

Signal vidéo numérique

par **Guy BALESTRAT**

Responsable de l'option Audiovisuel et Multimédia à l'École nationale supérieure des télécommunications (ENST)

1. Numérisation du signal vidéo.....	TE 5 330 - 2
1.1 Structures d'échantillonnage.....	— 2
1.2 Fréquences d'échantillonnage.....	— 8
2. Quantification	— 10
2.1 Luminance.....	— 11
2.2 Chrominance.....	— 11
3. Format 4:2:2	— 11
3.1 Fréquences d'échantillonnage.....	— 11
3.2 Structures d'échantillonnage.....	— 11
3.3 Quantification.....	— 11
4. Interface parallèle de l'UER	— 14
4.1 Signaux vidéo	— 14
4.2 Synchronisations	— 15
4.3 Interface physique	— 16
5. Sérialisation du format 4:2:2.....	— 16
6. Autres formats.....	— 16
7. Conclusion	— 17
Pour en savoir plus	Doc. TE 5 330

La numérisation du signal vidéo s'est développée dans les années 1970 pour aboutir à la publication en 1982 de l'avis 601 du Comité consultatif international des radiocommunications (CCIR), devenu en 1995 la recommandation UIT-R BT.601-5 sur le format 4:2:2. Il est légitime de se demander les raisons pour lesquelles le signal vidéo a été numérisé puisque la qualité du signal analogique est remarquable.

Le numérique présente plusieurs avantages. Pour l'enregistrement, par exemple, le bruit augmente à chaque génération en analogique, ce qui n'est pas le cas en numérique. La qualité reste constante en transmission. Dans ces deux cas, il n'y a pas de dégradation. De plus, le numérique permet de nombreux effets spéciaux et corrections colorimétriques, qui devaient être faits à la source auparavant. Par exemple, depuis que les incrustations se font en numérique, le « bleu » a disparu des cheveux du présentateur de la météo.

Le numérique a d'abord trouvé des applications ponctuelles comme le TBC (Time Base Corrector) pour les magnétoscopes, puis il a été introduit dans les chaînes analogiques (exemple du mélangeur truqueur numérique de Thomson, commercialisé avec entrées et sorties analogiques), ce qui a permis de profiter de certains des avantages du numérique sans avoir à changer entièrement d'équipement. Le principe des îlots numériques dans une chaîne analogique faisait intervenir un nombre important de conversions analogique-numérique et numérique-analogique, avec les problèmes afférents à ces conversions. Pour pallier cela, la chaîne de traitement doit être entièrement numérique, ce qui a été réalisé par le format 4:2:2 pour la production, la diffusion étant encore majoritairement analogique avec les systèmes NTSC, PAL et SECAM.

Avec trois codages analogiques pour la diffusion, identifiés à deux standards différents (625 lignes 25 images par seconde avec les systèmes PAL et SECAM et 525 lignes 30 images par seconde avec le système NTSC – 29,97 exactement), le risque était grand de se trouver dans une situation identique pour le numérique. Nous verrons comment cela a été évité.

La première étape de cette uniformisation a été de définir la nature du signal à numériser. En effet, le signal vidéo couleur peut se présenter sous forme « composite » comme le NTSC, le PAL ou le SECAM, ou bien en « composantes », c'est-à-dire un signal de luminance et deux signaux de différences de couleurs. C'est cette dernière solution, beaucoup plus souple, qui a été retenue. Dès lors, il a fallu définir le rapport de qualité à établir entre les informations de luminance et celles des composantes de couleur.

Cet article présente successivement les caractéristiques de la numérisation du signal vidéo puis les spécificités des recommandations.

1. Numérisation du signal vidéo

La numérisation d'un signal quelconque consiste dans un premier temps – et après filtrage pour éviter le repliement du spectre – à l'échantillonner.

Ces échantillons analogiques sont ensuite quantifiés sur un certain nombre de niveaux, ce qui donne un train binaire sur n bit en parallèles. Pour des facilités de transport, il s'ensuit une étape de sérialisation, ce qui permet de revenir au câble coaxial comme en analogique.

La numérisation d'un signal vidéo impose des contraintes supplémentaires en raison de son aspect tridimensionnel. Il faut échantillonner le signal, le quantifier mais aussi choisir une structure d'échantillonnage, c'est-à-dire tenir compte aussi bien de l'aspect spatial de l'image que de l'aspect temporel.

Rappelons quelques caractéristiques du signal vidéo. Nous évoluons dans un espace à quatre dimensions, trois spatiales et une temporelle. Lorsque la télévision est apparue, une dimension spatiale a disparu puisque l'objectif de la caméra projette le réel sur un plan. Il ne reste que deux dimensions spatiales (horizontale et verticale) et la dimension temporelle. Ce signal tridimensionnel a alors été transformé en un signal monodimensionnel de la forme $u = f(t)$. Pour ce faire, le signal a été discrétisé dans le plan temporel (comme au cinéma) en choisissant 25 ou 30 images par seconde suivant les systèmes et un entrelacement d'ordre deux, ce qui donne 50 ou 60 trames par seconde. Puis, sur le plan spatial, la dimension verticale a été discrétisée avec 625 ou 525 lignes par image.

Nota : initialement, d'autres formats existaient, comme le format français de 819 lignes, encore exploité en noir et blanc à la fin des années 1970.

Ainsi, la seule dimension continue est la dimension horizontale. Numériser un signal vidéo consiste donc dans un premier temps à l'échantillonner horizontalement et dans un deuxième temps à quantifier les échantillons ainsi obtenus. Mais le signal étant déjà discrétisé en temporel et en vertical, il faut choisir une structure d'échantillonnage adéquate.

Nous allons donc présenter le choix de la structure d'échantillonnage (§ 1.1) et de la fréquence d'échantillonnage (§ 1.2), puis nous aborderons la quantification du signal (§ 2).

1.1 Structures d'échantillonnage

L'échantillonnage d'un signal de télévision doit répondre à deux critères :

- la fréquence d'échantillonnage doit être égale à au moins deux fois la plus grande fréquence à transmettre du signal à numériser (critère de Nyquist) ;
- les échantillons doivent être prélevés en des points bien définis, formant une « structure d'échantillonnage ».

La structure d'échantillonnage est donc l'organisation des échantillons à la fois dans le plan image (écran de visualisation) et dans le temps (séquence animée). Si la dimension horizontale, continue, était la seule considérée, l'application de Shannon [1] déterminerait la fréquence d'échantillonnage à adopter et le problème s'en trouverait résolu. Malheureusement, les traitements à effectuer sur le signal vidéo sont aussi bien spatiaux que temporels, d'où l'importance du positionnement de ces échantillons. De plus, la vision étant isotrope, il convient de choisir des points qui ne privilégient pas une dimension spatiale par rapport à une autre.

Des tests de qualité ont permis de définir plusieurs contraintes au système. La position des échantillons doit être fixe dans l'image, ce qui implique que la fréquence d'échantillonnage doit être liée à la fréquence de balayage du système. De plus, la structure ainsi créée peut se présenter sous diverses formes en fonction du rapport qui existe entre la fréquence d'échantillonnage et la fréquence ligne. Trois structures ont été étudiées :

- structure orthogonale (§ 1.1.2) ;
- structure quinconce ligne (§ 1.1.3) ;
- structure quinconce trame (§ 1.1.4).

1.1.1 Paramètres

La structure d'échantillonnage dépend des paramètres suivants (figure 1a) :

- x_h est l'espace horizontal entre deux échantillons. Ce paramètre est donc directement lié à la fréquence d'échantillonnage, il correspond à l'échantillonnage d'un signal monodimensionnel ;

— x_ℓ est le décalage des échantillons en horizontal d'une ligne à la suivante ;

— x_t est le décalage des échantillons d'une trame à la suivante. Bien évidemment, puisqu'il s'agit d'un système entrelacé spatialement, ces deux lignes sont décalées verticalement ;

— x_i est le décalage d'une image à la suivante.

Les valeurs couramment attribuées aux paramètres x_i , x_ℓ et x_t sont 0 ou $x_h/2$, ce qui conduit à la structure orthogonale et aux structures quinconces.

1.1.2 Structure orthogonale

Les échantillons sont prélevés suivant une structure orthogonale, c'est-à-dire que les points sont alignés verticalement d'une ligne à la suivante. De plus, les échantillons sont prélevés à des positions fixes dans l'image d'une trame à la suivante et d'une image à la suivante (figure 1b).

L'échantillonnage de l'image engendre certains défauts (§ 1.1.6), notamment sur les transitions. Une transition verticale est simplement décalée vers la droite de l'image, ce qui n'est pas réellement gênant puisqu'il n'y a pas de déformation. Mais en ce qui concerne les transitions obliques, il y a apparition d'« escaliers » (cf. haubans des bateaux sur la figure 4b).

1.1.3 Structure quinconce ligne

Tout comme pour la structure orthogonale, dans la structure quinconce ligne stabilisée, les échantillons sont prélevés à des positions fixes d'une image à la suivante, mais en quinconce d'une ligne à la suivante. Ces échantillons donnent une meilleure définition que la structure orthogonale, le pas horizontal étant ramené à $x_\ell = x_h/2$ (figure 1c).

L'échantillonnage de l'image engendre aussi des défauts (§ 1.1.6). Une transition verticale est simplement décalée vers la droite de l'image mais plus ou moins suivant les lignes, ce qui donne un aspect en dents de scie (chaque dent correspond à deux lignes de deux trames successives). Les transitions obliques s'en trouvent améliorées par rapport à la structure orthogonale. La définition de l'image est elle aussi améliorée (cf. figure 4c).

1.1.4 Structure quinconce trame

Les échantillons sont prélevés à des positions fixes dans la trame d'une image à la suivante. Ces échantillons sont de plus prélevés suivant une structure orthogonale sur chaque trame et ce sont ces deux structures qui sont en quinconce (figure 1d). Le principal inconvénient de cette structure est une certaine activité temporelle dans l'image numérisée. Par contre, au niveau spatial, la définition est bien meilleure (cf. figure 4d). En effet, le principe du quinconce réduit le pas horizontal puisque $x_\ell = x_h/2$.

1.1.5 Restitution des transitions

Contours horizontaux

L'échantillonnage du signal n'intervient pas sur les contours horizontaux. En effet, le signal vidéo est échantillonné en vertical en 525 ou en 625 lignes. Les contours horizontaux sont donc restitués comme en analogique, quelle que soit la structure d'échantillonnage choisie.

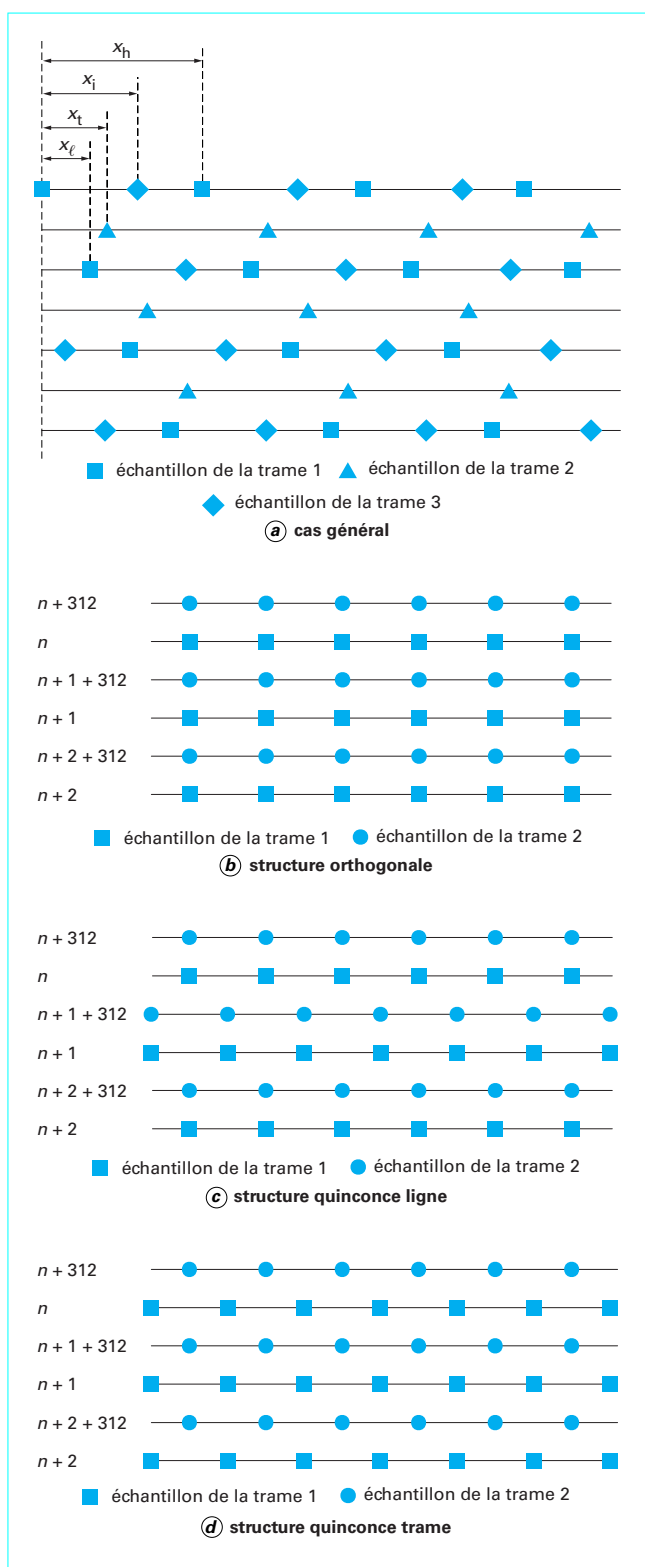


Figure 1 – Structures d'échantillonnage dans un système à 625 lignes

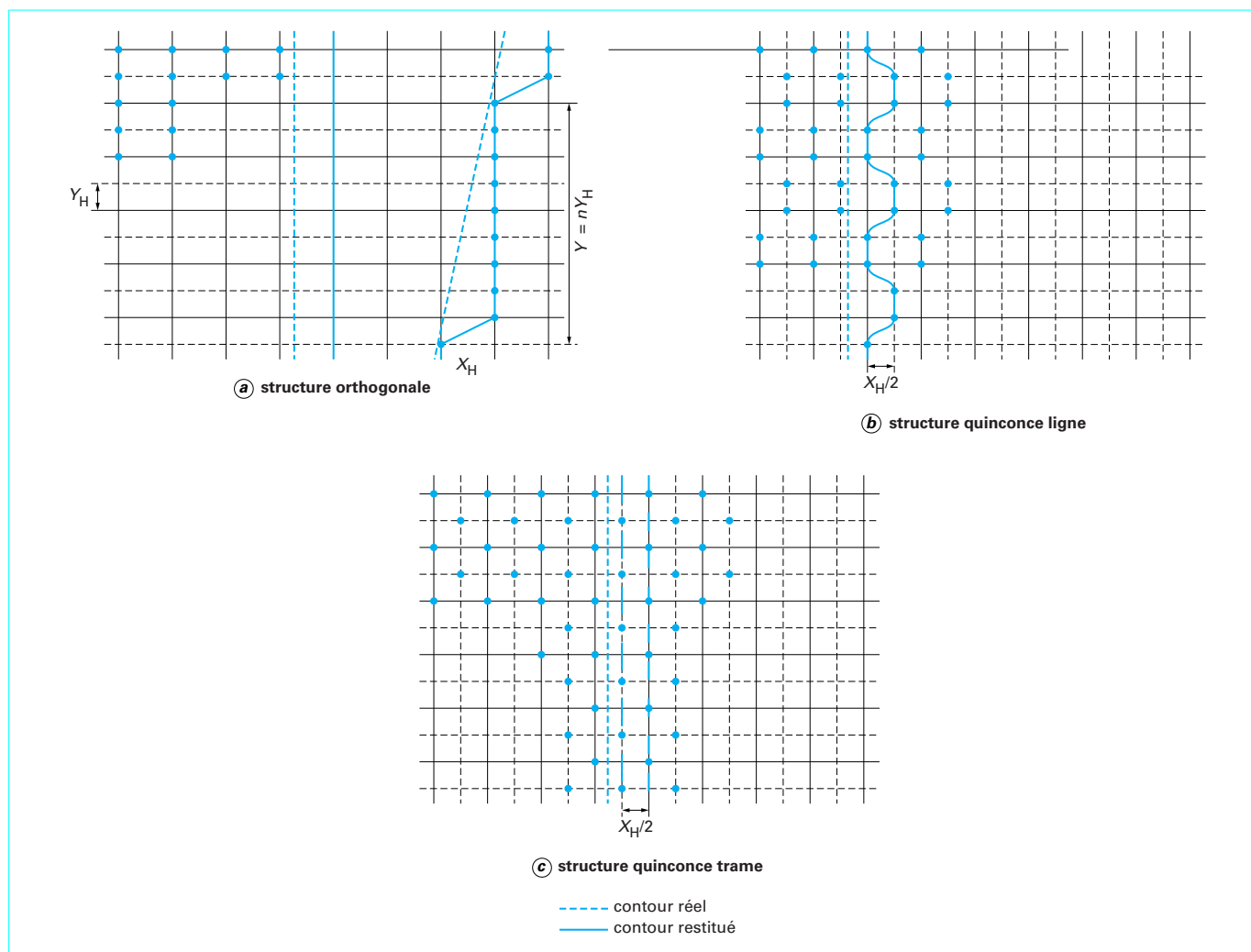


Figure 2 – Restitution des contours verticaux et quasi verticaux

■ Contours verticaux

Sont considérés aussi bien les contours verticaux que les contours quasi verticaux car dans une image naturelle, il est très peu probable de trouver un contour parfaitement vertical. Une maille carrée sur une trame est utilisée pour les figures, d'où une fréquence d'échantillonnage de 7,387 MHz :

$$288 \times 4/3 = 384 \text{ points par ligne}$$

ce qui donne

$$(52 \cdot 10^{-6}/384)^{-1} = 7,387 \cdot 10^6 \text{ Hz}$$

avec 288 lignes utiles dans une trame en numérique,
4/3 le format de l'image,
52 μ s la durée d'une ligne.

● La *structure orthogonale* ne présente pas de défaut visible pour un contour vertical. La transition subit un retard qui peut aller jusqu'à X_H (pas horizontal) pour un contour quasi vertical (figure 2a). Le pas vertical est $Y = nY_H$ où Y_H représente l'espace entre deux lignes. Plus le contour tend vers la verticale et plus n tend vers l'infini (ce qui donne le résultat du contour horizontal avec $n = 0$).

● La *structure quinconce ligne* (figure 2b) donne un contour en créneau avec une profondeur ramenée à $X_H/2$, le retard de la transition étant lui aussi de $X_H/2$ au maximum.

● La *structure quinconce trame* (figure 2c) donne également un contour en créneau avec une profondeur ramenée à $X_H/2$, le retard de la transition étant lui aussi de $X_H/2$ au maximum. Mais une importante différence avec la structure quinconce ligne apparaît à la visualisation. En effet, deux trames échantillonnées en orthogonal sont entrelacées et la position quinconce crée une activité temporelle assez gênante pour les faibles valeurs de la fréquence d'échantillonnage. Pour s'en rendre compte, il suffit de comparer la sensation perçue à celle d'un arrêt sur image avec un magnétoscope lors d'un mouvement rapide de l'image.

■ Contours obliques

Sont considérés les contours obliques inclinés de 45° et les contours à pratiquement 45° . La même structure que pour les contours verticaux est utilisée pour les figures (échantillonnage à 7,387 MHz).

● La *structure orthogonale* (figure 3a) restitue un contour en escalier avec une profondeur de $X_H/2\sqrt{2}$ et un pas de $X_H\sqrt{2}$.

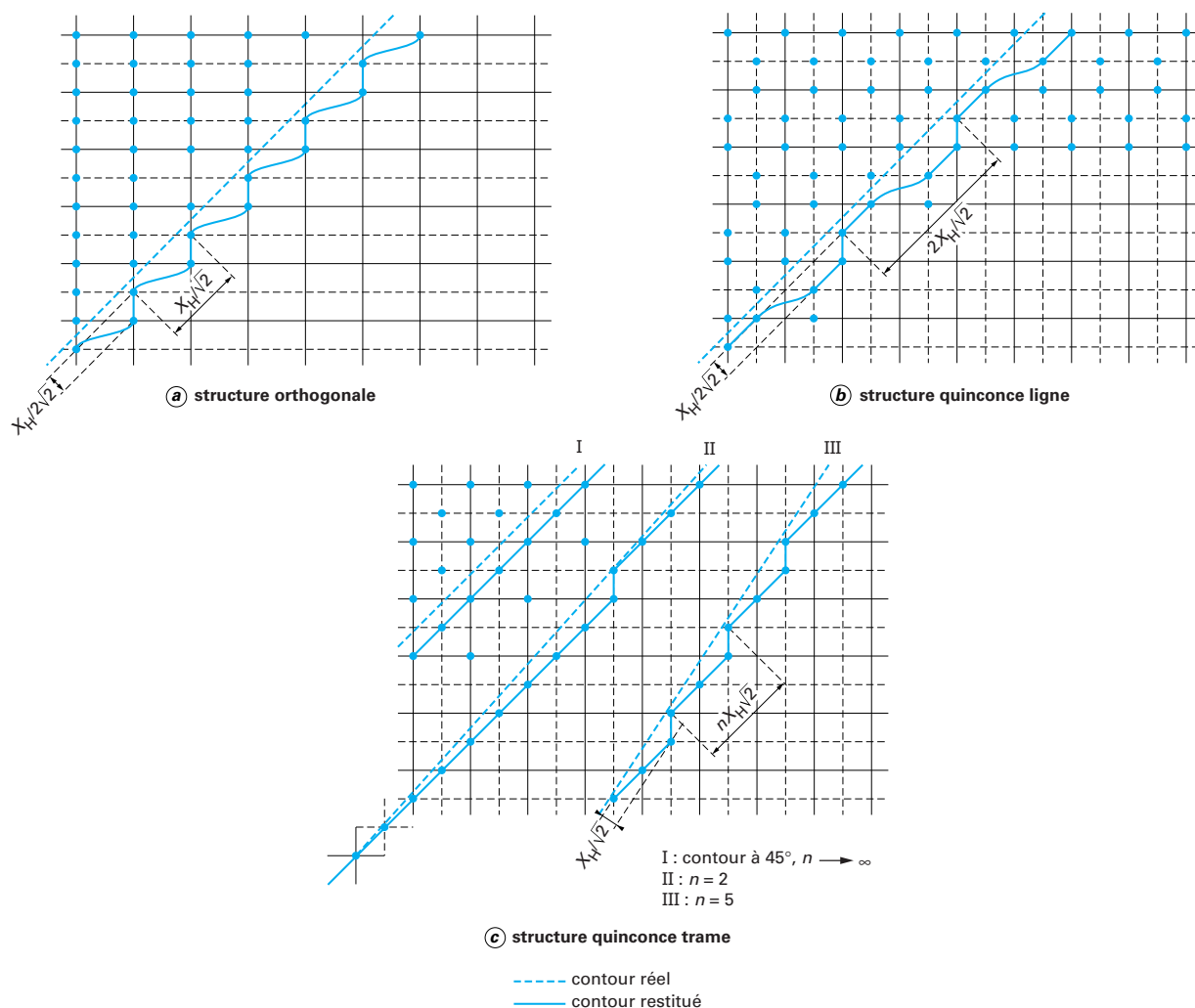


Figure 3 – Restitution des contours obliques

● La *structure quinconce ligne* (figure 3b) restitue un contour en escalier avec une période double de celle de la structure orthogonale, une profondeur de $X_H/2\sqrt{2}$ et un pas de $2X_H\sqrt{2}$.

● La *structure quinconce trame* (figure 3c) restitue a priori une transition correcte avec un léger retard pour 45° . En réalité, comme pour le contour vertical, il y a deux trames entrelacées, chacune étant échantillonnée en orthogonal. Il y a donc deux escaliers qui s'entrelacent, ce qui divise le pas par deux.

Dans le cas d'un contour pratiquement à 45° , est restitué un contour en escalier avec une profondeur de $X_H/\sqrt{2}$ et un pas de $nX_H\sqrt{2}$. Plus le contour tend vers 45° et plus n tend vers l'infini.

1.1.6 Visualisation des défauts

La figure 4 (pages 6 et 7) permet d'observer les défauts relatifs à chaque structure d'échantillonnage. L'image originale (figure 4a) est très intéressante pour l'échantillonnage du signal vidéo car les haubans sont diversement orientés.

Les contours en escaliers des transitions obliques sont observés pour la structure d'échantillonnage orthogonale (figure 4b).

En ce qui concerne la structure d'échantillonnage quinconce trame, chaque trame est échantillonnée en structure orthogonale et les deux structures sont placées en quinconce, ce qui donne sur le plan spatial un lissage des transitions obliques (figure 4). L'activité temporelle n'est malheureusement pas visible.

La structure quinconce ligne donne quant à elle un contour en dents de scie, particulièrement visibles sur les transitions telles que les haubans verticaux (figure 4c).

1.1.7 Choix de la structure

Ces structures d'échantillonnage donnent des résultats équivalents avec une fréquence d'échantillonnage supérieure à 10 MHz (figure 5). Le choix se porte donc sur la structure la plus simple à mettre en œuvre : la **structure d'échantillonnage orthogonale**. Il ne



(a) image originale



(b) échantillonnage en orthogonal à 4 MHz



© échantillonnage en quinconce ligne à 4 MHz



ⓓ échantillonnage en quinconce trame à 4 MHz

Figure 4 – Bateaux et phare. Visualisation des défauts des structures d'échantillonnage

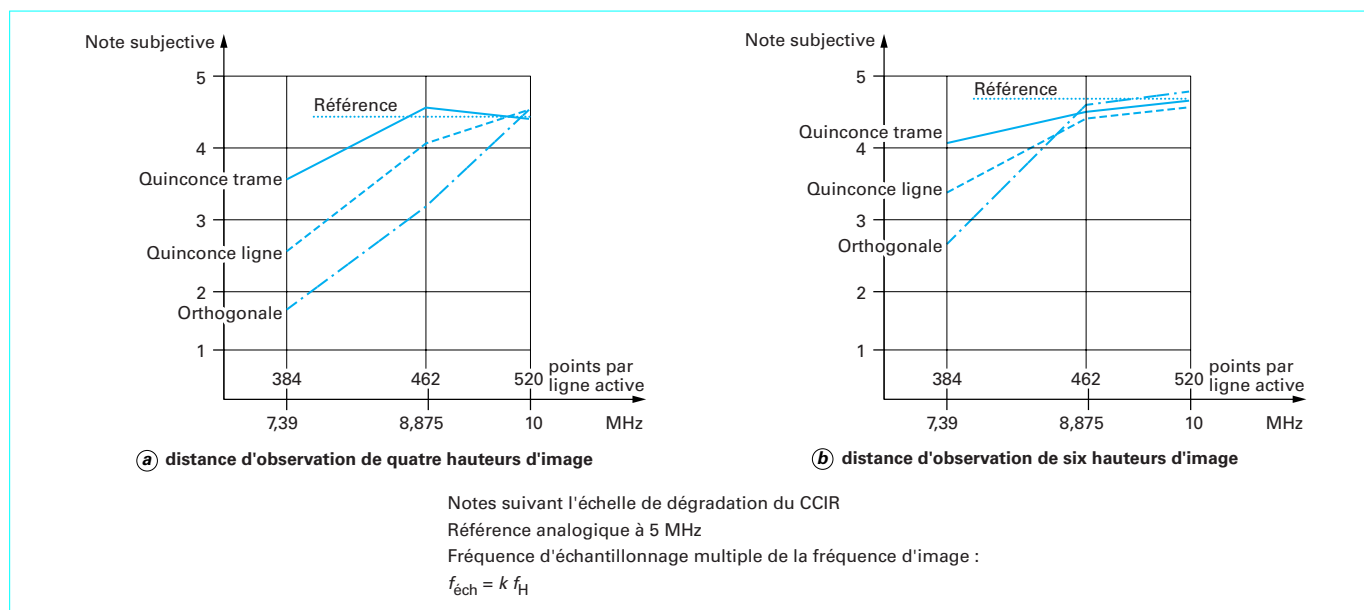


Figure 5 – Essais subjectifs sur l'échantillonnage de la luminance (d'après l'UIT-R)

faut en effet tenir compte ni de la parité ligne ($x_l = 0$) ni de la parité trame ($x_t = 0$).

1.2 Fréquences d'échantillonnage

Le choix de la fréquence d'échantillonnage doit répondre au critère de Nyquist (§ 1.1). Cette fréquence est directement liée d'une part à la **fréquence de balayage ligne** et d'autre part à la **largeur de bande du signal**. Les paragraphes 1.2.1 et 1.2.2 présentent le calcul des fréquences d'échantillonnage luminance et chrominance, donc le lien avec les fréquences lignes.

Auparavant, faisons un rappel sur le **repliement de spectre**. Selon Shannon [1], l'échantillonnage d'un signal doit se faire à une fréquence au moins deux fois la largeur de bande du signal à numériser. Si ce principe n'est pas respecté, il se produit un repliement de spectre (figure 6b). En présence d'un signal à trois dimensions, la représentation est similaire à celles de la figure 7, le troisième axe étant l'axe des temps. Les défauts visibles lors de la restitution du signal vidéo sont de deux types : temporel et spatial. Les défauts temporels sont bien connus et se voient régulièrement dans les films, que ce soit au cinéma ou à la télévision – c'est, par exemple, la roue de charrette qui tourne alternativement dans un sens puis dans l'autre quand la vitesse de rotation varie légèrement autour de la fréquence d'échantillonnage temporelle. Le repliement de spectre spatial est visible sur la figure 6. Le papier peint est une structure de raies verticales dont la fréquence croît en allant vers la gauche. La figure 6b a été obtenue par échantillonnage à 5 MHz et filtrage à 5 MHz du signal vidéo, ce qui permet de replier le spectre sur lui-même. Les artefacts dus au repliement de spectre sont surtout visibles sur la partie gauche de l'image (moirage).

1.2.1 Signal de luminance

Appliquons à la luminance les contraintes évoquées précédemment (§ 1.2).

La relation avec la **fréquence de balayage ligne** donne deux fréquences différentes puisqu'il existe un système européen à 625 lignes et 50 Hz, et un autre américain à 525 lignes et 59,94 Hz.

Nota : les Américains avaient du noir et blanc à 60 Hz. Lorsqu'il s'est agi de passer à la couleur, la fréquence de la sous-porteuse pour donner une bonne compatibilité directe a été choisie comme un multiple entier impair de la demi-fréquence ligne. Malheureusement, l'harmonique 286 de la fréquence ligne est très légèrement supérieure à la fréquence de la porteuse son et tombe dans la bande audio. Aussi, une nouvelle fréquence ligne a été déterminée de façon à être elle-même un sous-multiple de la porteuse son afin de ne pas interférer avec cette porteuse.

Fréquence ligne en télévision noir et blanc : $f_h = 525 \times 30 = 15\,750$ Hz

Fréquence de l'harmonique 286 de la fréquence ligne : $f_{286} = 15\,750 \times 286 = 4\,504\,500$ Hz

Fréquence ligne en télévision couleur : $f_h = 4\,500\,000/286 = 15\,734$ Hz

Il faut donc choisir une fréquence directement liée aux deux fréquences de balayage horizontal européenne et américaine, l'américaine étant prise à la valeur calculée ci-dessus de 15 734 Hz.

La **largeur de bande du signal** oblige à appliquer le « principe de Shannon » (§ 1.2). En conséquence, il faut trouver un compromis car la largeur de bande adoptée par les différents pays varie de 4,2 MHz à 6 MHz (4,2 MHz aux États-Unis, 5 MHz en Allemagne, 5,5 MHz au Royaume-Uni et 6 MHz en France), d'où un nombre élevé de fréquences d'échantillonnage.

Pour une fois, la solution retenue est unique et la fréquence a été déterminée de façon à respecter la structure d'échantillonnage orthogonale aussi bien pour les systèmes à 50 Hz que pour les pays à 60 Hz.

Dans les années 1970, différentes fréquences ont été envisagées car le principe de numérisation du signal composite était préféré, d'où des fréquences à deux, trois ou quatre fois la fréquence de sous-porteuse. Vers 1980, les Européens préconisaient un système de télévision numérique avec une fréquence d'échantillonnage à 12 MHz pour la luminance et 4 MHz pour les composantes de chrominance ; les Américains préféraient une fréquence d'échantillonnage d'environ 14,3 MHz, c'est-à-dire quatre fois la fréquence de la sous-porteuse du NTSC (3,58 MHz). Les expériences ont montré qu'il était possible de choisir dans une plage de fréquence de 12 à 14 MHz avec un optimum entre 13 et 13,5 MHz. Par un heureux hasard, la fréquence de 13,5 MHz possède une propriété tout à fait intéressante :



(a) image originale



(b) image filtrée à 5 MHz et échantillonnée à 5 MHz

Figure 6 – Clown. Repliement spatial de spectre

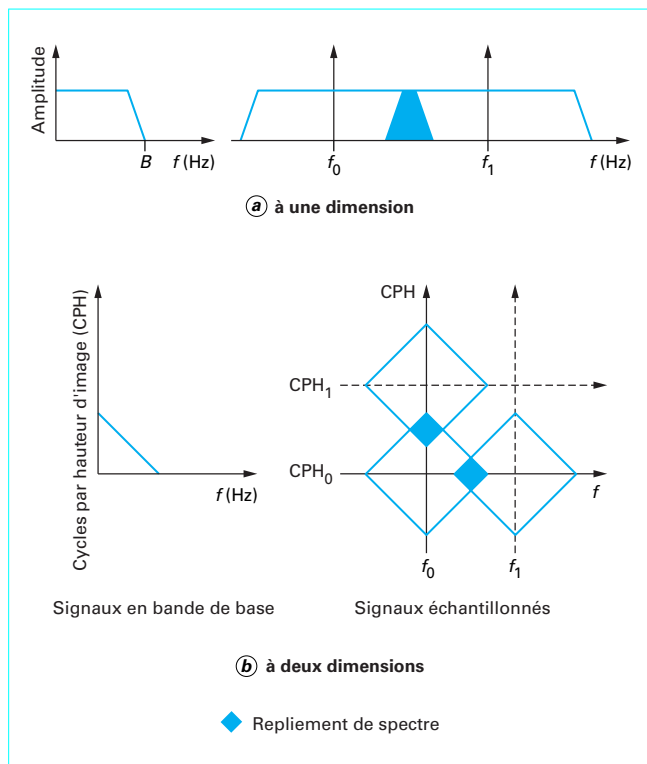


Figure 7 – Repliement de spectre

Fréquence ligne du système 625/50/2 : 25 images par seconde d'où :

$$f_{H(625)} = 625 \times 25 = 15\,625 \text{ Hz}$$

Fréquence ligne du système 525/60/2 : 30 images par seconde en télévision noir et blanc mais le choix de la sous-porteuse du NTSC a imposé aux Américains une fréquence ligne de 15 734 Hz, d'où une fréquence trame de 59,94 Hz, soit 29,97 images par seconde :

$$f_{H(525)} = 4\,500\,000/286 = 15\,734 \text{ Hz}$$

4,5 MHz est la fréquence de la porteuse son aux États-Unis (norme M) et est aussi un multiple de 15 625 Hz :

$$f_{H(625)} = 15\,625 \text{ Hz} \Rightarrow 4\,500\,000/15\,625 = 288$$

4,5 MHz est donc un multiple commun aux deux fréquences horizontales européenne et américaine mais en réalité, le PPCM (plus petit commun multiple) est 2,25 MHz. Il suffit donc de trouver un multiple de ce PPCM compris entre 12 et 14 MHz : c'est 13,5 MHz, soit six fois 2,25 MHz ou trois fois 4,5 MHz.

La fréquence retenue pour l'échantillonnage de la luminance est **13,5 MHz**. Afin que la structure d'échantillonnage soit fixe, il faut asservir cette fréquence sur le signal vidéo ce qui ne pose pas de problème car elle a été choisie pour cela.

1.2.2 Signal de chrominance

Lors de l'avènement de la télévision en couleur, il a fallu coder les informations à transmettre. Pour ce faire, une réduction de la bande de la chrominance au quart de celle de la luminance a été appliquée

(voir Bedford dans [2] et le mélange des fréquences élevées). Pour la numérisation des signaux D_R et D_B , le même principe est utilisé : la fréquence d'échantillonnage déterminée pour la luminance est divisée par deux sans occasionner de problème de repliement de spectre.

Pour chaque composante de couleur, la fréquence d'échantillonnage est de **6,75 MHz**.

Nota : le rapport 2 entre les deux fréquences se retrouve dans la dénomination 4:2:2 correspondant aux échantillons de Y , C_R et C_B .

2. Quantification

Le signal est maintenant échantillonné. L'étape suivante de la numérisation est celle de la quantification. Le débit binaire γ est directement lié.

Pour quantifier un signal, il faut définir une loi et une échelle de quantification. L'échelle de quantification a été fixée sur 256 niveaux, soit 8 bit, et la loi a été choisie linéaire. Le choix de mots de 8 bit peut paraître évident de par les composants disponibles et l'habitude de travailler sur 8 bit, mais il convient de le justifier. Pour le noir et blanc, 7 bit de quantification suffisent ; pour la couleur – à laquelle l'œil est moins sensible – 6 bit de quantification suffisent, d'où le choix de travailler sur des octets. Cela satisfait à la fois les contraintes de la luminance et celles de la chrominance.

De plus, un meilleur rapport signal/bruit est obtenu en choisissant une loi de compression logarithmique et le choix de cette loi linéaire peut paraître anormal. Mais le signal vidéo est déjà corrigé en γ . En réalité, la loi de quantification n'est donc pas linéaire mais en $1/\gamma$ (voir encadré) et permet de conserver un bon rapport signal/bruit (S/B) aux faibles niveaux de luminance.

Rappelons que $S/B = (6,02n + 1,76)$ où n est le nombre de bit.

Exemple :	$n = 6$	$S/B = 37,88 \text{ dB}$
	$n = 7$	$S/B = 43,90 \text{ dB}$
	$n = 8$	$S/B = 49,92 \text{ dB}$

Nota : en se reportant aux codages analogiques – pour lesquels le bruit est moins gênant en chrominance qu'en luminance avec un écart de 8 dB – notons que 6 ou 7 bit revient pratiquement au même, avec environ 6 dB d'écart en rapport signal/bruit.

Loi en $1/\gamma$

Weber [3] a démontré que la sensibilité de l'œil est logarithmique grâce à une expérience qui consiste à faire varier la luminance d'une plage uniforme de ΔL sur une autre plage uniforme et de luminance constante L . Il a ainsi démontré que plus la luminance est élevée, plus l'écart ΔL doit être important pour que l'œil perçoive cette différence, ce qui peut s'exprimer par la fraction de Weber :

$$\Delta L / L = C^{te}$$

ce qui correspond à la dérivée d'un logarithme.

Dans la loi en $1/\gamma$, la valeur prise par γ est habituellement de 2,2, ce qui donne :

$$E' = E^{1/2,2} = E^{0,45}$$

La figure 8 montre que la réponse logarithmique et celle en $1/\gamma$ ont alors la même allure.

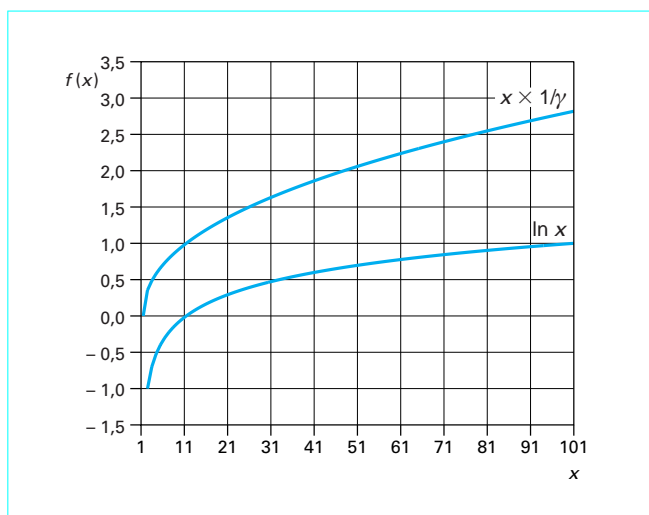


Figure 8 – Comparaison des corrections logarithmique et en $1/\gamma$

2.1 Luminance

L'effet du nombre de bits de quantification est visible sur la photographie des bateaux. En effet, le ciel et ses nuances demandent plus de précision qu'une scène quelconque. Sur l'image originale (figure 9a, page 12), le ciel est pratiquement uni, à l'exception de quelques nuages qui ne présentent qu'une faible différence de luminance par rapport à l'ensemble. Aussi, avec peu de bits de quantification, le pas de quantification est important et les petites différences ne sont pas prises en compte. Avec 3 bit, il y a huit niveaux de quantification, soit sept pas de quantification, ce qui donne pour un signal vidéo variant de 0 à 700 mV :

$$\text{pas de quantification} = 700/7 = 100 \text{ mV}$$

Sur l'image quantifiée sur 3 bit (figure 9b), le contenu est à peu près restitué mais il manque les nuances. Pour la luminance, 7 bit sont en fait nécessaires à la quantification. Cela donne 128 niveaux, soit 127 pas de quantification :

$$\text{pas de quantification} = 700/127 = 5,51 \text{ mV}$$

Avec 8 bit de quantification, soit 256 niveaux et 255 pas de quantification, on a :

$$\text{pas de quantification} = 700/255 = 2,75 \text{ mV}$$

En prenant 8 bit au lieu de 7 bit, le pas de quantification est diminué de moitié. Un nombre de **8 bit** est parfait pour l'enregistrement et la visualisation, mais pour certaines applications où de nombreux calculs sont nécessaires (par exemple, les effets spéciaux), il est préférable d'augmenter ce nombre et de passer à 10 bit de quantification. Cela laisse une marge de bits de poids faibles pour les calculs importants.

2.2 Chrominance

De la même manière, l'influence du nombre de bits de quantification pour la chrominance est étudiée. Dans ce cas, seulement **6 bit** sont nécessaires pour la quantification.

3. Format 4:2:2

3.1 Fréquences d'échantillonnage

Rappelons les valeurs normalisées des fréquences d'échantillonnages déterminées précédemment :

- pour la luminance Y : 13,5 MHz ;
- pour chaque composante de couleur C_R et C_B : 6,75 MHz.

L'échantillonnage ne peut se faire dans de bonnes conditions sans un **filtrage** adapté à l'application recherchée. Il est donc nécessaire de connaître les filtres qui vont le précéder. La recommandation UIT-R BT.601-5 [Doc. TE 5 330] donne les caractéristiques des filtres du signal de luminance et de ceux des signaux de chrominance.

Les résultats d'essais subjectifs sur images naturelles ont donné une note de 4,5/5 pour 4,5 MHz en luminance et plus de 1,5 MHz en couleur. Le choix s'est donc porté sur les filtres de la figure 10.

Les gabarits respectent la règle de Shannon avec des fréquences de 6,75 MHz et 3,375 MHz qui sont la moitié de 13,5 MHz et 6,75 MHz respectivement.

3.2 Structures d'échantillonnage

La structure d'échantillonnage orthogonale a été retenue aussi bien pour la luminance que pour la chrominance (§ 1.1.7). De plus, la fréquence d'échantillonnage de la chrominance est la moitié de celle de la luminance. Il faut donc positionner les structures d'échantillonnage des trois signaux les uns par rapport aux autres. Le nombre d'échantillons de luminance étant le plus important, la chrominance se place par rapport à ceux-ci. Le choix est de prendre les deux informations de chrominance sur le même échantillon. Il y aura donc un échantillon sur deux en luminance et un sur deux en luminance et chrominance (figure 11). Le premier échantillon en haut et à gauche de l'écran a les trois composantes.

3.3 Quantification

La quantification se fait sur 8 bit par composante ; de plus, elle est uniforme (le signal étant corrigé en γ à la source) (§ 2).

Il y a donc 256 niveaux de quantification équidistants disponibles de 0 à 255. Les niveaux 0 et 255 sont réservés à la signalisation des données de synchronisation. De plus, afin de préserver des marges de fonctionnement, 220 niveaux seulement sont utilisés pour la luminance. Le niveau du noir étant placé à 16, celui du blanc est donc à 235.

En ce qui concerne la chrominance, les valeurs sont aussi bien positives que négatives. La valeur la plus faible utilisée est toujours le niveau 16, les signaux étant positifs ou négatifs, le niveau attribué au zéro est le niveau 128 : cela donne $128 - 16 = 112$ niveaux de quantification. Il en faut aussi 112 pour les valeurs positives, soit au total 225 niveaux utilisés ($112 \times 2 + 1 = 225$), ce qui donne comme niveau maximum 239. Les composantes de chrominance varient donc de 16 pour $-0,5$ à 239 pour $+0,5$ avec 128 pour 0.

Ainsi, dans le format 4:2:2, tous les échantillons (Y, C_R et C_B) sont quantifiés sur 8 bit, C_R et C_B étant les signaux de différences de couleurs D_R et D_B pondérés. Nous allons déterminer les coefficients de pondération qui normalisent D_R et D_B .

Rappelons que les signaux Y, D_R et D_B s'expriment en fonction des composantes rouge, verte et bleue de la façon suivante :

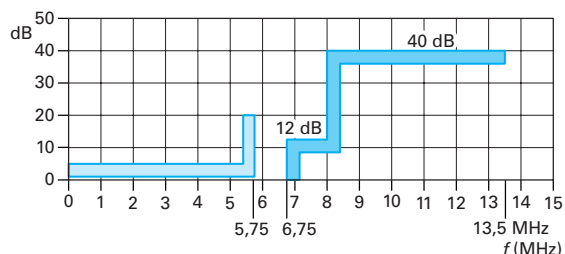


(a) image originale

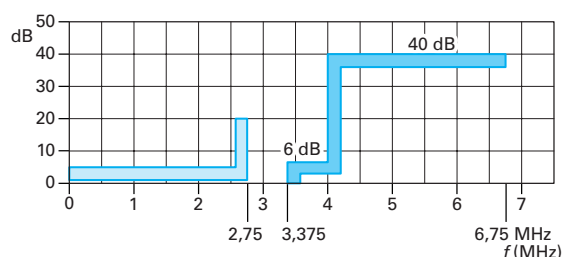


(b) image quantifiée en luminance sur 3 bit

Figure 9 – Bateaux et phare. Quantification



a) luminance



b) chrominance

Figure 10 – Gabarits des filtres de luminance et de chrominance (d'après l'UIT-R)

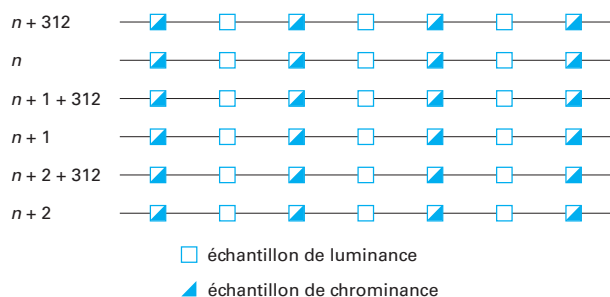


Figure 11 – Structure d'échantillonnage du format 4:2:2

– signal de luminance Y :

$$E'_Y = 0,299E'_R + 0,587E'_G + 0,114E'_B$$

– signaux de différences de couleurs $R - Y$ et $B - Y$:

$$D'_R = E'_R - E'_Y = E'_R - (0,299E'_R + 0,587E'_G + 0,114E'_B)$$

$$D'_R = E'_R - E'_Y = 0,701E'_R - 0,587E'_G - 0,114E'_B$$

$$D'_B = E'_B - E'_Y = E'_B - (0,299E'_R + 0,587E'_G + 0,114E'_B)$$

$$D'_B = E'_B - E'_Y = -0,299E'_R - 0,587E'_G + 0,886E'_B$$

Le tableau 1 donne les valeurs correspondant à une mire de barre à 100 %.

Tableau 1 – Valeurs des composantes des couleurs de la mire de barre à 100 %

	E'_G (1)	E'_R (1)	E'_B (1)	E'_Y	$E'_R - E'_Y$	$E'_B - E'_Y$
Blanc	1	1	1	1	0	0
Jaune	1	1	0	0,886	0,114	-0,886
Cyan	1	0	1	0,701	-0,701	0,299
Vert	1	0	0	0,587	-0,587	-0,587
Magenta	0	1	1	0,413	0,587	0,587
Rouge	0	1	0	0,299	0,701	-0,299
Bleu	0	0	1	0,114	-0,114	0,886
Noir	0	0	0	0	0	0

(1) 1 : signal présent ; 0 : signal absent

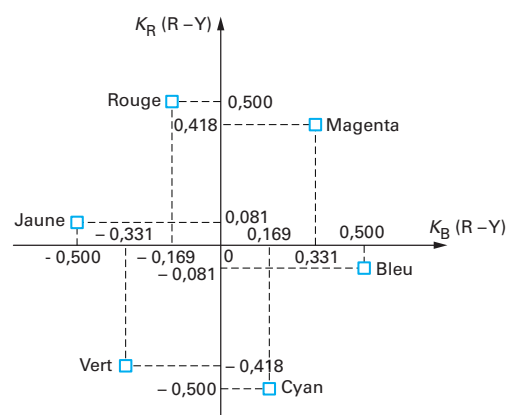


Figure 12 – Projection des couleurs de la mire de barre dans le plan de luminance nulle

Les signaux renormalisés C_R et C_B sont :

$$C_R = K_R(E'_{C_R} - E'_Y)$$

$$C_B = K_B(E'_{C_B} - E'_Y)$$

La projection des couleurs de la mire de barre dans le plan de luminance nulle suivant les axes renormalisés est représentée figure 12. Cette représentation n'est pas habituelle puisque normalement, les axes de référence sont E'_U et E'_V avec :

$$E'_U = \frac{1}{2,03} (E'_B - E'_Y)$$

$$E'_V = \frac{1}{1,14} (E'_R - E'_Y)$$

Nota : E est le signal électrique et E' le signal électrique corrigé en γ .

Pour déterminer les coefficients de pondération de $R - Y$ et de $B - Y$, on se place dans le plan de luminance nulle. Les valeurs de Y sont comprises entre 0 et 1 alors que celles de D_R sont comprises entre -0,701 (pour le cyan) +0,701 (pour le rouge), D_B ayant des

valeurs comprises entre $-0,886$ (pour le jaune) et $+0,886$ (pour le bleu). Pour ramener les signaux à l'unité, c'est-à-dire entre $-0,5$ et $+0,5$, on calcule les coefficients K_R (à partir du rouge) et K_B (à partir du bleu) :

$$E'_{C_R} = K_R(E'_R - E'_Y) \Rightarrow K_R = \frac{E'_{C_R}}{E'_R - E'_Y}$$

$$E'_{C_B} = K_B(E'_B - E'_Y) \Rightarrow K_B = \frac{E'_{C_B}}{E'_B - E'_Y}$$

Cela donne pour les coefficients de pondération, sachant que pour le rouge $E'_{D_R} = 0,701$ et que pour le bleu $E'_{D_B} = 0,886$:

$$K_R = \frac{E'_{C_R}}{E'_R - E'_Y} = \frac{0,5}{0,701} = 0,713$$

$$K_B = \frac{E'_{C_B}}{E'_B - E'_Y} = \frac{0,5}{0,886} = 0,564$$

$$E'_{C_R} = K_R(E'_R - E'_Y) = 0,500E'_R - 0,419E'_G - 0,081E'_B$$

$$E'_{C_B} = K_B(E'_B - E'_Y) = -0,169E'_B - 0,331E'_G + 0,500E'_R$$

En travaillant entièrement en numérique (la recommandation 601 donne des formules similaires pour le calcul en analogique), on obtient :

— primaires (expressions en numérique) :

$$E'_{R_D} = p(219E'_R) + 16$$

$$E'_{G_D} = p(219E'_G) + 16$$

$$E'_{B_D} = p(219E'_B) + 16$$

— composantes (expressions en numérique) :

$$Y = \frac{77}{256} E'_{R_D} + \frac{150}{256} E'_{G_D} + \frac{29}{256} E'_{B_D}$$

$$C_R = \frac{131}{256} E'_{R_D} - \frac{110}{256} E'_{G_D} - \frac{21}{256} E'_{B_D} + 128$$

$$C_B = -\frac{44}{256} E'_{R_D} - \frac{87}{256} E'_{G_D} + \frac{131}{256} E'_{B_D} + 128$$

avec p la partie entière.

Les termes en ' sont les signaux corrigés en γ .

4. Interface parallèle de l'UER

UER : Union européenne de radiodiffusion, plus connue sous le nom d'Eurovision.

L'interface parallèle est une interconnexion unidirectionnelle à neuf paires entre deux dispositifs. Il y a huit paires pour les données et une pour l'horloge synchrone. Mais en réalité, il y a onze paires car il y en a deux en réserve.

Les signaux de données sont :

- les signaux vidéo ;
- les signaux de référence temporelle ;
- les données auxiliaires ;
- les signaux d'identification.

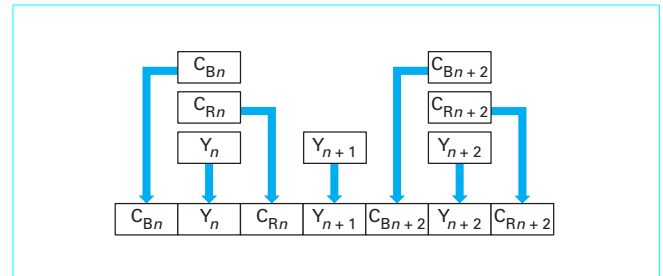


Figure 13 – Réalisation du train à 27 MHz

Ces signaux sont multiplexés temporellement et transmis en codage NRZ (non retour à zéro).

4.1 Signaux vidéo

On transmet les signaux Y , C_R et C_B de la recommandation UIT-R BT.601-5 de l'UIT.

Seulement 254 mots sont utilisables, les mots 0 et 255 étant réservés à l'identification des données. Le multiplexage commence par celui des octets de luminance et de chrominance qui se présentent en un train à 13,5 MHz pour la luminance Y et de deux trains à 6,75 MHz pour les informations de chrominance C_R et C_B .

Le signal multiplexé ainsi réalisé est à 27 MHz, ce qui donne le train numérique de la figure 13.

Jusqu'à présent, nous n'avons traité que le signal correspondant à l'image utile. Tout comme en analogique, il faut des synchronisations et, bien évidemment, celles-ci sont présentes dans le signal sous forme d'octets que nous allons détailler afin de mieux appréhender le fonctionnement global.

Tout comme pour les signaux analogiques, les mots de synchronisation sont insérés dans les suppressions. Le problème est de les repérer afin de les extraire et de les interpréter. En ce qui concerne le repérage – les valeurs 0 et 255 étant réservés à cet effet –, ce sont les mots 00 et FF en hexadécimal qui sont utilisés. Il y a deux séquences différentes suivant que l'on se trouve en présence de synchronisations ou de données. Les mots de synchronisation sont précédés de la séquence FF 00 00, alors que les données sont précédées de la séquence 00 FF FF. La synchronisation quant à elle est constituée d'un octet unique XY, que ce soit en ligne ou en trame. Les mots de synchronisation sont toujours positionnés au même instant caractéristique sur la ligne, et les demi-lignes disparaissent, ce qui a pour conséquence d'avoir deux suppressions trame différentes, l'une de 24 lignes et l'autre de 25 lignes. L'image utile passe donc de 575 lignes utiles à 576 lignes utiles, soit deux trames de 288 lignes utiles. Les échantillons quant à eux sont au nombre de 720 par ligne, alors que le calcul donne 702 échantillons pour la durée de la ligne analogique :

$52 \cdot 10^{-6} \times 13,5 \cdot 10^6 = 702$ points par ligne utile avec une durée de ligne utile en Europe de 52 μ s et une fréquence d'échantillonnage de 13,5 MHz.

La ligne numérique est donc un peu plus grande que la ligne analogique. Le nombre théorique d'échantillons par ligne pour une image 4/3 est : $576 \times 4/3 = 768$.

Or, une ligne utile n'a que 702 échantillons, ce qui a pour conséquence que les échantillons (ou pixels) télévision sont rectangulaires. Pour avoir des pixels carrés comme sur les ordinateurs, il en faudrait 768 par ligne.

Les pixels sont rectangulaires dans le rapport $768/702 = 1,09$.

Ayant déterminé le nombre de points par ligne utile, la figure 14 donne la position des échantillons de luminance et de chrominance par rapport à la synchronisation ligne, ainsi que les suppressions ligne analogique et ligne numérique. L'origine du temps n'est pas la même en analogique et en numérique. En effet, en analogique, c'est le front avant de synchronisation, en l'occurrence le front de descente, qui marque l'origine des temps, alors qu'en numérique, l'instant initial est au début de la ligne utile. La ligne vidéo compte 864 échantillons.

4.2 Synchronisations

Les synchronisations correspondent à une séquence de quatre octets. Ces mots de synchronisation ne sont pas placés au même instant caractéristique que les fronts de synchronisation analogique. Ils sont en réalité positionnés par rapport à l'image utile. Pour le signal numérique, les instants caractéristiques de l'image vidéo, qui correspondent au début et à la fin de la ligne utile et à la position de la ligne par rapport aux suppressions et à la trame, sont repérés. Trois bits sont nécessaires :

- F donne la parité trame. F est à 0 pour la trame impaire et à 1 pour la trame paire ;
- V précise la position par rapport à l'image utile et à la suppression. V est à 0 pendant l'image utile et à 1 pendant les suppressions ;
- H donne le début et la fin de ligne utile. H est à 0 pour indiquer le début de ligne utile et à 1 pour indiquer la fin de ligne utile. Ce sont les SAV (*start of active video*) et EAV (*end of active video*).

À cela s'ajoutent 4 bit de code correcteur d'erreurs et le huitième bit inutilisé est positionné à 1 en permanence (tableau 2).

Tableau 2 – Octets de synchronisation

2 ⁷	F	V	H	P ₃	P ₂	P ₁	P ₀
1	0	0	0	0	0	0	0
1	0	0	1	1	1	0	1
1	0	1	0	1	0	1	1
1	0	1	1	0	1	1	0
1	1	0	0	0	1	1	1
1	1	0	1	1	0	1	0
1	1	1	0	1	1	0	0
1	1	1	1	0	0	0	1

Exemple : ligne utile de la trame 1

Le tableau 2 donne pour les SAV et EAV, en utilisant la notation hexadécimale :

— séquence SAV : FF-00-00-80. La séquence FF-00-00 annonce un XY correspondant à un mot de synchronisation. L'octet XY a pour valeur 80_h. En effet, le bit de poids fort est à 1, F à 0 pour la trame 1, V à 0 pour l'image utile et H à 0 pour le début de ligne utile. Les bits correcteurs d'erreurs prennent la valeur 0_h ;

— séquence EAV : FF-00-00-9D. De la même manière que pour le SAV, la séquence FF-00-00 annonce un XY correspondant à un mot de synchronisation. Cette fois, l'octet XY a pour valeur 9D_h. En effet, le bit de poids fort est à 1, F à 0 pour la trame 1, V à 0 pour l'image utile et H à 1 pour le début de ligne utile. Les bits correcteurs d'erreurs prennent la valeur D_h.

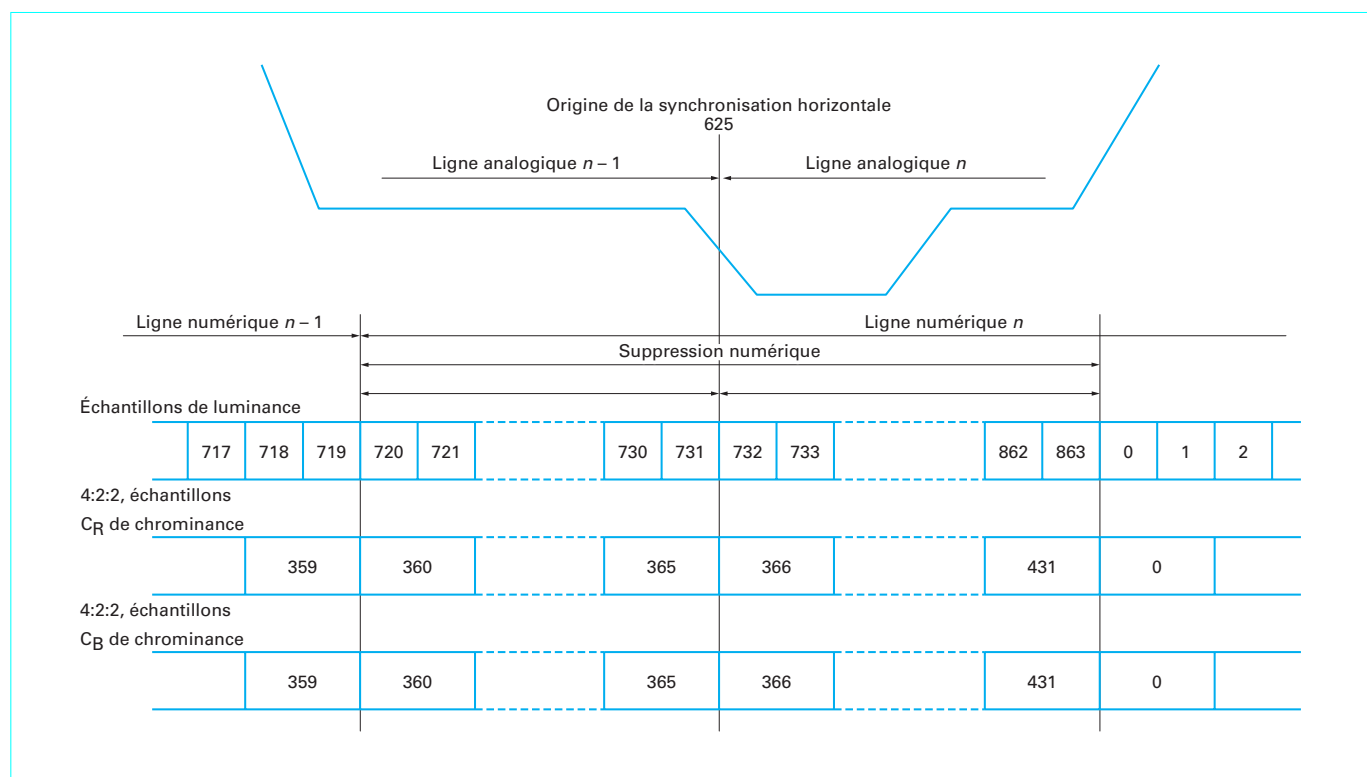


Figure 14 – Position des échantillons par rapport à la ligne analogique dans un système à 625 lignes (d'après l'UIT-R)

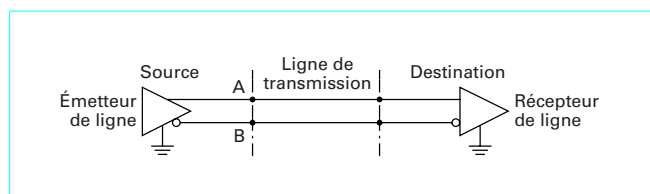


Figure 15 – Liaison numérique 4:2:2 (d'après l'UIT-R)

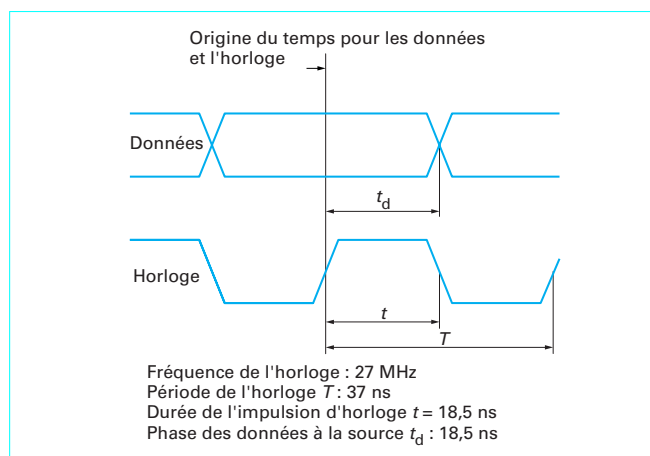


Figure 16 – Position de l'horloge par rapport aux données (d'après l'UIT-R)

D'après la figure 14, les échantillons de luminance 720 et 721 ainsi que les échantillons 360 de chrominance sont utilisés pour la séquence du EAV. Pour la séquence du SAV, les échantillons de luminance 862 et 863 ainsi que les échantillons 431 de chrominance sont utilisés.

4.3 Interface physique

Les échanges se font grâce à des câbles de onze paires. En effet, l'interface doit transporter l'horloge et 8 bit, mais il prévu une extension à 10 bit. Les signaux transitent en ECL (*emitter coupler logic*) ou compatible ECL et sont envoyés sur des paires torsadées en direct et complément (figure 15).

Les fronts d'horloge doivent arriver pendant que les données sont présentes, comme sur la figure 16.

Les connecteurs sont des connecteurs cannon à 25 points et l'impédance des câbles est de 110 Ω , le brochage est donné dans le tableau 3. Comme ils sont prévus pour une extension à 10 bit, les bits de poids faibles sont mis à 0.

5. Sérialisation du format 4:2:2

Le train binaire de la norme 4:2:2 est un train de mots de 8 bit dont le débit numérique se calcule :

- pour la luminance : 13,5 MHz sur 8 bit, soit $13,5 \times 8 = 108$ MHz ;

Tableau 3 – Brochage du connecteur 25 points

Contacts	Signaux	Contacts	Signaux
1	Horloge	14	Horloge
2	Masse A	15	Masse B
3	Donnée9	16	Donnée9
4	Donnée8	17	Donnée8
5	Donnée7	18	Donnée7
6	Donnée6	19	Donnée6
7	Donnée5	20	Donnée5
8	Donnée4	21	Donnée4
9	Donnée3	22	Donnée3
10	Donnée2	23	Donnée2
11	Donnée1	24	Donnée1
12	Donnée0	25	Donnée0
13	Blindage		

— pour chaque composante C_R et C_B : 6,75 MHz sur 8 bit, soit $6,75 \times 8 = 54$ MHz.

Cela donne, après multiplexage des informations, un train binaire à 27 MHz sur 8 bit, d'où un débit numérique de 216 Mbit/s.

Pour le transporter en série sur un seul support, il faut assurer la mise en forme spectrale par un codage supplémentaire de la synchronisation et permettre la récupération d'horloge. Le CCIR a recommandé un codage sur 9 bit, d'où un débit de 243 Mbit/s, alors que les Japonais ont réalisé des équipements avec un codage sur 10 bit, soit 270 Mbit/s. Ces équipements ayant été commercialisés, c'est le codage qui a été retenu.

Ce débit est bien trop grand pour s'inscrire dans un des canaux de débit normalisé à 34 ou 144 Mbit/s. Nous ne citerons que le procédé de réduction de débit qui consiste à quantifier non pas l'échantillon mais la différence entre la valeur réelle de cet échantillon et une valeur de « prédiction » déterminée soit en fonction d'autres échantillons de la même trame (critère spatial), soit en fonction d'échantillons des trames précédentes (critère temporel). Ce signal de différence peut alors être codé sur un nombre beaucoup plus réduit de bits et il est possible d'espérer une réduction sensible du débit binaire. Il ne s'agit là que d'un exemple et d'autres procédés de réduction de débit existent mais, dans tous les cas, il est nécessaire de mettre en œuvre des procédés de détection et de correction des erreurs de transmission. Un système à 34 Mbit/s a d'ailleurs été normalisé et est utilisé dans certaines liaisons satellites et même dans certaines liaisons par faisceaux hertziens. Une autre norme est utilisée pour la diffusion de la télévision, la norme MPEG2 DVB. Il est cependant important de souligner que l'UER transmet la vidéo par satellite à 34 Mbit/s.

6. Autres formats

■ Format 4:4:4

Ce format correspond à la numérisation des trois primaires à la fréquence de 13,5 MHz chacune. C'est d'ailleurs le format utilisé pour le calcul de la luminance et des composantes de chrominance lorsque celui-ci est fait en numérique. Il peut donner le format 4:2:2.

■ Format 4:4:4:4

Ce format est identique au 4:4:4 avec en plus un plan de découpe. Il est utilisé notamment en post-production et peut également donner le 4:2:2.

■ Format 4:2:0

Ce format s'obtient à partir du 4:2:2. La luminance est inchangée, on conserve l'ensemble des échantillons. En ce qui concerne la chrominance, le même principe s'applique en vertical qu'en horizontal. Un échantillon sur deux en chrominance est retenu ; il s'obtient par filtrage entre deux lignes consécutives. Ce format est celui utilisé par MPEG-2 MP@ML pour les systèmes à 50 Hz et 625 lignes.

■ Format 4:1:1

Ce format s'obtient à partir du 4:2:2. La luminance est inchangée et, contrairement au 4:2:0, pour lequel la définition verticale de la chrominance est divisée par deux, c'est ici la définition horizontale de la chrominance qui est divisée par deux. Ce principe se rapproche de celui des codages analogiques (NTSC, PAL, SECAM) pour lesquels la largeur de bande de la chrominance est le quart de celle de la luminance. Ce format est celui utilisé par MPEG-2 MP@ML pour les systèmes à 60 Hz et 525 lignes.

■ Format 4:2:2:4

Ce format est identique au 4:2:2 et présente, comme le 4:4:4:4, un plan de découpe. Il est utilisé en postproduction.

■ Format SIF

Le SIF (*source intermediate format*), tout comme le format 4:2:2, est supporté par les standards à 525 lignes et 29,97 images par seconde, et 625 lignes et 25 images par seconde. Les résolutions sont respectivement de 260×240 et 360×288 . En réalité, la résolution du format 4:2:2 est divisée par 4 (2 dans chacune des deux directions spatiales) ; en temporel, il y a suppression d'une trame sur deux, ce qui en fait un système progressif.

■ Format CIF

Le CIF (*commun intermediate format*) est un format commun au standard américain auquel il emprunte la définition temporelle et au standard européen auquel il emprunte la définition spatiale. Il s'agit donc d'un format à 360×288 et 29,97 images par seconde. C'est le format initial de la visiophonie.

■ Format QCIF

La résolution de ce format est encore réduite, elle reprend celle du format CIF divisée par 4 (2 dans chaque dimension spatiale) ; en temporel, elle divise par 2 ou par 4 celle du CIF (d'où son nom Quart-CIF).

Tous ces formats ne sont pas utilisés pour la télévision et d'autres peuvent encore apparaître. Il est cependant possible d'obtenir à partir du 4:2:2, des formats pour la télévision haute définition. En effet, puisque les Européens doublent le nombre de lignes, il faut également doubler le nombre de points et à terme appliquer un coefficient 4/3. La base étant le 4:2:2, cela donne $720 \times 2 = 1\,440$ points puis, avec application du coefficient, $1\,440 \times 4/3 = 1\,920$ points.

Afin d'établir quelques comparaisons, les tableaux 4 et 5 récapitulent les caractéristiques de ces différents formats.

Nota : ces formats seront repris et développés dans l'article consacré aux systèmes de compression.

Tableau 4 – Résolution des différents formats

Signal	4:2:2	4:2:0	4:1:1
Y	720×576	720×576	720×576
C _R	360×576	360×288	180×576
C _B	360×576	360×288	180×576

Tableau 5 – Comparaison des différents formats

Format	Lignes par image	Points par ligne	Images par seconde
4:2:2	576	720	25
SIF	288	360	25
CIF	288	360	29,97
QCIF	144	180	14,985

7. Conclusion

Cette étude de la numérisation du signal n'est certes pas exhaustive, mais elle donne les principes à connaître impérativement et qu'il convient de résumer ici.

La **fréquence d'échantillonnage** est commune aux standards 525/59,94 et 625/50 :

- 13,5 MHz pour la luminance ;
- 6,75 MHz pour la chrominance.

La **structure d'échantillonnage** orthogonale a été choisie.

Le signal corrigé en γ est **quantifié** sur 8 bit par composante.

Les problèmes de **repliement de spectre** sont résolus grâce au préfiltrage et au postfiltrage.

La connaissance de ces données permet d'aborder sereinement les principes de la **compression** du signal vidéo numérique, phénomène essentiel pour la télévision.