



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.920

(08/96)

SÉRIE P: QUALITÉ DE TRANSMISSION
TÉLÉPHONIQUE

Qualité audiovisuelle dans les services multimédias

**Méthodes d'essai interactives pour
communications audiovisuelles**

Recommandation UIT-T P.920

(Antérieurement Recommandation du CCITT)

RECOMMANDATIONS UIT-T DE LA SÉRIE P
QUALITÉ DE TRANSMISSION TÉLÉPHONIQUE

Vocabulaire et effets des paramètres de transmission sur l'opinion des usagers	Série	P.10
Lignes et postes d'abonnés	Série	P.30
		P.300
Normes de transmission	Série	P.40
Appareils de mesures objectives	Série	P.50
		P.500
Mesures électroacoustiques objectives	Série	P.60
Mesures de la sonie vocale	Série	P.70
Méthodes d'évaluation objective et subjective de la qualité	Série	P.80
		P.800
Qualité audiovisuelle dans les services multimédias	Série	P.900

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

RECOMMANDATION UIT-T P.920

METHODES D'ESSAI INTERACTIVES POUR COMMUNICATIONS AUDIOVISUELLES

Résumé

La présente Recommandation est destinée à définir des méthodes interactives d'évaluation permettant de quantifier l'influence des procédés de codage et des temps de transmission sur des communications audiovisuelles point à point ou multipoint. Cette méthode est fondée sur des essais d'opinion en conversation. Elle peut être considérée comme étant une extension des méthodes définies dans l'Annexe A/P.800.

Les points ci-après seront pris en considération dans les articles qui suivent:

- tâches à proposer aux évaluateurs;
- méthodes et modèles expérimentaux;
- questionnaires.

Source

La Recommandation UIT-T P.920, élaborée par la Commission d'études 12 (1993-1996) de l'UIT-T, a été approuvée le 30 août 1996 selon la procédure définie dans la Résolution n° 1 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 (Helsinki, 1^{er}-12 mars 1993).

Dans certains secteurs de la technologie 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 Recommandation, l'expression «Administration» est utilisée pour désigner de façon abrégée aussi bien une administration de télécommunications qu'une exploitation reconnue.

© UIT 1996

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 Introduction.....	1
2 Domaine d'application.....	1
3 Références.....	1
4 Modèle expérimental	2
4.1 Méthode de base et facteurs à étudier.....	2
4.2 Stimuli utilisés dans les essais de conversation.....	2
4.3 Conditions d'essai et modèles expérimentaux	3
4.4 Sujets.....	3
4.5 Apprentissage par les sujets et connexions de référence.....	4
4.6 Caractéristiques de la salle d'essai et de l'équipement.....	4
5 Recueil des opinions	5
Bibliographie.....	5
Appendice I – Exemples de tâches et de stimuli pour la conversation.....	5
I.1 Stimuli pour la conversation.....	5
I.2 Tâches visant à évaluer les effets du temps de propagation en conversation sur la qualité des communications.....	6
I.3 Tâches visant à évaluer les effets du temps de propagation des signaux audiovisuels sur la qualité des communications.....	6
I.4 Tâche visant à évaluer la synchronisation entre signaux audio et vidéo.....	6
Appendice II – Protocoles pour les stimuli conversationnels.....	7
II.1 Protocole pour la tâche du nom à deviner.....	7
II.2 Protocole pour la tâche de comparaison d'histoire.....	7
II.3 Protocole pour la tâche de comparaison d'images.....	8
Appendice III – Questionnaire relatif aux conditions d'essai.....	8
Appendice IV – Questions finales.....	9

Recommandation P.920

METHODES D'ESSAI INTERACTIVES POUR COMMUNICATIONS AUDIOVISUELLES

(Genève, 1996)

1 Introduction

Les présentes méthodes d'essai interactives pour communications audiovisuelles sont destinées à quantifier l'influence de grandeurs telles que les procédés de codage et le temps de transmission, qui peuvent affecter la capacité à réaliser une communication interactive audiovisuelle.

L'efficacité de ces essais dépend étroitement de la capacité de reproduire, dans des environnements de laboratoire, les conditions les plus proches des situations réelles. A ce propos, il faut prendre des précautions particulières lors du choix des tâches proposées aux sujets. En général, celles qui sont utilisées lors des essais de conversation pour évaluations téléphoniques ne conviennent pas pour l'évaluation audiovisuelle parce qu'elles attirent souvent l'attention du sujet ailleurs que sur l'écran de visualisation. De nouvelles tâches doivent donc être développées selon les critères illustrés dans la présente Recommandation.

D'importants travaux ont été effectués dans ce domaine, bien que tous les aspects de la qualité audiovisuelle n'aient pas été totalement élucidés.

La présente Recommandation reflète l'état d'avancement actuel de la recherche en matière d'évaluations interactives de la qualité audiovisuelle.

Au fur et à mesure des progrès accomplis dans ce domaine, il est certain que la compréhension de ces méthodes d'essai interactives ira en s'améliorant. La présente Recommandation sera révisée lorsque de nouvelles connaissances auront été obtenues.

2 Domaine d'application

La présente Recommandation est destinée à définir des méthodes interactives d'évaluation permettant de quantifier l'influence des procédés de codage et des temps de transmission sur des communications audiovisuelles point à point ou multipoint. Cette méthode est fondée sur des essais d'opinion en conversation. Elle peut être considérée comme étant une extension des méthodes définies dans l'Annexe A/P.800.

La présente Recommandation ne traite pas des sujets déjà définis dans d'autres Recommandations, comme les mesures objectives de la qualité d'une liaison déjà définies dans d'autres Recommandations de la série P.

3 Références

Les Recommandations et autres références suivantes contiennent des dispositions qui, par suite de la référence qui y est faite, constituent des dispositions valables pour la présente Recommandation. Au moment de la publication, les éditions indiquées étaient en vigueur. Toute Recommandation ou autre référence est sujette à révision; tous les utilisateurs de la présente Recommandation sont donc invités à rechercher la possibilité d'appliquer les éditions les plus récentes des Recommandations et autres références indiquées ci-après. Une liste des Recommandations ITU-T en vigueur est publiée régulièrement.

- Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission*.
- Recommandation UIT-R BT.812 (1994), *Evaluation subjective de la qualité des images alphanumériques et graphiques en télétexte et dans des services similaires*.

4 Modèle expérimental

4.1 Méthode de base et facteurs à étudier

Afin de quantifier l'incidence de facteurs (tels que le temps de transmission) qui peuvent affecter la capacité à conduire une communication interactive, l'approche proposée dans la présente Recommandation est fondée sur une évaluation de conversation active entre locuteurs. Par ailleurs, comme il est nécessaire d'exprimer ces opinions au moyen d'un système d'évaluation, plusieurs échelles d'évaluation à stimulus unique sont proposées.

4.2 Stimuli utilisés dans les essais de conversation

En général, dans les essais d'opinion conversationnels, on souhaite minimiser l'artificialité de l'environnement. Il est toutefois nécessaire, en même temps, de faire appel à une méthode d'incitation à la communication interactive utilisant les conditions qui sont en cours d'évaluation. Pour les évaluations téléphoniques, il est courant d'utiliser à cette fin un jeu de photographies ou d'autres imprimés. Pour les évaluations des performances de terminaux audiovisuels, de tels mécanismes sont toutefois susceptibles d'attirer l'attention du participant ailleurs que sur l'écran d'affichage, ce qui peut conduire à un mode de communication non naturel pour ce type de terminal.

Pour les applications générales, les directives suivantes sont proposées afin de concevoir des essais fondés sur des tâches à accomplir [3]:

- la tâche doit être conçue de manière que, au cours de leur conversation, les sujets maintiennent l'essentiel de leur attention sur le terminal audiovisuel;
- la tâche doit avoir une valeur intrinsèque suffisante, c'est-à-dire qu'elle doit suffisamment ressembler à une communication audiovisuelle réelle. Il est en particulier préférable que la tâche soit exécutée par deux sujets et non pas par un sujet avec un chef d'expérience;
- la tâche doit donner des résultats quantitatifs reproductibles, représentant des degrés suffisants d'efficacité de communication. Lorsque des durées sont en cause, les mesures de temps doivent figurer dans les résultats.

Une grande diversité de sujets, y compris des personnes âgées et malentendantes, doivent être en mesure d'accomplir la tâche.

Il est préférable que la tâche soit, par elle-même, suffisamment gratifiante pour les sujets, ce qui présente plusieurs avantages: les sujets apprennent la tâche plus rapidement et sont moins portés à la fatigue et à la perte de motivation.

Des expériences passées ont montré que l'on pouvait provoquer des conversations audiovisuelles animées si les participants à l'essai se connaissaient déjà. Par la suite, on peut faire appel à la présentation de matériel écrit en tant que source secondaire, plutôt que primaire, de stimulation. Contrairement à la téléphonie, une certaine familiarité entre paires de participants à la conversation est hautement souhaitable, pour ne pas dire essentielle.

On admet cependant que, pour des applications spécifiques, les tâches conversationnelles peuvent avoir à être modifiées pour tenir compte des services que le système en essai est destiné à fournir. Pour que l'on puisse obtenir des mesures représentatives des facteurs en cours d'étude, il est

recommandé que, dans de tels cas, les tâches conversationnelles soient structurées de façon à représenter les applications étudiées, en particulier pour ce qui est:

- a) du débit des informations échangées;
- b) et du taux d'utilisation du signal audio et vidéo.

Par exemple, pour tenir compte des attributs relevant de la première catégorie, les tâches pourraient aller d'une communication pratiquement dans un seul sens à une conversation libre ou à un rapide échange d'informations, que ce soit au moyen de signaux vidéo, de signaux audio ou des deux types de signaux. De même, pour tenir compte des attributs relevant de la deuxième catégorie, les tâches pourraient aller d'un travail effectué par les sujets sur un document imprimé placé devant eux (utilisation minimale des informations vidéo) à la lecture d'un texte en langue des signes par la liaison vidéo (utilisation maximale des informations vidéo). Les tâches proprement dites devront combiner des attributs des deux catégories. Ces directives ont été appliquées lors de la mise au point des tâches illustrées dans l'Appendice I, dont les protocoles sont exposés dans l'Appendice II.

4.3 Conditions d'essai et modèles expérimentaux

En général, au moins un facteur de dégradation des transmissions ou une condition d'essai est susceptible d'être évalué au cours d'un essai, en plus d'une condition de base (ou de référence) dans laquelle l'influence de ce facteur est minimale (lorsque l'on utilise la condition de référence, les participants doivent l'ignorer). Comme cependant les essais conversationnels prennent du temps, le nombre total de conditions doit être convenablement restreint, de façon à minimiser la fatigue des participants et à maximiser la précision expérimentale. Cette exigence doit être ajustée en fonction de la nécessité de faire en sorte que la durée de chaque conversation/condition soit d'au moins cinq minutes.

Comme dans le cas des essais conversationnels (communications audio), on peut estimer qu'un carré latin ou gréco-latin peut constituer un modèle expérimental approprié à cette fin [1]. Dans ce cas, les rangées du carré peuvent être associées aux participants à l'essai et les colonnes du carré à l'ordre dans lequel les conditions sont présentées au cours de l'essai.

D'autres traitements peuvent aussi convenir, selon les facteurs étudiés. Par exemple, des expériences effectuées ont paru indiquer qu'il pouvait y avoir une interaction entre la qualité du conduit de communication audiovisuelle et la grandeur perçue du temps de transmission. Par conséquent, il est parfois préférable d'appliquer deux traitements au moyen d'un carré gréco-latin, de manière que les lettres du premier alphabet soient associées à différentes valeurs du temps de transmission et que les lettres du deuxième alphabet soient associées à différents débits de codage des signaux vidéo ou audio.

D'autres modèles expérimentaux peuvent évidemment convenir. Le soin pourra donc être laissé à l'expérimentateur de les sélectionner de manière à atteindre des objectifs spécifiques de coût et de précision en fonction du nombre de conditions prises en compte.

Tout autres effets possibles concernant l'ordre dans lequel les tâches sont effectuées doivent être pris en compte.

4.4 Sujets

Au moins 16 sujets doivent participer à un essai. Le nombre exact dépendra du modèle expérimental et de la précision requise des résultats. Ces sujets ne devront pas être des experts et ne devront pas être directement impliqués, dans le cadre de leur activité normale, dans les techniques audio et/ou vidéo.

Dans les phases initiales du développement de systèmes de communications audiovisuelles et dans les expériences pilotes effectuées avant un assez grand essai, de petits groupes d'experts (4 à 8) ou d'autres sujets critiques peuvent apporter des résultats indicatifs, avec une fiabilité suffisante.

4.5 Apprentissage par les sujets et connexions de référence

Avant de commencer l'expérience, les sujets doivent recevoir un scénario de l'application prévue du système en essai. L'étendue et le type des dégradations doivent être indiqués au cours d'une phase préliminaire. Celle-ci peut aussi servir à effectuer un premier niveau de présentation des personnes par la liaison de communication dans la pire (ou la meilleure) condition expérimentale, tandis qu'ensuite on peut autoriser une discussion relative aux tâches attendues des participants dans la meilleure (ou la pire) condition expérimentale.

De nouveau, comme dans le cas de l'essai principal, les détails des conditions d'expérience ne doivent pas être révélés aux participants à l'essai.

4.6 Caractéristiques de la salle d'essai et de l'équipement

Les caractéristiques de bruit ambiant dans les salles d'essai doivent être représentatives des environnements bureautiques, tandis que l'isolement acoustique entre les deux environnements doit être meilleur que 60 dB. Les caractéristiques acoustiques des salles employées pour les évaluations conversationnelles et indiquées dans la Recommandation P.800 doivent également être applicables dans ce cas.

Les caractéristiques relatives à la vision ou à la luminance des salles, à la brillance de l'affichage vidéo et à la configuration matérielle par rapport à la (aux) source(s) d'éclairage et par rapport au(x) participant(s) sont les suivantes, conformément à la Recommandation UIT-R BT.812:

Distance de visualisation	4 à 8 H (Note)
Luminance (valeur de crête)	70 à 200 cd/m ²
Rapport de contraste de l'écran sans éclairage du fond	de 30 à 50
Rapport luminance du fond sur luminance maximale de l'écran	~0,25
Eclairage lumineux	environ 500 lx
Chromaticité générale	blanc

NOTE – La lettre H correspond à la hauteur de l'image.

Il y a lieu de définir la distance de visualisation en tenant compte non seulement des dimensions de l'écran mais aussi du type d'écran, du type d'application et du but de l'expérience. Pour une hauteur d'écran donnée, il est probable que la distance de visualisation préférée par les sujets augmentera lorsque la qualité visuelle se dégradera. Concernant ce point, il convient de prédéterminer la distance de visualisation préférée pour les essais de qualification. Pour les essais d'optimisation, il faudra utiliser des distances plus critiques.

Il est préférable d'utiliser toute la surface de l'écran pour afficher la scène. Si toutefois les séquences doivent, pour une raison ou pour une autre, être affichées dans une fenêtre de l'écran, la couleur du fond d'écran devra être le gris à 50 % correspondant à la valeur $Y=U=V=128$ (les valeurs U et V n'étant pas signées).

D'autres facteurs pouvant se rapporter au fonctionnement mains-libres peuvent être requis afin de minimiser les variations expérimentales résiduelles (pour complément d'étude).

5 Recueil des opinions

Comme dans le cas des évaluations conversationnelles téléphoniques, chaque participant doit aussi être invité séparément à donner son avis après l'application de chaque condition. Il est préférable de structurer d'avance le type de questionnement et de minimiser le nombre de questions posées après chaque condition ainsi que le nombre de variations résiduelles. On trouvera dans l'Appendice III un exemple de questionnaire sur les conditions d'essai qui pourrait être utilisé à cette fin.

Plusieurs échelles de jugement par catégorie peuvent être utilisées afin d'évaluer la performance des terminaux audiovisuels. Les échelles peuvent avoir diverses sensibilités selon les facteurs de dégradation de transmission.

Exemples d'échelle pouvant être utilisée à cette fin:

- qualité audiovisuelle globale, qualité vidéo, qualité audio. Ces qualités sont généralement évaluées au moyen des catégories suivantes: excellent, bon, satisfaisant, médiocre, mauvais;
- l'effort requis pour l'interruption est généralement évalué au moyen des catégories suivantes: pas d'effort, effort mineur, effort moyen, effort considérable, effort extrême;
- la difficulté et l'acceptabilité des communications sont généralement évaluées au moyen d'un choix binaire: oui ou non.

Bien que la sélection d'une échelle particulière puisse dépendre des buts d'une expérience individuelle, il importe que différents laboratoires utilisent ces échelles (et leur texte original ou traduit associé) de manière cohérente.

Si l'utilisation de terminaux audiovisuels est nouvelle pour la plupart des participants, il est recommandé de leur présenter des "Questions finales" après l'évaluation de la dernière condition d'essai. Dans de telles questions, on devra tenter d'intégrer tous les autres facteurs qui pourraient avoir été omis par inadvertance de l'expérience. L'Appendice IV donne un échantillon de telles questions.

Bibliographie

- [1] KIRK (R.E.): *Experimental design – Procedures for the behavioural sciences*, 2e édition, Brooks/Cole Publishing Co., Californie, 1982.
- [2] COM XII-85 (juillet 1991), *Proposals of quality assessment method due to delay and acceptable delay time*, NTT.
- [3] BRONKHORST (A.W.), VERHAVE (J.A.): *The effect of audio-video desynchronization on communication efficiency in videotelephony*, Etude pour PTT Telecom Netherlands, TNO Institute for Perception, Rapport IZF 1992 C-35, 1992.

Appendice I

Exemples de tâches et de stimuli pour la conversation

I.1 Stimuli pour la conversation

Les tâches suivantes diffèrent les unes des autres selon le degré auquel une conversation libre peut se produire [3]. Les protocoles relatifs à ces tâches sont décrits plus en détail dans l'Appendice II.

Tâche 1 Le nom à deviner – La tâche "nom à deviner" est un jeu de questions-réponses selon un protocole fixe, dont il résulte une conversation très restreinte.

- Tâche 2 La comparaison d'histoires – Dans la tâche "comparaison d'histoires", les sujets doivent découvrir un certain nombre de différences entre les deux versions d'une histoire. Ils sont autorisés à tenir une conversation sans restriction. Avant l'essai, les deux sujets doivent lire et mémoriser une brève histoire.
- Tâche 3 La comparaison d'images – Dans cette tâche, les sujets doivent mémoriser une image et ensuite déterminer s'ils ont reçu une image identique ou différente. La conversation n'est pas restreinte.

I.2 Tâches visant à évaluer les effets du temps de propagation en conversation sur la qualité des communications

Dans les tâches suivantes, l'émission de paroles augmente de la tâche 1 à la tâche 6, tandis que le rythme des inversions de sens de la conversation diminue [2]:

- 1) compter chacun à son tour;
- 2) lire chacun à son tour à haute voix des nombres aléatoires aussi rapidement que possible;
- 3) vérifier chacun à son tour à haute voix des nombres aléatoires aussi rapidement que possible;
- 4) compléter des mots comportant des lettres manquantes avec des lettres fournies par l'interlocuteur;
- 5) vérifier chacun à son tour des noms de ville aussi rapidement que possible;
- 6) déterminer la forme d'une figure décrite verbalement;
- 7) conversation libre.

Les tâches précédentes (à l'exception de la tâche 1 et de la tâche 7) ne peuvent pas être utilisées pour les évaluations de la qualité audiovisuelle parce que la plupart d'entre elles exigent que les sujets concentrent leur attention sur une feuille de papier et non sur un écran.

I.3 Tâches visant à évaluer les effets du temps de propagation des signaux audiovisuels sur la qualité des communications

Les tâches suivantes doivent attirer l'attention des évaluateurs sur le signal vidéo:

- 1) un des sujets montre et décrit un bloc de construction pour revêtement et l'autre sujet essaye de le reproduire;
- 2) un des sujets montre quelques exercices visant à soulager des douleurs cervicales et l'autre sujet essaye de les reproduire.

I.4 Tâche visant à évaluer la synchronisation entre signaux audio et vidéo

La tâche suivante a pour but d'attirer l'attention des évaluateurs sur la synchronisation entre signaux audio et vidéo:

- une personne applaudit tandis que l'autre vérifie la synchronisation entre le mouvement et le son.

La tâche 1 ne convient pas pour évaluer la synchronisation des lèvres.

D'autres tâches, visant à évaluer la synchronisation entre les deux signaux, feront l'objet de compléments d'étude.

Appendice II

Protocoles pour les stimuli conversationnels

II.1 Protocole pour la tâche du nom à deviner

En [3], le protocole suivant est utilisé pour la tâche du nom à deviner.

"La première tâche, celle du nom à deviner, est un jeu à questions-réponses selon un protocole fixe. Dans cette tâche, un des sujets reçoit trois éléments d'information: d'abord, soit le mot "Marque" ou le mot "Personne", pour indiquer s'il s'agit de deviner le nom d'une marque commerciale ou celui d'une personne (bien connue). Ensuite, dans le cas d'une marque, on décrit le produit ou, dans le cas d'une personne, la profession de celle-ci; en troisième lieu, le nom doit être deviné. Le sujet recevra par exemple le texte suivant:

- marque;
- cigarettes;
- camel."

"Pour deviner le nom de la marque ou de la personne, le deuxième sujet doit suivre le protocole suivant: la première question sera: "Est-ce une marque ou une personne?"; la deuxième, si c'est une marque, sera: "Quel est le produit?" ou, si c'est une personne: "Quelle est sa profession?". Ensuite, une hypothèse peut être formulée. Lorsque l'hypothèse est fautive, il faut demander les lettres consécutives du nom, avec possibilité d'hypothèse à chaque fois. Et ainsi de suite jusqu'à ce que le nom soit déduit correctement ou jusqu'à ce qu'il soit complètement épilé. Les sujets interrogés sont autorisés à consulter leur texte au cours de la conversation. La plupart des sujets ne jugent cependant pas cela nécessaire car ils n'ont eu aucune peine à retenir le peu d'information contenu dans ce texte.

Une fois le nom trouvé, on évalue le temps total nécessaire et le nombre de lettres qui ont été demandées. Plusieurs noms, de longueur et de difficulté variables, devront être trouvés dans chaque condition d'expérience. Une analyse de régression linéaire pourra ensuite être appliquée aux données de manière à obtenir des estimations du temps requis pour trouver directement un nom (aucune lettre suggérée) et le temps supplémentaire par lettre."

II.2 Protocole pour la tâche de comparaison d'histoire

En [3], le protocole suivant est utilisé pour la tâche de comparaison d'histoire.

"Dans la deuxième tâche, les sujets doivent découvrir un certain nombre de différences entre deux versions d'une histoire. Ils ont le droit d'avoir une conversation sans restriction. Avant l'essai proprement dit, les deux sujets doivent lire et mémoriser une brève histoire. On leur communique des histoires pratiquement semblables mais qui contiennent un certain nombre de différences, légères mais distinctes. (Par exemple des histoires d'environ 200 mots, contenant six différences.) En outre, les sujets reçoivent une liste de questions portant sur l'histoire, auxquelles ils doivent répondre eux-mêmes. Ces questions visent à améliorer la conservation de l'histoire en mémoire. Après la période de mémorisation, les sujets doivent entamer une conversation en vue de découvrir aussi rapidement que possible toutes les différences entre les deux histoires. Ils ne sont pas autorisés à consulter le texte pendant la conversation. Les sujets savent combien de différences il y a et ils sont informés en retour lorsqu'ils ont détecté une différence. On leur signale également le nombre de différences qu'ils ont détectées jusque là. La conversation se poursuit jusqu'à ce que toutes les différences soient détectées ou jusqu'à ce qu'aucune différence n'ait été détectée pendant un certain temps."

"L'instant de départ de la conversation et les instants où intervient la détection d'une différence sont enregistrés. Dans l'analyse, on détermine l'intervalle de temps entre le départ et la détection de la première différence ainsi que les intervalles entre les détections suivantes. Ces intervalles sont

ensuite ordonnés en fonction de leur durée, parce que l'intervalle de temps, c'est-à-dire le temps nécessaire pour la détection de la différence suivante, dépend à la fois de la stratégie utilisée par la paire de sujets et de l'emplacement des différences dans le texte. Le réordonnement des intervalles réduit la variance induite par ces facteurs. Il permet par ailleurs d'évaluer l'effet étudié en fonction de la durée de l'intervalle."

II.3 Protocole pour la tâche de comparaison d'images

En [3], on utilise le protocole suivant pour la tâche de comparaison d'images.

"Dans cette tâche, les sujets doivent mémoriser une image puis déterminer s'ils ont reçu des images identiques ou différentes. (On peut utiliser des images de divers objets, comme des paysages, des bâtiments et des sites urbains.) La conversation n'est pas restreinte. Les sujets ne sont pas autorisés à regarder l'image pendant la conversation.

Les résultats qui sont notés sont le temps total pris par les sujets pour parvenir à une décision et le fait que celle-ci est ou non correcte. Etant donné que les sujets ont normalement besoin de plus de temps lorsque les images sont identiques que lorsqu'elles sont différentes, les résultats doivent être séparés dans les deux cas. Bien que cette tâche ressemble un peu à celle de la comparaison d'histoires, il y a une différence dans le degré d'apparition d'une conversation libre parce que les sujets, en comparant des histoires, ont tendance à reprendre des phrases dans le texte, connues des deux interlocuteurs, tandis que la tâche de comparaison d'images les oblige à utiliser leurs propres énoncés."

Appendice III

Questionnaire relatif aux conditions d'essai

Ce questionnaire doit être utilisé après chaque condition d'essai.

L'administrateur: Je vais vous poser (de nouveau) quelques questions concernant votre opinion au sujet de la connexion par laquelle vous venez d'avoir une conversation. Etes-vous prêt?

- 1 Comment évaluez-vous la qualité audiovisuelle globale? (cerclez une réponse)
Excellent Bon Satisfaisant Médiocre Mauvais

- 2 Comment évaluez-vous la qualité vidéo de la connexion? (cerclez une réponse)
Excellent Bon Satisfaisant Médiocre Mauvais

- 3 Comment évaluez-vous la qualité audio de la connexion? (cerclez une réponse)
Excellent Bon Satisfaisant Médiocre Mauvais

4 Comment jugez-vous l'effort nécessaire pour interrompre votre (vos) interlocuteur(s)? (cerclez une réponse)

Pas d'effort	Effort mineur	Effort moyen	Effort considérable	Effort extrême
--------------	---------------	--------------	---------------------	----------------

5 Avez-vous rencontré une difficulté quelconque au cours de la connexion? (cerclez une réponse)

Oui Non

6 La connexion a-t-elle été acceptable? (cerclez une réponse)

Oui Non

L'administrateur: Merci de vos réponses. Nous allons maintenant reprendre l'expérience et rétablir notre connexion dans quelques minutes. Dès que la connexion sera établie, je vous dirai que vous pouvez commencer à converser avec votre interlocuteur.

Appendice IV

Questions finales

Les questions finales sont à poser à la fin de l'essai:

L'administrateur: Comme nous sommes arrivés à la fin de l'essai, je voudrais vous poser quelques questions supplémentaires. Etes-vous prêt?

Pouvez-vous m'indiquer, dans l'ordre d'importance, quels facteurs vous considérez comme suffisamment importants pour que vous souhaitiez qu'une amélioration leur soit apportée. (L'évaluateur donne alors lecture d'un certain nombre de facteurs.)

7a_____ 7b_____ 7c_____ 7d_____

Avez-vous d'autres observations au sujet de l'essai dans son ensemble, que vous voudriez nous communiquer?

Merci de votre coopération à l'essai. Je reviens dans quelques minutes pour fermer la session et vous accompagner vers la sortie.

SÉRIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série B	Moyens d'expression
Série C	Statistiques générales des télécommunications
Série D	Principes généraux de tarification
Série E	Réseau téléphonique et RNIS
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission
Série H	Transmission des signaux autres que téléphoniques
Série I	Réseau numérique à intégration de services
Série J	Transmission des signaux radiophoniques et télévisuels
Série K	Protection contre les perturbations
Série L	Construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	Maintenance: systèmes de transmission, de télégraphie, de télécopie, circuits téléphoniques, et circuits loués internationaux
Série N	Maintenance: circuits internationaux de transmission radiophoniques et télévisuels
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique
Série Q	Commutation et signalisation
Série R	Transmission télégraphique
Série S	Equipements terminaux de télégraphie
Série T	Equipements terminaux et protocoles 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 Z	Langages de programmation