SEALI: intra-acciones humano-máquina en la improvisación libre

Sección: Artículos
Recibido: 4/03/2026
Aceptado: 10/06/2026
DOI: 10.46530/virtualis.v18i31.449

SEALI: human-machine intra-actions in free improvisation

Aarón Arturo Escobar Castañeda
Tecnológico de Monterrey, México
ORCID: 0009-0002-5031-7391
correo: mailto:aaron.escobar@tec.mx

Resumen. Este artículo presenta el desarrollo de SEALI (Sistema de escucha automática para la libre improvisación), un sistema interactivo que utiliza la escucha de máquina para analizar elementos sonoros e incidir en la práctica de la libre improvisación musical. El sistema emplea técnicas algorítmicas aplicadas a la segmentación, el análisis y el entrenamiento de un corpus sonoro a través de redes neuronales, para generar modelos que describen la improvisación libre en términos tímbricos y formales. Para lograrlo, se analiza cómo el aprendizaje supervisado y sin supervisión pueden describir numéricamente esta práctica, así como las redes neuronales recurrentes—particularmente las redes LSTM—, pueden ofrecer una alternativa para la predicción de la forma musical. Estos modelos constituyen la base de este sistema, lo que posibilita la interacción en tiempo real con otros músicos. El artículo detalla las funcionalidades y hallazgos principales en torno a SEALI, aborda los estudios de caso que han contribuido a su evolución continua y retoma algunas discusiones sobre la agencialidad algorítmica en las prácticas artísticas.

Palabras clave: improvisación libre, escucha artificial, sistemas interactivos, redes neuronales, música algorítmica.

Abstract. This article presents the development of SEALI (Automatic Listening System for Free Improvisation), an interactive system that utilizes machine listening to analyze sound elements and influence the practice of free musical improvisation. The system employs algorithmic techniques applied to segmentation, analysis, and training of a sound corpus through neural networks to generate models that describe free improvisation in terms of timbral and formal characteristics. To achieve this, the article explores how supervised and unsupervised learning can numerically describe this practice, and how recurrent neural networks—particularly LSTM networks—can offer an alternative for predicting musical form. These models form the foundation of the system, enabling real-time interaction with other musicians. The article details SEALI's functionalities and key findings, discusses case studies that have contributed to its ongoing evolution, and revisits discussions on algorithmic agency in artistic practices.

Keywords: free improvisation, artificial listening, interactive systems, neural networks, algorithmic music.

Introducción

El presente artículo parte de mi investigación doctoral sobre SEALI (Sistema de escucha automática para la libre improvisación) el cual es un sistema sonoro interactivo que parte de la escucha artificial, modelada desde mi propia escucha, para analizar elementos sonoros en un ámbito muy acotado de la libre improvisación y, también, para interactuar a varios niveles temporales con otros improvisadores mediante la clasificación y la predicción sonora en el contexto de dicha práctica. Las funcionalidades de este sistema las pienso desde una perspectiva algorítmica y que, en conjunto, articulan la arquitectura de SEALI. A grandes rasgos, éstas comienzan en la selección de un corpus sonoro, posterior a ello, se aplicaron técnicas de segmentación, estructuración, análisis y entrenamiento de redes neuronales para generar una serie de supuestos abstractos a nivel numérico (modelos) que describan esta práctica musical. Finalmente, se desarrolló un sistema interactivo en tiempo real para establecer una serie de respuestas coherentes en la libre improvisación.

Cada una de estas funcionalidades, implica varios procesos entrelazados por bucles de retroalimentación donde mi escucha y el aprendizaje de máquina se unen con el objetivo de desarrollar la metodología y las técnicas algorítmicas que posibilitaron la realización de SEALI y su integración en contextos artísticos musicales. En este artículo abordaré a detalle las funcionalidades más generales de ese sistema, así como los estudios de caso que han sido útiles para continuar su evolución, bajo el entendido de que su constante introducción en la práctica artística sigue modificando sus funcionalidades para adecuarse cada vez más a las necesidades contextuales de una práctica en permanente cambio como la libre improvisación.

Definición y caracterización de la libre improvisación y su convergencia con SEALI

La improvisación libre es una forma de música que no sigue reglas ni estructuras preestablecidas, sino que se basa en la creatividad y la interacción de los músicos al momento de hacerla (Matthews, 2012). Aunque difícil de datar exactamente en su origen debido a sus raíces en diversas corrientes musicales y académicas, se consolidó como práctica en los años sesenta en Europa y Estados Unidos. Galiana (2018) define la improvisación libre como "un proceso creativo y no un producto acabado, manufacturado, listo para ser escuchado y consumido tantas veces como se desee". Sin embargo, algunos improvisadores han llegado a considerar que la libre improvisación fue la primera forma musical creada por los humanos (Bailey, 1980).

Inspirada por el *free jazz* y la música académica contemporánea, la improvisación libre surge como una práctica comprometida con la expresión artística y política que cuestionaba las injusticias sociales y económicas de la época. La práctica de la improvisación libre desafió las estructuras de poder en la industria musical, promoviendo la colaboración creativa y rompiendo con las convenciones musicales establecidas, dando lugar a nuevas técnicas y aproximaciones musicales. Los músicos de esta práctica exploran nuevas técnicas instrumentales, sonoridades, ritmos y formas de comunicación musical, rompiendo con las jerarquías y los géneros tradicionales. Algunos de los artistas más reconocidos de esta corriente son Evan Parker, John Coltrane, Peter Brötzmann, Derek Bailey, el grupo AMM, entre muchos otros.

Sí bien en la actualidad existen proyectos que comparten algunas características con SEALI (Escobar-Castañeda, 2022), este sistema se distingue por su enfoque en la convergencia entre la escucha artificial y la libre improvisación. En este sistema, la escucha artificial se concibe como un proceso modelado a través de algoritmos de aprendizaje y escucha de máquinas (Collins, 2012), partiendo de mi propia escucha, generando así un marco de referencia que toma mi bagaje como compositor e improvisador.

Inicialmente, este enfoque contempla el análisis de elementos sonoros dentro de la libre improvisación y, más adelante, la interacción dinámica entre la escucha artificial y la improvisación. Algunos elementos principales por destacar en este proceso son: la selección de un repertorio sonoro, la aplicación de técnicas algorítmicas para su análisis, la implementación de redes neuronales capaces de capturar numéricamente la esencia de la improvisación libre. Estas técnicas no solo posibilitan el análisis y la respuesta en tiempo real a la improvisación, sino también las interacciones a distintos niveles temporales con otros improvisadores, contribuyendo así a la evolución y expansión de la improvisación libre.

Desarrollo de SEALI

A continuación, presento una revisión de las funcionalidades que componen a SEALI. Estos aspectos se dividen en dos categorías: las funcionalidades relacionadas con la clasificación tímbrica y las relacionadas con la predicción de la estructura en la improvisación libre.

Corpus musical y base de datos

La primera tarea de SEALI, es analizar un corpus musical de improvisaciones libres que servirá como fuente de *conocimiento* a nivel tímbrico y de estructuras formales recurrentes en este contexto. Inicialmente, comencé con grabaciones de diversos

improvisadores de Europa, Estados Unidos y México desde los años 60 hasta la actualidad, sumando 350 archivos de audio, 29 GB de datos y 45 horas de reproducción. Cabe destacar, que este corpus sigue en evolución a lo largo de la investigación debido a la retroalimentación con otros músicos y las adecuaciones constantes al sistema.

Criterios de segmentación sonora

Una vez consolidado este corpus, el proceso subsecuente implica un análisis en dos fases: primero, la segmentación del corpus, y segundo, la construcción de una base de datos mediante la extracción de características de audio digital presentes en cada segmento. Como exploraremos más adelante, estas características se denominan descripciones o descriptores de audio (Vinet, 2002). Este proceso garantiza que los segmentos individuales mantengan coherencia entre sí y sean objetos significativos más allá de su contexto. La segmentación se basa en el análisis de gestos sonoros, elementos auditivamente reconocibles y significativos por sí mismos. Estos gestos pueden estar caracterizados por propiedades, como el timbre, la altura, el ritmo, el contenido, la densidad, la energía y la complejidad espectral. El desafío de esta tarea radica en que estos gestos pueden variar en altura, intensidad, duración, contenido espectral e intención al ser ejecutados, lo que crea un fenómeno multidimensional en constante evolución en función del contexto y la interpretación. Por lo tanto, es esencial entrenar al algoritmo de escucha y aprendizaje automático con una base de datos que contenga una amplia variedad de muestras sonoras representativas de la práctica.

El concepto de objeto sonoro, influenciado por la obra de Schaeffer (1996) desempeña un papel esencial en la segmentación del audio en SEALI. Schaeffer introdujo el término 'música concreta' para comprender la música como cualquier fenómeno sonoro perceptible y reproducible, independientemente de su origen o significado. El concepto de objeto sonoro se define como una manifestación sonora percibida como una entidad coherente, independientemente de su origen o significado. Los objetos sonoros se relacionan con el gesto, pensado como unidad sonora, que puede evocar emociones y significados sin depender de un contexto musical previo o futuro. Así, el gesto comunica una idea y contiene atributos comunicativos, su apertura a la reelaboración y recontextualización le otorgan un carácter polisémico y una curva energética.

Cabe mencionar que, a lo largo de la investigación, se examinaron distintas aproximaciones a la segmentación de audio dentro de la improvisación libre, lo que condujo a realizar ajustes y modificaciones en la metodología. Por limitaciones de espacio, en este artículo solo abordaré una de las aproximaciones más significativas.

Segmentación basada en cambios *MFCCs* mediante *SBIC*

A través de experimentaciones con diferentes descriptores y una validación auditiva, se determinó que los cambios tímbricos eran el criterio más adecuado para la segmentación en el contexto de la improvisación libre. Esta segmentación de audio se basó en el uso del algoritmo *SBic* (*Segmentation on Bayesian Information Criterion*) (Peeters & Rodet, 2004), que se enfoca en diferenciar segmentos en una cadena de audio utilizando un umbral principalmente espectral. Este algoritmo emplea un criterio de segmentación bayesiano en función de una matriz de descriptores de audio. En este caso, la matriz de descriptores *MFCCs* (*Mel-frequency cepstral coefficients*), propuesta por Davis y Mermelstein (1980), fue útil para establecer el criterio de segmentación, enfocado en cambios tímbricos sustanciales en el audio. El proceso de segmentación consta de tres fases: segmentación en bruto, segmentación fina y validación de los segmentos. Como resultado, se obtuvieron segmentos de audio que variaban en duración, desde centésimas de segundo hasta treinta segundos. Este criterio permitió establecer un puente con el concepto de objeto sonoro, ya que cada segmento contenía las características necesarias para identificarlo como una unidad coherente, lo que llevó a la creación de un corpus de improvisaciones de gestos sonoros.

Descriptores para el análisis digital de la señal de audio

El siguiente paso fue construir una base de datos derivada del corpus de segmentos de audio, donde cada elemento se describe numéricamente. Para llevar a cabo esta tarea, se empleó la librería *Essentia*, que ofrece una amplia gama de descriptores de audio y herramientas, que permitieron explorar diversas combinaciones. En este contexto, la elección de descriptores de audio relevantes se vuelve esencial para garantizar resultados coherentes y no generar contradicciones ni ruido en el proceso de clasificación (Johnston, 1988; Painter, 1997). La experiencia detallada reveló que la variedad y combinación de descriptores, como *MFCCs, Onset Detection, Loudness, Spectral Centroid, Spectral Flatness, Spectral Contrast, Dynamic Complexity, Spectral Complexity, Mel Bands* y *Chromagram* (Lerch, 2022), inicialmente generaron clasificaciones fluctuantes y en ocasiones poco claras. Este análisis, marcado por resultados aparentemente aleatorios, destacó la importancia de una cuidadosa selección de descriptores de audio. Tras exhaustivas pruebas, fue posible identificar cuatro descriptores —*flatness, loudness, spectral centroid* y *MFCCs*— como los más relevantes para la clasificación de la improvisación libre. Este proceso apunta a la necesidad de seleccionar solamente los datos más relevantes para optimizar la eficacia de los modelos de aprendizaje automático.

Clasificación

Posteriormente, un algoritmo de aprendizaje automático fue empleado para analizar la base de datos generada. Este algoritmo tiene la capacidad de discernir entre las particularidades tímbricas individuales cada segmento de la base de datos a través de la modificación de sesgos y pesos durante el entrenamiento. Para realizar este proceso, la base de datos se divide en dos partes: una se destina al entrenamiento, y la otra a probar el modelo computacional resultante. Durante el aprendizaje, el modelo proporciona un índice de certeza para clasificar o predecir diversos datos en relación con los datos del conjunto de entrenamiento.

SEALI toma dos enfoques: el aprendizaje supervisado y el aprendizaje sin supervisión (Kotsiantis et al., 2007). La distinción clave entre estos enfoques radica en que el primero requiere una base de datos anotada, mientras que el segundo usa datos no clasificados (Kotsiantis, 2007). En el campo del aprendizaje automático, la clasificación se considera un caso de aprendizaje supervisado en el que se dispone de un conjunto de datos previamente etiquetados, ya sea por humanos o por un modelo de agrupación de datos (Hastie et al., 2009). El aprendizaje no supervisado, por otro lado, se conoce como *clustering* (Xu & Wunsch, 2008) y consiste en agrupar datos en categorías que comparten ciertas características basadas en una medida de similitud. Este tipo de modelo se crea utilizando datos de entrenamiento y su fin es identificar la categoría a la que pertenece un elemento con características específicas. Cuando se presenta un nuevo ejemplo sin etiquetar a un modelo de clasificación, este determina la etiqueta en función de su posición relativa con respecto a los límites definidos por el modelo.

El propósito de la clasificación es asignar subcategorías a un conjunto de categorías previamente definidas. Estas categorías deben estar etiquetadas previamente, ya sea mediante una clasificación binaria—*one-hot-encode*—o una clasificación de múltiples clases. Generalmente, estos datos se analizan individualmente al extraer propiedades numéricas llamadas características o descriptores de audio (Zhang & Kuo, 2013). Los modelos de clasificación se construyen utilizando diversos enfoques, como umbrales simples, técnicas de regresión o redes neuronales (Alpaydin, 2004). El propósito del clasificador es generar una salida que corresponda a las categorías conocidas en los datos de entrenamiento. De esta manera, el modelo puede identificar elementos que no ha visto previamente, pero que comparten similitudes con los datos de entrenamiento. Otra aproximación implica entrenar al modelo para predecir valores futuros en función de secuencias de datos de entrada y datos históricos, a menudo utilizando redes neuronales recurrentes (Briot et al., 2019).

La principal función del aprendizaje automático y de los modelos es la de procesar y analizar datos sin la necesidad de definir reglas explícitas para cada problema. A diferencia de los sistemas expertos que requieren reglas definidas, los modelos de redes neuronales pueden abstraer y sintetizar información para identificar patrones y predecir nueva información. Esto es útil en una amplia variedad de problemas, incluyendo aquellos relacionados con aspectos sensibles de naturaleza humana, así como la creatividad y el arte. Los modelos se construyen a través del entrenamiento, que implica la actualización de pesos y sesgos, la optimización y la regularización de los datos. Estas funciones permiten a los modelos definir límites en el espacio de datos para categorizar nuevas observaciones.

Fiebrink (2011), añade que, los algoritmos de aprendizaje supervisado de uso general destinados a tareas de clasificación y regresión tienen la capacidad de aprender la relación entre las entradas y las salidas. A este respecto, un enfoque interesante es donde los usuarios pueden evaluar los modelos entrenados al ejecutarlos en tiempo real, observando cómo se comporta el sistema, por ejemplo, a través de la generación de sonidos sintetizados a medida que se introducen nuevas entradas. Además, los usuarios tienen la opción de ajustar y modificar de manera iterativa los ejemplos de entrenamiento y reentrenar los algoritmos utilizando los datos modificados. A aproximación podría llamarse aprendizaje automático interactivo y brinda a los usuarios la capacidad de corregir fácilmente múltiples errores del sistema mediante modificaciones *al vuelo* sobre los datos de entrenamiento. Por ejemplo, si el modelo entrenado produce un sonido incorrecto en respuesta a un gesto específico, el usuario puede registrar ejemplos adicionales de ese gesto junto con el sonido deseado y luego entrenar nuevamente el modelo. Esta flexibilidad también permite a los usuarios adaptar y modificar comportamientos del sistema a través del tiempo, como agregar de manera iterativa nuevas categorías de gestos hasta que la precisión aumente o no se requieran más categorías adicionales. El aprendizaje automático interactivo posibilita que las personas creen modelos precisos incluso con muy pocos ejemplos, logrando esto al insertar iterativamente nuevos ejemplos de entrenamiento en áreas del espacio de entrada que son críticas para mejorar el rendimiento del modelo (Fiebrink, 2011).

En un contexto de trabajo con compositores enfocados en el diseño de instrumentos controlados por gestos mediante el uso del aprendizaje automático interactivo Fiebrink (2010), observa una estrategia útil para explorar nuevos sonidos y relaciones entre gestos y sonidos. En este enfoque, los compositores primero establecen los “límites sonoros y gestuales del espacio compositivo”, definiendo valores para los parámetros de síntesis y las posiciones gestuales del controlador. Luego, crean un conjunto inicial

de datos de entrenamiento utilizando gestos y sonidos en estos límites establecidos. Después de entrenar las redes neuronales con estos datos, los modelos resultantes permiten a los compositores navegar por el espacio de gestos, descubriendo nuevos sonidos tanto dentro como fuera de los límites establecidos por los ejemplos iniciales. De manera análoga a lo que se describe en Fiebrink et al. (2012), en el contexto de SEALI, después de que el algoritmo *Mean Shift Clustering* analizó la base de datos y los agrupó en distintas carpetas según la clase asignada, los segmentos se sometieron nuevamente al análisis de los mismos descriptores de audio. Al finalizar este proceso, se logró construir una base de datos anotada definida por el algoritmo de agrupamiento *MSC*. Esta fase del proceso puede ser revisada posteriormente por un humano para verificar y corregir cualquier discrepancia. En este sentido, he decidido mantener en cierta medida la propuesta de clasificación del sistema, incorporando la colaboración humano-máquina, en la cual la máquina aporta su algoritmo derivado del proceso de toma de decisión—humana—orientada a la identificación de materiales sonoros.

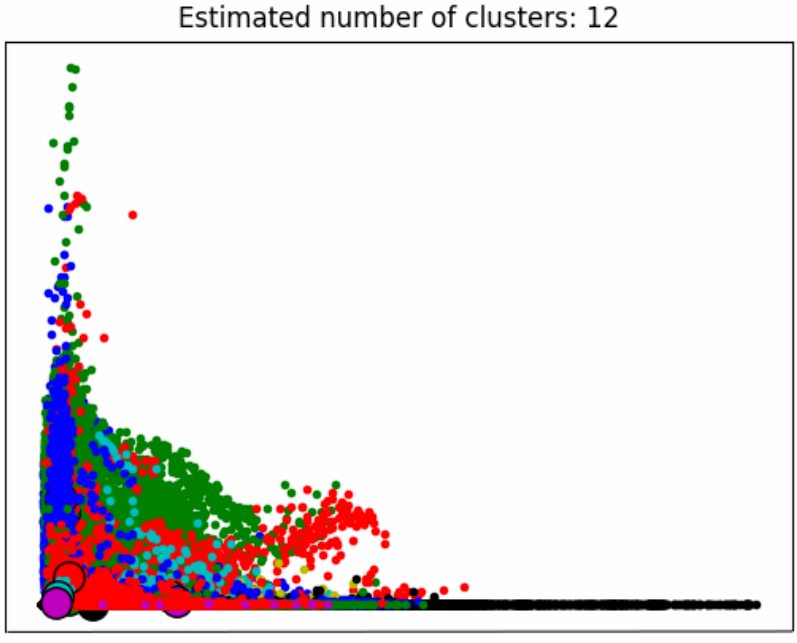
Creación de base de datos anotada con *Mean Shift Clustering*

Una de las razones que me llevó a elegir el algoritmo de *Mean Shift Clustering* (Comaniciu & Meer, 2002) fue la intención de evitar sesgos predefinidos para permitir que el proceso se basara en un sentido emergente, influenciado por la interacción algorítmica. Esto implicó que la aproximación se basara en una caracterización preconcebida o sesgada por teorías específicas relacionadas con alguna clasificación sonora o musical, sino generar un proceso que pudiera, en cierta medida, prescindir del juicio humano para distinguir entre las diferentes categorías sonoras dentro del corpus. La idea detrás de esta aproximación es utilizar medidas de similitud o diferencias sonoras para definir grupos basados en características tímbricas. En este sentido, se empleó el algoritmo de *MSC* para agrupar segmentos sonoros en diferentes clases. Lo interesante de este algoritmo es que no requiere una indicación explícita sobre el número de clases a crear, a diferencia del algoritmo *K-Means*, ya que se basa en encontrar concentraciones de elementos en el espacio de datos y actualizar iterativamente los puntos centrales en función de un umbral de decisión. Después, se filtran los puntos para eliminar duplicados cercanos y se forma un conjunto final de puntos centrales rodeados por elementos cercanos a cada conjunto.

El objetivo de incluir este enfoque es permitir que los algoritmos descubran patrones numéricos en los datos del audio para así evitar el sesgo que un humano podría introducir en este proceso y dejar al algoritmo tomar esa decisión. Si bien, esto refleja una diferencia fundamental entre la percepción humana y de máquinas debido a que en algunos casos la clasificación no se adecuaba a las decisiones que un humano

pudiera haber tomado, destaca la necesidad de comprender cómo funcionan los algoritmos de clasificación sin supervisión y cuáles son sus limitaciones en contextos complejos como los presentes en la improvisación libre. A continuación, presento los resultados de clasificación producidos por el algoritmo *MSC* utilizando esta base de datos según los criterios mencionados.

Figura 1
Clasificación de la base de datos con el algoritmo de MSCS



Análisis de la forma con redes neuronales recurrentes

Cuando se trata de analizar la forma musical, es decir el desarrollo que una obra sonora o musical tiene a lo largo del tiempo, el análisis de las secuencias de las bases de datos es útil para comprender la estructura y predecir posibles estados futuros a partir de una secuencia inicial. Para lograrlo, los datos se dividen en series de "n" pasos de tiempo (*timesteps*) y se someten al análisis de una red neuronal recurrente (*LSTM, long short*

term memory) (Hochreiter, 1997). Este proceso capacita a SEALI a discernir entre las particularidades tímbricas, pero no solo eso, también a organizar las estructuras numéricas presentes en el corpus a lo largo del tiempo. Como resultado, se genera un modelo computacional que puede ejecutarse en tiempo real para realizar nuevas clasificaciones o predicciones inicialmente sobre células sonoras y posteriormente sobre la forma musical.

El abordaje de las redes LSTM en SEALI es esencial para explorar las estructuras sintagmáticas en la improvisación musical, lo que proporcionaría una comprensión más profunda sobre distintos parámetros del desarrollo temporal del sonido, como la amplitud, el cambio espectral y el timbre. Esta perspectiva se enfoca en evaluar la capacidad de estos modelos para prever la dirección de la improvisación y su aplicación en tiempo real para interactuar en la ejecución musical. Este enfoque no solo busca reconstruir improvisaciones pasadas, sino también anticipar elementos inéditos de manera coherente, ampliando así las posibilidades creativas de SEALI en la improvisación musical.

Para validar su uso, realicé un ejercicio preliminar utilizando las *Variaciones Golberg* de J. S. Bach. El objetivo de este ejercicio fue evaluar de manera general las capacidades y limitaciones de esta herramienta, con el fin de identificar algunas de las variables de entrada y estado necesarias para generar reconstrucciones armónicamente coherentes dentro del lenguaje musical de Bach. Este análisis inicial resultó útil para demostrar que las redes LSTM pueden ofrecer una aproximación a la estructura musical, ya que, a partir de una secuencia inicial, pueden reconstruir los elementos subsecuentes utilizando la base de datos de entrenamiento. Además, pude identificar secciones musicales del tipo AABB presentes en las variaciones Goldberg, lo que contribuyó a mi comprensión sobre el funcionamiento de estas redes y su comportamiento en contextos musicales tonales. Para concluir este experimento, comparto algunos ejemplos generados con versiones del modelo en diferentes estados de entrenamiento, desde un 65 hasta un 95 por ciento de certeza, disponible en https://archive.org/details/goldberg-65/.

Posteriormente, el primer experimento en torno a la improvisación libre y las redes LSTM, partió del álbum “More74” del improvisador Derek Bailey, que consiste en 13 improvisaciones libres realizadas con guitarra eléctrica. Para analizar estas improvisaciones, segmenté los audios según cambios tímbricos utilizando los algoritmos *SBIC* y *MFCCs* de la librería *Essentia*, resultando en un total de 17802 segmentos de audio con duraciones de 100 milisegundos a 1600 milisegundos (o 1.6 segundos). A continuación, extraje las características de audio de estos segmentos

utilizando el descriptor *Flatness* para medir el nivel de ruido o sutileza tímbrica de cada uno y guardé estas características en un archivo *csv*. Posteriormente, transformé este archivo *csv* a formato *json*, redondeando todos los datos a un entero y dos decimales, lo que resultó en un máximo de 101 posibles elementos dentro de toda la base de datos. Los datos fueron dispuestos en secuencias con pasos de tiempo fijos estructurando así los 17802 elementos secuenciados en 100 pasos de tiempo, lo que resultó en 17802 listas de 100 elementos para análisis secuencial. Tras entrenar la red LSTM con estos datos durante 1723 iteraciones, la función de pérdida descendió hasta alcanzar un mínimo de 0.0100. Luego, realicé pruebas para verificar si el modelo podía reconstruir secuencias idénticas a las originales. Pasando la secuencia original del elemento 0 al 99, el modelo devolvió el elemento 100 como siguiente valor. Iterando este proceso 100 veces, donde las nuevas predicciones se convirtieron en nuevas entradas para las secuencias, se observó que el modelo reconstruyó perfectamente las secuencias derivadas de las originales.

Una vez completados estos experimentos, el siguiente paso fue dotar a SEALI de herramientas para realizar predicciones multivariables utilizando vectores compuestos por varias características de audio, esta técnica es conocida como secuencia a secuencia (*seq to seq)*. Para ello es crucial integrar otras capas que permitieran trabajar con un formato multivariable en la red, como las capas *Repeat Vector* y *Time distributed* (Brownlee, 2017). La implementación de estas capas permitió que el modelo predijera estructuras multiparamétricas y pudiera sistematizar los momentos de una improvisación libre más allá de lo que se lograba al predecir un solo vector.

Las redes LSTM *seq to seq* fueron fundamentales para dotar a SEALI de los mecanismos necesarios para analizar y predecir estructuras sonoras en improvisaciones libres. Los resultados obtenidos muestran el potencial de esta aproximación para la creatividad computacional en el contexto de la improvisación libre. En el arte, esta algoritmicidad Peeters (2020) puede representar una oportunidad para la exploración y la creatividad, aunque también plantea desafíos y cuestionamientos en cuanto a la interacción humano-máquina.

Primera aproximación interactiva de SEALI

El enfoque de las técnicas utilizadas en SEALI se centra en encontrar un equilibrio entre diversos componentes importantes en la improvisación musical, como la imprevisibilidad, el azar y la estructura formal. Se busca crear un sistema que integre estas perspectivas utilizando las posibilidades y limitaciones de la tecnología al momento de la creación de este proyecto y las capacidades de programación del autor

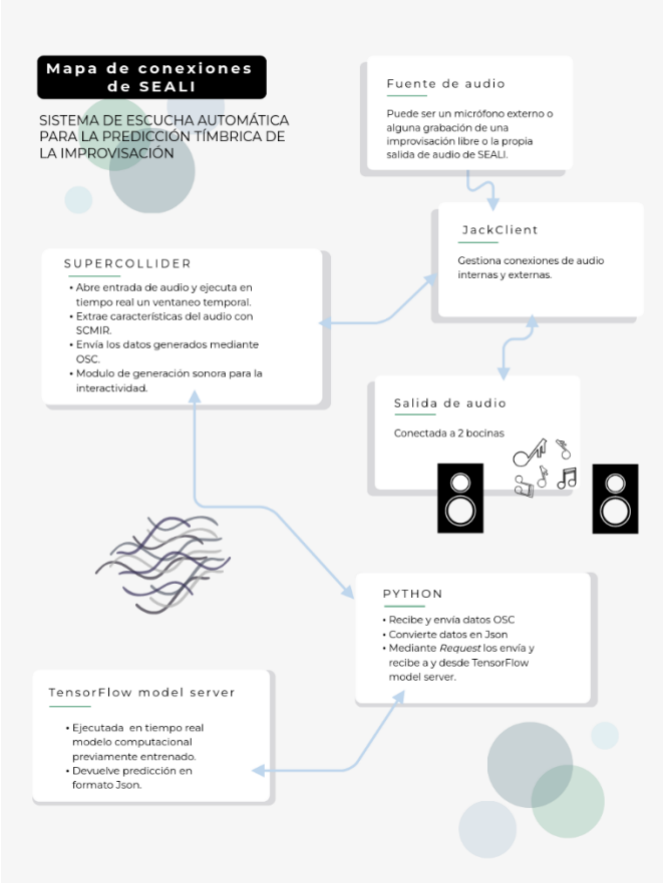
que escribe este texto. Sin embargo, en el estado del arte actual, aún se reconocen ciertas limitaciones tecnológicas en la generación de audio de alta fidelidad en tiempo real a través de redes neuronales profundas. Por lo tanto, en la parte reactiva de SEALI, se emplea síntesis de sonido en tiempo real y transformaciones de muestras de audio.
Para desarrollar un modelo computacional robusto capaz de interpretar nueva información en tiempo real a través de una base de datos anotada, se utiliza el aprendizaje supervisado mediante una red neuronal profunda creada con TensorFlow. Esta red consta de varias capas y miles de parámetros, lo que le permite realizar predicciones en tiempo real de nuevos ejemplos de audio a través de TensorFlow Server.

Una vez que el modelo está entrenado, es capaz de realizar predicciones en tiempo real sobre nuevos ejemplos de audio. Estos nuevos ejemplos se capturan a través de un micrófono o una conexión interna utilizando JackClient. Cada segmento de audio capturado se somete a un análisis con una ventana temporal previamente definida, en este caso, una duración de dos segundos. Las características de cada segmento se extraen mediante SuperCollider y la librería SCMIR, luego se envían a Python a través de OSC, donde se convierten en formato JSON. Luego, esta información se envía a TensorFlow Server para su clasificación y se devuelve a Python en formato OSC. Una vez que los datos de clasificación se reciben en SuperCollider, se procesan a través de un módulo de generación sonora, que utiliza reglas basadas en límites de certeza de clasificación. Este módulo activa archivos de audio de la base de datos clasificada, sometiéndolos a transformaciones como estiramientos temporales, cambios de tono, síntesis granular y síntesis concatenativa. Además, activa sintetizadores cuyos parámetros se modifican según los datos de clasificación, lo que se considera como un proceso de sonificación de las clasificaciones del modelo. A continuación, presento un diagrama del funcionamiento antes descrito.

Segunda aproximación interactiva de SEALI

En esta segunda aproximación, se entregan nuevos ejemplos de audio al modelo, los cuales se dividen en segmentos de dos segundos u otra ventana temporal predefinida que se decida. Mediante Python y a través de la librería *Essentia*, se extraen las características de estos segmentos y los datos se envían a TensorFlow Server para obtener una nueva predicción. Una vez que la información sobre el segmento de audio se clasifica, se envía a Python en formato JSON y Python la convierte en formato OSC para que SuperCollider la reciba.

Figura 2
Mapa de conexiones de SEALI para la predicción tímbrica e interacción en la improvisación libre



A partir de los datos de predicción generados por el modelo, existen varias formas de abordar la interactividad de SEALI. Una opción es sonificar cada una de las clases detectadas por el modelo. Esta estrategia podría denominarse 'aproximación continua', ya que se consideran todos los valores de la predicción, lo que resulta en una continuidad numérica en los niveles de detección para cada clase. Además, si el modelo fue entrenado con 10 clases, devolverá 10 valores diferentes que suman uno en total, donde el valor más alto se interpretará como la clase predominante en la predicción. A continuación, presento una improvisación generada de la interacción de SEALI con la

misma salida de audio que devuelve *supercollider*, es decir, escuchando sus propios procesos de retroalimentación (disponible en https://archive.org/details/seali-sola-23-nov-2020).

La segunda opción sería una 'aproximación discontinua con propiedades deterministas', aquí, se podría asignar la activación de un proceso relacionado con la detección de la clase predominante, en un momento sonoro específico, mediante reglas booleanas. Es importante destacar que estas dos posibilidades, continua y discontinua, podrían combinarse para enriquecer las capacidades interactivas de SEALI. Esto significa que cuando se detecta una clase específica, se podría activar un sintetizador u otro proceso, y, además, sus parámetros podrían modularse continuamente según los valores de predicción del modelo.

La metodología interna de SEALI abarca una serie de algoritmos que funcionan en línea tratando de emular un comportamiento en tiempo real, incluyendo la activación de entrada de audio, segmentación de audio, extracción de características, transformación de datos y generación sonora para la interacción. A pesar de que requiere escuchar un número de muestras de audio antes de interactuar, su capacidad de reacción es casi inmediata, lo que sugiere una respuesta instintiva y reactiva. En este estado de desarrollo, SEALI no tiene agencia sobre la toma de decisiones en este nivel interactivo, más allá de su capacidad para clasificar y detectar materiales. Además, la predicción de SEALI puede verse influenciada por factores externos, como ruidos ambientales, sonidos de los músicos y sonidos generados por SEALI, lo que afecta el resultado de la predicción. Una forma de abordar esta situación sería entrenar al sistema con ejemplos que incluyan estas texturas sonoras mixtas, lo que podría proporcionarle una comprensión más sólida de su propio proceso de retroalimentación y permitirle tomar acciones basadas en estas detecciones.

Intra-acciones y procesos de retroalimentación humano-máquina

Desde una perspectiva más teórica, un concepto clave que me sirvió para describir los procesos a los que estaba sujeto el desarrollo algorítmico de SEALI es la agencialidad algorítmica o *algoritmicidad* (Peeters, 2020). Este enfoque analiza cómo la algoritmicidad se inserta en los diferentes procesos dentro del ciclo de retroalimentación que contempla la práctica artística, el desarrollo tecnológico y la investigación. En este sentido, (Peeters, 2020) destaca cómo los algoritmos pueden ser considerados agentes activos en estos procesos, afectando activamente la interacción entre humanos y tecnología. Resulta interesante analizar a profundidad este ciclo de retroalimentación para comprender cómo la agencialidad algorítmica influye directamente en el desarrollo del proyecto y en los resultados sonoros generados. Estos

resultados parecen surgir de mis decisiones personales, sin embargo, tanto los sesgos algorítmicos como los sesgos humanos impregnan todo el proceso, desde la micro hasta la macro-escala (Gillespie, 2013).

Las funcionalidades de SEALI experimentan un proceso de afectación constante, influyendo de manera activa en la forma en que segmenta y clasifica la información recibida, en las predicciones de nuevos datos y en el resultado general que produce la interacción mutua de todas las partes implicadas en el proceso. Resulta relevante mencionar a Barad (2007) en este contexto, ya que introduce el concepto de intra-acción, el cual reconoce la constitución mutua de agencias entrelazadas, en contraposición a la noción de "interacción", que sugiere agencias individuales separadas antes de cualquier interacción. Debido a la escucha artificial/(asistida) de SEALI y su capacidad para analizar, sistematizar, clasificar, predecir y aportar elementos en la improvisación, surge la noción de una forma de escucha trans-humana. Esta perspectiva considera a SEALI como más que una herramienta tecnológica, sino como un agente capaz de *intra-accionar* con otros músicos, influyendo en la transducción de sus propias formas de escuchar, hacer y concebir la música libremente improvisada. Esta transducción, que está condicionada por la retroalimentación humano-algorítmica, se ve afectada por múltiples procesos recursivos que han configurado el sistema en su conjunto. En este sentido, el concepto de intra-acción de Barad cobra relevancia al reconocer la co-constitución de agencias entrelazadas en esta dinámica trans-humana de creación musical.

La escucha artificial de SEALI está intrínsecamente ligada a diversos aspectos que moldean su funcionamiento. Estos incluyen su conocimiento, que parte de la base de datos con la que se entrena, así como el sesgo algorítmico inherente que se deriva de la programación de los algoritmos. Además, se añaden a esta ecuación los múltiples bucles de retroalimentación que se generan a través de procesos de validación e interacción humana. Como resultado, el desarrollo de SEALI ha implicado una serie de intra-acciones que han requerido revisión, adecuación, modificación, reelaboración y reestructuración incluso de sus objetivos iniciales. Estos ajustes continuos reflejan la influencia constante de las intra-acciones humano-algoritmicas, resaltando así la importancia de considerar la performatividad de agencias entrelazadas en su propia praxis.

Al considerar su activa performatividad en torno a la improvisación libre, se puede observar un tipo de *transferencia de conocimiento* a través de las bases de datos con las que el sistema fue entrenado. La transferencia de conocimiento se limita a la predicción numérica y las múltiples filtraciones a las que se sometieron los datos de entrada, así como el proceso de comparación entre entrada y salida, en lugar de una transferencia

del conocimiento musical *real* sobre el estilo y las aproximaciones técnicas hacia la improvisación libre. Es evidente que una máquina no puede abstraer ni comprender lo que ocurre en la improvisación libre; en su lugar, puede sistematizar una serie de operaciones que le permiten identificar patrones superficiales del contorno de esta práctica musical. Además, fue necesario realizar numerosos procesos de abstracción en la información, lo que inevitablemente disminuye la resolución de las predicciones generadas por SEALI. Sin embargo, al restringir la escucha de máquina a ciertas descripciones numéricas del audio, es posible obtener resultados satisfactorios en las clasificaciones y predicciones del audio, así como la generación de réplicas numéricas estructurales idénticas a las versiones grabadas de algunas improvisaciones libres con las que trabajé durante su desarrollo.

Aplicaciones en Contextos Artísticos

SEALI fue sometido a diversas pruebas y etapas de desarrollo, entre ellas, la colaboración con improvisadores y músicos fue muy relevante. Estas colaboraciones proporcionaron valiosos aportes al proyecto, así como materiales sonoros que sirvieron para entrenar modelos computacionales adaptados a la práctica de improvisación libre y '*comprovisación*' de los participantes (Dudas, 2010). Además, estas interacciones dieron lugar al ciclo de conciertos "Resistencias Maquínicas", que exploró los procesos de colaboración humano-máquina en contextos de improvisación libre, esto ayudó a generar nuevos materiales musicales y mejorar la adaptabilidad de SEALI. Desde esta perspectiva, SEALI tuvo vínculos importantes con la escena de música experimental e improvisación libre en la Ciudad de México, enriqueciendo su desarrollo y llevándolo a interactuar de manera más asertiva con los músicos/improvisadores. Asimismo, se realizaron algunas colaboraciones en espacios como Toplap Barcelona y Hangar, esto ayudó a posicionar el proyecto a nivel internacional y recibir valiosa retroalimentación de comunidades interesadas en el desarrollo tecnológico y la creación musical, especialmente el livecoding. SEALI también tuvo presencia en el Programa Arte, Ciencia y Tecnología (ACT), en el Foro Internacional de Música Nueva Manuel Enríquez, en Música UNAM y en diversos foros independientes en la ciudad de Aguascalientes. Es importante mencionar que, si bien los desarrollos actuales de SEALI son el resultado del trabajo individual y la retroalimentación obtenida a través de la práctica improvisativa e investigativa, es necesario reconocer la valiosa participación en etapas tempranas del desarrollo de Diego Villaseñor, así como por Hugo Solis, Ivan Paz y Caleb Rascón.

El sistema está disponible para su instalación y uso bajo licencias libres en el repositorio de GitHub https://github.com/Atsintli/SEALI-V.I Este repositorio está

sujeto a la Licencia Pública General de GNU Versión 3 y se encuentra en constante evolución a medida que SEALI adquiere nuevas funcionalidades. La idea es fomentar la colaboración activa de futuro público interesado en el proyecto, músicos e improvisadores libres, con el fin de construir una comunidad en torno a esta iniciativa, sus posibles derivados y, especialmente, las técnicas de escucha y aprendizaje de máquinas aplicadas a la creación musical. Por otro lado, cabe mencionar que las bases de datos y los modelos creados, están abiertos y disponibles para realizar pruebas, experimentos futuros o sus posibles derivados.

Interdictos protésicos, estudio de caso

El desarrollo algorítmico de SEALI en su versión destinada a la comprovisación (Dudas, 2010), titulada "Interdictos Prostéticos" (https://archive.org/details/aaron-escobar-interdictos-protesicos), implicó desde el inicio una interacción creativa entre humanos y algoritmos. Esta versión toma una base de datos de gestos sonoros grabados por músicos e improvisadores, utilizando como inspiración los procesos de escucha automática de SEALI. Desde esta perspectiva, se busca establecer un vínculo entre la expresión sonora del músico/improvisador y la detección de elementos sonoros, formando un continuo orgánico-digital. Partiendo del concepto de escala, exploramos diversas aproximaciones a las cualidades acústicas y prácticas de los músicos, tomando tres parámetros: digitación, envolventes acústicas y contenido espectral, que a su vez incluyen brillo, complejidad, densidad y centralidad espectral. Los materiales generados fueron analizados para crear una base de datos de descriptores de audio y luego se agruparon con MSC. Posteriormente, estos datos reagrupados fueron analizados por una red neuronal profunda que generó un modelo capaz de identificar los materiales propuestos por cada músico/improvisador. A través de estos procesos de reconocimiento, SEALI toma decisiones en tiempo real en términos tímbricos, energéticos e interactivos, proponiendo gestos sonoros similares o contrastantes a los que escucha, de acuerdo con el conocimiento que tiene sobre el contexto de una improvisación. A continuación, presento una versión ensamblada con los materiales grabados por los músicos y las intra-acciones de SEALI: https://bit.ly/3qagvMh. Además de la versión que se estrenó en el MUAC, UNAM: https://www.youtube.com/watch?v=U3PrrrzxB4c

Esta obra permite su interpretación en diferentes modalidades, con distintas instrumentaciones e intérpretes. En el performance de "Interdictos Prostéticos" realizado por Miguel Ángel Cuevas (https://youtu.be/OUjAYWYenco), se observan varios fenómenos de retroalimentación que desafían tanto al improvisador como a SEALI. En ese performance, se destaca la propuesta musical estructural que aporta

SEALI, la cual se basa en las aproximaciones a la forma musical que proponen los improvisadores que formaron parte de los datos de entrenamiento del sistema. De esta manera, se conserva una esencia estructural que guía la improvisación, permitiendo que SEALI adapte su comportamiento a diversas situaciones sonoras. Este proceso puede considerarse como una abstracción del estilo musical, trascendiendo el tiempo y el espacio para integrarse en un flujo iterativo y continuo en múltiples contextos. Al conservar ciertos rasgos cadenciales reconocibles por un improvisador, la improvisación puede guiarse estructuralmente desde el conocimiento que caracteriza a ese estilo.

Resultados y discusión

La principal innovación introducida por SEALI radica en sus múltiples funciones y su capacidad para integrarlas, lo que hace útil a cada una de ellas tanto para músicos como para no músicos, improvisadores, investigadores musicales y programadores. Esta estructura ofrece la flexibilidad de derivar cada función en distintas mini-aplicaciones según las necesidades del proyecto. Aunque improvisar con SEALI requiere recursos computacionales relativamente modestos para ejecutar los modelos creados, el entrenamiento de ellos no fue sencillo. Se necesitó trabajar con servidores del IIMAS en la UNAM, proporcionados por el Dr. Caleb Rascón, aunque la capacidad limitada de estos impidió la realización de experimentos más amplios debido a restricciones de memoria para trabajar con bases de datos y arquitecturas neuronales más complejas.

Esta situación nos sitúa ante una nueva forma de tecnología que, si bien es accesible para cualquier persona con una computadora e internet, solo las grandes empresas o instituciones, en algunos casos universidades, pueden acceder a hardware especializado capaz de crear modelos mucho más grandes y complejos. A lo largo de esta investigación, se buscó tener acceso a algunos de los clústers de supercómputo de algunas instituciones en México, sin embargo, la falta de información y convocatorias vinculadas a proyectos de arte, ciencia y tecnología imposibilitó esta alternativa. Definitivamente, un mayor poder de cómputo ampliaría significativamente las capacidades de aprendizaje y memoria de SEALI.

Una de las discusiones abiertas en torno al desarrollo de este sistema es su falta de accesibilidad a público no familiarizado con la programación. Pese a ser de código abierto y estar disponible bajo licencias libres, la complejidad de la interconexión de sus módulos requiere simplificarse. Una posible solución es emplear plataformas como Docker o Kubernetes, que permitirían centralizar la gestión de los módulos en un servidor. Esto simplificaría enormemente el proceso de instalación y utilización,

haciendo que SEALI sea más accesible. En este contexto, al explorar diversas opciones, una idea es utilizar JavaScript, HTML y CSS para montar todo el sistema en una página web. Si bien esta aproximación tiene limitaciones en cuanto a la calidad y fluidez sonora, ampliaría la accesibilidad del sistema a personas que cuenten con acceso a internet y una computadora.

Conclusiones

El desarrollo de este sistema busca abordar la afectación como una propiedad intrínseca y necesaria para la activación de espacios sujetos a la intra-acción, compartida tanto por agentes digitales como humanos. En este sentido, es fundamental encontrar un estado de relación más equitativo con los dispositivos tecnológicos y seguir cuestionando aquellos elementos externos que determinan nuestras posibilidades de relación con la tecnología, usualmente desde una lógica heredera del "racionalismo modernista europeo, que [insiste] en la legitimidad simbólica de un modelo de sociedad capitalista" (Adler, 2021). Por ello, en el desarrollo tecnológico propuesto en esta investigación, se promueve el empleo del software libre y el código abierto, así como la posibilidad de replicar los desarrollos tecnológicos por parte de otros artistas o programadores.

Desde esta perspectiva, se considera que los desarrollos tecnológicos no necesariamente deben cumplir con criterios de funcionamiento óptimo o validación científica estricta para producir resultados estéticos o plantear cuestionamientos interesantes. Más bien, esta falta de rigor y la apuesta por reconocer los entrelazamientos agenciales humano-máquina, pueden ser aprovechados para transformar la práctica artística y las utilidades finales de estas herramientas, ampliando así el panorama de acción de cada área por separado y en conjunto.

En el ámbito musical, se apuesta por cuestionar las prácticas y herramientas predominantes en la creación musical académica, buscando desmantelar sus estados más predominantes. Esto implica mirar hacia la multiplicidad que ofrecen el código, el pensamiento algorítmico y la creación sonora, partiendo de una escucha profunda (Oliveros, 2005) en interacción con la investigación y la improvisación libre.

En su totalidad, esta investigación busca examinar el cruce del arte, la ciencia y la tecnología para abrir la posibilidad de aplicar herramientas tecnológicas a la creación de nuevas experiencias artísticas y otros posibles imaginarios creativos. Aunque aún hay aspectos de SEALI por explorar, como integrar más bases de datos, arquitecturas de redes neuronales más grandes y profundizar en sus mecanismos de interacción con

otros improvisadores, los resultados obtenidos hasta ahora, de acuerdo con la retroalimentación de los músicos/improvisadores, son satisfactorios. Los procesos de organización emergentes, característica inherente del aprendizaje automático, permiten que surjan, por medio de la correcta utilización de las herramientas, los rasgos distintivos sobre la forma musical y otras características concernientes a la práctica de la improvisación libre. Pese a que aún no puede adaptarse a otras prácticas musicales más allá de la improvisación libre, su capacidad para reaccionar de manera agnóstica a lo que escucha es un aspecto interesante que le permite activar sus respuestas sonoras en cualquier situación, considerando sus restricciones booleanas presentes en los procesos de sonificación de datos y de transformación sonora.

Sin duda, existen aspectos técnicos que requieren elaboración para lograr una integración más fluida entre los elementos sonoros propuestos por SEALI y los gestos sonoros proporcionados por los músicos/improvisadores, las improvisaciones presentadas aquí son evidencia de que es posible crear una forma musical diferente, incidir activamente en la praxis de la improvisación libre y en la forma en la que creamos en colaboración con las máquinas.

Obras citadas

Alpaydin, E. (2004). *Introduction to machine learning*. MIT Press.

Bailey, D. (1980). *Improvisation. its nature and practice in music*. Moorland.

Barad, K. M. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Briot, J., Hadjeres, G., y Pachet, F. (2019). *Deep Learning Techniques for Music Generation*. Springer International Publishing.

Brownlee, J. (2017). *Long short-term memory networks with Python: Develop sequence prediction models with deep learning*. Machine Learning Mastery.

Collins, N. (2006). Towards autonomous agents for live computer music: Realtime machine listening and interactive music systems (Tesis doctoral). University of Cambridge.

Collins, N. (2012). Automatic composition of electroacoustic art music utilizing machine listening. *Computer Music Journal*, *36*(3), 8-23.

Davis, S., y Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *28*(4), 357-366.

Comaniciu, D., y Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(5), 603-619.

Dudas, R. (2010). Comprovisation: The various facets of composed improvisation within interactive performance systems. *Leonardo Music Journal*, *20*, 29-31.

Escobar Castañeda, A. A. (2022). Resistencias maquínicas: Una caracterización a la improvisación libre con aprendizaje y escucha de máquina [Tesis doctoral, Universidad Nacional Autónoma de México]. Facultad de Música.

Fiebrink, R. A. (2011). Real-time human interaction with supervised learning algorithms for music composition and performance (Tesis doctoral). Princeton University, USA.

Fiebrink, R., Trueman, D., Britt, C., Nagai, M., Kaczmarek, K., Early, M., y Cook, P. (2012, marzo). *Toward understanding human-computer interaction in composing the instrument*. ICMC.

Galiana Gallach, J. L. (2018). De la naturaleza de la improvisación libre: Elementos esenciales para su identificación y diferencias con la composición escrita. Itamar. *Revista de Investigación Musical: Territorios para el Arte*, 0.

Gillespie, T. (2013). The relevance of algorithms. En T. Gillespie, P. J. Boczkowski y K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-194). MIT Press.

Hastie, T., Tibshirani, R., y Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Hochreiter, S., y Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Johnston, J. D. (1988). Transform coding of audio signals using perceptual noise criteria. *IEEE on Selected Areas in Communications*, *6*, 314–323.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, *31*(3), 249-268.

Kotsiantis, S. B., Zaharakis, I. D., y Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*, 159–190.

Lerch, A. (2022). *An introduction to audio content analysis: music information retrieval tasks and applications*. Wiley.

Matthews, W. (2012). *Improvisando: La libre creación musical*. Turner Publicaciones.

Oliveros, P. (2005). *Deep listening: A composer's sound practice*. iUniverse.

Painter, T., y Spanias, A. (1997). A review of algorithms for perceptual audio coding of digital audio signals.

Peeters, G., y Rodet, X. (2004). Automatically selecting signal descriptors for sound classification. En *Proceedings of the 2004 International Computer Music Conference* (pp. 464-467).

Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity*, *25*, 1-16.

Vinet, H., Herrera, P., y Pachet, F. (2002). The CUIDADO project. En 3rd International Society for Music Information Retrieval (ISMIR) Conference (pp. 197-203). Paris, France. Content Based Retrieval.

Schaeffer, P., y De Diego, A. C. (1996). *Tratado de los objetos musicales* (Serie Alianza música). Alianza.

Xu, R., Wunsch, D. (2008). *Clustering*. Reino Unido: Wiley.

Zhang, T., y Kuo, C. J. (2013). *Content-Based audio classification and retrieval for audiovisual data parsing*. Springer US.