

Reconocimiento de actividades infantiles utilizando sonido ambiental: Un enfoque preliminar

Antonio García-Domínguez, Carlos E. Galván-Tejada

Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica,
Zacatecas, Zac., México

antonio.garcia@uaz.edu.mx, ericgalvan@uaz.edu.mx

Resumen. El reconocimiento de actividades humanas es un nicho de interés por diferentes entidades, como lo son médicas, empresariales para-automatización y aquellas enfocadas a las casas habitación de los usuarios promedio, sin embargo, el estudio de actividades realizadas por infantes de 12 a 36 meses suele quedar fuera del estudio por ser un grupo complejo de focalizar. Debido a lo anterior, en este trabajo se propone el estudio de actividades infantiles y clasificarlas utilizando un modelo de agrupamiento jerárquico utilizando la información contenida dentro del espectro de los sonidos emitidos al realizar las actividades. Este estudio preliminar demuestra que el espectro tiene suficiente información para clasificar correctamente las actividades infantiles y estimar dicha actividad.

Palabras clave: reconocimiento de actividades infantiles, sonido ambiental, espectro de Fourier, agrupamiento jerárquico aglomerativo.

Child Activity Recognition Using Environmental Sound: A Preliminary Approach

Abstract. Human activity recognition is a topic of interest for different entities, such as medical, bussines for automation and those focused on the homes of average users, however, the study of activities performed by infants from 12 to 36 months usually be out of the study because it is a complex focus group. Due to the above, in this work we propose the study of child activities and classify them through a hierarchical grouping model using the information contained within the spectrum of the sounds emitted during the performance of the activity. This preliminary study shows that the spectrum has enough information to correctly classify children activities and to estimate such activity.

Keywords: child activity recognition, environmental sound, Fourier spectrum, agglomerative hierarchical grouping.

1. Introducción

En los últimos años, la inteligencia ambiental ha sido objeto de numerosos estudios e importantes investigaciones, con la finalidad de ofrecer soluciones enfocadas a facilitar la vida diaria de los seres humanos mediante una interacción automatizada y natural con el medio ambiente en el cual se desenvuelven. Uno de los temas analizados en esta área es el reconocimiento y clasificación de actividades humanas, el cual sirve como estudio base para numerosas aplicaciones planteadas y soluciones desarrolladas.

Un aspecto crucial a tomar en cuenta para el reconocimiento y clasificación de actividades humanas es la fuente de datos a utilizar, ya que de ello depende la elección de las actividades con las cuales se trabaja, además de las técnicas utilizadas para el estudio. La tendencia en los años recientes, para la adquisición de datos, ha sido la utilización de sensores incrustados en objetos de uso cotidiano [6,8]. Otros trabajos han explorado el uso de sensores presentes en los dispositivos móviles, principalmente el acelerómetro, como los presentados por Brayat et al. [3] y Casale et al. [5]. Buscando una mayor precisión y efectividad en el reconocimiento de actividades, se han realizado también trabajos que utilizan el sonido como fuente de datos [7,9,12], principalmente haciendo uso del micrófono con el que cuentan los dispositivos móviles.

Algunos de los trabajos en el área del reconocimiento y clasificación de actividades humanas se han enfocado en el monitoreo de grupos de personas con necesidades específicas, como lo son los adultos mayores y los niños, como los presentados por Shaikh et al. [11] y Boughorbel et al. [4], principalmente como apoyo para salvaguardar su integridad física. Cuando el estudio se centra en el reconocimiento de actividades infantiles, esta tarea se vuelve un enorme reto por la gran diversidad de actividades que realizan, dependiendo del rango de edad en la que se encuentran, debido a su desarrollo físico e intelectual natural. Dichas actividades van desde caminar o correr hasta actividades más complejas como la manipulación de juguetes o artículos potencialmente peligrosos.

Actualmente, la mayoría de los trabajos sobre reconocimiento de actividades infantiles hacen uso de acelerómetros, como el presentado por Nam et al. [10], para la identificación de actividades simples como caminar, correr o saltar. Dichos sensores utilizados deben ser colocados en las prendas que utilizan los niños cotidianamente, o en prendas especiales diseñadas específicamente para la experimentación.

Este trabajo se centra en el reconocimiento de actividades infantiles en niños de entre 12 y 36 meses, debido al comportamiento en común que comparten y que hace posible la identificación de actividades específicas para su reconocimiento y clasificación. La fuente de datos utilizada es el sonido ambiental, de manera que la adquisición de datos no interfiera con el comportamiento natural de los niños. Las actividades a tomar en cuenta para su análisis en el presente trabajo son: caminar, correr, llorar y manipular juguetes (bloques de plástico).

A continuación se describe cómo está organizado el artículo. Después de esta breve introducción, se describe la propuesta de metodología y materiales utilizados. En la tercera sección describimos los procesos de experimentación aplicados.

Seguido de la publicación de los resultados obtenidos durante el proceso. Por último, presentamos las conclusiones, y el trabajo a realizar en el futuro.

2. Materiales y métodos

En la presente sección se describen las características de los audios que modelan las actividades infantiles, así como los métodos utilizados para llevar a cabo el presente estudio.

2.1. Descripción del conjunto de datos

Para este trabajo sobre el reconocimiento de actividades infantiles utilizando sonido, el conjunto de datos (Data-Set) se encuentra formado por grabaciones de audio de cuatro actividades que realizan niños en el rango de edad especificado (12 a 36 meses), en un ambiente controlado (dentro de una habitación): caminar, correr, llorar y manipular juguetes (bloques de plástico). La tabla 1 muestra la descripción de cada una de las actividades a tomar en cuenta para el análisis.

Tabla 1. Descripción general de las actividades.

Actividad	Descripción
Caminar	Trasladarse de un lugar a otro a velocidad media
Correr	Trasladarse rápidamente de un lugar a otro
Llorar	Emitir sonido de llanto como reacción a algún suceso
Manipular juguetes	Jugar con bloques de plástico de modo que éstos emitan ruido al chocar

Dispositivos de grabación. Para el proceso de adquisición de los datos, los audios de las actividades fueron grabados utilizando los micrófonos presentes en dos dispositivos móviles. Las características principales de los dispositivos utilizados se muestran en la tabla 2.

Tabla 2. Dispositivos de grabación.

Dispositivo	Procesador	Sistema Operativo
Lanix Ilium s620	MediaTek MT6582 quad-core	Android 4.2.2
Motorola Moto G4	Snapdragon 617	Android 6.0.1

Metadatos. Los audios grabados se encuentran en formato WAV (WaveForm Audio File), con la finalidad de no experimentar la pérdida de calidad que

Tabla 3. Metadatos.

Actividad	Tasa de Muestreo	Canales	Resolución
Caminar	44,100 Hz	Stereo	16 bit
Correr	44,100 Hz	Stereo	16 bit
Llorar	44,100 Hz	Stereo	16 bit
Jugar	44,100 Hz	Stereo	16 bit

otros formatos presentan debido al proceso de compresión. Las características generales como tasa de muestreo, canales y resolución, se muestran en la tabla 3.

Las características presentadas anteriormente aseguran una calidad aceptable para los archivos de audio grabados, además de definir los parámetros necesarios para grabaciones futuras con la finalidad de expandir el Data-Set.

2.2. Espectro de Fourier

En el área de procesamiento de señales, la Transformada de Fourier es ampliamente utilizada debido a que, al expresar las señales en términos de las frecuencias que la componen, facilita su análisis y procesamiento para distintas aplicaciones.

En el presente trabajo se utiliza el espectro de Fourier de las muestras de audio para realizar el respectivo procesamiento dentro del modelo de clasificación utilizado.

2.3. Modelo de clasificación

Un aspecto importante a tomar en cuenta en el proceso de reconocimiento de actividades humanas es el modelo de clasificación utilizado, debido a que esto se reflejará en el comportamiento de los experimentos realizados y la precisión de los resultados obtenidos. Cuando se analizan señales de audio, como es el caso del presente trabajo, la utilización de un modelo adecuado para la clasificación de las muestras ayuda a obtener un procesamiento adecuado y confiable.

El modelo de clasificación elegido para este trabajo es un modelo de agrupamiento jerárquico aglomerativo, la elección se hizo tomando en cuenta la forma en que este método plantea la formación de grupos con individuos (muestras) de características similares. El modelo toma inicialmente tantas muestra se tengan, tratándolas como grupos individuales. El procesamiento consiste en ir formado grupos de forma ascendente, basándose en la similitud de sus individuos, hasta llegar a la cantidad de grupos definida para el estudio.

3. Experimentación

Mediante el proceso de captura de audios para las actividades a analizar en el presente trabajo, se obtuvieron un total de 23 grabaciones, en la tabla 4 se

muestran las cantidades de archivos de audio que se tienen para cada actividad. El procedimiento realizado a continuación fue llevado a cabo en el lenguaje de programación R, debido a que éste está enfocado al análisis de datos y modelado estadístico, con poderosas características que lo hacen uno de los lenguajes más utilizados en investigación.

Tabla 4. Archivos de audio por actividad.

Actividad	Grabaciones
Caminar	5
Correr	4
Lorar	7
Jugar	7

Para el análisis de los archivos, se trabajó con el espectro de frecuencias de las señales de audio, obtenido mediante el análisis de Fourier correspondiente. Ya que las actividades pueden variar en cuanto a su longitud temporal, debido a su naturaleza y a los infantes, los archivos grabados para el presente trabajo tienen una duración de entre 10 y 160 segundos. Para poder realizar el análisis de frecuencias mencionado, el primer paso fue dividir todos los archivos de audio en muestras de 10 segundos, con la finalidad de obtener un espectro de Fourier utilizando la transformada rápida de Fourier (FFT por sus siglas en inglés) [2] de igual longitud para todas las muestras y proceder al análisis respectivo. La tabla 5 muestra la cantidad total de muestras para cada actividad.

Tabla 5. Muestras por actividad.

Actividad	Muestras de 10 segundos
Caminar	24
Correr	6
Lorar	10
Jugar	42

Los datos de entrada al modelo de clasificación son cada uno de los espectros de frecuencias de las muestras de 10 segundos obtenidas a partir de los archivos de audio grabados de las actividades. El primer paso a realizar en el proceso de clasificación es formar la matriz de distancias con dichas muestras, esta matriz relaciona a todos los sujetos involucrados mediante una medida de su similitud. La obtención de la matriz de distancias se realizó con la función `dist()` contenida en el paquete `stats` de R Project [1].

La formación de grupos con individuos similares, de manera ascendente, se realiza en base a los datos contenidos en la matriz de distancias. Para este

proceso se sigue la estrategia de la distancia mínima o similitud máxima, en la cual las muestras con mayor similitud se juntan para formar grupos cada vez más grandes. El proceso finaliza cuando se ha llegado a obtener un solo grupo que contiene a todos los individuos del experimento o una vez que se ha alcanzado el número de grupos definido para el estudio.

La forma de procesar los datos de la matriz de distancias, formando grupos de manera jerárquica, es representada mediante un dendrograma, en el cual se puede visualizar el proceso completo y los grupos formados en cada paso. Este proceso se realizó utilizando la función `hclust()` contenida también en el paquete `stats` [1].

4. Resultados

En la figura 1 se observa el espectro de Fourier para una de las muestras de 10 segundos obtenidas de los archivos de audio. Este proceso fue realizado para cada una de las muestras totales. El espectro de Fourier permite visualizar la intensidad de cada frecuencia presente en la señal y, en base a esta característica, poder detectar el tipo de actividad al que pertenece la muestra de audio. Es importante mencionar que trabajar con el espectro de las señales de audio es un buen primer enfoque con el objetivo de reconocer actividades infantiles, sin embargo, sin un análisis más profundo, los experimentos realizados podrían parecer tender al caos, debido a la naturaleza de las grabaciones y varios factores más que podrían afectar, como el ruido de fondo.

Los espectros de Fourier de cada una de las muestras fueron tomados como datos de entrada para la elaboración de la matriz de distancia. Mediante esta matriz es posible conocer la medida de similitud entre las muestras de audio analizadas, basándose en la distancia euclidiana. En la tabla 6 se muestra una parte de la matriz de distancia resultante. Los nombres de las muestras están formados por la actividad a la que pertenece el sonido, el número de archivo y el número de muestra.

Tabla 6. Matriz de distancia para 5 muestras.

	Cry_1.wav_1	Cry_2.wav_1	Cry_2.wav_2	Cry_2.wav_3	Cry_3.wav_1	Cry_3.wav_2
Cry_2.wav_1	701.8916					
Cry_2.wav_2	683.3781	582.4608				
Cry_2.wav_3	681.9768	634.6091	581.2837			
Cry_3.wav_1	693.5581	748.2898	684.6646	637.5035		
Cry_3.wav_2	719.3180	746.5135	712.3953	670.5225	5733450	

Tomando la matriz de distancia obtenida en el paso anterior, se procesó utilizando el modelo de clasificación jerárquico utilizando el método de la distancia mínima entre muestras para el agrupamiento. La figura 2 muestra el dendrograma resultante.

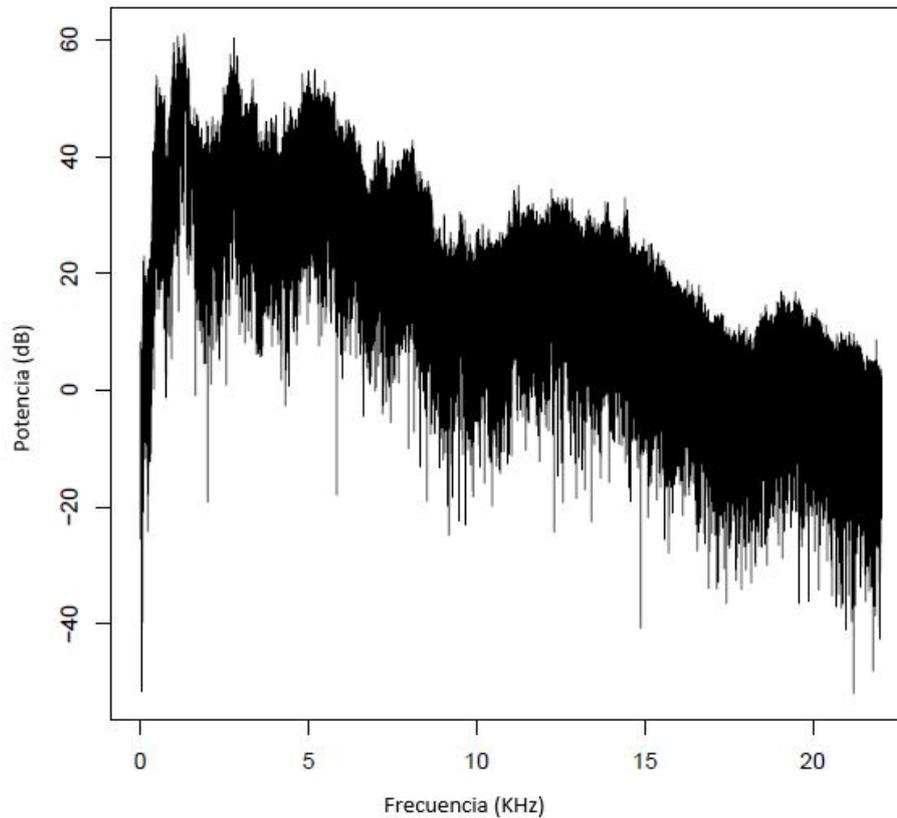


Fig. 1. Espectro de Fourier para una muestra de 10 segundos de duración.

5. Discusión y conclusiones

El enfoque principal de este trabajo de investigación es realizar un estudio preliminar para poder estimar las actividades realizadas por infantes de 12 a 36 meses (considerados biológicamente como bebés) utilizando la evolución espectral de archivos de audio con duración de 10 segundos. Una técnica de agrupamiento jerárquico es utilizada para dicho reconocimiento; de esta experimentación se puede discutir y concluir lo siguiente:

- El reconocimiento de actividades infantiles pueden ser clasificadas utilizando la información del espectro de la señal de audio: La información contenida dentro del espectro permite obtener un agrupamiento de las actividades, sin embargo, se requieren 10 ramas para obtener un agrupamiento correcto de las 4 actividades.
- Actividades similares, como lo son correr y caminar, pueden ser correctamente clasificadas: A pesar de que en su evolución temporal es similar

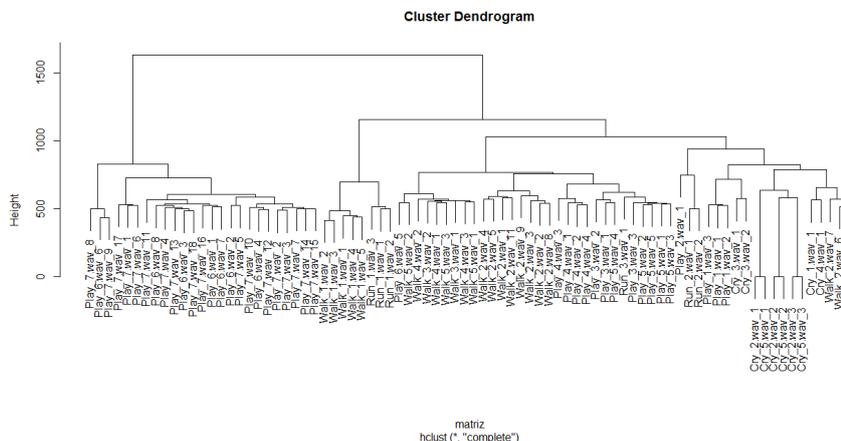


Fig. 2. Dendrograma resultante.

tanto visualmente como de manera auditiva, el espectro contiene suficiente información para diferenciar actividades que generan un sonido ambiental similar.

Por otro lado, se encontró indicios de un comportamiento que demuestra la necesidad de un data set más grande, sin embargo, una función de agrupamiento jerárquico tiene la capacidad de clasificar este tipo de señales correctamente, permitiendo así obtener un modelo de estimación de actividades infantiles utilizando solamente la información en sonidos ambientales.

6. Trabajo futuro

Como parte del trabajo futuro, se propone añadir más actividades que son comúnmente desarrolladas por infantes de 12 a 36 meses, así como un estudio mas profundo de características de la evolución espectral y temporal de la señal para clasificar dichas actividades. de manera adicional se propone los siguientes puntos específicos:

- Realizar un estudio de curvas de aprendizaje para estimar la necesidad de extender el conjunto de datos por cada actividad.
- Agregar más actividades realizadas por los infantes.
- Utilizar el enfoque de extracción de características para reducir la cantidad de datos necesarios así como optimizar la clasificación.
- Implementar técnicas complejas de aprendizaje por computadora.

Referencias

1. R: Distance matrix computation, <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/dist.html>
2. R: Fast discrete fourier transform (fft), <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/fft.html>
3. Bayat, A., Pomplun, M., Tran, D.A.: A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science* 34, 450–457 (2014)
4. Boughorbel, S., Breebaart, J., Bruekers, F., Flinsenberg, I., Kate, W.T.: Child-activity recognition from multi-sensor data. *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research - MB '10* (2010)
5. Casale, P., Pujol, O., Radeva, P.: Human activity recognition from accelerometer data using a wearable device. *Pattern Recognition and Image Analysis Lecture Notes in Computer Science* p. 289–296 (2011)
6. Cornacchia, M., Ozcan, K., Zheng, Y., Velipasalar, S.: A survey on activity detection and classification using wearable sensors. *IEEE Sensors Journal* 17(2), 386–403 (2017)
7. Galván-Tejada, C.E., Galván-Tejada, J.I., Celaya-Padilla, J.M., Delgado-Contreras, J.R., Magallanes-Quintanar, R., Martínez-Fierro, M.L., Garza-Veloz, I., López-Hernández, Y., Gamboa-Rosales, H.: An analysis of audio features to develop a human activity recognition model using genetic algorithms, random forests, and neural networks. *Mobile Information Systems* 2016, 1–10 (2016)
8. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials* 15(3), 1192–1209 (2013)
9. Lee, M., Yang, H., Han, D., Yu, C.: Crowdsourced radiomap for room-level place recognition in urban environment. *2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)* (2010)
10. Nam, Y., Park, J.W.: Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor. *IEEE Journal of Biomedical and Health Informatics* 17(2), 420–426 (2013)
11. Shaikh, M.A.M., Hirose, K., Ishizuk, M.: Recognition of real-world activities from environmental sound cues to create life-log. *The Systemic Dimension of Globalization* (Jan 2011)
12. Sim, J.M., Lee, Y., Kwon, O.: Acoustic sensor based recognition of human activity in everyday life for smart home services. *International Journal of Distributed Sensor Networks* 11(9), 679123 (2015)