

Perspectivas y enfoques para determinar medidas de similitud en interpretaciones musicales mediante algoritmos de análisis de datos



Perspectives and approaches to determine measures of similarity for musical performances using data analysis algorithms

Gustavo Adolfo Colmenares Pacheco.¹ & Zenaida Natividad Castillo Marrero.²

Recibido: 22-06-2020 / Aceptado: 20-07-2020 / Publicado: 07-08-2020

Abstract.

DOI: <https://doi.org/10.33262/concienciadigital.v3i3.1.1366>

The automatic characterization of a musical composition, and its interpretation, is a current line of research, due to its importance and the technological resources available and computational tools capable for detecting voice and sound. The recognition of an interpreter when listening a composition is simple for a human, but not as simple for machines, thus, this topic has kept a community of researchers in search of measures to compare and accurately recognize the composition characteristics and the interpreter. However, still a general measure has yet to be discovered, although various statistical-computational techniques have been generated and deserve to be evaluated and perhaps combined to strengthen any research on this topic. This work is the product of a comprehensive literature review that collects the main techniques and tools that have been used and proposed by researchers over the last two decades. The document will be of help to researchers who decide to undertake studies, evaluations and implementations of these tools, as well as those who wish to work on automatic recognition of interpreters and characteristics of musical compositions, or music information retrieval.

Keywords: Data analysis algorithm, Music information retrieval, Musical performance, Statistical Analysis, Mathematical and Computational models.

Resumen

La caracterización automática de obras musicales, e intérpretes de las mismas, es objeto de investigación en la actualidad. Debido principalmente a la importancia del tópico, pero

¹ Universidad de Investigación Experimental Yachay, Escuela de Matemáticas y Ciencias Computacionales. Urcuquí, Ecuador, gcolmenares@yachaytech.edu.ec

² Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Escuela de Estadística, Grupo CIED. Riobamba, Ecuador, zenaida.castillo@esepoch.edu.ec

también al auge tecnológico y la disposición de herramientas computacionales capaces de detectar voz y sonido. El reconocimiento automático de una composición, así como su intérprete y características relevantes desde el punto de vista musical, ha mantenido a una comunidad de investigadores en la búsqueda de medidas que permitan hacer comparaciones e inferir con precisión sobre las características de la composición y el intérprete; sin embargo, esta medida aún está por descubrirse, aunque se han generado diversas técnicas estadístico-computacionales que merecen ser evaluadas y quizás combinadas para fortalecer cualquier investigación en este tópico. Este trabajo, producto de una basta revisión bibliográfica, recoge las principales técnicas y herramientas que han sido utilizadas y propuestas por investigadores en las últimas dos décadas. El documento será de ayuda a los investigadores que decidan emprender estudios, evaluaciones e implementaciones de estas herramientas, así como también aquellos que deseen trabajar en el reconocimiento automático de obras musicales, sus características e intérpretes, o recuperación de información musical.

Palabras claves: Algoritmos de análisis de datos, Recuperación de información musical, Interpretación musical, Análisis Estadístico y Probabilístico, Modelos Matemáticos y Computacionales.

Introducción

A menudo nos preguntamos si es posible caracterizar automáticamente la interpretación de un pianista, considerando que cada uno tiene un lenguaje o una forma diferente de transmitir la obra que interpreta. La respuesta a esta pregunta ha sido objeto de estudio de muchos investigadores durante los últimos cuarenta años, luego de la aparición de los sistemas avanzados de grabación y el auge de microprocesadores de alta potencia que han abierto la brecha hacia el procesamiento de grandes volúmenes de datos y el uso de algoritmos matemáticos-computacionales complejos. En efecto, una grabación (señal digital) de una pieza musical corta tal como se muestra en la Figura 1, puede contener varios millones de datos. En el Cuadro 1 se muestra los datos correspondientes a la pieza musical de la Figura 1. De acuerdo a esta información puede inferirse que el procesamiento de estos datos resulta costoso a nivel de procesamiento y requiere de potentes algoritmos para el análisis.

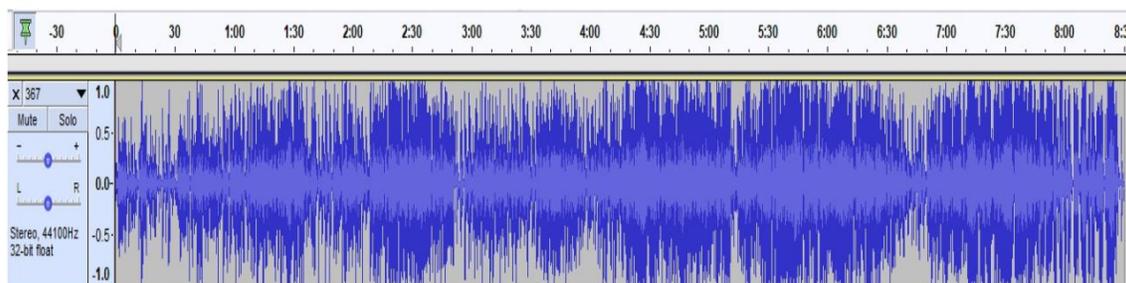


Figura 1. Señal digital de una obra completa en formato Wav (44.1 KHz)

Fuente: Audacity Software Editor

item	left	right	item	left	right	item	left	right
1	11616	11617	8	11794	11794	15	11929	11928
2	11685	11684	9	11821	11821	16	11884	11890
3	11718	11714	10	11783	11783	17	11945	11951
4	11681	11677	11	11846	11850	18	11971	11977
5	11745	11745	12	11876	11877	19	11930	11935
6	11771	11770	13	11839	11836	20	11996	12001
7	11731	11731	14	11905	11900	21	12022	12028

Cuadro 1. Extracto de los valores numéricos de las intensidades
Fuente: Software estadístico R (R Core Team, 2020)

La interpretación expresiva es una parte indispensable de la creación musical. Al interpretar una pieza, un pianista experto da forma a diversos parámetros (tempo, ritmo, dinámica, tono, articulación, etc.) de manera que no están prescritas en la partitura diseñada, produciendo así una interpretación expresiva que pone de manifiesto las cualidades dramáticas, afectivas, y emocionales que percibe oyente. Por ello, el estudio científico de la interpretación musical, a partir de grabaciones sonoras, se transforma en un análisis minucioso para extraer características y así determinar medidas que permitan establecer diferencias y similitudes interpretativas en ejecuciones similares, ver por ejemplo Figura 2. Capturar las propiedades estilísticas e interpretativas de un determinado pianista, ofrece, por tanto, una amplia variedad de caminos y enfoques dentro del área de la investigación científica.

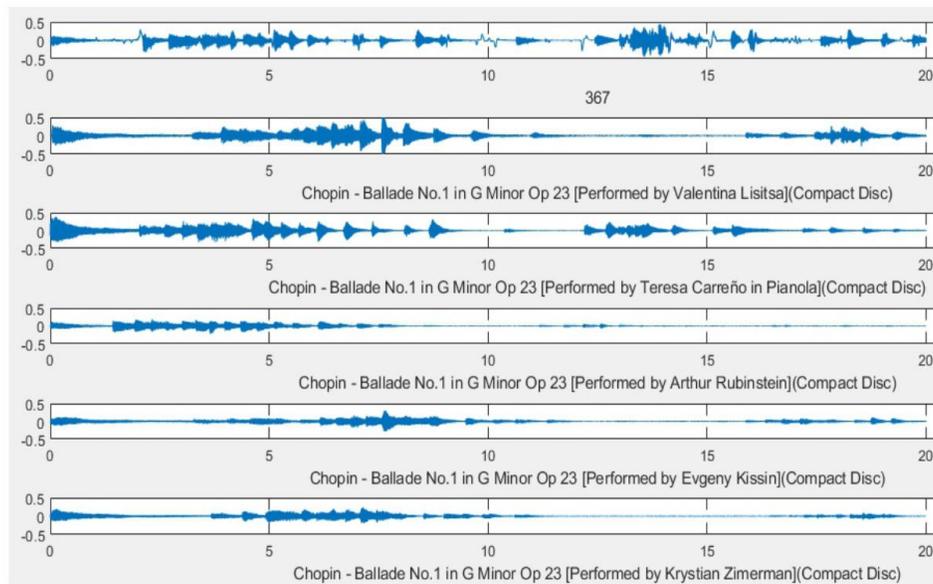


Figura 2. Comparación de señales de una obra con varios intérpretes
Fuente: Free tool Audacity Audio Software Editor

En este artículo se presenta un panorama actualizado del estado de la técnica en este contexto. Se exploran las tendencias recientes con un mayor énfasis en aquellas que usan algoritmos basados en el análisis de datos, a fin de establecer las limitaciones actuales y explorar nuevas

vías que conlleven al establecimiento de métodos que resulten eficaces en la caracterización de algunas obras musicales.

Metodología

Se plantea en este trabajo una revisión descriptiva sobre algoritmos y métodos existentes para el tratamiento de datos de composiciones musicales, con el fin de extraer información sobre la obra o el intérprete. El objetivo principal es el análisis del estado del arte en cuanto a técnicas estadístico-computacionales que están siendo usada por los investigadores y a su vez implementadas en software comerciales. Está dirigido a mostrar los principales desarrollos computacionales en el tópico.

La revisión fue sistemática, y se desarrolló de acuerdo a las siguientes etapas:

- 1) Definición de los objetivos y alcance.
- 2) Definición de las estrategias de búsqueda.
- 3) Búsqueda bibliográfica: Bases de datos; fuentes documentadas e información en la web.
- 4) Especificación y aplicación de criterios de selección.
- 5) Organización de la información. Detección de información relevante.
- 6) Escritura del artículo.

Características musicales e Interpretación

Una medida de la expresividad musical es una codificación numérica de una o varias características de la interpretación de un músico. Según Widmer y Goebel (2004) y Kirke y Miranda (2013) los descriptores más comunes que se pueden identificar en una onda de audio musical son intensidad, tiempo, ritmo, detección de picos, análisis de tonos, dinámica. En Raphael (2009) y Ohishi et al. (2014) se agregan otros parámetros como las diferencias tímbricas y análisis armónico. Todos estos descriptores vistos en su dimensión individual se denominan **características musicales**. Hablar de **interpretación musical** implica interacción multidimensional de las características mencionadas. Analizar similitudes y diferencias entre diferentes ejecutantes de una misma obra significa entonces encontrar una medida de la expresividad de cada intérprete. Un modelo computacional de la interpretación musical es por lo tanto cualquier función matemático-computacional que asigne una medida o permita hacer una predicción de las características de una obra musical en su ámbito multidimensional.

Para analizar la interpretación musical primero debe realizarse la detección y análisis de las características musicales (análisis espectral, análisis estadístico, modelos probabilísticos, redes neuronales artificiales, etc.) para posteriormente generar modelos que propongan medidas que caractericen la interpretación.

Análisis espectral y procesamiento digital de señales

Benetos et al. (2013), Dressler (2012) y Bay et al. (2009) aplican diversos algoritmos y métodos del análisis espectral como Music Information Retrieval Evaluation eXchange (MIREX)

y la estimación multiple-F0 para descubrir y analizar componentes de las frecuencias de audio. Estas investigaciones han producido resultados en la detección y extracción de características musicales como tonos, intensidades, acordes, duración, gradaciones de sonidos y detección de picos.

Para analizar las interpretaciones musicales y en particular la expresión del ejecutante se plantean estudios basados en la comparación de diferentes interpretaciones de la misma pieza por diferentes intérpretes, lo que permite estudiar similitudes y diferencias entre cada intérprete. Un enfoque sustentado en el procesamiento digital de señales busca codificaciones de cada intérprete a través de la comparación y análisis de parámetros expresivos en curvas alineadas. El uso de algoritmos de Alineamiento Temporal Dinámico DTW (*Dynamic Time Warping*, por sus siglas en inglés) ha contribuido a la detección de características musicales mediante la comparación de segmentos diferentes y cortos de la onda de audio [Adams, Bartsch, Shifrin, y Wakefield (2004), Muda, Begam, y Elamvazuthi (2010), Holzapfel y Stylianou (2008)].

DTW es un algoritmo que ha tenido gran popularidad desde su introducción a finales de la década de 1970 [Sakoe, H., & Chiba, S. (1978)] por ser computacionalmente eficiente y proveer una medida de similitud de series temporales que minimiza los efectos del desplazamiento y la distorsión en el tiempo, lo cual permite la transformación elástica de las series temporales para detectar formas similares con fases diferentes, como puede verse en la Figura 3. El algoritmo utiliza una técnica de programación dinámica para alinear dos series temporales X de tamaño m y Y de tamaño n con una plantilla determinada, de modo que se minimice una medida de distancia total. Para alinear estas secuencias se define una matriz de costo en la que cada celda (i, j) representa las distancias a pares entre el i -ésimo componente de X y el correspondiente j -ésimo Y . Generalmente se utiliza la distancia euclidiana para construir la matriz de costo, la función de costo (distancia) tiene un valor pequeño cuando las secuencias son similares y un valor grande si son diferentes. En la práctica, se aplican restricciones para acelerar el algoritmo, reduciendo el número de caminos a ser considerados durante el cómputo.

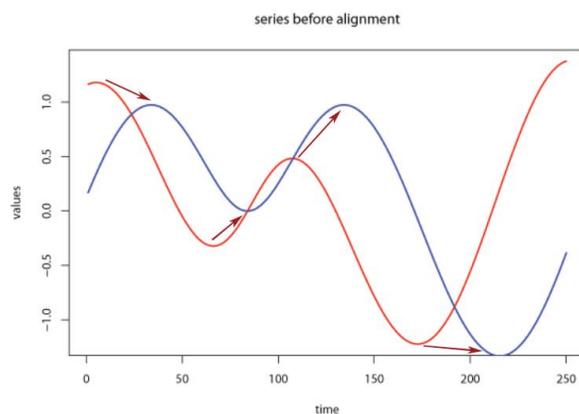


Figura 3. Alineación con DTW
Fuente: Senin, P. (2008).

Aunque DTW ha sido aplicado exitosamente para comparar palabras aisladas y diversos patrones del habla en el reconocimiento de voz automatizado y también reconocimiento de gestos corporales y faciales, su aplicación en el contexto del análisis de señales de audio presenta un alto costo computacional, es por ello que las investigaciones que sustentan el uso de este algoritmo se limitan a extractos musicales de pocos segundos. Cuando se buscan similitudes o diferencias interpretativas en ejecuciones similares de una misma obra, los segmentos sonoros a comparar deben ser escogidos con criterio musicológico (frases, temas o ideas musicales) lo cual se traduce en archivos de datos muy grandes (un minuto de una onda grabada con una frecuencia de muestreo de 44.1 KHz generan un vector de 2.646.000 datos).

La complejidad de DTW es orden n^2 lo cual lo transforma en un problema de big data, por lo tanto, su implementación requiere máquinas de alto rendimiento o HPC (High Performance Computing, por sus siglas en inglés).

Métodos estadísticos, distribuciones de probabilidad y cadenas de Markov

Los enfoques estadísticos y probabilísticos, por otra parte, intentan describir las características expresivas presentes en una señal musical mediante el análisis de parámetros estadísticos y la modelización de la distribución de probabilidad de los datos. Las señales musicales reales son muy variables, pero tienen una estructura intrínsecamente numérica.

Sapp (2007) presenta un enfoque gráfico comparativo, análisis de series temporales y medición de parámetros estadísticos como la media y el coeficiente de correlación para establecer diferencias interpretativas de 89 interpretaciones de una misma obra. En Temperley (2009) se propone un modelo probabilístico de la música polifónica tomando un patrón de notas como entrada y combinando tres aspectos técnicos de la música (análisis métrico, análisis armónico y flujo). El modelo también arroja una estimación de la probabilidad del propio patrón de notas. En Fu et al. (2015) se analizan acordes musicales usando estimación de tiempo local y análisis de varianza.

En Yang, Lin, Su, y Chen (2008) se analiza un conjunto de datos a partir de 195 canciones populares, especialmente escogidas por sus características expresivas. Los segmentos son convertidos a un formato uniforme (22.050 Hz, 16 bits, and mono channel PCM WAV) y normalizado al mismo nivel de volumen. Luego se usa el algoritmo DWCH (Daubechies Wavelet Coefficient Histograms) y algoritmos estándares de contraste espectral, disponibles en paquetes de software comercial, para extraer características musicales y construir un espacio multidimensional. Las entradas se utilizan para entrenar tres algoritmos de regresión: regresión lineal múltiple (MLR), regresión de soporte vectorial (SVR) y AdaBoost (BoostR).

En Stamatatos (2002) se presenta una comparación de características para discriminar entre los diferentes intérpretes de una misma pieza. Basándose en el cálculo de la media simple y la media ponderada, sobre un conjunto de datos de una pieza interpretada por 22 pianistas,

en dos grupos, donde uno de ellos es tomado como referencia. Se encuentra un estándar de interpretación o interpretación promedio entre los pianistas y la partitura de la obra, calculado a partir de las desviaciones en términos de tiempo, articulación y dinámica. En cuanto a la dinámica, el autor aclara que es definida considerando únicamente el nivel de sonido o intensidad de la nota tocada.

Kim et al. (2011) se basan en la afirmación de que la expresión en la música de piano se puede representar sólo con tempo, volumen (intensidad) y duración, para desarrollar un sistema automático de interpretación pianística denominado Polyhymnia. La entrada de este programa es un archivo digital de una partitura de piano en formato MusicXML, un tipo de archivo similar al MIDI (Musical Instrument Digital Interface) que es muy utilizado para archivar partituras musicales en formato digital. Las marcas de dinámica son mapeadas de los valores de ataque de las notas (MIDI velocity). Se hace una formulación probabilística condicional y se utilizan Campos Aleatorios Condicionales (*Conditional Random Fields* por sus siglas en inglés). De acuerdo a los autores, Polyhymnia puede aprender y predecir la expresión polifónica e interpretar los símbolos musicales automáticamente.

Cabe hacer notar que en estas investigaciones se escogen características individualizadas de la ejecución musical, pero ninguno ha afrontado el problema de la multidimensionalidad de la expresividad musical.

Otro enfoque ampliamente propuesto en diversas investigaciones es la utilización de variantes de los Modelos Ocultos de Markov (HMM por sus siglas en inglés) para describir la evolución temporal de interpretaciones musicales y reconocimiento de voz.

Grindlay y Helmbold (2006) utilizan las HMM combinadas con *Dynamic Bayesian Network* para modelar la relación estadística entre la estructura de la composición musical y las variaciones interpretativas. Para producir estas relaciones automáticamente, a partir de un cuerpo de datos de entrenamiento, se intenta que el modelo pueda ser utilizado para generar actuaciones expresivas de piezas nuevas, no incluidas en el conjunto de entrenamiento. En la investigación sólo se evalúan las variables de tiempo y velocidad de la música en dos conjuntos de datos en formato MIDI. El primero con 28 interpretaciones de la pieza *Traümerei* de Robert Schumann, extraídos de discos compactos comerciales de reconocidos pianistas. El segundo, extraído de grabaciones realizadas a estudiantes de piano.

Teramura et al. (2008); Gu y Raphael (2012); Okumura et al. (2014) plantean modelos que suponen la distribución de probabilidad subyacente de los elementos expresivos de una interpretación musical como gaussiana. Un enfoque diferente es el adoptado por Kim et al. (2013) y Moulieras y Pachet (2016), que utilizan modelos de entropía para modelar la distribución de probabilidad de los datos musicales.

Molina-Solana et al. (2008) identificaron intérpretes de violín basados en análisis de descriptores (energía & duración) extrayendo elementos expresivos y construyendo modelos de tendencias. En la fase de entrenamiento se extraen descriptores de audio por segmentos

melódicos etiquetados, usando el modelo *Implication-Realization* (IR) de E. Narmour, y se procede al cálculo de distribuciones probabilísticas para cada patrón y descriptor IR (generación de tendencias). En la fase de identificación, el sistema analiza el rendimiento de entrada y busca el modelo más similar aprendido previamente. Los resultados reportados, muestran que el enfoque es capaz de alcanzar altas tasas de identificación.

Fukayama et al. (2011) usaron campos aleatorios condicionales y cadenas de Markov en archivos MusicXML para modelar tono, duración y nota. Asumieron como expresión melódica: tempo instantáneo, volumen y duración. Alvarado et al. (2016) trabajaron con un enfoque de regresión bayesiana usando procesos gaussianos para calcular la distribución posterior, sobre una función estocástica que estima tonos y cambios de la envolvente en el tiempo. Kumar Ashis Pati et al. (2018) construyeron modelos de regresión basados en redes neuronales profundas capaces de predecir calificaciones dadas por jueces humanos expertos para interpretaciones de instrumentos de viento.

Machine Learning y Principal Component Analysis

De particular interés, en las últimas décadas, han resultado los algoritmos basados en análisis de componentes principales y Machine Learning. En Tobudic (2002) y Widmer (2003) los autores proponen el enfoque de aprendizaje automático basado en datos para la elaboración de modelos de rendimiento de la interpretación musical. A través del aprendizaje por instancias y la programación lógica inductiva, el sistema aprende reglas explícitas simples y complejas de rendimiento a partir de grandes colecciones de datos reales, por ejemplo, notas, fraseo, estructuras rítmicas, etc. Una vez terminada la fase de aprendizaje, el sistema puede inferir mapeos de puntuación de rendimiento. En contraste con esto, el enfoque basado en reglas fue propuesto en el modelo KTH desarrollado en el KTH Royal Institute de Estocolmo (Friberg et al., 2006), donde el sistema realiza una combinación de reglas de rendimiento, las cuales han sido diseñadas manualmente a partir de juicios de expertos y pruebas de escucha.

El éxito de las Redes Neuronales Artificiales y métodos de Deep Learning, como la visión por ordenador y el procesamiento del lenguaje natural, han sido explorados en Humphrey et al. (2012) y en Schlüter (2017) para la recuperación de información musical. Otros enfoques, tratados por Cancino Chacón y Grachten, (2015) y Giraldo S. y Ramírez (2016) intentan modelar la expresión de un intérprete musical utilizando redes neuronales de avance (FFNN por sus siglas en inglés) para predecir características expresivas en base a una puntuación. Es de hacer notar que en estos sistemas se agregan características que describen el contexto rítmico/melódico subyacente a una nota.

Nakamura et al. (2017) aplican HMM y aprendizaje no supervisado para la estimación de los valores de nota en la transcripción musical de una interpretación polifónica proveniente de un archivo MIDI, mientras que Patricio y Chew (2014) exploran la cuantificación de las diferencias de estilo interpretativo utilizando HMM en doce interpretaciones del Preludio de la suite de J. S. Bach en Sol mayor, repartidas entre violoncello barroco y violoncello

moderno. Se consideraron en esa investigación sólo las variables musicales de tiempo y volumen.

Las investigaciones que aplican Análisis de Componentes Principales (PCA, por sus siglas en inglés) muestran importantes resultados para reconocer determinadas formas en el espectro de una onda. Lo importante de esta herramienta de análisis es que reduce en gran medida la dimensionalidad, un factor crucial para mejorar el costo computacional de ondas con grandes volúmenes de datos. Liem y Hanjalic (2015) utilizan PCA para alinear diferentes fragmentos y detectar variaciones y similitudes en interpretaciones de grabaciones orquestales. Zurek, Gamarra (2016) identifican similitudes acústicas en ondas sonoras reduciendo la dimensión de características extraídas con la implementación posterior de un clasificador basado en el método de los k-vecinos más cercanos (KNN) y una red neuronal artificial.

Resultados y Conclusiones

El resultado de estudios particulares sobre un tópico generalmente es sesgado hacia las herramientas que se usan; y el conocimiento de diferentes herramientas y estudios sobre un tópico capacita al investigador para discernir y combinar con el propósito de lograr mejores resultados en su campo de investigación.

El presente documento presentó:

- 1) Una investigación bibliográfica sobre las principales metodologías, algoritmos y tratamiento de datos provenientes de obras musicales.
- 2) El resultado de esta investigación permitió resumir y presentar, en forma precisa, elementos y aspectos del tópico de reconocimiento automático de características de obras musicales y sus interpretaciones.
- 3) Se contrastan algunas técnicas computacionales presentando fortalezas y debilidades de las mismas.
- 4) De acuerdo a las características de las composiciones musicales se orienta hacia el uso de técnicas apropiadas según el estado del arte.
- 5) Se soporta la información con bibliografía selecta, específica y relevante, que puede ser utilizada en lo sucesivo para abordar el tema en otras investigaciones de carácter teórico-práctico.

Referencias Bibliográficas

- Adams, N. H., Bartsch, M. A., Shifrin, J., & Wakefield, G. H. (2004, October). Time Series Alignment for Music Information Retrieval. In ISMIR.
- Alvarado, P. A., & Stowell, D. (2016, September). Gaussian processes for music audio modelling and content analysis. In 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.
- Bay, M., Ehmann, A. F., & Downie, J. S. (2009, October). Evaluation of multiple-f0 estimation and tracking systems. In ISMIR (pp. 315-320).

- Benetos, E., Dixon, S., Giannoulis, D., Kirchhoff, H., & Klapuri, A. (2013). Automatic music transcription: challenges and future directions. *Journal of Intelligent Information Systems*, 41(3), 407-434.
- Dressler, K. (2012). Multiple fundamental frequency extraction for MIREX 2012. Eight Music Information Retrieval Evaluation Exchange (MIREX).
- Friberg, A., Bresin, R., and Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Adv. Cogn. Psychol.* 2, 145–161.
- Fu, M., Xia, G., Dannenberg, R. B., & Wasserman, L. A. (2015). A Statistical View on the Expressive Timing of Piano Rolled Chords. In *ISMIR* (pp. 578-583).
- Giraldo, S., & Ramirez, R. (2016). A machine learning approach to ornamentation modeling and synthesis in jazz guitar. *Journal of Mathematics and Music*, 10(2), 107-126.
- Grindlay, G., & Helmbold, D. (2006). Modeling, analyzing, and synthesizing expressive piano performance with graphical models. *Machine learning*, 65(2-3), 361-387.
- Gu, Y., & Raphael, C. (2012, October). Modeling Piano Interpretation Using Switching Kalman Filter. In *ISMIR* (pp. 145-150).
- Holzappel, A., & Stylianou, Y. (2008, March). Rhythmic similarity of music based on dynamic periodicity warping. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 2217-2220). IEEE.
- Humphrey, E. J., Bello, J. P., & LeCun, Y. (2012, October). Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. In *ISMIR* (pp. 403-408).
- Kim, T. H., Fukayama, S., Nishimoto, T., & Sagayama, S. (2011). Polyhymnia: An Automatic Piano Performance System with Statistical Modeling of Polyphonic Expression and Musical Symbol Interpretation. In *NIME* (pp.96-99).
- Kim, T. H., Fukayama, S., Nishimoto, T., & Sagayama, S. (2013). Statistical approach to automatic expressive rendition of polyphonic piano music. In *Guide to Computing for Expressive Music Performance* (pp. 145-179). Springer, London.
- Liem, C. C., & Hanjalic, A. (2015). Comparative analysis of orchestral performance recordings: an image-based approach. In *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015, Malaga, Spain, October 26-20, 2015*. Eds. Muller, M., Wiering, F. IEEE.
- Molina-Solana, M., Arcos, J. L., & Gomez, E. (2008). Using Expressive Trends for Identifying Violin Performers. In *ISMIR* (pp. 495-500).
- Moulieras, S., & Pachet, F. (2016). Maximum entropy models for generation of expressive music. *arXiv preprint arXiv:1610.03606*.

- Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv preprint arXiv:1003.4083.
- Nakamura, E., Yoshii, K., & Dixon, S. (2017). Note value recognition for piano transcription using markov random fields. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(9), 1846-1858.
- Ohishi, Y., Mochihashi, D., Kameoka, H., & Kashino, K. (2014, May). Mixture of Gaussian process experts for predicting sung melodic contour with expressive dynamic fluctuations. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3714-3718). IEEE.
- Okumura, K., Sako, S., & Kitamura, T. (2014). Laminae: A stochastic modeling-based autonomous performance rendering system that elucidates performer characteristics. In *ICMC*.
- Patricio, C. V., & Chew, E. Application of Hidden Markov Models to music performance style classification via timing and loudness. *Scientific committee*, 40.
- Raphael, C. (2009, July). Representation and Synthesis of Melodic Expression. In *IJCAI* (pp. 1475-1480).
- R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1), 43-49.
- Senin, P. (2008). Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855(1-23), 40.
- Schlüter, J. (2017). Deep learning for event detection, sequence labelling and similarity estimation in music signals/submitted by Jan Schlüter (Doctoral dissertation, Universität Linz).
- Sapp, C. S. (2007, September). Comparative Analysis of Multiple Musical Performances. In *ISMIR* (pp. 497-500).
- Stamatatos, E. (2002). Quantifying the differences between music performers: Score vs. norm. In *ICMC*.
- Teramura, K., Okuma, H., Taniguchi, Y., Makimoto, S., & Maeda, S. I. (2008). Gaussian process regression for rendering music performance. *Proc. ICMPC*, 167-172.
- Temperley, D. (2009). A unified probabilistic model for polyphonic music analysis. *Journal of New Music Research*, 38(1), 3-18.
- Yang, Y. H., Lin, Y. C., Su, Y. F., & Chen, H. H. (2008). A regression approach to music

- emotion recognition. IEEE Transactions on audio, speech, and language processing, 16(2), 448-457.
- Tobudic, A., and Widmer, G. (2006). Relational IBL in classical music. Mach. Learn. 64, 5–24. doi: 10.1007/s10994-006-8260-4.
- Widmer, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. Artificial Intelligence, 146(2), 129-148.
- Widmer, G., & Goebel, W. (2004). Computational models of expressive music performance: The state of the art. Journal of New Music Research, 33(3), 203-216.
- Zurek, E. E., Gamarra, M. R., Escorcía, J. R., Gutierrez, C., Bayona, H., Pérez, R., & García, X. (2016). Análisis Espectral Para El Reconocimiento De Huellas Acústicas. Revista de Investigaciones Universidad del Quindío, 28(1), 116-122.

PARA CITAR EL ARTÍCULO INDEXADO.

Colmenares Pacheco, G. A., & Castillo Marrero, Z. N. (2020). Perspectivas y enfoques para determinar medidas de similitud en interpretaciones musicales mediante algoritmos de análisis de datos. *ConcienciaDigital*, 3(3.1), 75-86.
<https://doi.org/10.33262/concienciadigital.v3i3.1.1366>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Conciencia Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Conciencia Digital**.

