



UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

P.931

(12/98)

SÉRIE P: QUALITÉ DE TRANSMISSION
TÉLÉPHONIQUE, INSTALLATIONS TÉLÉPHONIQUES
ET RÉSEAUX LOCAUX

Qualité audiovisuelle dans les services multimédias

**Mesure du temps de transmission, de la
synchronisation et du débit de trames dans
les communications multimédias**

Recommandation UIT-T P.931

(Antérieurement Recommandation du CCITT)

RECOMMANDATIONS UIT-T DE LA SÉRIE P
**QUALITÉ DE TRANSMISSION TÉLÉPHONIQUE, INSTALLATIONS TÉLÉPHONIQUES ET RÉSEAUX
LOCAUX**

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.931

MESURE DU TEMPS DE TRANSMISSION, DE LA SYNCHRONISATION ET DU DEBIT DE TRAMES DANS LES COMMUNICATIONS MULTIMEDIAS

Résumé

Les véritables systèmes de communications multimédias sont caractérisés par une capacité qui les distingue d'un simple bouquet de voies indépendantes: celle de maintenir la relation temporelle entre les différents médias.

La présente Recommandation spécifie les paramètres et les méthodes de mesure permettant d'évaluer la synchronisation relative entre voies médias, ainsi que deux autres paramètres importants de la qualité temporelle. Le temps de transmission (ou de propagation) dans une voie est un paramètre critique lorsque l'on évalue l'aptitude d'un système à transporter des conversations et d'autres signaux interactifs. La durée entre arrivées de trames et son inverse, le débit (ou la fréquence) de trames, caractérisent la capacité d'un système à acheminer des informations de manière continue et cohérente.

Les systèmes multimédias combinent des voies vidéo, des voies audio et des voies de données pour améliorer les communications. La présente Recommandation traite de tous ces médias. Le temps de transmission vidéo peut varier largement sur de courtes séquences. Les séquences audio et vidéo peuvent subir des distorsions en cours de transmission et les trains de données peuvent avoir une structure légère ou inexistante, assortie d'erreurs sur les bits. Bien que chaque média pose des problèmes de mesurage uniques, les méthodes ici spécifiées y répondent et les résolvent. La méthode fondée sur la variance concerne et mesure d'éventuelles variations instantanées du temps de transmission vidéo. La méthode de mesure du temps de transmission audio concerne des voies dans lesquelles la forme d'onde originale des paroles n'est pas préservée. Il existe également des méthodes applicables aux voies de données qui tirent parti des structures natives et qui tolèrent des erreurs sur les bits. Toutes ces méthodes impliquent des signaux d'essai qui sont représentatifs des applications prévues des systèmes.

Ces méthodes visent la saisie de séquences de trames à l'entrée et à la sortie des médias avec une échelle temporelle commune. Elles effectuent des comparaisons de trames afin de déterminer les trames de sortie actives (non répétées). Elles font correspondre les trames de sortie actives à des trames d'entrée uniques afin de déterminer le temps de transmission et la synchronisation. Ces méthodes permettent de constituer des ensembles contenant les paramètres souhaités dans leur forme élémentaire, en termes de temps de transmission, de dérive temporelle et de répartition des durées entre arrivées de trames.

Source

La Recommandation UIT-T P.931, élaborée par la Commission d'études 12 (1997-2000) de l'UIT-T, a été approuvée le 3 décembre 1998 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.

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 Recommandation, 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 Recommandation 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 Recommandations.

A la date d'approbation de la présente Recommandation, l'UIT avait été avisée de l'existence d'une propriété intellectuelle protégée par des brevets à acquérir pour mettre en œuvre la présente Recommandation. 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	1
1.1	1
1.2	2
1.2.1	2
1.2.2	2
1.2.3	4
1.2.4	4
1.2.5	4
2	4
3	5
3.1	5
3.2	7
4	8
5	9
5.1	9
5.2	10
5.3	11
6	12
6.1	12
6.1.1	12
6.1.2	12
6.1.3	16
6.2	16
6.2.1	16
6.2.2	18
6.2.3	19
6.2.4	20
6.2.5	20
6.2.6	22
6.2.7	22
6.2.8	23
6.3	24
7	24

	Page
7.1	Acquisition de trames audio pour les mesures 24
7.1.1	Description des trames audio 24
7.1.2	Conversion d'analogique en numérique 24
7.1.3	Affectation des marqueurs temporels..... 25
7.2	Mesurage du temps de transmission pour les trames audio 25
7.2.1	Généralités..... 26
7.2.2	Préparation du signal..... 27
7.2.3	Phase approchée 27
7.2.4	Phase précise 30
7.2.5	Combinaison des résultats des phases approchée et précise 33
8	Considérations relatives à la combinaison des mesures audio et vidéo 34
8.1	Mesures d'activité et de synchronisation des voies audio et vidéo..... 34
8.2	Association de mesures individuelles..... 34
9	Mesures sur les voies de données..... 35
9.1	Acquisition de trames de données pour les mesures 35
9.1.1	Considérations relatives à la définition des trames de données 35
9.1.2	Affectation des marqueurs temporels..... 36
9.2	Mesure du temps de transmission des données 36
9.2.1	Mise en correspondance de données structurées..... 36
9.2.2	Mise en correspondance de données non structurées..... 36
10	Prescriptions relatives à la stabilité des temporisateurs et à la synchronisation..... 37
10.1	Résolution..... 37
10.2	Précision et stabilité (erreur admissible sur la durée)..... 37
10.3	Erreur de réglage du temps..... 40
	Appendice I – Bibliographie..... 40
	Appendice II – Symboles mathématiques et conventions adoptées 41

Introduction

Les véritables systèmes de communications multimédias sont caractérisés par une capacité qui les distingue d'un simple bouquet de voies indépendantes: celle de maintenir la relation temporelle entre les différents médias.

La présente Recommandation spécifie les paramètres et les méthodes de mesure permettant d'évaluer la synchronisation relative entre voies médias, ainsi que deux autres paramètres importants de la qualité temporelle. Le temps de transmission (ou de propagation) dans une voie est un paramètre critique lorsque l'on évalue l'aptitude d'un système à transporter des conversations et d'autres signaux interactifs. La durée entre arrivées de trames et son inverse, le débit de trames, caractérisent la capacité d'un système à acheminer des informations de manière continue et cohérente.

Les systèmes multimédias actuels combinent des voies vidéo, des voies audio et des voies de données pour améliorer les communications. La présente Recommandation traite de tous ces médias. Le temps de transmission vidéo peut varier largement sur de courtes séquences. Les séquences audio et vidéo peuvent subir des distorsions en cours de transmission et les trains de données peuvent avoir une structure légère ou inexistante, assortie d'erreurs sur les bits. Bien que chaque média pose des problèmes de mesurage uniques, les méthodes ici spécifiées y répondent et les résolvent. La méthode fondée sur la variance concerne et mesure d'éventuelles variations instantanées du temps de transmission vidéo. La méthode de mesure du temps de transmission audio concerne des voies dans lesquelles la forme d'onde originale des paroles n'est pas préservée. Il existe également des méthodes applicables aux voies de données qui tirent parti des structures natives et qui tolèrent des erreurs sur les bits. Toutes ces méthodes impliquent des signaux d'essai qui sont représentatifs des applications prévues des systèmes.

Recommandation P.931

MESURE DU TEMPS DE TRANSMISSION, DE LA SYNCHRONISATION ET DU DEBIT DE TRAMES DANS LES COMMUNICATIONS MULTIMEDIAS

(Genève, 1998)

1 Domaine d'application et interfaces

1.1 Domaine d'application

La présente Recommandation traite des méthodes d'essai pour systèmes de transmission multimédia utilisant des ressources de transport numériques. Elle définit un ensemble de paramètres métrologiques caractérisant les aspects suivants de la performance d'un système:

- 1) la durée s'écoulant entre les arrivées de trames vidéo actives, qui est l'inverse du débit de trames élémentaire;
- 2) le temps de transmission dans une voie vidéo ou temps de transmission vidéo;
- 3) le temps de transmission dans une voie audio ou temps de transmission audio;
- 4) le temps de transmission dans une voie de données ou temps de transmission de données (ainsi que la durée entre arrivées de trames);
- 5) la synchronisation temporelle des voies.

Le domaine d'application des mesures est limité aux cas dans lesquels des interfaces appropriées sont présentes aux entrées et aux sorties des médias ou aux cas dans lesquels ces interfaces peuvent être offertes avec des capacités d'essai facultatives.

Les applications suivantes sont hors du domaine d'application de la présente Recommandation:

- 1) aspects métrologiques de la performance d'un système autres que le temps de transmission, la synchronisation et le débit de trames. Les mesures temporelles ne caractérisent pas complètement la qualité d'un système de transmission multimédia. Par exemple, le maintien de la qualité de reproduction des trames vidéo entre l'entrée et la sortie est également d'une importance évidente pour les utilisateurs. L'optimisation de cette performance subjective pour tous les paramètres qualitatifs peut avoir priorité sur l'optimisation des résultats des mesures paramétriques relevées conformément à la présente Recommandation;
- 2) choix illimité de sources à contenu utile et représentatif. Les méthodes de mesure spécifiées ici imposent des restrictions aux signaux des sources choisies pour ces essais. Les sources dont les séquences vidéo contiennent beaucoup d'activité cinétique provoquent souvent une augmentation des temps de transmission, une diminution du débit de trames et une dérive du synchronisme audiovisuel dans certaines applications multimédias. Les mesures doivent donc représenter des scènes expérimentales qui soient réalistes pour l'application du système multimédia évalué. D'autres limitations sont indiquées dans les paragraphes concernant chaque méthode de mesure;
- 3) mesure des niveaux de performance des systèmes lorsque les interfaces d'entrée et de sortie ne sont pas accessibles;
- 4) limites des paramètres. Ces limites sont hors du domaine d'application de la présente Recommandation pour le moment car celle-ci ne présente que les méthodes permettant de mesurer ces paramètres sans indiquer de valeurs d'évaluation.

1.2 Interfaces

1.2.1 Voies d'utilisateur à utilisateur

Théoriquement, le mesurage du temps de transmission sera effectué aux interfaces avec l'utilisateur, de façon à caractériser la totalité du temps de transmission d'utilisateur à utilisateur. La voie d'utilisateur à utilisateur complète commence et se termine par les dispositifs assurant l'interface avec les utilisateurs. Soit par exemple une voie vidéo avec sa caméra et ses dispositifs d'affichage, comme indiqué sur la Figure 1.

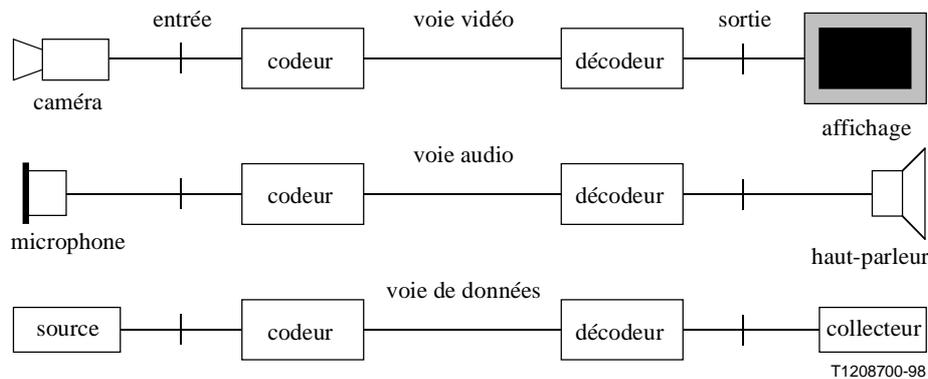


Figure 1/P.931 – Voies d'utilisateur à utilisateur dans un système multimédia

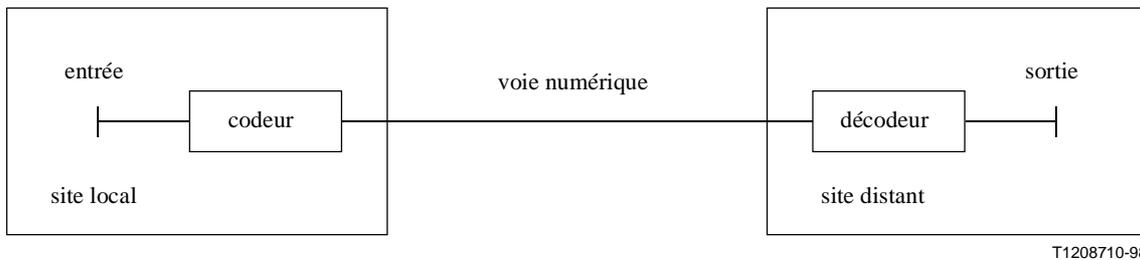
Malheureusement, les signaux entrent dans ces voies et en sortent sous forme de lumière, ce qui rend difficile la production et l'acquisition de signaux appropriés aux mesurages. Pour simplifier l'interconnexion de l'équipement de mesure avec la voie, l'on spécifie la voie numérique entre les connecteurs d'interface électrique entre la sortie de la caméra et l'entrée de l'affichage. Cela présente l'avantage de donner à l'interface d'essai une structure plus physique et plus logique. Le temps de transmission supplémentaire apporté par une caméra et par un affichage peut être évalué séparément (l'on s'attend que ce temps de transmission supplémentaire sera limité à l'intervalle échantillon/affichage, qu'il pourra être constant pour les affichages et qu'il sera habituellement indépendant du signal d'essai). Ce temps de transmission supplémentaire sera ajouté aux mesures de temps de transmission variable relevées conformément à la présente Recommandation.

Il est possible de distinguer des interfaces d'entrée et des interfaces de sortie dans les voies audio et de données d'utilisateur à utilisateur.

1.2.2 Configurations applicables

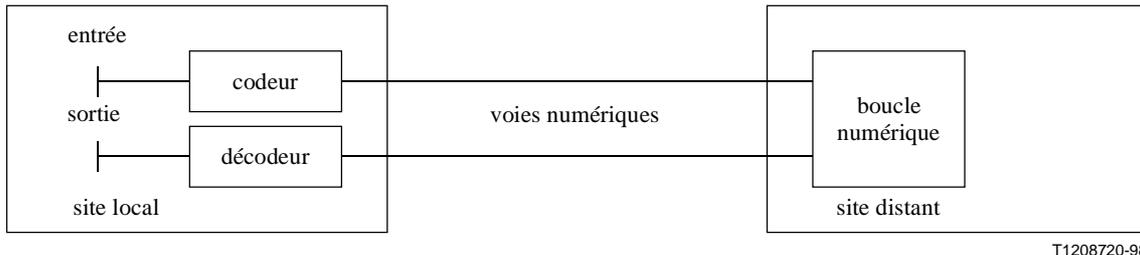
Les configurations de voie suivantes sont des applications appropriées de la présente Recommandation. Chaque figure montre les interfaces d'entrée et de sortie nécessaires.

Les Figures 2, 3 et 4 ne montrent, par souci de simplification, qu'un codeur, un décodeur et une voie numérique. Les éléments pouvant constituer la voie média dans ces essais ne sont pas strictement limités.



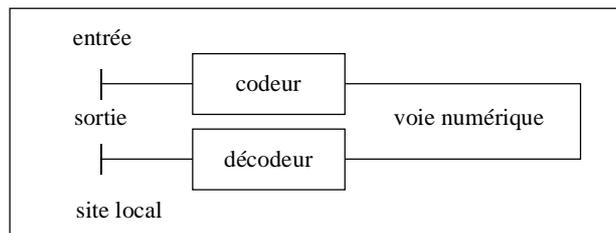
T1208710-98

Figure 2/P.931 – Mesure d'extrémité à extrémité¹



T1208720-98

Figure 3/P.931 – Mesure distante de boucle numérique

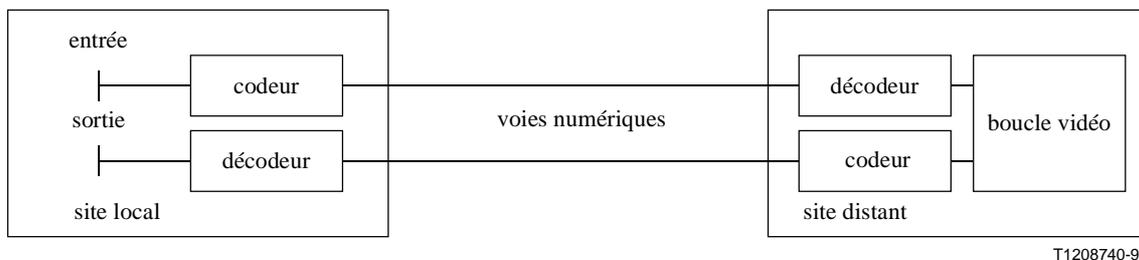


T1208730-98

Figure 4/P.931 – Mesure du système local

La Figure 5 montre une configuration de mesure sur voies vidéo avec application limitée. Il s'agira d'une mesure du temps de transmission dans les deux sens sur deux systèmes à un seul sens, au moyen d'un seul appareil de mesure.

¹ A noter que si la voie numérique contient des composants de traitement, il faut prendre des précautions lors de l'interprétation des résultats afin de tenir compte des effets d'interaction.



T1208740-98

Figure 5/P.931 – Mesure d'une boucle vidéo distante

La boucle vidéo n'est pas appropriée aux cas où les voies numériques ont un temps de transmission asymétrique. Dans les systèmes de téléconférence vidéo à débit de codage inférieur à 30 trames par seconde, le codage de la voie aller peut avoir une influence sur le temps de transmission de la voie de retour, ce qui rendrait l'essai de boucle inapproprié dans ce cas. Dans ces circonstances, la mesure dans les deux sens ne révélera pas l'évaluation dans un seul sens recherché.

NOTE – L'évaluation dans un seul sens d'un système symétrique est simplement égale à la moitié du temps de transmission dans les deux sens.

1.2.3 Interfaces applicables – Vidéo

Les travaux qui ont permis d'élaborer la présente Recommandation ont été initialement conduits avec utilisation d'interfaces composites pour signaux vidéo analogiques. Cependant, la conception des systèmes de vidéoconférence évolue rapidement pour passer d'un ensemble d'éléments (utilisant l'interface composite) à des systèmes plus intégrés. Par ailleurs, les demandes en produits vidéo de haute qualité excèdent les possibilités du signal analogique composite. Il sera nécessaire d'appliquer la présente Recommandation à des interfaces nouvelles afin de rester en phase avec le progrès technique. Les interfaces à composantes numériques, les interfaces RVB à contrôle informatique et les interfaces à caméra numérique sont des candidates probables.

Les paramètres de mesure définis ici nécessitent seulement la capacité de fournir des trames vidéo à l'entrée et de comparer les trames vidéo à la sortie d'une voie vidéo. Aucune technique n'impose à proprement parler d'interface composite.

Pour faciliter les mesures décrites dans la présente Recommandation sur des systèmes à intégration totale, il faudra des caractéristiques facultatives d'accès d'interface pour prendre en charge les essais. La complexité des systèmes individuels peut ne pas subir d'augmentation appréciable si presque toutes les fonctions additionnelles qui sont requises pour mettre en œuvre ces interfaces sont contenues dans le sous-système facultatif. Ces caractéristiques d'interface peuvent être utiles dans d'autres activités, comme le découplage des pannes et l'assurance de la qualité de fabrication. Elles devraient être recherchées par les constructeurs sur cette base.

1.2.4 Interfaces applicables – Audio

Cette spécification est applicable à toutes les interfaces audio normalisées.

1.2.5 Interfaces applicables – Données

Cette spécification est applicable à toutes les interfaces normalisées de données.

2 Références normatives

Les Recommandations UIT-T 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 UIT-T en vigueur est publiée régulièrement.

- Recommandation UIT-T P.84 (1993), *Méthode d'essai d'écoute subjective pour évaluer les équipements de multiplication de circuit numérique et les systèmes téléphoniques avec mise en paquets.*
- Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission.*
- Recommandation UIT-T P.861 (1998), *Mesure objective de la qualité des codecs vocaux fonctionnant en bande téléphonique (300-3400 Hz).*
- Recommandation UIT-T P.910 (1996), *Méthodes subjectives d'évaluation de la qualité vidéographique pour les applications multimédias.*
- Recommandation UIT-T P.920 (1996), *Méthodes d'essai interactives pour communications audiovisuelles.*
- Recommandation UIT-T P.930 (1996), *Principes d'un système de dégradation de référence pour signaux vidéo.*
- Recommandation UIT-T X.25 (1996), *Interface entre équipement terminal de traitement de données et équipement de terminaison de circuit de données pour terminaux fonctionnant en mode paquet et raccordés par circuit spécialisé à des réseaux publics pour données.*
- Recommandation UIT-R BT.470-6 (1998), *Systèmes de télévision classiques.*
- Recommandation UIT-R BT.500-9 (1998), *Méthodologie d'évaluation subjective de la qualité des images de télévision.*
- Recommandation UIT-R BT.601-5 (1995), *Paramètres de codage en studio de la télévision numérique pour des formats standards d'image 4:3 (normalisé) et 16:9 (écran panoramique).*

3 Termes et définitions

La présente Recommandation définit les termes suivants:

3.1 Termes généraux

3.1.1 système de communication multimédia: système qui gère plusieurs flux médias d'une manière synchronisée du point de vue de l'utilisateur. Un tel système permet l'interconnexion de correspondants multiples, l'établissement de connexions multiples et l'adjonction/suppression de ressources et d'utilisateurs à l'intérieur d'une même session de communication.

3.1.2 flux média: séquence d'unités de présentation destinée à acheminer un contenu spécifique donné.

3.1.3 niveaux hiérarchiques de codage: unités imbriquées de représentation de signal en lesquelles un flux média peut être décomposé.

3.1.4 niveaux hiérarchiques de contenu: unités informationnelles imbriquées en lesquelles un flux média peut être décomposé.

3.1.5 unité de présentation: plus petite subdivision pratique d'un flux média (défini par le système de mesure) qui transporte une unité de contenu indépendante et autonome, à partir des niveaux de hiérarchie de contenu présents dans le flux.

3.1.6 trame vidéo: unité de présentation de la voie vidéo. Niveau le plus bas de la hiérarchie de contenu dans un flux média vidéo, où les différences entre unités séquentielles apparaissent dans l'unité de présentation. La hiérarchie de contenu décrite dans la présente Recommandation peut au besoin réutiliser ces termes dans certaines hiérarchies de codage. Pour les systèmes à 525 lignes et à 625 lignes, une trame vidéo est définie comme étant composée d'un seul balayage (conformément à la Recommandation BT.470-6).

3.1.7 trame audio: unité de présentation de la voie audio. Groupe d'échantillons audio consécutifs. Le nombre préféré d'échantillons dans une trame audio dépend de la fréquence d'échantillonnage audio et est indiqué au paragraphe 5. Les trames audio n'ont pas de relation avec les trames construites par certains codecs audio/vocaux.

3.1.8 trame de données: unité de présentation de la voie de données. Groupe de bits de données consécutifs. Le nombre préféré de bits dans une trame de données dépend de l'application de la voie de données.

3.1.9 voie numérique: moyen d'acheminement de l'information d'un point à un autre sous forme numérique. Une voie numérique peut être implémentée dans un réseau composé d'éléments de communication numériques.

3.1.10 voie vidéo: moyen d'acheminement de trames vidéo d'un point à un autre. Une séquence de trames injectée à l'entrée de cette voie correspond à une séquence similaire (quoique non nécessairement identique) de trames à la sortie de cette voie. La voie vidéo peut comporter les éléments suivants: dispositifs de conversion de format vidéo, codeurs (compresseurs) et décodeurs (décompresseurs), tampons de régularisation du débit, multiplexeurs et démultiplexeurs, modulateurs et démodulateurs, ressources de transmission, commutateurs, ponts de conférence et autres composants nécessaires pour obtenir les caractéristiques de voie souhaitées.

3.1.11 voie audio: moyen d'acheminement de signaux audio d'un point à un autre. Une forme d'onde audio injectée à l'entrée de cette voie correspond à une forme d'onde semblable (quoique non nécessairement identique) à la sortie de cette voie. La voie audio peut comporter les éléments suivants: codeurs (compresseurs) et décodeurs (décompresseurs), tampons, multiplexeurs et démultiplexeurs, modulateurs et démodulateurs, ressources de transmission, commutateurs, ponts de conférence et autres composants nécessaires pour obtenir les caractéristiques de voie souhaitées.

3.1.12 voie de données: moyen d'acheminement de données d'un point à un autre. Une séquence de bits de données injectée à l'entrée de cette voie correspond à une séquence similaire (quoique non nécessairement identique) à la sortie de cette voie. La voie de données peut comporter les éléments suivants: dispositifs de conversion de format, codeurs (compresseurs) et décodeurs (décompresseurs), tampons, dispositifs de segmentation et de réassemblage du flux, multiplexeurs et démultiplexeurs, modulateurs et démodulateurs, ressources de transmission, commutateurs, ponts de conférence et autres composants nécessaires pour obtenir les caractéristiques de voie souhaitées.

3.1.13 intégrité des balayages: attribut d'une voie vidéo qui est présent lorsque le contenu des balayages impairs (pairs) dans une séquence de source au format de 525 ou de 625 lignes est acheminé dans les balayages impairs (pairs) à la sortie.

3.1.14 trame vidéo répétée: trame vidéo de sortie qui ne peut pas être distinguée de la ou des trames qui la précèdent dans la séquence (lorsque les trames de la séquence d'entrée correspondante possèdent des différences perceptibles). Une trame répétée est censée être produite à un point intermédiaire de la voie vidéo. Comme les trames répétées n'ont pas parcouru la voie de l'entrée jusqu'à la sortie, elles ne sont pas utilisées pour le calcul de la répartition des temps de transmission

dans la voie vidéo. Les trames répétées n'acheminent pas non plus de nouveaux stimuli visuels et sont exclues du calcul de la durée entre arrivées de trames (et donc du débit de trames élémentaire).

3.1.15 trame vidéo non répétée (trame active): une trame vidéo de sortie qui est différenciable de la ou des trames qui la précèdent dans la séquence (lorsque les trames de la séquence d'entrée correspondante possèdent des différences perceptibles). Une trame active est censée avoir parcouru la voie de l'entrée jusqu'à la sortie et son temps de transmission peut être inclus dans le calcul de la répartition des temps de transmission dans la voie vidéo. Comme les trames actives acheminent de nouveaux stimuli visuels, elles sont à la base du calcul de la durée entre arrivées de trames (et donc du débit de trames élémentaire).

NOTE – Toute trame interpolée issue d'un décodeur sera interprétée comme une trame active dans ce processus.

3.1.16 correspondance des trames: processus de comparaison d'une séquence de trames avec une autre séquence de trames afin de déterminer la correspondance entre trames dans chaque séquence ainsi que la correspondance de trames individuelles.

NOTE – Un moyen permettant de contrôler la correspondance entre deux trames vidéo consiste à comparer leurs représentations numériques au niveau des pixels puis à calculer le carré de la somme des différences par rapport à la moyenne sur tous les pixels (méthode dite de la variance).

3.1.17 identification des trames vidéo répétées: processus consistant à comparer chaque trame vidéo de sortie à la ou aux trames précédentes dans la séquence et à quantifier le degré de correspondance entre chaque paire de trames. Lorsque la correspondance entre une paire de trames est élevée (les seules différences pouvant être attribuées au bruit de mesure), la paire n'est pas différenciable. Lorsque la séquence de trames d'entrée correspondante est différenciable, la trame en cause est dite répétée. Voir la Note du 3.1.16.

3.1.18 identification des trames vidéo actives: processus consistant à comparer chaque trame vidéo de sortie à la ou aux trames précédentes dans la séquence et à quantifier le degré de correspondance entre chaque paire de trames. Lorsque la correspondance entre une paire de trames est limitée (au point que les différences mesurées puissent être différenciées du bruit de mesure) et que la séquence de trames d'entrée correspondante est différenciable, la trame en cause est dite active. Voir la Note du 3.1.16.

3.2 Paramètres mesurables

Le présent sous-paragraphe contient les définitions de haut niveau des paramètres de mesure essentiels. Il contient également des définitions d'ordre méthodologique.

3.2.1 temps de transmission: temps pris par une trame particulière pour parcourir une voie de transmission. Ce temps est calculé par enregistrement des instants auxquels les trames sont injectées dans la voie, puis par détection d'une trame de sortie qui a parcouru la voie et enregistrement de son instant d'arrivée en sortie. Cette trame de sortie est ensuite mise en correspondance univoque avec une trame d'entrée. Le temps de transmission est alors égal au temps d'arrivée moins le temps d'entrée.

NOTE – Lorsqu'il y a peu ou pas d'activité dans la voie, les méthodes décrites ici poseront des difficultés pour effectuer des mesurages valides. Ce paramètre perd son importance après l'affichage de la première trame, suivie de trames toutes identiques (absence d'activité ou vidéo statique).

3.2.2 distribution des temps de transmission de flux médias: ensemble des temps de transmission calculés pour une séquence de trames de sortie, exprimé de façon que toute variation entre mesures individuelles soit bien mise en évidence. Des statistiques de récapitulation classiques peuvent également être fournies, le cas échéant.

3.2.3 durée inter-frames actives à l'arrivée: temps écoulé entre trames actives successives à la sortie de la voie. Ce temps est calculé par sélection d'une trame active (ou d'une trame non répétée qui a parcouru la voie) et par enregistrement de son instant d'arrivée à la sortie. La plus récente (précédente) trame active doit ensuite être détectée et son instant d'arrivée est enregistré. La durée entre arrivées de trames actives dans la voie est alors égale au temps d'arrivée de la trame actuelle moins le temps d'arrivée précédent. Voir la Note du 3.2.1.

3.2.4 distribution des durées entre arrivées de trames actives: ensemble des durées entre arrivées qui a été calculé pour une séquence de trames de sortie actives, exprimé de manière que toute variation entre mesures individuelles soit bien mise en évidence. Des statistiques de récapitulation classiques peuvent également être fournies, le cas échéant.

3.2.5 débit (fréquence) de trames élémentaire: inverse de la durée entre arrivées de trames actives pour la trame active présente. Le débit de trames élémentaire est égal à 1 divisé par la différence entre les temps d'arrivée des trames actives présente et précédente.

3.2.6 données statistiques du débit de trames: ensemble de valeurs statistiques calculé pour une séquence de trames actives de sortie et exprimé de façon que toute variation soit bien mise en évidence. Lorsque des statistiques récapitulatives du débit de trames sont signalées, leur calcul doit utiliser la répartition des durées entre arrivées et l'inverse de ces durées.

3.2.7 rapport élémentaire de décalage de trames: quotient entre la durée inter-frames à l'entrée et la durée inter-frames à la sortie (ou rapport des débits de trame élémentaires). Mesure de la variation entre les débits des trames actives d'entrée et de sortie.

4 Abréviations

La présente Recommandation utilise les abréviations suivantes:

ANSI	Institut national américain de normalisation (<i>American National Standards Institute</i>)
ATM	mode de transfert asynchrone (<i>asynchronous transfer mode</i>)
cc	courant continu
DFT	transformée de Fourier discrète (<i>discrete Fourier transform</i>)
EAV	fin de ligne vidéo active (<i>end of active video</i>)
FFT	transformée de Fourier rapide (<i>fast Fourier transform</i>)
fps	trames par seconde (<i>frames per second</i>)
GPS	système mondial de localisation (<i>global positioning system</i>)
IIR	réponse impulsionnelle infinie (<i>infinite impulse response</i>)
MSE	erreur quadratique moyenne; variance (<i>mean square error</i>)
MTIE	erreur maximale sur la durée (<i>maximum time interval error</i>)
PSD	densité spectrale de puissance; puissance spectrique (<i>power spectral density</i>)
PSNR	crête du rapport signal sur bruit (<i>peak signal-to-noise ratio</i>)
RMS	écart type (<i>root mean square</i>)
RVB	rouge, vert, bleu
SAV	début de ligne vidéo active (<i>start of active video</i>)

SMPTE	Société des ingénieurs en images animées et télévision (<i>society of motion picture and television engineers</i>)
TIE	erreur sur la durée (<i>time interval error</i>)

5 Calculs chronologiques pour médias de communications générales

Le présent paragraphe présente un modèle général pour le calcul des paramètres contenus dans la présente spécification. Il applique ce modèle à chaque support traité. La Figure 6 montre le modèle général. Deux séquences d'**unités de présentation**, P et P', pénètrent dans l'interface d'entrée de la voie et quittent l'interface de sortie de la voie. Lorsque la dernière partie de chaque unité de présentation traverse l'interface, le système de mesure relève un temporisateur T et associe la valeur T(n) à l'unité de présentation entrante n, P(n). A l'interface de sortie, le système de mesure relève un temporisateur T' et associe la valeur T'(m) à l'unité de présentation sortante, P'(m).

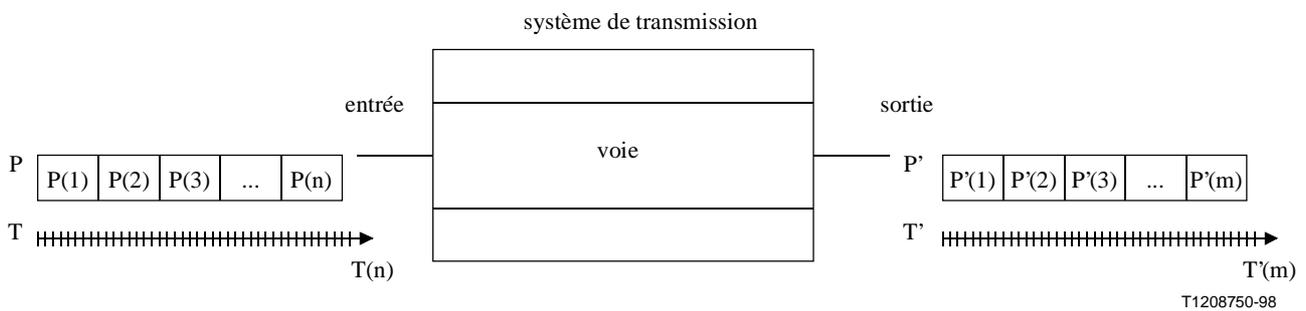


Figure 6/P.931 – Modèle général pour les unités de présentation

On notera que les temporisateurs T et T' peuvent constituer un temporisateur unique dans certaines configurations de mesure (par exemple les configurations locales et de boucle décrites par les Figures 3, 4 et 5). Si ce n'est pas le cas (par exemple dans les configurations d'extrémité à extrémité comme dans la Figure 2), les temporisateurs doivent être synchronisés. De même, un seul temporisateur fournit des marqueurs temporels quasi-simultanés pour toutes les voies d'une interface donnée, ce qui permet d'effectuer des calculs directs entre voies et à l'intérieur de celles-ci.

Le début et la fin des unités de présentation ne coïncident pas toujours, comme indiqué sur la Figure 6. Des marqueurs temporels doivent être associés à la fin d'une unité de présentation afin de régler les problèmes aussi bien théoriques que pratiques. Les marqueurs temporels initiaux peuvent être mémorisés séparément en vue de mesures additionnelles.

On notera que les unités P et P' sont représentées ci-dessus comme des flux continus. Mais des arrivées non périodiques sont possibles avec de nombreux médias. En fait, des unités de présentation complètes peuvent nécessiter plus de temps pour sortir du système que pour y entrer. C'est une des raisons d'affecter des marqueurs temporels à la fin des unités de présentation. Une autre raison est d'ordre pratique car la fin d'une unité de présentation est souvent plus facile à prévoir que son début.

5.1 Calculs pour une voie unique

Le temps de transmission dans une voie pour l'unité de présentation P'(m) est, après détermination de la correspondance de P'(m) avec P(n):

$$t_p(m) = T'(m) - T(n)$$

La correspondance des trames est normalement triviale dans une voie sans pertes. Mais les méthodes ici définies traiteront aussi bien de la distorsion que de la perte totale d'unités de présentation. Si les unités de présentation qui se succèdent à l'entrée ne sont pas différentes, les résultats des mises en correspondance seront ambigus.

La durée entre arrivées de l'unité de présentation $P'(m)$ en sortie de voie est donnée par la relation:

$$b'_p(m) = T'(m) - T'(m-1)$$

où $m-1$ est l'indice de l'unité de présentation précédente.

Le débit de trames élémentaire pour l'unité $P'(m)$ en sortie de voie est:

$$f'_p(m) = \frac{1}{T'(m) - T'(m-1)}$$

Lorsque les systèmes ignorent systématiquement des unités de présentation (comme c'est le cas dans les voies numériques vidéo à faible débit), les débits de trame diffèrent à l'entrée et à la sortie. Le rapport élémentaire de décalage de trames est:

$$\frac{b'_p(m)}{b_p(n)} = \frac{f_p(n)}{f'_p(m)}$$

où $b_p(n) = T(n) - T(n-1)$

et où $P'(m)$ correspond à $P(n)$

Des ensembles complets de mesures individuelles de temps de propagation et de durée entre arrivées, représentés par:

$$t_p = \{t_p(1), t_p(2), t_p(3), t_p(4), \dots, t_p(M)\}$$

et par

$$b'_p = \{b'_p(2), b'_p(3), b'_p(4), \dots, b'_p(M)\}$$

peuvent être collectés pour analyse statistique et graphique.

5.2 Trames de flux médias

Dans la voie vidéo, les unités de présentation sont des trames vidéo. Le paragraphe 6 définit ces unités pour chaque interface vidéo sous la forme d'un simple balayage au format de 525 lignes. V et V' sont par exemple les variables pour les unités des flux vidéo d'entrée et de sortie.

Dans la voie audio, les unités de présentation sont des trames audio. Celles-ci sont un groupe d'échantillons numérisés qui représentent le flux audio (voir le Tableau 1). Le paragraphe 7 définit ces trames pour les interfaces audio. A et A' sont les variables pour les unités des flux audio d'entrée et de sortie. La longueur recommandée des trames audio est à peu près de la même durée que la **trame vidéo** associée, si elle est présente (par exemple 16,66 ... ms pour le format 525 lignes).

Dans une voie de données, les unités de présentation sont des trames de données. Les unités de données sont des mots d'essai uniques ou des groupes d'éléments binaires représentant une certaine unité de présentation pour l'application de l'utilisateur du flux de données. Le paragraphe 9 définit ces unités pour les interfaces de données. D et D' sont les variables pour les unités des flux d'entrée et de sortie de données.

Tableau 1/P.931 – Longueur préférée des trames audio

Fréquence d'échantillon	longueur préférée d'échantillon et durée
8 000	128, 16 ms
16 000	256, 16 ms
32 000	512, 16 ms
44,1 k	512, 11,61 ms
48 k	512, 10,66 ms

5.3 Calculs de synchronisation

A titre d'exemple de calcul chronologique des paramètres de synchronisation, considérons les flux médias dans lesquels:

- les variables $A(m)$ et $V(n)$ sont associées, soit par construction de l'appareil de mesure soit par connaissance *a priori* de l'association;
- les variables $V'(q)$ et $V(n)$ sont mises en correspondance au moyen des méthodes décrites dans la présente Recommandation;
- les variables $A'(p)$ et $A(m)$ sont mises en correspondance au moyen des méthodes décrites dans la présente Recommandation.

Le décalage temporel entre trames audio et vidéo associées est, à l'entrée:

$$O_{AV}(m,n) = T_A(m) - T_V(n)$$

Ce paramètre indique la position temporelle de la trame audio par rapport à la trame vidéo.

Le décalage temporel entre trames audio et vidéo associées est, à la sortie:

$$O'_{AV}(p,q) = T'_A(p) - T'_V(q)$$

L'on notera que $O'_{AV}(p,q)$ est la relation de synchronisation qui est perçue par un utilisateur en sortie de voie.

La dérive temporelle entre trames audio et vidéo associées, due à la voie de transmission du système, est, à la sortie:

$$S_{AV}(p,q) = O'_{AV}(p,q) - O_{AV}(m,n)$$

Conformément à la convention des écarts temporels positifs, un retard de synchronisation a une valeur positive. Si la deuxième voie est en avance, le paramètre a une valeur négative.

Pour les systèmes fonctionnant à faible débit de trames, une dérive temporelle différente de zéro peut assurer une synchronisation des lèvres perçue comme plus précise. Dans les systèmes à retard vidéo variable, il est habituellement non souhaitable de faire varier le retard audio afin de conserver une dérive nulle car une telle manipulation peut rendre inintelligibles les données audio.

Si un système de transmission possède la capacité de voies audio multiples, ces calculs sont appropriés afin d'évaluer la synchronisation entre voies audio.

6 Mesures vidéo

6.1 Acquisition de trames vidéo pour les mesures

Le présent paragraphe décrit la hiérarchie des éléments qui peuvent exister dans une séquence vidéo. Il spécifie le niveau de chaque élément hiérarchique du codage vidéo qui servira d'unité fondamentale pour les comparaisons. La présente Recommandation définit ces unités fondamentales comme étant des **trames vidéo**, avec au besoin une redéfinition des termes de la hiérarchie de codage. Ce sujet sera d'abord traité au niveau théorique afin de préparer l'extension des méthodes au-delà des interfaces électriques spécifiques qui seront traitées ensuite.

6.1.1 Description des trames vidéo

De nombreux niveaux de granularité peuvent être présents dans une séquence vidéo. Les informations contenues dans les séquences vidéo peuvent se répartir comme indiqué sur la Figure 7:

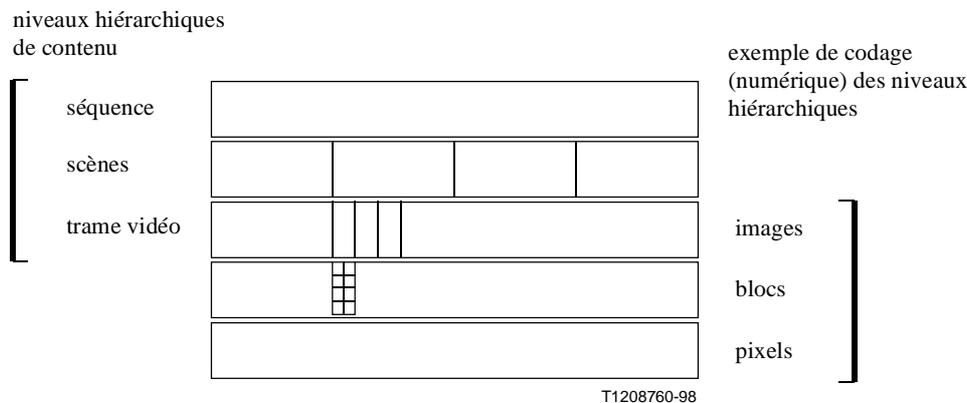


Figure 7/P.931 – Exemple de hiérarchie dans une séquence vidéo

La Figure 7 montre une séquence vidéo complète, composée de quatre scènes à contenu varié. Chaque scène se compose de nombreuses images ou trames complètes qui diffèrent afin de traduire un mouvement ou un changement. Les images constituent le plus bas niveau autonome de cet exemple de hiérarchie de contenu dans le domaine temporel, en ce sens que les plus petits changements de contenu interviennent à ce niveau dans le temps. Ce niveau est l'unité de présentation fondamentale pour l'utilisateur. L'on définira les **trames vidéo** à ce niveau, où les différences entre unités séquentielles apparaissent à l'intérieur de l'unité de présentation.

Les trames vidéo peuvent également occuper le niveau le plus élevé de la hiérarchie de codage vidéo. La Figure 7 illustre ce cas en montrant un exemple de hiérarchie de codage numérique. Si cependant il est possible que le codage dépende de groupes d'images, les hiérarchies de contenu et de codage peuvent se superposer sur plusieurs niveaux. Les trames vidéo continueront à être définies au plus bas niveau de contenu.

6.1.2 Trames vidéo aux interfaces composites

La Figure 8 montre la hiérarchie de séquence vidéo pour des signaux au format de 525 lignes et au format de 625 lignes.

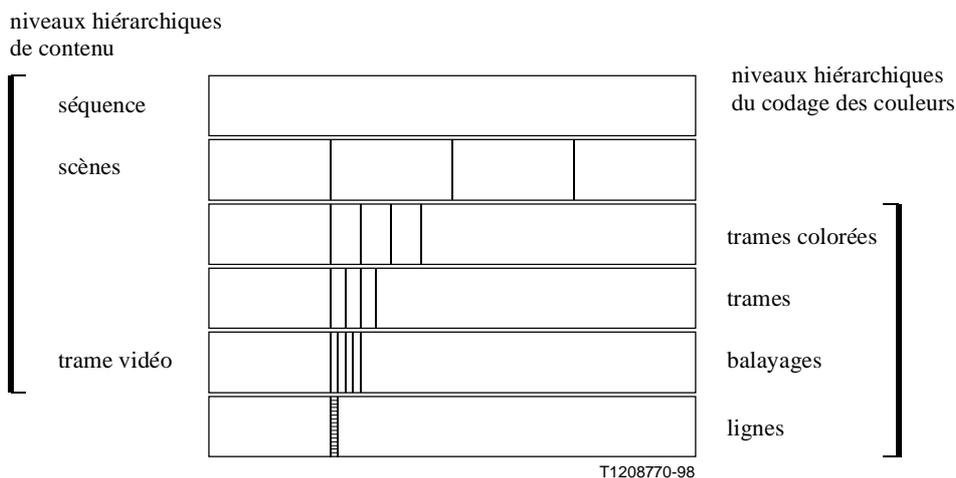


Figure 8/P.931 – Hiérarchie de séquence vidéo à 525 ou à 625 lignes

Pour les formats à 525 ou 625 lignes entrelacées, une **trame vidéo** est définie comme étant constituée d'un seul balayage, celui-ci étant spécifié dans la BT.470-6. Des balayages successifs peuvent acheminer de nouvelles informations jusqu'à l'utilisateur au moyen de l'unité de présentation, mais sans utiliser toutes les lignes de celle-ci. De même, étant donné que la fréquence de balayage est supérieure à la fréquence des trames complètes, cette définition permet un échantillonnage plus précis de la séquence vidéo.

Cette définition est également nécessaire pour tenir compte des systèmes de transmission qui (en raison de restrictions inhérentes concernant par exemple le débit de transmission) acheminent les trames vidéo actives à des débits inférieurs au débit de trames de l'interface vidéo et qui affichent chaque nouvelle trame active dès que possible. Cette méthode de rafraîchissement de l'affichage peut se traduire par de notables différences entre balayages successifs.

6.1.2.1 Conversion d'analogique en numérique

Les méthodes de mesure décrites dans les paragraphes suivants nécessitent une numérisation du signal composite analogique. La Recommandation UIT-R BT.601-5 propose une méthode à haute fiabilité pour échantillonner la zone analogique des lignes actives des signaux de luminance à 525 ou 625 lignes. L'expérience a montré que le signal de luminance fournit suffisamment d'informations pour comparer et faire correspondre des trames vidéo afin d'effectuer les mesures décrites dans la présente Recommandation.

La Figure 9 montre comment l'on peut organiser les échantillons numérisés qui correspondent aux trames vidéo.

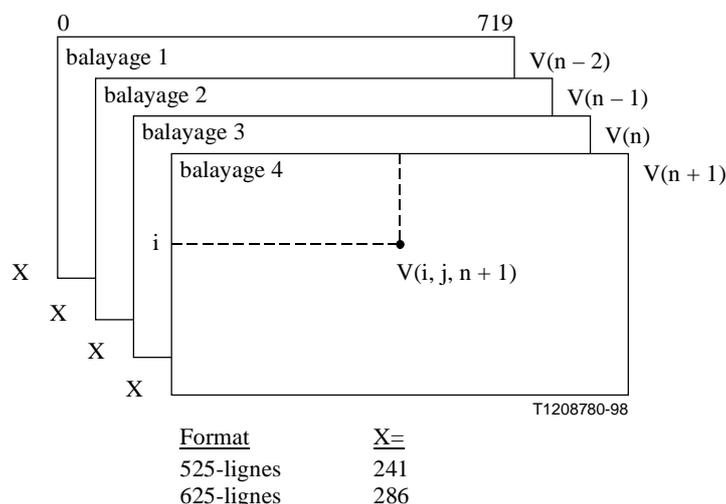


Figure 9/P.931 – Coordonnées recommandées pour la numérisation des trames vidéo

La variable $V(n)$ représente une trame vidéo au temps $T(n)$. $V(n - 1)$ est la trame vidéo précédente et $V(n + 1)$ la trame vidéo suivante. $V(i,j,n)$ est le (i,j) ^{ème} pixel de luminance dans la trame vidéo n au temps $T(n)$.

D'autres systèmes de coordonnées sont possibles mais il y a lieu d'utiliser la numérotation allant du coin supérieur gauche au coin inférieur droit afin de faciliter les comparaisons entre différentes implémentations métrologiques. L'implémentation de l'interface numérique à composantes au 6.1.3 facilitera également ces comparaisons. Exemple d'autres systèmes de coordonnées: des numéros de ligne spécifiques peuvent être attribués en partant des lignes de synchronisation verticale [10].

6.1.2.2 Affectation des marqueurs temporels

Dans les systèmes à 525 lignes, le temps $T(n)$ associé à la trame vidéo n doit être relevé immédiatement après la numérisation du dernier pixel de la trame (coordonnées 242, 719) et avant le début de la ligne suivante (10,222 μ s).

Dans les systèmes à 625 lignes, le temps $T(n)$ associé à la trame vidéo n doit être relevé immédiatement après la numérisation du dernier pixel de la trame (coordonnées 286, 719) et avant le début de la ligne suivante (10,666 μ s).

6.1.2.3 Alignement du gain, de la zone active et des coordonnées spatiales

Pour assurer la précision, la présente Recommandation nécessite l'application, avant le mesurage, d'un facteur de correction de la variation du gain (g), de la différence de niveau (l), du décalage horizontal (h) et du décalage vertical (v).

La référence bibliographique [3] définit une méthode permettant de mesurer la variation moyenne du gain et la différence moyenne des niveaux. Elle définit également des méthodes manuelles et automatiques pour mesurer la zone vidéo active et le décalage vidéo actif. Ces méthodes peuvent être utilisées afin de garantir la qualité des comparaisons de trames et des mises en correspondance de trames. Lorsqu'elles sont mesurées et utilisées, les coordonnées de la zone vidéo active et du décalage vidéo actif doivent être spécifiées.

Si le système de transmission redimensionne les trames d'entrée (c'est-à-dire s'il élargit ou rétrécit l'échelle dans le sens horizontal ou dans le sens vertical ou dans les deux sens), de façon que de nouvelles différences soient présentes lors de la comparaison des séquences d'entrée et de sortie, il est parfois souhaitable d'utiliser un facteur de correction afin de tenir compte de la variation

dimensionnelle (z) et de minimiser les différences avant la mise en correspondance des trames. Lorsque le redimensionnement est délibéré et étendu, il est essentiel de le compenser. Il reste à déterminer si un redimensionnement notable prévaut dans les systèmes de communication multimédia et, dans ce cas, dans quelle mesure le redimensionnement apporte un bruit mesurable. Il est toutefois possible de compenser le redimensionnement et de poursuivre les processus spécifiés ici.

Des erreurs de mesurage peuvent apparaître si ces méthodes ne sont pas employées. Par exemple, si la voie présente un décalage vidéo de 4 pixels, un panoramique de caméra dans le même sens se traduirait par une correspondance incorrecte des trames et par des temps de transmission imprécis. S'ils n'étaient pas compensés, les décalages et les variations de gain ou de niveau augmenteraient également le niveau de bruit tout en réduisant l'efficacité de la mise en correspondance des trames par la méthode de la variance.

Il y a des circonstances où des méthodes simplifiées fournissent des mesures précises. Le sous-paragraphe suivant indique les cas où des simplifications métrologiques sont possibles.

6.1.2.4 Options admises

Une option admise dans la présente Recommandation consiste à définir une zone de sous-trame par ses coordonnées x-y et à effectuer toutes les opérations de comparaison et de correspondance sur cette sous-trame. En variante, la sous-trame peut être définie par la zone d'action sûre et/ou par la zone de titre sûr décrite [7]. Cela permet d'exclure des mesures les limites de trame sans déterminer exactement la zone vidéo active ainsi que d'améliorer le seuil de capture de bruit en évitant les erreurs en limite de trame. Si l'une de ces options est adoptée, les coordonnées de la sous-trame doivent être spécifiées.

Lorsque l'on définit une zone de sous-trame, les éléments suivants doivent être pris en considération:

- 1) il faut éviter les bordures qui ne contiennent pas d'image;
- 2) il faut inclure les zones statiques possédant des bordures horizontales et verticales ainsi qu'une gamme complète de valeurs de pixel si possible, afin de faciliter la détermination des facteurs de correction des coordonnées spatiales;
- 3) il faut inclure les zones comportant des mouvements sinon les mesures de temps de transmission donneront des résultats ambigus;
- 4) les dimensions de la zone de sous-trame, mesurées en pixels, sont proportionnelles au temps de calcul. Par ailleurs, des zones de petites dimensions peuvent améliorer l'opération de mise en correspondance des trames lorsque des mouvements sont repérés, mais ces dimensions doivent être assez grandes pour éviter toute ambiguïté.

Une autre option consiste à utiliser une plus faible fréquence d'échantillonnage horizontal lors de la numérisation d'enregistrement. L'expérience a montré qu'une définition horizontale de 320 pixels dans la période de ligne numérique active pouvait autoriser un calcul de qualité suffisante pour ces mesures, sans qu'il soit nécessaire d'apporter des corrections d'alignement spatial. Si cette option est adoptée, les pixels horizontaux contenus dans la période de ligne numérique active doivent être spécifiés.

Dans les systèmes de mesure où l'on sait que la fréquence d'image est inférieure à 30 images par seconde, il est possible de mémoriser un balayage sur deux afin de diminuer la capacité nécessaire aux acquisitions, ce balayage pouvant être pair ou impair. Un exemple de ces systèmes est donné par les terminaux compatibles avec la H.320, avec codage vidéo conforme à la H.261. Si cette option est adoptée, le balayage sélectionné doit être spécifié.

6.1.3 Trames vidéo aux interfaces numériques en composantes

Lorsque l'on traite des trames vidéo en composantes, l'on définit de nouveau une trame vidéo comme étant constituée d'un seul balayage au format 525 ou 625 lignes.

6.1.3.1 Organisation du signal

Les signaux vidéo aux interfaces numériques (voir [8] pour les interfaces bits-parallèles et [9] pour les interfaces bits-séries) ont déjà été convertis à partir de leur forme analogique et aucune nouvelle numérisation n'est requise ici. Ces interfaces multiplexent des échantillons de 8 ou 10 bits, extraits des composantes vidéo, dans l'ordre (C_B, Y, C_R, Y, C_B, Y, ...). Les systèmes de mesure peuvent n'utiliser que les échantillons de luminance (Y).

Les mots d'information de la séquence de synchronisation sont insérés dans le flux binaire afin d'identifier la ligne numérique active. Les lignes actives commencent après le mot de début de ligne vidéo active (SAV, *start of active video*) et se terminent après le mot de fin de ligne vidéo active (EAV, *end of active video*). Elles contiennent 720 échantillons de luminance. Les coordonnées recommandées des échantillons sont les mêmes que pour les interfaces avec les signaux composites.

6.1.3.2 Affectation des marqueurs temporels

Les mêmes prescriptions (voir 6.1.2.2) que pour les interfaces composites s'appliquent ici.

6.1.3.3 Alignement du gain, de la zone active et des coordonnées spatiales

Les mêmes prescriptions (voir 6.1.2.3) et options (voir 6.1.2.4) que pour les interfaces composites s'appliquent ici.

6.2 Méthodes de mesure utilisant la variance des trames vidéo

Le présent sous-paragraphe décrit les méthodes de mesure pour un système utilisant une approche par la variance. Les implémentations de ces méthodes doivent pouvoir fournir des séquences de trames vidéo appropriées à l'entrée de la voie. Ces méthodes nécessitent également une acquisition et, au besoin, une numérisation de la composante de luminance des séquences de trames vidéo aux interfaces avec la voie. La fourniture et l'acquisition doivent être effectuées conformément aux dispositions du 6.1 pour les interfaces utilisées. Une fois que les trames actives et leurs trames correspondantes dans la séquence d'entrée ont été trouvées, les calculs chronologiques souhaités doivent être effectués comme spécifié dans le paragraphe 5.

Ces méthodes tiennent compte de plusieurs circonstances spéciales, y compris le fait que les systèmes de transmission de haute qualité conservent l'intégrité des balayages entrelacés et utilisent des séquences vidéo de source extraites de films à 24 fps par le procédé de décimation à 3:2.

6.2.1 Généralités

La détection de trames actives dans une séquence de trames vidéo et le repérage de trames en correspondance entre séquences nécessitent une méthode de comparaison normalisée. Cette méthode compare les trames vidéo pixel par pixel et exprime la différence entre une paire de trames sous la forme de la variance calculée sur tous les pixels concernés. Pour une paire de trames (l'une provenant de la séquence d'entrée l'autre de la séquence de sortie), la variance (MSE, *mean square error*) sera donc donnée par la relation:

$$M[V'(m), V(n)] = \frac{1}{K_s} \sum_{j=J_{\min}}^{J_{\max}} \sum_{i=I_{\min}}^{I_{\max}} [V'(i, j, m) - V(i, j, n)]^2$$

où $V'(i,j,m)$ est la valeur du pixel i,j dans la trame de sortie au temps $T'(m)$, et où $V(i,j,n)$ est la valeur du pixel i,j dans la trame d'entrée au temps $T(n)$. K_s est le nombre total de pixels dans la sous-trame rectangulaire examinée, donné par:

$$K_s = (I_{\max} - I_{\min} + 1) \times (J_{\max} - J_{\min} + 1)$$

On notera que $V'(i,j,m)$ a subi une correction pour tenir compte de toute variation de gain, différence de niveaux, décalage horizontal, décalage vertical et redimensionnement spatial (si nécessaire) entre entrée et sortie (avec les facteurs de correction correspondants g , l , h , v , et z):

$$V'(i, j, m) = \frac{V^*(x + v, y + h, m) - 1}{g}$$

où $V^*(x,y,m)$ est le pixel de sortie avant application des facteurs de correction. Si la trame vidéo de sortie doit être redimensionnée pour correspondre à l'entrée, on a:

$$V'(i, j, m) = \frac{V^{**}(\hat{i} + v, \hat{j} + h, m) - 1}{g}$$

où $V^{**}(m) = f(V^*(m), z)$

et où $f(V^*(m), z)$ représente une fonction de redimensionnement.

Pour les comparaisons entre trames adjacentes à l'intérieur d'une séquence (par exemple pour détecter des trames actives à l'interface de sortie), $V(i,j,n)$ devient $V'(i,j,m - 1)$ dans l'équation ci-dessus pour la variance.

La variance (MSE) est un facteur important pour le calcul du rapport signal sur bruit de crête (PSNR, *peak signal-to-noise ratio*) comme en [3]:

$$PSNR = 20 \log_{10} \left[\frac{V_{crête}}{\sqrt{M[V'(m), V(n)]}} \right] \text{dB}$$

La Figure 10 illustre le processus de mesure à haut niveau pour la méthode MSE.

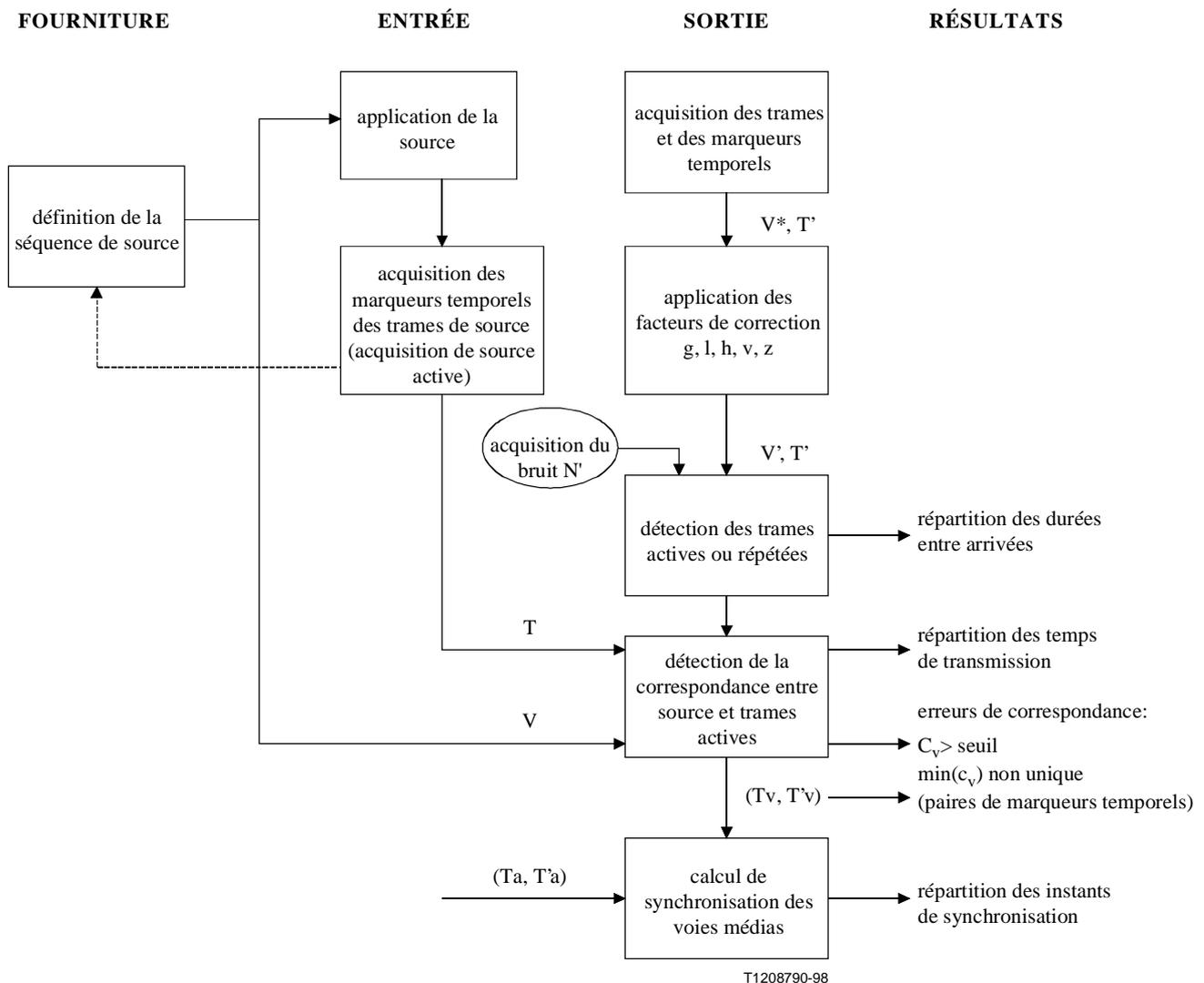


Figure 10/P.931 – Schéma fonctionnel pour les mesures vidéo par la méthode MSE

6.2.2 Etalonnage de la différence minimale détectable entre trames

Le présent sous-paragraphe spécifie la méthode permettant de déterminer le bruit (ou la variation parasite) dans les processus de numérisation et de mémorisation qui recueillent les séquences de trames vidéo afin de les comparer. Ce niveau de bruit dépend des options spécifiques qui ont été choisies (par exemple le format de numérisation). Il doit être connu pour que les mesures soient valides.

Les conditions d'essai pour étalonner le bruit d'acquisition sont les suivantes:

- 1) appliquer une scène **vidéo statique** à l'entrée de la voie. La vidéo statique est définie comme étant une "imagerie vidéo qui ne contient ni mouvement ni changement" [2]. Il importe de conserver le même rapport signal sur bruit vidéo à l'entrée pendant l'étalonnage et le mesurage. L'une des techniques fait appel à une séquence source composée d'une seule trame répétée qui est extraite d'une ou de plusieurs des séquences vidéo animées devant faire l'objet d'essais ultérieurs. Cette technique ne reproduit pas le bruit contenu dans la séquence source et ne convient que pour les cas où le bruit d'acquisition est nettement supérieur au bruit de source. Pour certaines séquences d'essai et trames vidéo en direct, il est possible de subdiviser spatialement la trame vidéo et de définir une sous-trame statique (par exemple l'arrière-plan) pour cet étalonnage ainsi qu'une sous-trame cinétique pour d'autres mesures.

Des signaux d'essai non cinétiques (barres chromatiques SMPTE) ont aussi été utilisés avec succès (là encore, le bruit de source présent dans le signal d'essai statique doit être le même que dans la séquence d'essai);

- 2) acquisition (numérisation et mémorisation) de la séquence correspondante de trames en sortie de voie. De 30 à 60 trames devraient suffire. Lorsque la voie fait appel à la compression numérique, il doit être possible d'obtenir un niveau de qualité stable sur la vidéo statique, ce qui évite toute **réponse en coupure de scène** qui dégraderait la mesure du bruit.

En général, le bruit acquis à l'entrée sera différent du bruit à la sortie. Certains codecs éliminent par filtrage le bruit de source pour améliorer le signal à coder.

Pour une séquence de 30 trames, l'on calcule l'ensemble de $30 - 1 = 29$ valeurs de variance sur trames adjacentes:

$$M[V'(m), V'(m - 1)].$$

Le niveau de bruit acquis en sortie est la variance maximale de l'ensemble ci-dessus.

$$v'(m) = M[V'(m), V'(m - 1)] \text{ pour } m = 2, 3, \dots, 30$$

où $v'(m)$ est la valeur de variance pour la trame $V'(m)$ et le bruit acquis, N' , est

$$N' = \max(v')$$

où v' est l'ensemble des valeurs de variance pour la séquence V' . La variation à l'intérieur de l'ensemble des valeurs de variance est en principe faible (< 20%) étant donné que la moyenne a été calculée sur un grand nombre de pixels pour chaque valeur de l'ensemble. Pour la séquence d'entrée, l'on a $N = \max(v)$.

Pour laisser une certaine marge entre le niveau de bruit acquis et un seuil de détection des trames actives, l'on détermine que les trames de sortie dont la variance est $v'(m) = M[V'(m), V'(m - 1)] \leq 1,5 \times N'$ sont des trames répétées. Pour une séquence de source, l'on détermine que les trames dont la variance est $v(n) = M[V(n), V(n - 1)] \leq 1,5 \times N$ sont des **trames indifférenciables**. Il peut exister de petites différences entre trames répétées ou indifférenciables mais le système de mesure ne peut pas les détecter de façon fiable. La sélection et la qualification des séquences de source pour les essais doivent tenir compte de ce seuil. L'on n'est pas censé détecter des trames actives lorsque les trames de la séquence de source sont indifférenciables par l'appareil de mesure. Cette marge va également dans le sens d'un plus grand degré de confiance dans la détection des trames actives.

6.2.3 Détection des trames différenciables dans une séquence

Dans une séquence vidéo V , l'on calcule l'ensemble des valeurs de variance v et l'on compare chaque élément de cet ensemble au seuil de différenciation des trames ($1,5 \times N$). Toutes les trames $V(n)$ dont la variance $M[V(n), V(n - 1)] > 1,5 \times N$ est différenciable des trames qui la précèdent. Une voie à l'essai doit être fournie avec des trames d'entrée différenciables afin d'y détecter des trames actives et des trames répétées.

Si l'on considère des séquences de source destinées à des systèmes de transmission de haute qualité conservant l'intégrité des balayages, il est plus approprié de comparer la **trame vidéo** actuelle, $V(n)$, à la trame $V(n - 2)$ afin d'apparier les balayages équivalents et d'éviter des erreurs de comparaison issues d'un décalage spatial entre balayages successifs.

La procédure suivante décrit des essais conditionnels permettant de vérifier qu'une trame vidéo possède des différences détectables (dans des séquences de source à balayages entrelacés).

- 1) Calculer la valeur $M[V(n), V(n - 1)]$.
- 2) Si le résultat est $\leq 1,5N$, déclarer la trame indifférenciable sinon continuer.
- 3) Calculer $M[V(n), V(n - 2)]$.
- 4) Si le résultat est $\leq 1,5N$, déclarer la trame indifférenciable sinon continuer.
- 5) La trame $V(n)$ possède des différences détectables.

Le 6.2.6 contient d'autres considérations relatives à la caractérisation des scènes de source ou d'entrée. Noter que des séquences créées au moyen d'un processus de décimation à 3:2 ne pourront pas faire l'objet de cette qualification mais pourront encore être utilisées selon les dispositions du 6.2.7.

6.2.4 Caractérisation des trames actives et des trames répétées

Dans une séquence vidéo de sortie, V' , l'on calcule l'ensemble des valeurs de variance $M[V'(m), V'(m - 1)]$ et l'on compare chaque valeur de cet ensemble au seuil de différenciation des trames ($1,5 \times N'$).

Noter que de nombreux systèmes de transmission de haute qualité conservent l'intégrité des balayages tout en introduisant une distorsion minimale. Pour ces systèmes, il est également approprié de comparer la **trame vidéo** actuelle, $V'(m)$ à la trame $V'(m - 2)$ afin d'apparier les balayages équivalents et d'éviter une erreur de comparaison due au décalage spatial entre balayages. Lors des essais à des interfaces non entrelacées ou utilisant les options reconnues pour le débit d'acquisition réduit et la basse résolution, la comparaison avec $V'(m - 2)$ est probablement inutile.

Une trame $V'(m)$ dont le calcul de variance donne $M[V'(m), V'(m - 1)]$ et $M[V'(m), V'(m - 2)] > 1,5 \times N'$, en réponse à une séquence d'entrée possédant des différences détectables possède une correspondance limitée avec ces deux trames et doit être caractérisée comme étant une **trame active**.

Une trame $V'(m)$ dont le calcul de variance donne $M[V'(m), V'(m - 1)]$ et $M[V'(m), V'(m - 2)] \leq 1,5 \times N'$, en réponse à une séquence d'entrée possédant des différences détectables possède une correspondance étroite avec $V'(m - 1)$ ou $V'(m - 2)$ et doit être caractérisée comme étant une **trame répétée**.

6.2.5 Contrôle de correspondance entre trames (mise en correspondance de trames)

Pour une trame active m et une séquence d'entrée de X trames, calculer l'ensemble des X valeurs de variance $M[V'(m), V]$. La trame d'entrée qui présente la meilleure correspondance est celle qui produit la valeur minimale de variance dans l'ensemble de:

$$c_v(x) = M[V'(m), V(x)] \text{ pour } 1 \leq x \leq X$$

où c_v est l'ensemble des valeurs de variance pour la trame $V'(m)$ en *comparaison* avec chaque trame de la séquence V . La trame d'entrée qui concorde le mieux avec la trame $V'(m)$ est définie comme étant

$$C_v = \min(c_v)$$

[l'erreur (variance MSE) minimale représente la correspondance maximale ou la meilleure concordance entre trames].

Un ensemble de règles peut améliorer le processus de mise en correspondance et diminuer l'ambiguïté. Il peut y avoir des cas où une seule trame active correspond de près à plusieurs trames d'entrée. De tels cas devraient être minimisés par des critères de correspondance fondés sur des méthodes de comparaison de pixels (variance). Mais certaines circonstances, comme les suivantes, augmentent la probabilité d'ambiguïté:

- distorsion spatiale extrême due à un faible débit de transmission, à l'utilisation d'un format de trame numérique à faible résolution, etc.;
- contenu de source à haut degré de mouvement (provoquant du maculage ou une autre distorsion), à mouvements répétitifs, à intervalles fixes dans une séquence;
- faible débit de trames actives de sortie, augmentant les probabilités de correspondance avec des trames de source;
- interpolation de trames, pouvant rendre la mise en correspondance plus difficile.

L'expérience a montré que l'implémentation de cette méthode pouvait résoudre ces exemples de circonstances difficiles.

Les règles suivantes peuvent être utiles pour faciliter l'automatisation de cette méthode et pour résoudre des correspondances ambiguës:

- 1) prescrire une correspondance biunivoque: une trame active et une seule peut correspondre à une trame d'entrée donnée. Une double correspondance peut s'expliquer par le fait qu'une trame active a été détectée par erreur. Si les résultats de la comparaison font apparaître une double correspondance, cette condition d'erreur apparente doit être signalée;
- 2) appliquer la séquence suivante par exemple (avec des trames non entrelacées): si la trame $V'(m)$ correspond à la trame $V(n)$, la prochaine trame active $V'(m + 2)$ doit correspondre à la trame $V(n + 1)$ ou $V(n + 2)$ ou $V(n + 3)$, etc. Il n'est pas permis que la trame $V'(m + 2)$ corresponde à la trame $V(n - 1)$ ou $V(n)$ mais un tel événement doit être étiqueté comme une erreur possible;
- 3) reconnaître un temps de transmission minimal: aucune correspondance n'est admise si le temps de transmission est inférieur à t_{\min} . Un temps négatif est exclu par la condition $t_{\min} \geq 0$;
- 4) reconnaître la situation de non-correspondance: certaines trames actives peuvent contenir trop de distorsion pour correspondre à la séquence émise. De telles trames doivent être comptées et signalées, ainsi que le seuil de non-correspondance utilisé. Les utilisateurs du système de mesure doivent déterminer l'étendue normale des valeurs de variance pouvant concorder dans le système de transmission en essai et doivent régler ce seuil au-dessus de cette étendue;
- 5) effectuer des diagnostics: le processus de mise en correspondance peut être répété à partir de l'extrémité opposée de la séquence afin de vérifier si moins de correspondances ambiguës ou moins de non-correspondances se produisent. Les règles peuvent aussi être appliquées en sens inverse;
- 6) contrôler la trame suivante: si la trame active suivante dans la séquence de sortie est en correspondance univoque avec la séquence émise, utiliser cette correspondance et appliquer les règles ci-dessus à la trame active précédente;
- 7) sélectionner aléatoirement la meilleure valeur: si une ambiguïté existe encore après l'application des règles ci-dessus, une sélection aléatoire peut être utilisée. Il est toutefois recommandé de calculer la variance avec une résolution suffisante afin de minimiser de telles occurrences. Les erreurs introduites dans la répartition par un processus aléatoire devraient s'annuler en fin de séquence et certaines statistiques récapitulatives resteront inchangées. De telles sélections doivent être comptées et signalées;
- 8) si les résultats obtenus avec une scène donnée tendent à nécessiter une intervention et une résolution poussées pour appliquer ces règles, il y a lieu de tenter d'effectuer le mesurage avec une scène différente.

6.2.6 Qualification de la séquence de source pour les méthodes de variance

Le succès des méthodes de variance dépend de l'utilisation de séquences de source appropriées. Comme indiqué ci-dessus, la séquence de source doit posséder des différences intertrames qui soient différenciables par l'appareil de mesure; elle ne doit pas comporter de mouvements répétitifs ni d'intervalles vidéostatiques qui provoqueraient certainement une ambiguïté de correspondance. Lors de l'utilisation d'une zone de sous-trame, le processus doit adopter une zone similaire comme base de qualification. La procédure suivante peut déterminer les séquences vidéo appropriées au mesurage:

- 1) sélectionner la première trame vidéo de la séquence de source et la comparer à toutes les autres trames de cette séquence;
- 2) enregistrer le nombre et la position de toutes les trames indifférenciables (comme décrit au 6.2.2);
- 3) effectuer l'analyse suivante: l'intervalle entre toutes les trames indifférenciables doit être suffisant pour résoudre les ambiguïtés d'alignement entre entrée et sortie au moyen d'estimations préalables de la durée entre arrivées de trames et d'autres informations. Par exemple, si l'on sait que le temps de transmission est inférieur à 2 s, des trames indifférenciables peuvent apparaître à un intervalle supérieur ou égal à 2 s;
- 4) répéter les étapes ci-dessus avec au moins les X premières trames (avec par exemple X = 60). L'on préfère les premières trames de la séquence parce que certains résultats des règles de comparaison peuvent dépendre des résultats antérieurs.

On notera également que:

des séquences de source comportant des trames adjacentes similaires (différenciables mais à un faible degré) auront une plus grande probabilité de produire des correspondances ambiguës avec des trames actives. Un contrôle des différences entre trames révélera l'étendue de ces analogies, comme décrit au 6.2.3.

La Bibliographie [1] spécifie un ensemble de séquences de source qui sont appropriées pour l'évaluation de systèmes de télévidéoconférence et de télévidéophonie. Certaines séquences contiennent toutefois quelques sections vidéostatiques rémanentes qui doivent être évitées pour le mesurage du temps de transmission (mais qui sont souhaitables pour l'étalonnage). D'autres séquences de scènes sont disponibles et peuvent convenir à d'autres applications. Même des séquences vidéo en direct sont admises, à condition que le système de mesure puisse acquérir et mémoriser aussi bien les séquences d'entrée que les séquences de sortie au cours d'un mesurage et que la séquence d'entrée soit ensuite évaluée quant à sa convenance. Toutes les séquences d'entrée doivent être évaluées au moyen de l'application des mesures spécifiques selon les procédures indiquées ici et au 6.2.3.

6.2.7 Considérations relatives à l'utilisation de séquences de source décimées à 3:2

Il existe plusieurs conditions qui peuvent se traduire par une duplication de trames (et donc à leur indiscernabilité) dans une séquence de source:

- 1) le contenu de la séquence est statique (sans mouvement ni changement);
- 2) un processus de décimation à 3:2 a créé la séquence vidéo à partir d'un film. Chaque cinquième balayage (dans un format de 525 lignes) est une répétition du balayage précédent (pair ou impair);
- 3) après un processus de décimation à 3:2, la séquence vidéo est éditée, ce qui produit des répétitions de balayages non périodiques.

Plus précisément, la présence de l'une quelconque de ces trois conditions provoquerait l'échec de la séquence de source aux essais de qualification. Il est impossible d'effectuer des mesurages de temps

de transmission ou de débit de trames avec une séquence statique. Certains mesurages sont cependant possibles avec une décimation à 3:2.

Deux problèmes sont rencontrés lors de l'application de la méthode de mesure par variance à des sources décimées à 3:2:

- le processus de mise en correspondance de trames actives avec leur contrepartie dans la séquence source est plus complexe si certaines trames de source sont répétées. Si des correspondances ambiguës apparaissent, les calculs de temps de transmission sont douteux pour cette trame. L'ambiguïté peut être résolue par examen de la séquence de source pour y détecter la présence de trames répétées;
- le débit de trames réel peut être supérieur à celui qui a été calculé au moyen des seules trames actives. Certaines trames de la séquence de sortie, caractérisées comme étant "répétées", peuvent avoir parcouru la voie. Ces trames peuvent être détectées au moyen d'un processus supplémentaire de mise en correspondance.

Pour effectuer des mesurages avec une séquence de source décimée à 3:2, il faut créer un enregistrement des trames répétées dans la séquence de source, comme décrit au 6.2.4 (mais appliqué à la séquence de source). Après la caractérisation des trames actives/répétées d'après la séquence de source, il est à prévoir que le processus de mise en correspondance appariera les trames actives aux trames de source non répétées ainsi que, dans quelques cas, avec des trames de source répétées. La règle 1 du 6.2.5 prescrit une correspondance de trames biunivoque de sorte que l'on déterminera la meilleure correspondance en consultant l'enregistrement des trames de source répétées et en y choisissant la trame de source non répétée. Comme les trames originales qui produisent des trames actives apparaissent toujours avant leurs répétitions, le processus de mise en correspondance doit toujours suivre la séquence normale du début à la fin (la règle 5 ne doit jamais être appliquée).

Une fois que le processus de mise en correspondance a été appliqué à toutes les trames actives, un deuxième processus de mise en correspondance peut commencer pour les trames répétées. Cette passe tient compte à la fois des trames de source répétées et de toutes trames non mises en correspondance au cours de la première passe, ce qui évite la plupart des trames de source non répétées. Si une correspondance univoque existe entre des trames répétées de la source et des séquences de sortie (après mise en correspondance des trames actives adjacentes de sortie), l'instant d'arrivée de la trame répétée doit être inclus dans la répartition des durées entre arrivées de trame en sortie. De nouveau, une mise en correspondance biunivoque des trames doit être effectuée si une trame de source répétée est dupliquée par le système de transmission. Aucun calcul chronologique n'utilisera de trames de sortie répétées sans trames de source en correspondance univoque.

6.2.8 Facteurs d'influence sur la précision et la stabilité des mesures

Dans de nombreux systèmes de transmission vidéo, le décodeur doit fournir des trames vidéo à sa sortie en fonction d'un rythme d'affichage périodique (comme une interface avec des signaux analogiques composites). Si les horloges d'affichage d'entrée et de sortie ne sont pas synchronisées, un tampon doit être ajouté dans le décodeur. Lorsque celui-ci dispose d'une trame vidéo prête à l'affichage, il doit attendre la prochaine possibilité de sortie et augmente donc le temps de transmission global du système. Cet intervalle d'attente par le décodeur sera appelé *délai de sortie*.

Le délai de sortie est borné par la fréquence de rafraîchissement de l'affichage. Dans les systèmes de transmission à interfaces composites qui peuvent effectuer un rafraîchissement en limite de balayage, le délai de sortie maximal est de 16,7 ms. Pour des rafraîchissements aux limites de balayage sur 525 lignes, le délai maximal est de 33 ms. Le délai de sortie réel sera une valeur aléatoire comprise entre 0 et le maximum.

Lorsque les horloges des signaux d'entrée et de sortie ont un faible décalage de fréquence, le délai de sortie variera dans le temps. Si le décalage de fréquence est constant, le délai de sortie dérivera dans

son étendue de manière cyclique. Si les horloges sont synchronisées sur des oscillateurs indépendants à commande par faisceau de césium, le délai de sortie présentera une variation inférieure à 1 ms sur 13 900 heures. Si la précision d'horloge est fondée sur des oscillateurs à quartz indépendants (avec un décalage de $2 \cdot 10^{-6}$), le délai de sortie parcourra cycliquement (toutes les 4,58 heures) l'étendue entière de 33 ms.

Des marqueurs temporels à résolution de sous-trame permettent de caractériser le délai de sortie comme un composant intrinsèque du temps de transmission global du système.

6.3 Méthodes de mesure pour trames vidéo à codage temporel interne

Dans certaines situations, il est possible d'insérer des symboles visibles dans le flux vidéo d'entrée, qui peuvent être utilisés pour identifier de manière univoque chaque trame d'entrée. Ces symboles seront acheminés vers la sortie du système et pourront être utilisés de façon simple pour mesurer le débit de trames et le temps de transmission. Ces méthodes, ainsi que les restrictions relatives à leur emploi, feront l'objet d'un complément d'étude.

7 Mesures audio

7.1 Acquisition de trames audio pour les mesures

7.1.1 Description des trames audio

Une trame audio est un groupe d'échantillons audio consécutifs. Le nombre préféré d'échantillons dans une trame audio dépend de la fréquence d'échantillonnage audio et est indiqué dans le paragraphe 5. La Figure 11 ci-dessous décrit la position des trames audio dans les hiérarchies de contenu et de codage audio (dans cet exemple, les trames de codage ont une durée inférieure à celle des trames audio).

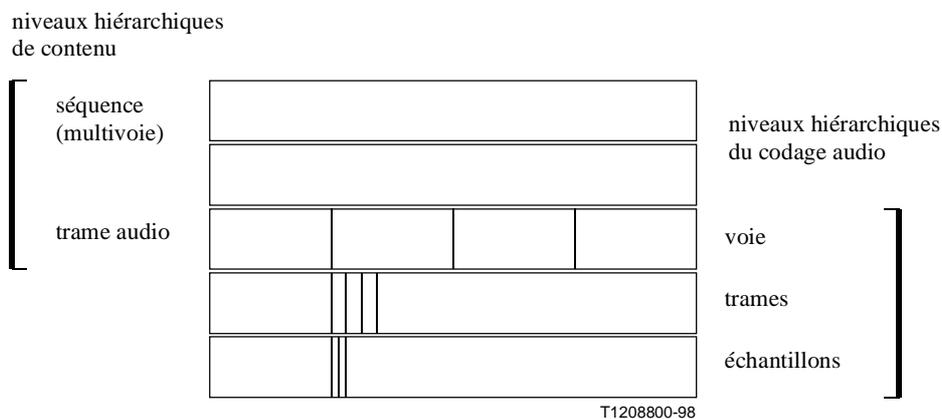


Figure 11/P.931 – Hiérarchies de contenu et de codage audio

7.1.2 Conversion d'analogique en numérique

Les méthodes de mesure décrites dans les sous-paragraphe ci-dessous nécessitent la numérisation du signal audio analogique. Le processus de numérisation produit des échantillons audio qui peuvent ensuite être groupés en trames audio. La fréquence d'échantillonnage audio est déterminée par la largeur de bande requise pour les mesures subséquentes. Conformément au théorème de Nyquist, la fréquence d'échantillonnage doit être au moins égale au double de cette largeur de bande de mesure.

Pour les signaux audio qui sont limités à la parole, une fréquence d'échantillonnage de 8000 échantillons par seconde est suffisante. Des fréquences supérieures peuvent être requises pour d'autres signaux audio. Le processus de numérisation peut autoriser une précision d'au moins 8 éléments binaires. Selon le rapport signal sur bruit du signal audio, une précision additionnelle pouvant atteindre 16 éléments binaires sera souvent utile pour les méthodes de mesure ci-après. Le processus de numérisation doit comporter un filtrage passe-bas approprié afin d'éviter le crénelage et doit être adapté à l'impédance et à la symétrie interne des signaux audio.

7.1.3 Affectation des marqueurs temporels

Le temps, $T(n)$, associé à la trame audio n doit être lu immédiatement après la numérisation du dernier échantillon contenu dans la trame audio n et avant la numérisation de l'échantillon suivant. Les marqueurs temporels pour chaque échantillon contenu dans une trame peuvent être calculés à partir du marqueur temporel de la trame audio étant donné que la fréquence d'échantillonnage est connue.

7.2 Mesurage du temps de transmission pour les trames audio

De nombreux systèmes de transmission à retenir éventuellement possèdent la capacité d'acheminer des signaux audio utilisables sans protection des formes d'onde audio entre l'entrée et la sortie. Des exemples de spécifications de codage audio peuvent être trouvés dans les Recommandations G.728 (codage CELP à 16 kbit/s), G.729 (codage CELP à 8 kbit/s) et G.723 (6,4 et 5,3 kbit/s). En d'autres termes, une mesure constante du temps de transmission ne doit pas se fonder seulement sur les formes d'onde audio. La méthode de mesure décrite ici présente une phase approchée qui fait appel aux enveloppes audio et une phase précise qui fait appel aux densités spectrales de puissance (PSD, *power spectral density*). Les enveloppes audio et les densités PSD audio sont globalement préservées par la plupart des voies. La Figure 12 représente l'organigramme du processus de mesurage.

Ce processus en deux phases est efficace parce que la phase approchée explore une vaste étendue de valeurs possibles de temps de transmission mais à basse résolution. Si cette même étendue était explorée à haute résolution, un nombre beaucoup plus grand de calculs seraient requis. Une fois que la phase approchée a été menée à bonne fin, sa mesure à basse résolution peut souvent être affinée en mesure à haute résolution par la phase précise qui suit. La phase précise n'impose qu'une exploration dans une étendue étroite de valeurs possibles de temps de transmission, compatible avec l'incertitude du mesurage approché. Pour certaines voies, les densités PSD des signaux audio ne sont pas adéquatement préservées et des mesures fines ne sont pas possibles. Dans ces situations, la mesure approchée devient, avec son incertitude intrinsèque, la mesure finale du temps de transmission.

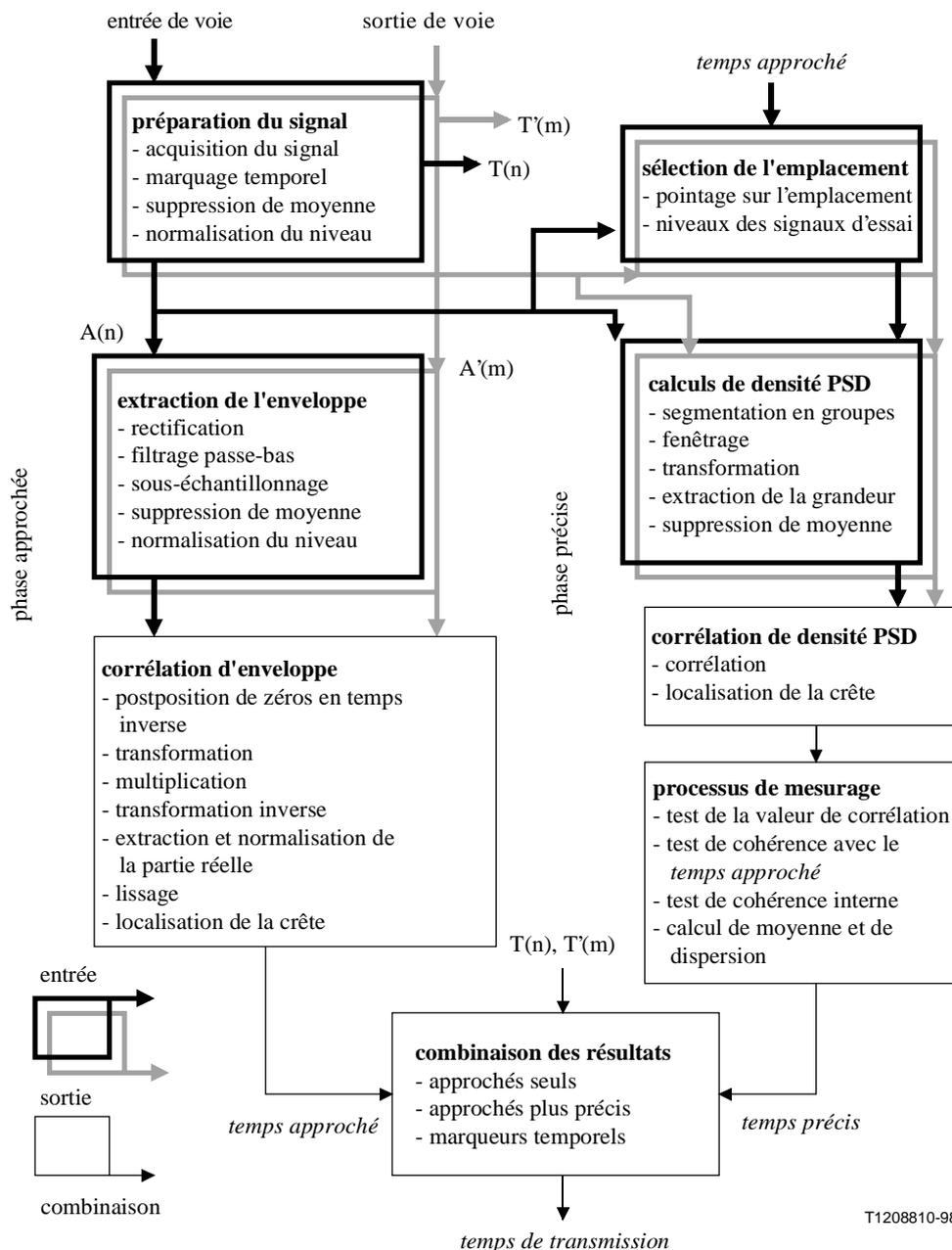


Figure 12/P.931 – Organigramme du mesurage du temps de transmission audio

7.2.1 G n ralit s

Avant que le mesurage puisse commencer, une s rie de trames audio doit  tre recueillie   l'entr e et   la sortie de la voie. L'utilisation d'un plus grand nombre de trames augmente   la fois la fiabilit  et la complexit  du mesurage. Si un groupe de trames audio ne contient que des s quences de silence entre des mots ou des phrases d'une conversation parl e, ou des tonalit s continues, aucun mesurage fiable ne pourra avoir lieu. Pour d tecter une situation de niveau audio insuffisant, le niveau quadratique moyen des  chantillons audio du groupe de trames audio extraites de l'entr e de voie doit  tre compar  au niveau quadratique moyen nominal de cette entr e de voie. Ce niveau nominal peut  tre repris des sp cifications d'entr e de voie ou  tre mesur : s'il est   plus de 30 dB au-dessous du niveau nominal d'entr e de voie, il faut acqu rir un plus grand nombre de trames audio avant que le mesurage puisse commencer. De m me, le niveau quadratique moyen des  chantillons audio dans le groupe des trames audio extraites de la sortie de voie doit  tre compar  au niveau quadratique nominal de cette sortie de voie: s'il est   plus de 30 dB au-dessous du niveau nominal de sortie de

voie, il faut acquérir un plus grand nombre de trames audio avant que le mesurage puisse commencer.

Pour des signaux de parole normaux, l'utilisation de plus grands groupes de trames audio diminue la probabilité que ces groupes ne contiennent que du silence. Par ailleurs, un plus grand nombre de trames audio apporte plus de données à mesurer et la mesure effectuée sera plus fiable. Pour les signaux audio qui sont limités à la parole, il est préférable que 256 trames audio soient extraites de l'entrée de voie et de la sortie de voie. Le mesurage sera également possible avec 128 ou 64 trames. Lorsque la fréquence d'échantillonnage est de 8000 échantillons par seconde et que chaque trame contient 128 échantillons, ces choix correspondent respectivement à environ 4 s, 2 s ou 1 s de parole. Le temps de transmission audio escompté ne doit pas dépasser 25% de la durée du signal vocal utilisé dans ce mesurage. Lorsque 256 trames (4 s) de signal vocal sont utilisées, des temps pouvant atteindre 1 s peuvent être mesurés. Lorsque 64 trames (1 s) sont utilisées, on ne mesure que 250 ms de temps de transmission. Les calculs métrologiques sont les plus efficaces lorsque le nombre de trames acquises est une puissance de deux.

7.2.2 Préparation du signal

Lorsqu'un mesurage du temps de transmission audio est requis, la trame audio extraite le plus récemment de l'entrée de voie, $A(n)$, est concaténée avec un certain nombre de trames audio déjà acquises [par exemple $A(n)$, $A(n-1)$, ... $A(n-255)$] afin de former la chronologie la plus récente des échantillons d'entrée de voie. De même, la trame audio extraite le plus récemment de la sortie de voie $A'(m)$ est concaténée au même nombre de trames audio déjà acquises [par exemple $A'(m)$, $A'(m-1)$, ... $A'(m-255)$] afin de former la chronologie la plus récente des échantillons de sortie de voie. Comme indiqué dans le paragraphe 5, la différence de temps entre ces deux processus d'acquisition est exprimée par $T'(m) - T(n)$. Des valeurs positives de cet intervalle indique que l'acquisition en sortie de voie intervient après l'acquisition en entrée de voie.

Les échantillons d'entrée sont placés dans une matrice appelée *ref*, qui contient les échantillons $ref(1)$, $ref(2)$, ... $ref(L1)$. Les échantillons de sortie sont placés dans une matrice de mêmes dimensions appelée *test*, qui contient les échantillons $test(1)$, $test(2)$, ... $test(L1)$. La valeur moyenne de chaque matrice est ensuite supprimée afin d'éliminer toute composante aperiodique dans les signaux audio numérisés:

$$\begin{cases} ref(i) = ref(i) - \frac{1}{L1} \cdot \sum_{j=i}^{L1} ref(j) \\ test(i) = test(i) - \frac{1}{L1} \cdot \sum_{j=1}^{L1} test(j) \end{cases} \quad \text{pour } 1 \leq i \leq L1$$

Ensuite, chaque matrice est normalisée en fonction d'un niveau quadratique moyen commun:

$$\begin{cases} ref(i) = ref(i) \cdot \left[\frac{1}{L1-1} \sum_{j=1}^{L1} ref(j)^2 \right]^{-\frac{1}{2}} \\ test(i) = test(i) \cdot \left[\frac{1}{L1-1} \sum_{j=1}^{L1} test(j)^2 \right]^{-\frac{1}{2}} \end{cases} \quad \text{pour } 1 \leq i \leq L1$$

7.2.3 Phase approchée

La méthode de mesure commence par une phase approchée qui extrait et met en intercorrélacion les enveloppes audio. Celles-ci sont à peu près préservées par la plupart des voies.

7.2.3.1 Extraction des enveloppes

Les enveloppes audio sont calculées comme suit. Les signaux audio numérisés dans les matrices *ref* et *test* sont rectifiés au moyen de la valeur absolue de chaque échantillon. Etant donné que les signaux audio numérisés à l'origine dans *ref* et *test* seront nécessaires pour la phase précise, les signaux rectifiés sont mémorisés, avec d'autres résultats intermédiaires subséquents, dans les matrices temporaires *ref_temp* et *test_temp*:

$$ref_temp(i) = |ref(i)|, test_temp(i) = |test(i)|, 1 \leq i \leq L1$$

Les signaux rectifiés subissent ensuite un filtrage passe-bas afin de créer des enveloppes audio ayant une largeur de bande d'environ 125 Hz. C'est ce filtrage passe-bas et le sous-échantillonnage subséquent qui donnent à la phase approchée sa résolution réduite et sa charge de calcul réduite. Le facteur de réduction de la largeur de bande et le facteur de sous-échantillonnage sont tous les deux spécifiés par la variable *B*. Le Tableau 2 donne les valeurs appropriées de *B* pour certaines fréquences d'échantillonnage audio courantes.

Tableau 2/P.931 – Valeurs du facteur de réduction de largeur de bande, B

Fréquence d'échantillonnage audio (éch./s)	B
8 000	32
16 000	64
32 000	128
44 100	176
48 000	192

Lorsque la fréquence d'échantillonnage audio est de 8000 échantillons par seconde, la largeur de bande doit être réduite d'un facteur *B* = 32 pour passer d'une largeur de bande nominale de 4000 Hz à une largeur de bande nominale de 125 Hz. La réduction de largeur de bande requise peut être convenablement approchée au moyen d'un filtre passe-bas de Butterworth à réponse impulsionnelle infinie (IIR, *infinite impulse response*) ayant un point -3 dB à 125 Hz. La réalisation de forme directe est la suivante:

$$out(i) = \sum_{j=0}^7 b_j \cdot in(i-j) - \sum_{j=1}^7 a_j \cdot out(i-j), 1 \leq i \leq L1$$

où

$$out(i) = in(i) = 0, i \leq 0$$

Il faut veiller à éliminer tous les échantillons de sortie de filtre qui contiendraient une transitoire d'ouverture de filtre. Pour la réalisation de forme directe montrée ici, la transitoire d'ouverture est limitée à environ 400 échantillons. Les coefficients du filtre sont indiqués dans le Tableau 3.

Tableau 3/P.931 – Valeurs des coefficients pour un filtre passe-bas de Butterworth à réponse IIR du 7^e ordre

j	a_j	b_j
0	1,00000000	$0,00553833 \times 10^{-7}$
1	-6,55883158	$0,03876830 \times 10^{-7}$
2	18,44954612	$0,11630512 \times 10^{-7}$
3	-28,85178274	$0,19384125 \times 10^{-7}$
4	27,08958968	$0,19384206 \times 10^{-7}$
5	-15,27097592	$0,11630465 \times 10^{-7}$
6	4,78557610	$0,03876843 \times 10^{-7}$
7	-0,64312159	$0,00553831 \times 10^{-7}$

Les deux matrices *ref_temp* et *test_temp* subissent un filtrage passe-bas au moyen de ce filtre. Elles sont ensuite sous-échantillonnées par conservation de tous les $B^{\text{ièmes}}$ échantillons seulement, ce qui donne un total de $L2$ échantillons. Par exemple, lorsque $B = 32$, les échantillons 1, 33, 65, etc. seront conservés. Dans le cas d'un groupe de 256 trames audio, dont 128 échantillons de chaque trame sont utilisés comme entrée en phase approchée, le processus de sous-échantillonnage donnera $L1 = 32\ 768$ et $L2 = L1/B = 1024$ échantillons. Les deux matrices *ref_temp* et *test_temp* contiennent alors les enveloppes audio, qui seront finalement normalisées. La valeur moyenne de chaque matrice est supprimée et chaque matrice est divisée par son écart type pour la normaliser au niveau quadratique moyen commun.

$$ref_temp(i) = ref_temp(i) - \frac{1}{L2} \cdot \sum_{j=1}^{L2} ref_temp(j)$$

$$test_temp(i) = test_temp(i) - \frac{1}{L2} \cdot \sum_{j=1}^{L2} test_temp(j)$$

$$ref_temp(i) = ref_temp(i) \cdot \left[\frac{1}{L2-1} \sum_{j=1}^{L2} ref_temp(j)^2 \right]^{-\frac{1}{2}}$$

$$test_temp(i) = test_temp(i) \cdot \left[\frac{1}{L2-1} \sum_{j=1}^{L2} test_temp(j)^2 \right]^{-\frac{1}{2}}, 1 \leq i \leq L2$$

7.2.3.2 Intercorrélation des enveloppes audio

On calcule l'intercorrélacion des enveloppes audio dans les matrices *ref_temp* et *test_temp* au moyen d'une convolution circulaire. Celle-ci est calculée à son tour au moyen des transformées de Fourier discrète (DFT, *discrete Fourier transform*) ou rapide (FFT, *fast Fourier transform*). La matrice *ref_temp* est d'abord étendue de la longueur $L2$ à la longueur $2 \cdot L2$ par postposition de $L2$ zéros. Dans l'exemple ci-dessus, $L2=1024$ zéros seraient ajoutés pour que la matrice finale ait une dimension de 2048. Ensuite, la matrice *test_temp* subit une inversion temporelle. Pour effectuer cette opération interne, les échantillons 1 et $L2$ de *test_temp* sont échangés, ainsi que les échantillons 2 et $L2 - 1$, les échantillons 3 et $L2 - 2$, etc. Lorsque le nombre $L2$ est pair, l'échange final s'effectue entre les échantillons $L2/2$ et $L2/2 + 1$. Lorsque le nombre $L2$ est impair, l'échange final s'effectue

entre les échantillons $L2/2 - 1/2$, et $L2/2 + 3/2$. Après cette inversion temporelle, la matrice *test_temp* est étendue de la longueur $L2$ à la longueur $2 \cdot L2$ par postposition de $L2$ zéros.

Ensuite, les matrices *ref_temp* et *test_temp* sont transformées par DFT ou FFT. Lorsque la longueur d'une matrice, $2 \cdot L2$, est une puissance de deux, la transformation FFT peut être utilisée. Si la longueur $2 \cdot L2$ n'est pas une puissance de deux, la transformation DFT peut être utilisée. En variante, le nombre de zéros postposés dans l'étape précédente peut être augmenté de façon que la longueur de la matrice soit une puissance de deux et que la transformation FFT puisse alors être utilisée. De toute façon, un algorithme de transformation interne peut être utilisé afin d'obtenir des versions transformées des matrices *ref_temp* et *test_temp* qui remplaceront les versions antérieures. Ces transformations produisent des nombres complexes.

Les échantillons complexes enregistrés dans les matrices *ref_temp* et *test_temp* sont ensuite multipliés un à un et les résultats complexes sont introduits dans une nouvelle matrice appelée *cross_corr*, qui a la même longueur que *ref_temp* et *test_temp*:

$$cross_corr(i) = ref_temp(i) \cdot test_temp(i), \text{ pour } i = 1 \text{ pour } 2 \cdot L2$$

La matrice *cross_corr* subit ensuite une transformée de Fourier rapide inverse (IFFT, *inverse fast Fourier transformed*) ou une transformation de Fourier discrète inverse (IDFT, *inverse discrete Fourier transformed*) selon ce qui correspond à sa longueur. Une transformation interne peut être utilisée. Théoriquement, le contenu résultant de *cross_corr* devrait être constitué de nombres réels. En pratique, les calculs à précision finie donnent une petite composante imaginaire. A ce point, la partie réelle de *cross_corr* est conservée et la partie imaginaire est ignorée. Ensuite, chaque résultat contenu dans *cross_corr* est normalisé:

$$cross_corr(i) = cross_corr(i)/(L2 - 1), 1 \leq i \leq 2 \cdot L2$$

On notera que cette normalisation est requise pour obtenir des valeurs d'intercorrélation vraie entre -1 et $+1$. Mais elle n'a pas d'influence sur les étapes suivantes de lissage ou de détection de crête.

La matrice *cross_corr* contient les valeurs d'intercorrélation entre les enveloppes vocales contenues dans les matrices *ref_temp* et *test_temp* pour toute variation possible de ces enveloppes. Ces résultats sont ensuite lissés par un filtre passe-bas symétrique à réponse impulsionnelle finie (FIR) du deuxième ordre. Ils sont ensuite enregistrés dans une matrice d'intercorrélations lissées:

$$cross_corr_s(i) = .25 \cdot cross_corr(i - 1) + .5 \cdot cross_corr(i) + .25 \cdot cross_corr(i + 1), \\ 2 \leq i \leq 2 \cdot L2 - 1, cross_corr_s(i) = cross_corr(i), i = 1, 2 \cdot L2$$

Après ce lissage, la plus grande valeur contenue dans *cross_corr_s* est prise comme indication du temps de transmission approché:

$$coarse_delay = (L2 - j) \cdot B \text{ échantillons}$$

où:

$$cross_corr_s(j) > cross_corr_s(i), 1 \leq i \leq 2 \cdot L2, i \neq j$$

L'incertitude sur la valeur de *coarse_delay* à ce point est prise égale à $\pm B$ échantillons. Si la matrice *cross_corr_s* ne contient pas une valeur maximale unique, le mesurage doit être refait avec de nouveaux échantillons audio.

7.2.4 Phase précise

L'incertitude de $\pm B$ échantillons, inhérente au mesurage approché du temps de transmission audio, peut souvent être réduite par une phase précise du mesurage de ce temps.

7.2.4.1 Sélection de l'emplacement

La phase précise est effectuée à $n1$ emplacements dans le signal audio acquis. Lorsque les signaux audio sont limités à la parole et que 256 trames audio sont utilisées pour le mesurage du temps de transmission audio, la valeur de $n1$ est égale à 6. D'autres valeurs de $n1$ peuvent mieux convenir pour d'autres signaux audio. A chaque emplacement, l'on explore une étendue de valeurs possibles du temps de transmission, de $-3 \cdot B$ à $3 \cdot B$ échantillons.

Les emplacements où la phase précise est appliquée sont sélectionnés aléatoirement. A chaque emplacement, $8 \cdot B$ échantillons sont extraits de la matrice *ref* et sont enregistrés dans la matrice *ref_temp*. Puis $2 \cdot B$ échantillons sont extraits de la matrice *test* et sont enregistrés dans la matrice *test_temp*. Les échantillons extraits de la matrice *test* sont décalés par le temps de transmission approché qui a été mesuré:

$$ref_temp(i) = ref(location - 4 \cdot B - 1 + i), 1 \leq i \leq 8 \cdot B,$$

$$test_temp(i) = test(location + coarse_delay - B - 1 + i), 1 \leq i \leq 2 \cdot B,$$

où *location* est une variable pseudo-aléatoire à répartition uniforme qui est extraite de l'intervalle:

$$[\max(4 \cdot B + 1, 1 - coarse_delay + B), \min(L1 - 4 \cdot B + 1, L1 - coarse_delay - B + 1)]$$

Le mesurage précis du temps de transmission ne sera pas possible dans les régions à séquences de silence ou à tonalités continues. Deux essais de niveau pour détecter les régions de silence sont effectués à chaque emplacement pour veiller à ce que le signal audio qui y est présent ne s'écarte pas de plus de 30 dB du niveau moyen du signal audio:

$$-30 \leq 10 \cdot \log_{10} \left[\frac{1}{8 \cdot B - 1} \sum_{i=1}^{8 \cdot B} ref_temp(i)^2 \right], -30 \leq 10 \cdot \log_{10} \left[\frac{1}{2 \cdot B - 1} \sum_{i=1}^{2 \cdot B} test_temp(i)^2 \right]$$

Si l'un ou l'autre de ces essais de niveau est défavorable, un nouvel emplacement doit être sélectionné.

7.2.4.2 Calculs de la densité spectrale de puissance

La phase précise consiste à calculer les intercorrélations entre densités spectrales de puissance (PSD, *power spectral density*) audio à chacun des emplacements sélectionnés. Les densités PSD sont calculées comme suit. Les $8 \cdot B$ échantillons contenus dans *ref_temp* sont répartis en groupes de $2 \cdot B$ échantillons. Il y a $6 \cdot B + 1$ groupes de ce type. Chaque groupe d'échantillons est enregistré dans une matrice appelée *ref_temps_i*:

$$ref_temp_i(j) = ref_temp(i + j - 1), 1 \leq i \leq 6 \cdot B + 1, 1 \leq j \leq 2 \cdot B$$

Chaque matrice *ref_temp_i*, ainsi que la matrice *test_temp*, est multipliée par une fenêtre de Hamming puis transformée en domaine de fréquence par DFT ou FFT sur une de $2 \cdot B$. Ces étapes peuvent être effectuées sans transfert:

$$ref_temp_i(j) = ref_temp_i(j) \cdot \{.54 - .46 \cdot \cos(2\pi(j - 1)/(2 \cdot B - 1))\}, 1 \leq i \leq 6 \cdot B + 1, 1 \leq j \leq 2 \cdot B$$

$$test_temp(j) = test_temp(j) \cdot \{.54 - .46 \cdot \cos(2\pi(j - 1)/(2 \cdot B - 1))\}, 1 \leq j \leq 2 \cdot B$$

$$ref_temp_i = FFT (ref_temp_i), 1 \leq i \leq 6 \cdot B + 1$$

$$test_temp = FFT (test_temp)$$

Dans le domaine fréquentiel, seuls les $B + 1$ premiers échantillons complexes contenus dans chaque matrice sont uniques, de sorte que seuls ces échantillons sont sauvegardés. La grandeur de chaque échantillon conservé est relevée et donne la racine carrée de la densité spectrale de puissance de chaque trame. Ces résultats sont dénommés, par souci de simplicité, PSD.

$$ref_temp_i(j) = |ref_temp_i(j)|, 1 \leq i \leq 6 \cdot B + 1, 1 \leq j \leq B + 1$$

$$test_temp(j) = |test_temp(j)|, 1 \leq j \leq B + 1$$

La valeur moyenne de chaque PSD est ensuite supprimée:

$$ref_temp_i(j) = ref_temp_i(j) - \frac{1}{B+1} \cdot \sum_{j=1}^{B+1} ref_temp_i(j), 1 \leq i \leq 6 \cdot B + 1, 1 \leq j \leq B + 1$$

$$test_temp(j) = test_temp(j) - \frac{1}{B+1} \cdot \sum_{j=1}^{B+1} test_temp(j), 1 \leq j \leq B + 1$$

7.2.4.3 Intercorrélation

Une valeur d'intercorrélation est calculée entre la PSD enregistrée dans la matrice *test_temp* et chacune des $6 \cdot B + 1$ densités PSD enregistrées dans les matrices *ref_temp_i*.

$$cross_corr(i) = \frac{\left(\sum_{j=1}^{B+1} ref_temp_i(j) \cdot test_temp(j) \right)}{\left(\sum_{j=1}^{B+1} ref_temp_i(j)^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{B+1} test_temp(j)^2 \right)^{\frac{1}{2}}}, 1 \leq i \leq 6 \cdot B + 1$$

La matrice *cross_corr* contient donc les valeurs des intercorrélations entre la PSD de référence et les PSD d'essai pour chaque variation dans le domaine temporel. On notera que le deuxième terme du dénominateur de l'équation pour la matrice *cross_corr* est une constante normalisante qui est nécessaire pour obtenir des valeurs d'intercorrélation vraie entre -1 et $+1$. Cette constante n'a pas d'incidence sur la phase suivante de détection de crête mais en a une sur le traitement postérieur des mesures du temps de transmission précis. La plus grande valeur contenue dans *cross_corr* est prise comme une indication du temps de transmission précis:

$$fine_delay_k = (3 \cdot B + 1) - j \text{ échantillons,}$$

$$corr_k = cross_corr(j), 1 \leq k \leq n1,$$

où

$$cross_corr(j) > cross_corr(i), 1 \leq i \leq 6 \cdot B + 1, i \neq j$$

Si *cross_corr* ne contient pas de valeur maximale unique, la procédure de phase précise doit être répétée sur un nouvel emplacement. L'ensemble de cette phase précise, à partir de la sélection d'un emplacement, est répété $n1$ fois, ce qui donne $n1$ mesures du temps de transmission précis, enregistrées respectivement dans *fine_delay_1*, *fine_delay_2*, ... *fine_delay_n1*, ainsi que $n1$ valeurs de corrélation correspondantes, enregistrées respectivement dans *corr_1*, *corr_2*, ... *corr_n1*. On notera que chacune des estimations de temps précis s'inscrira entre $-3 \cdot B$ et $3 \cdot B$ inclus.

7.2.4.4 Traitement des mesures de temps de transmission précis

Une fois que l'on a calculé les $n1$ mesures de temps précis et les valeurs d'intercorrélation correspondantes, ces données subissent un traitement complémentaire afin de déterminer la façon dont elles seront utilisées.

Chacune des $n1$ valeurs de corrélation est d'abord comparée à un seuil:

$$\sqrt{\frac{1}{2}} \leq corr_k \Rightarrow \text{la mesure } fine_delay_k \text{ est retenue, } 1 \leq k \leq n1$$

Ce processus permet de ne conserver que les mesures de temps précis où au moins la moitié de la variance de densité PSD est prise en compte. Le nombre de mesures de temps précis qui satisfont à cette condition est n_2 et les mesures sont donc renumérotées $fine_delay_1, fine_delay_2, \dots, fine_delay_{n_2}$. Si $n_2 < n/2$, la phase précise ne donnera pas de résultat utile. Dans ce cas, la valeur de $fine_delay$ est mise à "non valide" et la phase précise est terminée.

Si $n_2 \geq n/2$, la phase précise continue et contrôle les n_2 mesures de temps précis restantes quant à leur cohérence avec la mesure de temps approché. Étant donné que l'incertitude sur la mesure du temps approché est de $\pm B$ échantillons et que le temps approché a été supprimé, seules les mesures de temps précis entre $-B$ et B échantillons sont retenues:

$$|fine_delay_k| \leq B \Rightarrow \text{la mesure } fine_delay_k \text{ est retenue, } 1 \leq k \leq n_2$$

Le nombre de mesures de temps précis qui satisfont à ce contrôle est n_3 , et les mesures sont alors renumérotées en $fine_delay_1, fine_delay_2, \dots, fine_delay_{n_3}$. Si $n_3 < n/2$, la phase précise ne produira pas de résultat utile. Dans ce cas, la valeur de la mesure $fine_delay$ est mise à "non valide" et la phase précise est terminée.

Si $n_3 \geq n/2$, la phase précise continue et contrôle quant à leur cohérence les n_3 mesures de temps précis restantes. Cet essai nécessite une exploration de tous les sous-ensembles possibles de dimension $n_3, n_3 - 1$, et ainsi de suite en descendant jusqu'à $n/2$. Il existe un seul sous-ensemble possible de dimension $n_3, n_3 - 1$ sous-ensembles possibles de dimension $n_3 - 1, n_3 \cdot (n_3 - 1)/2$ sous-ensembles possibles de dimension $n_3 - 2$, et ainsi de suite. Pour chaque sous-ensemble, la dispersion des mesures de temps précis est contrôlée:

$$\max_i \{fine_delay_i\} - \min_i \{fine_delay_i\} \leq \frac{B}{2}, fine_delay_i \in \text{sous-ensemble actuel}$$

Le plus grand sous-ensemble qui satisfait à ce contrôle est appelé sous-ensemble final. La phase précise ne peut pas fournir de résultat utile lorsque:

- aucun sous-ensemble ne satisfait à ce contrôle;
- ou lorsque le sous-ensemble qui satisfait à ce contrôle n'est pas le seul ni le plus grand.

Si l'une de ces deux conditions est vérifiée, la valeur de la mesure $fine_delay$ est mise à "non valide" et la phase précise est terminée.

Le nombre de mesures de temps précis dans le sous-ensemble final est n_4 . Ces n_4 mesures de temps précis sont alors renumérotées en $fine_delay_1, fine_delay_2, \dots, fine_delay_{n_4}$. La valeur moyenne de ces n_4 mesures de temps précis est prise comme étant la mesure finale du temps de transmission précis:

$$fine_delay = \frac{1}{n_4} \cdot \sum_{i=1}^{n_4} fine_delay_i$$

La dispersion des n_4 mesures dans le sous-ensemble final est conservée comme une indication d'incertitude sur la mesure finale du temps de transmission précis:

$$spread = \max_i \{fine_delay_i\} - \min_i \{fine_delay_i\}, fine_delay_i \in \text{sous-ensemble final}$$

7.2.5 Combinaison des résultats des phases approchée et précise

Si la phase précise n'a pas permis d'obtenir une mesure utile du temps de transmission précis, la phase précise aura mis la mesure $fine_delay$ à "non valide". Dans ce cas, la seule mesure approchée devient la mesure du temps de transmission. Si la phase précise a permis d'obtenir une mesure utile du temps de transmission précis, la mesure approchée est augmentée de cette mesure précise et l'incertitude est réduite comme suit, par rapport à celle de la seule mesure approchée:

$$fine_delay = "invalid" \Rightarrow delay = coarse_delay \pm B \text{ échantillons}$$

$$fine_delay \neq "invalid" \Rightarrow delay = coarse_delay + fine_delay \pm spread \text{ samples}$$

Ces valeurs de temps de transmission ne sont correctes que lorsque les acquisitions d'échantillons audio à l'entrée et à la sortie de la voie sont simultanées. Après conversion du temps de transmission en secondes, des marqueurs temporels peuvent être utilisés pour corriger la mesure du temps de transmission afin de tenir compte d'une acquisition non simultanée:

$$delay = delay/sample_rate \text{ secondes}$$

$$delay = delay + T'(m) - T(n) \text{ secondes}$$

8 Considérations relatives à la combinaison des mesures audio et vidéo

Les concepts et méthodes définis dans les paragraphes 6 et 7 permettent d'examiner certains problèmes de mesurage pour voies multiples.

8.1 Mesures d'activité et de synchronisation des voies audio et vidéo

Pour donner des résultats valides, les deux méthodes de mesure, vidéo et audio, nécessitent des niveaux minimaux du signal dans les voies en essai. Chaque méthode a ses propres exigences. Les méthodes de type vidéo exigent des différences détectables entre trame vidéo actuelle et trame vidéo précédente, alors que les méthodes audio exigent que le niveau quadratique moyen d'un groupe de trames audio soit suffisant par rapport aux niveaux nominaux des interfaces.

Pour qu'il soit possible de comparer les mesures de temps de transmission audio et vidéo, les conditions d'activité nécessaires doivent être présentes simultanément à chaque entrée et des mesurages valides doivent être effectués dans les deux voies (qui impliquent une activité en sortie). Sinon, les calculs de synchronisation ne sont pas possibles et une opportunité différente doit être recherchée.

8.2 Association de mesures individuelles

Les calculs de synchronisation du paragraphe 5 nécessitent une association entre trames d'entrée ainsi qu'un processus de mise en correspondance de trames fournissant des paires de marqueurs temporels pour chaque voie contrôlée. Le calcul du décalage temporel entre voies et de l'erreur temporelle introduite par la voie de transmission ne comporte pas la mesure du temps de transmission comme étape intermédiaire. Plus précisément, cela évite de comparer des temps de transmission mesurés à deux moments différents pour évaluer la synchronisation de deux voies.

Les voies audio constituent cependant une exception logique car elles transportent un média isochrone. Lorsque la variation mesurée du temps de transmission audio est limitée à l'erreur expérimentale escomptée, le temps de transmission audio moyen peut être considéré comme une valeur constante représentative. Ce temps moyen peut être comparé à une répartition du temps de transmission dans une voie vidéo ou télématique afin d'obtenir une répartition de la dérive temporelle entre les voies. Cette méthode permet d'effectuer une comparaison entre mesures audio et vidéo relevées isolément mais dans des conditions de source (etc.) identiques. De cette manière, les appareillages d'essai qui ne peuvent pas effectuer de mesurages simultanés sur des voies multiples peuvent encore fournir des informations utiles si les conditions le permettent.

Si la variation mesurée du temps de transmission audio est supérieure à l'erreur expérimentale prévue, des mesures sur voie individuelle ne peuvent pas être utilisées.

9 Mesures sur les voies de données

Le présent paragraphe spécifie les méthodes de mesure du temps de transmission dans les voies de données qui font partie de systèmes de communications multimédias. De nombreuses Recommandations sur les performances en termes de transmission de données définissent le temps de transmission ou de propagation pour des protocoles de communication spécifiques (comme la Recommandation X.25 et les diverses Recommandations traitant du mode ATM).

9.1 Acquisition de trames de données pour les mesures

Les voies de données implémentées dans les systèmes de communications multimédias peuvent varier largement selon leurs objectifs et leurs attributs spécifiques. Plutôt que d'essayer de traiter toutes les possibilités, le présent sous-paragraphe présente les méthodes générales qui sont applicables aux voies de données, selon deux catégories principales (définies ci-dessous).

9.1.1 Considérations relatives à la définition des trames de données

La présente Recommandation traite de la voie de données à un niveau logique, au-dessus de la couche Physique et de ses interfaces électriques. Des systèmes pouvant effectuer les mesurages décrits dans la présente Recommandation disposeront cependant de capacités conformes à une ou plusieurs normes sur les interfaces électriques. Les observations faites à ces interfaces électriques formeront la base des mesures.

La présente Recommandation se rapporte à une séquence de Z bits consécutifs formant une trame de données. Une trame de données n est représentée par la variable $D(n)$ et le premier bit de $D(n)$ est $D(1,n)$. La longueur d'une trame de données peut être déterminée par l'application de la voie de données.

L'on distingue deux configurations possibles pour les données d'utilisateur:

- 1) les utilisateurs fournissent des bits d'information imbriqués dans une *structure* normalisée. Ces structures peuvent être des paquets, des cellules ou des trames;
- 2) les utilisateurs fournissent des flux *non structurés* de bits. Le système de communication multimédia peut effectuer sa propre segmentation dans ce flux binaire.

Lorsque l'utilisateur fournit un flux binaire structuré et que cette structure permet de reconnaître les trames individuelles à chaque interface de voie, la structure originale est considérée comme étant la trame de données pour les mesures des systèmes multimédias. Une exception peut se produire lorsque la structure originale contient un grand nombre de bits et que le temps d'insertion de structure est grand par rapport à la longueur des trames audio et des trames vidéo. Dans ce cas, il est parfois plus efficace de traiter le flux binaire comme s'il n'avait pas de structure. La situation idéale est un temps d'insertion égal pour les trames de tous les médias, ce qui autorise une correspondance biunivoque.

Lorsque la voie de données permet l'absence de structure des bits d'entrée et que le système de mesure peut fournir les bits, un générateur de séquences pseudo-aléatoires peut être utilisé, ce qui offre plusieurs avantages:

- la séquence peut être produite facilement au site local et au site distant;
- l'intervalle de répétition de la séquence peut être choisi de façon à éviter des correspondances ambiguës;
- la longueur des trames de données peut être réduite à un simple multiple de la longueur du registre à décalage linéaire de réinjection; elle peut également être choisie de façon à s'approcher de la longueur d'autres trames de média.

Lorsque la voie de données nécessite une grande structure, le générateur de séquences pseudo-aléatoires peut fournir les bits d'information acheminés par cette structure.

Le système de mesure peut également acquérir le flux binaire d'entrée à partir de la source normale de la voie de données. Dans ce cas, la longueur des trames de données coïncidera généralement avec la structure originale. Si la source de données produit un codage vide, il est très peu probable que les mesures seront fidèles.

9.1.2 Affectation des marqueurs temporels

Le temps, $T_D(n)$, associé à la trame de données n , doit être lu immédiatement après la communication du dernier bit contenu dans la trame traversant l'interface et avant que le bit suivant soit communiqué par l'intermédiaire de l'interface.

Comme la durée du signal d'insertion dans une trame est également une mesure utile en transmission de données, des marqueurs temporels additionnels peuvent être associés au premier bit d'une trame de données et doivent être lus avant que le bit suivant soit communiqué par l'intermédiaire de l'interface. Dans certains systèmes, la durée du signal d'insertion à l'entrée peut être différente de la durée du signal d'insertion à la sortie. Par ailleurs, la durée du signal d'insertion n'est pas toujours constante.

9.2 Mesure du temps de transmission des données

Le présent sous-paragraphe indique deux méthodes pour mettre en correspondance les trames de données à l'entrée et à la sortie.

9.2.1 Mise en correspondance de données structurées

La référence bibliographique [4] présente des méthodes de détermination de la correspondance entre paquets X.25 qui sont applicables à de nombreuses formes de données structurées. Habituellement, les bits d'en-tête contiennent des informations suffisantes (comme un numéro de séquence) pour que l'on puisse différencier les paquets les uns des autres. Ces identificateurs imbriqués sont une base valide pour la mise en correspondance de trames de données.

Si les informations d'en-tête d'un protocole spécifique ne sont pas suffisantes, il est parfois possible de fonder la correspondance sur une identification additionnelle insérée dans le champ de données d'utilisateur.

9.2.2 Mise en correspondance de données non structurées

La référence bibliographique [5] définit la correspondance binaire par comparaison de séquences binaires. Les bits $D'(m)$ correspondent à une séquence d'entrée de longueur égale s'il existe des entiers n et d tels que:

$$D'(i,m) = D(i + d,n) \text{ pour presque tous les entiers } 1 \leq i \leq Z - d$$

et

$$D'(i,m) = D(i - Z + d,n + 1) \text{ pour presque tous les entiers } Z - d + 1 \leq i \leq Z$$

où d est le décalage binaire de l'entier positif ($d < Z$) qui peut exister entre les attributions de trame d'entrée, $D(n)$, et de sortie, $D'(m)$.

La Figure 13 illustre la correspondance aux limites des trames d'entrée.

Si les trames de données d'entrée ou de sortie sont réalignées de façon que leur décalage binaire soit fixé à $d = 0$, le test de correspondance se simplifie en:

$$D'(i,m) = D(i,n) \text{ pour presque tous les entiers tels que } 1 \leq i \leq Z$$

Si les bits sont communiqués de part et d'autre des interfaces de façon périodique, il est possible de calculer le marqueur temporel pour tout bit contenu dans une trame de données ce qui rend les marqueurs temporels disponibles pour le premier ou le dernier bit dans toute nouvelle définition de trame $D(n)$ telle que $d = 0$.

La possibilité d'une correspondance sur *presque tous* les bits d'une trame de données est également une possibilité de correspondance correcte en présence d'erreurs limitées sur les bits. Lorsqu'il n'y a pas d'erreurs sur les bits, il doit y avoir égalité pour tous les entiers i dans l'étendue $1 \leq i \leq Z$.

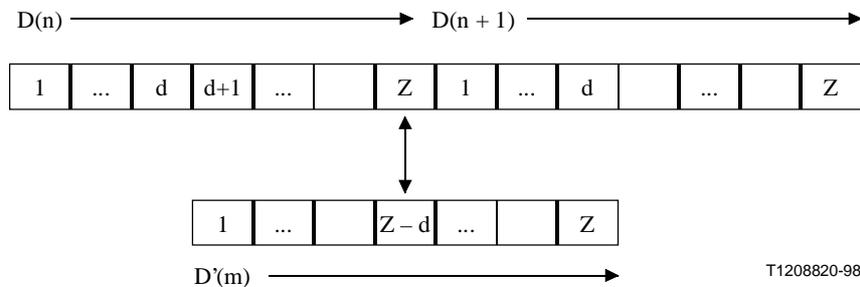


Figure 13/P.931 – Correspondance entre trames de données à l'entrée et à la sortie

10 Prescriptions relatives à la stabilité des temporisateurs et à la synchronisation

Le présent paragraphe contient les prescriptions minimales concernant les temporisateurs ou horloges internes fournissant les marqueurs temporels pour les trames. Il y a deux configurations d'horloge à considérer:

- 1) une horloge unique fournit les marqueurs temporels d'entrée et de sortie (habituellement utilisés par les applications de mesure en boucle locale ou distante). Dans ce cas, seules les prescriptions de précision, de stabilité et de résolution sont applicables car elles caractérisent complètement la qualité de fonctionnement d'une horloge;
- 2) deux horloges, se trouvant éventuellement dans des endroits différents (distants) fournissent les marqueurs temporels à l'entrée et à la sortie du système de transmission. Cette configuration s'applique aux mesures de bout en bout. Toutes les prescriptions du présent paragraphe s'appliquent à cette configuration.

10.1 Résolution

La résolution minimale de l'échelle temporelle pouvant être incluse dans les marqueurs temporels est de $0,1 \mu\text{s}$ (10^{-7} s). Il s'agit de la résolution de mémorisation interne prévue pour les mesurages. Bien que cette résolution totale ne doive pas être signalée si elle n'est pas compatible avec la précision de l'horloge interne, elle offre aux concepteurs une base précise pour l'évaluation de la stabilité et de la précision des horloges.

10.2 Précision et stabilité (erreur admissible sur la durée)

La précision et la stabilité de l'échelle temporelle de l'horloge interne est entièrement contrainte par la prescription relative à l'erreur admissible sur la durée (TIE, *time interval error*), qui est définie comme étant la variation temporelle des lectures indiquées par une horloge donnée par rapport à une échelle temporelle idéale pendant une période d'observation donnée, S . L'erreur maximale sur la durée (MTIE, *maximum time interval error*) est la plus grande TIE pour tous les intervalles de mesurage possibles dans la période d'observation.

En pratique, les mesurages de temps de transmission effectués conformément à la présente Recommandation dureront au moins 1 s. La spécification d'erreur MTIE commencera donc à partir d'un intervalle d'observation de 0,01 s (des intervalles plus courts ne sont pas spécifiés).

Dans de nombreuses applications de la présente Recommandation, l'horloge ou les horloges seront synchronisées sur un signal de référence de temps tel que le système mondial de localisation (GPS, *global positioning system*). Dans ce cas, l'erreur MTIE est décrite par l'équation suivante:

$$\text{MTIE, ns} \leq 10^{-2} S + 150$$

comme illustré dans la Figure 14.

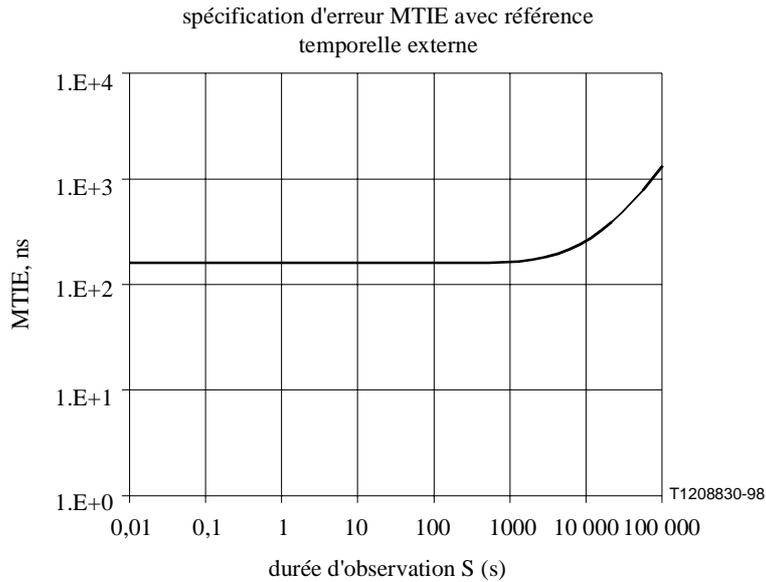


Figure 14/P.931 – Erreur MTIE avec référence temporelle

D'autres applications utiliseront une unique horloge interne ou une horloge distante pouvant être synchronisée avant le mesurage mais dépendant ensuite de sa précision interne (également appelée en régime libre) pour garder le temps. Cette méthode permet d'effectuer des mesurages pendant des intervalles de temps limités si une base de temps primaire (comme le GPS) n'est pas disponible. Pour les horloges internes de type A, l'erreur MTIE est contrainte par:

$$\text{MTIE, ns} \leq 10S + 150$$

comme illustré à la Figure 15.

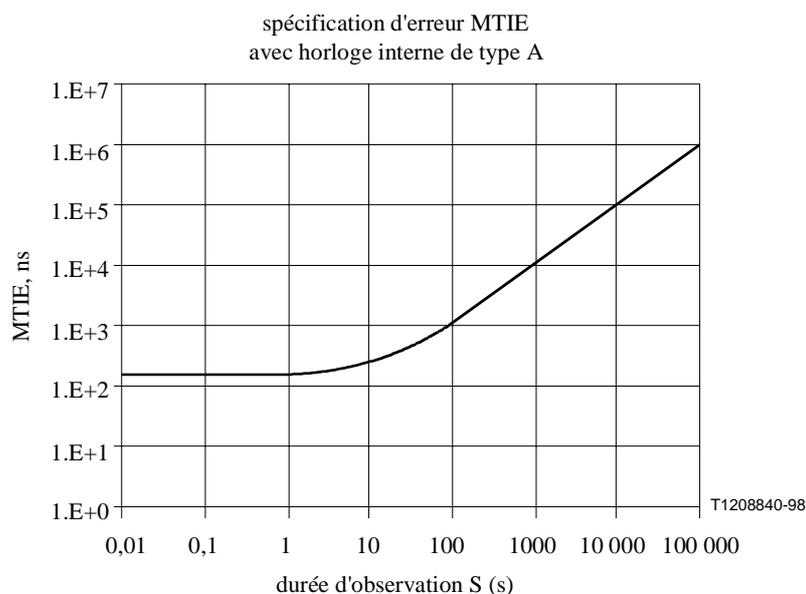


Figure 15/P.931 – Spécification d'erreur MTIE avec horloge interne de type A (en cours de mesure)

Pour les horloges internes de type B, l'erreur MTIE est contrainte par:

$$\text{MTIE, ns} \leq 138,9 \times S + 150$$

comme illustré à la Figure 16.

Dans tous les cas, les comptes-rendus de mesure doivent être accompagnés de l'erreur maximale (déterminée par la présence d'une source de référence), y compris le temps écoulé depuis que la référence est disponible et l'intervalle de mesure réel.

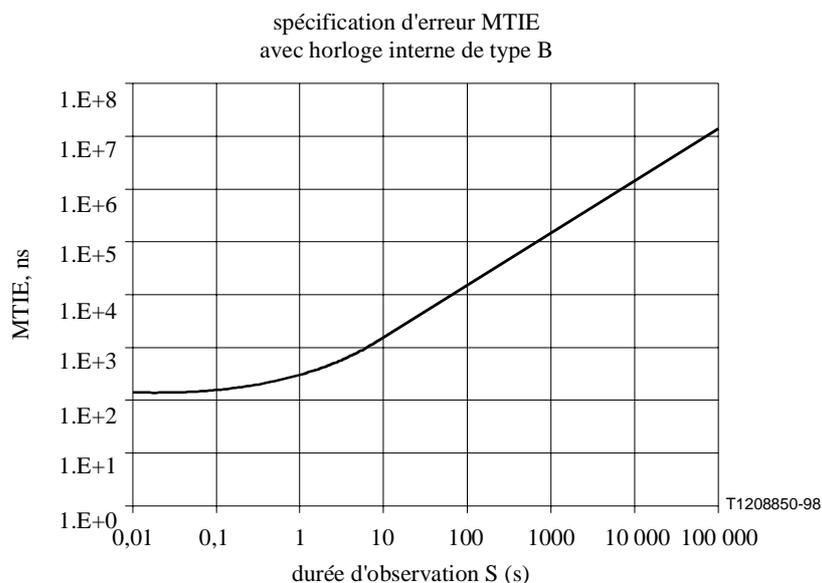


Figure 16/P.931 – Spécification d'erreur MTIE avec horloge interne de type B (en cours de mesure)

10.3 Erreur de réglage du temps

Si plusieurs horloges sont utilisées lors d'un mesurage, elles doivent être synchronisées. Si elles sont synchronisées directement ou en fonction d'une troisième horloge de référence, l'erreur de réglage maximale sera de $\pm 0,075 \mu\text{s}$ (7.5×10^{-8} s).

APPENDICE I

Bibliographie

- [1] ANSI T1.801.01-1995, *American National Standard for Telecommunications – Digital Transport of Video Teleconferencing/Video Telephony Signals – Video Test Scenes for Subjective and Objective Performance Assessment.*
- [2] ANSI T1.801.02-1996, *American National Standard for Telecommunications – Digital Transport of Video Teleconferencing/Video Telephony Signals – Performance Terms, Definitions, and Examples.*
- [3] ANSI T1.801.03-1996, *American National Standard for Telecommunications – Digital Transport of One-Way Video Signals – Parameters for Objective Performance Assessment.*
- [4] ANSI T1.504a-1991, *American National Standard for Telecommunications – Packet Switched Data Communication Service – Performance Measurement Methods.*
- [5] ANSI T1.517-1995, *American National Standard for Telecommunications – Performance Parameters and Objectives for Integrated Services Digital Networks.*
- [6] ANSI T1.314-1991, *Video Codec for Audiovisual Services at $p \times 64$ kbits.*
- [7] SMPTE RP 27.3-1989², *Recommended Practice, Specifications for Safe Title Areas, Test Pattern for Television Systems.*
- [8] SMPTE 125M-1992, *SMPTE Standard for Television – Component Video Signal 4:2:2 – Bit-Parallel Digital Interface.*²
- [9] SMPTE 259M-1993, *SMPTE Standard for Television – 10-Bit 4:2:2 Component and 4f_{sc} NTSC Composite Signals – Bit-Parallel Digital Interface.*²
- [10] SMPTE 170M-1994, *SMPTE Standard for Television – Composite Analog Video Signal – NTSC for Studio Applications*².
- [11] Contribution UIT-T COM 12-75-E, "Visual Channel Delay and Frame Rate Measurement – Initial Results with a Prototype System," *Union internationale des télécommunications – Secteur de la normalisation des télécommunications, Période d'études 1993-1996, Commission d'études 12, Contribution normale, source AT&T, mars 1996.*
- [12] NETRAVALI (A.N.) et HASKELL (B.G.), *Digital Pictures: Représentation et Compression, Plenum Publishing Corporation, New York, NY, 1988.*

² Disponible auprès de la Société des ingénieurs en images animées et télévision (SMPTE), 595 West Hartsdale Ave. White Plains, NY, 10607.

APPENDICE II

Symboles mathématiques et conventions adoptées

V	séquence de trames vidéo adjacentes à l'entrée d'une voie
V'	séquence de trames vidéo adjacentes à la sortie d'une voie
V'(m)	trame vidéo de sortie au temps T'(m)
V'(i,j,m)	pixel de luminance (i,j) dans une trame vidéo de sortie m au temps T'(m)
V*(i,j,m)	pixel de luminance en sortie avant application des facteurs de correction
g	facteur de correction du gain
l	facteur de correction de différence de niveau
h	facteur de correction de décalage horizontal
v	facteur de correction de décalage vertical
z	facteur de correction de dimension de trame
A	séquence d'unités audio adjacentes à l'entrée de voie
A'	séquence d'unités audio adjacentes à la sortie de voie
D	séquence d'unités de données adjacentes à l'entrée de voie
D'	séquence d'unités de données adjacentes à la sortie de voie
T	flux du "média de temporisation" à l'entrée
T'	flux du "média de temporisation" à la sortie
T _P	ensemble des marqueurs temporels associés aux unités de présentation générale à l'entrée
T' _P	ensemble des marqueurs temporels associés aux unités de présentation à la sortie
T' _{P(m)}	valeur de temporisation (marqueur temporel) associée à l'unité de présentation m d'un flux média général (à la sortie)
T' _{P(m-1)}	valeur de temporisation de l'unité de présentation qui précède l'unité m
T _A	ensemble des marqueurs temporels associés aux trames audio d'entrée
T' _A	ensemble des marqueurs temporels associés aux trames audio de sortie
T' _{A(m)}	valeur de temporisation (marqueur temporel) associée à la trame audio de sortie m
t _{P(m)}	temps de transmission dans une voie pour l'unité de présentation m
t _P	ensemble de temps de transmission dans une voie mesurés pour un flux média
b _{P(m)}	durée entre arrivées pour l'unité de présentation m
b _P	ensemble des durées entre arrivées mesurées pour un flux média
f _{P(m)}	débit de trames élémentaire pour l'unité de présentation m
M[V'(m), V'(m-1)]	variance (MSE) entre deux trames adjacentes
M[V'(m), V(n)]	variance MSE entre une trame d'entrée et une trame de sortie
K _S =(I _{max} -I _{min} + 1) × (J _{max} -J _{min} + 1)	nombre total de pixels dans la sous-région spatiale pour la variance MSE

PSNR	rapport signal sur bruit de crête calculé avec le niveau vidéo de crête, V_{peak}
v'	ensemble des valeurs de variance MSE pour des trames de sortie adjacentes dans une séquence vidéo
c	ensemble des valeurs de variance MSE comparant une trame de sortie à la séquence d'entrée
c_v	ensemble des valeurs de variance MSE comparant une trame vidéo de sortie à la séquence d'entrée
a'	ensemble des valeurs de comparaison pour des trames de sortie adjacentes dans un flux audio
d'	ensemble des valeurs de comparaison pour des trames de sortie adjacentes dans un flux de données
N'	bruit d'acquisition en sortie étalonné
C_P	valeur de comparaison pour l'unité de présentation qui correspond au mieux à une unité active donnée
$O_{AV}(m,n)$	décalage temporel entre trames audio et vidéo associées à l'entrée
$O'_{AV}(p,q)$	décalage temporel entre trames audio et vidéo associées à la sortie
$S'_{AV}(m,n)$	dérive temporelle entre trames audio et vidéo associées à la sortie, due au système/à la voie de transmission

Variables utilisées dans le paragraphe 7

B	facteur de réduction de la largeur de bande et facteur de sous-échantillonnage
coarse_delay	temps de transmission tel que mesuré par la phase approchée
cross_corr	matrice temporaire contenant les valeurs d'intercorrélations finales
cross_corr(i)	$i^{\text{ème}}$ élément de la matrice cross_corr
cross_corr_s	version lissée de la matrice cross_corr dans la phase approchée
cross_corr_s(i)	$i^{\text{ème}}$ élément de la matrice cross_corr_s
delay	valeur finale de sortie du mesurage en deux phases du temps de transmission
fine_delay	temps de transmission tel que mesuré par la phase précise
fine_delay_k	$k^{\text{ème}}$ mesurage du temps de transmission précis
L1	nombre d'échantillons audio d'entrée mesurés
L2	nombre d'échantillons audio après sous-échantillonnage
location	emplacement de mesurage en phase précise
n1	nombre de mesurages effectués en phase précise
n2	nombre de mesurages retenus après le premier contrôle en phase précise
n3	nombre de mesurages retenus après le deuxième contrôle en phase précise
n4	nombre de mesurages retenus après le troisième contrôle en phase précise
ref	matrice des échantillons audio prélevés à l'entrée de voie
ref(i)	$i^{\text{ème}}$ élément de la matrice ref

ref_temp	matrice d'enregistrement temporaire pour les échantillons d'entrée de voie traités
ref_temp(i)	$i^{\text{ème}}$ élément de la matrice ref_temp
ref_temp_i	matrice d'enregistrement temporaire pour les échantillons d'entrée de voie traités
ref_temp_i(j)	$j^{\text{ème}}$ élément de la matrice ref_temp_i
sample_rate	fréquence de numérisation des échantillons à l'entrée et à la sortie de la voie
spread	dispersion dans le sous-ensemble final des mesures de temps de transmission précis
test	matrice des échantillons audio extraits de la sortie de voie
test(i)	$i^{\text{ème}}$ élément de la matrice test
test_temp	matrice d'enregistrement temporaire des échantillons audio de sortie de voie traités
test_temp(i)	$i^{\text{ème}}$ élément de la matrice test_temp

Variables utilisées dans le paragraphe 9

Z	longueur d'une trame de données en bits
D(n)	trame de données d'entrée n
D(i,n)	bit i dans la trame de données d'entrée n
d	décalage en bits entre trames de données d'entrée et de sortie lors de la détermination de la correspondance

Variables utilisées dans le paragraphe 10

MTIE,ns	erreur maximale sur la durée (MTIE) exprimée en nanosecondes
S	intervalle d'observation pour les mesures d'erreur MTIE

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
Série Z	Langages de programmation