RECOMENDACIÓN UIT-R BS.1657

Procedimiento para probar la calidad de funcionamiento de los sistemas de identificación automática de audio

(Cuestión UIT-R 8/6)

(2003)

La Asamblea de Radiocomunicaciones de la UIT,

considerando

a) que en el futuro los metadatos acompañarán a la mayoría de las transmisiones de radiodifusión de audio;

b) que la generación automática de metadatos será necesaria para ofrecer un servicio completo y rentable en el futuro;

c) que la identificación automática de los elementos de audio permite el rastreo de los programas transmitidos;

d) que hoy en día se han desarrollado ya diversos esquemas para la extracción de metadatos de audio;

e) que el ISO/IEC JTC 1/SC 29/WG 11 está concluyendo actualmente esquemas de codificación de metadatos para datos multimedio;

f) que hasta el momento no se han normalizado procedimientos de evaluación de la calidad de los esquemas de extracción de metadatos de audio,

recomienda

1 que para evaluar la calidad de funcionamiento de los sistemas de identificación automática de audio se utilice el procedimiento descrito en el Anexo 1.

Anexo 1

Procedimiento para evaluar la calidad de funcionamiento de los sistemas de identificación automática de audio

1 Introducción

En una época en la que cada vez hay más bases de datos sobre contenido musical, ya sea material genuino de audio o sus metadatos («datos sobre los datos»), también hay una demanda cada vez más apremiante de aplicaciones para mantener ese gran volumen de datos. A esta demanda no sólo contribuyen profesionales, también usuarios comunes de Internet y melómanos que buscan en la web información sobre su estilo musical preferido. Para facilitar la recuperación de la información deseada se distinguen dos niveles de abstracción:

– Búsqueda de metadatos que se pueden extraer más o menos automáticamente del contenido audio, por ejemplo instrumentación, melodía, ritmo. Como ejemplos de aplicación se pueden citar un sistema de búsqueda a partir de una canción tarareada o la clasificación por géneros comúnmente utilizada en aplicaciones que hacen recomendaciones.

– Identificación automática de títulos, cuando no se dispone de metadatos o éstos son insuficientes o poco fidedignos. Se extrae una «característica» de la información de audio y se compara con una base de datos con información conocida, que de este modo permite identificar los correspondientes metadatos, por ejemplo nombre del artista, título de la canción, etc.

Si bien la principal aplicación del primer nivel mencionado es la interacción humana, el segundo también se puede aplicar para proteger los derechos de autor mediante la supervisión de los programas de radio y las transacciones en Internet. Es sobre todo en este último contexto en el que los algoritmos que corresponden a ese perfil se denominan técnicas de «impresión digital».

2 Motivación

Para satisfacer la demanda de la industria discográfica, la velocidad de identificación de la tecnología de impresión digital utilizada debe ser alta y debe soportar las alteraciones y modificaciones comunes del contenido de audio original. A este respecto, la industria discográfica se ha dado cuenta de la necesidad de garantizar la calidad de los sistemas de identificación de audio y ha formulado recientemente una solicitud de información sobre tecnologías de impresión digital de audio.

La gravedad del problema y la urgencia de su solución ha quedado patente por el hecho de que recientemente han aparecido una serie de soluciones diferentes, muchas de ellas protegidas por el derecho de autor. Sin embargo, todos los métodos se encuentran con los mismos problemas en lo que atañe a la robustez cuando se modifica o deteriora el material original. Aunque éste haya sido modificado mediante una serie de procedimientos o se haya deteriorado, deberá no obstante reconocerse el derecho de propiedad intelectual del artista y el compositor. De lo anterior se deduce que la identificación automática de música debe ser en el mejor de los casos tan precisa y tolerante a las modificaciones de la señal como la capacidad humana de percepción e identificación. Además de la robustez a las alteraciones de la señal, un buen sistema de impresión digital debe utilizar una impresión de tamaño reducido (habida cuenta de que ciertas aplicaciones quizá necesiten almacenar millones de impresiones digitales), debe permitir la rápida extracción e identificación de las impresiones digitales y debe tener otras propiedades recomendables. Cabe observar que la robustez en lo que concierne a las alteraciones de la señal y el formato reducido de la impresión digital son dos requisitos antagónicos que los sistemas han de conciliar.

Por consiguiente, para evaluar la calidad de un sistema de identificación automática de audio se ha de definir un entorno de prueba en el que se especifiquen los diferentes tipos de degradación de la señal y los múltiples grados de severidad de la misma y que describa cómo determinar otros parámetros esenciales del sistema. Para poder evaluar objetivamente los sistemas de identificación se necesita un procedimiento de prueba unificado.

3 Parámetros de calidad

En los sistemas de identificación de audio se han de considerar los siguientes parámetros de calidad:

- Tamaño del segmento del material audio que se ha de identificar:
  - ¿qué porción de un elemento es necesaria para realizar la identificación?
- Tamaño de la impresión digital:
  - ¿cuántos datos (bytes) por elemento se han de almacenar en la base de datos?
  - el tamaño de la impresión digital ¿es constante o variable (con respecto a la longitud del elemento)?
- Tamaño de la base de datos:
  - ¿cuántos elementos puede tratar simultáneamente el sistema?

- Modo de identificación:
  - ¿El sistema permite la identificación de subconjuntos de material audio seleccionados aleatoriamente («impresión digital continua») o la identificación está restringida a segmentos de impresión digital cortos? En este último caso, ¿cuál es el tamaño del segmento?
- Velocidad de identificación:
  - ¿cuánto tiempo se tarda en identificar un elemento?
  - ¿cuánto varía este tiempo en función del número de elementos en la base de datos?
- Calidad de funcionamiento de la identificación con material original y alterado:
  - ¿cuánta distorsión soporta sin que afecte considerablemente la velocidad de reconocimiento?
  - ¿cuánto varía la velocidad de reconocimiento en función del número de elementos en la base de datos y del grado de distorsión?
- Velocidad de generación de impresiones digitales:
  - ¿a qué velocidad se pueden generar las impresiones digitales en una determinada plataforma?
  - ¿cuántos recursos son necesarios para generar la impresión digital (por ejemplo velocidad de la unidad de procesamiento central, cantidad de memoria viva, necesidad de unidad de procesamiento de coma flotante)?
- Velocidad de aprendizaje:
  - ¿cuánto se tarda en añadir elementos a la base de datos? ¿Cuánto varía este tiempo en función del número de elementos que contiene la base de datos?

Para evaluar correctamente estas propiedades y demostrar así la adecuación del sistema para su aplicación en el mundo real, el entorno de prueba debe tener unas condiciones de contorno constantes en lo que respecta a las características que se están probando.

Las condiciones de prueba importantes son el tamaño y el contenido de la base de datos de referencia, el tamaño (duración de la reproducción) y el número de elementos de prueba, las reglas de modificación exacta para los elementos de prueba, y la plataforma informática, que incluye la especificación de la unidad de procesamiento central, la memoria y el sistema operativo. En el conjunto de elementos de prueba también debe haber una serie de títulos de prueba que no estén en la base de datos de referencia para comprobar adecuadamente que el sistema no los identifica.

4 Selección del material de prueba y del tamaño de la base de datos

La base de datos de referencia debe contener todos los estilos y géneros musicales, y los géneros más oídos deben aparecer en mayor cantidad. Para hacer una evaluación realista se sugiere que la base de datos tenga entre 10 000 y 100 000 entradas.

Definición de términos:

- Por elemento duplicado se entiende un elemento que es la reproducción exacta del original salvo en una cierta cantidad de muestras de valor cero añadidas al principio o al final. Esta circunstancia se puede dar a veces cuando la «misma» canción figura en diferentes recopilaciones o álbumes.
- Por elemento similar se entiende una mezcla (remix) diferente, una nueva versión/interpretación o una nueva grabación (en directo) de otro elemento de la base de datos.

Requisitos para seleccionar el material de prueba:

– Se debe poner cuidado para que no haya *elementos duplicados* en la base de datos.

– La base de datos contendrá algunos elementos similares (un mínimo de 20 pares). Ejemplo:

  – diez grabaciones en directo de un artista de la misma canción en diferentes conciertos;

  – diez pares de original/mezcla de una canción de artistas diferentes;

  – diez versiones originales/nuevas interpretaciones de una canción por diferentes artistas.

– La base de datos se definirá antes del primer experimento. No se permite modificar la base de datos según los resultados obtenidos en la prueba.

5 Método de prueba

Dado que la velocidad de cálculo puede depender del grado de distorsión del elemento de prueba es obligatorio medir por separado la velocidad de extracción y la de búsqueda (clasificación) en cada experimento (1, 2, 3a) a 3i)).

5.1 Experimento 1

En la primera prueba todos los títulos de la base de datos de referencia deben estar inalterados para su identificación. Por tanto, el porcentaje de elementos identificados correctamente por el sistema sometido a prueba debe ser del 100%.

El tamaño medio de la impresión digital se calcula a partir de todos los elementos de referencia. El resultado será un tamaño medio por elemento o un tamaño por longitud del elemento, en función del tipo de impresión digital utilizada por el sistema sometido a prueba. Los datos de los sistemas que no realizan la impresión digital continua se examinarán por separado de los datos de los sistemas que si la utilizan.

5.2 Experimento 2

A continuación se añadirán al conjunto de prueba dos fragmentos o pasajes de 1 000 elementos que no figuran en la base de datos de referencia y por consiguiente desconocidos por el sistema, con una longitud de 5 y 30 s, respectivamente. Estos 2 000 fragmentos se someten a la identificación por el sistema para averiguar si funciona la identificación negativa y comprobar las posibles identificaciones falsas. En este conjunto de 2 000 elementos debe haber al menos diez del tipo «elementos similares» (a un elemento correspondiente en la base de datos de referencia).

5.3 Experimento 3

Para probar la robustez en lo que concierne a las piezas musicales modificadas se elige un conjunto de 1 000 elementos del conjunto de referencia. La primera prueba se realizará de conformidad con lo descrito en 3a). Las otras pruebas (3b) a 3i)) se realizarán a partir de los fragmentos creados en 3a), es decir, resultantes de combinar la distorsión específica de cada prueba con el efecto «de corte» descrito en 3a). La razón por la que se combinan las demás distorsiones con el efecto «de corte» se debe a que de esta manera se elimina la hipótesis poco realista de impresiones digitales perfectamente alineadas.

Se recomienda utilizar los siguientes procedimientos de modificación:

3a) Corte/traslación

Se utilizarán únicamente subsegmentos pequeños del elemento de prueba. Se deberá variar la muestra de inicio del fragmento (elegido aleatoriamente pero constante para todos los sistemas de prueba). La longitud de pasaje debe ser de 5, 10 y 20 s, respectivamente.

3b) Compresión y expansión dinámicas

Se seleccionarán los parámetros de conformidad con la configuración habitual utilizada en la radiodifusión.

3c) Ajuste del nivel

Se aplicará a la señal de entrada un cierto factor de escala, por ejemplo –6 dB y 10 dB, evitándose los recortes de la misma.

3d) Ecualización

Se utilizará una ecualización de banda de una octava con atenuaciones de las bandas adyacentes de –6 dB y +6 dB.

3e) Adición de ruido

Se añadirá ruido blanco o ruido rosa con una relación *S*/*N* total de 10 y 20 dB, respectivamente.

3f) Conversión de la frecuencia de muestreo y variación del tono

Se utilizarán desviaciones de +5% y –5% en la frecuencia de muestreo.

3g) Codificación de sonido y marcado de contenido

Se evaluarán los efectos de la codificación de sonido utilizando una señal codificada con MPEG-1/2 Capa-3 con las siguientes combinaciones de velocidad binaria/canal: 24 kbit/s (mono), 64 kbit/s (estéreo), 96 kbit/s (estéreo) y 128 kbit/s (estéreo).

3h) Limitación de la banda

Se limitará la banda de la señal de entrada para que no tenga un límite de frecuencia superior de 4 kHz.

3i) Transmisión acústica

Se comprobarán las imperfecciones causadas por la reproducción acústica en condiciones acústicas moderadas: la señal se transmite por un altavoz y se vuelve a grabar utilizando un micrófono. La distancia recomendada entre ambos es de 50 cm. No es necesario que el altavoz o el micrófono sean de alta calidad. La prueba se debe realizar en una habitación normal (sin acondicionamiento ni aislamiento acústico).

Los parámetros de cada prueba de modificación se han ajustado de manera que la percepción acústica humana equivalente se clasificaría desde una «ligera alteración» a una «fuerte variación» de la pieza original. En el caso de la codificación de audio correspondería a una codificación del formato MP3 a 128 kbit/s (estéreo) para una ligera alteración del material original y a 24 kbit/s (mono) para una fuerte variación. Se recomienda la codificación intermedia a 96 kbit/s (estéreo) y 64 kbit/s (estéreo) ya que éstas son las velocidades binarias que se utilizan comúnmente en las transacciones Internet. No se deben utilizar más de cinco niveles de degradación[1].

---

[1] Se considera que no es necesario incluir las codificaciones MPEG-1/2 Capa-2, MPEG-2/4 AAC, Dolby-E, etc., que se utilizan frecuentemente en la radiodifusión ya que éstas normalmente no se utilizan mal en un entorno de estudio como sucede frecuentemente con la codificación MPEG-1/2 Capa-3 (MP3).

6 Plataforma de prueba

Se recomienda que la plataforma informática y sistema operativo utilizados correspondan con los equipos más modernos disponibles por el usuario normal. Un ejemplo de plataforma válido para 2002 es un ordenador con procesador Pentium que funcione a 1 GHz, con 512 MB de memoria viva y con Windows 2000™ o Linux.

7 Variación de los parámetros del sistema

Durante las diferentes pruebas se pueden ajustar los sistemas de impresión digital que permiten variar el grado de robustez o la compacidad de la impresión digital en función de la configuración de los parámetros de extracción, con objeto de optimizar la calidad de funcionamiento de cada tarea/prueba. Ahora bien, en ese caso cada combinación de sistema y configuración se debe considerar como un sistema distinto con un determinado ámbito de aplicación, formato de impresión digital propio y proceso de extracción. Esto no es aplicable a los sistemas en los que se puede obtener una base de datos de impresiones digitales más compactas o menos robustas, a partir de una representación menos compacta o más robusta, mediante un proceso de transcodificación autocontenido, es decir cuando basta con un solo proceso de extracción de impresión digital del material audio de referencia para habilitar todas las funciones mostradas en las pruebas.

8 Informe de la prueba

En los informes de prueba debe figurar, lo más claramente posible, las razones del estudio, los métodos utilizados y las conclusiones obtenidas. Deben tener un grado de detalle suficiente para que una persona entendida pueda, en principio, reproducir el estudio con objeto de comprobar empíricamente los resultados. Los lectores informados deben poder comprender y elaborar una crítica de los detalles más importantes de la prueba, por ejemplo las razones subyacentes del estudio, los métodos de diseño experimentales y ejecución, y los análisis y conclusiones.

Se debe poner especial cuidado en los siguientes aspectos:

- la especificación y selección de los elementos de referencia y de prueba;
- la selección de los segmentos similares y los correspondientes resultados de las pruebas para estos elementos especiales;
- la descripción detallada de los parámetros de las diferentes distorsiones;
- la descripción detallada de los parámetros utilizados en la prueba de los sistemas;
- la explicación detallada de todas las conclusiones que se han alcanzado.