



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

# UIT-T

SECTOR DE NORMALIZACIÓN  
DE LAS TELECOMUNICACIONES  
DE LA UIT

## Serie H

### Suplemento 1

(05/99)

SERIE H: SISTEMAS AUDIOVISUALES Y  
MULTIMEDIOS

---

**Perfil de aplicación – Utilización de la  
comunicación en vídeo a baja velocidad binaria  
para la conversación en tiempo real mediante el  
lenguaje de signos y la lectura labial**

Recomendaciones UIT-T de la serie H – Suplemento 1

(Anteriormente Recomendaciones del CCITT)

---

RECOMENDACIONES UIT-T DE LA SERIE H  
**SISTEMAS AUDIOVISUALES Y MULTIMEDIOS**

Características de los canales de transmisión para usos distintos de los telefónicos	H.10–H.19
Utilización de circuitos de tipo telefónico para telegrafía armónica	H.20–H.29
Utilización de circuitos o cables telefónicos para transmisiones telegráficas de diversos tipos o transmisiones simultáneas	H.30–H.39
Utilización de circuitos de tipo telefónico para telegrafía facsímil	H.40–H.49
Características de las señales de datos	H.50–H.99
CARACTERÍSTICAS DE LOS SISTEMAS VIDEOTELEFÓNICOS	H.100–H.199
INFRAESTRUCTURA DE LOS SERVICIOS AUDIOVISUALES	
Generalidades	H.200–H.219
Multiplexación y sincronización en transmisión	H.220–H.229
Aspectos de los sistemas	H.230–H.239
Procedimientos de comunicación	H.240–H.259
Codificación de imágenes vídeo en movimiento	H.260–H.279
Aspectos relacionados con los sistemas	H.280–H.299
Sistemas y equipos terminales para los servicios audiovisuales	H.300–H.399
Servicios suplementarios para multimedios	H.450–H.499

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

## **SUPLEMENTO 1 A LAS RECOMENDACIONES UIT-T DE LA SERIE H**

### **PERFIL DE APLICACIÓN – UTILIZACIÓN DE LA COMUNICACIÓN EN VÍDEO A BAJA VELOCIDAD BINARIA PARA LA CONVERSACIÓN EN TIEMPO REAL MEDIANTE EL LENGUAJE DE SIGNOS Y LA LECTURA LABIAL**

#### **Resumen**

El lenguaje de signos y la lectura labial son dos importantes terrenos de aplicación de la comunicación en vídeo. Para transmitir con éxito los componentes del lenguaje visual deben satisfacerse algunos requisitos de calidad.

El presente Suplemento es un perfil de aplicación de documento que sirve de base para los requisitos y suministra una orientación acerca de su cumplimiento. No tiene el propósito de presentar nuevos sistemas de codificación de vídeo sino indicar de qué manera podrían aplicarse con buenos resultados en dichos terrenos los sistemas de codificación de vídeo actuales y futuros.

#### **Orígenes**

El Suplemento 1 a las Recomendaciones de la serie H del UIT-T ha sido preparado por la Comisión de Estudio 16 (1997-2000) del UIT-T y fue aprobado por el procedimiento de la Resolución N.º 5 de la CMNT el 27 de mayo de 1999.

## 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 Conferencia Mundial de Normalización de las Telecomunicaciones (CMNT), 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 N.º 1 de la CMNT.

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 *empresa de explotación reconocida (EER)* designa a toda persona, compañía, empresa u organización gubernamental que explote un servicio de correspondencia pública. Los términos *Administración*, *EER* y *correspondencia pública* están definidos en la *Constitución de la UIT* (Ginebra, 1992).

## 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 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 1999

Es propiedad. Ninguna parte de esta publicación puede reproducirse o utilizarse, de ninguna forma o por ningún medio, sea éste electrónico o mecánico, de fotocopia o de microfilm, sin previa autorización escrita por parte de la UIT.

## ÍNDICE

### Página

1	Alcance .....	1
2	Abreviaturas .....	1
3	Definiciones .....	1
4	Referencias.....	1
5	Necesidades básicas para la reproducción del lenguaje de signos y la lectura labial .	2
5.1	Características básicas .....	2
5.2	Requisitos de resolución temporal.....	2
5.2.1	Alfabeto dactilológico .....	2
5.2.2	Lenguaje de signos general.....	3
5.2.3	Lectura labial .....	3
5.2.4	Adaptación.....	3
5.2.5	Análisis del requisito de velocidad de cuadros.....	3
5.2.6	Granularidad de la resolución temporal.....	6
5.3	Requisitos de resolución espacial .....	6
5.4	Fidelidad .....	7
5.5	Retardo.....	7
5.6	Sincronización .....	7
5.7	Conclusiones sobre los requisitos de calidad de funcionamiento.....	8
6	Comprobación de la calidad de funcionamiento.....	9
6.1	Material de referencia .....	9
6.2	Evaluaciones de la calidad de funcionamiento .....	9
7	Consejo a los diseñadores de terminales.....	9
8	Consejo al usuario.....	10
9	Ampliación del alcance.....	10
Apéndice I – Declaración de derechos de autor y descripción técnica del material de prueba del Suplemento 1 a las Recomendaciones de la serie H .....		10
I.1	Copyright .....	10
I.2	Servicios de apoyo .....	10
I.3	Detalles de la secuencia en vídeo.....	11

CD-ROM incluido: Videosecuencia "Irene"

## **Introducción**

Millones de sordos emplean el lenguaje de signos como su primer idioma y están ansiosos por poder utilizarlo en las conversaciones a larga distancia. La velocidad de conversación en el lenguaje de signos es comparable a la de la conversación vocal.

Las personas con diferente grado de insuficiencia auditiva pueden adquirir muy buenos indicios para "percibir" la conversación observando el rostro del que habla e interpretando el movimiento de sus labios.

El presente Suplemento describe la importancia de diferentes factores que deben tenerse en cuenta en la aplicación de una codificación de vídeo a baja velocidad binaria para su utilización adecuada en el lenguaje de signos y la lectura labial.

Los requisitos mencionados en este Suplemento se elaboraron experimentalmente con usuarios y no deberían considerarse valores fijos o absolutos. En diferentes situaciones, las exigencias pueden ser tanto mayores como menores.

## Suplemento 1 a las Recomendaciones de la serie H

### PERFIL DE APLICACIÓN – UTILIZACIÓN DE LA COMUNICACIÓN EN VÍDEO A BAJA VELOCIDAD BINARIA PARA LA CONVERSACIÓN EN TIEMPO REAL MEDIANTE EL LENGUAJE DE SIGNOS Y LA LECTURA LABIAL

(Ginebra, 1999)

#### 1 Alcance

El presente perfil de aplicación para el lenguaje de signos y la lectura labial suministra las características necesarias de un sistema de comunicación en vídeo para la conversación de persona a persona mediante el lenguaje de signos y la interpretación del movimiento de los labios, acompañado o no de lenguaje articulado audible.

Establece requisitos de calidad de funcionamiento que es importante satisfacer para asegurar que la conversación tenga éxito.

Describe cómo puede evaluarse la calidad de funcionamiento para el lenguaje de signos y la lectura labial.

Sugiere los factores que han de manejarse externamente respecto del protocolo de codificación de vídeo en el diseño del terminal, así como el entorno en el que se utilizan los terminales para el lenguaje de signos y la lectura labial.

Este Suplemento incluye la secuencia de prueba "Irene" para la evaluación de la comunicación en vídeo para el lenguaje de signos.

#### 2 Abreviaturas

En este Suplemento se utilizan las siguientes siglas.

CIF	Formato de intercambio común (352 × 288 píxels) ( <i>common interchange format</i> )
fps	Cuadros por segundo; imágenes por segundo ( <i>frames per second; pictures per second</i> )
QCIF	Cuarto de CIF (176 × 144 píxels) ( <i>quarter CIF</i> )
SQCIF	Sub QCIF (112 × 96 píxels)

#### 3 Definiciones

En este Suplemento se define el termino siguiente.

**3.1 cuadro:** Una imagen completa reproducida en vídeo se llama un "cuadro". En algunos sistemas, los cuadros están constituidos por dos medias imágenes, en que cada una contiene la mitad de la información de los cuadros. Estas medias imágenes se denominan campos.

#### 4 Referencias

- [1] HELLSTRÖM, DELEVERT, REVELIUS; Quality requirements on Videotelephony for Sign Language, *Swedish National Association of the Deaf*, 1997.
- [2] Recomendación UIT-T G.114 (1996), *Tiempo de transmisión en un sentido*.

- [3] FROWEIN; Improved speech reception through videotelephony, *IEEE journal on Selected Areas in Communication*, mayo de 1991.
- [4] Recomendación UIT-T P.931 (1998), Retardo de las comunicaciones multimedios, sincronización y medición de la velocidad de tramas.

## **5 Necesidades básicas para la reproducción del lenguaje de signos y la lectura labial**

### **5.1 Características básicas**

Los componentes expresivos del lenguaje de signos son movimientos y posiciones de las manos, los ojos, la boca, el rostro y el cuerpo.

En la interpretación del movimiento de los labios, los componentes son los movimientos del rostro. Frecuentemente la lectura de los labios se apoya mediante la voz. En otros casos, se combina con el lenguaje de signos. Hay también personas con formas pronunciadas de sordera que no usan el lenguaje de signos y dependen por completo en la conversación de la interpretación del movimiento de los labios.

A los efectos de la codificación de vídeo, se considera que la escena en que una persona habla con signos o articula palabras presenta una cantidad mediana de movimiento, que puede llegar a ser bastante grande.

### **5.2 Requisitos de resolución temporal**

Tanto el lenguaje de signos como el labial requieren una buena reproducción de los gestos. Suponiendo que un sistema reproduce los gestos mediante imágenes distribuidas uniformemente, han de observarse las siguientes condiciones:

- El lenguaje de signos y la lectura labial se consideran practicables a 20 cuadros por segundo (fps, *frames per second*) [1], [3].
- Dentro de ciertos límites, es posible utilizar una velocidad de cuadros de 12 fps o mayor [1].
- Para la lectura labial, un aumento constante de la capacidad de uso se observa al aumentar la velocidad de cuadros hasta 15. Después de 15 fps, el incremento prosigue pero de manera menos pronunciada [3].
- Se ha observado cierto grado de practicabilidad entre 8 y 12 fps, pero con una fuerte degradación perceptiva o de velocidad.
- Por debajo de 8 fps, el lenguaje labial o de signos no es practicable.

#### **5.2.1 Alfabeto dactilológico**

Los requisitos de resolución temporal para el lenguaje de signos pueden ilustrarse con el ejemplo del alfabeto dactilológico. El alfabeto dactilológico es una técnica por la cual cada letra del alfabeto corresponde a una posición particular de la mano. Las posiciones en el alfabeto dactilológico varían según los países. La comunicación tiene lugar presentando dichas posiciones en una rápida secuencia, para formar palabras. Las palabras así deletreadas son generalmente nombres propios y otras expresiones que los signos generales del lenguaje de signos no abarcan. El alfabeto dactilológico es muy rápido y puede llegar hasta 10 letras por segundo. Para una reproducción fiable, deberían reproducirse por lo menos dos imágenes por letra. Dicho de otro modo, puede concluirse que la reproducción legible del alfabeto dactilológico requiere por lo menos 20 cuadros por segundo.



### 5.2.2 Lenguaje de signos general

El alfabeto dactilológico es sólo una parte del lenguaje de signos. La mayor parte del lenguaje de signos está constituida por signos que representan conceptos completos, partes de frases, relaciones gramaticales y nombres comunes. Hay muchos lenguajes de signos en el mundo. Incluso siendo diferentes, los conceptos comunes son bastantes similares como para que las ideas de este Suplemento sean válidas para todos ellos. También en el curso del lenguaje de signos general tienen lugar rápidos movimientos de la mano y breves parpadeos de los ojos que transmiten información de carácter gramatical. En muchos casos, los requisitos de resolución temporal son similares a los del lenguaje dactilológico.

### 5.2.3 Lectura labial

Una cuantificación aproximada de los requisitos para la interpretación del movimiento de los labios puede calcularse a partir de la velocidad de fonemas en el lenguaje corriente. Una velocidad normal es 10 fonemas por segundo. Para poder percibirlos en el lenguaje visual, deberían reproducirse por lo menos a 20 imágenes por segundo de la secuencia.

### 5.2.4 Adaptación

Tanto en el caso del lenguaje labial como de signos, la velocidad de la producción de lenguaje puede disminuirse un poco a voluntad. Esto explica por qué es posible a veces utilizar una velocidad de 12 a 15 cuadros por segundo. Los que están habituados a la lectura labial y al lenguaje de signos consiguen captar el sentido basándose en sus experiencias anteriores y la repetición de los gestos. Esto hace posible que algunos usuarios puedan efectuar breves conversaciones por conexiones de menor calidad que los requisitos antes mencionados.

### 5.2.5 Análisis del requisito de velocidad de cuadros

Un análisis de la secuencia de prueba "Irene" explica más en detalle las necesidades.

#### Lenguaje dactilológico

El cuadro 1 muestra una representación aproximada de una secuencia en alfabeto dactilológico en la secuencia de prueba "Irene". Las imágenes correspondientes a esta secuencia se reproducen en la figura 1.

**Cuadro 1 – Ejemplo de representación dactilológica en cuadros a 25 y 12,5 cuadros por segundo**

cuadro N.º	308		310					315					320					325					330					335	336
25 fps	e	e	e	–	d	s	s	s	s	–	v	v	v	–	i	–	–	k	k	k	–	e	n	n	n	n	n	n	n
12,5 fps		e		–		s		s		–		v		–		–		k		k		e		n		n		n	

Las cifras de la primera fila corresponden a los números de cuadro desde el comienzo de la secuencia. Las letras indican las situaciones en que las letras se expresan claramente mediante gestos de la mano. El guión indica que no se ha formado ninguna letra clara, como transición entre letras. La palabra en cuestión es "Edsviken", nombre de una localidad.

De estas ocho letras, tres pueden percibirse con claridad únicamente en un cuadro, por lo que podría perderse en una transmisión a 12,5 cuadros por segundo. Esta velocidad de cuadros se produce cuando cada segundo cuadro se salta en el sistema de codificación. Un ejemplo de muestreo a 12,5 fps aparece en la fila inferior del cuadro. Puede verse que, de la palabra original "Edsviken" sólo queda "Esvken". Esto demuestra claramente el riesgo de pérdida del contenido del lenguaje cuando la velocidad de cuadros es inferior a 20 fps.

La distribución de letras en la secuencia a 25 fps es:

1 cuadro    3 letras;

2 cuadros   0 letra;

3 cuadros   3 letras;

4 cuadros   1 letra;

7 cuadros   1 letra (fin de la frase).

Longitud media dentro de las frases: 2,3 cuadros por letra.

### **Conclusión**

En este ejemplo, las letras que forman las palabras varían entre 1 y 4 cuadros en el tiempo, en que cada cuadro representa 40 ms. La longitud media representa una visibilidad de 2,3 cuadros por letra. El ejemplo es demasiado corto para llegar a conclusiones estadísticas reales. Sin embargo, puede observarse que con esta velocidad de lenguaje dactilológico parece suficiente una velocidad de cuadros de 25 fps, mientras que la velocidad de 12,5 fps requiere cierto esfuerzo de asociación para comprender las palabras expresadas en alfabeto dactilológico.



**Figura 1 – Cuadros que contienen la palabra dactilológica "Edsviken" registrada a 25 cuadros por segundo**

## **Lenguaje de signos general**

Buena parte de la película "Irene" se expresa mediante signos generales sin recurrir al alfabeto dactilológico.

Se ha realizado un simple análisis de una frase. La frase en cuestión se presenta a continuación. Se transcribe signo por signo indicando con un número entre paréntesis la cantidad de cuadros que ocupa cada signo.

La secuencia aparece entre los cuadros 406 y 529 de "Irene".

"SHE(7) TELLS(7) SELF(11) HOW(4) SHE(2) FELT(11) EXPERIENCED(13) ADOLESCENCE(16)".

Ningún signo de esta secuencia ocupó menos de 2 cuadros ni representó gestos más rápidos que los signos del alfabeto dactilológico. Hay signos que se expresan con gestos más largos y, por lo tanto, presentan otras exigencias para la codificación del vídeo.

## **Marcas gramaticales**

Hay un parpadeo de carácter gramatical en los cuadros 394 y 395, que indica una marca de fraseo.

Dicha marca es de una longitud suficiente para un muestreo a 12,5 fps.

### **5.2.6 Granularidad de la resolución temporal**

En la mayoría de los casos la cámara de vídeo utilizada para la comunicación se ajusta a las normas de vídeo generales. Es decir, entrega 25 ó 30 cuadros por segundo. Esto añade granularidad a las velocidades de cuadro útiles. Cuando se emplean tales cámaras, no tiene mucho sentido referirse a velocidades de cuadros entre 12,5 y 25 fps o entre 15 y 30 fps. Estas velocidades de cuadros intermedias significan que los intervalos de imagen de origen variarán entre 40 y 80 ms o entre 33 y 66 ms, respectivamente, lo cual comporta ciertos riesgos de pérdida de algunos detalles del gesto. En conclusión, para satisfacer el requisito de 20 cuadros por segundo mediante cámaras comunes, la velocidad de cuadros en destino debería ser de 25 ó 30 cuadros por segundo.

## **5.3 Requisitos de resolución espacial**

En lo que respecta a la resolución espacial, se indica que se han encontrado las siguientes necesidades para las comunicaciones de persona a persona por el lenguaje de signos [1]:

- Es posible emplear la resolución QCIF, pero se pierden algunos detalles menores, como los movimientos de los ojos. Esto aumenta el desconcierto del escucha.
- La resolución CIF es adecuada. La elevación de QCIF a CIF proporciona una mejor percepción del lenguaje.
- La resolución SQCIF es demasiado aproximada para una percepción fiable, aunque en algunos casos pueden captarse algunos signos.
- De emplearse resoluciones diversas en distintas partes de la imagen, serán las manos y el rostro los elementos que requerirán la más alta resolución. En los sistemas así diseñados, debe cuidarse de no introducir distorsiones en las demás partes de la imagen que puedan distraer al usuario.

Puede realizarse una sencilla comprobación teórica. En la figura de medio cuerpo que suele utilizarse en la conversación de persona a persona mediante signos, el tamaño de un dedo corresponde aproximadamente a 1/50 del ancho de la imagen. Para una resolución fiable de los dedos en una imagen, cada dedo debería ocupar por lo menos 3 píxels. Esto fija el requisito mínimo de resolución espacial en QCIF, que contiene 176 píxels de ancho. También la dirección de la mirada es importante en el lenguaje visual y requiere una mayor resolución. Por eso es mejor la opción CIF y debe preferirse.

Para la lectura labial, la figura en las llamadas de persona a persona puede reducirse a un plano algo mayor que la cabeza. En este caso, se indica que QCIF sería una resolución suficiente para la interpretación del movimiento de los labios [3]. Cuando se emplea una resolución QCIF, el usuario del terminal debe asegurarse de que la pantalla se observe a una distancia conveniente, para que la resolución relativamente más baja no produzca demasiadas molestias a la percepción.

#### **5.4 Fidelidad**

En la comunicación por vídeo suelen producirse borrosidades en el movimiento.

Los modelos que describen las borrosidades no están todavía bien desarrollados. Existen grandes variaciones en cuanto al tipo de borrosidad y su efecto para la percepción. Por tal motivo, nos limitaremos aquí a efectuar una breve comparación de formas admisibles de borrosidad en distintas situaciones.

Se indica al vídeo VHS como suficiente para una buena percepción del lenguaje de signos y la lectura labial. En las grabaciones en vídeo, los objetos en rápido movimiento suelen aparecer de manera considerablemente borrosa, ya que las velocidades del obturador se sitúan normalmente entre 1/50 y 1/60 de segundo. Esto indica que la borrosidad es aceptable en objetos que se mueven rápidamente como parte de grandes movimientos de gesticulación.

En el curso de movimientos prolongados puede producirse ocasionalmente borrosidad. La resolución espacial en esos momentos no debería ser nunca inferior a la correspondiente a SQCIF. Esta conclusión se deduce del hecho que SQCIF constituye una resolución demasiado aproximada para obtener una percepción fiable.

A los efectos de una buena percepción, cuando la resolución espacial básica es CIF, la borrosidad ocasional introducida no debería ser mayor que la que se percibe en una resolución QCIF.

#### **5.5 Retardo**

El retardo entre la cámara de origen y la pantalla de recepción es un factor decisivo en la aplicación de conversación. Son preferibles valores por debajo de 0,4 s, aumentando la preferencia al descender a 0,1 s [3].

Los valores por encima de 0,8 s se consideran inconvenientes para una buena conversación mediante el lenguaje de signos [1].

Los requisitos para el lenguaje de signos y la lectura labial parecen ser similares a los de la conversación vocal. El tiempo entre una expresión y la visualización o audición de la reacción esperada es por lo menos el doble del retardo. Incluso el límite de 0,4 s especificado por la Recomendación G.114 [2] parece muy largo, puesto que evidentemente involucra un retardo de la respuesta en 0,8 s.

#### **5.6 Sincronización**

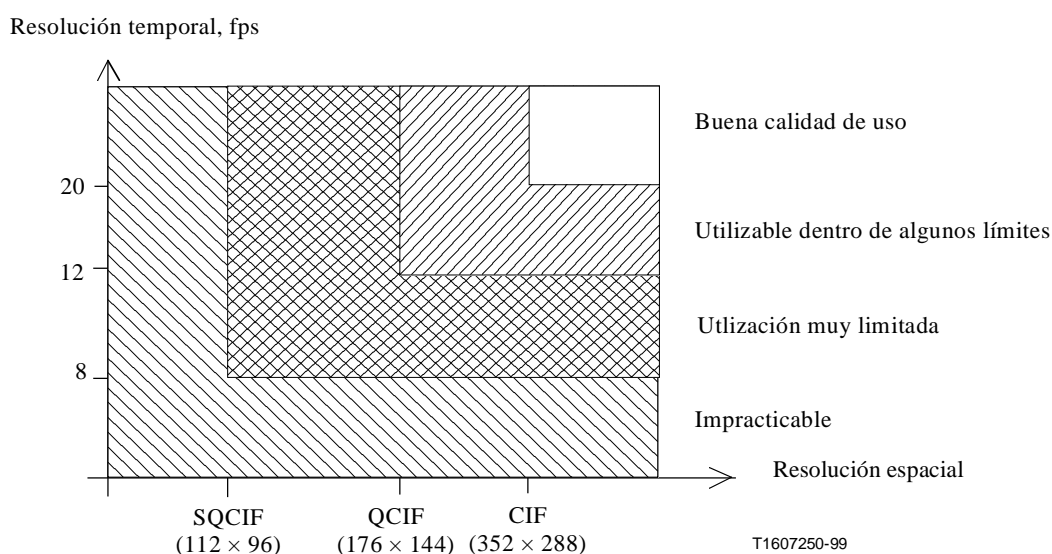
Para un lenguaje labial apoyado con la voz, la sincronización entre sonido y vídeo es esencial. Se consideran aceptables diferencias de tiempo de hasta 100 ms [1].

Para las personas que pueden emplear tanto la voz como la lectura labial, esta combinación puede ser muy eficaz para la percepción [3].

## 5.7 Conclusiones sobre los requisitos de calidad de funcionamiento

Para la aplicación de transmisiones de conversaciones persona a persona mediante el lenguaje de signos y la lectura labial, valen las siguientes metas básicas de calidad de funcionamiento:

- Un objetivo de 25 a 30 cuadros por segundo en resolución CIF y un retardo máximo de 0,4 s, con una borrosidad ocasional admitida inferior a la correspondiente a la resolución QCIF durante movimientos de mediana intensidad.
- Cuando es necesario en los entornos de velocidad binaria muy baja, aceptación de 12-15 fps en QCIF en movimiento de mediana intensidad y una degradación ocasional correspondiente a SQCIF durante gestos prolongados del lenguaje de signos.
- El mantenimiento de una sincronización sonora mejor que 100 ms.
- El retardo extremo a extremo debe ser inferior a 0,4 s. Si no se puede evitar, se aceptará un retardo de hasta 0,8 s.



NOTA – Han de observarse los valores en presencia de movimientos de lenguaje de signos o lenguaje labial.

**Figura 2 – Requisitos de resolución para la conversación persona a persona mediante el lenguaje de signos y la lectura labial**

**Cuadro 2 – Resumen de la degradación de utilidad por efecto del retardo y la borrosidad**

Calidad de uso	Retardo extremo a extremo	Borrosidad ocasional durante gestos prolongados	
		Para resolución CIF	Para resolución QCIF
Buena	<0,4 s	No	–
Utilizable dentro de algunos límites	0,4 – 0,8 s	Degradación a $\cong$ QCIF	No
Utilización limitada	0,8 – 1,2 s	Degradación a $\cong$ SQCIF	Degradación a $\cong$ SQCIF
Impracticable	>1,2 s	Degradación a < SQCIF	Degradación a < SQCIF

## **6 Comprobación de la calidad de funcionamiento**

### **6.1 Material de referencia**

En este Suplemento figura un CD-ROM que contiene una presentación en vídeo de lenguaje de signos para la evaluación de la calidad de funcionamiento. La videosecuencia "Irene", extraída de un programa de la TV sueca, contiene suficientes ejemplos de gestos en el lenguaje de signos. Muestra también la rapidez normal de los gestos.

Swedish Educational Broadcasting Company tiene la propiedad intelectual de esta videosecuencia. En el apéndice I se reproduce el fichero Readme del CD-ROM, que contiene la declaración de derechos de autor y la descripción técnica de los ficheros electrónicos.

### **6.2 Evaluaciones de la calidad de funcionamiento**

Se prueba un códec o una instalación terminal transmitiendo la escena de evaluación a través de un códec o aparatos de videofonía conectados a la red. El resultado se registra y evalúa. La Recomendación P.931 [4] especifica una evaluación del método.

Se evalúa la velocidad de cuadros durante el lenguaje de signos.

Se registra la resolución estática seleccionada.

Toda borrosidad adicional que aparezca durante un movimiento de intensidad mediana se mide comparando los cuadros registrados con las imágenes de la misma escena con resolución reducida a QCIF y SQCIF. Sólo se evalúa la borrosidad en las manos y el rostro.

Se mide el retardo.

Se mide la sincronización de audio (voz) y vídeo (movimientos de los labios).

A partir de estos registros puede evaluarse la calidad de funcionamiento y compararse con las metas descritas anteriormente.

Para la evaluación aproximada de estos valores en caso de no disponerse de equipos de laboratorio, puede utilizarse un sencillo instrumento de evaluación ideado por la Asociación Nacional Sueca de Sordos.

## **7 Consejo a los diseñadores de terminales**

Con el fin de satisfacer las necesidades de los usuarios, el terminal debería ajustarse a algunas características.

- Debería suministrar una interfaz para la activación de sistemas externos de alerta, por ejemplo, luces intermitentes, vibrador de bolsillo, vibrador pulsera o fuertes generadores de sonido.
- Los usuarios pueden tener necesidad a veces de volver a la conversación mediante un texto. Por tal motivo, es aconsejable aplicar en el terminal el protocolo T.140 de conversación por texto.
- La preferencia por una velocidad superior a 20 fps y un retardo inferior a 0,4 s pide la utilización de un algoritmo sin saltos de cuadros. Una mayor velocidad de cuadros proporciona automáticamente la posibilidad de lograr un retardo razonable.
- Pueden admitirse excepciones de todos los requisitos de calidad durante 2 s después de un cambio de escena.

## 8 Consejo al usuario

El usuario debería acondicionar un entorno con buenas condiciones de iluminación y un fondo liso detrás suyo.

## 9 Ampliación del alcance

En caso de utilizarse el equipo para aplicaciones de lenguaje de signos o lectura labial en videoconferencia, multidifusión, radiodifusión o recuperación de informaciones, las circunstancias siguientes modifican los requisitos.

- La escena suele ser más amplia, comprendiendo tanto a las personas que gesticulan como a otros objetos. Esto indica una capacidad de uso a partir de la resolución espacial CIF.
- El usuario tiene menos posibilidades de devolver informaciones a los efectos de controlar la percepción ejerciendo su influencia en la persona que habla o gesticula. De ahí que se requiera una velocidad de cuadros igual o superior a 20 fps.
- Los requisitos en materia de retardo son menos estrictos. En el caso de radiodifusión o recuperación de informaciones se admiten retardos de varios segundos. En el caso de conferencias, los requisitos de retardo son similares a los de la utilización del sistema para conversación.
- Los requisitos exactos de cada aplicación están fuera del alcance de este perfil de aplicación.

## APÉNDICE I

### **Declaración de derechos de autor y descripción técnica del material de prueba del Suplemento 1 a las Recomendaciones de la serie H**

*En este apéndice se reproduce el contenido del fichero Readme.txt del CD-ROM.*

Suplemento 1 a las Recomendaciones UIT-T de la serie H (06/1999).

**Videosecuencia "Irene"**

Versión 1.0, junio de 1999.

### **I.1 Copyright**

Todos los derechos reservados. El material sólo puede utilizarse para investigación y desarrollo de productos destinados a personas sordas. El material no puede incluirse en productos comercializables sin autorización de Swedish Educational Broadcasting Company. Se prohíbe toda otra utilización del material.

### **I.2 Servicios de apoyo**

Para la distribución de programas informáticos actualizados, diríjase a:

Departamento de Ventas de la UIT  
Place des Nations  
CH-1211, Ginebra 20  
SUIZA  
Correo electrónico: sales@itu.int



Para resolver problemas técnicos, comuníquese con el servicio Helpdesk de la TSB:

Servicio Helpdesk de la TSB  
UIT  
Place des Nations  
CH-1211, Ginebra 20  
SUIZA  
fax: +41 22 730 5853  
Correo electrónico: tsbedh@itu.int

### **I.3 Detalles de la secuencia en vídeo**

La secuencia en vídeo presenta elementos del lenguaje de signos destinados a utilizarse como material de prueba para la codificación de vídeo. Contiene expresiones del lenguaje de signos ejecutadas a una velocidad natural.

La secuencia se denomina "Irene" por el nombre de la persona que actúa. Se expresa en lenguaje de signos sueco y fue producida originariamente por la Swedish Educational Broadcasting Company. Presenta la figura de medio cuerpo que suele utilizarse en la comunicación personal mediante signos a través de videófonos. Está grabada en PAL a 25 fps. Existe en tres formatos:

- 1) Sign\_Irene.mpeg (3261 Kbytes)  
MPEG-1 codificado para resolución CIF a 25 fps.
- 2) Sign\_Irene.cif (80 190 Kbytes)  
Formato YCbCr 4:2:0 para CIF a 25 fps.
- 3) Sign\_Irene.qcif (20 048 Kbytes)  
Formato YCbCr 4:2:0 para QCIF a 25 fps.

### **Contenido del lenguaje de sordomudos (alfabeto dactilológico)**

Ésta es una representación aproximada de dos secuencias del lenguaje de sordomudos en "Irene". Los números son números de cuadros desde el inicio de la versión MPEG. Las letras indican cuándo se forman claramente con la mano. Un guión indica que no se ha formado ninguna letra con suficiente claridad en la transición entre letras. La primera representación es "Pia Wickman", en que la última "a" sólo puede visualizarse en los labios.

"Pia Wickman"									
Cuadro	Letra	Cuadro	Letra	Cuadro	Letra	Cuadro	Letra	Cuadro	Letra
29	p	39	-	49	-	59	-	69	-
30	p	40	a	50	-	60	-	70	n
31	p	41	a	51	w	61	k	71	n
32	p	42	a	52	w	62	k	72	n
33	-	43	a	53	-	63	k	73	n
34	-	44	a	54	i	64	-	74	n
35	i	45	a	55	-	65	-	75	n
36	i	46	a	56	c	66	-	76	n
37	i	47	-	57	c	67	-	77	n
38	i	48	-	58	c	68	m		

" Edsviken "									
Cuadro	Letra	Cuadro	Letra	Cuadro	Letra	Cuadro	Letra	Cuadro	Letra
308	e	315	s	322	i	329	e	336	n
309	e	316	s	323	-	330	n		
310	-	317	-	324	-	331	n		
311	-	318	v	325	k	332	n		
312	d	319	v	326	k	333	n		
313	s	320	v	327	k	334	n		
314	s	321	-	328	-	335	n		

### Contenido general de los signos

La última frase de la videosecuencia está señalada con signos (sin recurrir al alfabeto dactilológico), comparable a palabras.

La frase en cuestión se transcribe signo por signo indicando con un número entre paréntesis la cantidad de cuadros que ocupa cada signo "SHE(7) TELLS(7) HERSELF(11) HOW(4) SHE(2) FELT(11) EXPERIENCED(13) ADOLESCENCE(16)".

La secuencia aparece entre los cuadros 406 y 529 en la versión MPEG.

### Componentes gramaticales

La videosecuencia contiene un cierto número de parpadeos, que son componentes gramaticales típicos del lenguaje de signos utilizados como delimitadores de oraciones. Son breves y, en muchos casos, sólo ocurren en una o dos tramas. Hay un parpadeo gramatical en los cuadros 394 y 395.

## **SERIES DE RECOMENDACIONES DEL UIT-T**

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión: definiciones, símbolos, clasificación
Serie C	Estadísticas generales de telecomunicaciones
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
<b>Serie H</b>	<b>Sistemas audiovisuales y multimedios</b>
Serie I	Red digital de servicios integrados
Serie J	Transmisiones de señales radiofónicas, de televisión 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	RGT y mantenimiento de redes: sistemas de transmisión, circuitos telefónicos, telegrafía, facsímil y circuitos arrendados internacionales
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica, instalaciones telefónicas y redes locales
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 y comunicación entre sistemas abiertos
Serie Y	Infraestructura mundial de la información y aspectos protocolo Internet
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación