

Indexación de audio: Recuperación de Información Musical (Music Information Retrieval, MIR)

< audias >

Audio, Data Intelligence and Speech
<http://audias.ii.uam.es>

Daniel Ramos Castro

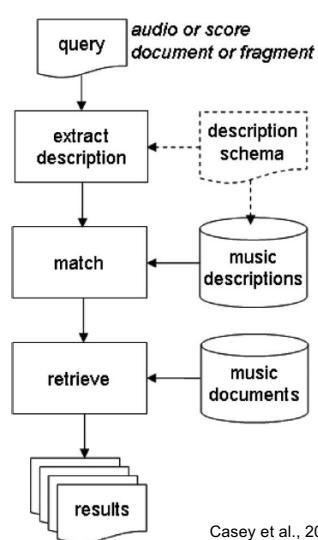
Reuperación de Información Musical

- Music Information Retrieval (MIR)
 - Conjunto de tecnologías orientadas al acceso a información relacionada con la música
- Potenciales usuarios
 - Cliente particular
 - Empresas (productoras, discográficas)
 - Profesionales (autores, musicólogos, profesores, etc.)

Aproximaciones a MIR

- Área organizada a partir de
 - Tipos de búsqueda
 - Tipo de comparación
 - Tipo de salida
- Tipos de búsqueda y salida
 - Información textual (meta-datos)
 - Fragmentos musicales o grabaciones de audio
 - Partituras
 - ...
- Tipo de comparación
 - Exacta
 - Aproximada

MIR: Arquitectura Básica



Casey et al., 2008. "Content-based MIR: current directions and future challenges." Proc. Of the IEEE 96(4), pp. 668-696.

Alcance de MIR: Especificidad

- Alta especificidad
 - Identificación muy precisa del contenido
 - Tendiente a estar basada en contenido del audio
- Baja especificidad
 - Características generales
 - Tendiente a estar basada en meta-datos

Casey et al., 2008. "Content-based MIR: current directions and future challenges." Proc. of the IEEE 96(4), pp. 668-696.

Use Case	Speci-ficity	Description
Music Identification	H	Identify a compact disk, provide metadata about an unknown track, mobile music information retrieval: e.g. shazam.com
Plagiarism detection	H	Identify mis-attribution of musical performances, mis-appropriation of music intellectual property.
Copyright monitoring	H	Monitor music broadcast for copyright infringement or royalty collection
Versions	H/M	Remixes, live vs. studio recordings, cover songs. Used for database normalization and near-duplicate results elimination
Melody	H/M	Find works containing a melodic fragment
Identical Work / Title	M	Retrieve performances of same opus number or song title
Performer	M	Find music by a specific artist
Sounds like	M	Find music that sounds like a given recording
Performance Alignment	M	Mapping one performance onto another independent of tempo and repetition structure
Composer	M	Find works by one composer
Recommend-ation	M/L	Find music that matches the user's personal profile
Mood	L	Find music using emotional concepts: Joy, Energetic, Melancholy, Relaxing
Style / Genre	L	Find music that belongs to a generic category: Jazz, Funk, Female Vocal
Instrument(s)	L	Find works with same instrumentation
Music-Speech	L	Radio broadcast segmentation, Music archives cataloguing



MIR: Meta-Datos y Contenido

- MIR basado en meta-datos
 - Datos añadidos al audio musical
 - Autor, estilo, álbum, año, etc.
 - Enfoque más común en la actualidad
 - Muchísimas páginas web de acceso a música lo implementan
 - Con mayor o menor funcionalidad
- MIR Basado en Contenido
 - Basado en el contenido del propio fichero de audio
 - Enfoque complementario al basado en meta-datos
 - No pretende sustituirlo
 - En este tema nos concentraremos en este enfoque



Estrategias Basadas en Meta-datos

- **Meta-datos objetivos (*factual metadata*)**
 - ▣ **Verdades indiscutibles sobre un contenido musical**
 - Artista
 - Álbum
 - Año de publicación
 - ...
 - ▣ **Ventajas**
 - No sujetos a subjetividad
 - ▣ **Inconvenientes**
 - Pueden describirse con diferentes expresiones textuales
 - Ejemplo: "Price", "The Artist Formerly Known as Prince".
 - Ejemplo: "Purple Rain", "P. Rain".
 - Reduce muchísimo las prestaciones

Estrategias Basadas en Meta-datos

- **Meta-datos subjetivos (*subjective metadata*)**
 - ▣ **Opiniones acerca del fichero**
 - Variables individuo a individuo
 - ▣ **Complementan a los meta-datos objetivos**
 - ▣ **Ventajas**
 - Aumentan el rendimiento en precisión y en carga computacional
 - ▣ **Inconvenientes**
 - Pueden variar incluso para la misma pieza

MIR Basado en Meta-datos

- Ventajas
 - ▣ Recuperación de información textual
 - Tecnologías maduras
- Inconvenientes
 - ▣ Incoherencia y ambigüedad entre meta-datos
 - Debido a diferentes etiquetados
 - Subjetividad (por ejemplo, estilo de una pieza)
 - ▣ Etiquetado detallado muy costoso en tiempo

MIR Basado en Contenido (Content-based)

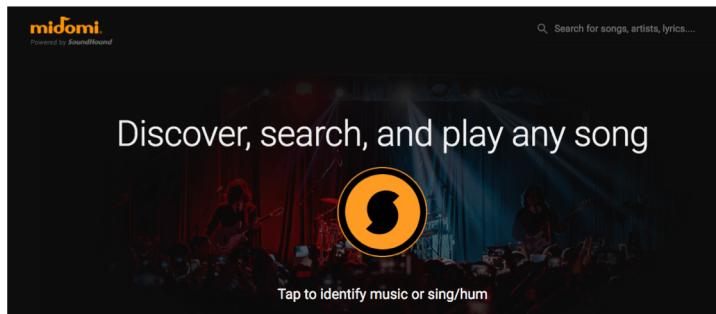
- Ejemplos
 - ▣ www.shazam.com
 - ▣ Recuperación de metadatos a partir de una muestra del audio
 - Incluso ruidosa



MIR Basado en Contenido (Content-based)

- Ejemplos

- www.midomi.com
- Recuperación de metadatos a partir de tarareo o canto
 - *Query by humming, query by singing*



MIR Basado en Contenido (Content-based)

- Ventajas

- El usuario no tiene que saber meta-datos acerca de lo que está buscando
 - Título
 - Año
 - ...

- Inconvenientes

- Es necesario acceder a la información que hay en el audio
- Tratamiento del audio añade mucha complejidad
- Dos tipos de estrategias (complementarias)
 - Descriptores de alto nivel
 - Descriptores de bajo nivel

Estrategias Basadas en Descriptores de Alto Nivel

- Información de Alto Nivel: una definición
 - ▣ Información que un oyente formado extrae de la escucha de una pieza musical si no recibe información previa acerca de la misma
 - Ritmo
 - Melodía principal
 - Inicio de instrumento (*onset*)
 - Orquestación (instrumentos que participan en una pieza)
 - Armonía y tonalidad
 - ...

Casey et al., 2008. "Content-based MIR: current directions and future challenges." Proc. Of the IEEE 96(4), pp. 668-696.



Estrategias Basadas en Características de Bajo Nivel

- Características de bajo nivel
 - ▣ Medidas del audio que contienen información sobre un trabajo musical
 - ▣ No se pueden determinar con una escucha por parte de un oyente formado
- Estrategias de segmentación (división del audio para su análisis)
 - ▣ Basados en tramas (*frame-based*)
 - Enventanado de señal en períodos de entre 10 y 1000 ms.
 - ▣ Sincronizados rítmicamente (*beat-synchronous*)
 - ▣ Distribuciones estadísticas de las características
 - *Bag-of-features models*



Estrategias Basadas en Características de Bajo Nivel

□ Características de bajo nivel de uso en MIR

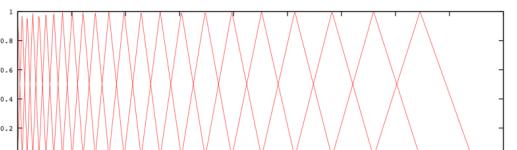
■ Espectro logarítmico (escala Mel, MFCC)

■ Banco de filtros en escala Mel

■ Ideal para distinguir el tipo

de sonido (timbre)

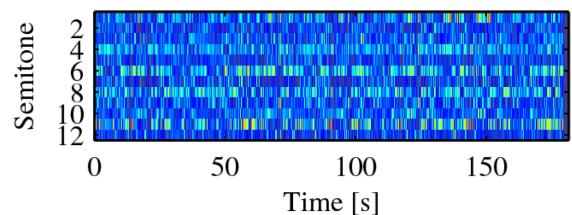
- Timbre: lo que diferencia dos sonidos con mismo volumen y nota musical



■ Perfil pitch-class (cromagrama)

■ Diferencia entre notas musicales

■ Información tonal, armónica o cromática



J. H. Jensen et al., 2008 "A Tempo-Insensitive Distance Measure for Cover Song Identification based on Chroma Features." Proc. ICASSP-08, pp. 2209-2212.



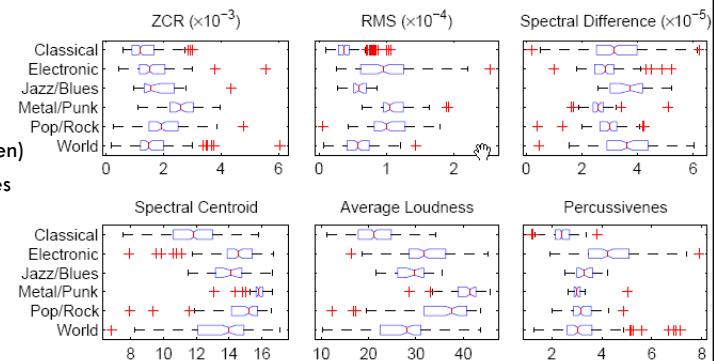
Estrategias Basadas en Características de Bajo Nivel

□ Distribuciones estadísticas de las características (bag-of-features)

■ Ejemplo para clasificación de género musical con características bag-of-features

(cada “punto” representa a un fichero de audio)

- ZCR: tasa de cruces por cero, medida general de cantidad de frecuencias “agudas”
- RMS: potencia de señal
- Spectral difference, spectral centroid: medidas de la distribución espectral
- Average loudness: intensidad percibida (volumen)
- Percusiveness: medida de la cantidad de golpes rítmicos de percusión
- ...



E. Pampalk, 2006. "Computational Models of Music Similarity and their Application in MIR." PhD. Thesis, Vienna Univ.



Análisis de Audio Musical

- Extraer información (de alto nivel) acerca de las características de una pieza musical
 - ▣ A partir del procesado del audio (a bajo nivel)
- Consideraciones:
 - ▣ La señal de audio musical presenta dos dimensiones
 - Dimensión tonal (frecuencia)
 - Dimensión temporal
 - ▣ La música tiene restricciones en ambas dimensiones
 - Tono (notas musicales, dimensión tonal)
 - Ritmo (distribución de pulsos rítmicos, dimensión temporal)

Análisis de Audio Musical

- Tipo de información de alto nivel a extraer
 - ▣ Melodía
 - ▣ Ritmo
 - ▣ Tonalidad
 - ▣ Tempo
 - ▣ Estructura
 - ▣ ...
- Aplicaciones
 - ▣ Seguimiento rítmico
 - ▣ Estimación de melodía principal y acompañamiento
 - ▣ Reconocimiento de acordes y tonalidad
 - ▣ Información acerca de la estructura de la pieza
 - ▣ ...

Ejemplo: Seguimiento Rítmico

- Problema típico en MIR por su tremenda utilidad
 - ▣ Generador de meta-datos
 - Ritmo
 - Tempo
 - ▣ Segmentador de contenido del audio
 - Acotados por pulsos rítmicos
 - Útil para otras tareas
 - Reconocimiento de estructura
 - Reconocimiento de acordes/melodía/acompañamiento
 - ▣ Búsqueda basada en ritmo
 - Sin necesidad de meta-datos

Seguimiento Rítmico

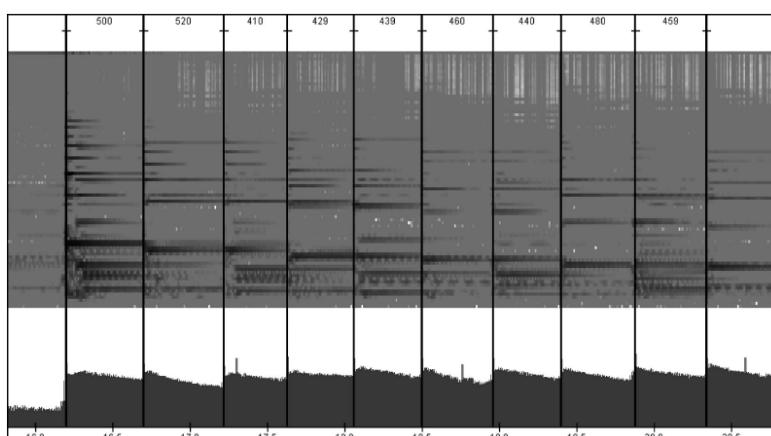
- Primer paso
 - ▣ Reconocimiento de inicios (*onset recognition*)
 - En algunas piezas es fácil de estimar ritmo a partir de los inicios
 - Música con percusión muy marcada y ritmos simples
 - Pop, pop-rock, etc.
 - <http://www.youtube.com/watch?v=jBDF04fQKtQ>
- En otras no es tan obvio
 - Música clásica, jazz, rock progresivo/sinfónico
 - https://www.youtube.com/watch?v=LdpMpfp-J_I

Seguimiento Rítmico: Reconocimiento de Inicios

- Primer paso
 - ▣ Reconocimiento de activaciones de instrumentos (*onset recognition*) Suposición
- Es muy frecuente que en los inicios instrumentales ocurran pulsos rítmicos
- Consecuencia
 - ▣ El ritmo se puede definir a partir de los intervalos entre inicios
 - *Inter Onset Interval, IOI*

Seguimiento Rítmico: Reconocimiento de Inicios

- Ejemplo donde el paso de IOI a ritmo es sencillo



Casey et al., 2008. "Content-based MIR: current directions and future challenges." Proc. of the IEEE 96(4), pp. 668-696.

Seguimiento Rítmico

- Técnicas básicas
 - ▣ Histograma de IOI entre dos posiciones de inicios
 - Se elige el IOI más probable
 - Se puede refinar dentro del intervalo del histograma
- Suelen ser ineficientes con
 - ▣ Ritmos no uniformes
 - ▣ Tempos variables
 - Accelerandos, retardandos
 - Cambios de tempo bruscos y frecuentes

http://www.youtube.com/watch?v=9X_ViIPA-Gc



Seguimiento Rítmico

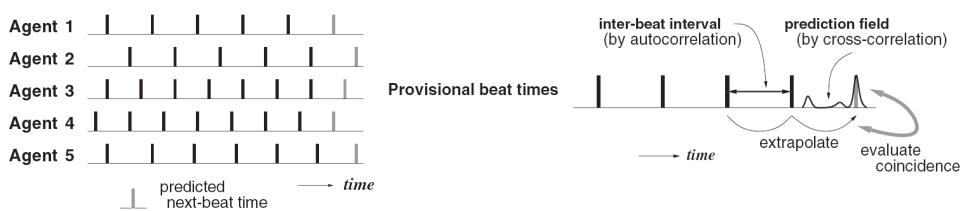
- Problema típico: ambigüedad
 - ▣ Varios inicios corresponden a un ritmo
 - ▣ Varios ritmos son posibles
- Se requieren soluciones complejas
 - ▣ Varias hipótesis (una por posible ritmo)
 - ▣ Algoritmos tipo beam-search o multiagente
 - ▣ Algoritmos de tipo probabilístico
 - Máximo A Posteriori (MAP)
 - Estimación bayesiana



Seguimiento Rítmico

□ Ejemplo: enfoque multiagente

- Diferentes estimaciones de ritmo (hipótesis)
- Fiabilidad (probabilística) dependiente de la predicción de los siguientes acentos



M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds." J. New Music Res., vol. 30, no. 2, pp. 159–171, 2001.



Similitud de Audio Musical

□ Similitud de especificidad media

- Reconocimiento de versiones (covers)

- Tarea compleja

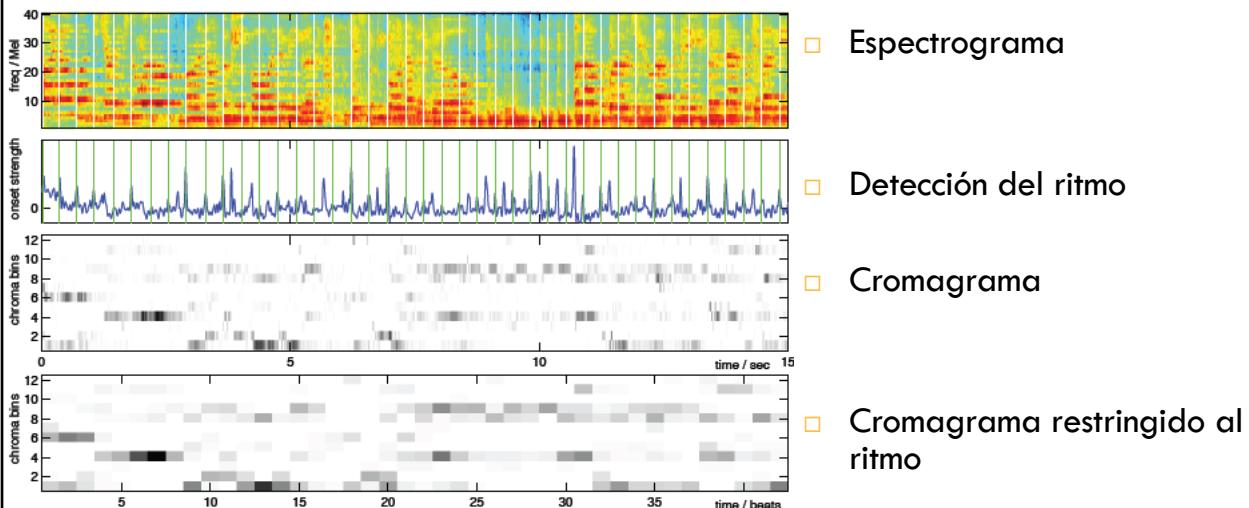
- Ejemplo:

- Chroma features
- Restringidos al tempo de la canción
 - Necesario seguimiento del tempo
- Medidas de correlación cruzada en todo el tema
 - Evita la no similitud en la estructura

D. Ellis et al., "Identifying Cover Songs with Beat-Synchronous Chroma Features," in Proc. ICASSP 2007.



Similitud de Audio Musical

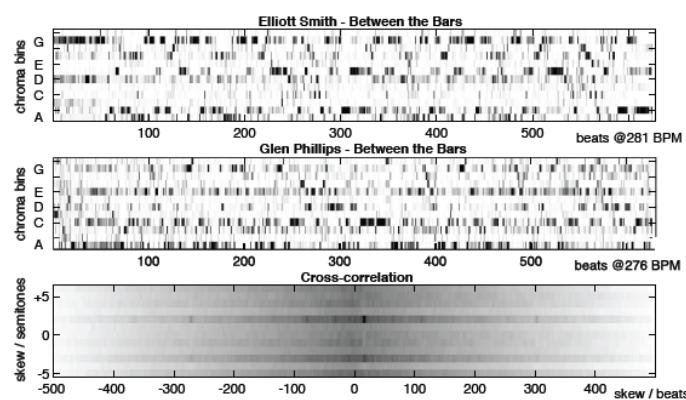


D. Ellis et al., "Identifying Cover Songs with Beat-Synchronous Chroma Features , " in Proc. ICASSP 2007.



Similitud de Audio Musical

- Correlación cruzada en todo el tema
- Desplazamiento para máxima correlación



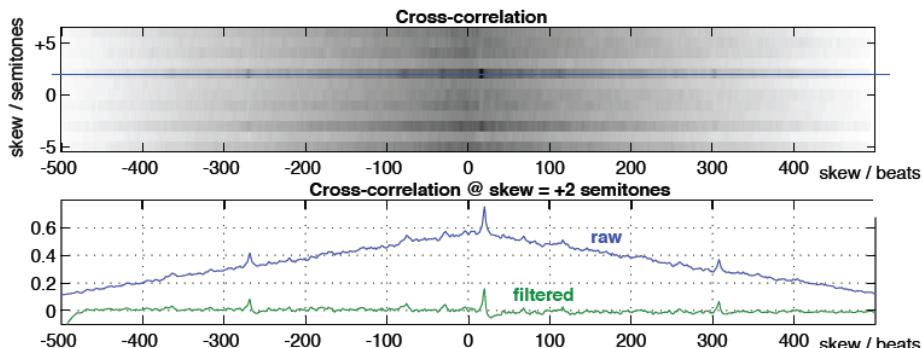
D. Ellis et al., "Identifying Cover Songs with Beat-Synchronous Chroma Features , " in Proc. ICASSP 2007.



Similitud de Audio Musical

□ Picos locales importantes

■ Similitudes debido a cambios en la estructura

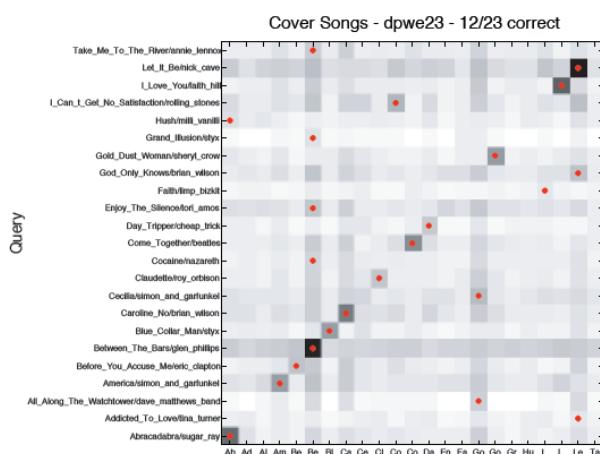


D. Ellis et al., "Identifying Cover Songs with Beat-Synchronous Chroma Features , " in Proc. ICASSP 2007.



Similitud de Audio Musical

□ Resultados (Corpus UsPop2002)

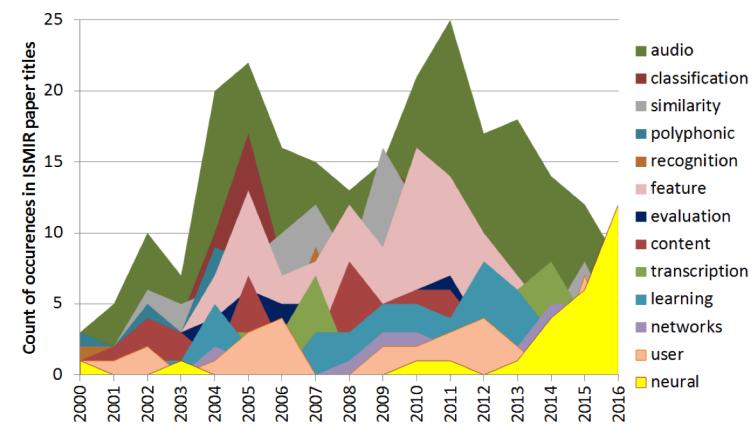


D. Ellis et al., "Identifying Cover Songs with Beat-Synchronous Chroma Features , " in Proc. ICASSP 2007.



Deep Learning para MIR

- Como en todos los campos de las tecnologías de la información y las comunicaciones, el Deep Learning entra con fuerza en MIR a partir de la década de 2010

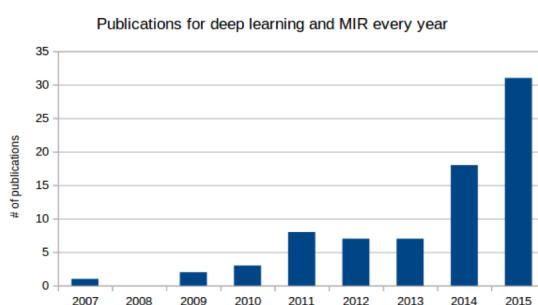


E. Gomez et al., Tutorial at ISMIR 2016: "Music Information Retrieval: Overview, Recent Developments and Future Challenges".



Deep Learning para MIR

- Deep Learning aplica a múltiples tareas en MIR
 - Detección de inicios (onsets), clasificación (género, artista, música vs. voz...), detección de voz cantada, estimación de tonalidad, similitud musical, recomendación, separación de fuentes de audio, extracción de melodía...



E. Gomez et al., Tutorial at ISMIR 2016: "Music Information Retrieval: Overview, Recent Developments and Future Challenges".



Deep Learning para MIR

- Ejemplo: detección de voz y música

de Benito-Gorron et al. EURASIP Journal on Audio, Speech, and Music Processing (2019) 2019:9
<https://doi.org/10.1186/s13636-019-0152-1>

EURASIP Journal on Audio, Speech, and Music Processing

RESEARCH

Open Access

Exploring convolutional, recurrent, and hybrid deep neural networks for speech and music detection in a large audio dataset

Diego de Benito-Gorron*, Alicia Lozano-Diez, Doroteo T. Toledano and Joaquin Gonzalez-Rodriguez



Abstract

Audio signals represent a wide diversity of acoustic events, from background environmental noise to spoken communication. Machine learning models such as neural networks have already been proposed for audio signal modeling, where recurrent structures can take advantage of temporal dependencies. This work aims to study the implementation of several neural network-based systems for speech and music event detection over a collection of 77,937 10-second audio segments (216 h), selected from the Google AudioSet dataset. These segments belong to YouTube videos and have been represented as mel-spectrograms. We propose and compare two approaches. The first one is the training of two different neural networks, one for speech detection and another for music detection. The second approach consists on training a single neural network to tackle both tasks at the same time. The studied architectures include fully connected, convolutional and LSTM (long short-term memory) recurrent networks. Comparative results are provided in terms of classification performance and model complexity. We would like to highlight the performance of convolutional architectures, specially in combination with an LSTM stage. The hybrid convolutional-LSTM models achieve the best overall results (85% accuracy) in the three proposed tasks. Furthermore, a distractor analysis of the results has been carried out in order to identify which events in the ontology are the most harmful for the performance of the models, showing some difficult scenarios for the detection of music and speech.

Keywords: Acoustic event detection, Speech activity detection, Music activity detection, Neural networks, Convolutional networks, LSTM



Máster en Big Data y Data Science

Indexación, búsqueda y análisis en repositorios multimedia: MIR

32



Máster en Big Data y Data Science

Indexación, búsqueda y análisis en repositorios multimedia

Indexación de audio: Recuperación de Información Musical (Music Information Retrieval, MIR)

< audias >

Audio, Data Intelligence and Speech

<http://audias.ii.uam.es>

Daniel Ramos Castro