



Dr. Roberto Carlos Luján Villar

Universidad Del Valle, Cali, Colombia

@ janlujazz@gmail.com

ID 0000-0001-6435-4412

Dr. Juan David Luján Villar

Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

@ lujanvillar@gmail.com

ID 0000-0001-6435-4412

■ Recibido / Received
5 de mayo de 2019

■ Aceptado / Accepted
30 de junio de 2019

■ Páginas / Pages
De la 59 a la 82

■ ISSN: 1885-365X

Reconocimiento de emociones musicales a través de datos y tecnologías digitales

Music emotion recognition through data and digital technologies

Resumen

Este estudio explora el campo de investigación denominado reconocimiento de emociones musicales, una perspectiva transdisciplinaria que aborda la investigación de los estados anímicos y las emociones musicales a través de la recuperación de datos y diversos tipos de análisis informáticos y analógicos. Se plantean algunas preguntas con la finalidad de exponer los principales presupuestos y recientes perspectivas, lo cual guía a nivel metodológico dos experimentos que exponen los planteamientos centrales de este tipo de enfoques y evidencian una serie de posibles aplicaciones prácticas. Los resultados indican diversas posibilidades de investigación que apuntan a la consideración de este tipo de enfoques como alternativas complementarias para afrontar los retos actuales del conocimiento de la música, una de las principales manifestaciones culturales que incide en la alteración de la experiencia emocional de una persona.

PALABRAS CLAVE: reconocimiento de emociones musicales, emociones, estados anímicos musicales, computación afectiva.

Abstract

This study explores the field of Music Emotion Recognition, a transdisciplinary perspective that addresses the investigation of moods and musical emotions through the recovery of data and various types of computer and analog analysis. Some questions are posed in order to present the main assumptions and new perspectives, which guides at a methodological level two experiments that expose the central approaches of this type of providers and show a series of practical applications. The results indicate research possibilities that point to the consideration that it is complementary alternatives to face the learning challenges of music, one of the main cultural manifestations that affects the alteration of a person's emotional experience.

KEY WORDS: music emotion recognition, emotions, music moods, affective computing.

1. Introducción

El reconocimiento de emociones musicales (REM) (*Music Emotion Recognition MER*) es un campo de investigación científica reciente y en plena evolución. Se deriva de dos campos mayores: 1) la recuperación de información musical (RIM) (*Music Information Retrieval MIR*); y 2) la computación afectiva (*Affective Computing*) (Calvo, D'Mello, Gratch y Kappas, 2014; Picard, 2000). A grandes rasgos se puede decir que el REM gira en torno a varias ideas respecto a la comprensión psicológica de la relación entre el afecto humano y la música, de lo cual se desprenden tres ideas centrales de investigación; 1) la emoción es *percibida* e identificada con la escucha musical; 2) la emoción es *inducida* y experimentada mediante la representación de la escucha musical, lo cual difiere de la emoción percibida; y 3) la emoción es transmitida mediante una intención comunicativa explícita (Yang y Chen, 2011; Yang, Dong y Li, 2018). Como se puede comprender a partir de lo anterior, las emociones percibidas son más intersubjetivas y clasificables que las emociones experimentadas más intrasubjetivas, arraigadas de manera cultural y dependientes de la memoria (Aljanaki, 2016, p. 22-23; Panda, Malheiro y Paiva, 2018a).

Una de las ideas centrales del REM radica en la capacidad de poder determinar mediante sistemas automáticos ingresando diversos datos (señales musicales) y variables (parámetros computacionales), cuáles y qué tipo de emociones son percibidas desde las composiciones musicales, e intentan percibir cómo cada una de las formas de sus rasgos estructurales pueden producir cierto tipo de reacciones características en los oyentes.



Tabla 1. Rasgos acústicos y conceptos musicales

Concepto musical	*Rasgos
Timbre	<i>MFCCs, Δ MFCCs, $\Delta\Delta$ MFCCs, spectral features (centroid, shape, spread, skewness, kurtosis, contrast, flatness), tristimulus, brightness, 95% rolloff, zero crossing rate, octave-based spectral contrast, Daubechies wavelets coefficient histogram (DWCH), auditory modulation features, inharmonicity, roughness, dissonance, odd to even harmonic ratio</i>
Volumen	<i>RMS energy, specific loudness on Bark critical bands</i>
Armonía	<i>Chromagram, chromagram peak, key, mode, key clarity, harmonic change, chords, HCDF</i>
Altura tonal	<i>Spectral centroid, low energy</i>
Ritmo	<i>Tempo (bpm), beat histograms, rhythm regularity (autocorrelation on onset detection curve), rhythm strength, onset rate</i>
Articulación	<i>Attack slope, attack time</i>

*Los nombres de los algoritmos se mantienen en su idioma original

Fuente: elaborado a partir de Aljanaki (2016, p. 52).

La Tabla 1 presenta los principales conceptos musicales indagados en el REM y sus rasgos (algoritmos) convencionales usados por la comunidad de especialistas en este campo. Los estudios revelan que las personas relacionan los estados anímicos percibidos

y experimentados con etiquetas y adjetivos, aunque se advierte que el factor temporal es primordial en su intensidad y características de duración, por estas razones a nivel metodológico en este contexto, a las emociones se le asigna un atributo que mide la intensidad de la emoción en una escala determinada (Deng, Leung, Milani y Chen, 2015).

En la actualidad, existe una diferencia específica proveniente de la psicología de la música sobre la concepción de los términos emoción (*emotion*) y estados anímicos (*mood*) en el abordaje de los modelos REM, una discusión que se extiende de manera general al campo de la psicología de la música (Juslin, 2019, p. 46-48). En este sentido, las emociones son concebidas como una reacción breve a nivel temporal (minutos u horas) e intensas a propósito de la percepción derivada de un estímulo, sujeto u objeto específico. A diferencia, los estados anímicos precisan estados emocionales más duraderos (días, semanas o meses), son menos intensos, y reflejan una emotividad definida por una inclinación afectiva global, debido a que no dependen de la reacción a un objeto puntual o un estímulo contextual.

Las perspectivas del REM giran en torno a las ciencias computacionales y la teoría psicológica de la música, sobre lo cual existen como mínimo tres enfoques y abordajes:

- Los modelos categóricos basados en la propuesta inicial de Hevner (1936);
- Los modelos dimensionales diseñados por Russell (1980) y Thayer (1989); y
- Las propuestas híbridas que configuran la detección de la variación músico-emocional (DVME).

Entre las estrategias más usuales se encuentran los conjuntos de anotaciones músico-emocionales realizadas por psicólogos especialistas sobre extensos corpus musicales y la clasificación a través de etiquetas o *tags* emocionales, realizadas por comunidades de usuarios a partir de canciones específicas, sobre lo cual se han creado estrategias como juegos o portales interactivos de Internet y redes sociales dedicadas a recomendaciones musicales, un campo ligado a las exploraciones REM (Deng et al., 2015, Yang et al., 2018). Por ejemplo, aplicaciones como *Emotion Face Clock* (Schubert, Ferguson, Farrar, Taylor y McPherson, 2013), *MoodSwings* (Speck, Schmidt, Morton y Kim, 2011), o concursos de evaluación de modelos investigativos, algoritmos de optimización e investigaciones empíricas entre las cuales figuran el anual *Music Information Retrieval Evaluation eXchange* (MIREX), una comunidad ampliamente reconocida en la evaluación RIM, y el workshop *MediaEval Database for Emotional Analysis in Music* (DEAM) (Aljanaki, 2016), dedican sus esfuerzos a focalizar a través de diferentes perspectivas, la medición y predicción de emociones musicales y procesos relacionados.

Los sistemas dedicados a la recuperación de datos basados en las emociones o *Emotion-Based Retrieval Tools* conforman un importante campo de la investigación y práctica del REM. Algunas de estas estrategias de aplicación son los sistemas de recomendación musical a través de la creación de comunidades de melómanos, críticos y músicos, la generación de *playlists* o listados musicales automáticos con fuertes rasgos que parten de la inteligencia artificial y la categorización musical con fines de evaluación, consumo y crítica musical.



2. Modelos investigativos en el campo del REM

Entre los modelos y enfoques de investigación del REM se encuentran los modelos que incluyen perspectivas a las cuales se denomina; categóricas, dimensionales y mixtas. Entre los modelos mixtos –menos descritos y abordados–, tanto a nivel de objetos (aspectos musicales, textuales y material videasonoro) y enfoques se encuentran, por ejemplo, el modelo de detección de la variación músico-emocional (*Music Emotion Variation Detection*, MEVD), el cual es una sofisticada perspectiva multimetodológica. La Tabla 2 presenta un cuadro descriptivo general de las metodologías usuales del REM, la conceptualización empleada en el campo de las emociones y la descripción objetiva de cada estrategia investigativa.

Tabla 2. Comparación de la investigación sobre la predicción automática de emociones musicales

Metodología	Conceptualización de la emoción	Descripción
MER categórico	Categórico	Predice las clases <i>discretas</i> de etiquetas de piezas musicales
MER dimensional	Dimensional	Predice los valores numéricos emocionales de piezas musicales
Detección de la variación músico-emocional (<i>music emotion variation detection</i> , MEVD)	Dimensional	Predice la variación emocional continua de cada pieza musical

Fuente: elaborado a partir de Yang y Chen (2011, p. 22).



En todos estos modelos se pueden encontrar diversos procesos algorítmicos, la mayoría de acceso abierto al público a través de plataformas como MATLAB o Python. Un punto de referencia son los trabajos que emplean un gran número de algoritmos de clasificación; regresión de vectores de soporte (*Support Vector Regression*, SVR), regresión lineal (*Linear Regression*), filtrado Kalman (*Kalman Filtering*), procesos gaussianos (*Gaussian Processes*), campos aleatorios condicionales (*Conditional Random Fields*, CRF), campos neuronales condicionales continuos (*Continuous Conditional Neural Fields*, CCFN), y memoria a corto y largo plazo en redes neuronales recurrentes (*Long Short-Term Memory Recurrent Neural Networks*), entre muchos otros procesos de computo (Aljanaki, 2016; Grekow, 2018; Laurier, 2011; Panda, Malheiro y Paiva, 2018b; Schedl, Gómez y Urbano, 2014; Yang y Chen, 2011). Sin embargo como advierte Grekow (2018, p. vii) el REM requiere familiarización con diversos campos distantes –en la mayoría de los casos–, corrientes heterogéneas que plantean campos problemáticos a la altura de los recursos disponibles, una tarea compleja y tremendamente emocionante.

3. Modelos categóricos

Entre los modelos categóricos más socorridos en el MER se encuentran el clásico mapeado de los ocho *clusters* emocionales propuestos tempranamente por Hevner (1936). Sobre esta perspectiva se realizaron diversos abordajes en los cuales se cuantificó cada uno de los *clusters* encontrados sobre el valor afectivo y la expresividad de la música, a partir de las categorías discretas propuestas en esta exploración inicial. Esta génesis de la utilidad de los modelos categóricos demostró la influencia de los modos en las percepciones musicales con relación a su papel afectivo. Según Hevner el modo mayor está fuertemente asociado con la felicidad, la alegría, la diversión y la honestidad. Por el contrario, el modo menor se asocia con la tristeza, el anhelo sentimental, y demás efectos sensibles.

Figura 1. Palabras que figuran en el círculo de Hevner aglutinadas en ocho grupos

7	6	5
exaltado	alegre	humorístico
altísimo	jubiloso	juguetón
triumfal	jovial	caprichoso
dramático	feliz	fantasioso
apasionado	animado	pintoresco
sensacional	brillante	animado
agitado		delicado
emocionante		ligero
impetuoso		agraciado
8		4
vigoroso		lírico
robusto		sin prisa
enfático		satisfactorio
marcial		tranquilo
pesado		sereno
majestuoso		calmado
exaltado		
1	2	3
espiritual	patético	ensueño
sublime	dolorido	rendimiento
imponente	triste	tierno
digno	afligido	sentimental
sagrado	trágico	nostálgico
solemne	melancólico	anhelo
sobrio	frustrado	suplicante
serio	depresivo	lamentable
	sombrio	
	intenso	
	oscuro	



Fuente elaboración a partir de Hevner (1936, p. 249).

En estos sistemas las emociones son categorías y están definidas por un número discreto de adjetivos, palabras clave y etiquetas emocionales y afectivas que describen a nivel estadístico las emociones musicales descritas. Aunque el potencial de los enfoques categóricos en algunos casos se reduce a una lógica descriptiva donde una categoría contiene un conjunto completo de varios tonos de emociones, lo cual puede llegar a un extremo de simplificación, las bondades de la categorización son evidentes en el sentido de otorgar un orden coherente con la percepción y expresividad musical humana (Juslin, 2019, p. 63), a propósito de los diseños experimentales de laboratorio.

La Figura 1 presenta el resultado del experimento de Hevner (1936) quién agrupó según la afinidad de cada emoción un total de 66 adjetivos en 8 *clusters*. Este es un modelo de orden distributivo y se basa en una estadística donde el proceso de combinar los adjetivos en grupos homogéneos es decir, semejantes a nivel emocional, busca su agrupamiento para un tratamiento estadístico ulterior, lo cual establece el carácter sistemático de los significados musicales. Los avances a este respecto no se detienen y desarrollos categóricos posteriores tomaron como base el trabajo de Hevner. Existe un reto sobre la brecha lingüística existente entre el inglés y el español como lenguas que varían de manera notoria entre un concepto y otro; a este respecto es fundamental sostener sobre una base epistemológica y teórica la pertinencia de los conceptos y sus significados contextualizados.

Tabla 3. Cinco clusters de estados anímicos usados en el MIREX

Cluster	Etiquetas de los estados anímicos
MIREX_C_1	Apasionado, entusiasta, confiado, bullicioso, ruidoso.
MIREX_C_2	Rizado, alegre, divertido, dulce, amable/afable
MIREX_C_3	Alfabetizado, conmovedor, melancólico, agridulce, otoñal, melancólico
MIREX_C_4	Chistoso, tonto, exagerado, peculiar, caprichoso, ingenioso, irónico
MIREX_C_5	Agresivo, ardiente, tenso/ansioso, intenso, volátil, visceral

Fuente: elaborado a partir de Hu, Downie, Laurier, Bay y Ehmann (2008).

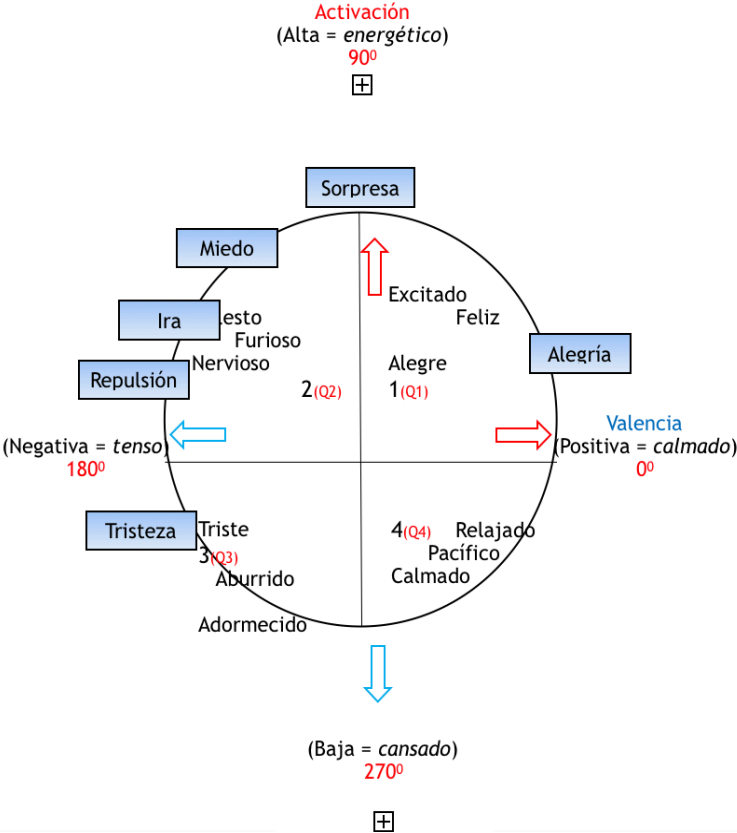
La Tabla 3 presenta el modelo de 5 *clusters* recurrentes ideado por Hu et al. (2008) el cual es de base empírica, y aunque no posee un apoyo teórico psicológico fue utilizado con rigor por diversas investigaciones como la propuesta realizada por Laurier, Grivolla y Herrera (2008) y la tesis doctoral de Laurier (2011). Una de las exploraciones significativas a este respecto fue la elaboración de Shubert (1999) quién cuantificó a través de series temporales las emociones musicales de diversos oyentes. En este documento de lectura obligada en el campo del REM Shubert (1999) articuló un espacio bidimensional emocional (*Two-Dimensional Emotion Space*) o 2DES el cual a su vez fue administrado por el programa *EmotionSpace Lab*, una pieza de software ideada para registrar y computar datos músico-emocionales según las ideas de Hevner, Russell y demás teóricos. Esta investigación pionera buscó articular pruebas de carácter experimental para categorizar las emociones musicales en diversas clases, aplicar procesos de reconocimiento de patrones, estadísticas de diversa índole y muestreos aplicados a un modelo de clasificación que combina en el plano (cartesiano) 2DES los diversos descriptores emocionales de las piezas analizadas.



4. Modelos dimensionales

En los modelos dimensionales se destacan las perspectivas de Russell (1980) y su modelo circunflejo del afecto que luego fueron modificadas de manera experimental por Thayer (1989) y el modelo de placer-activación-dominancia (PAD) (The Pleasure-Arousal-Dominance, PAD) de Mehrabian (1996). La importancia de ambas perspectiva radica en la capacidad de computación posible en la implementación de estas ideas. El modelo circunflejo VA combina los niveles de valencia (valence = estados afectivos positivos/negativos) y activación (arousal = nivel energético) (Russell, 1980), a través de valores que representan diferentes estados emocionales los cuales figuran en un espacio bidimensional. En lo correspondiente a los modelos REM estas configuraciones son elegidas como base teórica y conceptual para representar la emocionalidad musical desplegada y situada, como un punto o distribución de probabilidad continua en el espacio de la emoción humana.

Figura 2. El modelo circunflejo de Russell modificado, en el cual aparecen las seis emociones básicas, y los cuadrantes de Thayer



Fuente: adaptado de Yang, Lin, Su y Chen (2008, p. 448) y Thayer (1989, p. 135).

Los modelos dimensionales se diferencian de los demás modelos propuestos por su eficacia cuando se necesita realizar contrastaciones sobre anotaciones realizadas por expertos o participantes a nivel experimental. Por ejemplo en los procesos de etiquetado y descripción emocional para la construcción de datos verdaderos (*ground-truth data*), a través de diferentes modalidades que investigan los rasgos musicales en su relación con la producción de emociones y estados anímicos. Entre estos métodos usuales se encuentran

- El método de anotación de adjetivos (*Adjectives Annotation Method, AA*),
- El método de anotación de características musicales (*Music Features Annotation Method, MFA*)
- El método de anotación de modelos dimensionales (*Dimensional Models Annotation Method, DMA*).

En estas modalidades se emplean con regularidad segmentos de piezas musicales para su evaluación y etiquetación, las cuales son evaluadas a través de estimaciones numéricas para ser posteriormente clasificadas y optimizadas con métodos de regresión y regresión continua con el ánimo de encontrar y establecer predicciones emocionales. Preguntas como ¿qué sentimientos percibe?, o ¿cómo califica el tempo de un segmento?, guían el proceso completo de extracción de características musicales (*Music Feature Extraction*) y la fase de modelado del REM.

Entre los métodos de clasificación se utilizan (con sus siglas en inglés) los siguientes procesos; análisis cuadrático discriminante (QDA), vecino K más cercano (KNN), árbol, bosque aleatorio (RF), empaquetado (Bag), refuerzo (Boost), máquina de vectores de soporte con núcleo radial (SVM-R), máquina de vectores de soporte con núcleo polinomial (SVM-P), máquina de vectores de soporte con núcleo lineal (SVM-L), L1-LR, L2-LR, L1-SVM, L2-SVM y análisis discriminante regularizado de centroides reducidos (SCRDA) (Zhang, Huang, Yang, Xu y Sun, 2015) ¿Cómo se pueden asociar los rasgos musicales con los fenómenos emocionales? Para responder esta pregunta se hace necesario reconocer los presupuestos teóricos que asocian los diferentes rasgos musicales y la predicción emocional.



Tabla 4. Tipos de acordes y sus efectos emocionales asociados

Tipo de Acorde	Ejemplo	Emociones asociadas
Mayor	C	Felicidad, alegría, confianza, satisfacción, brillo.
Menor	Cm	Tristeza, oscuridad, mal humor, aprensión, melancolía, depresión, misterio.
Séptima	C7	Obscuro, nerviosismo moderado, alma.
Séptima mayor	Cmaj7	Romance, suavidad, jazzidad, serenidad, regocijo, tranquilidad.
Séptima menor	Cm7	Suavidad, melancolía, jazzidad.
Novena	C9	Optimismo.
Disminuida	Cº	Miedo, choque, suspenso.
Cuarta suspendida	Csus4	Tensión encantadora.
Novena menor	C9b	Escalofriante, ominoso, miedo, oscuridad.
Novena adherida	Cadd9	Falto de emoción, austeridad.

La Tabla 4 presenta a partir de diversas evidencias de carácter empírico, la funcionalidad de los tipos de acordes a propósito de las emociones asociadas que son percibidas por los oyentes de manera general. Como se explicó, las investigaciones que soportan esta serie de correlaciones entre los acordes, las estructuras musicales y los estados emocionales se cuentan por centenares tal como lo constatan Gabriellsson y Lindström (2001) y recientemente Juslin (2019). ¿Cuáles son los rasgos musicales relevantes en los enfoque del REM?



Tabla 5. Rasgos musicales relevantes en el REM

Rasgos	Ejemplos
Timing	Tempo, tiempo, variación, duración, constaste
Dinámicas	Nivel global, crescendo/decrecendo, acentos
Articulación	Global (staccato, legato), variabilidad
Timbre	Sonoridad espectral, sonoridad armónica,
Tono	Grave o agudo (nivel de alturas)
Intervalo	Rango (corto o largo), dirección (ascendente o descendente)
Tonalidad	Cromático-atonal, orientado al tono
Ritmo	Regular, irregular, suave, firme, fluido, ruidoso
Modo	Mayor o menor
Volumen	Alto o bajo
Forma musical	Complejidad, repetición, interrupción, condensación
Vibrato	Extendido, rango, rápido

Fuente: elaborado a partir de Panda, Malheiro y Paiva (2018b).

La Tabla 5 presenta los principales rasgos musicales relevantes en los enfoques investigativos del REM. Como se puede observar existe una serie de rasgos significativos que complementan los modelos de Hevner y sus ocho categorías y el modelo de Russell y Thayer adaptado a una perspectiva bidimensional de activación (A) y valencia (V) denominada como 2D-AV. En este contexto, la valencia es un mecanismo regulatorio que corresponde a sentimientos positivos por ejemplo, la alegría, la felicidad, la relajación, frente a otros negativos por ejemplo, emociones negativas como el miedo, la ira y la tristeza. Por otra parte, se puede decir que el *arousal* fisiológico y psicológico es el nivel de la activación o pasividad de la emoción percibida, lo cual es susceptible de inferirse a partir de los rasgos expuestos en las Tablas 5 y 6.

Tabla 6. Rasgos musicales conectados con 5 categorías emocionales específicas

Rasgos musicales	Felicidad	Tristeza	Rabia	Miedo	Ternura
<i>Tempo</i>	Rápido, poca variabilidad	Lento	Rápido pequeña variabilidad	Rápido poca variabilidad	Lento
<i>Modo</i>	Mayor	Menor	Menor	Menor	Mayor
<i>Armonía</i>	Simple y consonante	Disonante	Atonalidad, disonante	Disonante	Consonante
<i>Volumen</i>	Medio-alto, poca variabilidad	Bajo, variabilidad moderada	Alto, poca variabilidad	Bajo, largos niveles de variabilidad, cambios rápidos	Medio-bajo, poca variabilidad
<i>Tono</i>	Alto, mucha variabilidad, amplio rango ascendente	Bajo, rango estrecho, descendente	Alto, poca variabilidad, ascendente	Alto, ascendente, amplio rango, largos contrastes	Bajo, rango estrecho justo
<i>Entonación</i>	Creciente	Sostenido, caída	Acento en la tonalidad con notas inestables	-	-
<i>Formantes de los cantantes</i>	Aumentado	Disminuido	Aumentado	-	Disminuido
<i>Intervalos</i>	Cuarta perfecta y quinta	Corta (segunda menor)	Séptima mayor y cuarta aumentada	-	-
<i>Articulación</i>	<i>Staccato</i> , variabilidad larga	<i>Legatto</i> , poca variabilidad	<i>Staccato</i> , variabilidad moderada	<i>Staccato</i> , variabilidad larga	<i>Legatto</i> , poca variabilidad
<i>Ritmo</i>	Suave y fluido	<i>Ritardando</i>	Complejo, cambios súbitos, acelerando	Sacudido	-
<i>Timbre</i>	Brillante	Apagado	Sostenido	Suave	Suave
<i>Ataques de los tonos</i>	Rápido	Lento	Rápido	Suave	Lento
<i>Variabilidad del timing</i>	Corto	Largo (<i>rubato</i>)	corto	Muy largo	Moderado
<i>Vibrato</i>	Ritmo medio-rápido, medio extendido	Lento, poco extendido	Ritmo medio-rápido, bastante extendido	Ritmo rápido, poco extendido	Medio rápido, poco extendido
<i>Contrastes entre las notas largas y cortas</i>	Sostenido	Suave	Sostenido	-	Suave
<i>M i c r o -estructuras</i>	Regularidades	Irregularidades	Irregularidades	Irregularidades	Regularidades
<i>Otros</i>	-	Pausas	Ruido espectral	Pausas	Acento en la tonalidad con notas estables

Fuente: elaborado a partir de Juslin y Laukka, (2004, p. 221) y Laurier (2011, p. 16).



La Tabla 6 presenta en detalle los rasgos musicales respecto a 5 emociones: felicidad, tristeza, rabia, miedo y ternura. Esta es una detallada elaboración que establece la correspondencia directa entre la música y su expresividad. Este resumen intenta responder a la pregunta ¿cómo expresa la música diferentes emociones? (Juslin y Laukka, 2004, p. 220). Gabrielsson y Lindström (2001) desde una perspectiva psicológica consideran que la dimensión de activación está relacionada con aspectos musicales en sus dimensiones opuestas, como el *tempo* (rápido o lento), el tono (alto o bajo), el nivel de sonoridad (alto o bajo) y el timbre (brillante o suave), y por el contrario la dimensión de valencia está relacionada con el modo (mayor o menor) y la armonía (consonante o disonante); en este sentido las funciones de la música son correlacionadas con las dimensiones psicológicas de la percepción musical humana (Yang et al., 2018).

5. Modelos multimodales

El empleo de estos modelos aborda la investigación de problemas relacionados con las emociones musicales a partir de las estructuras musicales, el significado de las letras y la clasificación RIM. Esta metodología se basa en un enfoque de clasificación y regresión, y procura demostrar las dinámicas variables de las emociones musicales y sonoras, al utilizar tipos de datos de verdad a partir de detallados procesos numéricos y de etiquetado que emplean perspectivas tanto categóricas como dimensionales. Además, estos modelos tienen como objetivo describir la detección de la variedad de emociones musicales y formular comparaciones entre diferentes acervos y segmentos musicales a través de la formación de curvas de variación de las emociones (Yang et al., 2008), y el pre-procesamiento de etiquetado individual emocional segundo a segundo para indicar las estimaciones de las distribuciones y predicciones afectivas (Schmidt y Kim, 2010).

Yang y Chen (2012, p. 18) recuerdan que un enfoque REM dimensional o multimodal de tipo detección de la variación músico-emocional (DVME) depende del problema en el cual se diseñe la investigación. Esto significa que a nivel computacional la investigación puede abordar la clasificación de segmentos de piezas musicales por lo general de 30 segundos de duración en el los modelos REM aplicado en un plano 2D-AV, o en su lugar, determinar los valores (puntos) de cada segundo de una pieza musical, también posicionados en un plano 2D-AV. Macdorman, Ough, y Ho (2007) presentaron una propuesta considerada DVME en la cual buscaron establecer la relación existente entre las propiedades físicas de las señales acústicas y su impacto emocional, a través de las clasificaciones de ochenta y cinco participantes sobre 100 extractos de canciones de pocos segundos. Demostraron cómo el ritmo, el tono y la intensidad se correlacionan en los procesos de apreciación del contenido emocional musical. Su idea se basa en la necesidad de realizar un estudio granular, es decir, utilizar muestras musicales micro que permitan reconocer las dinámicas músico-emocionales con un amplio nivel de detalle. Lo que significa que a menor tamaño del segmento de la canción se infiere que será más homogénea la emoción, y por lo tanto se obtenga una mayor consistencia entre las evaluaciones de los sujetos de la prueba y los datos verdaderos. Sin embargo, surgen algunos problemas a nivel metodológico y aplicativo.



¿Cuál es el tamaño idóneo de una muestra (segmento) musical para ser investigado? Esta pregunta fue abordada por Xiao, Dellandrea, Dou y Chen (2008), quienes a través de diferentes pruebas con música clásica concluyeron que el tamaño de los segmentos que presentan un mayor estado emocional estable es de 8 a 16 segundos. Es decir, que este tiempo valida con eficiencia el rendimiento de clasificación del sistema y su optimización. Sin embargo, en músicas populares los tiempos de las muestras habitualmente se efectúan con 25 a 30 segundos, tiempo en el cual tiene duración, por ejemplo, el coro de una canción de pop. Juslin (2019, p. 49-52) plantea la dificultad de saber qué tipo de modelo es mejor, si los encuadres categóricos o dimensionales. Debido a que ambas perspectivas son dominantes en la investigación psicológica, y además, los modelos dimensiones se pueden ampliar de manera considerable por ejemplo, cuando se presentan tres dimensiones; activación de energía-activación tensión-valencia, estos modelos también presentan límites categóricos cuando cambian de un estado a otro, por ejemplo *negativo o positivo*, debido a que se establecen ambas perspectivas de manera complementaria, lo cual se denomina en la literatura especializada del REM como modelos multimodales, a los cuales Juslin (2019) no se refiere.

6. Aplicaciones prácticas basadas en el REM



6.1. EXPERIMENTO 1

Para presentar una aplicación práctica de un abordaje desde el REM se analizó a nivel musical un conjunto de canciones, que fue abordado de manera inicial a nivel letrístico en otro espacio. El *corpus* del presente estudio se basa en el análisis de 165 canciones completas de rap puertorriqueño, las cuales proceden de 12 álbumes musicales de diversos artistas, y seleccionadas de manera no aleatoria (ver Tabla 7). Las canciones fueron procesadas en formato mp3 a una frecuencia de muestreo de 44100 hz, 16 bit, en formato estéreo con diferentes calidades de reproducción; 128, 192, y 320 kilobits por segundo (Kbps), a través de un equipo de cómputo con procesador Intel CORE i5 de 32 bits.

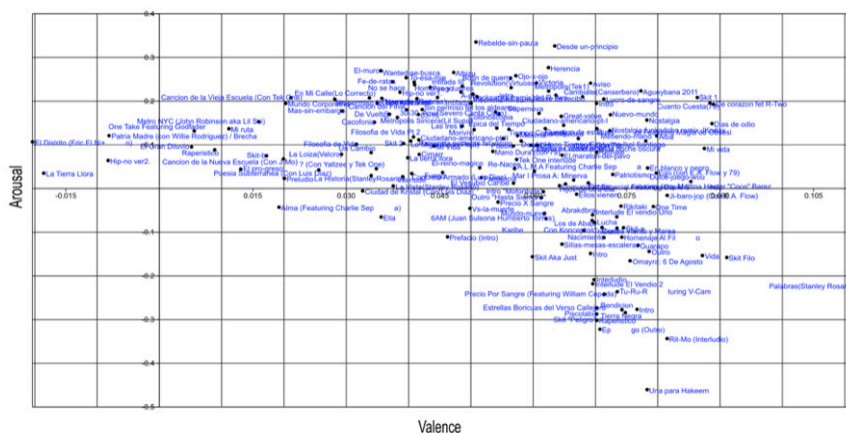
Tabla 7. Álbumes de rap puertorriqueño que conforman la muestra escogida

Artista	Álbum	Año
Intifada	<i>Intifada I</i>	2004
Intifada	<i>Mundo Nuevo</i>	2006
Intifada	<i>Intifada III</i>	2011
Luis Díaz	<i>Rebelde sin pausa</i>	2018
Siete Nueve	<i>El pro-greso</i>	2003
Siete Nueve	<i>Trabuco</i>	2007
Siete Nueve	<i>Antología the mixtape</i>	2008
Siete Nueve	<i>Antología</i>	2012
Siete Nueve	<i>El patio sessions, vol. 1</i>	2018
R-Two	<i>Vida, lucha y victoria</i>	2018
Nuff Ced	<i>La metrópolis</i>	2013

En este abordaje se utilizó la aplicación MooDetector (Cardoso, Panda y Paiva, 2011) la cual emplea la extracción de los rasgos musicales acústicos (ver Tabla 1) con el algoritmo del programa Marsyas, y los ubica en los cuadrantes del espacio 2D-AV según el modelo de Thayer (1989), utilizando máquinas de vectores de soporte (SVM) y regresión de vectores de apoyo (SVR), para la clasificación ulterior de los datos obtenidos. En el caso de cada canción la aplicación permite obtener los resultados individuales es una colección de valores 2D-AV, uno para cada segmento en ventanas de 1,5 segundos en este caso. Además de obtener a partir de estos valores un proceso definido como rastreo del estado anímico (Mood Tracking) en cada pieza o segmento analizado. Este proceso se orienta de manera técnica, cuando los resultados se miden calculando la proporción de coincidencia entre las anotaciones y las predicciones de clasificación, lo cual significa el porcentaje de tiempo con concordancia entre la anotación y la predicción (Cardoso et al., 2011; Panda y Paiva, 2011). Es importante recalcar que el presente estudio expone una metodología que ofrece resultados automáticos a partir de los datos adecuados (muestras musicales digitales) en el análisis realizado, y no es el objetivo realizar una experimentación basada en un conjunto específico de sujetos partícipes como es usual en este campo para correlacionar ambos niveles de análisis.



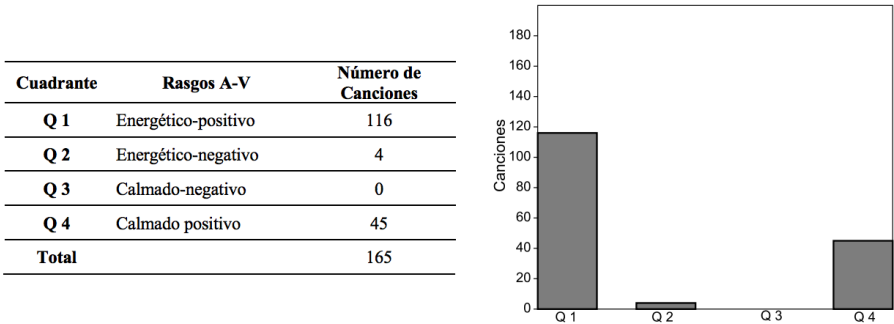
A.



B.

Roberto Carlos Luján Villar y Juan David Luján Villar. "Reconocimiento de emociones musicales a través de datos y tecnologías digitales". *Comunicación y Hombre*. 2020, n° 16, pp 59-82.

Tabla 8. Resultados de la muestra abordada según el número de canciones en la librería ubicadas en cada cuadrante (Q) y rasgos en el plano 2D-AV



La Tabla 8 expone la similitud en las canciones de manera previsible, lo cual indica la existencia de pocas diferencias o puntos extremos a nivel compositivo en la muestra total. No obstante, a nivel lírico en cada canción los contenidos pueden ser diametralmente opuestos, aunque la mayoría de las muestras se sitúan en el Q1 es decir, en una categoría *energético-positivo*. Consideramos que este es un indicador fiable según las descripciones realizadas en otro espacio sobre las canciones de los artistas Siete Nueve e Intifada, quienes forman parte de la presente muestra.

El presente análisis se fundamenta en el análisis basado en contenido (*Content-Based Analysis*) (Kim et al., 2010) o REM basado en contenido (*Content-Based MER*) (Grekow, 2018), como se denomina el análisis de los rasgos acústicos musicales, que intenta representar los aspectos emocionales de un *corpus* musical abordado. Aunque diversos tipos de datos se pueden emplear en este tipo de análisis por ejemplo, las letras y el audio o las imágenes de carátulas, sólo aplicamos el contenido emocional de la música desde sus rasgos acústicos, un reclamo de Adams (2015) quién manifiesta que las investigaciones sobre el rap y el hip-hop se realizaban de manera recurrente sobre sus aspectos textuales, en detrimento de sus rasgos musicales. Es importante señalar que en la muestra global los niveles de activación (*arousal*) son mayormente positivos, de lo cual podemos inferir que en el nivel de información sobre el contenido emocional en la música rap de Puerto Rico se mantiene un estilo de hip-hop convencional y *underground*, en el que se presentan líricas ingeniosas y contestatarias a nivel de sus mensajes, que se alimenta de la crítica social, las problemáticas del colonialismo y las perspectivas ancestrales en esta nación caribeña.



6.2. EXPERIMENTO 2

Para demostrar un enfoque complementario a nivel metodológico, se analizó la trayectoria emocional de una canción que coincide plenamente con los datos verdaderos o *ground-truth data*, obtenidos según el procedimiento técnico del experimento 1. La canción se titula *Una para Hakeem* del rapero R-Two, y fue abordada a través de 8 muestras de 30 segundos cada

una, que sumadas componen el total de la canción. Los segmentos fueron procesados en formato mp3 a una frecuencia de muestreo de 44100 hz, 16 bit, en formato estéreo con una calidad de 192 Kbps.

La idea central de esta aplicación es trazar una línea que establezca los posibles cambios emociones y el estado final emocional de la canción. Aunque este proceso indica que podría ser un modelo multimodal, continúa como un modelo dimensional, ya que investigar segmentos de 30 segundos de duración no consiste en rigor con una implementación DVME como se explicó arriba. Sin embargo, debido a que la música rap por lo general sostiene un mismo *tempo*, patrón armónico, timbre y modo, sus cambios y variaciones estructurales no se efectúan a través de los segundos transcurridos de cada pieza, si no que se presentan cambios a nivel de las partes de cada canción, lo cual es una forma decidida de abordar la música popular y sus características emocionales (Williams, 2015).

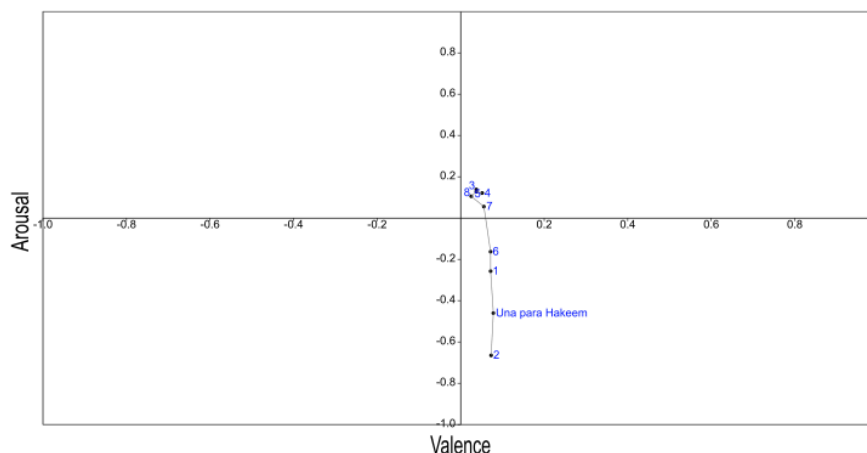
Tabla 9. Segmentos de la canción Una para Hakeem de R-Two

Muestra	Artista	Arousal	Valence	Etiqueta
1	R-Two	-0.257334	0.0717131	Relajado
2	R-Two	-0.664968	0.0729323	Relajado
3	R-Two	0.138304	0.0375354	Alegre
4	R-Two	0.121218	0.0515131	Alegre
5	R-Two	0.127759	0.037486	Alegre
6	R-Two	-0.162812	0.0716285	Relajado
7	R-Two	0.0554472	0.0555759	Alegre
8	R-Two	0.105538	0.0252576	Alegre
Una para Hakeem	R-Two	-0.460348	0.0782986	Relajado

La Tabla 9 presenta las diferentes anotaciones de la canción *Una para Hakeem* y las diferencias que se presentan a lo largo de su desarrollo. Como se puede observar, las diferentes partes de la canción se sitúan en dos cuadrantes: 1 (Q1) y 4 (Q4), aunque mayoritariamente en el cuadrante 1 (Q1). Sin embargo, cuando se analiza la canción en su totalidad su configuración aparece en el cuadrante 4 (Q4) con una etiqueta de *relajado*, es decir, la canción se sitúa en una categoría *calmado-positivo* (ver Figura 4), lo cual es coherente con la intencionalidad del tema; una dedicatoria del artista a su hijo.



Figura 4. Plot de *Valence* y *Arousal* de las muestras de la canción *Una para Hakeem* de R-Two y la canción completa, con el recorrido total

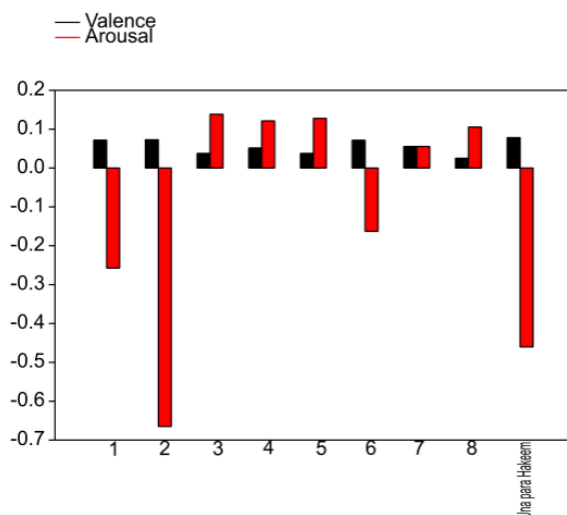


La Figura 4 presenta los tránsitos anímicos de la canción analizada. En el plano 2D-VA se observa el árbol de expansión mínimo, esto es, el conjunto de líneas con una longitud total mínima que conecta todos los puntos. En esta visualización se utilizan longitudes euclidianas en 2D para trazar la distribución espacial de todos los puntos que representan cada uno de los segmentos, lo cual traza la distancia entre los 8 puntos que conforman la canción, incluyendo la versión completa. A partir de esto, se puede decir que la suma de los diferentes segmentos establece una comunicación polivalente a nivel emocional, aunque la variación estructural sea menor, el beat sea cíclico e iterativo y su reconocimiento tonal se mantenga, esta es una muestra de lo complejo que puede ser un análisis emocional basado en los rasgos acústicos de este tipo de música.

La música hip-hop posee –de manera general– una diferencia estructural con el resto de la música Occidental, esto es, su carácter cíclico (Adams, 2015). Esta es una característica que quizá favorece el enfoque utilizado en este espacio, debido a que la naturaleza del rap se basa en el desarrollo de un ritmo circular, y aunque los estados afectivos de las líricas varían de manera constante debido a su expresividad, la circularidad rítmica y armónica en su estructura, las figuras musicales específicas no necesariamente se tienen que correlacionar para contener un tipo de letra específica. Por esta razón, el hip-hop no puede verse a nivel emocional ni musical como una obra de arte unificada (Adams, 2015), lo que dificulta su análisis e interpretación unidireccional. A nivel lírico también se dificulta establecer análisis lineales si recordamos que el rap en estado puro evidencia por lo general un conjunto complejo de sincopaciones y notas osciladas (subdivisiones rítmicas) conjugadas a través de un forma de canción y poesía de manera simultánea.



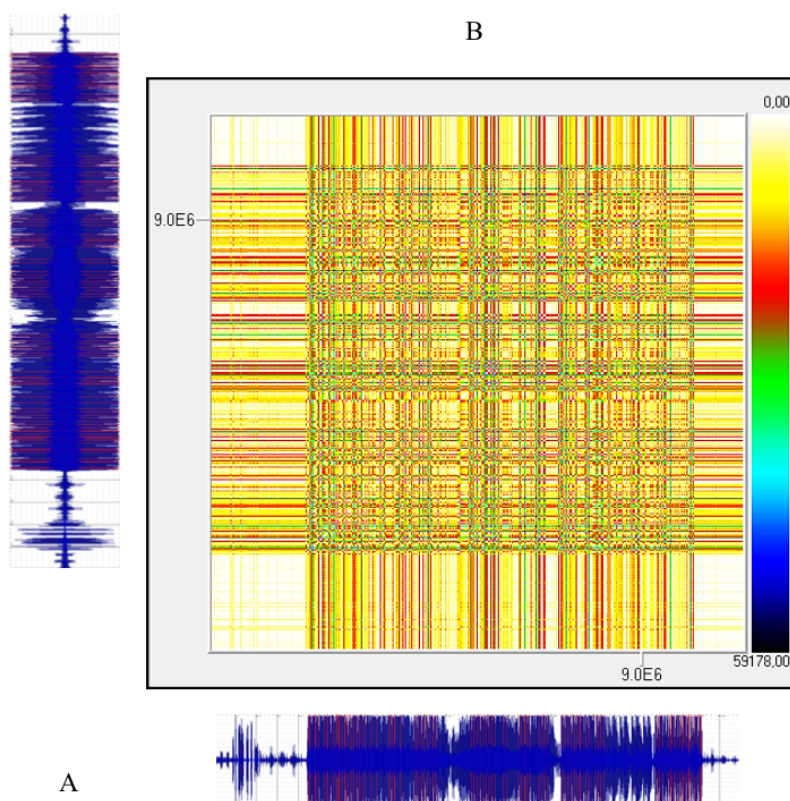
Figura 5. Gráfico de barras de *valence* y *arousal* de las muestras de la canción *Una para Hakeem* de R-Two y la canción completa



La Figura 5 expone los niveles de *valence* o valencia, que indican el grado de emoción positiva o negativa (ver Figura 2) y *arousal* o activación-excitación de las muestras analizadas que componen la canción estudiada. Los resultados indican que los niveles de activación son mayores que los de valencia, los cuales son en un sentido negativos (A)-positivos (V), una energía relajada-positiva, es decir *calmado-positivo*. La activación de la música hip-hop a nivel internacional exhibe según diversos estudios de manera general (Hu y Downie, 2007) una clasificación como música altamente energética o up-beat, con un estado anímico ingenioso (*Witty*) e inteligente callejero (*Street Smart*), y se agrupa bajo categorías como chistoso, peculiar, caprichoso, ingenioso, irónico o agresivo, ardiente, tenso/ansioso, intenso, volátil o visceral.

Los niveles de activación emocional presentan diversas correlaciones con dimensiones identificadas en la música como la potencia, la energía y la tensión. A nivel biomusicológico los niveles de activación se ven afectados por la intensidad de los sonidos y factores acústicos complementarios como el flujo espectral y la entropía espectral, lo cual tiene como particularidad una naturaleza intercorrelacionada y redundante es decir, no excluyente entre estos factores. Esto significa que la música emplea en este nivel, señales parcialmente superpuestas, aspecto que fue demostrado a nivel experimental por Gingras, Marin y Fitch (2014). En nuestra exploración estas indagaciones tienen relevancia debido a que la música rap examinada presenta una configuración basada en la música tonal occidental. Los datos actuales indican que la activación emocional surge debido a las características físicas del estímulo, lo cual brinda relevancia y objetividad metodológica a los estudios RIM y MER.

Figura 6. Tracking de la canción *Una para Hakeem* de R-Two



La Figura 6A exhibe la gráfica o *waveform* de la canción *Una para Hakeem*, las líneas rojas presentan los cambios emocionales de la canción, los cuales coinciden con los sonidos de los redobles o *snare*s del beat. La Figura 6B presenta el *plot* de recurrencia con los parámetros de distancia euclidiana, radio = 1.5, y espacio de fase de dimensión = 1. En los cuadros morados se observan las partes de la canción donde hay silencios o dinámicas de pausas o diálogos, los cuadros negros presentan las partes líricas con el beat, los cuadros morados también delimitan los coros y puentes, y la línea roja cruzada presenta el tiempo de la canción. Los valores del mapa de recurrencia, como se observa en la imagen, van desde los tonos amarillos más bajos hasta los tonos azules más altos es decir, que las regiones más oscuras representan mayores niveles de similitud o alta recurrencia.

La Figura 6B además, presenta los rasgos acústicos de la canción bajo escrutinio, convertidos en series temporales, susceptible de visualizar cambios a partir de la auto-similitud, proximidad y continuidad que las señales musicales manifiestan (Schedl et al., 2014, p. 167-168). El método se conoce como análisis de auto-similitud (*Self-Similarity Analysis*), donde el resultado es una matriz de auto-similitud (*Self-Similarity Matrix*) en la cual las regiones de alta auto-similitud aparecen como cuadrados brillantes a lo largo de la diagonal principal marcada por una línea roja (Cooper y Foote, 2002) (ver Figura 6B).

Bharadwaj, Hegde, Dutt y Rajan (2018) investigaron a través del procesamiento de señales con la técnica figurativa de *plot* de recurrencia el procesamiento musical en el cerebro en 20 sujetos. En sus resultados presentaron datos y las visualizaciones de las regiones frontales implicadas en la escucha de las fases de seis ragas de música clásica del norte de India/Hindustaní clasificadas como feliz y triste. Los resultados de la matriz de auto-similitud ofrecen una medida que establece qué tan similar es cada segmento al conjunto. En efecto, se pueden apreciar patrones difíciles de observar de otra manera, a partir de los datos de las piezas analizadas o de datos provenientes de las neurociencias.

7. Conclusión

En la actualidad el REM presenta el aumento de servicios y bases de datos de información robustas sobre emociones musicales disponibles de manera abierta para el público en general y la comunidad científica. Aunque no podemos decir que a nivel de modelos investigativos se encuentra en etapa embrionaria aún continúa en su etapa de maduración metodológica, y de modo similar ocurre con los protocolos de experimentación.

Los modelos del REM realizan un esfuerzo para establecer la predicción de las emociones evocadas por la música. En este trabajo intentamos introducir este campo como una opción viable la cual es parte del inmenso campo del RIM. Las perspectivas enunciadas posibilitan complementar metodologías clásicas y convencionales de la etnomusicología y la musicología, a partir de exploraciones basadas en largas series de datos, la clasificación emocional de productos musicales es posible en medio de la complejidad social que involucran los consumos socioculturales y las prácticas estéticas.

Las perspectivas actuales en el campo del REM parten de marcos interdisciplinarios y procuran a nivel experimental demostrar la importancia de las emociones en la comunicación social, quizá un determinante del progreso y el avance humano. Estas ideas despliegan un gran impacto de aplicación si reconocemos la era en la cual nos encontramos, dominada por los sistemas inteligentes autónomos, las redes sociales de diversa naturaleza, la masificación de la información y su inmediatez y los consumos culturales a escala global. Para afrontar esta situación, los aspectos metodológicos en el reconocimiento de las emociones subjetivas y sociales experimentadas, este tipo de exploraciones se basan en las funciones de la música en la vida cotidiana de maneras innovadoras, una apuesta que al parecer tiene mucho que ofrecer en el contexto global actual.

8. Bibliografía

ADAMS, Kyle (2015). The musical analysis of hip-hop. En Williams, Justin A. (ed). *The Cambridge companion to hip-hop*, 118-134. Cambridge U.K.: Cambridge University Press. doi: 10.1017/CCO9781139775298.012

- ALJANAKI, Anna (2016). *Emotion in Music: representation and computational modeling* (tesis doctoral). Utrecht University, Utrecht. Recuperado de <https://dspace.library.uu.nl/handle/1874/339995>
- BHARADWAJ, Sushrutha, HEGDE, Shantala, DUTT, Narayana y RAJAN, Anand Prem (2018). Application of nonlinear signal processing technique to analyze the brain correlates of happy and sad music conditions during listening to raga elaboration phases of Indian classical music. En Richard Parncutt y Sabrina Sattmann (eds.), *Proceedings of ICMPC15/ESCOM10* (pp. 83-84). Graz, Austria: Centre for Systematic Musicology, University of Graz.
- CALVO, Rafael A., D'MELLO, Sidney, GRATCH, Jonathan y KAPPAS, Arvid (eds.). (2014). *The Oxford handbook of affective computing*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199942237.001.0001
- CARDOSO, Luis, PANDA, Renato y PAIVA, Rui Pedro (2011). MOODetector: A prototype software tool for mood-based playlist generation. *Simpósio de Informática – INForum 2011*, 124.
- CHASE, Waine (2006). *How music really works!*, Vancouver, Canada: Roedy Black.
- COOPER, Matthew y FOOTE, Jonathan (2002). Automatic music summarization via similarity analysis. En Michael Fingerhut (ed.). *Proc Int Society Music Information Retrieval Conf (ISMIR)*, 81–85. Paris: Ircam - Centre Pompidou. Recuperado de <https://pdfs.semanticscholar.org/709a/d98062c8e02e122a34bf9e4b9ed722afcafc.pdf>
- DENG, James J., LEUNG, Clement H. C., MILANI, Alfredo y CHEN, Li (2015). Emotional states associated with music: Classification, prediction of changes, and consideration in recommendation. *ACM Trans. Interact. Intell. Syst. (TiiS)*, 5(1), 1-36. doi: 10.1145/2723575
- GABRIELSSON, Alf y LINDSTRÖM, Erik (2001). The influence of musical structure on emotional expression. En Patrik Juslin y John A. Sloboda (eds.). *Music and emotion: Theory and research* (pp. 223-249). Oxford: Oxford University Press.
- GREKOW, Jacek (2018). *From content-based music emotion recognition to emotion maps of musical pieces*. Cham, Switzerland: Springer. doi: 10.1007/978-3-319-70609-2_2
- GRINGAS, Bruno, MARIN, Manuela M. y FITCH, W. Tecumseh (2014). Beyond intensity: Spectral features effectively predict music-induced subjective arousal. *The quarterly journal of experimental psychology*, 67(7), 1428-1446. doi: 10.1080/17470218.2013.863954
- HEVNER, Kate (1936). Experimental studies of the elements of expression in music. *The American journal of psychology*, 48(2), 246-268. doi: 10.2307/1415746
- HU, Xiao y DOWNIE, J. Stephen (2007). Exploring mood metadata: Relationships with genre, artist and usage metadata. En *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007* (pp. 67-72). Vienna: ISMIR. Recuperado de <http://ismir2007.ismir.net/>



proceedings/ISMIR2007_p067_hu.pdf

HU, Xiao, DOWNIE, J. Stephen, LAURIER, Cyril, BAY, Mert y EHMANN, Andreas F. (2008). The 2007 MIREX audio mood classification task: Lessons learned. En J. P. Bello, E. Chew, y D. Turnbull (eds.). *Proceedings of the International Symposium on Music Information Retrieval (ISMIR) 2008* (pp. 462-467). Philadelphia: Drexel University. Recuperado de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.153.9934&rep=rep1&type=pdf>

JUSLIN, Patrik N. (2019). *Musical emotions explained. Unlocking the secrets of musical affect*. Oxford: Oxford University Press.

JUSLIN, Patrik N. y LAUKKA, Petri (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, 33(3), 217–238. doi: 10.1080/0929821042000317813

LAURIER, Cyril (2011). *Automatic classification of musical mood by content-based analysis* (tesis doctoral). Universitat Pompeu Fabra, Barcelona. Recuperado de <http://mtg.upf.edu/node/2385>

KIM, Youngmoo E., SCHMIDT, Erik M., MIGNECO, Raymond, MORTON, Brandon G., RICHARDSON, Patrick, SCOTT, Jeffrey, SPECK, Jacquelin, TURNBULL, Douglas (2010). Music emotion recognition: a state of the art review. En Downie, J. Stephen y R. C. Veltkamp (eds). *International Society for Music Information Retrieval Conference - ISMIR'2010* (pp. 255-266). Utrecht: ISMIR.

LAURIER, Cyril, GRIVOLLA, Jens y HERRERA, Perfecto (2008). Multimodal music mood classification using audio and lyrics. En M.A. Wani (ed.). *Proceedings of the International Conference on Machine Learning and Applications*, San Diego, CA (pp. 688–693). Piscataway, NJ: IEEE. doi: 10.1109/ICMLA.2008.96

MACDORMAN, Karl F., OUGH, Stuart y HO, Chin-Chang (2007). Automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison. *Journal of new music research*, 36(4), 281-299. doi: 10.1080/09298210801927846

MEHRABIAN, Albert (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current psychology*, 14(4), 261-292. doi: 10.1007/BF02686918

PANDA, Renato, MALHEIRO, Ricardo y PAIVA, Rui Pedro (2018a). Musical texture and expressivity features for music emotion recognition. En E. Gómez, X. Hu, E. Humphrey, y E. Benetos (eds.). *19th International Society for Music Information Retrieval Conference, Paris, France, 2018* (pp. 383-391). Paris: ISMIR. Recuperado de <https://archives.ismir.net/ismir2018/paper/000250.pdf>

PANDA, Renato, MALHEIRO, Ricardo y PAIVA, Rui Pedro (2018b). Novel audio features for music emotion recognition. *IEEE transactions on affective computing*. doi: 10.1109/TAFFC.2018.2820691



- PANDA, Renato y PAIVA, Rui Pedro (2011). Using support vector machines for automatic mood tracking in audio music. En Audio Engineering Society (eds.). *Proceedings of the 130th Audio Engineering Society Convention, 2011 - AES 130, London, UK* (pp. 579-586). New York: AES-Curran.
- PICARD, Rosalind W. (2000). *Affective computing*. Cambridge: MIT Press.
- RUSSEL, James A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161-1178. doi: 10.1037/h0077714
- SCHEDL, Markus, GÓMEZ, Emilia y URBANO, Julian (2014). Music information retrieval: Recent developments and applications. *Foundations and trends in information retrieval* 8(2-3), 127-261. Recuperado de <http://mtg.upf.es/system/files/publications/059-music-information-retrieval-recent-developments-applications.pdf>
- SCHUBERT, Emery (1999). *Measurement and time series analysis of emotion in music* (tesis de Ph.D.). School of Music Music Education, Univ. New South Wales, Sydney, NSW, Australia. Recuperado de <http://unsworks.unsw.edu.au/fapi/datastream/unsworks:500/SOURCE01?view=true>
- SCHUBERT, Emery, FERGUSON, Sam, FARRAR, Natasha, TAYLOR, David y MCPHERSON, Gary E. (2013). The six emotion-face clock as a tool for continuously rating discrete emotional responses to music. In: M. Aramaki, M. Barhet, R. Kronland-Martinet S. Ystad (eds). *From Sounds to Music and Emotions*. CMMR 2012 (pp. 1-18). Springer: Berlin, Heidelberg.
- SCHMIDT, Erik M. y KIM, Youngmoo E. (2010). Prediction of time-varying musical mood distributions from audio. En J. Stephen Downie, y R. C. Veltkamp (eds). *International Society for Music Information Retrieval Conference - ISMIR'2010* (pp. 465-470). Utrecht: ISMIR. doi: 10.1109/ICMLA.2010.101
- SPECK, Jacquelin A., SCHMIDT, Erik M., MORTON, Brandon G., y KIM, Youngmoo E. (2011). A comparative study of collaborative vs. traditional musical mood annotation. En A. Klapuri y C. Leider (eds.), *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, Florida, (pp. 549-554). Miami: ISMIR. Recuperado de <https://archives.ismir.net/ismir2011/paper/000098.pdf>
- THAYER, Robert E. (1989). *The biopsychology of mood and arousal*. Oxford University Press, Oxford.
- WILLIAMS, Justin A. (2015). Intertextuality, sampling, and copyright. En Justin A. Williams (ed). *The Cambridge companion to hip-hop* (pp. 206-220). Cambridge U.K.: Cambridge University Press. doi: 10.1017/CCO9781139775298.018
- XIAO, Zhongzhe, DELLANDREA, Emmanuel, DOU, Weibei y CHEN, Liming (2008). What is the best segment duration for music mood analysis? En *2008 International Workshop on Content-Based Multimedia Indexing* (pp. 17-24). New York: IEEE. doi: 10.1109/CBMI.2008.4564922



- YANG, Yi-Hsuan y CHEN, Homer H. (2012). Machine recognition of music emotion: a review. *ACM transactions on intelligent systems and technology*, 3(3), Article 40, 1-30. doi: 10.1145/2168752.2168754
- YANG, Yi-Hsuan y CHEN, Homer H. (2011). *Music emotion recognition*. Boca Raton, Florida: CRC Press. doi: 10.1201/b10731
- YANG, Xinyu, DONG, Yizhuo y LI, Juan (2018). Review of data features-based music emotion recognition methods. *Multimedia systems*, 24(4), 365-389. doi: 10.1007/s00530-017-0559-4
- YANG, Yi-Hsuan,, LIN, Yu-Ching, SU, Ya-Fan y CHEN, Homer H. (2008). A regression approach to music emotion recognition. *IEEE transactions on audio, speech and language processing*,. 16(2), 448-457. doi: 10.1109/TASL.2007.911513
- ZHANG, Jiang-Long, HUANG, Xiang-Lin, YANG, Li-Fang, XU, Ye y SUN, Shu-Tao (2015). Feature selection and feature learning in arousal dimension of music emotion by using shrinkage methods. *Multimedia systems*, 23(2), 251-264. doi: 10.1007/s00530-015-0489-y

