

## TIMBRE, DESCRIPCIÓN SONORA Y CLASIFICACIÓN AUTOMÁTICA

Carlos Gustavo Román Echeverri

Universidad de La Salle  
carroman@hotmail.com

**Resumo:** El presente trabajo desarrolla algunas consideraciones teóricas sobre el timbre, para luego explicarlo en el contexto de las nuevas tecnologías y el área del *Music Information Retrieval*, mencionando algunas aplicaciones concretas en descripción sonora y clasificación tímbrica a través de métodos de procesamiento de señales y aprendizaje automático.

**Palavras-chave:** timbre, sonido, procesamiento de señales, aprendizaje de máquina, recuperación de información musical

**Timbre, sound description and automatic classification.**

**Abstract:** This paper presents theoretical considerations on timbre within the context of new technologies and the Music Information Retrieval field, mentioning some concrete applications on sound description and timbral classification by means of signal processing methods and machine learning techniques.

**Keywords:** Timbre, sound, signal processing, machine learning, music information retrieval

### SOBRE EL TIMBRE

El timbre sigue constituyendo un problema interesante y relevante en diversas áreas del conocimiento, ya que para su comprensión global se deben tener en consideración aspectos de índole multidisciplinaria: acústicos y físicos, que configuran características específicas del sonido; aspectos musicales y estéticos, relacionados con los instrumentos, la fuente del sonido o la textura determinada de una pieza; aspectos perceptivos y cognitivos, que definen la manera en la que el sonido se procesa a nivel fisiológico; aspectos técnicos, que pueden llegar a definir posibles aplicaciones tecnológicas que procesen o identifiquen el timbre de un sonido determinado. La noción de timbre no es un universal antropológico, ni ha estado siempre presente en el desarrollo del pensamiento musical occidental. Según explica Rafael Ferrer (2008), los nuevos descubrimientos e inventos generan giros conceptuales mentales que a su vez generan cambios en la manera en la que los mecanismos corporales asimilan una realidad específica. Precisamente, la idea de una conciencia sobre lo tímbrico puede rastrearse al periodo de la Ilustración francesa con la figura de Jean-Philippe Rameau, quien propuso que la distinción entre oír y escuchar podría conducir a la comprensión efectiva de un cuerpo sonoro, planteando de esta manera la necesidad de realizar un esfuerzo consciente para concentrar el acto de escucha en aspectos particulares del sonido. Durante el siglo XIX, Hermann von Helmholtz fue pionero en buscar relaciones entre los atributos perceptivos del sonido y propiedades físicas específicas. Posteriormente el psicólogo Carl Seashore propuso al timbre como la cualidad más importante y compleja del sonido, por encima de la altura, el volumen o la duración (FERRER, 2009). En la actualidad, la dificultad de explicar el timbre desde un punto de vista estrictamente científico ha sido señalada en diversos estudios (ver SETHARES, 1999; O'CALLAGHAN, 2007). Las definiciones estándar resultan ser incompletas o negativas, como aquella de la ANSI (American National Standards Institute) que determina que es la cualidad que permite diferenciar dos sonidos con la misma altura y volumen; o reduccionistas,

como decir que el timbre está definido exclusivamente por la envolvente espectral o por un conjunto de sobretonos. El timbre como una diferencia auditiva ha sido ejemplificado metafóricamente como el equivalente al *rostro* de un sonido específico, donde un conjunto de rasgos sonoros está configurado de tal manera que le permite ser identificable como una unidad (O'CALLAGHAN, 2007).

En música, el timbre está determinado por las propiedades físicas del instrumento que origina el sonido, así como también el amplio rango de posibilidades de producción de sonidos con un propósito musical. El timbre de un instrumento musical específico se percibe como constante frente a variaciones de frecuencia o volumen, a pesar de que el mismo instrumento puede producir una sensación tímbrica diferente ante la variación de las alturas o intensidades. En la historia de la música occidental el timbre ha sido una preocupación relativamente reciente. Obras como *El arte de la fuga* de Johann Sebastian Bach, por ejemplo, prescindieron de detalles en la instrumentación, posicionando otros aspectos musicales como la relación entre alturas por encima de lo tímbrico. En 1911, Arnold Schoenberg propuso en su Tratado de Armonía la *Klangfarbenmelodie* (literalmente sonido-color-melodía) o melodía de timbres, la idea de la valoración de lo tímbrico como parte inherente a la creación musical, con la que se crean sucesiones coherentes de timbres que ya no están subordinados a la altura como elemento constitutivo de la música (SCHOENBERG, 1974). El surgimiento y desarrollo de movimientos como la *Musique concrète* -que consideraba cualquier fuente sonora en soporte de grabación para uso como material musical y por lo tanto exigía una escucha desligada de los parámetros tradicionales- o la música electrónica -que permitía una generación de timbres inéditos y control minucioso de parámetros sobre los mismos- también hablan de una preocupación creciente por los aspectos tímbricos musicales y por una actitud de escucha diferente hacia los mismos. Pero uno de los principales retos frente al timbre ha sido precisamente el de encontrar formas consistentes de describir el sonido.

### **Descripción sonora**

La realidad musical se puede definir como aquella que surge como resultado de una experiencia de inmersión corporal dentro de energía sonora (LEMAN, 2008). Pero para aproximarse a la plétora de fenómenos complejos que emergen de esta experiencia musical, las descripciones constituyen un vehículo inmediato para alcanzar una comprensión racional de los mismos. Las descripciones proveen una significación dentro de un contexto cultural específico, teniendo en cuenta que toda experiencia musical es subjetiva, y que la materia a describir no es siempre directamente observable. Históricamente, el campo de la musicología se ha referido a este problema de interpretar la música a través de descripciones basadas en lo lingüístico, que es una de las maneras de codificar la experiencia musical a través de dispositivos de comunicación simbólica. La descripción del timbre desde un punto de vista perceptivo, usualmente requiere el uso de descriptores semánticos asociados con propiedades y atributos de otros sentidos, por ejemplo con lo visual (color, brillo, opacidad) o lo táctil (blando, punzante). Sin embargo, no existe una correspondencia directa entre rasgos acústicos medibles y timbres definidos, por lo que los timbres no pueden ser representados por medio de una magnitud física escalar unidimensional en la que todos los posibles timbres puedan ser ubicados y ordenados. El paradigma del estructuralismo cognitivo (FERRER, 2008) en las últimas décadas ha logrado que la investigación en timbre se enfoque en el escalamiento multidimensional basado en pruebas de similaridad, constituyendo el método más adecuado para tratar de encontrar modelos computacionales que representen los sistemas de percepción tímbrica humana.

Por otro lado, el desarrollo de tecnologías del sonido ha dado a lugar a la aparición de nuevas herramientas para el análisis y la comprensión de la música. La aparición de los dispositivos para la grabación y reproducción sonora, por ejemplo, ayudaron a replantear la noción misma de timbre, permitiendo no sólo almacenar y encapsular el evento sonoro sino

también emplearlo para su posterior análisis, tal como lo planteara Béla Bartók (1979) con su idea de usar el material musical grabado como medio objetivo para describir y analizar sutilezas que no aparecen en la partitura o en otro medio de representación. Hoy en día, la cultura musical es casi completamente dependiente de la infraestructura tecnológica, especialmente en las áreas de producción, creación y distribución de música, la cual está disponible en cantidades cada vez mayores a través de redes mundiales. Esto representa una oportunidad para emplear los medios tecnológicos no solamente como plataforma para acceder a la música, sino también como una herramienta para describirla. En los últimos años, el campo del *Music Information Retrieval* (MIR, recuperación de información musical) ha tratado de solucionar el problema de categorizar, procesar, clasificar y etiquetar archivos musicales en bases de datos, teniendo en cuenta el permanente incremento en la cantidad de información y la naturaleza pluralista y multicultural del material musical. Una manera de mirar al área del MIR es como una de las principales tecnologías masivas que tratan de reducir la brecha entre el mundo físico del sonido y el campo perceptivo de los sentidos. Una de las herramientas computacionales más importantes que se tienen para tratar de conectar el mundo de información digital abstracta de archivos de audio con conceptos semánticos bien definidos relacionados con la percepción humana, es la de los descriptores sonoros. Diversos rasgos temporales y espectrales son decodificados por los humanos desde la cóclea hasta la corteza auditiva primaria para discriminar la fuente sonora, que es posteriormente clasificada en centros auditivos superiores (HERRERA et al, 2006). Con métodos computacionales, algunos de estos rasgos pueden ser extraídos, cuantificados y codificados de señales de audio brutas, obtenidos de la señal en el dominio del tiempo o de su espectro en el dominio de la frecuencia. Es importante conocer las características acústicas y perceptivas más relevantes no sólo del instrumento musical en sí mismo, sino también de los descriptores asociados con un sonido en particular. La cantidad de descriptores usados en diversos técnicas estandarizadas de MIR es muy amplia, pero pueden ser clasificados principalmente en tres categorías: descriptores de energía (como el valor RMS), descriptores temporales (como el centroide temporal o el zero-crossing rate) y descriptores espectrales (como el centroide espectral o la irregularidad espectral). Precisamente, el timbre sigue siendo básicamente una sensación humana, por lo que la idea de enseñar a una máquina a describirlo y categorizarlo de manera similar resulta un problema interesante.

### **Clasificación automática**

Una de las maneras de crear y consolidar un cuerpo de conocimiento en cualquier área es a través de métodos de clasificación. La clasificación en música puede ser vista como abstracciones acerca de las funciones sociales de aspectos musicales para una cultura específica en un periodo de tiempo determinado, y así el método de clasificación sólo puede ser comprendido dentro de ese contexto. Precisamente, uno de las tareas más relevantes en la descripción de contenido en audio es la clasificación automática de acuerdo a diversos criterios. Estos sistemas de clasificación se pueden referir a elementos sonoros y musicales específicos o a descripciones semánticas más abstractas y culturalmente subjetivas. La clasificación de instrumentos musicales ha sido una constante en el desarrollo y consolidación de diversas culturas musicales a través de la historia, como lo de muestra el hecho de que uno de los primeros sistemas de clasificación de instrumentos fue implementado en uno de los dispositivos de clasificación más antiguos que se conozcan, el mandala (KARTOMI, 1990). En el contexto del MIR, una de las metas principales para estas tareas de clasificación es la de encontrar codificaciones específicas de energía física que pueden relacionarse a descripciones de alto nivel, en este caso, instrumentos musicales (LEMAN, 2008). Aunque muchos de estos modelos históricos tienen sus bases en ideas sociales, culturales o religiosas, desde el punto de vista perceptivo, un instrumento musical se relaciona intrínsecamente a la sensación tímbrica que produce. La descripción automática de una pieza musical mediante la detección de un instrumento o conjunto de instrumentos, requiere analizar directamente la fuente del

sonido físico grabado en un archivo de audio y la manera en que puede ser agrupado lingüísticamente. Puede resultar de interés entonces para los campos de la musicología, la psicoacústica o para aplicaciones musicales comerciales, poder recuperar y clasificar automáticamente de una base de datos piezas de música que hagan uso de cierto instrumento musical, sin importar el estilo musical, el género, el periodo histórico o la ubicación geográfica, o sin tener en cuenta ningún tipo de metadatos adicionales. El timbre puede ser usado como una huella acústica, al encontrar los rasgos acústicos y sus respectivos descriptores que hacen que el sonido de un instrumento sea identificable o resalte en un contexto musical específico. Perceptivamente, los instrumentos y sus timbres son relevantes a la cantidad de información en el proceso de audición. La presencia de un solo instrumento o combinación de instrumentos puede llegar a definir la textura general o la atmósfera en una pieza de música. Similarmente, la inclusión de un instrumento en una sección específica de una pieza puede crear un contraste o distinción que puede llegar a caracterizarla. Construyendo bases de datos de archivos digitales de audio polifónico que contengan timbres específicos y empleando técnicas de aprendizaje automático supervisado (*supervised machine learning*), es posible construir modelos para clasificar automáticamente piezas musicales de acuerdo al timbre. A continuación se describe una metodología propuesta para resolver este problema. Es posible entrenar clasificadores con descriptores de audio (integrados temporalmente de los valores brutos extraídos de los datos de audio polifónico) usando bases de datos extensas (FUHRMANN et al, 2009). El procedimiento es el siguiente (Fig. 1):

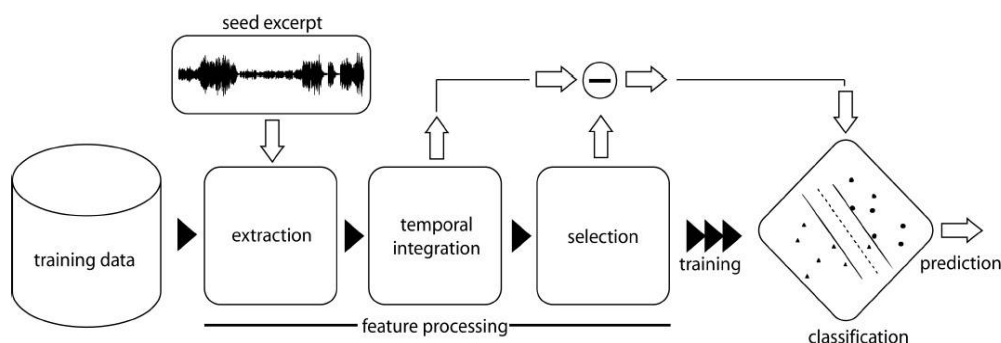


Fig. 1 Diagrama de flujo para detección automática y clasificación en audio polifónico (tomado de Furhmann, Haro and Herrera, 2009)

- Construir las bases de datos para el instrumento o timbre seleccionado con una anotación adecuada, así como una base de datos con ejemplos que no contengan el instrumento.
- Extraer los descriptores y computarlos en el tiempo mediante análisis estadístico de las bases de datos. No es necesario pre-procesamiento en los archivos, la extracción de descriptores se hace directamente sobre los archivos.
- Seleccionar los descriptores más relevantes empleando técnicas de selección de atributos.
- Se ejecutan las fases de entrenamiento, pruebas y clasificación para el conjunto de descriptores seleccionados empleando diversas técnicas de aprendizaje de máquina.
- Finalmente se analizan, evalúan y comparan descriptores, modelos, técnicas y resultados de clasificación.

Un proyecto específico que siguió esta metodología realizado en el Music Technology Group de la Universitat Pompeu Fabra en Barcelona, España, (ROMÁN, 2011) mostró resultados promisorios. Esta metodología presenta una serie de ventajas en comparación con otras técnicas que han sido investigadas en el área: se aplica a música en escenarios reales,

es decir, señales de audio polifónico que manejan una diversidad de fuentes sonoras creando una mezcla multi-tímbrica, en lugar de las técnicas de clasificación de audio monofónico en las que los instrumentos están específicamente aislados; no requiere de pre-procesamiento, por lo que se reduce considerablemente el tiempo de computación en comparación con otros métodos; otras técnicas que tratan de resolver este problema de clasificación automática, como separación de fuentes (source separation), todavía están en etapas incipientes, por lo que constituye una metodología simple y con un buen balance costo/beneficio en términos de implementación; una vez un modelo para un instrumento específico haya sido establecido, no requiere de más información del archivo de audio, es decir, puede emplearse en archivos que no contengan metadatos; la metodología puede extrapolarse a otras categorías, por ejemplo combinaciones de instrumentos, que pueden ayudar a clasificar los datos de acuerdo a taxonomías predefinidas, o a instrumentos de cualquier cultura musical del planeta, siguiendo las aproximaciones multiculturales que se están tratando de implementar en el campo del MIR.

## REFERENCIAS

- BARTÓK, B. **Escritos sobre música popular**. México D.F.: Siglo XXI Editores, 1979.
- FERRER, R. "Embodied cognition applied to timbre and musical appreciation: Theoretical foundation". **British Postgraduate Musicology**, 10, 2009.
- FUHRMANN, F., HARO M., & HERRERA P. "Scalability, generality and temporal aspects in automatic recognition of predominant musical instruments in polyphonic music". **Conference of the International Society for Music Information Retrieval (ISMIR)**, 2009.
- HERRERA, P., KLAPURI, A., DAVY, M., "Automatic Classification of Pitched Musical Instrument Sounds," In Klapuri, A., Davy, M. (Eds.) **Signal Processing Methods for Music Transcription**, Springer, New York, 2006.
- KARTOMI, M. **On Concepts and Classifications of Musical Instruments**. Chicago: University of Chicago Press, 1990.
- LEMAN, M. **Embodied Music Cognition and Mediation Technology**. Cambridge: MIT Press, 2008.
- O'CALLAGHAN, C. **Sounds**. New York: Oxford University Press, 2007.
- ROMÁN, C. **Detection of Genre-Specific Musical Instruments**. Tesis de Maestría. Barcelona: UPF, 2011.
- SCHOENBERG, A. **Tratado de Armonía**. Madrid: Real Musical, 1974.
- SETHARES, W. A. **Tuning, Timbre, Spectrum, Scale**. New York : Springer, 1997