RAPPORT 1204

SYNCHRONISATION AUTOMATIQUE DES SIGNAUX VIDEO ET AUDIO APRES TRANSMISSION

(Question 47/10)

(1990)

1. Introduction

Chaque fois que les composantes vidéo et audio d'un signal de télévision sont transmises sur des trajets distincts, il peut arriver que leurs temps de transmission diffèrent. Si leur différence devient telle que le spectateur perçoit une séparation entre ce qu'il voit et ce qu'il entend, son plaisir peut être gâché. Ceci a été étudié en détail en Australie (voir [CCIR, 1986-90a]).

On peut apprécier la différence de temps admissible d'après les normes adoptées pour le cinéma, où l'image peut précéder le son de deux images (83 ms) au maximum ou être en retard sur lui d'une image (42 ms) au maximum. La raison de cette asymétrie des différences de temps admissibles est que l'arrivée du son après celle de l'image est chose habituelle, contrairement à un son précédant la vision.

Des synchroniseurs d'image sont de plus en plus utilisés dans les transmissions à grande distance parce qu'ils assurent une grande souplesse opérationnelle. Néanmoins, de fortes différences de temps de propagation peuvent être facilement introduites si l'on ne prend pas de dispositions pour que le son soit retardé d'une même quantité que la vidéo dans le synchroniseur.

Des différences de temps de propagation sont fréquemment compensées manuellement par un opérateur s'efforçant d'apprécier la correction à faire avant que d'autres observateurs ne la perçoivent, mais c'est un processus difficile. Une méthode est à l'étude, qui permettrait de corriger automatiquement les retards relatifs entre vision et son.

Il est facile de retarder un signal audio sous forme numérique en l'injectant dans un enregistreur à décalage et on peut modifier ce retard en changeant l'adresse d'où le signal de sortie est pris. La Figure 1 montre que cela augmente seulement le retard de la composante audio.

Un retard similaire du signal vidéo est plus difficile à réaliser. Mais un tel retard est à peine nécessaire d'une part parce qu'on tolère beaucoup plus facilement le retard du son sur l'image, d'autre part parce qu'un tel retard est moins vraisemblable. La seule cause prévisible d'un retard accru sur le trajet audio est l'insertion d'un filtre passe-bas numérique FIR à coupure brusque. Un filtre de ce type pourrait, par exemple, être nécessaire pour convertir la fréquence d'échantillonnage audio de 48 kHz en studio à 32 kHz pour la transmission, mais cela ne procure qu'un retard de 5,3 ms.

2. Génération de codes temporels

Aucune méthode de correction n'est proposée pour le son sous forme analogique, une méthode ayant déjà été proposée à cette fin [Cooper, 1988]. Cependant, il est prévu que les signaux son et image seront presque entièrement transmis sous forme numérique à l'avenir. Pour autant que les signaux son et vidéo soient transmis dans le même train de bits, il ne saurait se poser aucun problème de synchronisation. Mais, lorsque le signal vidéo est traité, que ce soit pour la synchronisation de trame ou pour toute autre fin, il faut appliquer une méthode de rétablissement du synchronisme.

Selon l'interface audionumérique pour studios de radiodiffusion spécifiée au § 3.5 de la Recommandation 647, les données de la voie de signalisation acheminent, dans les multiplets 18-21, un code binaire de 32 bits qui transmet le numéro du premier échantillon du bloc en cours. Même à la fréquence d'échantillonnage de 48 kHz, les 2^{32} états de ce code peuvent transporter des numéros uniques d'échantillons pendant une période de plus de 24 heures. Ainsi, dans les données de la voie de signalisation de l'interface audio existe une indication de temps, qui est présente dans tous les blocs de 192 échantillons, c'est-à-dire toutes les 6 ms avec une fréquence d'échantillonnage de 32 kHz, ou toute les 4 ms avec une fréquence d'échantillonnage de 48 kHz, avec une précision de 31,25 μs ou de 20,83 μs.

Cependant, l'interface audionumérique est conçue principalement pour transmettre des programmes monophoniques ou stéréophoniques dans un environnement de studio et il est peu probable qu'une telle interface soit utilisée pour la transmission entre studios. De ce fait, l'information de code temporel devrait être transcodée dans le train de bits transmis.

Cela étant, il est proposé qu'un code de Synchronisation Audio/Vision (SAV) soit généré et transmis avec chaque signal: code SAV (A) dans le train de bits audio et SAV(V) dans l'intervalle vertical de la vidéo. On estime que ce procédé nécessiterait 11 bits par bloc, soit un supplément de 1,83 kbit/s avec la fréquence d'échantillonnage 32 kHz ou 2,75 kbit/s avec l'échantillonnage à 48 kHz.

Ces 11 bits décriraient la synchronisation du premier échantillon du bloc actuel, à la milliseconde près. Pour ce faire, on prendrait l'adresse d'échantillon binaire pour chaque bloc déjà présent dans les octets 18 à 21 des données de la voie de signalisation de l'interface audionumérique, et on la diviserait par la fréquence d'échantillonnage en kilohertz, par exemple 32, 44,1 ou 48 selon le cas.

Le quotient binaire ainsi obtenu déterminerait l'instant de début de chaque bloc du signal audionumérique en unités précises de 1 ms pour l'échantillonnage à 32 kHz ou 48 kHz, et avec une erreur maximale de $\pm$ 0,5 ms pour l'échantillonnage à 44,1 kHz. Avec 11 bits, on peut spécifier au total 2 048 états différents, c'est-à-dire que le code temporel aurait un cycle complet de 2 048 secondes, le comptage allant de zéro jusqu'à 2 047 pour revenir ensuite à zéro.

Le code temporel de l'UER, spécifié pour la vidéo, donne seulement un comptage d'images, à savoir une image toutes les 40 ms en PAL. Pour le système NTSC, on pourrait avoir une image toutes les 33,37 ms. Il semble donc impossible d'établir une relation entre le code temporel audio et le code temporel vidéo classique. En conséquence, il est proposé d'insérer, dans chaque intervalle vertical de la vidéo, un signal supplémentaire de 11 bits, semblable à celui du signal audio, avec indication de l'instant (en unités de 1 ms) où il a été inséré. Il faut noter que ce code temporel de Synchronisation Audio/Vidéo est totalement indépendant du code temporel vidéo classique.

Il n'est pas nécessaire de transmettre de l'information sur la fréquence d'échantillonnage dans le code SAV, du fait qu'il a déjà été pourvu à cette fonction dans la division du numéro d'adresse binaire de l'interface audionumérique par la fréquence d'échantillonnage en kHz. Cependant, il est possible de récupérer l'information fréquence d'échantillonnage sur le code SAV, en calculant la différence entre les numéros de code temporel SAV des blocs successifs.

Si cette différence est calculée entre des numéros séparés par deux blocs, l'intervalle de temps sera de 12,0 ms pour l'échantillonnage à 32 kHz, et la différence entre les numéros sera par conséquent de 11, 12 ou 13. Pour l'échantillonnage à 48 kHz, la période sera de 8,0 ms, donnant une différence de 7, 8 ou 9. La logique qui reconnaît l'un ou l'autre des groupes de différences pourra donc identifier la fréquence d'échantillonnage. Une procédure similaire s'applique aux signaux audio échantillonnés à d'autres débits.

Le procédé proposé est tributaire d'une grande stabilité d'horloge dans les codes temporels. La stabilité des signaux de rythme de télévision est spécifiée dans le Rapport 624: $\pm$ 5 Hz sur 4,433 MHz pour le système PAL; celle des signaux audionumériques est spécifiée dans la Recommandation 646: $\pm$ 1 x 10^{-5}. On a admis cette tolérance relativement large afin d'assurer les signaux audio émanant d'équipements portables éloignés.

Toutefois, dans le même document, le point 3 sous "RECOMMANDE" stipule ce qui suit: "lorsqu'un équipement audionumérique fonctionne de façon autonome, la tolérance maximale sur la fréquence d'échantillonnage interne doit être de $\pm$ 1 x 10^{-5}. Lorsque plusieurs équipements audionumériques sont interconnectés, en radiodiffusion sonore ou en télévision, il doit être possible de synchroniser la fréquence d'échantillonnage interne sur une fréquence d'échantillonnage extérieure (par exemple: signaux de synchronisation de télévision, horloge mère de la maison de la radio, horloge de haute précision d'un réseau de télécommunications)".

Le système proposé devrait pouvoir fonctionner de façon satisfaisante avec ces tolérances.

3. Correction des retards

A la réception, chacun des deux codes temporels SAV (qui arrivent à des moments différents) d'un bloc audio ou d'une trame vidéo sert à synchroniser son propre code temporel à fonctionnement continu, comme le montre la Figure 1. Ces deux codes temporels dérivés sont ensuite comparés et leur différence sert de signal de correction pour faire varier le retard audio jusqu'à ce que la différence s'annule.

On pourrait penser que, avec une durée de cycle couvrant 2 048 états, le système serait capable de corriger des différences de temps allant jusqu'à 2,048 secondes, mais la différence à la sortie, N(D), est égale à N(V) - N(A). Tant que la vidéo est en retard sur l'audio, N(D) est positif, mais si le son vient à prendre du retard sur la vidéo, N(D) devient négatif et sa valeur est indiquée par un grand nombre, N(D) + 2 048. Par exemple, si le retard du son est de 100 ms, l'indication sera 1 948.

Cela étant, nous proposons ce qui suit: si la différence entre les deux codes temporels dérivés est comprise entre les nombres décimaux 1 536 et 2 047, on considérera que cette différence est comprise entre -512 et -1, et elle sera par conséquent utilisée pour annuler le retard du son; cette annulation serait accompagnée d'une indication "vidéo en avance sur audio". Le système est donc capable de corriger les erreurs correspondant à une avance du son sur la vidéo; la correction maximum est de 1,535 seconde, c'est-à-dire suffisante pour les plus grandes erreurs prévisibles.

Les sync vidéo fonctionnent de façon asynchrone par rapport au code temporel dérivé de l'audio; de ce fait, les temps indiqués pour la vidéo peuvent être entachés d'une erreur maximale de ± 0,5 ms, et la différence de temps indiquée, N(D), entre audio et vidéo peut aussi varier de cette quantité. Il est proposé, par conséquent, d'introduire dans le système une hystérésis telle que le nombre exprimant la correction du retard, N(C), ne varie pas tant que sa différence avec N(D) ne dépasse pas ± 2 ms et que, une fois la variation opérée, elle se poursuive jusqu'à ce que N(C) = N(D). Cette caractéristique n'est pas représentée dans la Figure 1.

Il est proposé également que les erreurs initiales supérieures à 3 ms soient corrigées immédiatement, mais que les erreurs de ± 2 ms le soient à raison d'une sur 500, c'est-à-dire sur un intervalle de temps d'une seconde.

Le retard variable peut être introduit par l'application de techniques déjà connues.

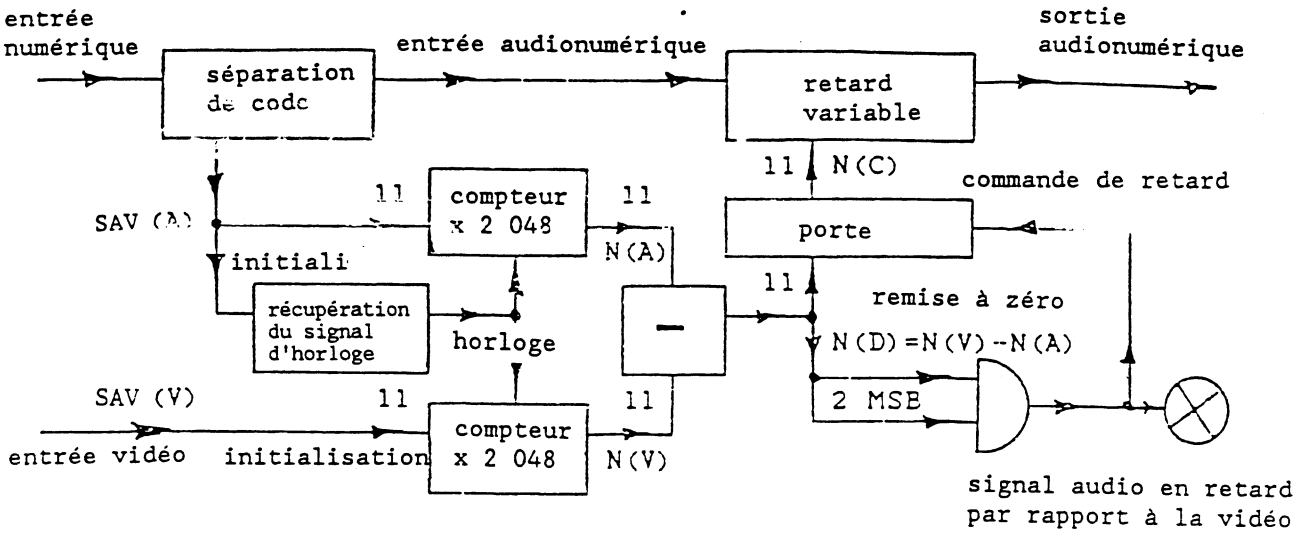

FIGURE 1 - Représentation schématique du synchroniseur audio/vidéo

4. Transmission du code de synchronisation A/V

On trouvera ci-après une méthode possible pour transmettre le code SAV dans chacun des signaux composants.

4.1 Composante son

Dans le document [CCIR, 1986-90b], on propose une méthode grâce à laquelle des signaux radiophoniques de haute qualité, avec une paire stéréo échantillonnée à 48 kHz, peuvent être acheminés dans le RNIS, soit à 1 920 kbit/s (pour accès H12), soit à 1 536 kbit/s (pour accès H11).

Avec la méthode proposée dans le document en question, les données de la voie de signalisation, fournies par l'interface audionumérique et contenant le temps binaire à 32 bits de l'horloge "jour", sont transmises avec une dilatation de temps de 48:1. En d'autres termes, ces données sont transmises en transparence, mais un bloc entier de la voie de signalisation est transmis toutes les 192 ms, au lieu de toutes les 4 ms dans le cas de l'interface studio. Le code de synchronisation A/V à 11 bits peut être dérivé de ces données, comme il a été expliqué plus haut, le seul inconvénient étant le suivant: la détection d'une variation éventuelle du temps de transmission ne peut intervenir qu'après un délai pouvant atteindre 192 ms. Mais il ne semble pas que cela soit important.

Une fois établie la différence de temps entre le code temporel restitué et le code temporel initial - différence qui pourrait atteindre 192 ms - il faut prévoir des moyens pour lui appliquer un ajustement fixe, éventuellement nécessaire, avant de dériver le code SAV. Cette question n'est pas traitée en détail dans [CCIR, 1986-90b], si ce n'est la remarque suivante que l'on trouve au § 10, "Voie de signalisation": "Etant donné qu'on transmet seulement un bloc de données sur 48, les deux compteurs (échantillon local et code d'adresse de l'heure du jour) doivent être incrémentés de la quantité voulue dans le décodeur."

4.2 Composante vidéo

Avant de pouvoir insérer un code à 11 bits dans l'intervalle de suppression trame, il faudra se mettre d'accord sur l'endroit de l'intervalle où le code sera placé. Le meilleur endroit devrait être à peu près le même que dans le code de livraison de programme proposé dans [UER, 1989].

5. Conclusion

Nous pensons que la méthode proposée permet de synchroniser des composantes audio et vidéo transmises sur des trajets différents, en occupant un intervalle minimum dans chacun des deux signaux.

REFERENCES BIBLIOGRAPHIQUES

COOPER, J.C. [septembre 1988] - Video-to-Audio Synchrony Monitoring and Correction - JSMPTE, pages 695 - 698.

UER [1989] - Specification of the domestic video programme delivery control (PDC) EBU SPB 459rév., mars 1989.

Documents du CCIR:

[1986-90] a. 10/315 (Australie); b. GTI CMTT/4-8 (GTI CMTT/4).