrid, España. Ha publicado varios libros y más de 60 artículos científicos en revistas académicas, y ha dirigido un video documental sobre introducción a la realidad virtual. Fue profesor visitante en la Facultad de Negocios e Informática de la Universidad de Ontario, Canadá, donde participó en un proyecto de investigación sobre la lucha contra las desviaciones sociales con la tecnología. Fue profesor asistente en el Colegio de Telemática de la Universidad de Colima. México Actualmente, Miguel es profesor asociado del Departamento de Matemáticas e Informática de la Universidad de Algoma, Canadá. El Dr. García Ruiz fue miembro del Sistema Nacional de Investigadores (SNI) de México del Consejo Nacional de Ciencia y Tecnología (CONACYT), y fue galardonado con el Educador de Innovación del Año, otorgado por Sault Ste. Centro de Innovación Marie (SSMIC) en 2015.



**Rocío Maciel** es responsable de vinculación y gestión del talento en el Centro de Innovación en Ciudades Inteligentes de la Universidad de Guadalajara. Tiene amplia experiencia en los procesos de educación, particularmente aquellos en modalidades en línea, ha impulsado dichos programas en la Universidad de Guadalajara y en la Secretaría de Educación Jalisco con más de

10,000 usuarios. Graduada del programa de doctorado y

maestría en Administración y Metodología de la enseñanza del Instituto Mexicano de Estudios Pedagógicos IMEP. Tiene más de 25 años de experiencia en Tecnologías de Información, directora de Informática, coordinadora del área de Diseño Gráfico, coordinadora de la licenciatura en Sistemas de Información, consultor jurídico, asesor técnico-pedagógico. Coordinadora del Doctorado en Tecnologías de Información. Profesora de Programas de Doctorado, Maestría y Licenciatura en Tecnologías de Información.