

INFORME 1204

SINCRONIZACIÓN AUTOMÁTICA DE LAS SEÑALES DE VÍDEO Y DE AUDIO
DESPUÉS DE LA TRANSMISIÓN

(Cuestión 47/10)

(1990)

1. Introducción

Cuando las componentes de vídeo y de audio de una señal de televisión se transmiten por trayectos separados, existe la posibilidad de que sus tiempos de transmisión sean distintos. Cuando la diferencia es tal que el espectador percibe una separación entre la imagen y el sonido, la apreciación del programa puede resultar dañada. Este asunto se ha estudiado con detalle en Australia y se refleja en [CCIR, 1986-90a].

Puede tomarse como norma del valor de las diferencias de tiempo admisibles, la de las películas cinematográficas, en la que la imagen puede preceder al sonido en no más de dos cuadros (83 ms) o ir por detrás de éste en no más de un cuadro (42 ms). El motivo de la asimetría en estas diferencias de tiempo admisibles es que cuando el sonido llega después de la imagen se produce una experiencia familiar, mientras que la llegada del sonido antes de la imagen no tiene ese carácter.

En las transmisiones de larga distancia se utilizan de forma creciente los sincronizadores de cuadro porque confieren una mayor flexibilidad operacional. Sin embargo, las grandes diferencias de retardo que pueden aparecer fácilmente, a menos que se adopten medidas para retardar el sonido igualmente con la imagen en el sincronizador.

Las diferencias de sonido las compensa a menudo de forma manual un operador que trata de evaluar la corrección antes de que los demás espectadores puedan percibirla, si bien se trata de un proceso difícil. Está en estudio un método mediante el cual puedan corregirse automáticamente los retardos relativos entre la imagen y el sonido.

Una señal de audio en forma digital se retarda fácilmente aplicándola a un registrador de desplazamiento y el retardo puede variarse modificando el punto del que se toma la salida. El esquema que se muestra en la fig. 1 produce únicamente un aumento del retardo para la componente de audio.

Un retardo similar de la señal de vídeo es más difícil de conseguir. No obstante, no hay gran necesidad de efectuar este retardo del vídeo, porque se tolera más fácilmente que el sonido siga a la imagen, y el primer caso es menos probable que ocurra. La única causa previsible de aumento del retardo en el trayecto de la señal de audio, es la inserción de un filtro digital paso bajo de respuesta impulsiva finita con un corte agudo. Un filtro de este tipo puede necesitarse, por ejemplo, para convertir la velocidad de muestreo de audio de 48 kHz en el estudio, a 32 kHz para la transmisión; pero esto solo introduciría un retardo de 5,3 ms.

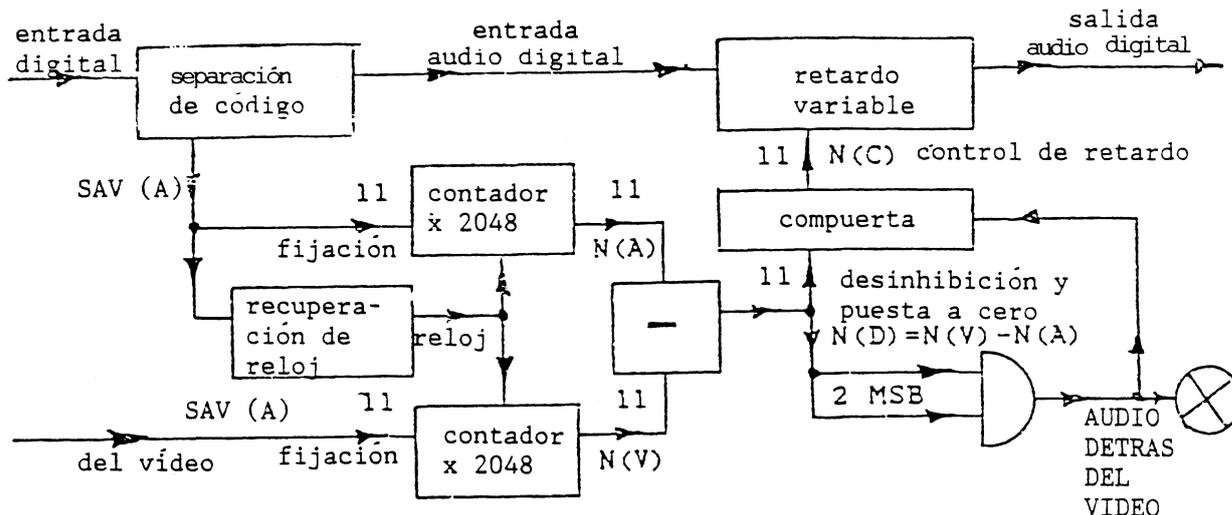


FIGURA 1 - Esquema del sincronizador automático audio/vídeo

2. Generación de códigos temporales

No se propone ningún método de corrección para la señal de sonido en forma analógica porque se ha propuesto un método para este fin (Cooper, 1987). No obstante, puede avanzarse que en el futuro, el sonido y la imagen se transmitirán casi enteramente por medios digitales. Mientras que el sonido y la imagen se transmitan en el mismo tren binario, no deben surgir problemas de sincronismo. Una vez procesada la señal de imagen a través de la sincronización de cuadro o de cualquier otra forma, se necesitará un método de resincronismo.

En el interfaz audio digital para los estudios de radiodifusión que figura en el punto 3.5 de la Recomendación 647, los datos del estado de canal llevan, en los octetos 18-21, un código binario de 32 bits que indica el número de la primera muestra del bloque actual. Con una velocidad de muestreo de 48 kHz, los 2^{32} estados de este código pueden indicar números únicos de muestras durante un periodo de más de 24 horas. Así pues, en los datos de estado de canal del interfaz de audio, se presenta una indicación temporal cada bloque de 192 muestras, es decir, cada 6 ms, con una frecuencia de muestreo de 32 kHz, o de 4 ms con 48 kHz, y con una precisión de 31,25 μ s o de 20,83 μ s.

No obstante, el interfaz digital audio está concebido principalmente para cursar programas monofónicos o estereofónicos en un entorno de estudio y no es probable que se utilice para la transmisión entre estudios. Así pues, toda información de código temporal tendrá que transcodificarse para tomar su sitio en el tren binario transmitido.

Se propone, por tanto, la generación de un código de Sincronización Audio/Vídeo (SAV) que se transmita con cada señal, el código SAV (A) en el tren binario de audio y el código SAV (V) en el intervalo vertical del vídeo. Se estima que esto exigirá 11 bits por bloque, es decir, una velocidad adicional de 1,83 kbit/s con una frecuencia de muestreo de 32 kHz, o de 2,75 kbit/s con un muestreo de 48 kHz.

Estos 11 bits describirían la temporización de la primera muestra del bloque actual, redondeando al milisegundo más próximo. Ello se realizaría tomando la dirección de la muestra binaria para cada bloque ya presente, en los octetos 18-21 de los datos de estado del canal del interfaz de audio digital, tal como se ha descrito anteriormente, y dividiendo por la velocidad de muestreo en kilohertzios, por ejemplo, por 32, 44,1 ó 48, según convenga.

El cociente binario resultante accionaría el principio de cada bloque audio digital en unidades precisas de 1 ms con un muestreo de 32 kHz ó 48 kHz y dentro de un error máximo de $\pm 1/2$ ms, con muestreo de 44,1 kHz. Con 11 bits, pueden especificarse un total de 2,048 estados distintos, es decir, el código temporal tendría un ciclo completo de 2 048 segundos, desde la cuenta cero a la cuenta 2047, y vuelta a cero de nuevo.

El código temporal especificado por la UER para el video indica una cuenta sólo de imágenes, es decir, una cada 40 ms en PAL. Para el NTSC, la cuenta sería de una cada 33,37 ms. De esta manera, no parece haber forma de relacionar el código temporal audio con el código temporal video convencional. Por tanto, se propone la inserción de una señal de 11 bits adicional similar a la incorporada en la señal de audio, en cada intervalo vertical del video, indicando el tiempo, en unidades de 1 ms, en el que se inserta. Hay que tener en cuenta que este código temporal de sincronización audio/video es completamente independiente del código temporal video convencional.

No es necesario cursar información en la frecuencia de muestreo dentro del código SAV porque esto ya se ha previsto en la división del número de dirección binario del interfaz digital audio, mediante la frecuencia de muestreo, en kHz. No obstante, la información sobre la frecuencia de muestreo puede recuperarse del código SAV, contando la diferencia entre los números del código temporal SAV de los bloques sucesivos.

Si la diferencia se toma entre los números de dos bloques separados, el tiempo será 12,0 ms para la frecuencia de muestreo de 32 kHz, de forma que la diferencia de números será 11, 12 ó 13. Para el muestreo de 48 kHz, el periodo será de 8,0 ms, lo que da lugar a diferencias de 7, 8, ó 9. Así pues, reconociendo de forma lógica uno u otro grupo de diferencias de números, se identificará la frecuencia de muestreo. Para señal audio digital muestreada a otras velocidades se aplica un procedimiento similar

El procedimiento propuesto depende de una gran estabilidad del reloj en los códigos temporales. La estabilidad de las señales de temporización en televisión viene especificada en el Informe 624 como ± 5 Hz en 4,433 MHz para el PAL, y la de la señal audio digital, en la Recomendación 646 como ± 1 parte en 10^5 . Esta tolerancia comparativamente amplia fue permitida en previsión de las señales de audio procedentes de algunos equipos portátiles distantes.

No obstante, el punto 3 de la Recomendación mencionada indica que "cuando un equipo audio digital funciona en forma autónoma, la tolerancia máxima de la frecuencia de muestreo interna debe ser de ± 1 en 10^5 . Cuando están interconectados, en radiodifusión sonora o en televisión, varios equipos audio digitales, debe ser posible sincronizar la frecuencia de muestreo interna en una frecuencia de muestreo externa (por ejemplo, señales de sincronización de televisión, reloj piloto de la casa de la radio, reloj de alta precisión de una red de telecomunicaciones)."

El esquema propuesto actuaría de forma satisfactoria con estas tolerancias.

3. Corrección de los retardos

En la recepción, cada uno de los dos códigos temporales SAV, que llega únicamente a intervalos de un bloque de audio o de una trama de video, se utiliza para sincronizar su propio código temporal de funcionamiento continuo, tal como se indica en la fig. 1. Estos dos códigos temporales derivados se comparan y sus diferencias se utilizan como señal de corrección para variar el retardo de audio, hasta que la diferencia sea cero.

Puede parecer que, con un ciclo temporal de 2048 estados, el sistema podría corregir diferencias de retardo de hasta 2,048 segundos. Sin embargo, como diferencia de salidas, $N(D)$, se toma $N(V) - N(A)$. En tanto que el vídeo va por detrás del audio, $N(D)$ es positivo pero si el sonido llega a pasar detrás del vídeo, $N(D)$ será negativo y quedará indicado por un gran número, $N(D) + 2048$. Así pues, si el sonido se retarda en 100 ms, la indicación será de 1948.

Por tanto, se propone que si la diferencia entre los dos códigos temporales derivados se encuentra entre los números 1536 a 2047, debe interpretarse como que cae entre -512 y -1, y utilizarse por tanto para fijar a cero el retardo, junto con una indicación de "el vídeo precede al audio". De esta manera, el sistema puede corregir errores de la señal de audio que preceda a la de vídeo en un máximo de 1,535 segundos, lo cual es suficiente para los errores mayores que cabe prever.

Como el sincronismo de vídeo funciona asincrónicamente a partir del código temporal derivado del audio, los tiempos indicados del vídeo pueden caer en errores de $\pm 1/2$ ms como máximo, de forma que la diferencia de tiempo indicada $N(D)$ entre el audio y el vídeo variará también en esta cantidad. Se propone por tanto, incorporar una histéresis en el sistema, de forma que el número de corrección de retardo $N(C)$ no varíe hasta que $N(D)$ difiera de él en más de ± 2 ms, y que una vez que se produzca el cambio, se proceda hasta que la corrección de $N(C)$ sea igual a $N(D)$. Este aspecto no se indica en la fig. 1.

También se propone que los errores iniciales mayores de 3 ms se corrijan de forma inmediata, pero que los errores de ± 2 ms se corrijan a una velocidad de 1 en 500, es decir, a lo largo de un periodo de un segundo.

El retardo variable puede aplicarse utilizando técnicas ya conocidas.

4. Transmisión del código de sincronización A/V

En los puntos que siguen se indica un posible método de transmisión del código SAV en cada una de las señales componentes.

4.1 Audio

Un documento resultante del GIT CMTT/4 [CCIR, 1986-90b] propone un método mediante el cual las señales de programa radiofónico de gran calidad que comprenden un par estereofónico muestreado a 48 kHz, pueden cursarse por la RDSI, a 1 920 kbit/s (para acceso H12) o a 1 536 kbit/s (para acceso H11).

En el modelo que propone el documento, los datos del estado del canal del interfaz audio digital, que incluyen el reloj de la hora del día en código binario de 32 bit, se transmiten en la expansión temporal 48:1. En otras palabras, los datos se transmiten transparentemente, pero se transmite un bloque completo del estado del canal cada 192 ms, en lugar de cada 4 ms, como en el caso del interfaz de estudio. El código de sincronismo A/V de 11 bit puede obtenerse a partir de estos datos, tal como se ha descrito anteriormente, y el único inconveniente es que todo cambio en el retardo de transmisión se detecta únicamente después de un retardo de hasta 192 ms. No obstante, no parece ser ésta una consideración importante.

Una vez establecida la diferencia de temporización entre el código temporal recuperado y el código temporal inicial, que puede ser hasta de 192 ms, se han de prever todos los ajustes fijos necesarios de ella antes de obtener el código SAV. Este aspecto no se aborda con todo detalle en [CCIR, 1986-90b], más allá de la observación en el punto 10, estado del canal, en el sentido de que "como sólo se transmite un bloque de datos de cada 48, los dos contadores (muestra local y código de dirección de la hora del día) deben aumentarse en el decodificador en la cantidad adecuada".

4.2 Vídeo

Antes de poder insertar un código de 11 bit en el intervalo vertical, es necesario llegar a un acuerdo sobre el lugar en que debe ponerse. El emplazamiento óptimo podría situarse por ejemplo en el código de entrega de programas propuesto en [UER,1989].

5. Conclusión

Se considera que el método propuesto puede sincronizar las componentes de audio y vídeo transmitidas por trayectos distintos, sin ocupar más de un mínimo de "espacio" en cualquiera de las señales.

REFERENCIAS BIBLIOGRÁFICAS

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

UER [marzo de 1989] Specification of the domestic video programme delivery control (PDC). EBU 459 Rev.

Documentos del CCIR

[1986-90]: a. 10/315 (Australia); b. GIT CMTT/4-8 (GIT CMTT/4).

INFORME 1071

CONVERSIÓN Y SINCRONIZACIÓN DE LAS FRECUENCIAS DE MUESTREO DE LAS SEÑALES DE SONIDO DIGITALES

(Cuestión 51/10, Programa de Estudios 51B/10)

(1986)

Cuando se combinan dos o más señales de sonido digitales en un equipo de tratamiento digital, las frecuencias de muestreo de las señales deben ser sincronas para que se puedan efectuar las operaciones aritméticas necesarias.

En [CCIR, 1982-86a] se describe un sincronizador de retardo variable para señales de sonido digitales que tienen la misma frecuencia de muestreo nominal.

El sincronizador consta de una memoria tampón en la que se introducen las muestras de sonido, que son leídas a la frecuencia necesaria de muestreo a la salida. La memoria tampón opera efectivamente como un dispositivo de retardo variable para comprimir o dilatar muy ligeramente en el tiempo la señal de entrada a fin de conseguir la sincronización.

De cuando en cuando hay que repetir las muestras o suprimirlas para evitar que la memoria se llene o se vacíe completamente (y, por consiguiente, no pueda seguir cumpliendo su función de elemento sincronizador). No se produce ninguna degradación audible siempre que la repetición o la supresión de palabras de muestra tenga lugar durante las «pausas» (o silencios) naturales.

Se ha demostrado que un sincronizador experimental de frecuencia de muestreo que se funda en el principio del retardo variable puede sincronizar satisfactoriamente señales de sonido digitales con frecuencias de muestreo que difieren en hasta $\pm 1 \times 10^{-5}$ de su frecuencia nominal. Sin embargo, se recomienda que las frecuencias de muestreo tengan una estabilidad de $\pm 1 \times 10^{-6}$ con respecto a la frecuencia nominal cuando se utiliza este tipo de sincronizador.

Los resultados experimentales obtenidos demuestran que la utilización de este tipo de sincronizador no produce degradación perceptible de la señal. Esta técnica tampoco causa ningún aumento del ruido de cuantificación.