RECOMENDACIÓN UIT-R BS.1693

Procedimiento para probar la calidad de los sistemas automatizados de consulta cantada

(Cuestión UIT-R 8/6)

(2004)

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 los sistemas de consulta cantada constituyen una forma natural de interrogar a los bancos de datos de audio;

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

e) que la Recomendación UIT-R BS.1657 – Procedimiento para probar la calidad de funcionamiento de los sistemas de identificación automática de audio, describe un procedimiento para las pruebas de calidad de los sistemas de identificación automática;

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

g) 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 automatizados de consulta cantada se utilice el procedimiento descrito en el Anexo 1.

Anexo 1

Procedimiento para evaluar la calidad de funcionamiento de los sistemas por pseudomelodía automatizados de interrogación

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 Red información sobre su estilo musical preferido. Para facilitar la recuperación de la información deseada se distinguen dos niveles de abstracción:

– La búsqueda de metadatos de nivel superior, tal como un oyente humano describiría el contenido, por ejemplo, la melodía, el ritmo, el timbre, la instrumentación o el género. Como ejemplo de aplicación se puede citar un sistema de consulta cantada, el cual puede utilizarse como referencia para posibles recomendaciones.

– La extracción de metadatos de nivel medio para la identificación automática de ciertas interpretaciones de contenidos musicales. Descripciones de los aspectos técnicos de los datos de audio (contenido espectral, etc.) se extrae y compara con un banco de datos de material conocido, creando con ello un enlace a metadatos tales como los de artista o nombre de la canción.

Para una panorámica del estado actual de la técnica de los sistemas de consulta cantada, véase el documento ISMIR 2002 (3rd International Conference on Music Information Retrieval, IRCAM – Centre Pompidou Paris, France, octubre de 2002.

2 Motivación

Para satisfacer la demanda de la industria discográfica, la velocidad de identificación de la tecnología utilizada de consulta cantada debe ser alta y debe soportar las alteraciones y modificaciones habituales de las representaciones almacenadas en el banco de datos de canciones.

Este problema se aborda mediante una serie de soluciones distintas, a menudo patentadas, surgida recientemente[Clarisse y otros, 2002], [Ghias y otros, 1995], [Haus y Pollastri, 2001], [Heinz y Brückmann, 2003], si bien, todos los métodos se enfrentan a los mismos problemas relacionados con su inmunidad ante las modificaciones del material original. Ello lleva a la propuesta de que los sistemas automatizados de consulta cantada deben ser en teoría tan precisos y tolerantes ante las modificaciones de la señal como la percepción y la identificación humanas. Por tanto, un sistema avanzado de consulta cantada tiene que tener una gran inmunidad ante las distintas distorsiones respecto a la calidad de la señal y las variaciones respecto a las entradas de melodía ideal. Además, debe incorporar un tratamiento fiable de grandes bancos de datos de canciones compuestos por varios miles de ellas.

Por consiguiente, para evaluar la calidad de un sistema de consulta cantada se ha de definir un entorno de prueba en que abarque los diferentes tipos de modificaciones de la señal 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

Para la evaluación de los sistemas de consulta cantada se han de considerar los siguientes parámetros de calidad:

Entrada de audio requerida:

– ¿Es necesario cantar una cierta parte de la canción o es posible cantar cualquier parte?

– ¿Cuál es la longitud mínima de la entrada para dar un resultado fiable?

Tamaño de la representación de los datos:

– ¿Cuántos datos (bytes) por canción han de almacenarse en un banco de datos musical?

Tamaño del banco de datos musical:

– ¿Cuántas canciones pueden guardarse en un banco de datos musical?

Modo de identificación:

– ¿Cómo influye en la velocidad de identificación y en la calidad el tipo de entrada, tal como el canto en lengua materna, el tarareo o los modos de cantar del tipo «la-la-la», etc.?

Velocidad de identificación de la melodía:

– ¿Cuánto tiempo lleva identificar una melodía?

– ¿Cómo se conjuga ello con el número de canciones del banco de datos musical?

– ¿Cómo se conjuga ello con la calidad de los datos de entrada?

Para evaluar estas propiedades de forma sensible y mostrar con ello la conveniencia de un sistema para aplicaciones del mundo real, un entorno de pruebas debe tener condiciones de contorno constantes en relación con las características que se prueban.

Las condiciones de prueba pertinentes son:

– el tamaño y contenido del banco de datos musical (véase el § 4);

– el tamaño de la interrogación (en referencia a la duración de la melodía) y el número de elementos de prueba (véase el § 4);

– las reglas exactas de modificación de los elementos de prueba (véanse los § 5 y 6); y

– la plataforma de cálculo, que incluye la especificación de la unidad de procesamiento central (CPU), la memoria y el sistema operativo (véase el § 7).

4 Selección del material de prueba y del tamaño del banco de datos musical

Debe definirse un banco de datos de muestras musicales de referencia respecto al que plantean su interrogación todos los sistemas. El banco debe contener una mezcla de distintos estilos musicales (canciones populares de diferentes países, clásica, ...) con prevalencia de las canciones más familiares a nivel mundial. Debe adoptarse una protección especial para evitar la duplicación de elementos en el banco de datos (nuevas grabaciones, etc.).

Para una evaluación estadísticamente fiable y pertinente se sugiere un tamaño del banco de datos musical comprendido entre 500-1 000 canciones.

Como la preparación de representaciones abstractas de gran calidad de canciones musicales en la forma necesaria para la búsqueda en el banco de datos es un procedimiento complicado y costoso, la construcción del banco de datos de referencia musical se deja a los participantes. Ello conducirá a un criterio implícito de calidad que hallará su significado en los resultados de prueba obtenidos. Todos los participantes son libres de elegir el formato propio del banco de datos que depende del algoritmo de búsqueda.

Debe definirse un conjunto de elementos de prueba (banco de datos de muestras de interrogación) que cumpla los requisitos siguientes: para evitar toda calibración relativa a un conjunto especial de interrogaciones, cada participante debe aportar un total de 200 melodías de interrogación. Una adaptación de los parámetros de los sistemas de consulta cantada a un banco de datos de interrogación con criterios subjetivos puede quedar sobrepasado por esta demanda. Los elementos de prueba deben tener una buena calidad de audio, incluyendo teóricamente la ausencia de distorsión de la señal. Las entradas deben contener tipos diferentes, tales como letras cantadas, melodías aproximadas (tatareadas) y entradas instrumentales. Todo ello debe realizarse mediante una distribución representativa de diversos cantantes e instrumentalistas.

Todos los elementos de prueba deben constituir representaciones de las melodías que figuran en el banco de datos de referencia. El diseño de un comportamiento de rechazo no es adecuado debido a los grados variables de similitud entre melodías.

A medida que aumenta en el tiempo el número de nuevos sistemas probados de consulta cantada irá aumentando el tamaño del banco de datos de muestras de interrogación. Por tanto, será necesario efectuar una repetición de las pruebas a fin de comparar la calidad, conforme a un banco de datos de interrogaciones estadísticamente más expresivo. Se recomienda un procedimiento automatizado de pruebas.

5 Modificaciones

Para ser más realista respecto a las aplicaciones del mundo verdadero, deben modificarse los elementos de prueba de gran calidad (véase el § 4) que utilizan fuentes comunes de polución acústica:

– compresión audio (mp3, aac, ...);
– limitación de la anchura de banda (telefonía, ...);
– cuantificación (modulación por impulsos codificados (MIC), ley A, ...);
– distorsión GSM (sistema mundial para comunicaciones móviles) (velocidad plena, ...);
– ruido de fondo (público, restaurante, tienda de música, ...).

En el § 6 se indica una lista de las reglas exactas.

6 Método de prueba

El parámetro principal para estimar la calidad de los sistemas en cuestión será el porcentaje de melodías determinadas correctamente. Este parámetro puede dividirse en dos categorías:

– el elemento buscado se señala en primer lugar de la lista de resultados presentados;
– el elemento buscado se encuentra entre las diez melodías que el sistema estima más similares.

Estas cifras, así como la velocidad de la extracción y del proceso de búsqueda (determinación) tiene que medirse por separado para cada experimento.

6.1 Experimento 1

En la primera prueba, todos los títulos del banco de datos de referencia deben quedar inalterados para su identificación. De esta manera, se obtienen condiciones óptimas respecto a la calidad de audio y los resultados deben mostrar un índice elevado de identificaciones correctas.

6.2 Experimento 2

Para ensayar el comportamiento en términos de inmunidad del sistema que se considera, se aplican modificaciones diversas a los elementos del banco de datos de muestras de consulta. Se han de elegir modificaciones que representen distorsiones acústicas producidas habitualmente en la realidad.

– Distorsión GSM:

  Se han de procesar los elementos de prueba mediante tres técnicas diferentes de codificación de señales vocales utilizadas para la telefonía móvil (GSM de «velocidad plena», «velocidad plena mejorada» y «semivelocidad»).

– Compresión audio:

  Los ejemplos se han de comprimir/descomprimir utilizando códecs de audio «MPEG-1/2 Capa 3» que aplican velocidades de codificación de 64, 96 y 128 kbit/s. Se recomienda el códec Fraunhofer original.

– Cuantificación:

  Han de someterse los elementos de consulta a una cuantificación no lineal de ley A (8 kHz, 8 bits).

– Limitación de la anchura de banda:

  Se limita la entrada mediante un filtro paso banda conforme a la calidad convencional de la telefonía, es decir de 300-3 400 Hz. Las características del filtro paso banda utilizado deben cumplir el requisito de una caída mínima de 12 dB/octava.

– Ruido de fondo:

  Para disponer de un banco de datos con distorsión casi normalizada de señales vocales de la vida real y el ruido del tipo murmullo, se ha de utilizar el contenido del «CD ICRA ruido» [Dreschler, y otros]. Se han de componer dos clases de señales de ruido diferentes con los datos de la interrogación original, es decir, utilizando las pistas de ruido no modificadas y las versiones atenuadas (–6 dB), respectivamente:

  – murmullo de 2 personas (voz normal, pista 6);

  – murmullo de 6 personas (voz elevada, pista 8).

7 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 2004 es un ordenador con procesador Pentium 4/Athlon que funcione a 2,4 GHz, con 512 Mbits de memoria viva y con Windows XP™ o Linux.

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:

– una especificación y selección del banco de datos musical y del banco de datos de muestras de audio;

– una descripción detallada de los sistemas que se prueban;

– una descripción detallada de todas las conclusiones obtenidas.

Referencias Bibliográficas

CLARISSE, L. P., MARTENS, J. P., LESAFFRE, M., DE BAETS, B., DE MEYER, H. y LEMAN, M. [octubre de 2002] An Auditory Model Based Transcriber of Singing Sequences. ISMIR 2002, 3rd International Conference on Music Information Retrieval, IRCAM – Centre Pompidou Paris, France, p. 116-123.

GHIAS, A., LOGAN, J., CHAMERLIN, D. y SMITH, B. C. [1995] Query By Humming. Musical Information Retrieval in an Audio Database. Procs. ACM Multimedia, p. 231-236.

HAUS, G. y POLLASTRI, E. [2001] An Audio Front End for Query-by-Humming Systems. Procs. ISMIR 2001, p. 65-72.

HEINZ, Th. y BRÜCKMANN, A. [marzo de 2003] Using a Physiological Ear Model for Automatic Melody Transcription and Sound Source Recognition. AES 114th Convention. Amsterdam, Países Bajos.

DRESCHLER, W. A., VERSCHUURE, H., LUDVIGSEN, C. y WESTERMANN, S. ICRA Noises: Artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment. *Audiology*, 40, p. 148-157.