



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

Série H

Supplément 1

(05/99)

SÉRIE H: SYSTÈMES AUDIOVISUELS ET
MULTIMÉDIAS

**Profil d'application – Utilisation des
vidéocommunications à faible débit pour les
conversations en temps réel par langage signé
et lecture labiale**

Recommandations UIT-T de la série H – Supplément 1

(Antérieurement Recommandations du CCITT)

RECOMMANDATIONS UIT-T DE LA SÉRIE H
SYSTÈMES AUDIOVISUELS ET MULTIMÉDIAS

Caractéristiques des canaux de transmission pour des usages autres que téléphoniques	H.10–H.19
Emploi de circuits de type téléphonique pour la télégraphie à fréquence vocale	H.20–H.29
Circuits et câbles téléphoniques utilisés pour les divers types de transmission télégraphique et de transmissions simultanées	H.30–H.39
Circuits de type téléphonique utilisés en béliographie	H.40–H.49
Caractéristiques des signaux de données	H.50–H.99
CARACTÉRISTIQUES DES SYSTÈMES VISIOPHONIQUES	H.100–H.199
INFRASTRUCTURE DES SERVICES AUDIOVISUELS	
Généralités	H.200–H.219
Multiplexage et synchronisation en transmission	H.220–H.229
Aspects système	H.230–H.239
Procédures de communication	H.240–H.259
Codage des images vidéo animées	H.260–H.279
Aspects liés aux systèmes	H.280–H.299
Systèmes et équipements terminaux pour les services audiovisuels	H.300–H.399
Services complémentaires en multimédia	H.450–H.499

Pour plus de détails, voir la Liste des Recommandations de l'UIT-T.

SUPPLÉMENT 1 AUX RECOMMANDATIONS UIT-T DE LA SÉRIE H

PROFIL D'APPLICATION – UTILISATION DES VIDEOCOMMUNICATIONS À FAIBLE DÉBIT POUR LES CONVERSATIONS EN TEMPS RÉEL PAR LANGAGE SIGNE ET LECTURE LABIALE

Résumé

Le langage signé et la lecture labiale sont deux domaines d'application importants de la communication vidéo, mais il est nécessaire de satisfaire à certaines exigences de qualité pour assurer la transmission correcte des éléments du langage "visuel".

Le présent supplément contient un profil d'application qui établit les éléments de base des prescriptions et fournit les orientations sur la manière d'y répondre. Le présent supplément ne vise pas à proposer de nouveaux systèmes de codage vidéo, mais plutôt à montrer comment les systèmes actuels et futurs peuvent donner de bons résultats dans le domaine en question.

Source

Le Supplément 1 aux Recommandations UIT-T de la série H, élaboré par la Commission d'études 16 (1997-2000) de l'UIT-T, a été approuvé le 27 mai 1999 selon la procédure définie dans la Résolution n° 5 de la CMNT.

AVANT-PROPOS

L'UIT (Union internationale des télécommunications) est une institution spécialisée des Nations Unies dans le domaine des télécommunications. L'UIT-T (Secteur de la normalisation des télécommunications) est un organe permanent de l'UIT. Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

La Conférence mondiale de normalisation des télécommunications (CMNT), qui se réunit tous les quatre ans, détermine les thèmes d'études à traiter par les Commissions d'études de l'UIT-T, lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution n° 1 de la CMNT.

Dans certains secteurs des technologies de l'information qui correspondent à la sphère de compétence de l'UIT-T, les normes nécessaires se préparent en collaboration avec l'ISO et la CEI.

NOTE

Dans la présente publication, le terme *exploitation reconnue (ER)* désigne tout particulier, toute entreprise, toute société ou tout organisme public qui exploite un service de correspondance publique. Les termes *Administration*, *ER* et *correspondance publique* sont définis dans la *Constitution de l'UIT (Genève, 1992)*.

DROITS DE PROPRIÉTÉ INTELLECTUELLE

L'UIT attire l'attention sur la possibilité que l'application ou la mise en œuvre de la présente publication puisse donner lieu à l'utilisation d'un droit de propriété intellectuelle. L'UIT ne prend pas position en ce qui concerne l'existence, la validité ou l'applicabilité des droits de propriété intellectuelle, qu'ils soient revendiqués par un Membre de l'UIT ou par une tierce partie étrangère à la procédure d'élaboration des publications.

A la date d'approbation de la présente publication, l'UIT avait été avisée de l'existence d'une propriété intellectuelle protégée par des brevets à acquérir pour mettre en œuvre la présente publication. Toutefois, comme il ne s'agit peut-être pas de renseignements les plus récents, il est vivement recommandé aux responsables de la mise en œuvre de consulter la base de données des brevets du TSB.

© UIT 1999

Droits de reproduction réservés. Aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'UIT.

TABLE DES MATIÈRES

	Page
1	Domaine d'application..... 1
2	Abréviations 1
3	Définitions..... 1
4	Références 1
5	Besoins fondamentaux: reproduction du langage signé et lecture labiale..... 2
5.1	Caractéristiques de base 2
5.2	Besoins de résolution temporelle 2
5.2.1	La dactylogogie..... 2
5.2.2	L'expression signée générale..... 2
5.2.3	Lecture labiale..... 3
5.2.4	Adaptation..... 3
5.2.5	Analyse du besoin de fréquence d'image..... 3
5.2.6	Granularité de la résolution temporelle..... 5
5.3	Besoins de résolution spatiale 5
5.4	Fidélité..... 6
5.5	Temps de transmission 6
5.6	Synchronisme 6
5.7	Conclusions relatives au niveau de qualité 6
6	Vérification du niveau de performances 7
6.1	Matériel de référence..... 7
6.2	Evaluation du niveau de qualité 8
7	Conseils pour l'implémentation du terminal 8
8	Conseil pour l'utilisateur 8
9	Elargissement du domaine 8
Appendice I – Déclaration de copyright et description technique s'appliquant à la	
	séquence vidéo "Irene" 9
I.1	Copyright..... 9
I.2	Support 9
I.3	Détails sur la séquence vidéo 9

CD-ROM inclus: vidéoclip "Irene"

Introduction

Des millions de sourds utilisent le langage signé comme première langue et désirent ardemment utiliser ce mode d'expression dans la conversation à distance. Le débit de conversation du langage signé est sensiblement égal à celui de la conversation verbale.

L'observation du visage de l'interlocuteur pour la lecture labiale facilite considérablement la compréhension par les malentendants, indépendamment de leur degré de perte d'acuité auditive.

Le présent supplément décrit l'importance des divers facteurs intervenant dans l'application du codage vidéo à faible débit afin de rendre son utilisation satisfaisante pour le langage signé et la lecture labiale.

Les prescriptions décrites dans le présent supplément ont été établies sur la base de l'expérience avec les utilisateurs, mais elles ne doivent pas être considérées comme absolues et invariables. Selon les situations, elles peuvent être assouplies ou rendues plus strictes.

Supplément 1 aux Recommandations de la série H

PROFIL D'APPLICATION – UTILISATION DES VIDEOCOMMUNICATIONS À FAIBLE DÉBIT POUR LES CONVERSATIONS EN TEMPS RÉEL PAR LANGAGE SIGNE ET LECTURE LABIALE

(Genève, 1999)

1 Domaine d'application

Le présent profil d'application pour le langage signé et la lecture labiale donne les caractéristiques dont un système de communication vidéo doit disposer pour les besoins de la conversation de deux personnes en langage signé avec lecture labiale, avec ou sans parole audible.

Il établit des prescriptions de qualité de fonctionnement qu'il est important de respecter pour le bon déroulement de la conversation.

Il décrit la manière d'évaluer le niveau d'efficacité du langage signé et de la lecture labiale.

Il propose des facteurs extérieurs au protocole de codage vidéo, se rapportant à la conception des terminaux et à l'environnement d'utilisation du langage signé et de la lecture labiale.

Le présent supplément comporte la séquence de test "Irene" pour l'évaluation des vidéocommunications pour le langage signé.

2 Abréviations

Le présent supplément utilise les abréviations suivantes:

CIF	format intermédiaire commun (352×288 pixels) (<i>common interchange format</i>)
fps	images par seconde (<i>frames per second</i>)
QCIF	quart de CIF (176×144 pixels)
SQCIF	sous-quart de CIF (112×96 pixels)

3 Définitions

Le présent supplément définit le terme suivant:

3.1 image: image complète en reproduction vidéo. Dans certains systèmes, elle est constituée de deux demi-images contenant chacune la moitié des informations de l'image complète. Ces demi-images sont appelées trames.

4 Références

- [1] HELLSTRÖM, DELEVERT, REVELIUS: Quality requirements on Videotelephony for Sign Language, *Swedish National Association of the Deaf*, 1997 (Prescriptions de qualité visiophonique pour le langage signé).
- [2] Recommandation UIT-T G.114 (1996), *Temps de transmission dans un sens*.
- [3] FROWEIN: Improved speech reception through videotelephony, *IEEE journal on Selected Areas in Communication*, mai 1991 (Réception améliorée de la parole par la visiophonie).
- [4] Recommandation UIT-T P.931 (1998), *Mesure du temps de transmission, de la synchronisation et du débit de trames dans les communications multimédias*.

5 Besoins fondamentaux: reproduction du langage signé et lecture labiale

5.1 Caractéristiques de base

Dans le langage signé, les éléments porteurs de signifiants sont les mouvements et les positions des mains, des yeux, de la bouche, du visage et du corps.

En lecture labiale, ce sont les mouvements du visage. Dans beaucoup de cas, la lecture labiale est assistée par la voix. Dans d'autres, elle est utilisée en combinaison avec le langage signé. Il existe également des sourds profonds qui ne signent pas et qui communiquent exclusivement par lecture labiale.

En termes de codage vidéo, une scène avec un interlocuteur s'exprimant en langage signé est considérée comme étant d'une animation moyenne occasionnellement élevée.

5.2 Besoins de résolution temporelle

Une bonne reproduction des mouvements est nécessaire, tant pour le langage signé que pour la lecture labiale. Partant du principe qu'un système reproduit les mouvements avec une répartition uniforme des images, on a observé:

- que l'utilisation du système pour le langage gestuel et la lecture labiale est bonne à 20 images par seconde (fps) [1], [3];
- que moyennant quelques restrictions, on peut utiliser des fréquences d'image à partir de 12 images par seconde (fps) [1];
- qu'en lecture labiale, l'augmentation de la fréquence d'image jusqu'à 15 fps se traduit par une augmentation rapide de la qualité; au-delà, l'amélioration se poursuit mais de façon moins marquée [3];
- qu'entre 8 et 12 fps, la possibilité d'utilisation est très limitée avec une forte dégradation de la perception ou de la vitesse;
- que les cadences inférieures à 8 fps ne présentent aucune utilité pratique pour la lecture labiale et le langage signé.

5.2.1 La dactylogogie

On peut déterminer les besoins en résolution temporelle du langage signé à partir de celui de la dactylogogie, technique dans laquelle chaque lettre de l'alphabet est représentée par une position déterminée de la main (variable d'un pays à l'autre). La dactylogogie consiste à montrer les positions en succession rapide de manière à former des mots. Les mots épelés sont généralement des noms de personnes et d'autres noms propres qui ne sont pas couverts par les signes du langage signé. La dactylogogie est très rapide, atteignant souvent 10 lettres à la seconde. Pour obtenir une reproduction fidèle, il faut au moins deux images par lettre. On peut donc en conclure que la bonne reproduction de la dactylogogie nécessite au minimum 20 images par seconde.

5.2.2 L'expression signée générale

La dactylogogie n'est qu'une partie du langage signé, qui est constitué de signes représentant des notions globales, des parties de phrases, des éléments de grammaire ou des noms communs. Il y a beaucoup de langages signés, mais en dépit de leurs différences, leurs concepts communs sont suffisamment proches pour que le raisonnement proposé dans le présent supplément soit applicable à tous. Par ailleurs, l'expression signée générale est accompagnée de mouvements rapides de la main ainsi que de battements rapides des paupières donnant des informations d'ordre grammatical. Souvent, les besoins de résolution temporelle sont les mêmes que ceux de la dactylogogie.

5.2.3 Lecture labiale

On peut calculer approximativement le besoin de résolution de la lecture labiale à partir du débit de phonèmes du discours parlé, soit 10 phonèmes par seconde. Pour pouvoir lire les phonèmes visibles, il faut les reproduire à 20 fps au moins.

5.2.4 Adaptation

Dans les deux cas – lecture labiale et langage gestuel – on peut diminuer légèrement le débit, ce qui explique que des cadences de 12-15 fps sont parfois possibles. Comme les utilisateurs confirmés de ces deux techniques tablent beaucoup sur leur expérience et sur les redondances, ils peuvent établir des conversations brèves sur des connexions dont la qualité est inférieure à celle des besoins signalés plus haut.

5.2.5 Analyse du besoin de fréquence d'image

On trouvera des précisions à ce sujet dans le paragraphe relatif à la séquence d'essai "Irene".

Dactylogogie

Le Tableau 1 donne une représentation approximative de la séquence de dactylogogie de la séquence d'essai "Irene". Les images correspondantes sont montrées dans la Figure 1.

Tableau 1 – Exemple de représentation de dactylogogie à 25 et à 12,5 fps

n° d'image	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	

Les nombres de la rangée supérieure sont les numéros d'image comptés depuis le début de la séquence. Dans les deux rangées inférieures, une lettre indique que la dactylogogie est claire, un tiret qu'aucune lettre n'a été clairement formée au cours de la transition entre deux épellations. Le mot représenté est "Edsviken", un nom de lieu.

De ces huit lettres, trois seulement n'apparaissent clairement que sur une seule image. Il y a donc risque de perte de lettres à 12,5 fps, fréquence d'image obtenue par suppression d'une image sur deux. Un exemple d'échantillonnage à 12,5 fps apparaît dans la rangée inférieure. Il montre que le mot initial "Edsviken" est devenu "Esvken", et donc qu'il y a risque de perte de contenu aux fréquences inférieures à 20 fps.

Dans la séquence à 25 fps, la distribution des lettres est la suivante:

3 lettres sont représentées par 1 image;

0 lettre est représentée par 2 images;

3 lettres sont représentées par 3 images;

1 lettre est représentée par 4 images;

1 lettre est représentée par 7 images (fin de la phrase).

Longueur moyenne des lettres à l'intérieur du mot: 2,3 images.

Conclusion

Dans cet exemple, les lettres formant le mot occupent de 1 à 4 images d'une durée de 40 ms chacune. La durée moyenne de visibilité est de 2,3 images par lettre. L'exemple n'est pas assez long pour permettre d'en tirer des conclusions statistiques concrètes, mais on constate qu'à la vitesse de la dactylogogie, une cadence de 25 fps paraît suffisante alors qu'une cadence de 12,5 fps nécessite, pour la compréhension des mots exprimés par dactylogogie, une certaine intuition.

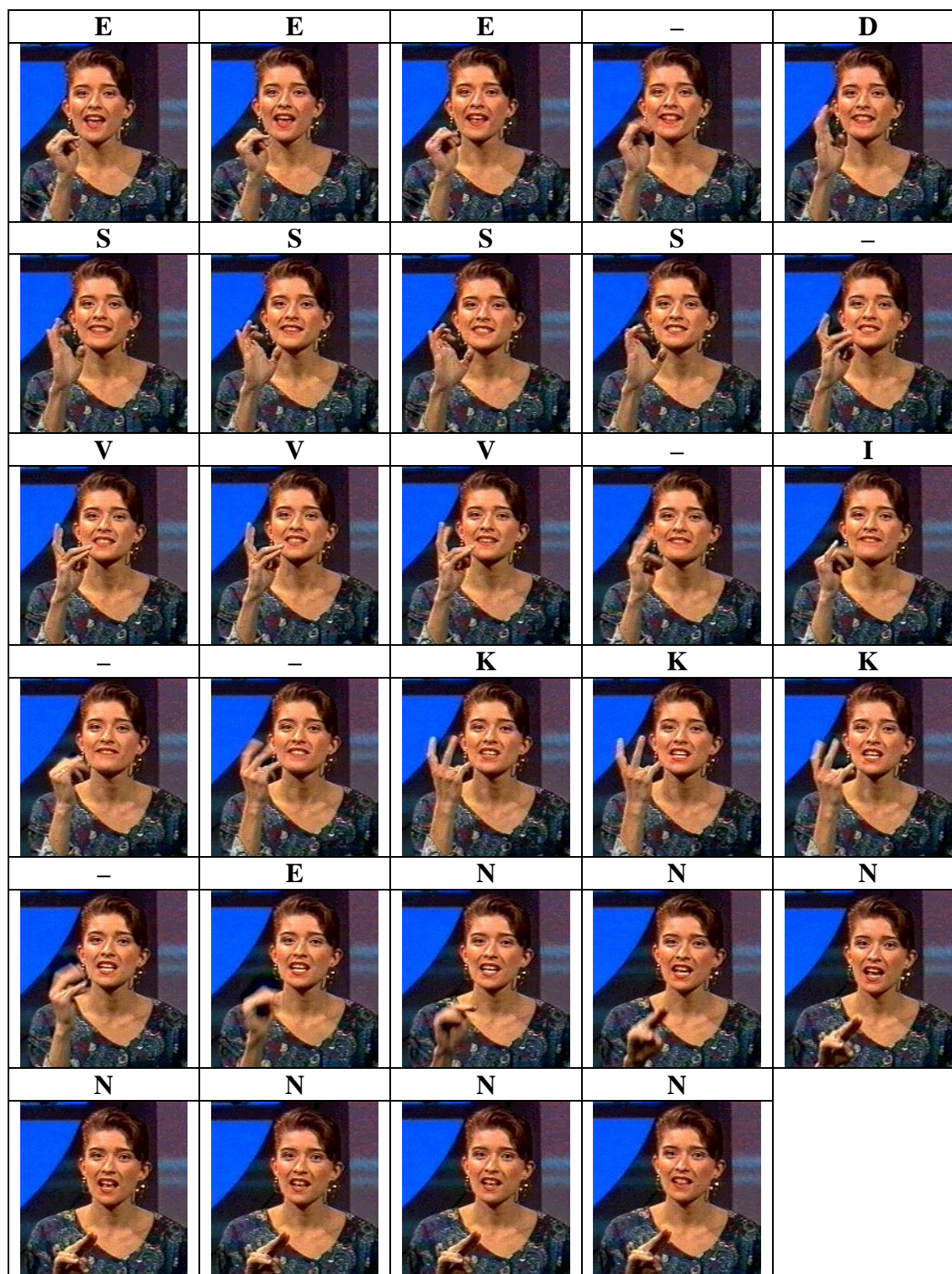


Figure 1 – Images de la dactylogogie du mot "Edsviken" enregistrées à 25 fps

L'expression signée générale

Des parties importantes de l'extrait du vidéoclip "Irene" sont exprimées au moyen de signes dans lesquels la dactylogogie n'intervient pas.

Une analyse simple portant sur une seule phrase est présentée ci-après. Elle est transcrite signe par signe avec, entre parenthèses, le nombre d'images qu'occupe chacun de ces signes.

La séquence occupe les images 406 à 529 du vidéoclip en question.

"ELLE(7) DIT(7) ELLE-MÊME(11) COMBIEN(4) ELLE(2) RESSENT(11) VECU(13) ADOLESCENCE(16)".

Aucun signe de la séquence n'occupe moins de deux images et aucun ne comporte de gestes plus rapides que ceux de la dactylogogie. Certains signes comportent des gestes plus amples, qui présentent d'autres exigences au niveau du codage vidéo.

Marqueurs grammaticaux

Les trames 394 et 395 contiennent un "clignement grammatical" indiquant une fin de phrase.

Ce marqueur a une longueur suffisante pour un échantillonnage à 12,5 fps.

5.2.6 Granularité de la résolution temporelle

Dans la plupart des cas, la communication vidéo se fait au moyen d'une caméra vidéo conforme aux normes habituelles, notamment la fréquence de 25 ou de 30 fps, qui introduit une certaine granularité aux fréquences utilisables. À l'aide d'une telle caméra, il n'est guère utile d'envisager les fréquences comprises entre 12,5 et 25 fps ou 15 et 30 fps, dont les intervalles de l'image source varient entre 40 et 80 ms ou 33 et 66 ms, respectivement, ce qui présente des risques de perte de détail du mouvement. On en conclut que pour répondre au besoin de 20 fps avec des caméras de télévision courantes, il convient d'utiliser la fréquence de 25 ou de 30 fps.

5.3 Besoins de résolution spatiale

En ce qui concerne la résolution spatiale, on a constaté, pour les communications de personne à personne au moyen du langage signé: [1]

- que l'on peut utiliser la résolution QCIF, mais que celle-ci ne rend pas les détails les plus fins, montrant la direction du regard, ce qui demande un effort à l'interlocuteur;
- que la résolution CIF est satisfaisante. Le passage de la résolution QCIF à la résolution CIF assure une meilleure compréhension de la parole;
- que la résolution SQCIF est trop grossière pour une représentation fiable, mais que certains signes sont parfois compris;
- que si l'on utilise des résolutions différentes selon les parties de l'image, il faudra accorder la résolution la plus élevée aux mains et au visage. Dans de tels systèmes, il faut prendre garde de ne pas introduire dans d'autres parties de l'image des distorsions susceptibles de détourner l'attention de l'interlocuteur.

Une vérification théorique simple est la suivante: dans le plan américain (personnage coupé à mi-corps) généralement utilisé dans les conversations de personne à personne au moyen du langage signé, un doigt correspond à 1/50 environ de la largeur d'image. Pour restituer un doigt avec une résolution suffisante, il doit être représenté par trois pixels au moins. Cela fixe la résolution spatiale minimale au niveau de la QCIF, qui occupe 176 pixels en largeur. La direction du regard, elle aussi importante dans le langage visuel, nécessite une résolution plus élevée. On préférera la CIF, mieux adaptée.

Dans le cas de la lecture labiale en communication de personne à personne, on peut resserrer le champ de manière à obtenir un plan rapproché du visage. On a constaté que dans de telles

conditions, la résolution QCIF est suffisante pour la lecture labiale [3], mais l'interlocuteur doit s'assurer que l'écran de visualisation se trouve à une distance telle que la résolution relativement modeste n'entraîne pas une représentation trop vague.

5.4 Fidélité

En communication vidéo, l'image comporte souvent du flou.

On ne dispose guère de modèles pour décrire le flou. Comme le type de flou et ses effets sur la représentation peuvent varier considérablement, le présent examen sera limité à une brève description du flou acceptable en différentes circonstances.

On admet généralement que la qualité de la vidéo VHS est suffisante pour une bonne représentation du langage signé et de la lecture labiale. Dans les enregistrements vidéo, les objets animés de mouvements rapides présentent souvent un flou marqué résultant de la vitesse d'obturation, généralement de 1/50 à 1/60 s. Cela montre que le flou est acceptable pour des éléments animés de mouvements rapides tels que des gestes larges utilisés dans le langage signé.

Les gestes larges comportent parfois du flou. Quand cela se produit, la résolution spatiale ne doit jamais être inférieure à la SQCIF, car celle-ci est trop grossière pour une représentation fiable.

Pour que la représentation soit bonne lorsque la résolution spatiale de base est la résolution CIF, le flou occasionnel ne devrait pas entraîner une qualité inférieure à celle perçue en résolution QCIF.

5.5 Temps de transmission

Le temps de transmission de bout en bout des données vidéo, de la caméra jusqu'à l'écran de réception, est déterminant dans la communication de personne à personne. Les durées inférieures à 0,4 seconde sont les plus recommandées, mais elles sont d'autant plus indiquées qu'elles se rapprochent de 0,1 seconde [3].

Les durées supérieures à 0,8 seconde se sont avérées gênantes dans les conversations au moyen du langage signé [1].

Il apparaît que les conditions requises pour le langage signé et la lecture labiale sont analogues à celles de la conversation verbale. Le temps qui s'écoule entre la prononciation ou l'expression gestuelle et la réaction attendue est au moins deux fois plus long que le temps de transmission. Même la durée de 0,4 seconde spécifiée par la Recommandation G.114 [2] semble longue car il faut, dans ce cas, un temps d'attente de 0,8 seconde pour obtenir la réponse.

5.6 Synchronisme

Le synchronisme entre le son et l'image est déterminant pour la lecture labiale accompagnée de la parole. On a constaté que les écarts étaient acceptables jusqu'à 100 ms [1].

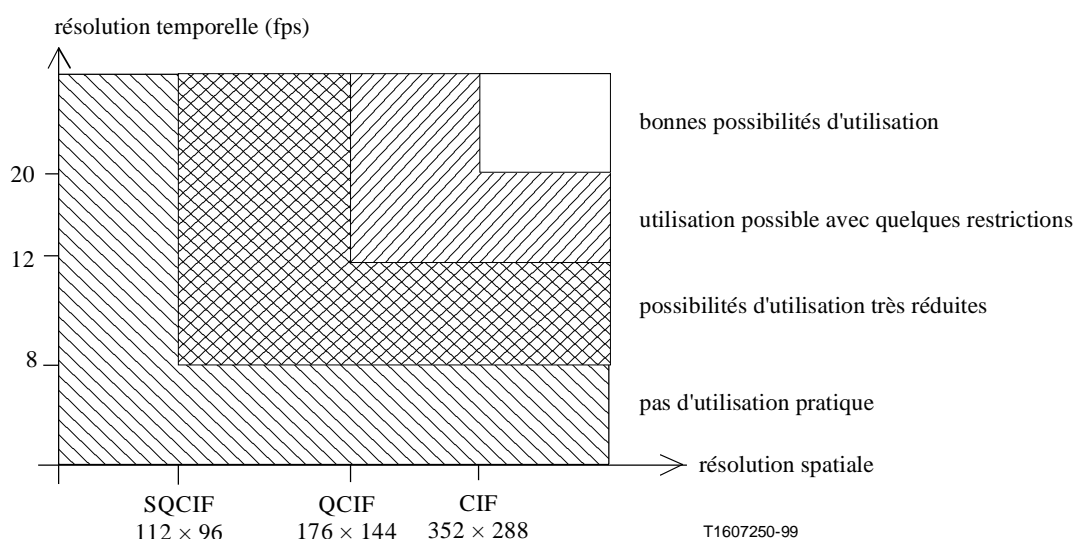
La combinaison parole/lecture labiale, lorsqu'elle est possible, est très efficace au niveau de la compréhension [3].

5.7 Conclusions relatives au niveau de qualité

Les objectifs de qualité suivants sont applicables à la transmission des conversations de personne à personne au moyen du langage signé et de la lecture labiale:

- choisir de préférence une fréquence de 25-30 fps à la résolution CIF et un temps de transmission maximal de 0,4 seconde, tout en acceptant un flou occasionnel des gestes de moyenne amplitude qui abaisse la qualité au niveau QCIF;
- en transmission à très faible débit, accepter si nécessaire une fréquence de 12-15 fps avec résolution QCIF pour les gestes de moyenne amplitude et un abaissement occasionnel au niveau SQCIF pour les gestes larges;

- maintenir le synchronisme avec le son au-dessous de 100 ms;
- le temps de transmission de bout en bout doit être inférieur à 0,4 seconde, mais si cela est impossible, accepter jusqu'à 0,8 seconde.



NOTE – Les valeurs doivent être respectées en présence de mouvements du langage signé ou de lecture labiale.

Figure 2 – Besoins de résolution pour le langage signé et la lecture labiale dans les conversations de personne à personne

Tableau 2 – Dégradation des possibilités d'utilisation causée par le temps de propagation et le flou

Possibilité d'utilisation	Temps de bout en bout	Flou occasionnel au cours des gestes larges	
		En résolution CIF	En résolution QCIF
Bonne	<0,4 s	Non	–
Utilisable avec quelques restrictions	0,4 – 0,8 s	Dégradation à \cong QCIF	Non
Utilisation limitée	0,8 – 1,2 s	Dégradation à \cong SQCIF	Dégradation à \cong SQCIF
Pas d'utilisation pratique	>1,2 s	Dégradation à < SQCIF	Dégradation à < SQCIF

6 Vérification du niveau de performances

6.1 Matériel de référence

Le présent supplément comporte un CD-ROM qui contient un vidéoclip servant à évaluer les performances de transmission de langage signé. Ce vidéoclip, intitulé "Irene", est repris d'un programme de la télévision suédoise, et contient un échantillonnage représentatif de mouvements de langage signé de rapidité normale.

Les droits de ce clip sont réservés à la compagnie de radiodiffusion éducative suédoise (*swedish educational broadcasting company*). L'Appendice I reproduit le fichier `Readme.txt` du CD-ROM,

qui contient la déclaration de droits d'auteur et la description technique du contenu des fichiers électroniques.

6.2 Évaluation du niveau de qualité

On soumet à des essais un codec ou un terminal par la transmission de la scène à évaluer via un codec ou une paire de visiophones connectés au réseau. Les résultats sont enregistrés et évalués. La Recommandation P.931 [4] spécifie une méthode d'évaluation.

Évaluation de la fréquence d'image pendant la séquence signée.

Consignation de la résolution statique sélectionnée.

Mesure du flou de bougé des gestes de moyenne amplitude par comparaison avec les images de la même scène enregistrées avec une résolution abaissée à QCIF ou SQCIF. Le flou est uniquement évalué sur les mains et le visage.

Mesure du temps de transmission.

Mesure du synchronisme du son (voix) et de l'image (mouvement des lèvres).

On peut évaluer, à partir de ces enregistrements, les niveaux de qualité et les comparer aux objectifs ci-dessus.

Pour faire une évaluation approximative de ces valeurs en l'absence d'équipement de laboratoire, on peut recourir à un outil d'évaluation simple de *national swedish association of the deaf*.

7 Conseils pour l'implémentation du terminal

Afin de répondre aux besoins des utilisateurs, le terminal doit présenter certaines caractéristiques:

- il doit disposer d'une interface pour activer des systèmes d'avertissement extérieurs tels que: témoin clignotant, vibreur de poche ou de la taille d'une montre, puissant générateur de bruit;
- étant donné que les utilisateurs souhaitent parfois revenir à la conversation en mode texte, il est souhaitable d'implémenter dans le terminal le protocole T.140 correspondant.
- la préférence pour une fréquence supérieure à 20 fps et un temps de transmission inférieur à 0,4 seconde nécessite l'emploi d'un algorithme sans saut d'image. Une fréquence rapide permet d'obtenir automatiquement un temps de transmission raisonnable;
- des exceptions à toutes les prescriptions de qualité sont tolérées pendant la période de 2 secondes suivant un changement de scène;

8 Conseil pour l'utilisateur

L'utilisateur doit créer un environnement disposant de bonnes conditions d'éclairage, sur fond uni.

9 Elargissement du domaine

Si l'équipement doit être utilisé pour les applications de langage signé et de lecture labiale en visioconférence, multidiffusion, radiodiffusion ou recherche d'informations, les éléments ci-après modifient les prescriptions:

- le champ est généralement plus large, englobant les personnes communiquant par langage verbal et des objets. Cela indique que la possibilité d'utilisation commence au niveau de la résolution spatiale CIF;
- l'utilisateur a moins d'occasions de donner des retours d'information permettant de mieux gérer la perception en informant l'orateur ou la personne qui communique par langage signé. Pour cette raison, il faut une fréquence supérieure à 20 fps;

- les prescriptions en matière de temps de transmission sont moins rigoureuses. En cas de radiodiffusion ou de recherche d'informations, des temps de transmission de plusieurs secondes sont acceptables. En cas de conférence, elles sont analogues à celles qui s'appliquent à la conversation;
- les prescriptions exactes pour chaque application ne relèvent pas du présent profil d'application.

APPENDICE I

Déclaration de copyright et description technique s'appliquant à la séquence vidéo "Irene"

Le présent appendice est la traduction du contenu du fichier Readme.txt du CD-ROM

Supplément 1 aux Recommandations UIT-T de la série H (06/1999)

Vidéoclip "Irene"

Version 1.0, juin 1999

I.1 Copyright

Tous droits réservés. Ce matériel ne peut être utilisé qu'à des fins de recherche et de développement portant sur des produits destinés aux malentendants. Il ne peut être inclus dans un produit commercial sans la permission de *swedish educational broadcasting company*. Toute autre utilisation du matériel est interdite.

I.2 Support

Pour la distribution et la mise à jour du logiciel, prière de contacter:

Département des ventes
UIT
Place des Nations
CH-1211 Genève 20
SUISSE
Courrier électronique: sales@itu.int

Pour tout problème technique, prière de contacter le service support du TSB:

TSB service support
UIT
Place des Nations
CH-1211 Genève 20
SUISSE
Télécopie: +41 22 730 5853
Courrier électronique: tsbedh@itu.int

I.3 Détails sur la séquence vidéo

Ce clip vidéo contient une séquence de langage signé à utiliser comme matériel de test pour le codage vidéo. Les signes sont exécutés à vitesse normale.

La séquence est intitulée "Irene" d'après le nom de la personne filmée. Cette séquence, en langage signé suédois, est extraite d'une production de la compagnie de radiodiffusion éducative suédoise. Le

champ couvre le plan américain généralement adopté par les personnes qui utilisent les visiophones pour signer. La séquence est enregistrée en système PAL, à 25 fps. Elle est fournie en trois formats:

- 1) Sign_Irene.mpeg (3261 Koctets)
codage MPEG-1 en résolution CIF à 25 fps.
- 2) Sign_Irene.cif (80 190 Koctets)
Format YCbCr 4:2:0 en résolution CIF à 25 fps.
- 3) Sign_Irene.qcif (20 048 Koctets)
Format YCbCr 4:2:0 en résolution QCIF à 25 fps.

Contenu dactylogologique

Le tableau suivant est une représentation approximative de deux séquences dactylogologiques du clip "Irene". Les nombres sont les numéros des images depuis le début de la version MPEG. Les lettres sont indiquées lorsqu'elles sont clairement formées par la main. Un tiret indique une transition pendant laquelle aucune lettre n'est clairement formée. Le premier nom est "Pia Wickman", dans lequel le dernier "a" n'est visible qu'en labial.

"Pia Wickman"									
Fr.	Ltr.	Fr.	Ltr.	Fr.	Ltr.	Fr.	Ltr.	Fr.	Ltr.
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"									
Fr.	Ltr.	Fr.	Ltr.	Fr.	Ltr.	Fr.	Ltr.	Fr.	Ltr.
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		

Contenu signé général

La dernière phrase du clip est signée en langage signé symbolique (non dactylogologique), dans lequel chaque signe correspond à un mot.

La phrase est retranscrite ci-dessous signe par signe, avec indication entre parenthèses du nombre d'images occupées par chaque signe: "ELLE(7) DIT(7) ELLE-MEME(11) COMBIEN(4) ELLE(2) RESSENT(11) VECU(13) ADOLESCENCE(16)".

La séquence occupe les images de 406 à 529 de la version MPEG.

Composants grammaticaux

Le clip contient un certain nombre de clignements d'yeux, typiques de la syntaxe du langage signé, utilisés comme délimiteurs syntaxiques. Ils sont rapides, et n'occupent souvent qu'une ou deux images. Un clignement syntaxique apparaît dans les images 394 et 395.

SERIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série B	Moyens d'expression: définitions, symboles, classification
Série C	Statistiques générales des télécommunications
Série D	Principes généraux de tarification
Série E	Exploitation générale du réseau, service téléphonique, exploitation des services et facteurs humains
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission, systèmes et réseaux numériques
Série H	Systèmes audiovisuels et multimédias
Série I	Réseau numérique à intégration de services
Série J	Transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	RGT et maintenance des réseaux: systèmes de transmission, de télégraphie, de télécopie, circuits téléphoniques et circuits loués internationaux
Série N	Maintenance: circuits internationaux de transmission radiophonique et télévisuelle
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux
Série Q	Commutation et signalisation
Série R	Transmission télégraphique
Série S	Equipements terminaux de télégraphie
Série T	Terminaux des services télématiques
Série U	Commutation télégraphique
Série V	Communications de données sur le réseau téléphonique
Série X	Réseaux pour données et communication entre systèmes ouverts
Série Y	Infrastructure mondiale de l'information et protocole Internet
Série Z	Langages et aspects informatiques généraux des systèmes de télécommunication