

Unión Internacional de Telecomunicaciones

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

P.800.2

(07/2016)

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

Métodos de evaluación objetiva y subjetiva de la
calidad vocal y de video

**Interpretación y notificación de las notas
medias de opinión**

Recomendación UIT-T P.800.2

UIT-T



RECOMENDACIONES UIT-T DE LA SERIE P
**CALIDAD DE LA TRANSMISIÓN TELEFÓNICA, INSTALACIONES TELEFÓNICAS
Y REDES DE LÍNEAS LOCALES**

Vocabulario y efectos de los parámetros de transmisión sobre la opinión de los clientes	P.10–P.19
Características de los terminales vocales	P.30–P.39
Sistemas de referencia	P.40–P.49
Aparatos para mediciones objetivas	P.50–P.59
Medidas electroacústicas objetivas	P.60–P.69
Medidas relativas a la sonoridad vocal	P.70–P.79
Métodos de evaluación objetiva y subjetiva de la calidad vocal	P.80–P.89
Características de los terminales vocales	P.300–P.399
Aparatos para mediciones objetivas	P.500–P.599
Métodos de evaluación objetiva y subjetiva de la calidad vocal y de video	P.800–P.899
Calidad audiovisual en servicios multimedia	P.900–P.999
Aspectos de calidad de transmisión y de calidad de servicio en los puntos extremos de redes de protocolo Internet	P.1000–P.1099
Comunicaciones implicando vehículos	P.1100–P.1199
Modelos y herramientas para la evaluación de la calidad de los medios secuenciados	P.1200–P.1299
Evaluación de las telerreuniones	P.1300–P.1399
Directrices para el análisis, la evaluación y la información estadísticos de las mediciones de la calidad	P.1400–P.1499
Métodos de evaluación objetiva y subjetiva de la calidad de servicios distintos a los servicios de voz o de vídeo	P.1500–P.1599

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

Recomendación UIT-T P.800.2

Interpretación y notificación de las notas medias de opinión

Resumen

La Recomendación UIT-T P.800.2 presenta algunos de los tipos más comunes de nota media de opinión (MOS) y describe la información mínima que debería adjuntarse a los valores MOS para poderlos interpretar correctamente.

Historia

Edición	Recomendación	Aprobación	Comisión de Estudio	ID único*
1.0	ITU-T P.800.2	2013-05-14	12	11.1002/1000/11934
2.0	ITU-T P.800.2	2016-07-29	12	11.1002/1000/12973

Palabras clave

ACR, experimento subjetivo, índice por categorías absolutas, modelo objetivo, MOS, nota media de opinión, notificación.

* Para acceder a la Recomendación, sírvase digitar el URL <http://handle.itu.int/> en el campo de dirección del navegador, seguido por el identificador único de la Recomendación. Por ejemplo, <http://handle.itu.int/11.1002/1000/11830-en>.

PREFACIO

La Unión Internacional de Telecomunicaciones (UIT) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones y de las tecnologías de la información y la comunicación. El Sector de Normalización de las Telecomunicaciones de la UIT (UIT-T) 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 Recomendación, 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 Recomendación es voluntaria. Ahora bien, la Recomendación 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 Recomendación 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 Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT no ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. 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 en la dirección <http://www.itu.int/ITU-T/ipr/>.

© UIT 2019

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

ÍNDICE

	Página
1 Alcance	1
2 Referencias	1
3 Definiciones	1
3.1 Términos definidos en otros documentos	1
3.2 Términos definidos en esta Recomendación	1
4 Abreviaturas y acrónimos	1
5 Convenios	2
6 Introducción	2
7 Valores MOS subjetivos	2
8 Interpretación de los valores MOS	4
9 Consideraciones relativas al vídeo	5
10 Análisis estadístico de MOS	6
11 Valores MOS objetivos	6
12 Notificación de los valores MOS subjetivos	6
13 Notificación de los valores MOS objetivos	8
14 Notas	8
Bibliografía	9

Recomendación UIT-T P.800.2

Interpretación y notificación de las notas medias de opinión

1 Alcance

Esta Recomendación presenta algunos de los tipos más comunes de nota media de opinión (MOS) y describe la información mínima que debería adjuntarse a los valores MOS para poderlos interpretar correctamente.

El presente texto no pretende ofrecer orientaciones definitivas para la realización de pruebas subjetivas u objetivas. En la bibliografía que figura al final de esta Recomendación se proporciona información más detallada.

2 Referencias

Las siguientes Recomendaciones del UIT-T y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes. En esta Recomendación, la referencia a un documento, en tanto que autónomo, no le otorga el rango de una Recomendación.

[UIT-T P.800.1] Recomendación UIT-T P.800.1 (2006), *Terminología de las notas medias de opinión*.

3 Definiciones

3.1 Términos definidos en otros documentos

Ninguno.

3.2 Términos definidos en esta Recomendación

En la presente Recomendación se definen los siguientes términos:

3.2.1 condición: uno de una serie de casos de uso que se evalúan en un experimento subjetivo; a menudo se le denomina circuito de referencia hipotético (HRC) en experimentos de vídeo.

3.2.2 subcondición: subconjunto de una condición que se define por una característica específica del caso de uso, por ejemplo, material vocal de un hablante en particular.

3.2.3 sujeto: participante en un experimento subjetivo.

3.2.4 voto: respuesta de un sujeto a una pregunta en una escala de evaluación para una muestra de prueba individual o interacción.

4 Abreviaturas y acrónimos

En esta Recomendación se utilizan las siguientes siglas y acrónimos:

ACR Índices por categorías absolutas (*absolute category rating*)

DCR Índices por categorías de degradación (*degradation category rating*)

DMOS	Nota media de opinión sobre las degradaciones (<i>degradation mean opinion score</i>)
HRC	Circuito de referencia hipotético (<i>hypothetical reference circuit</i>)
MOS	Nota media de opinión (<i>mean opinion score</i>)
MUSHRA	Ensayo multiestímulo con referencia y patrón ocultos (<i>multi-stimulus test with hidden reference and anchor</i>)
QCIF	Cuarto de formato intermedio común (<i>quarter common intermediate format</i>)
SSCQE	Método de evaluación de calidad continua de estímulo único (<i>single stimulus continuous quality evaluation</i>)
VGA	Matriz gráfica de vídeo (<i>video graphics array</i>)

5 Convenios

Ninguno.

6 Introducción

La cuantificación de la calidad de audio y de vídeo es inherentemente subjetiva. Esto implica que el valor de referencia de la calidad de audio y de vídeo se determinará en función de la opinión del usuario. Ahora bien, la opinión sobre lo que es "bueno" puede ser muy diferente de una persona a otra – no hay opiniones correctas o incorrectas.

Antes de implantar una nueva tecnología de transmisión de audio o vídeo, conviene evaluar la calidad de transmisión mediante uno o más experimentos subjetivos. El propósito de un experimento subjetivo es recabar la opinión de numerosas personas ("sujetos") acerca del funcionamiento del sistema respecto de una serie de casos de uso bien definidos ("condiciones")¹. La nota media de opinión (MOS) para una condición dada es simplemente el promedio de las opiniones ("votos") recabadas para ese caso de uso particular.

Los algoritmos de medición objetiva de la calidad tienen por objeto predecir, en un experimento subjetivo, el valor MOS que produciría una señal de entrada determinada. Así, al interpretar un valor MOS obtenido por métodos objetivos, es importante entender el diseño básico del experimento que se pretende predecir.

Existen diversos tipos de valores MOS y de metodologías de prueba que permiten obtenerlos. El propósito de esta Recomendación es ofrecer al lector una reseña de los principales aspectos que se han de tener en cuenta al interpretar los valores de MOS y de la información mínima que debe adjuntarse a los valores MOS en el momento de su notificación.

7 Valores MOS subjetivos

Tipos de MOS

Si bien existe la falsa creencia de que los valores MOS solo se refieren a los servicios de voz, el proceso de evaluación de la calidad por los sujetos del experimento puede aplicarse también a los servicios de vídeo y audio en general. También es posible pedir a los sujetos que evalúen la calidad audiovisual global de un servicio. La UIT ha elaborado varias normas en las que se describen diferentes aspectos de las pruebas subjetivas para aplicaciones de vídeo y audio en general, además de las aplicaciones de voz. Esas normas se enumeran en la bibliografía.

¹ En los experimentos de vídeo, las condiciones se denominan a menudo circuitos de referencia hipotéticos (HRC).

Los experimentos subjetivos se dividen en general en dos tipos: pasivos e interactivos. En un experimento subjetivo pasivo, se presentan a los sujetos muestras de prueba pregrabadas que representan las condiciones de interés. Se pide a los sujetos que escuchen y/o vean pasivamente el material de la prueba y expresen su opinión al respecto utilizando la escala de evaluación que se les ha suministrado. En un experimento interactivo, dos o más sujetos entablan una conversación utilizando equipos diseñados para emular los casos de uso de interés. A menudo se les asignan tareas para estimular la conversación y la interacción. La mayoría de los experimentos tienden a ser de tipo pasivo. Sin embargo, hay algunos aspectos de la experiencia del usuario, como los efectos del retardo y el eco, que solo se manifiestan en el contexto de una conversación.

Metodología de pruebas y escalas de evaluación

En un experimento subjetivo, se pide a los sujetos que expresen su opinión utilizando una "escala de evaluación". La finalidad de la escala es traducir la evaluación de la calidad de un sujeto en valores numéricos que pueden promediarse entre sujetos y otros factores del experimento.

Existen varias escalas de evaluación de uso común, si bien el examen de las ventajas relativas de las diferentes escalas no entra dentro del ámbito de aplicación de la presente Recomendación. La escala más comúnmente utilizada es la de los índices por categorías absolutas (ACR) de 5 puntos:

Excelente	5
Buena	4
Regular	3
Mediocre	2
Mala	1

La escala ACR es una escala discreta, lo que significa que la respuesta del sujeto debe ajustarse a uno de los cinco valores mencionados. Sin embargo, como se utiliza el método de promediación para combinar los resultados extraídos de diferentes sujetos, los valores MOS no se limitan a valores enteros. Algunas escalas de evaluación incluyen más de cinco posiciones discretas, mientras que otras permiten que el sujeto proporcione respuestas intermedias en puntos entre las posiciones.

El término "absoluto" en el acrónimo ACR hace referencia a la petición que se hace a los sujetos de evaluar cada muestra de forma independiente. En algunas escalas de evaluación, como la de los índices por categorías de degradación (DCR), se pide la opinión de los sujetos sobre la diferencia entre una muestra procesada mediante la condición de interés y una versión no procesada de la misma muestra. El valor MOS obtenido en un experimento de este tipo se denomina a menudo MOS de degradación o DMOS.

En la mayoría de los diseños experimentales, se pide a los sujetos que evalúen la calidad de muestras cortas de audio o de vídeo. La duración de las muestras suele oscilar entre 6 y 10 segundos, ya que esto proporciona tiempo suficiente para que el sujeto se forje una opinión y al mismo tiempo impide eventuales sesgos en el tramo final de la muestra. Dado que una sola muestra de esta duración difícilmente puede representar una condición completa, se suele pedir a los sujetos que evalúen múltiples muestras de prueba relativas al mismo caso de uso. Por ejemplo, en un experimento de voz, cada condición de red objeto de prueba podría representarse con muestras de voz de tres hablantes masculinos y tres femeninos. En consecuencia, se pueden obtener valores MOS para toda la condición, calculando el promedio entre sujetos y hablantes, o solo para una subcondición, por ejemplo, un hablante en particular o el género del hablante.

Existen métodos de prueba, como la evaluación de calidad continua de estímulo único (SSCQE), que utilizan muestras de prueba de una duración mucho más larga y que, por tanto, requieren que el sujeto actualice continuamente su opinión sobre la calidad durante la reproducción de la muestra de prueba. En este caso, el resultado que se obtiene es una secuencia cronológica de notas sobre la calidad evaluada por cada sujeto y no un único valor de opinión.

Algunas metodologías de prueba requieren que el sujeto responda a un gran número de preguntas. Además de proporcionar más información sobre las condiciones objeto de prueba, esto puede considerarse una parte necesaria del diseño de la prueba. Por ejemplo, en el método de prueba UIT-T P.835 se requiere que el sujeto evalúe por separado la calidad de la voz y la calidad del ruido de una muestra antes de proporcionar una puntuación de calidad global. Se ha comprobado que este método permite obtener resultados sobre los sistemas de supresión de ruido más estables que los del método de prueba ACR de pregunta única.

Conviene destacar que algunas preguntas pueden no estar directamente relacionadas con la calidad y pueden referirse a un aspecto diferente de las comunicaciones, por ejemplo, en el método [b-UIT-T P.800] se define una escala de *esfuerzo* de escucha para los experimentos de voz. Asimismo, en algunos experimentos de conversación, se pide a los sujetos que opinen sobre su experiencia mientras hablan y no mientras escuchan.

8 Interpretación de los valores MOS

Si bien el examen que sigue a continuación se centra principalmente en los valores MOS de voz, muchos de los aspectos planteados en las subsecciones se aplican también a los valores MOS de vídeo, audio y audio-vídeo. En el siguiente apartado se examinan las principales diferencias aplicables al vídeo.

La idea de que un códec vocal específico tiene una valoración MOS específica es otro error común. Este error se debe entre otras cosas al uso generalizado de los modelos objetivos de evaluación de la calidad, que producen resultados con una buena reproducibilidad. Estos modelos están concebidos para predecir o estimar los resultados de los experimentos subjetivos; sin embargo, para un códec determinado con una velocidad binaria determinada, el valor MOS obtenido en un experimento subjetivo puede variar sustancialmente de un experimento a otro. Las razones son múltiples.

En primer lugar, los valores exactos de MOS obtenidos para una condición particular en un experimento subjetivo pueden verse influidos por un gran número de factores, entre los cuales:

- las instrucciones que se dan a los sujetos y la formulación utilizada en la escala de opinión;
- el equipo que se utiliza para presentar el material (microteléfonos, auriculares, altavoces);
- presentación monoaural, dióptica binaural o estereofónica;
- nivel de presentación;
- entorno acústico;
- preparación de los sujetos;
- perfil del sujeto, por ejemplo, edad y nivel de exposición a la tecnología;
- diferencias culturales a la hora de interpretar y usar las escalas de evaluación;
- material de voz (contenido fonético y características del hablante);
- idioma (presencia/ausencia, prevalencia e importancia de determinados sonidos y transiciones).

En segundo lugar, el valor exacto de MOS que se obtiene para una condición determinada en un experimento subjetivo depende de la calidad de las demás condiciones del experimento. Por ejemplo, una condición de códec vocal de la Recomendación UIT-T G.729 podría obtener una puntuación superior a 3,9 en un experimento ACR cuando la mayoría de las demás condiciones son de peor calidad que la UIT-T G.729; por el contrario, la condición de la UIT-T G.729 podría obtener una puntuación significativamente inferior a 3,9 cuando la mayoría de las demás condiciones presentan una calidad superior.

En tercer lugar, si se realiza un experimento con códecs que funcionan en diferentes anchos de banda de audio, la presencia de condiciones de mayor ancho de banda reducirá el MOS obtenido para condiciones con menor ancho de banda de audio. La presencia de un mayor ancho de banda de audio en un experimento de voz se denomina a menudo el "contexto" del experimento. Por ejemplo, una condición de códec vocal de la Recomendación UIT-T G.711 a menudo arrojará una puntuación superior a 4,0 en un experimento ACR de banda estrecha (300-3 700 Hz), mientras que hay más probabilidades de que arroje una puntuación entre 3,5 y 3,7 en un experimento ACR de banda ancha (50-7 000 Hz), debido a la presencia de muestras de banda ancha de mayor calidad.

Estos dos últimos aspectos apuntan al hecho de que en los experimentos los sujetos tienden a adaptar su uso de las escalas de evaluación al contenido del experimento. De hecho, los experimentos mejor diseñados incluyen un periodo de práctica al comienzo del experimento en el que los sujetos escuchan ejemplos de una variedad de condiciones, que abarcan las mejores y las peores.

Una de las principales conclusiones que se derivan de las consideraciones descritas antes es que la comparación directa de los valores MOS obtenidos de experimentos separados solo es pertinente si esos experimentos se han concebido expresamente para ser comparados, e incluso en ese caso, los datos deben analizarse estadísticamente a fin de garantizar la validez de esa comparación.

9 Consideraciones relativas al vídeo

Varias de las anteriores consideraciones relativas a los experimentos subjetivos de voz también se aplican a los experimentos de vídeo. Las condiciones experimentales, a menudo denominadas circuitos de referencia hipotéticos (HRC), definen en general diversas combinaciones de códec de vídeo, velocidad binaria, velocidad de trama y condiciones de transmisión. Entre los factores que influyen en los valores exactos de MOS obtenidos para una condición en particular cabe citar los siguientes:

- el equipo utilizado para presentar el material (tecnología de visualización, frecuencia de actualización, contraste, etc.);
- el entorno de la visualización (color, temperatura y nivel de iluminación);
- distancia de visualización (generalmente expresada como la relación entre la distancia de visualización y la altura de la pantalla);
- contenido de vídeo.

Este último punto es particularmente importante para los experimentos de vídeo. La elección del material de prueba es un factor mucho más importante en los experimentos de vídeo que en los experimentos de voz. Esto se debe a que el contenido de una secuencia de vídeo puede incidir de forma importante en la eficacia de su codificación. Por ejemplo, el contenido informativo en una secuencia deportiva de movimientos rápidos es mucho más alto que en una secuencia de videoconferencia centrada en cabezas y hombros.

En los experimentos de vídeo, el contexto primario viene determinado por la resolución de la imagen de vídeo. En general, los experimentos subjetivos no mezclan diferentes resoluciones y, por lo tanto, los valores MOS de vídeo se refieren a una resolución particular, por ejemplo, 480p o 1080p. Cuando las resoluciones son mixtas, el contexto del experimento se define por la resolución con el mayor número de líneas. En ese caso, es importante determinar si las resoluciones más pequeñas que se muestran son las originales o si se redimensionan ajustándose a la resolución más grande del experimento.

10 Análisis estadístico de MOS

El análisis estadístico de los valores MOS subjetivos no entra dentro del ámbito de aplicación de la presente Recomendación. Sin embargo, es importante adjuntar a los valores MOS la información necesaria para poder realizar un análisis estadístico básico, por ejemplo, el cálculo de un intervalo de confianza para cada condición. Dicha información comprenderá, para una condición o subcondición determinada, el número de votos, el promedio y la desviación típica de los votos.

11 Valores MOS objetivos

El propósito de un modelo de calidad objetivo es predecir el valor MOS que obtendría una señal de audio o vídeo en un experimento subjetivo. Como se mencionó antes, el valor exacto de MOS que se obtiene en un experimento determinado para un códec o cadena de transmisión en particular depende de una multitud de aspectos que tienen que ver con el diseño y la ejecución del experimento. Por consiguiente, los diseñadores de modelos objetivos tienen que predecir experimentos idealizados, es decir, experimentos que se realizan conforme a una metodología de prueba específica, generalmente ACR, y que incluyen una muestra equilibrada de las distorsiones que aparecerán en la zona de interés de la aplicación.

Por ejemplo, para la correspondencia definida en [b-UIT-T P.862.1] se toman los resultados brutos del modelo objetivo de la Recomendación UIT-T P.862 y se comparan con un rango que se habrá determinado promediando los resultados de un gran número de experimentos subjetivos realizados según el método ACR, tal como se describe en [b UIT-T P.800]. Se ha incorporado una función de correspondencia similar en la etapa de salida del modelo UIT-T P.863.

Una de las ventajas del modelo objetivo es que los resultados son reproducibles y, por lo tanto, las mediciones realizadas en diferentes momentos y lugares pueden compararse directamente. Con todo, cierta prudencia se impone, ya que factores como la elección del material de prueba y cualquier procesamiento previo o posterior pueden introducir un sesgo en los resultados.

Por razones que parecerán ahora evidentes, diferentes modelos objetivos pueden producir diferentes valores previstos de MOS para las mismas condiciones. Por ejemplo, los modelos de la UIT-T P.862.1 y la UIT-T P.863 no producen exactamente los mismos valores previstos de MOS para la voz codificada a la que se refiere la Recomendación UIT-T G.729, aunque este códec está dentro del alcance de ambos modelos. Esto se debe en parte a que los dos modelos han sido entrenados y optimizados usando diferentes experimentos subjetivos. Por este motivo, al comparar las predicciones objetivas de MOS con los umbrales, por ejemplo, en el marco de la supervisión de un acuerdo de nivel de servicio o para activar una alarma, se deben elegir esos umbrales en función del modelo que produce la predicción.

12 Notificación de los valores MOS subjetivos

En el Cuadro 1 se describe la información que debe facilitarse al notificar los valores MOS subjetivos y la información adicional que se recomienda suministrar.

Si se ha realizado un experimento de conformidad con una Recomendación de la UIT, la información sobre la metodología puede incluir en general una simple referencia a la norma pertinente y al método específico utilizado, aunque es conveniente señalar, en su caso, las variaciones con respecto a los procedimientos normalizados.

Es importante facilitar en todos los casos información acerca de las muestras de prueba utilizadas para los experimentos pasivos. Si se trata de muestras de vídeo, puede ser útil suministrar información más detallada en la que se indique, por ejemplo, si determinadas secuencias contienen cambios de panorámica o de escena.

**Cuadro 1 – Información mínima para la notificación
de los valores MOS subjetivos**

Información	Tipo de experimento	Disposición
<p>Metodología</p> <ul style="list-style-type: none"> Pasiva o interactiva Evaluación por muestreo o continua Evaluación absoluta o relativa de las muestras Instrucciones y pregunta(s) presentada(s) a los sujetos Posiciones en la escala de evaluación Escala de evaluación discreta o continua Duración de la muestra <p>O</p> <ul style="list-style-type: none"> Recomendación UIT y método utilizado 	Todos	Obligatoria
<p>Plan de la prueba</p> <ul style="list-style-type: none"> Propósito del experimento Fecha y lugar de la prueba Procesamiento de la información Diseño del experimento, por ejemplo, diseño de bloqueo Número de sesiones y duración Número de sujetos Perfil de los sujetos, desglose por edad y sexo Tipo de sujetos entrevistados (por ejemplo, inexperto o experto) Información sobre el equipo utilizado Entorno de presentación, es decir, ruido de fondo, nivel de luz, etc. 	Todos	Recomendada
<p>Información sobre la condición/HRC</p> <ul style="list-style-type: none"> Número de condiciones Lista de condiciones Promedio de votos por condición (MOS) Desviación típica de los votos por condición Número de votos por condición 	Todos	Obligatoria
<p>Información sobre las subcondiciones</p> <ul style="list-style-type: none"> Lista de factores de la subcondición Valores MOS para las subcondiciones Número de votos y varianza por subcondición 	Todos	Facultativa
<p>Presentación de audio</p> <ul style="list-style-type: none"> Ancho de banda del audio Canales de audio, por ejemplo, monofónico, estereofónico, etc. Nivel de presentación de audio Método de presentación de audio, por ejemplo, altavoces, auriculares (monoaural, diótico binaural, etc.) 	Voz, audio, AV	Obligatoria
<p>Presentación vídeo Resolución de la imagen de vídeo (Notas 1 y 2)</p> <ul style="list-style-type: none"> Distancia de visualización en función de la altura, por ejemplo, 3H Tipo y tamaño del dispositivo, por ejemplo, televisor, tableta, teléfono, etc. Aplicación, es decir, videoteléfono, video a la carta, TV lineal, etc. 	Video, AV	Obligatoria

Cuadro 1 – Información mínima para la notificación de los valores MOS subjetivos

Información	Tipo de experimento	Disposición
Idioma	Pasivo, voz, AV	Obligatoria
Número y género de los hablantes	Pasivo, voz, AV	Obligatoria
Tipo de material vídeo, por ejemplo, deportes, cabeza y torso	Pasivo, voz, AV	Obligatoria
Tipo de audio, por ejemplo, música clásica, música popular, bandas sonoras de películas	Pasivo, voz, AV	Obligatoria
<p>NOTA 1 – Deberá indicarse el uso de imágenes entrelazadas.</p> <p>NOTA 2 – Si el experimento contiene diferentes resoluciones de imagen, debe indicarse si las resoluciones de imagen más pequeñas se presentan en su forma original o a escala superior.</p>		

13 Notificación de los valores MOS objetivos

Si se notifica un valor MOS obtenido por un modelo objetivo del UIT-T, por lo general bastará con comunicar el modelo utilizado y los ajustes no predeterminados. Si se trata de modelos no normalizados, se deberá indicar ese dato en la casilla "Metodología" del Cuadro 1 y se describirá el diseño experimental objeto de predicción. También se recomienda facilitar información sobre el tipo de material de prueba utilizado en los experimentos para ensayar y entrenar el modelo objetivo.

14 Notas

En [UIT-T P.800.1] se proporciona una notación genérica y de alto nivel que puede ayudar a identificar la fuente de un valor MOS. Si bien la notación de P.800.1 es útil para tener una visión general del contexto en el que se generó un valor MOS, es preferible contar con una descripción detallada del contexto con arreglo a esa Recomendación, que debe facilitarse siempre que sea posible.

Bibliografía

La UIT ha normalizado una serie de métodos de prueba subjetivos para diferentes aplicaciones. A continuación se enumeran algunos de los más utilizados.

[b-UIT-T G.729] Recomendación UIT-T G.729 (2012), *Codificación de la voz a 8 kbit/s mediante predicción lineal con excitación por código algebraico de estructura conjugada (CS-ACELP)*.

La serie UIT-T P.800 incluye numerosas Recomendaciones relativas a la evaluación subjetiva y objetiva de la calidad de voz, entre las que cabe destacar:

[b-UIT-T P.800] Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión*.

[b-UIT-T P.805] Recomendación UIT-T P.805 (2007), *Evaluación subjetiva de la calidad de la conversación*.

[b-UIT-T P.835] Recomendación UIT-T P.835 (2003), *Metodología de prueba subjetiva para evaluar los sistemas de comunicación vocal que utilizan un algoritmo de cancelación de ruido*.

[b-UIT-T P.862.1] Recomendación UIT-T P.862.1 (2001), *Función de correspondencia para convertir los resultados brutos de la prueba P.862 en nota media de opinión de la calidad de escucha objetiva*.

[b-UIT-T P.863] Recomendación UIT-T P.863 (2014), *Predicción de la calidad de escucha objetiva por percepción*.

La serie UIT-T P.900 incluye Recomendaciones relativas a la evaluación de multimedios:

[b-UIT-T P.910] Recomendación UIT-T P.910 (2008), *Métodos subjetivos de evaluación de la calidad vídeo para aplicaciones multimedios*.

[b-UIT-T P.911] Recomendación UIT-T P.911 (1998), *Métodos de evaluación subjetiva de la calidad audiovisual para aplicaciones multimedios*.

[b-UIT-T P.912] Recomendación UIT-T P.912 (2016), *Métodos de evaluación subjetiva de la calidad de vídeo para tareas de reconocimiento*.

El UIT-R también ha publicado Recomendaciones relativas a la evaluación subjetiva de la calidad del audio y del vídeo:

[b-UIT-R BS.1116-1] Recomendación UIT-R BS.1116-1 (1997), *Métodos para la evaluación subjetiva de pequeñas degradaciones en los sistemas de audio incluyendo los sistemas de sonido multicanal*.

[b-UIT-R BS.1534-1] Recomendación UIT-R BS.1534-1 (2003), *Método para la evaluación subjetiva del nivel de calidad intermedia de los sistemas de codificación*.

[b-UIT-R BT.500-13] Recomendación UIT-R BT.500-13 (2012), *Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión*.

[b-UIT-R BT.710-4] Recomendación UIT-R BT.710-4 (1998), *Métodos de evaluación subjetiva de la calidad de las imágenes en sistemas de televisión de alta definición*.

NOTA – No todas las metodologías normalizadas descritas por el UIT-R suelen medir las notas medias de opinión. Las Recomendaciones UIT-R publicadas ofrecen una documentación completa y referencias sobre todas las metodologías pertinentes. Para una descripción y explicación más detalladas de las metodologías de prueba mencionadas, se remite al lector a las diferentes Recomendaciones publicadas por el UIT-R.

En el manual que se menciona a continuación se examinan con más detalle los métodos de prueba subjetivos y las prácticas idóneas:

[b-UIT-T manual] *Procedimientos prácticos para pruebas subjetivas* (2011).

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie D	Principios de tarificación y contabilidad y cuestiones económicas y políticas de las telecomunicaciones/TIC internacionales
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 multimedia
Serie I	Red digital de servicios integrados
Serie J	Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedia
Serie K	Protección contra las interferencias
Serie L	Medio ambiente y TIC, cambio climático, ciberdesechos, eficiencia energética, construcción, instalación y protección de los cables y demás 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
Serie P	Calidad de la transmisión telefónica, instalaciones telefónicas y redes de líneas locales
Serie Q	Conmutación y señalización, y mediciones y pruebas asociadas
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, redes de próxima generación, Internet de las cosas y ciudades inteligentes
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación