



ANGLATÈCNIC, S.L.

Ingeniería y Desarrollo de Sistemas para los
sectores Broadcast y Tecnologías de la Información

Ref.: I/263-v2
1/09/11

**Informe I/263-v2: “Flujo de trabajo para la
creación de Audiodescripciones”**



Informe I/263-v2: “Flujo de trabajo para la creación de Audiodescripciones”

Introducción: Generación de la pista de audiodescripción

La generación de una pista de audiodescripción se puede descomponer en distintos pasos. No todos los pasos son indispensables y algunos se pueden realizar simultáneamente en función de la aplicación y el contexto. Los pasos son:

1. Grabación de segmentos El locutor, mediante el equipo necesario, genera segmentos de voz relacionados con el TC (código de tiempo) dónde se deberán insertar utilizando el vídeo original o una copia de menor calidad como referencia.
2. Almacenamiento Los distintos segmentos de voz que componen la audiodescripción se digitalizan e indexan según el TC de inserción. Como formato de almacenamiento de los segmentos de voz se puede utilizar cualquier códec (p.e.mp3 o WAV)
3. Transporte Si es necesario, se deberá hacer llegar la audiodescripción al sistema que realice los pasos posteriores.
4. Mezcla Generación de una sola pista de audio con el audio de programa (audio original del vídeo) y la audiodescripción
5. Distribución al cliente final En función del medio de distribución se deberá procesar la pista de audio de forma distinta si se distribuye mediante un DVD o se emite por TDT desde una emisora de televisión.

Nos interesa evaluar los distintos escenarios en función de su versatilidad y calidad. Para ello, analizaremos cada paso individualmente y finalmente mostraremos algunos posibles escenarios. Un punto a destacar es que los segmentos de audio contienen voz. Este tipo de audio tiene un ancho de banda muy limitado y otras características que simplifican su captura.

Paso 1: Grabación de segmentos

La calidad vendrá determinada por el equipamiento de grabación y la ubicación del locutor. Se puede separar la creación del guión de la locución. Para una calidad óptima será necesario equipamiento de audio profesional de alta calidad (micrófonos, ecualizadores, filtros, etc...) y un locutorio. Con equipamiento más sencillo como un micrófono para PC y una habitación



silenciosa también se pueden conseguir resultados aceptables. Como siempre tendremos un compromiso entre:

- **Precio:** El equipamiento más profesional será más caro lo que limitará el número de equipos que podremos instalar con un presupuesto fijo.
- **Disponibilidad:** Un entorno insonorizado nos va a limitar en disponibilidad, ya que no tendremos disponibilidad inmediata y obligará a desplazamientos del locutor.

Si se necesitan muchas estaciones deberemos optar por equipos más baratos y entornos menos insonorizados a costa de una calidad menor de los segmentos de audio. El locutor puede necesitar un vídeo de referencia para marcar los puntos de inserción del segmento.

Paso 2: Almacenamiento

Normalmente, los segmentos de audio se deberán almacenar temporalmente para su posterior utilización en la generación de la pista de audiodescripción. También se deberá relacionar con el punto del vídeo original dónde se debe almacenar mediante un fichero de índice o método equivalente. La calidad de los segmentos se puede ver afectada por el formato y el medio de almacenamiento. Si se utiliza un formato digital, la calidad vendrá determinada también por el códec utilizado. La necesidad de un transporte posterior también determina el tamaño de la audiodescripción. En este caso el compromiso se encuentra entre:

- **Calidad:** Utilización de códecs sin pérdidas y altas frecuencias y bits de muestreo, pero que ocupan mucho tamaño. Por ejemplo PCM a 48000 Hz .
- **Tamaño:** Utilización de códecs con compresión y pérdidas adaptados al oído humano o al tipo de audio. Algunos ejemplos son el mp3 o el speex. Este último está especialmente adaptado a la voz.

Paso 3: Transporte

Un sistema versátil debe permitir separar cada paso en el tiempo y en el espacio. De esta forma se pueden adecuar los recursos necesarios a la carga de trabajo (equipamiento, personal, etc...). El método de almacenamiento de la audiodescripción determinará las limitaciones en el sistema de transporte. Un medio físico obligará a la utilización de métodos de transporte tradicionales. Por contra, la utilización de ficheros digitales permitirá la utilización de redes como internet. En este último caso, la distribución es mucho más ágil y depende del ancho de banda de las redes utilizadas y del tamaño del fichero de almacenamiento.



Paso 4: Mezcla

El usuario u oyente debe oír simultáneamente el audio del vídeo audiodescrito y los segmentos de la audiodescripción. Para ello se deben mezclar los dos audios. Una audiodescripción de calidad debe permitir al oyente seguir cómodamente los diálogos y efectos del vídeo original y los segmentos de audiodescripción. Para ello, normalmente se debe adaptar el volumen de ambos audios en el momento de la mezcla. Por ejemplo, un explosión puede dificultar la comprensión del segmento de la audiodescripción o un segmento de audiodescripción oculte un sonido ambiente relevante. El equipamiento utilizado para la mezcla determinará la calidad de la pista generada, pero vendrá limitado por el método de distribución del vídeo y la audiodescripción. Podemos distinguir los sistemas según si la mezcla se realiza Offline o Online:

- **Mezcla Offline** (ver imagen 2): Se almacena la pista de audio mezclada para su posterior distribución. Este sería el caso típico en DVD o sistema de podcast.
- **Mezcla Online** (ver imagen 3): El audio del vídeo y la audiodescripción se mezclan en el momento de su emisión. No existe la pista de audiodescripción, ya que se genera 'en directo'. Éste es un caso típico en sistemas Broadcast

Paso 5: Distribución al cliente final

Si se trata de una pista de audiodescripción almacenada para su posterior reproducción, se deberán aplicar la mismas restricciones que en el paso 2 (Almacenamiento) con la particularidad que si se debe reproducir sincrónicamente con el vídeo.

Escenario Tradicional

Se genera la pista de audiodescripción mediante un equipo de edición profesional en un locutorio insonorizado. Se consigue una pista de audiodescripción de alta calidad pero utilizando una infraestructura de alto coste. Cómo se puede ver en el esquema adjunto, los pasos 1 a 4 se realizan simultáneamente mediante un reproductor y un grabador.

- El audio de programa en formato AES se introduce en el sistema mezclador de audio.
- La salida del micrófono en formato AES se introduce también en el mezclador de audio.
- La salida del mezclador en formato AES con la audiodescripción se introduce en el grabador en formato AES.



- El grabador añade la nueva pista de audio al vídeo SDI con el audio AES original. Tenemos un SDI con dos pistas de audio AES
- Se dispone de una salida de vídeo de referencia para el locutor.

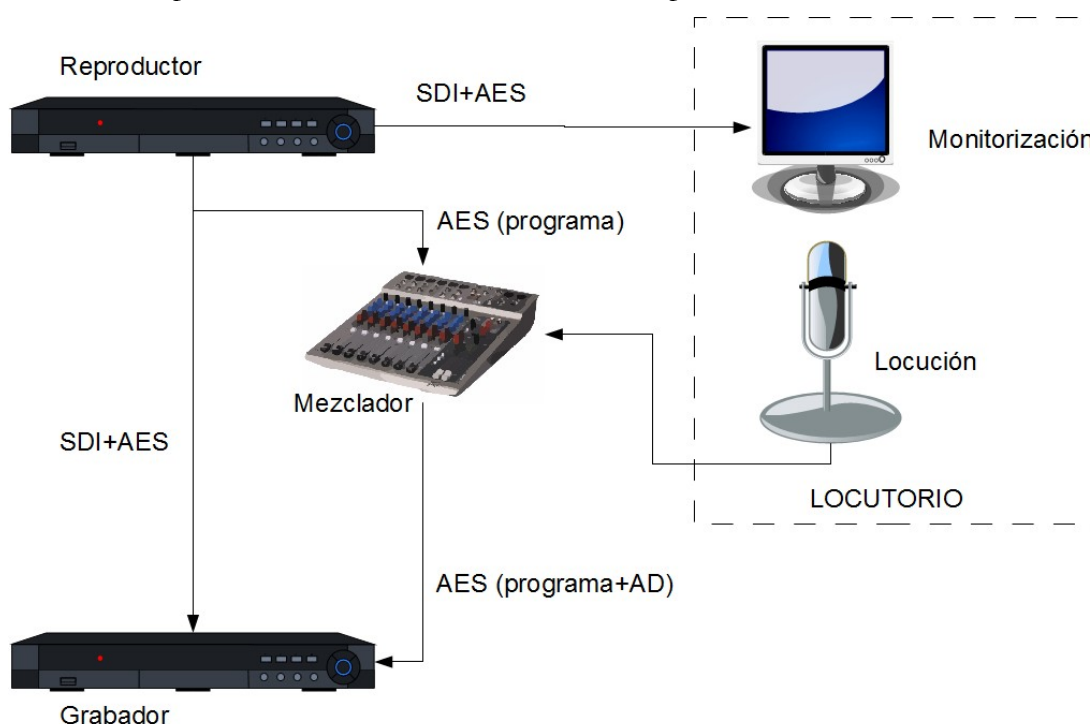


Imagen 1: Sala de producción de Audiodescripciones (en un locutorio insonorizado)

Escenario con Fingertext

Se distribuyen todos los pasos de la generación de la pista de audiodescripción. El sistema está compuesto de uno o varios equipos de bajo coste y un equipo profesional de alto coste. Los equipos de bajo coste se componen de un equipo PC con una tarjeta de audio conectada a unos auriculares y a un micrófono. La calidad de estos 3 elementos determinarán la calidad de los segmentos generados.

- Pasos 1 y 2: Se generan y se almacenan los segmentos en equipos de bajo coste que pueden estar distribuidos geográficamente. Se trata de los pasos que requieren más tiempo por parte del locutor.



- Paso 3: Se envía mediante Internet la audiodescripción con los segmentos codificados y comprimidos.
- Paso 4: En una estación de alto coste se genera la mezcla offline (ver Imagen 2) o online (ver Imagen 3).

Con este esquema conseguimos múltiples ventajas:

- **Rentabilidad de la inversión** en la estación de alto coste, ya que se minimiza el tiempo de ocupación. Una sola estación de alto coste en el estudio o la emisora puede generar las pistas de audiodescripción generadas por múltiples estaciones de bajo coste en un tiempo mínimo.
- Se puede estar trabajando simultáneamente en diversos programas.
- **Compatible con el teletrabajo** por parte del locutor a partir de un vídeo de calidad más baja como referencia.
- Las mismas estaciones de bajo coste nos permiten trabajar para un estudio (offline, ver imagen 2) o una emisora (online, ver imagen 3), ya que será la estación de alto coste la encargada de generar la pista de audiodescripción a partir del mismo tipo de fichero.

Algunas consideraciones sobre la Calidad de la pista de audiodescripción:

- Al tratarse de segmentos de voz, los requerimientos de ancho de banda son mucho menores que los del audio del vídeo.
- Estos se mezclarán con el audio del vídeo, por lo que los segmentos de voz no se oirán aisladamente.

Estas características permiten utilizar sistemas de compresión de audio (mp3,speex) para facilitar el transporte de la audiodescripción. No obstante, esta compresión es opcional y nada nos impide utilizar sistemas de codificación digital de audio sin pérdidas (wav, pcm) a cambio de un transporte mas costoso pero sin merma de calidad por la codificación.

Escenario Fingertext con EBU-WAVE

Actualmente existe un formato estandarizado para la distribución de pistas de audiodescripción, el BWF (Broadcast Wave Format) o EBU-WAVE. De forma sencilla, se trata de un fichero RIFF WAVE (.WAV) con metainformación específica en un chunk especial. De estos metadatos, el mas importante es el TC Inicial del fichero para poderlo sincronizar con el vídeo en emisión.



Este formato genera un sólo fichero de audio sin comprimir (PCM) o comprimido (MPEG-1 Audio, Dolby, etc..) con la pista de audiodescripción. Se puede utilizar en el escenario Fingertext convirtiendo los segmentos a una sola pista EBU-WAVE sin mezclar el audio de programa mediante un proceso de renderización.

Las ventajas de utilizar este formato son:

- Formato **estándar** de distribución con acceso a librerías y catálogos de programas ya audiodescritos.
- Posibilidad de reproducción con software estándar al tratarse de un formato (WAV) ampliamente extendido, sobretodo cuando se utiliza el códec PCM.

Los inconvenientes de utilizar este formato son:

- **Tamaño** de los ficheros muy grandes, ya que también se incluyen los tiempos de silencio. Limitará la distribución de estos en redes de bitrate limitado o medios pequeños. Además requerirá muchos recursos de almacenamiento.
- Dificultades en la **edición** de los segmentos si se deben ajustar los tiempos cuando se realizan cortes o hay diferencias entre el programa de referencia y el que se emitirá. Se deberá utilizar un editor de audio genérico.

Para evitar los inconvenientes de la utilización de este formato se puede utilizar el formato propio de Fingertext (.NAR) sin perder las ventajas del BWF, ya que el sistema Fingertext también permite trabajar con este tipo de ficheros:

- Generamos y almacenamos la pista de audiodescripción en formato NAR, lo cual nos ofrece ventajas entre las cuales podemos destacar:
 - flexibilidad del formato: nos permite grabar/borrar segmentos individuales, desplazar segmentos, etc.
 - optimizado de los recursos de espacio utilizados (almacenamiento, transporte, archivo, etc.).
- En el momento que se necesite generar el fichero BWF para su **emisión** o para su **intercambio** con otros sistemas, se puede utilizar el editor para generar la pista (renderizar) en el formato EBU-WAVE (BWF) siguiendo la especificación EBU tech3285 1.0
- Si es necesario, el sistema Fingertext permite además el almacenamiento, gestión y catalogación, y **emisión** de pistas de audiodescripción en este formato cuando se reciban de sistemas externos.

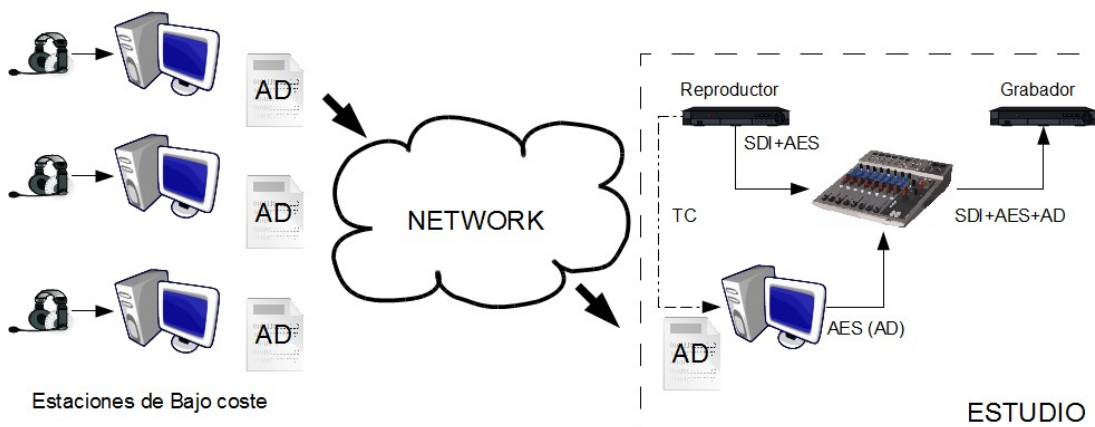


Imagen 2: Esquema de sistema Fingertext con estación Offline

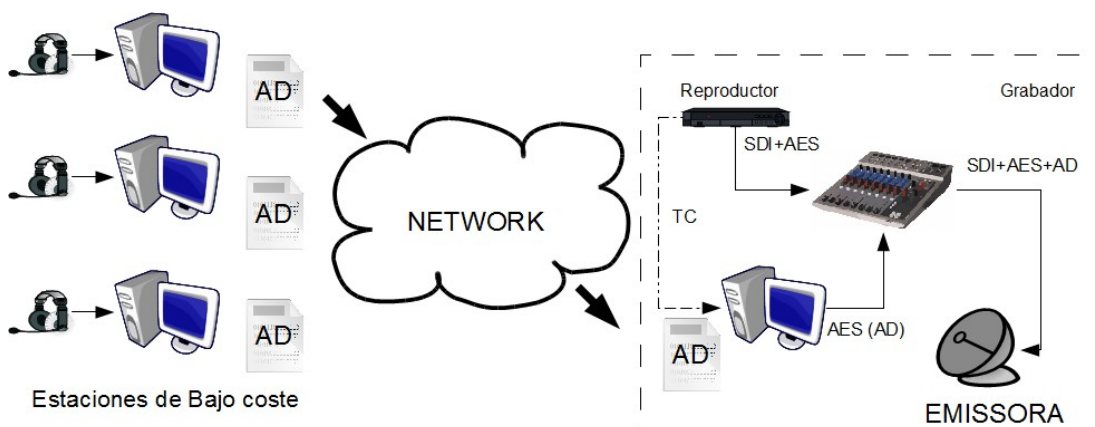


Imagen 3: Esquema de sistema Fingertext con estación Online