

Unión Internacional de Telecomunicaciones

**UIT-T**

SECTOR DE NORMALIZACIÓN  
DE LAS TELECOMUNICACIONES  
DE LA UIT

**Serie P**  
**Suplemento 24**  
(10/2005)

SERIE P: CALIDAD DE TRANSMISIÓN TELEFÓNICA,  
INSTALACIONES TELEFÓNICAS Y REDES LOCALES

---

**Parámetros que describen la interacción con  
sistemas de diálogo oral**

Recomendaciones UIT-T de la serie P – Suplemento 24

UIT-T



## RECOMENDACIONES UIT-T DE LA SERIE P

### CALIDAD DE TRANSMISIÓN TELEFÓNICA, INSTALACIONES TELEFÓNICAS Y REDES LOCALES

Vocabulario y efectos de los parámetros de transmisión sobre la opinión de los clientes	Series	P.10
Líneas y aparatos de abonado	Series	P.30
		P.300
Patrones de transmisión	Series	P.40
Aparatos para mediciones objetivas	Series	P.50
		P.500
Medidas electroacústicas objetivas	Series	P.60
Medidas relativas a la sonoridad vocal	Series	P.70
Métodos de evaluación objetiva y subjetiva de la calidad	Series	P.80
		P.800
Calidad audiovisual en servicios multimedios	Series	P.900
Aspectos de calidad de transmisión y de calidad de servicio en los puntos extremos de redes de protocolo Internet	Series	P.1000

*Para más información, véase la Lista de Recomendaciones del UIT-T.*

## **Suplemento 24 a las Recomendaciones UIT-T de la serie P**

### **Parámetros que describen la interacción con sistemas de diálogo oral**

#### **Resumen**

En el presente Suplemento se define un conjunto de parámetros que pueden extraerse de los servicios basados en sistemas de diálogo oral, a partir de interacciones registradas del usuario (experimental) con el servicio del caso. Estos parámetros cuantifican el flujo de la interacción, el comportamiento del usuario y del sistema y la calidad de funcionamiento de los dispositivos de tecnología de voz que participan en la interacción. Asimismo, facilitan información útil para el desarrollo, optimización y mantenimiento del sistema, y complementan las opiniones subjetivas sobre la calidad recabadas de conformidad con la Rec. UIT-T P.851.

#### **Orígenes**

El Suplemento 24 a las Recomendaciones UIT-T de la serie P fue aceptado el 21 de octubre de 2005 por la Comisión de Estudio 12 (2005-2008) del UIT-T.

#### **Palabras clave**

Comprensión automática de voz, evaluación, generación de voz, gestión de diálogo, oral, parámetro de interacción, reconocimiento automático de voz, sistema de diálogo tecnología de voz.

## PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Asamblea Mundial de Normalización de las Telecomunicaciones (AMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

## NOTA

En esta publicación, la expresión "Administración" se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

La observancia de esta publicación es voluntaria. Ahora bien, la publicación puede contener ciertas disposiciones obligatorias (para asegurar, por ejemplo, la aplicabilidad o la interoperabilidad), por lo que la observancia se consigue con el cumplimiento exacto y puntual de todas las disposiciones obligatorias. La obligatoriedad de un elemento preceptivo o requisito se expresa mediante las frases "tener que, haber de, hay que + infinitivo" o el verbo principal en tiempo futuro simple de mandato, en modo afirmativo o negativo. El hecho de que se utilice esta formulación no entraña que la observancia se imponga a ninguna de las partes.

## PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente publicación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de publicaciones.

En la fecha de aprobación de la presente publicación, la UIT no ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta publicación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB.

© UIT 2006

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

## ÍNDICE

	<b>Página</b>
1 Alcance .....	1
2 Referencias .....	1
3 Definiciones.....	1
4 Abreviaturas, siglas o acrónimos .....	2
5 Introducción.....	3
6 Características de los parámetros de interacción .....	4
7 Descripción general de los parámetros de interacción .....	4
7.1 Parámetros relacionados con el diálogo y la comunicación .....	5
7.2 Parámetros relacionados con la metacomunicación .....	7
7.3 Parámetros relacionados con la cooperatividad .....	9
7.4 Parámetros relacionados con tareas .....	10
7.5 Parámetros relacionados con la adquisición de voz .....	12
7.6 Otros parámetros .....	15
8 Interpretación de los valores de los parámetros de interacción .....	16
BIBLIOGRAFÍA .....	17



## Suplemento 24 a las Recomendaciones UIT-T de la serie P

### Parámetros que describen la interacción con sistemas de diálogo oral

#### 1 Alcance

En el presente Suplemento se describen parámetros que facilitan información sobre la interacción con servicios basados en sistemas de diálogo oral, desde el punto de vista del ingeniero de sistema y del operador de servicios. Los sistemas de diálogo oral objeto de este Suplemento permiten la interacción mediante lenguaje oral con un usuario humano por la red telefónica funcionando por turnos, y disponen de funciones de reconocimiento automático de voz, comprensión de voz, gestión del diálogo, generación de respuestas y reproducción de voz. Estos sistemas permiten acceder a información almacenada en bases de datos o realizar diferentes tipos de transacciones.

Los parámetros que se definen en este documento cuantifican el flujo de interacción, el comportamiento del usuario y del sistema y la calidad de funcionamiento de los dispositivos de tecnología de voz que participan en la interacción. Para extraer todos los parámetros, el sistema de diálogo oral tiene que ser accesible como una caja transparente (caja blanca); no obstante, algunos parámetros también pueden extraerse considerando el sistema como caja negra, es decir, sin acceder a los distintos componentes que lo integran. La extracción puede realizarse en parte automáticamente y en parte por una persona que escriba y anote las interacciones en ficheros registros. Los parámetros describen la calidad de funcionamiento del sistema desde la perspectiva del ingeniero de sistemas; por esa razón, contienen información complementaria a la que se obtiene mediante los experimentos de evaluación subjetiva de los sistemas de diálogo oral, que se especifican en la Rec. UIT-T P.851. Para mayor información sobre los métodos de evaluación subjetiva en general y sobre la evaluación de dispositivos de reproducción de voz en particular, véanse las Recs. UIT-T P.800 y P.85, así como el Manual de telefonometría. Los parámetros que figuran en este Suplemento no se aplican específicamente a las degradaciones que pudiera introducir el canal de transmisión. Estos efectos quedan en estudio en la CE 12 del UIT-T.

#### 2 Referencias

- Recomendación UIT-T P.85 (1994), *Método para la evaluación subjetiva de la calidad vocal de los dispositivos generadores de voz.*
- Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión.*
- Recomendación UIT-T P.851 (2003), *Evaluación de la calidad subjetiva de los servicios telefónicos basados en sistemas conversacionales.*
- *Manual de telefonometría del UIT-T* (1992).

#### 3 Definiciones

Las definiciones que no aparecen en esta lista figuran en la Rec. UIT-T P.10.

**3.1 intervención:** Capacidad de una persona de hablar tras la invitación o reproducción del sistema [10].

**3.2 diálogo:** Conversación o intercambio de información. En el contexto de unidad de evaluación: uno de los posibles trayectos a través de la estructura de diálogo.

**3.3 eficiencia:** Medida de la precisión y la completitud de las tareas específicas que realiza un sistema con respecto a los recursos empleados por el mismo (por ejemplo, tiempo, intervención humana).

- 3.4 intercambio de palabras:** Un par de turnos contiguos y relacionados, en los que cada vez habla una de las partes que intervienen en el diálogo [8].
- 3.5 funcionalidad:** Capacidad del sistema de realizar funciones que satisfacen requisitos estipulados y definidos cuando se emplea en condiciones específicas.
- 3.6 metacomunicación:** Comunicación sobre la comunicación, por ejemplo para resolver ambigüedades ("¿lo entendiste correctamente?") o para llegar a un acuerdo sobre la utilización del idioma.
- 3.7 calidad de funcionamiento:** Capacidad de una unidad de realizar la función para la cual ha sido concebida.
- 3.8 tecnología de voz:** Disciplina relativa a la investigación y desarrollo de sistemas de adquisición y reproducción de lenguaje oral, que emplea resultados de otras disciplinas afines tales como la acústica, ingeniería eléctrica, estadística, fonética y procesamiento de lenguaje natural, y que consiste en la especificación de requisitos, el diseño, la materialización y evaluación del sistema, el procesamiento del corpus y del registro lingüístico y la evaluación del producto desde el punto de vista del usuario [10].
- 3.9 sistema de diálogo oral:** Sistema informático en el que el usuario humano y dicho sistema interactúan por turnos mediante lenguaje oral.
- 3.10 tarea:** Todas las actividades que debe realizar un usuario para alcanzar un determinado objetivo en un dominio concreto.
- 3.11 diálogo temático:** Diálogo sobre un tema específico, destinado a lograr un objetivo explícito (por ejemplo, la resolución de un problema o la obtención de información concreta) [8].
- 3.12 transacción:** Parte de un diálogo destinada a una sola tarea de alto nivel (por ejemplo, reservar un viaje o consultar el saldo de una cuenta bancaria). Una transacción podrá tener la misma duración que un diálogo o bien un diálogo podrá consistir en más de una transacción [8].
- 3.13 turno:** enunciado. Unidad de conversación, recitada por una parte en el diálogo, desde que la parte comienza a hablar hasta que comienza a hablar la siguiente [1].
- 3.14 enunciado:** Véase turno.

#### 4 Abreviaturas, siglas o acrónimos

ASR	Reconocimiento automático de la voz ( <i>automatic speech recognition</i> )
AVM	Matriz atributo-valor ( <i>attribute-value matrix</i> )
AVP	Par atributo-valor ( <i>attribute-value pair</i> )
DARPA	Organismo de proyectos de investigación avanzada para la defensa ( <i>Defense Advanced Research Projects Agency</i> )
DP	Programación dinámica ( <i>dynamic programming</i> )
DTMF	Multifrecuencia bitono ( <i>dual tone multiple frequency</i> )
IVR	Respuesta vocal interactiva ( <i>interactive voice response</i> )
MOS	Nota media de opinión ( <i>mean opinion score</i> )
SDS	Sistema de diálogo oral ( <i>spoken dialogue system</i> )
WoZ	El mago de Oz ( <i>Wizard-of-Oz</i> )

## 5 Introducción

Es posible que los sistemas de diálogo oral (SDS), es decir, los sistemas informáticos que admiten la comunicación de usuarios humanos en lenguaje oral, por turnos, lleguen a ser parte integral de las redes telefónicas modernas. Estos sistemas permiten acceder a bases de datos y realizar transacciones a través del aparato telefónico, por ejemplo, obtener información sobre horarios de trenes o aviones, movimientos de bolsa, información turística, realizar operaciones en una cuenta bancaria o reservar una habitación de hotel. A diferencia de los sistemas sencillos de respuesta vocal interactiva (IVR) con entrada DTMF, los sistemas SDS ofrecen una gama completa de capacidades de interacción de voz, en particular reconocer la voz del usuario, asignar significado a las palabras reconocidas, decidir cómo continuar el diálogo, formular una respuesta lingüística, y generar una voz destinada al usuario. De este modo se logra una interacción oral más o menos "natural" entre el usuario y el sistema.

Para evaluar la calidad de los servicios basados en SDS desde la perspectiva del usuario, la CE 12 del UIT-T elaboró la Rec. UIT-T P.851, publicada en 2003, en la que se describen métodos para realizar experimentos de evaluación subjetiva con el fin de determinar la calidad *desde el punto de vista del usuario*, en los que se considera el SDS como una caja negra. Los experimentos realizados de acuerdo con la Rec. UIT-T P.851 permiten obtener información valiosa sobre la calidad que percibe el usuario. Sin embargo, resulta difícil determinar la contribución de cada componente del sistema a la calidad general que experimenta el usuario, por ejemplo, determinar qué componentes es necesario mejorar en caso de problemas de interacción. Así pues, la evaluación debería complementarse con información relativa a la calidad de funcionamiento del sistema *desde el punto de vista del ingeniero de sistemas y del operador del servicio*.

La información sobre el sistema puede describirse mediante los denominados *parámetros de interacción*. Estos parámetros ayudan a cuantificar el flujo de la interacción, el comportamiento del usuario y el sistema y la calidad de funcionamiento de los dispositivos de tecnología de voz que intervienen en la interacción. Describen la calidad de funcionamiento del sistema desde el punto de vista del ingeniero de sistemas y del operador del servicio y, por consiguiente, ofrecen información complementaria de los datos de evaluación subjetiva. Para extraer algunos de los parámetros, el sistema de diálogo oral tiene que ser accesible como una caja transparente; no obstante, algunos parámetros también pueden extraerse considerando el sistema como una caja negra, es decir, sin acceder por separado a los componentes que lo integran.

Este Suplemento se describe un conjunto de parámetros de interacción que se han venido utilizando en los últimos 15 años para evaluar los SDS. Estos parámetros guardan relación con la comunicación general de información entre el usuario y el sistema, la metacomunicación en caso de malentendido, la cooperatividad del sistema, las tareas que pueden realizarse con la ayuda del sistema y las capacidades de adquisición de voz del sistema. Todavía no existe una descripción paramétrica de la calidad de generación de voz (por ejemplo, la calidad de la voz sintetizada). El conjunto de parámetros se basa en la investigación teórica que se describe en [17].

No todos los parámetros de interacción guardan una relación directa con la calidad percibida de los servicios basados en SDS. Generalmente, los distintos parámetros tienen poco que ver con la valoración de calidad del usuario, pero resulta conveniente disponer de un vasto conjunto de parámetros que describan la interacción entre el usuario y el sistema, por cuanto estos parámetros aportan la mayor parte de la información potencialmente pertinente sobre la calidad percibida desde la perspectiva del ingeniero de sistemas. Asimismo, estos parámetros ofrecen información útil para el desarrollo, optimización y mantenimiento del sistema.

Una vez definidos estos parámetros y utilizados en experimentos de evaluación en diferentes laboratorios de prueba, será posible estimar su incidencia en la calidad percibida para una gran variedad de sistemas y servicios, lo que permitirá elaborar algoritmos para predecir la calidad a partir de los parámetros de interacción. Este asunto se está estudiando en la CE 12 del UIT-T y otros foros.

## **6 Características de los parámetros de interacción**

Los parámetros de interacción pueden extraerse durante la interacción de usuarios efectivos o experimentales con el servicio. La extracción puede realizarse en parte automáticamente y en parte con la ayuda de ficheros registro que transcribe y anota una persona experta. Algunos parámetros sencillos, tales como la duración de la interacción o de los enunciados, pueden medirse automáticamente mediante instrumentos, utilizando para ello algoritmos adecuados. Por otra parte, es necesario que una persona transcriba y anote los resultados cuando no sólo es importante la forma (las señales de voz), sino también el contenido y significado de los enunciados del sistema o del usuario (por ejemplo, para determinar la precisión de una palabra o concepto).

La complejidad de los SDS exige que la descripción del comportamiento del sistema y la comparación entre sistemas o versiones de éstos se base en muchos parámetros diferentes [24]. Por esta razón, resulta necesario extraer los parámetros de interacción de las dos maneras (automática y por un experto) a fin de obtener la mayor cantidad de información posible. Partiendo de la información recopilada, los sistemas de diálogo oral podrán optimizarse y mantenerse con eficacia.

Dado que se obtienen de la interacción entre el usuario y el sistema, los parámetros de interacción dependen de las características del sistema, del usuario y de la interacción entre ambos. Por regla general, no es posible distinguir cuál de estas características es la determinante, puesto que el comportamiento del usuario depende sobremedida del sistema (por ejemplo, las preguntas formuladas por éste) y viceversa (por ejemplo, el vocabulario y el habla del usuario afecta a la precisión de reconocimiento y comprensión del sistema). Por consiguiente, los parámetros de interacción dependen en gran medida de las características del grupo de usuarios con el que se han realizado los registros.

Los parámetros de interacción se determinan en un laboratorio de pruebas en condiciones controladas o en pruebas de campo. En este último caso, quizá no sea posible extraer todos los parámetros, puesto que no se podrá recabar toda la información necesaria. Por ejemplo, si se desea determinar el éxito de una cierta interacción temática (por ejemplo, la obtención de un horario de trenes), será necesario conocer exactamente lo que desea consultar el usuario. Esta información sólo puede extraerse en un laboratorio de pruebas, por ejemplo, del modo que se describe en la Rec. UIT-T P.851. Si no se dispone de un sistema completamente integrado, se pueden recopilar los parámetros mediante una simulación denominada "mago de Oz" (WoZ), en la que una persona reemplaza las partes que faltan del sistema sometido a prueba. Al interpretar los parámetros obtenidos será necesario tomar en consideración las características de esta simulación.

Los parámetros de interacción pueden calcularse a nivel de palabra, enunciado u oración, o bien a nivel de interacción o diálogo completos. En el caso de parámetros de nivel de palabra o enunciado, es frecuente calcular valores medios para cada diálogo. Los parámetros recopilados con un grupo específico de usuarios pueden analizarse con respecto a la incidencia del sistema (versión), el grupo de usuarios y la configuración experimental (casos, entorno experimental, etc.) utilizando para ello los métodos estadísticos normales. En la Rec. UIT-T P.851 se caracterizan estos efectos.

## **7 Descripción general de los parámetros de interacción**

Partiendo de la vasta literatura sobre el tema, se han determinado los parámetros utilizados en diferentes experimentos de evaluación durante los últimos 15 años. La literatura empleada puede consultarse en [2][3][4][6][7][8][9][11][12][14][16][21][22][23][24][25][26][27][28][30][31][32], y los parámetros se resumen en [17]. Los parámetros pueden clasificarse a grandes rasgos del modo siguiente:

- parámetros relacionados con el diálogo y la comunicación;
- parámetros relacionados con la metacomunicación;
- parámetros relacionados con la cooperatividad;

- parámetros relacionados con las tareas;
- parámetros relacionados con la adquisición de voz.

Estas categorías se explican brevemente en las siguientes cláusulas. Para cada categoría se indican los correspondientes parámetros, junto con su definición, el nivel de interacción al que corresponde el parámetro (palabra, enunciado o diálogo), así como el método de medición (instrumental o por un experto).

### 7.1 Parámetros relacionados con el diálogo y la comunicación

Los parámetros relacionados con el diálogo general o la comunicación de información indican muy aproximadamente la forma de la interacción. Estos parámetros no especifican en detalle la función comunicativa de cada enunciado. Los parámetros que pertenecen a esta categoría se enumeran en el cuadro 1 y son parámetros relacionados con la duración (duración total del diálogo, duración de los turnos del sistema y del usuario, tiempo de respuesta del sistema y del usuario), y parámetros relacionados con la palabra y el turno (número medio de turnos del sistema y usuario, número medio de palabras por turno del sistema y del usuario, número de preguntas del sistema y del usuario).

Cabe destacar los dos parámetros que se proponen en [11], a saber, la *densidad de preguntas*, que es un indicador de la eficacia con la que el usuario puede facilitar nueva información al sistema, y la *eficiencia de comprensión*, que describe la eficiencia con la que el sistema puede asimilar la información comunicada por el usuario. Si bien estos parámetros también están relacionados con la capacidad de comprensión de lenguaje del sistema, se han incluido en esta cláusula por cuanto son el resultado de las capacidades de interacción del sistema en su totalidad, y no únicamente de las capacidades de comprensión del lenguaje.

Todos los parámetros de esta categoría son de carácter general y se refieren al diálogo en su totalidad, aunque algunos se calculan a nivel de enunciado. Los parámetros generales resultan a veces problemáticos, dado que las diferencias de capacidad cognoscitiva pueden ser grandes en relación con las diferencias originadas por el sistema, y porque las personas pueden aprender estrategias para solucionar tareas e influir así considerablemente sobre los parámetros generales.

**Cuadro 1 – Parámetros de interacción relacionados con el diálogo y la comunicación**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
<i>DD</i>	Duración del diálogo	Duración total del diálogo en [ms], véase por ejemplo [8][6][12][21].	Diálogo	Instrumento
<i>STD</i>	Duración del turno del sistema	Duración media de un turno del sistema, desde que el sistema comienza a hablar hasta que para, en [ms]. Un turno es un enunciado, es decir, una unidad de discurso que dice una de las partes que intervienen en el diálogo. [8]	Enunciado	Instrumento
<i>UTD</i>	Duración del turno del usuario	Duración media de un turno del usuario, desde que el usuario comienza a hablar hasta que para, en [ms]. [8]	Enunciado	Instrumento
<i>SRD</i>	Tiempo de respuesta del sistema	Tiempo medio de respuesta del sistema, desde que el usuario termina de hablar hasta que comienza el sistema, en [ms]. [22]	Enunciado	Instrumento
<i>URD</i>	Tiempo de respuesta del usuario	Tiempo medio de respuesta del usuario, desde que el sistema termina de hablar hasta que comienza el usuario, en [ms]. [22]	Enunciado	Instrumento

**Cuadro 1 – Parámetros de interacción relacionados con el diálogo y la comunicación**

<b>Sigla</b>	<b>Nombre</b>	<b>Definición</b>	<b>Nivel de interacción</b>	<b>Método de medición</b>
# <i>turns</i>	Número de turnos	Número total de turnos que integra un diálogo. [30]	Diálogo	Instrumento/ Experto
# <i>system turns</i>	Número de turnos del sistema	Número total de turnos del sistema durante un diálogo. [30]	Diálogo	Instrumento/ Experto
# <i>user turns</i>	Número de turnos del usuario	Número total de turnos del usuario durante un diálogo. [30]	Diálogo	Instrumento/ Experto
<i>WPST</i>	Palabras por turno del sistema	Número medio de palabras por turno del sistema en un diálogo. [6]	Enunciado	Instrumento/ Experto
<i>WPUT</i>	Palabras por turno del usuario	Número medio de palabras por turno del usuario en un diálogo. [6]	Enunciado	Instrumento/ Experto
# <i>system questions</i>	Número de preguntas del sistema	Número total de preguntas que formula el sistema por diálogo.	Diálogo	Experto
# <i>user questions</i>	Número de preguntas del usuario	Número total de preguntas que formula el usuario por diálogo. [12][21]	Diálogo	Experto
<i>QD</i>	Densidad de preguntas	<p>Número medio de nuevos conceptos introducidos (posiciones, véase 7.4) por pregunta de usuario. Sea <math>n_d</math> el número de diálogos, <math>n_q(i)</math> el número total de preguntas en el <math>i</math>-ésimo diálogo, y <math>n_u(i)</math> el número de conceptos nuevos "comprendidos" correctamente por el sistema en el <math>i</math>-ésimo diálogo, en ese caso:</p> $QD = \frac{1}{n_d} \sum_{i=1}^{n_d} \frac{n_u(i)}{n_q(i)}$ <p>Un concepto no se cuenta en <math>n_u(i)</math> si el sistema ya lo comprendió en uno de los enunciados anteriores. [11]</p>	Conjunto de diálogo	Experto
<i>CE</i>	Eficiencia de comprensión	<p>Número medio de turnos que es necesario para que el sistema "entienda" cada concepto. Sea <math>n_d</math> el número de diálogos, <math>n_u(i)</math> el número de conceptos nuevos "comprendidos" correctamente por el sistema en el <math>i</math>-ésimo diálogo, y <math>n_c(i)</math> el número total de conceptos en el <math>i</math>-ésimo diálogo, en ese caso:</p> $CE = \frac{1}{n_d} \sum_{i=1}^{N_d} \frac{n_u(i)}{n_c(i)}$ <p>Un concepto formulado por el usuario se cuenta tantas veces como aparezca hasta que el sistema lo comprenda. [11]</p>	Conjunto de diálogo	Experto

## 7.2 Parámetros relacionados con la metacomunicación

La metacomunicación, es decir, la comunicación sobre la comunicación, es especialmente importante para la interacción oral con los sistemas cuyas capacidades de reconocimiento, comprensión y razonamiento son limitadas. En este caso, se necesitan enunciados de corrección y aclaración a fin de resolver los malentendidos.

Los parámetros que pertenecen a este grupo cuantifican el número de enunciados del sistema y del usuario que forman parte de la metacomunicación. La mayoría de los parámetros se calculan contando el número absoluto de enunciados en un diálogo relacionados con un determinado problema de interacción y luego se promedia para un conjunto de diálogos. Los parámetros son el número de veces que el usuario pide ayuda, el número de mensajes del sistema expirados, el número de enunciados del usuario rechazados por el sistema por no haber podido extraer contenido semántico de los mismos (rechazos ASR), el número de mensajes de error del sistema para diagnóstico, el número de intentos de intervenir del usuario y el número de intentos del usuario de cancelar una acción anterior.

La capacidad del sistema (y del usuario) de recobrase tras producirse problemas de interacción puede describirse de dos maneras, a saber, explícitamente mediante la tasa de corrección, es decir, el porcentaje de todos los turnos (del sistema o del usuario) que se utilizan principalmente para corregir problemas de interacción, o implícitamente mediante el parámetro *recuperación implícita*, que cuantifica la capacidad del sistema de recobrar los enunciados que no pudieron reconocerse o comprenderse totalmente.

A diferencia de las mediciones generales, la mayoría de los parámetros relacionados con la metacomunicación describen la función de los enunciados del sistema y del usuario en el proceso de comunicación. Así pues, la mayoría de los parámetros han de determinarse con la ayuda de un experto que tome notas. Los parámetros se enumeran en el cuadro 2.

**Cuadro 2 – Parámetros de interacción relacionados con la metacomunicación**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
# <i>help request</i>	Número de veces que el usuario pide ayuda	Número total de veces que el usuario pide ayuda en un diálogo. El Experto anota cada vez que el usuario pide explícitamente ayuda, ya sea en forma de pregunta (por ejemplo, ¿cuáles son las opciones disponibles?) o de exclamación ("¡Dime las opciones disponibles!"). [30]	Enunciado	Experto
# <i>system help</i>	Número de mensajes de ayuda del sistema para diagnóstico	Número total de mensajes de ayuda generados por el sistema en un diálogo. Un mensaje de ayuda es un enunciado del sistema que informa al usuario de las opciones disponibles en un determinado punto del diálogo.	Enunciado	Instrumento/ Experto
# <i>time-out</i>	Número de invitaciones expiradas	Número de invitaciones en un diálogo para las que expiró el temporizador por falta de respuesta del usuario. [30]	Enunciado	Instrumento
# <i>ASR rejection</i>	Número de rechazos ASR	Número total de rechazos ASR en un diálogo. Un rechazo ASR se define como una indicación del sistema de que éste es incapaz de "oír" o "comprender" al usuario, es decir, que el sistema no puede extraer el significado de lo enunciado por el usuario. [30]	Enunciado	Instrumento

**Cuadro 2 – Parámetros de interacción relacionados con la metacomunicación**

<b>Sigla</b>	<b>Nombre</b>	<b>Definición</b>	<b>Nivel de interacción</b>	<b>Método de medición</b>
<i># system error</i>	Número de mensajes de error del sistema para diagnóstico	Número total de mensajes de error generados por el sistema en un diálogo. Un mensaje de error de diagnóstico se define como un enunciado del sistema mediante el cual indica que es incapaz de realizar una determinada tarea o facilitar cierta información. [22]	Enunciado	Instrumento/ Experto
<i># barge-in</i>	Número de intentos de intervenir del usuario	Número de veces que el usuario trata de intervenir en un diálogo. Se cuenta como un intento de intervenir cada vez que el usuario trata intencionadamente de intervenir mientras el sistema sigue hablando. No se cuentan, en cambio, los enunciados del usuario que no están destinados a influir en el curso del diálogo (risas, expresiones de ira o cortesía). [30]	Enunciado	Experto
<i># cancel</i>	Número de intentos de cancelar del usuario	Número total de veces que el usuario trata de cancelar el diálogo. Por intento de cancelar se entiende un turno del usuario en el que éste trata de recomenzar el diálogo desde el principio o volver explícitamente atrás uno o varios niveles en la jerarquía del diálogo. [16][23]	Enunciado	Experto
<i>SCT, SCR</i>	Número de turnos de corrección del sistema, tasa de corrección del sistema	Número total de veces (SCT) o porcentaje (SCR) de turnos del sistema en un diálogo destinados principalmente a rectificar un "problema", es decir, turnos en los que no se enuncia nuevo contenido, sino que se interrumpe el flujo del diálogo. El "problema" puede deberse a errores de comprensión o reconocimiento de voz, o a enunciados ilógicos, contradictorios o indefinidos emitidos por el usuario. En caso de que el usuario no responda a una pregunta del sistema, la correspondiente respuesta del sistema se clasifica como turno de corrección del sistema, salvo cuando el usuario solicita información que el sistema no tiene o pida que el sistema haga una acción que no puede realizar en ese momento. [8][24][9][7]	Enunciado	Experto
<i>UCT, UCR</i>	Número de turnos de corrección del usuario, tasa de corrección del usuario	Número total de veces (UCT) o porcentaje (UCR) de turnos del usuario en un diálogo destinados principalmente a rectificar un "problema", es decir, turnos en los que no se enuncia nuevo contenido, sino que se interrumpe el flujo del diálogo (véase SCT, SCR). [8][24][9][7]	Enunciado	Experto

**Cuadro 2 – Parámetros de interacción relacionados con la metacomunicación**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
IR	Recuperación implícita	<p>Capacidad del sistema de recobrar los enunciados del usuario que no pudieron reconocerse o comprenderse totalmente. Viene dado por la siguiente expresión que determina si la respuesta del sistema a los enunciados parcialmente comprendidos fue o no "adecuada" (véase la definición de PA:PA en la cláusula 7.5):</p> $IR = \frac{\# \text{ enun. con respuesta adecuada del sistema}}{PA:PA}$ <p>La definición de "adecuación" figura en 7.3. [7]</p>	Enunciado	Experto

### 7.3 Parámetros relacionados con la cooperatividad

La cooperatividad se ha clasificado como uno de los factores fundamentales para lograr una buena interacción con un sistema de diálogo oral [1]. Lamentablemente, resulta difícil de cuantificar la cooperatividad del sistema. Varios parámetros relacionados con el diálogo y la metacomunicación están en cierto modo relacionados con la cooperatividad del sistema, pero, sin embargo, no sirven para cuantificarla.

Para medir directamente la cooperatividad se utilizan los parámetros de adecuación contextual definidos por Simpson y Fraser [24]. Varios expertos analizan cada enunciado del sistema para determinar si viola una o varias de las máximas de Grice relativas a la cooperatividad (véase [13]), a saber:

- *Cantidad* de información: incluya en sus enunciados toda la información necesaria (a los efectos de la comunicación en ese preciso momento); no incluya más información de la necesaria.
- *Calidad*: trate de que sus enunciados sean verdaderos; no diga lo que considere que es falso, ni algo de lo que no esté seguro.
- *Pertinencia*: sea pertinente.
- *Forma*: sea perspicuo; evite expresiones oscuras y ambigüedades (evite prolijidad innecesaria); sea breve y metódico.

Bernsen y Dybkjær [1] han adaptado estos principios para su aplicación a los sistemas de diálogo oral.

Los enunciados se clasifican en las categorías de adecuados (no vulneran las máximas de Grice), inadecuados (infringen una o varias máximas), adecuado/inadecuado (los expertos no logran convenir en su clasificación), incomprensible (el contenido del enunciado no puede comprenderse en el contexto del diálogo), o fallo total (el sistema no emite respuesta lingüística). Valga decir que la clasificación no siempre resulta sencilla, por lo que puede ser necesario aplicar principios de interpretación.

**Cuadro 3 – Parámetros de interacción relacionados con la cooperatividad**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
<i>CA:AP,</i> <i>CA:IA,</i> <i>CA:TF,</i> <i>CA:IC,</i>  <i>%CA:AP,</i> <i>%CA:IA,</i> <i>%CA:TF,</i> <i>%CA:IC</i>	Adecuación en el contexto	<p>Número total o porcentaje de enunciados que se consideran adecuados en el contexto del diálogo inmediato, el cual se determina clasificando los enunciados según cumplan o infrinjan algunas de las máximas de Grice sobre cooperatividad:</p> <ul style="list-style-type: none"> <li>• <i>CA:AP</i>: adecuado, no vulnera las máximas de Grice, no es inesperadamente notorio ni resulta extraño de alguna manera.</li> <li>• <i>CA:IA</i>: inadecuado, infringe una o varias de las máximas de Grice.</li> <li>• <i>CA:TF</i>: fallo total, sin respuesta lingüística.</li> <li>• <i>CA:IC</i>: incomprensible, el experto que anota no puede entender el contenido.</li> </ul> <p>Para mayor información, véase [24][8][9]; la clasificación es similar a la utilizada en [14].</p>	Enunciado	Experto

#### 7.4 Parámetros relacionados con tareas

Los servicios modernos actuales permiten la interacción orientada a tareas entre el sistema y el usuario, y el éxito de las mismas es fundamental para que el servicio resulte útil. El lugar ideal para determinar el éxito de las tareas es un laboratorio, en el que se hagan pruebas específicas a tal efecto (véase la Rec. UIT-T P.851). Ahora bien, las mediciones realistas deberán tener en cuenta las posibles variaciones con respecto al caso ideal: el usuario no ha prestado la debida atención a las instrucciones del caso, se ha despistado y no ha escuchado lo enunciado por el sistema o ha sido imposible realizar la tarea y ha sido necesario modificarla en el curso del diálogo.

La modificación de la tarea experimental se contempla en la mayoría de las definiciones de éxito de tareas que se encuentran en la documentación. El resultado será satisfactorio si se facilita la respuesta correcta a las condiciones establecidas en las instrucciones, si se relajan las condiciones del sistema o del usuario (o ambos) o se establece que no existe solución para la tarea en cuestión. El fallo de tarea puede atribuirse provisionalmente al comportamiento del sistema o del usuario, estando este último influenciado por el del sistema.

Un método diferente para determinar el éxito de una tarea es el coeficiente  $\kappa$ . Se utiliza un método de comprensión de voz basado en atributos (conceptos, posiciones) a los que se asignarán valores permitidos en el curso del diálogo entre el sistema y el usuario. Los pares de atributos y valores asignados se denominan pares atributo-valor (AVP). El conjunto de todos los atributos disponibles y sus correspondientes valores asignados por la tarea (conjunto denominado matriz atributo-valor, AVM) describe completamente una tarea, que podrá realizarse con la ayuda del sistema. A fin de determinar el coeficiente  $\kappa$  se crea una matriz de errores  $M(i,j)$  para los atributos en las respuestas correctas (definición del caso) y en la solución notificada (fichero registro del diálogo). Posteriormente podrá calcularse a partir de esta matriz la concordancia entre las respuestas correctas y los resultados obtenidos  $P(A)$  y la coincidencia por azar  $P(E)$  (véase el cuadro 4). Esta matriz  $M(i,j)$  puede calcularse por separado para cada diálogo o para un conjunto de diálogos correspondientes a un sistema específico o una determinada configuración del sistema.

El coeficiente  $\kappa$  se basa en la disponibilidad de un plan sencillo de codificación de tareas, a saber, la matriz AVM. No obstante, algunas tareas no pueden caracterizarse tan sencillamente, en cuyo caso habrá que utilizar un procedimiento más elaborado, que dependerá del tipo de tarea en cuestión.

**Cuadro 4 – Parámetros de interacción relacionados con las tareas**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
TS	Éxito de la tarea	<p>La clasificación del éxito de una tarea de acuerdo con el resultado obtenido por el usuario al final del diálogo, siempre que este resultado pueda lograrse con la ayuda del sistema. La clasificación indica si se alcanzó o no el objetivo, y el supuesto origen de los problemas:</p> <ul style="list-style-type: none"> <li>• <i>TS:S</i>: alcanzado (existen soluciones para esta tarea).</li> <li>• <i>TS:SCs</i>: alcanzado gracias a la relajación de condiciones por el sistema.</li> <li>• <i>TS:SCu</i>: alcanzado gracias a la relajación de condiciones por el usuario.</li> <li>• <i>TS:SCsCu</i>: alcanzado gracias a la relajación de condiciones por el sistema y el usuario.</li> <li>• <i>TS:SN</i>: alcanzado en el sentido que se ha descubierto que no existe solución.</li> <li>• <i>TS:F<sub>s</sub></i>: No alcanzado debido al comportamiento del sistema o su incapacidad.</li> <li>• <i>TS:F<sub>u</sub></i>: No alcanzado debido al comportamiento no cooperativo del usuario.</li> </ul> <p>Véase también [8][7][24].</p>	Diálogo	Experto
κ	Coeficiente kappa	<p>Porcentaje que indica en qué medida se realiza satisfactoriamente la tarea, de acuerdo con la estadística kappa, la cual se calcula analizando si el resultado obtenido de la AVM al final del diálogo coincide con la AVM del caso (respuestas correctas). Se crea una matriz de errores <math>M(i,j)</math> para los atributos en el resultado y en las respuestas esperadas, siendo <math>T</math> el número de totales en <math>M</math>, y <math>t_i</math> la suma de los totales en la columna <math>i</math> de <math>M</math>. Así pues:</p> $\kappa = \frac{P(A) - P(E)}{1 - P(E)}$ <p>siendo <math>P(A)</math> la proporción de veces que la AVM del diálogo real coincide con la respuesta correcta, <math>P(A) = \sum_{i=1}^n \frac{M(i,i)}{T}</math>.</p> <p><math>P(E)</math> puede calcularse a partir de la proporción de veces que pueden coincidir por azar,</p> $P(E) = \sum_{i=1}^n \left(\frac{t_i}{T}\right)^2.$ <p>[31][4]</p>	Diálogo o conjunto de diálogo	Experto

## 7.5 Parámetros relacionados con la adquisición de voz

La capacidad de adquisición de voz de un sistema de diálogo oral viene dada por su capacidad de reconocer palabras y enunciados, y de extraer el significado de la cadena reconocida ("comprensión de voz"). Pueden distinguirse dos tipos de reconocimiento automático de voz, a saber, el sistema reconocimiento de palabras, que consiste en extraer sólo palabras de la voz cuando el usuario dice palabras sueltas (reconocimiento de palabras sueltas) o habla normalmente (detección de palabras clave) y, por otra parte, el sistema reconocimiento continuo de voz, que consiste en reconocer frases o enunciados enteros. Muchos sistemas de comprensión de voz se basan en los pares atributo-valor (véase la cláusula 7.4). Los parámetros que se describen en el párrafo siguiente se utilizan tanto en el reconocimiento de voz como en la comprensión de voz.

Por regla general, los sistemas de reconocimiento continuo de voz generan como resultado una cadena de palabras posible. Para juzgar si la cadena representa correctamente lo que ha dicho el usuario, el experto tendrá que facilitar una transcripción de referencia. Para cada enunciado, se realiza en primer lugar una alineación a nivel de palabra entre la cadena posible y la de referencia, utilizando para ello el algoritmo de concordancia de programación dinámica (DP) [19] [20]. Partiendo de esta alineación se cuenta el número de palabras determinadas correctamente ( $c_w$ ), el número de sustituciones ( $s_w$ ), el número de inserciones ( $i_w$ ) y el número de supresiones ( $d_w$ ). Estas cifras se comparan con el número total de palabras en el texto de referencia ( $n_w$ ), a fin de obtener dos medidas distintas de la calidad del reconocimiento, a saber, la tasa de palabras erróneas *WER* y la tasa de palabras acertadas *WA*, véase el cuadro 5.

Es posible definir medidas complementarias de la calidad a nivel de frase, esto es la tasa de frases correctas *SA* o la tasa de frases erróneas *SER* (véase el cuadro 5). En general, *SA* es menor que *WA*, dado que una sola palabra mal reconocida en una frase repercute en el parámetro *SA*. No obstante, en algunos casos puede ser mayor que la tasa de palabras acertadas, especialmente cuando se reconocen correctamente muchas frases de una sola palabra. En Strik y otros [26] [27] se pone de relieve el hecho de que la *SER* y la *SA* perjudican a todo un enunciado cuando falla únicamente el reconocimiento de una sola palabra; el problema puede sortearse mediante los parámetros *NES* y *WES* (véase el cuadro 5). Cuando los enunciados no están separados en frases, las mediciones de frases se podrán aplicar a todo el enunciado.

Los sistemas de reconocimiento de palabras sueltas generan como resultado una palabra o enunciado posible. Las palabras obtenidas y las pronunciadas pueden compararse directamente a fin de obtener indicadores de calidad similares a los del caso de reconocimiento continuo, sin contar las inserciones. En lugar de las inserciones, pueden contarse el número de falsas alarmas en un determinado periodo de tiempo (véase Leeuwen y Steeneken [28]). También es posible calcular *WA* y *WER* para las palabras clave únicamente, cuando el sistema de reconocimiento funciona en modo detección de palabras clave.

Pueden distinguirse dos métodos comunes para evaluar la comprensión de la voz. El primero consiste en clasificar las respuestas del sistema a las preguntas formuladas por el usuario en las categorías de respuesta correcta, parcialmente correcta, incorrecta o fallo. Las distintas categorías de respuestas pueden combinarse en las mediciones que se han utilizado en el programa US DARPA (véase el cuadro 5). El segundo método consiste en clasificar la capacidad de análisis del sistema, en cuanto a los enunciados analizados correctamente o los pares atributo-valor (AVP) identificados correctamente. A partir de los AVP identificados pueden calcularse indicadores generales, tales como la precisión de conceptos (*CA*), la tasa de conceptos erróneos (*CER*) o la precisión de comprensión (*UA*). Todos los parámetros se describen en el cuadro 5.

**Cuadro 5 – Parámetros de interacción relacionados con la adquisición de voz**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
<i>WER, WA</i>	Tasa de palabras erróneas, tasa de palabras acertadas	<p>Porcentaje de palabras que se han reconocido correctamente, según la ortografía del enunciado obtenido y el de referencia (transcripción), una vez alineados mediante el algoritmo "sclite" (véase [18]). Sea <math>n_w</math> el número total de palabras de todos los enunciados del usuario durante un diálogo y sean <math>s_w</math>, <math>d_w</math> e <math>i_w</math> el número de palabras sustituidas, suprimidas e insertadas, respectivamente, la tasa de palabras erróneas y la tasa de palabras acertadas vienen dadas por las siguientes expresiones:</p> $WER = \frac{s_w + i_w + d_w}{n_w}$ $WA = 1 - \frac{s_w + i_w + d_w}{n_w} = 1 - WER$ <p>Véase [24]; para mayor información sobre cómo calcular estos parámetros en el caso de reconocimiento de palabras sueltas, véase [28].</p>	Palabra	Instrumento/ Experto
<i>SER, SA</i>	Tasa de frases erróneas, tasa de frases acertadas	<p>Porcentaje de frases enteras que se han reconocido correctamente. Sea <math>n_s</math> el número total de frases, y sean <math>s_s</math>, <math>i_s</math> y <math>d_s</math> el número de frases sustituidas, insertadas y suprimidas, respectivamente:</p> $SER = \frac{s_s + i_s + d_s}{n_s}$ $SA = 1 - \frac{s_s + i_s + d_s}{n_s} = 1 - SER$ <p>Véase [24].</p>	Enunciado	Instrumento/ Experto
<i>NES</i>	Número de errores por frase	<p>Promedio de errores de reconocimiento en una frase. Sean <math>s_w(k)</math>, <math>i_w(k)</math> y <math>d_w(k)</math> el número de palabras sustituidas, insertadas y suprimidas en la frase <math>k</math>, en ese caso:</p> $NES(k) = s_w(k) + i_w(k) + d_w(k)$ <p>El promedio <i>NES</i> viene dado por la expresión:</p> $NES = \frac{\sum_{k=1}^{\# \text{ turnos usuario}} NES(k)}{\# \text{ turnos usuario}} =$ $= \frac{WER \cdot \# \text{ palabras usuario}}{\# \text{ turnos usuario}}$ <p>Véase [26].</p>	Enunciado	Instrumento/ Experto

**Cuadro 5 – Parámetros de interacción relacionados con la adquisición de voz**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
<i>WES</i>	Palabras erróneas por frase	<p>Idéntico al <i>NES</i> pero normalizado al número de palabras en la frase <math>k</math>, <math>w(k)</math>:</p> $WES(k) = \frac{NES(k)}{w(k)}$ <p>El promedio <i>WES</i> viene dado por la expresión:</p> $WES = \frac{\sum_{k=1}^{\# \text{turnos usuario}} WES(k)}{\# \text{turnos usuario}}$ <p>Véase [26].</p>	Palabra	Instrumento/ Experto
<i>AN:CO</i> , <i>AN:IN</i> , <i>AN:PA</i> , <i>AN:FA</i> , <i>%AN:CO</i> , <i>%AN:IN</i> , <i>%AN:PA</i> , <i>%AN:FA</i>	Número o porcentaje de respuestas del sistema correctas, incorrectas, parcialmente correctas y sin respuesta	<p>Número o porcentaje total de preguntas del usuario que el sistema respondió</p> <ul style="list-style-type: none"> <li>• correctamente (<i>AN:CO</i>);</li> <li>• incorrectamente (<i>AN:IC</i>);</li> <li>• correctamente en parte (<i>AN:PA</i>);</li> <li>• no respondió (<i>AN:FA</i>),</li> </ul> <p>por diálogo, véase [21][12][14].</p>	Enunciado	Experto
<i>DARPA<sub>s</sub></i> , <i>DARPA<sub>me</sub></i>	Nota DARPA, error modificado DARPA	<p>Mediciones con arreglo a la iniciativa de comprensión de voz DARPA, modificada por Skowronek [25][17] para tener en cuenta las respuestas parcialmente correctas:</p> $DARPA_s = \frac{AN : CO - AN : IC}{\# \text{ preguntas usuario}}$ $DARPA_{me} = \frac{AN : FA + 2 \cdot (AN : IC + AN : PA)}{\# \text{ preguntas usuario}}$ <p>[21][12][25]</p>	Enunciado	Experto
<i>PA:CO</i> , <i>PA:PA</i> , <i>PA:IC</i> , <i>%PA:CO</i> , <i>%PA:PA</i> , <i>%PA:IC</i>	Número de enunciados del usuario analizados correctamente, en parte correctamente, incorrectamente	<p>Evaluación del número de conceptos (pares atributo-valor, AVP) de un enunciado extraídos por el sistema:</p> <ul style="list-style-type: none"> <li>• <i>PA:CO</i>: el sistema ha entendido correctamente todos los conceptos de un enunciado del usuario.</li> <li>• <i>PA:PA</i>: el sistema ha entendido correctamente al menos uno de los conceptos enunciados por el usuario (no los ha entendido todos).</li> <li>• <i>PA:IC</i>: el sistema no ha comprendido ningún concepto de un enunciado del usuario.</li> </ul> <p>Se expresa como el número o porcentaje total de enunciados del usuario en un diálogo que se han analizado correctamente, correctamente en parte o incorrectamente. [7]</p>	Enunciado	Experto

**Cuadro 5 – Parámetros de interacción relacionados con la adquisición de voz**

Sigla	Nombre	Definición	Nivel de interacción	Método de medición
<i>CA, CER</i>	Precisión de conceptos, tasa de conceptos erróneos	<p>Porcentaje de unidades semánticas comprendidas correctamente por diálogo. Los conceptos se definen mediante pares atributo-valor (AVP), siendo <math>n_{AVP}</math> el número total de AVP, y <math>s_{AVP}</math>, <math>i_{AVP}</math> y <math>d_{AVP}</math> el número de AVP sustituidos, insertados y suprimidos, respectivamente. La precisión de conceptos y la tasa de conceptos erróneos se calculan mediante las siguientes expresiones:</p> $CA = 1 - \frac{s_{AVP} + i_{AVP} + d_{AVP}}{n_{AVP}}$ $CER = \frac{s_{AVP} + i_{AVP} + d_{AVP}}{n_{AVP}}$ <p>Véase [9][24][3][2].</p>	Enunciado	Experto
<i>UA</i>	Precisión de comprensión	<p>Porcentaje de enunciados de usuario para los cuales se han extraído correctamente todas las unidades semánticas (AVP):</p> $UA = \frac{PA:CO}{\# \text{ turnos usuario}}$ <p>Véase [32].</p>	Enunciado	Experto

## 7.6 Otros parámetros

La mayoría de los parámetros de interacción enumerados en los cuadros describen el comportamiento del sistema, dado que el objetivo que se persigue es, obviamente, evaluar la calidad del sistema y del servicio. Sin embargo, es posible definir parámetros relacionados con el usuario que, aún siendo específicos del grupo de usuarios que realiza las pruebas, pueden guardar una estrecha relación con la calidad percibida por el usuario.

Si se consideran los distintos aspectos de la calidad de un servicio basado en SDS con arreglo a lo indicado en 5.3/P.851, se observa que varios aspectos de calidad no quedan contemplados en los parámetros de interacción. No existen parámetros directamente relacionados con la utilidad, la satisfacción del usuario, la aceptabilidad o la calidad de voz reproducida. Hasta ahora ha habido muy pocos métodos de análisis paramétrico de la calidad de la voz reproducida (ya sea por concatenación o sintetizada). Se han definido mediciones instrumentales relacionadas con la inteligibilidad de la voz, por ejemplo, en la Norma CEI 60268-16 [15], pero todavía no se han aplicado a la telefonía. Se han propuesto mediciones de la dificultad de concatenación que pueden calcularse a partir del texto de entrada y la base de datos de voz de un sistema de síntesis por concatenación [5]. Si bien estas mediciones muestran grandes correlaciones con las notas medias de opinión (MOS) obtenidas mediante experimentos de audición, tales mediciones dependen sobremanera del sintetizador vocal y su corpus de concatenación.

## **8 Interpretación de los valores de los parámetros de interacción**

Si bien los parámetros de interacción, como los definidos en este suplemento, son importantes en el diseño, optimización y mantenimiento del sistema, no están directamente relacionados con la calidad que percibe el usuario humano. Por consiguiente, el conjunto de parámetros de interacción debería complementarse con opiniones de usuarios sobre diferentes aspectos de calidad, tal y como se describe en la Rec. UIT-T P.851. Sólo así podrá obtenerse información válida sobre la calidad de servicios basados en sistemas de diálogo oral.

Una posible interpretación de los valores de los parámetros de interacción podría basarse en resultados experimentales, que, valga decir, a menudo son específicos del sistema o servicio del caso. Por ejemplo, un aumento en el número de mensajes de expiración podría indicar que el usuario no sabe qué decir en ciertos puntos del diálogo, o que está desconcertado por las acciones del sistema [29]. Un aumento de las tentativas de intervención podría significar simplemente que el usuario ha aprendido que puede interrumpir al sistema. En cambio, si el número de tentativas se reduce podría indicar que el usuario no sabe qué decir al sistema. Los enunciados prolongados del usuario podrían deberse a que se le ha atribuido mucha iniciativa al usuario. La disminución de los valores de los parámetros relacionados con la metacomunicación (en particular, la metacomunicación iniciada por el usuario) podría atribuirse a una mayor robustez del sistema, mayor fluidez del diálogo y más eficiencia en la comunicación [1].

## BIBLIOGRAFÍA

- [1] BERNSEN, N.O., DYBKJÆR, H., DYBKJÆR, L.: *Designing interactive speech systems: From first ideas to user testing*, Springer, DE-Berlín, 1998.
- [2] BILLI, R., CASTAGNERI, G., DANIELI, M.: Field trial evaluations of two different information inquiry systems, *Proc. 3rd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA'96)*, US-Basking Ridge NJ, pp. 129-134, 1996.
- [3] BOROS, M., ECKERT, W., GALLWITZ, F., GORZ, G., HANRIEDER, G., NIEMANN, H.: Towards understanding spontaneous speech: Word accuracy vs. concept accuracy, *Proc. 4th Int. Conf. on Spoken Language Processing (ICSLP'96)*, IEEE, US-Piscataway NJ, 2, pp. 1009-1012, 1996.
- [4] CARLETTA, J.: Assessing agreement of classification tasks: The kappa statistics, *Computational Linguistics*, Vol. 22(2), pp. 249-254, 1996.
- [5] CHU, M., PENG, H.: An objective measure for estimating MOS of synthesized speech, *Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia)*, DK-Aalborg, 3, pp. 2087-2090, 2001.
- [6] COOKSON, S.: Final evaluation of VODIS – Voice operated data inquiry system, *Proc. of Speech'88, 7th FASE Symposium*, UK-Edinburgh, 4, pp. 1311-1320, 1988.
- [7] DANIELI, M., GERBINO, E.: Metrics for evaluating dialogue strategies in a spoken language system, *Empirical Methods in Discourse Interpretation and Generation. Papers from the 1995 AAAI Symposium*, US-Stanford CA, AAAI Press, US-Menlo Park CA, pp. 34-39, 1995.
- [8] FRASER, N.: Assessment of interactive systems, *Handbook on Standards and Resources for Spoken Language Systems* (D. Gibbon, R. Moore and R. Winski, eds.), Mouton de Gruyter, DE-Berlín, pp. 564-615, 1997.
- [9] GERBINO, E., BAGGIA, P., CIARAMELLA, A., RULLENT, C.: Test and evaluation of a spoken dialogue system, *Proc. Int. Conf. Acoustics Speech and Signal Processing (ICASSP'93)*, IEEE, US-Piscataway NJ, 2, pp. 135-138, 1993.
- [10] GIBBON, D., MOORE, R., WINSKI, R., Eds.: *Handbook on Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, DE-Berlín, 2000.
- [11] GLASS, J., POLIFRONI, J., SENEFF, S., ZUE, V.: Data collection and performance evaluation of spoken dialogue systems: The MIT experience, *Proc. 6th Int. Conf. on Spoken Language Processing (ICSLP 2000)*, CN-Beijing, 4, pp. 1-4., 2000.
- [12] GOODINE, D., HIRSCHMAN, L., POLIFRONI, J., SENEFF, S., ZUE, V.: Evaluating interactive spoken language systems, *Proc. 2nd Int. Conf. on Spoken Language Processing (ICSLP'92)*, CA-Banff, 1, pp. 201-204, 1992.
- [13] GRICE, H.P.: Logic and conversation, *Syntax and Semantics, Vol. 3: Speech Acts* (P. Cole and J.L. Morgan, eds.), Academic Press, US-New York NY, pp. 41-58, 1975.
- [14] HIRSCHMAN, L., PAO, C.: The cost of errors in a spoken language system, *Proc. 3rd Europ. Conf. on Speech Communication and Technology (Eurospeech'93)*, DE-Berlín, 2, pp. 1419-1422, 1993.
- [15] IEC 60268-16 (2003), *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*. International Electrotechnical Commission, CH-Ginebra.

- [16] KAMM, C.A., LITMAN, D.J., WALKER, M.A.: From novice to expert: The effect of tutorials on user expertise with spoken dialogue systems, *Proc. 5th Int. Conf. on Spoken Language Processing (ICSLP'98)*, AU-Sydney, 4, pp. 1211-1214, 1998.
- [17] MÖLLER, S.: *Quality of telephone-based spoken dialogue systems*. Springer, US-New York NY, 2005.
- [18] NIST Speech Recognition Scoring Toolkit, *Speech recognition scoring toolkit*, National Institute of Standards and technology, <http://www.nist.gov/speech/tools>, US-Gaithersburg MD, 2001.
- [19] PICONE, J., DODDINGTON, G.R., PALLETT, D.S.: Phone-mediated word alignment for speech recognition evaluation, *IEEE Trans. Acoustics Speech and Signal Processing*, Vol. 38(3), pp. 559-562, 1990.
- [20] PICONE, J., GOUDIE-MARSHALL, K.M., DODDINGTON, G.R., FISHER, W.: Automatic text alignment for speech system evaluation, *IEEE Trans. Acoustics Speech and Signal Processing*, Vol. 34(4), pp. 780-784, 1986.
- [21] POLIFRONI, J., HIRSCHMAN, L., SENEFF, S., ZUE, V.: Experiments in evaluating interactive spoken language systems, *Proc. DARPA Speech and Natural Language Workshop*, US-Harriman CA, pp. 28-33, 1992.
- [22] PRICE, P.J., HIRSCHMAN, L., SHRIBERG, E., WADE, E.: Subject-based evaluation measures for interactive spoken language systems, *Proc. DARPA Speech and Natural Language Workshop*, US-Harriman CA, pp. 34-39, 1992.
- [23] SAN-SEGUNDO, R., MONTERO, J.M., COLÁS, J., GUTIÉRREZ, J., RAMOS, J.M., PARDO, J.M.: Methodology for dialogue design in telephone-based spoken dialogue systems: A Spanish train information system, *Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia)*, DK-Aalborg, 3, pp. 2165-2168, 2001.
- [24] SIMPSON, A., FRASER, N.M.: Black box and glass box evaluation of the SUNDIAL system, *Proc. 3rd Europ. Conf. on Speech Communication and Technology (Eurospeech'93)*, DE-Berlin, 2, pp. 1423-1426, 1993.
- [25] SKOWRONEK, J.: *Entwicklung von Modellierungsansätzen zur Vorhersage der Dienstqualität bei der Interaktion mit einem natürlichsprachlichen Dialogsystem*. Diploma thesis (unpublished), Institut für Kommunikationsakustik, Ruhr-Universität, DE-Bochum, 2002.
- [26] STRIK, H., CUCCHIARINI, C., KESSENS, J.M.: Comparing the performance of two CSRs: How to determine the significance level of the differences, *Proc. 7th Europ. Conf. on Speech Communication and Technology (Eurospeech 2001 – Scandinavia)*, DK-Aalborg, 3, pp. 2091-2094, 2001.
- [27] STRIK, H., CUCCHIARINI, C., KESSENS, J.M.: Comparing the recognition performance of CSRs: In search of an adequate metric and statistical significance test, *Proc. 6th Int. Conf. on Spoken Language Processing (ICSLP 2000)*, CN-Beijing, 4, pp. 740-743, 2000.
- [28] VAN LEEUWEN, D., STEENEKEN, H.: Assessment of recognition systems, *Handbook on Standards and Resources for Spoken Language Systems* (D. Gibbon, R. Moore and R. Winski, eds.), Mouton de Gruyter, DE-Berlin, pp. 381-407, 1997.
- [29] WALKER, M.A., FROMER, J., DI FABBRIZIO, G., MESTEL, C., HINDLE, D.: What can I say?: Evaluating a spoken language interface to email, *Human Factors in Computing Systems. CHI'98 Conference Proc.*, US-Los Angeles CA, ACM, US-New York NY, pp. 582-589, 1998.

- [30] WALKER, M.A., LITMAN, D.J., KAMM, C.A., ABELLA, A.: Evaluating spoken dialogue agents with PARADISE: Two case studies, *Computer Speech and Language*, Vol. 12(3), pp. 317-347, 1998.
- [31] WALKER, M.A., LITMAN, D.J., KAMM, C.A., ABELLA, A.: PARADISE: A framework for evaluating spoken dialogue agents, *Proc. of the 35th Ann. Meeting of the Assoc. for Computational Linguistics*, ES-Madrid, pp. 271-280, 1997.
- [32] ZUE, V., SENEFF, S., GLASS, J.R., POLIFRONI, J., PAO, C., HAZEN, T.J., HETHERINGTON, L.: JUPITER: A telephone-based conversational interface for weather information, *IEEE Trans. Speech and Audio Processing*, Vol. 8(1), pp. 85-96, 2000.





## SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie D	Principios generales de tarificación
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedios
Serie I	Red digital de servicios integrados
Serie J	Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedios
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	Gestión de las telecomunicaciones, incluida la RGT y el mantenimiento de redes
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
<b>Serie P</b>	<b>Calidad de transmisión telefónica, instalaciones telefónicas y redes locales</b>
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos, comunicaciones de sistemas abiertos y seguridad
Serie Y	Infraestructura mundial de la información, aspectos del protocolo Internet y Redes de la próxima generación
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación