

Sinestes.IA

Alejandro Aristizábal Medina

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE ARTES
DEPARTAMENTO DE MÚSICA
CARRERA DE ESTUDIOS MUSICALES
INGENIERÍA DE SONIDO
BOGOTÁ
2019

ABSTRACT

Sinestes.IA es un proyecto de creación que explora el uso de *Algoritmos de aprendizaje profundo* como herramientas interactivas y creativas. Inspirado en el fenómeno de la *Sinestesia*, este proyecto busca traducir, en tiempo real, los movimientos corporales en música. Con esto en mente, se indaga la relación existente entre la música y la danza, desarrollando un sistema de composición a partir de los *esfuerzo-acciones* de la *teoría Laban*. Por otro lado, se hace uso del concepto de *Modelos del Mundo* como guía para el algoritmo de aprendizaje. Por esta razón, *Sinestes.IA* es capaz de comprender el baile, tanto visual como temporalmente, y de responder a este con instrucciones musicales. Este proyecto demuestra las capacidades, posibilidades y limitaciones que se pueden presentar al momento de desarrollar herramientas de creación interactivas e inteligentes, y propone distintas maneras para obtener mejores resultados.

Palabras clave: *Inteligencia Artificial, Aprendizaje de Máquina, Aprendizaje Profundo, Modelos del mundo, Sinestesia, Teoría Laban, Esfuerzo-acciones, Música Algorítmica, Música Interactiva.*

Sinestes.IA is a creative project that explores the usage of *Deep Learning Algorithms* as creative and interactive tools. Inspired in the psychological phenomenon of *Synesthesia*, this projects aims to translate, in real time, body movements into music. With this in mind, the existing relation between dance and music is examined, developing a composition system based on the *Action-Efforts* described in *Laban's method*. On the other hand, the concept of *World Models* is used as a guideline for the learning algorithm. Because of this, *Sinestes.IA* is capable of understanding movement, both visually and temporarily, and is able to respond to this with musical instructions. This project showcases the capabilities, possibilities, and limitations that may be present when developing smart and interactive tools in the creative means, and proposes different ways for obtaining better results.

Keywords: *Artificial Intelligence, Machine Learning, Deep Learning, World Models, Synesthesia, Laban's Theory, Action-efforts, Algorithmic Music, Interactive Music.*

1. INTRODUCCIÓN	3
2. CONTEXTO	4
3. MARCO TEÓRICO	6
3.1. Redes neuronales	6
3.1.1. Redes neuronales completamente conectadas	7
3.1.2. Redes neuronales convolucionales	7
3.1.3. Redes neuronales recurrentes	7
3.2. Sistemas de tracking	8
3.3 Los esfuerzo-acciones de Laban	9
4. OBJETIVOS	11
5. METODOLOGÍA	11
5.1. Definir el problema de aprendizaje	11
5.2. Obtener bases de datos que reflejen el problema	12
5.2.1. Cómo capturar el video	12
5.2.2. Cómo capturar el sonido	12
5.2.3. Componer los ejemplos sonoros	13
5.2.4. Capturar los datos	14
5.3. Definir un modelo de aprendizaje base	15
5.3.1. Modelo de comprensión espacial	16
5.3.2. Modelo de comprensión temporal	18
5.3.3. Modelo de acción	18
5.4. Entrenar el modelo de aprendizaje	19
6. RESULTADOS	19
6.1. Modelo de comprensión espacial	19
6.2. Modelo de comprensión temporal	20
6.3. Modelo de acción	20
6.4. Resultado final	21
7. CONCLUSIONES	21
8. BIBLIOGRAFÍA	22

1. INTRODUCCIÓN

Pocos fenómenos sensoriales son atribuidos tanto misterio y tanto arte como la sinestesia. Sinestesia es un fenómeno poco comprendido, que se describe como *"la confusión de sensaciones, donde la excitación de un sentido genera estimulación en otra modalidad sensorial completamente diferente"*. (Ternaux, 2003, p. 1, t.d.a). Artistas como Rimbaud, Baudelaire y Kandinsky tuvieron la suerte de comprender medios artísticos por fuera de sus capacidades, y expresarlos a través de sus habilidades.

Lamentablemente, la sinestesia es una condición poco común, razón por la cual el arte de Kandinsky, o los poemas de Baudelaire se destacan del ordinario. La sinestesia, como herramienta artística, tiene una característica interesante, y es que reduce la brecha entre la destreza y la creatividad, permitiendo a escritores expresarse en términos de luz, a pintores expresarse en términos de sonido, entre otros. Si este fenómeno pudiera replicarse, ¿cuántas personas disfrutarían de la capacidad de expresarse en medios que desconocen?

Se han realizado diversos estudios, con el objetivo de comprender las razones por las que este tipo de fenómenos se dan, y en qué manera se expresan; y si bien se han llegado a conclusiones sobre la causa física, los mecanismos cerebrales relacionadas a su expresión siguen siendo desconocidos (Ternaux, 2003, p. 2, t.d.a). Uno de los esfuerzos más reconocibles de comprender la asociación artística entre percepciones en un artista ha sido con Kandinsky, quien logró especificar una teoría bajo la cual su percepción musical se relacionaba con su expresividad visual. Por desgracia, esta descripción es bastante subjetiva y única, no extendiéndose a una generalidad del fenómeno.

"El azul [...] es comparable con el sonido de una flauta. El rojo es cálido, [...] comparable con el sonido de una tuba. El naranja [...] sonido de una campana..." (Di Marco & Spadaccini, 2015, p. 25)

Todo pareciera indicar que estamos lejos de comprender este fenómeno, mucho menos de replicarlo. Pero, hoy en día, tenemos herramientas tecnológicas que podrían permitirnos cerrar la brecha entre la destreza y la creatividad. El aprendizaje de máquina es una de esas herramientas. Esta área de la inteligencia artificial permite hallar patrones entre una gran cantidad de datos, sean estos patrones conocidos previamente o no. Esta capacidad computacional, que a primera instancia suena sencilla, ha permitido una expansión

acelerada al uso de la tecnología en el día a día. Cosas como reconocimiento facial, reconocimiento de voz, procesamiento de lenguaje natural, y mucho más; son algunos ejemplos de aprendizaje de máquina siendo usado en la cotidianidad.

Es probable que no podamos comprender con totalidad la expresión de la sinestesia, pero podemos generar herramientas que busquen simularla. El desarrollo de estas herramientas tiene el potencial de aproximar la expresión creativa a la sociedad, permitiendo a personas sin formación artística expresarse de manera instintiva. Lograr este objetivo implica generar herramientas de creación con funcionalidades intuitivas, donde la curva de aprendizaje sea sutil. Cumplir este objetivo requiere de mucha experimentación y exploración en distintas técnicas y estrategias para facilitar la expresividad artística. Es debido a esto que se crea Sinestes.IA, proyecto artístico con la intención de indagar en el uso de inteligencia artificial como instrumento de expresión musical.

Sinestes.IA propone la exploración de un ambiente musical a través de un medio de expresión muy común: el movimiento. Se ha escogido el movimiento debido a que es una acción natural e intuitiva. Además, su expresión artística, la danza, es un medio que se encuentra altamente ligado a la música debido a que ambas están subordinadas al factor de la temporalidad. Ahora bien, desarrollar una herramienta que pueda traducir el movimiento en música plantea varios problemas desde el punto de vista técnico como conceptual. Entre estos se destacan realizar un sistema que comprenda tanto la danza como la música, y desarrollar una función que logre traducir el movimiento a sonidos. Estos obstáculos serán enfrentados a lo largo del proyecto, con el objetivo de analizarlos y de establecer las soluciones que se implementarán para lograr el objetivo de Sinestes.IA.

2. CONTEXTO

Sinestes.IA se ha visto influenciado por proyectos previos que he desarrollado. Estos trabajos han indagado tanto en el uso de inteligencia artificial de manera artística, como en el manejo del movimiento como interfaz musical. Entre estos proyectos se encuentra *Viaje Musical Programable (MVP)*, el cual exploraba el uso de Kinect para interactuar con unos parámetros musicales desarrollados en Pure Data. La interacción permitía cambiar volumen y paneo de varios loops, al igual que activar, desactivar o cambiar pistas sonoras a través de gestos predeterminados. La intención de este proyecto era crear un instrumento que permitiera a cualquier usuario interactuar con la música, pero su desarrollo fue poco intuitivo, y tenía una curva de aprendizaje acentuada, requiriendo tiempo y habilidad poder expresarse por este medio.

Otro proyecto desarrollado recientemente fue *Música Autómata*. Este buscaba indagar el uso de inteligencia artificial para interactuar con un ambiente sonoro. *Música Autómata* consistía en desarrollar un programa que observaba los gestos faciales del oyente, con el objetivo de maximizar un gesto específico. El programa solo podía interactuar con parámetros sonoros, por lo que debía explorar qué sonidos generaban el gesto facial deseado en el oyente. Gracias a esto el oyente tenía el mismo control sobre la música que el programa, involucrándose de manera pasiva en la composición. El programa está basado en un algoritmo de aprendizaje reforzado, y el ambiente sonoro fue desarrollado en Pure Data.

La idea de explorar el movimiento como interfaz musical no es nueva, y ha sido ampliamente explorada por varios artistas a través de la historia. Entre los experimentos pasados se destaca el uso de Theremins por John Cage y Merce Cunningham en 1965 con la intención de capturar los movimientos de bailarines y transformarlos en sonido en la obra *Variations V*.

En otros ejemplos más recientes se encuentra Imogen Heap con los guantes Mi.Mu, los cuales le permiten captar los movimientos de sus manos con alta precisión y controlar parámetros musicales con ellos. Los guantes Mi.Mu le permiten al artista grabar pistas, agregar efectos, controlar instrumentos y mucho más a través de gestos predefinidos por el intérprete.

Ahora bien, el uso de inteligencia artificial en áreas creativas como el arte y la música también ha sido ampliamente explorado. Entre estos experimentos se destaca la extracción y adición de estilos artísticos en imágenes y pinturas, permitiendo al usuario visualizar cómo una imagen, foto y dibujo se verían bajo el estilo de distintos artistas. Algoritmos como estos le permiten a cualquier persona explorar estilos artísticos al igual que variar su intensidad o relevancia en la composición final de la imagen, y hoy en día se ha demostrado la aplicabilidad de estos algoritmos en áreas de diseño y creación.

En cuanto a la música, herramientas similares se han desarrollado, que estudian el estilo musical de un artista a través de múltiples ejemplos y lo recrean para crear nuevas composiciones, o para transferirlo a motivos o ideas musicales previas. Bajo esta premisa se han construido herramientas y compañías, como Flow Machines y AIVA (Artificial Intelligence Virtual Artist) con resultados comparables e indistinguibles a los compuestos por humanos.

Por último, y probablemente la referencia más cercana a este proyecto, fue la demostración realizada por Yamaha presentada bajo el nombre *Mai Hi Ten Yu*. En esta demostración,

Yamaha desarrolló una inteligencia artificial que capturaba los movimientos musculares de un bailarín para transformarlos en música a través de un piano.

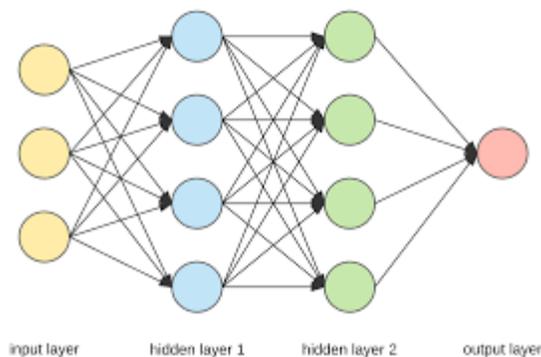
3. MARCO TEÓRICO

Antes de continuar es necesario comprender los conceptos y tecnologías que se piensan manejar durante el desarrollo del proyecto, con el objetivo de comprender su aplicación y relación con Sinestes.IA.

Sinestes.IA es un proyecto que abarca distintas temáticas y conceptos con la intención de definir las herramientas que se usarán para el desarrollo del mismo. Estos temas van desde el tipo de inteligencia artificial y las arquitecturas, como las teorías de comprensión del movimiento y su relación con la música. Siendo así, estos son los conceptos que se cubrirán:

3.1. Redes neuronales

Las redes neuronales son algoritmos usados ampliamente en el área de aprendizaje de máquina e inteligencia artificial. Estos algoritmos fueron conceptualizados con la intención de imitar la funcionalidad del cerebro y la manera en la que aprende. Las redes neuronales son máquinas de cálculo moldeables, que reciben información y sacan información, por lo que se pueden entender como una función matemática. Lo que hace a las redes neuronales especiales comparadas a cualquier otra función matemática es la capacidad de “entrenarse” o adaptarse a los requerimientos de los datos entregados. Esto permite construir modelos en los cuales se sepa con exactitud los datos que debe recibir y generar el algoritmo, aún cuando la manera en que estos datos se relacionan sean desconocidos previamente.



Entrenar una red neuronal requiere de algún método para calificar su rendimiento. Al igual que en la academia se le da un puntaje a los trabajos o exámenes realizados según algún criterio, las redes neuronales necesitan una manera de medir si cumplen o no el objetivo esperado. Conociendo su rendimiento actual, se le pone la tarea a la red neuronal de maximizar su rendimiento, o en el caso del examen, se busca obtener la mejor nota. Es con base en este objetivo, y a través de la prueba y error, que la red neuronal logra aprender.

Para dar un ejemplo de este proceso, supongamos que se le asigna la tarea a alguien de obtener el mejor resultado en un examen desconocido de opción múltiple. Esta persona sólo podrá saber qué ha respondido y cuál es su calificación final. Al inicio, esta persona no tendrá mejor opción que adivinar las respuestas, pero a medida que va repitiendo el proceso, empezará a observar qué respuestas son correctas y cuáles incorrectas, a través del efecto que sus elecciones dan en el resultado final. Tras un número de repeticiones de este proceso, la persona será capaz de responder el examen perfectamente, aún cuando desconoce el significado de sus respuestas.

Las redes neuronales vienen en muchas formas o arquitecturas, cada una con alguna diferencia que la hace mejor para distintas situaciones. Entre estas se destacan para este proyecto las redes neuronales completamente conectadas, las redes neuronales convolucionales y las redes neuronales recurrentes.

3.1.1. Redes neuronales completamente conectadas

Estas son las redes neuronales más básicas y comunes. Suelen funcionar bastante bien para casos sencillos de aprendizaje, pero no son muy eficientes al momento de analizar imágenes o datos secuenciales, comparado a otras arquitecturas.

3.1.2. Redes neuronales convolucionales

Este tipo de redes neuronales son ampliamente usadas en el contexto de análisis de imágenes, ya que tienen la característica de poder analizar cada pixel en su contexto, a diferencia de las redes neuronales completamente conectadas, que consideran cada pixel un dato independiente. Este es el tipo de red neuronal que se suele usar en aplicaciones como reconocimiento facial o generación de imágenes.

3.1.3. Redes neuronales recurrentes

Esta arquitectura fue especialmente diseñada para tratar con datos secuenciales, tales como sonido o video. Lo que las diferencia del resto es la capacidad de retener memoria, lo cual

les permite usar información pasada para tomar decisiones en el futuro. Estas redes neuronales son usadas en herramientas como en análisis de voz, comprensión del lenguaje natural, o en generación de texto, audio y video.

3.2. Sistemas de tracking

Ya explicados los conceptos base de la inteligencia artificial, es hora de hablar de los sistemas que suelen ser usados en música interactiva para hacer seguimiento del intérprete. Roger T. Dean, autor de *The Oxford Handbook of computer music*, determina que existen cinco métodos para hacer seguimiento del movimiento, cada uno con sus ventajas, desventajas y requerimientos.

El primer tipo de sistema de tracking es el denominado *Inside-in*, el cual consiste de sensores ubicados en el cuerpo a medir. Este tipo de sensores tienen la ventaja de que pueden capturar movimientos finos, y no requieren de una posición fija en el escenario. Entre sus desventajas se encuentra la posibilidad de obstruir la movilidad del objeto, además que dificulta la captura de información de la posición del objeto en relación al entorno.

Luego se encuentran los sensores *Inside-out*, los cuales son ubicados en el cuerpo y proveen información externa del objeto. Estos, contrario a los *Inside-in*, facilitan la captura de información del entorno, lo que puede proveer posicionamiento con respecto al escenario. Al ser sensores montados en el objeto, también tienen la desventaja de poder limitar la movilidad del objeto, además de no proveer información del objeto per-se.

Los siguientes sensores son aquellos que van montados externamente al objeto, y se denominan *Outside-in*. Estos sensores captan información del objeto de interés desde una perspectiva externa al objeto. Tienen la ventaja de capturar simultáneamente información del objeto y su entorno, pero los datos capturados son menos finos comparados a los anteriores tipos de sensores. Por otro lado, estos sensores limitan la movilidad del objeto al marco de referencia que pueden capturar, y sufren de oclusión.

Está claro que los sistemas de tracking deben hacer uso de alguno de estos tipos de sensores, pero no se encuentran limitados al uso único de cada tipo. Por esta razón, se clasifica el próximo sistema de tracking como sistema híbrido, y sencillamente consiste en la combinación de los sistemas anteriormente mencionados.

Por último, se menciona como otro sistema de tracking la visión computacional, que si bien no es un tipo de sensor empodera la captura de datos debido a la capacidad de procesamiento y análisis de datos que provee.

Para este proyecto se hará uso del sistema *Outside-in*, en conjunción con visión computacional. Se ha escogido este tipo de sistema debido a la accesibilidad de este tipo de sensores, y ha que requieren preparación previa mínima con respecto al objeto a medir. Por otro lado, debido al interés de explorar el uso de algoritmos de aprendizaje, el sistema de visión computacional será desarrollado como parte de dicha exploración.

3.3 Los esfuerzo-acciones de Laban

Rudolf Von Laban es actualmente aclamado por muchos como uno de los padres de la danza contemporánea. Además de haber sido un gran coreógrafo, sus aportes al análisis del espacio y movimiento abrieron paso a nuevas maneras de comprender la danza y el arte escénico. Laban, movido por la filosofía expresionista del momento, buscó otorgarle un sentido científico a la expresión artística del movimiento, formulando estrategias para analizar, comprender y especificar el movimiento, no solo como arte sino como ciencia. Entre su trabajo sobre el movimiento, Laban postula que todo movimiento se puede describir de acuerdo a su espacialidad (E), temporalidad (T) y peso (P). Cada uno de estos aspectos se pueden combinar, y su combinación describe algún tipo de Esfuerzo-acción.

- **Espacio (E)**

La espacialidad describe el uso del espacio en el movimiento, y puede alternar entre un espacio directo e indirecto. El uso de un espacio directo implica un manejo eficiente del espacio, en el que todo movimiento toma el rumbo más corto para llegar a su objetivo. Por otro lado, el espacio indirecto implica un manejo más amplio del movimiento, en el que se exploran distintas rutas para llegar a un mismo objetivo o posición.

- **Tiempo (T)**

La temporalidad describe tanto la longitud del movimiento como la periodicidad de una sucesión de movimientos. Este aspecto puede alternar entre un tiempo súbito y uno sostenido. El tiempo súbito es corto e impredecible, y se percibe como movimientos erráticos. En cuanto al tiempo sostenido se observan movimientos largos y predecibles.

- **Peso (P)**

El peso es análogo a la energía del movimiento, y puede alternar entre enérgico y ligero. Un peso enérgico refleja movimientos firmes y con fuerza, mientras que un peso ligero se refleja como movimientos libres sin fuerza.

ESFUERZO-ACCIÓN	TIEMPO	PESO	ESPACIO
Flotar	Sostenido	Ligero	Indirecto
Deslizar	Sostenido	Ligero	Directo
Retorcer	Sostenido	Enérgico	Indirecto
Presionar	Sostenido	Enérgico	Directo
Sacudir	Súbito	Ligero	Indirecto
Tocar	Súbito	Ligero	Directo
Atacar	Súbito	Enérgico	Indirecto
Golpear	Súbito	Enérgico	Directo

Siendo que solo hay 3 aspectos para el movimiento, y que su descripción es binaria, solo existen 8 combinaciones posibles de estos estados¹. Según el método de Laban todo movimiento se puede caracterizar por alguna de estas combinaciones, y cada combinación refleja un estado emocional característico.

¹ Si bien Laban agrega otra dimensión o característica al movimiento, denominada Flujo, esta no afecta sustancialmente la concepción de los Esfuerzo-acciones, por lo que no será discutida en este proyecto.

4. OBJETIVOS

Sinestes.IA tiene varios objetivos de carácter técnico, artístico e investigativo, que concluyen en un resultado creativo. Estos objetivos son:

- Generar música en tiempo real a partir de la danza, haciendo uso de la inteligencia artificial como medio para traducir ambas formas de arte.
- Explorar el uso de algoritmos de aprendizaje de máquina como medios de expresión artística.
- Estudiar la relación entre las artes escénicas y la música, y establecer una conexión directa.
- Componer ejemplos musicales como herramienta de entrenamiento para los algoritmos de aprendizaje.
- Desarrollar algoritmos de aprendizaje que logren transformar el movimiento a instrucciones musicales.

5. METODOLOGÍA

Para lograr cumplir el objetivo principal de Sinestes.IA es necesario comprender qué tipo de proyecto conforma y cuál es el método que se suele aplicar. Sinestes.IA es, en gran medida, un proyecto de aprendizaje de máquina, enmarcado en un carácter artístico que lo define y complementa. Por esta razón la metodología se va a ver basada en los requerimientos de estos proyectos, y en la manera de proceder de los mismos.

Para desarrollar un proyecto de aprendizaje de máquina hay que implementar varios aspectos. Estos son:

5.1. Definir el problema de aprendizaje

La definición del problema de aprendizaje se basa en la obtención de una información deseada a partir de datos previamente conocidos. Esto se suele ver reflejado como un problema de clasificación, como por ejemplo, clasificar la imagen de una persona según la identidad de la misma; o regresión, como por ejemplo, la predicción del precio de un apartamento a través de sus características. En el caso de Sinestes.IA, el problema de aprendizaje se define como predecir el sonido que debe acompañar una sucesión de imágenes de alguien bailando. Esta definición da indicios al tipo de arquitecturas que es necesario manejar, debido al uso de imágenes y secuencias como datos previamente obtenidos.

5.2. Obtener bases de datos que reflejen el problema

Para un proyecto de aprendizaje de máquina, la obtención de las bases de datos es el requerimiento más relevante, debido a que es con base en esto que el algoritmo podrá aprender. Para el problema que Sinestes.IA busca explorar, la base de datos debe consistir en ejemplos de la información que el programa recibirá, al igual que el resultado que se espera de dichos datos. Como ejemplo, si se desea entrenar una red neuronal para predecir los precios de apartamentos, la base de datos debe consistir de las características del apartamento como elementos de entrada, al igual que el precio de dicho apartamento como salida.

Como fue señalado en la definición del problema, Sinestes.IA manejará como datos de entrada una sucesión de imágenes, y de salida sonidos. Ambos tipos de datos deben tener correspondencia, lo que significa que por cada imagen debe haber un dato sonoro, y que cada dato sonoro debe reflejar lo que está ocurriendo en la imagen. Para poder obtener estos datos es necesario plantear varios puntos a resolver:

5.2.1. Cómo capturar el video

La captura de video se realizará usando el sensor Kinect, desarrollado para la consola de videojuegos Xbox 360. Se ha escogido esta herramienta debido a que fue diseñado con el objetivo de capturar los movimientos e interactuar con entornos virtuales, para lo que se le incorporó un sensor de profundidad.



5.2.2. Cómo capturar el sonido

El tratamiento del sonido en aplicaciones de inteligencia artificial no es trivial, debido a que un segundo de sonido suele representarse por decenas de miles de valores o samples. Poner a una red neuronal a generar los samples de manera secuencial requeriría de recursos y tiempos por fuera de los estipulados para este proyecto. Aproximaciones de este estilo a la generación musical se pueden observar con Wavenet de DeepMind, red neuronal que

genera resultados excepcionales, a costa de los tiempos de generación, que con altos valores de computación puede tomar hasta 90 minutos generar un segundo de audio (Cahill, P. 2018). Siendo que este proyecto requiere de generación de sonido en tiempo real es necesario buscar otras aproximaciones.

La mejor opción para solucionar el problema anterior es codificar la información musical. Siendo que los sonidos son generados o sintetizados por el computador, es posible representar la información sonora por medio de mensajes que indiquen la activación y desactivación de sonidos. Este tipo de protocolos ya existen, siendo MIDI (Musical Instrument Digital Interface) el más usado. El protocolo MIDI codifica información de altura, velocidad y canal en su forma más básica, pero no incluye información alguna sobre cómo debería sonar o qué instrumento usar. Estos datos son relevantes para este proyecto, ya que se busca que el programa pueda reproducir una gran variedad de sonoridades. Por esta razón se implementó un entorno sonoro en Pure Data, con la posibilidad de representar hasta 6 instrumentos simultáneos, cada uno absolutamente moldeable a través de una lista de parámetros. Cada instrumento puede representar desde un sonido ligero y etéreo hasta sonidos súbitos y estridentes, dependiendo de los parámetros especificados (ver anexo: [entorno sonoro](#)).

5.2.3. Componer los ejemplos sonoros

Siendo que el objetivo del proyecto es generar música en relación a la danza, dicha relación debe estar representada en los datos de entrenamiento. Para lograr esto es necesario parametrizar la música, de tal forma que pueda ser fielmente representada en la danza al momento de capturar la información. Dicha parametrización no es sencilla, debido a que no existe una relación objetiva entre el movimiento y la música, de tal manera que se pueda expresar un solo sonido específico con un solo tipo de movimiento (Gualdrón, 2013).

Aún así, se ha observado que conocimientos en lenguaje corporal suelen facilitar la expresividad de carácter y dinámica en estudiantes de dirección. Trabajos como *Las Ventajas de Aprender Técnica de Dirección por Medio del Análisis de Movimiento Laban* de Deyanira Gualdrón exploran cómo los esfuerzo-acciones de Rudolf Von Laban pueden ejemplificar ciertos tipos de características musicales al ser usados en un ambiente de dirección. Además, la carencia de una relación directa entre la danza y la música abre la posibilidad a la creatividad, permitiendo definir un estilo propio para establecer una conexión entre ambos tipos de arte. Es debido a esto que se usarán los esfuerzo-acciones de Laban como guía para definir un estilo musical en coherencia con el movimiento.

Para establecer el estilo y su conexión con los esfuerzo-acciones, se relacionarán los aspectos del movimiento (espacio, tiempo y peso) con características sonoras y musicales, para así guiar tanto la síntesis como la composición. Esta parametrización es subjetiva, pero fue pensada para mantener cierta generalidad, a través de opiniones tanto de compositores como de bailarines. Como ejemplo, el aspecto del peso en el movimiento suele verse relacionado con la altura y el timbre de los sonidos, de tal forma que sonidos agudos y redondos se ven reflejados por movimientos ligeros, mientras que sonidos estridentes y bajos se reflejan con movimientos enérgicos.

Con base en estos parámetros se compusieron 8 pequeñas piezas, con duraciones de máximo minuto y medio, donde se buscaba reflejar cada esfuerzo-acción haciendo uso del entorno sonoro implementado anteriormente. De esta manera, se puede especificar el tipo de movimiento específico que se desea para cada obra, y así grabar los datos necesarios para el entrenamiento (ver anexo: [composiciones](#)).

5.2.4. Capturar los datos

Debido a que los datos de video y audio deben estar en sincronía, y a que la información sonora es representada de una forma no tradicional, fue necesario diseñar un sistema propio para capturar la información. Esto implicó que era necesario definir limitaciones de velocidad de captura y tamaño de buffer para una grabación fiel y constante.

La velocidad de captura depende de la frecuencia máxima de la información a grabar, y según el teorema de Nyquist la frecuencia de muestreo debe estar por encima del doble de la frecuencia máxima de los datos. La frecuencia para video estándar para reproducción es de 24fps (frames por segundo), siendo 16fps el mínimo considerado para la percepción de movimiento continuo.

En cuanto a los datos sonoros, es necesario establecer una velocidad máxima de acciones para definir la frecuencia de muestreo necesaria. El estudio de Justin London denominado *Cognitive Constraints in Metric Systems* examina la velocidad máxima de impulsos sonoros independientes que el oído humano puede percibir, y establece que una distancia entre impulsos de 50ms (20Hz) es la velocidad teórica máxima de percepción, debido a que este es el límite mínimo de captación de tonos en la escucha humana. A partir de esto, se puede definir la velocidad de muestreo de las acciones sonoras como 40Hz, y usar esta velocidad para la grabación tanto de sonido como de video en este proyecto.

Ahora bien, al momento de capturar los datos, se trabajó con 7 bailarines para realizar múltiples grabaciones de cada composición. Se requirieron múltiples bailarines para tener variedad de estilos de danza y facilidad de coordinación de tiempos con cada bailarín. Debido a las diferencias de tiempo, las piezas más cortas fueron grabadas más veces para compensar su duración. Una vez finalizadas las grabaciones se obtuvo una cantidad de 240 capturas individuales, que si bien son pocas para un proyecto de aprendizaje profundo, son suficientes para obtener resultados interesantes para este proyecto.

5.3. Definir un modelo de aprendizaje base

El diseño del modelo de aprendizaje se basa sustancialmente en el comportamiento de los datos. Siendo que existen tantas arquitecturas y tantas investigaciones realizadas para distintas aplicaciones, es necesario definir algún sistema inicial para evaluar su rendimiento e iterar según los requerimientos del proyecto. Esto significa que si bien es importante tener claro cómo aproximarse al problema, los detalles del modelo se pueden refinar a través de experimentación.

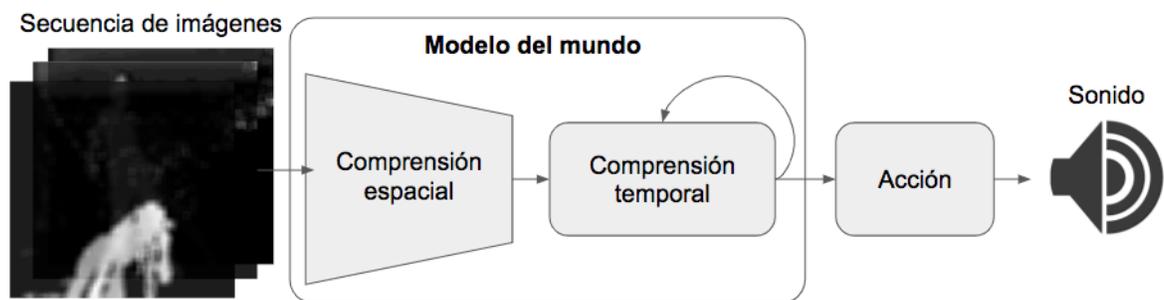
El modelo de aprendizaje de Sinestes.IA debe ser capaz de analizar el baile desde una perspectiva espacial como temporal, debido a que son secuencias de imágenes. Además de esto, debe ser capaz de responder a las imágenes de baile con acciones sonoras a interpretar. Este tipo de problemas es similar al usado en videojuegos, donde de entrada se alimenta con una secuencia de imágenes del videojuego, y el modelo responde con las acciones que debe tomar para continuar jugando. Es por esta razón que el modelo de inteligencia artificial de Sinestes.IA va a estar basado en un estudio denominado *World Models* realizado por David Ha et. al. Este estudio se basa en el concepto de Modelos Mentales de Jay Wright Forrester descrito por él como:

“La imagen del mundo a nuestro alrededor; que llevamos en nuestra cabeza, es solo un modelo. Nadie se imagina en su cabeza todo el planeta, gobierno o país. Solo se tiene una cantidad selecta de conceptos, y relaciones entre ellos, y usamos estos para representar el sistema real”
(J.W. Forrester; 1971, t.d.a)

Esto implica que es necesario separar el modelo en dos partes: el modelo del mundo, y el modelo de acciones. El modelo del mundo debe ser capaz de comprender una secuencia de imágenes, lo que implica que debe consistir de una red neuronal convolucional y una red neuronal recurrente, para así poder lidiar con el comportamiento espacial y temporal del

baile. Es debido a esto que la estructura macro del proyecto consistirá de un modelo de comprensión espacial, un modelo de comprensión temporal y un modelo de acción.

De esta manera, el flujo de información iniciaría con la secuencia de imágenes entrando al modelo de comprensión espacial, para luego transmitir el resultado de dicha comprensión al modelo temporal. Con esto, la inteligencia artificial sacará una conclusión de lo que está ocurriendo en el baile, y esa conclusión la enviará al modelo de acción, donde generará una hipótesis de los sonidos que deberían acompañar el baile.



Flujo y estructura del modelo de aprendizaje

5.3.1. Modelo de comprensión espacial

Como ha sido mencionado anteriormente, el análisis de imágenes suele realizarse con una arquitectura de red neuronal convolucional. Estos modelos suelen recibir la imagen y entrenarse a partir de lo que se desee obtener de ellas. El problema actual es que se desconoce el resultado que se espera obtener de esta sección. Este tipo de problemas se conoce como aprendizaje no supervisado, ya que son problemas en los que el resultado deseado no está previamente definido.

El sistema que se usa en el estudio de *World Models*, y que también será usado en este proyecto, se denomina como *Autoencoder*. Los autoencoders tienen la labor de codificar y decodificar la información entrante, buscando generar en la salida la misma información que en la entrada. De esta manera, el entrenamiento consiste en evaluar la similitud de los datos entrantes con los generados por el algoritmo.

Ahora bien, lo que hace estos modelos especiales, es que en el proceso obligan al modelo a tener que representar los datos en un espacio menor al de la información de entrada.

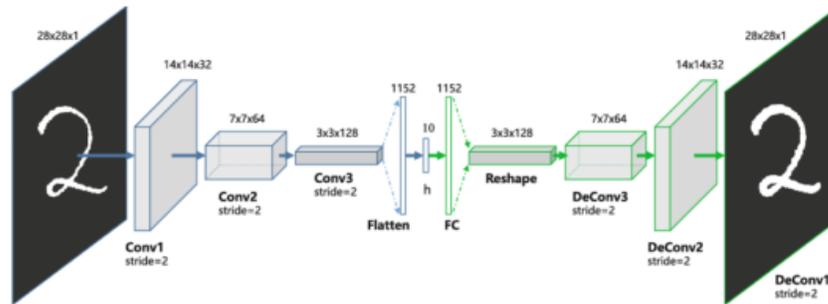
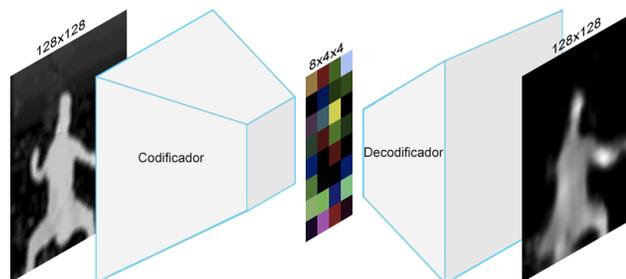


Imagen obtenida de <https://www.edureka.co/blog/autoencoders-tutorial/>. Acá la imagen del número 2 (izquierda) se comprime en la capa h (centro), y de esta capa se trata de reconstruir la imagen original (derecha).

Como analogía, consideremos la labor de resumir toda una película en un solo párrafo de 500 palabras, de tal forma que luego se pueda usar ese resumen para regrabar toda la película lo más fiel a la original. Ahora consideremos que no solo debe reconstruir una película, sino una gran variedad de películas usando la misma estructura. Si esta tarea se realiza, obliga que cada palabra sea crucial para la reconstrucción de la (o las) películas, y ninguna palabra será malgastada.

Este proceso logra reducir la información visual de las imágenes a características relevantes de las mismas, al igual que comprime la cantidad de datos que entran al sistema, facilitando el aprendizaje para las próximas capas.

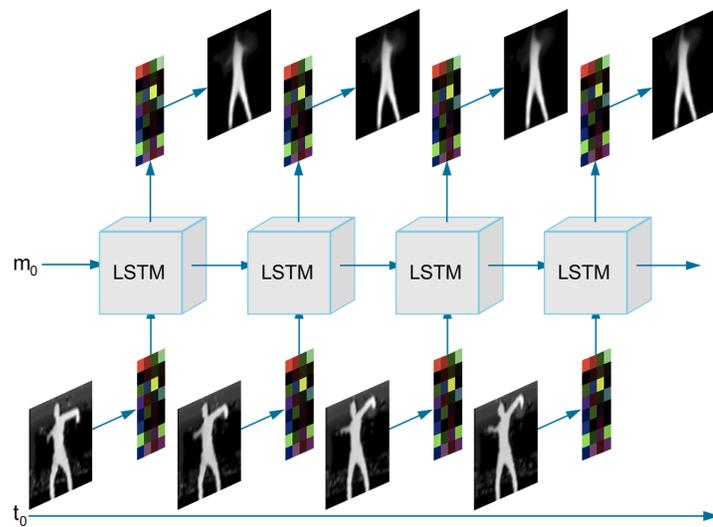


Demostración de la primera etapa de Sinestes.IA. En este caso la imagen original es comprimida a 128 datos (8x4x4).

5.3.2. Modelo de comprensión temporal

Obtenida la versión comprimida de las imágenes, es posible usar esta representación para analizar la evolución del baile en el tiempo. Para lograr esto, el modelo debe tener algún

tipo de memoria para recordar la información pasada. Es debido a esto que se utilizará una red neuronal recurrente, que además de recibir la imagen comprimida, recibe información de los procesos anteriores. En este caso específico, se utilizará la arquitectura LSTM (Long Short Term Memory), que ha demostrado ser buena para considerar información tanto cercana como lejana temporalmente. Para generar la comprensión temporal se entrena a la red neuronal a predecir lo que ocurrirá en el futuro según la información pasada. De esta manera, si el algoritmo predice correctamente y de manera consistente, implica que comprende el comportamiento temporal de los datos.



Demostración de la segunda etapa de Sinestes.IA. Se puede observar cómo se procesa una secuencia de imágenes individualmente (flujo vertical), mientras que cada capa comparte información de los procesos anteriores (flujo horizontal).

De esta capa nos interesa más la salida intermedia de la red neuronal, ya que es la que representa la comprensión temporal del baile. Una vez obtenido esto, es posible usar esta información como entrada para el modelo de acción.

5.3.3. Modelo de acción

Este modelo se encarga de recibir el análisis del baile, y a partir de esto proponer la acción sonora que lo acompaña. Debido a que las capas anteriores realizan una gran cantidad de trabajo, esta capa puede ser considerablemente más sencilla. Por esta razón, el modelo de

acción será una red neuronal completamente conectada, la arquitectura más básica². En este caso, la red neuronal recibe los datos de las salidas intermedias del modelo anterior, y como resultado busca predecir las acciones sonoras que acompañan el baile. Es en este punto en el que se usan los datos sonoros grabados en la base de datos para establecer la relación entre la danza y la música.

5.4. Entrenar el modelo de aprendizaje

Para entrenar los modelos anteriores se usaron los datos capturados necesarios en cada etapa, empezando por el modelo de comprensión espacial, para luego entrenar el modelo de comprensión temporal y por último entrenar el modelo de acción. Esto es debido a que cada modelo consecuente se entrena a partir de los resultados del anterior, por lo que un entrenamiento continuo de todas las etapas no era productivo.

Aunque se tiene una arquitectura base del modelo, los detalles se definen en el momento de entrenamiento a través de experimentación constante. Se busca elegir los parámetros que generen el mejor resultado en cada etapa, por lo que el proceso consiste en entrenar varias veces cada modelo con parámetros distintos, y observar el desempeño en cada labor específica. Una vez obtenido el mejor resultado, se guardan estos parámetros para continuar con la siguiente etapa.

6. RESULTADOS

Debido a que Sinestes.IA consiste en múltiples modelos de aprendizaje individuales, se analizarán los resultados de estas partes al igual que el resultado final.

6.1. Modelo de comprensión espacial

Luego de 25 experimentos en esta sección específica, se lograron comprimir todas las imágenes del baile, cuyo tamaño es de 128x128 píxeles a imágenes de 8x4, con cuatro canales de color (RGBA). Esto representa una compresión de 16384 valores a 128 totales. Con esta compresión se logran reconstruir todas las posiciones del baile representadas en los datos, al igual que posiciones no conocidas por el modelo, en un nivel significativo.

² Se ha escogido un modelo sencillo en coherencia con el estudio en el que está basado este proyecto. Es probable que esta opción no genere los mejores resultados, debido a que las acciones sonoras también son datos secuenciales. Como los modelos secuenciales son los más complicados de entrenar, no se optará por usar otra red neuronal recurrente en esta capa.

Uno de los problemas principales que se encontró en esta etapa fue la reconstrucción de los brazos. Debido a que los brazos representaban una porción muy pequeña de la silueta, y que variaban significativamente en cada posición en contraste con el resto de la figura, el modelo presentó dificultades para representar esta información de manera eficiente. Es por esto que en la mayoría de los casos la figura no tiene brazos o se ve acompañado de manchas borrosas en donde deberían mostrarse estos.

6.2. Modelo de comprensión temporal

Una manera de observar el rendimiento de un modelo secuencial predictivo como este, es ponerlo a generar secuencias de longitud indefinida. A esto se le suele llamar generar sueños, y permite observar qué comportamientos temporales captura el modelo tanto a corto plazo como a largo plazo.

Analizando el método anterior, se puede observar que el modelo temporal sabe cómo se desarrolla el baile a corto plazo a partir de posiciones iniciales aleatorias. Pero, al momento de generar secuencias largas las predicciones tienden a estancarse en una posición estándar. Esto demuestra que el modelo toma en consideración una cantidad muy reducida de información en el pasado, mientras que ignora el desarrollo del baile ocurrido mucho antes.

6.3. Modelo de acción

Una vez entrenada esta última etapa, se observó su rendimiento haciendo uso tanto de los sueños generados en la etapa anterior como de la captura de imagen en tiempo real. Los sonidos generados demuestran una gran habilidad por parte del algoritmo de hacer uso de las sonoridades encontradas dentro de la base de datos, y relacionarlas con posiciones o gestos específicos del movimiento. Además, se puede observar que el modelo de inteligencia artificial relacionó satisfactoriamente el silencio con el reposo o quietud.

Una de las limitaciones encontradas al final del proceso, fue la incapacidad de combinar instrumentos que no habían sido combinados previamente en los datos de entrenamiento. Esto significa que el modelo no se encontró en capacidad de reproducir sonidos percutidos con sonidos armónicos, debido a que esta conjunción no se presenta en los datos. Una manera de poder solucionar esto es transformando el modelo de acción en seis modelos independientes, cada uno encargado de un instrumento específico. Esta solución tiene el inconveniente de eliminar toda relación existente entre los instrumentos, lo que implica que el concepto de armonías o texturas sería ignorado por el algoritmo final.

6.4. Resultado final

Para estudiar los resultados del sistema, se realizó una exploración exhaustiva de la relación movimiento-música, buscando comprender la manera en que el algoritmo interpreta los movimientos y los representaba en sonidos. Esta exploración reveló cierta relación entre uso de instrumentos y sonoridades con ubicaciones espaciales de la figura, de tal forma que dependiendo del lugar en el que uno se encuentra parado se obtendrían distintos timbres y comportamientos temporales del sonido. Se observó, además, que los sonidos percusivos suelen estar relacionados con el cambio abrupto de ubicación, lo cual demuestra cierta coherencia con el carácter súbito y enérgico de *Golpear*. Adicional a esto, los movimientos leves dentro de una misma posición suelen estar acompañados de sonidos agudos y largos, reminiscentes de los esfuerzos con temporalidades sostenidas y pesos ligeros.

Ahora bien, al momento de observar los resultados, se evidenció la influencia de las limitaciones obtenidas en cada etapa del modelo. La falta de detalle y la reconstrucción parcial de las posiciones de baile, por parte del modelo de comprensión visual, limitan la variedad de movimientos que pueden afectar el sonido generado. Por otro lado, la inhabilidad del modelo de comprensión temporal para recordar secuencias a largo plazo afectan directamente la temporalidad de la música, ya que ésta tampoco tendrá la capacidad de encadenar secuencias melódicas o armónicas extensas.

Si bien los resultados no demuestran una capacidad compositiva compleja por parte del sistema, su utilidad como instrumento de exploración musical es clara. La aproximación sin prejuicios por parte de terceros reveló comportamientos sonoros interesantes al momento de investigar movimientos o posiciones no representados en los datos, tales como tapar la cámara con la mano, o la exploración constante en los límites de percepción del sensor, causaron que el sistema reaccionara con hipótesis sonoras por fuera de lo que se entrenó.

7. CONCLUSIONES

Si bien el experimento realizado a través de Sinestes.IA no demostró capacidades musicales complejas a nivel de armonía y desarrollo en el tiempo, se evidenciaron las posibilidades que la inteligencia artificial ofrece para generar herramientas que puedan transformar medios de expresión en otros. Con los suficientes recursos, tiempo y conocimiento, es posible alcanzar mejores resultados a través de la obtención de más datos y la construcción de modelos más complejos.

Algo que cabe resaltar de este experimento es la distribución de control sobre el resultado final. Sinestes.IA es un proyecto que combina el arte interactivo con el arte algorítmico, lo que implica que hay 3 agentes independientes con cierto nivel de control sobre la pieza: El diseñador del algoritmo, el algoritmo y el usuario o intérprete. Este fue un proyecto concebido con la idea de entregar el mayor grado de libertad al algoritmo, y se refleja claramente en los resultados, donde tanto el intérprete como el diseñador se aproximan al algoritmo con la intención de entenderlo, mas no de controlarlo. Esto es debido a como fue conceptualizado y a que este no se reajusta al momento de ser usado. Una aproximación interesante al problema del control es hacer uso de sistemas de aprendizaje constante, también denominados aprendizaje reforzado. David Rokeby señala a estos sistemas como la posición extrema a la entrega de control, ya que pueden ser diseñados para aprender constantemente del usuario y así aportar mayor dirección al intérprete sobre el resultado.

Por último, Sinestes.IA es tan solo una primera aproximación a la idea de construir herramientas artísticas que se adapten al usuario. Alcanzar este objetivo, como se menciona anteriormente, requiere una gran cantidad de experimentación, y este proyecto ha logrado revelar tanto las complicaciones como las posibilidades de la inteligencia artificial como herramienta de apoyo creativo. De este proyecto surgirán nuevas aproximaciones al problema, y servirá como referencia inicial a los procesos y resultados que se pueden obtener y superar para proyectos posteriores.

8. BIBLIOGRAFÍA

- Ternaux, Jean-Pierre. (2003). *Synesthesia: A Multimodal Combination of Senses*. *Leonardo*. 36. 321-322. 10.1162/002409403322258790.
- Di Marco, P., & Spadaccini, A. (2015). *Wassily Kandinsky y la Sinestesia que lo condujo a la Música Visual*. 2-27. Retrieved April 6, 2019, from <https://alessiaspadaccini.com/wp-content/uploads/2015/05/Wassily-Kandinsky-y-la-Música-Visual.pdf>.
- Dean, R. T. (2011). *The Oxford handbook of computer music*. Oxford: Oxford University Press.
- Ng, A. (n.d.). *Machine Learning*. Retrieved 2017, from <https://www.coursera.org/learn/machine-learning>
- Ng, A. (n.d.). *Deep Learning*. Retrieved January 14, 2018, from <https://www.coursera.org/specializations/deep-learning>

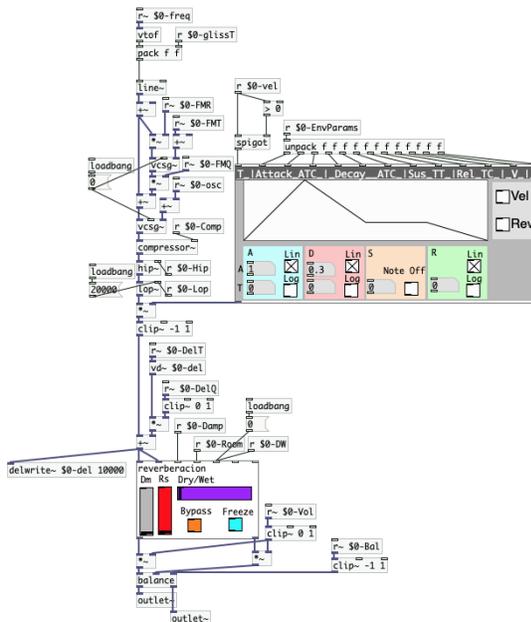
- Dörr, E., & Lantz, L. (2003). *Rudolf von Laban: The "Founding Father" of Expressionist Dance*. *Dance Chronicle*, 26(1), 1-29. Retrieved from <http://www.jstor.org.ezproxy.javeriana.edu.co:2048/stable/1568111>
- Gualdrón Ordoñez, D. (2013). *Las Ventajas De Aprender Técnica De Dirección Por Medio Del Análisis De Movimiento Laban: Una comparación entre la terminología tradicional y Laban*. University of Washington.
- Ha, D., & Schmidhuber, J. (2018). *World Models*. *ArXiv*, 1803(10122), 4th ser., 1-10. doi:10.5281/zenodo.1207631
- Forrester, J.W. (1971). *Counterintuitive Behavior of Social Systems*.
- Rokeby, D. (1996, July 3). *Transforming Mirrors: Interaction in the context of Art*. Retrieved May 12, 2019, from http://www.sunypress.edu/pdf/David_Rokeby_Transforming_Mirrors.pdf

ANEXOS:

ENTORNO SONORO

El ambiente sonoro de Sinestes.IA debe ser diseñado con el propósito de ofrecer la mayor cantidad de variedad tímbrica y textural en un número reducido de parámetros. De esta manera, el modelo de inteligencia artificial no se verá limitado por las posibilidades, sino por sus capacidades de aprendizaje. Por otro lado, el entorno debe poder ser completamente representado en todo momento por los parámetros que el programa va a manipular, lo que significa que el algoritmo tendrá control absoluto de lo que está ocurriendo dentro del ambiente sonoro.

Para lograr representar este entorno, los parámetros no solo deben capturar la información de intensidad y altura de cada voz que se piense manejar, sino que además deben capturar las características tímbricas y temporales de cada sonido. Esto lo diferencia del protocolo MIDI, ya que esta solo encapsula las instrucciones de interpretación que debe manejar el instrumento escogido. Para lograr esto, se desarrolló un instrumento como un componente aislado del entorno, el cual incluye todos los requerimientos anteriores. De esta forma, se puede construir el ambiente sonoro a partir de múltiples instrumentos, cuya construcción es idéntica, reduciendo la complejidad y maximizando la variedad.



Los parámetros del instrumento deben contener información que indique la intensidad, altura, envolvente, timbre, volumen, paneo y demás agregados que afectan el sonido. De esta manera, el instrumento recibe como entrada en el siguiente orden:

1. **VELOCITY:** Similar al protocolo MIDI, este es un valor de 0 a 127 que representa la intensidad interpretativa del sonido.
2. **AMPLITUD DE ATAQUE:** Este es un valor de 0 a 1 que representa el porcentaje de intensidad al que llega el instrumento durante el ataque de la envolvente.
3. **TIEMPO DE ATAQUE:** Indica la duración de la etapa de ataque en milisegundos.
4. **CURVA DE ATAQUE:** Es un valor binario (0,1) que indica si la curva es lineal o logarítmica respectivamente.
5. **AMPLITUD DE DECAY:** Representa el porcentaje de intensidad en la etapa de decay. Esta será la misma amplitud usada para el sustain.
6. **TIEMPO DE DECAY:** Indica la duración de la etapa de decay en milisegundos.
7. **CURVA DE DECAY:** Valor binario que indica el tipo de curva de la etapa de decay.
8. **TIEMPO DE SUSTAIN:** Indica la duración de la etapa de sustain.
9. **NOTE-OFF TRIGGER:** Indica si el tiempo de sustain dependerá de un mensaje de note-off, que en este caso sería un Velocity de cero. De ser así, el tiempo de sustain es ignorado, para depender del velocity.
10. **TIEMPO DE RELEASE:** Indica el tiempo que se demora el sonido en silenciarse desde que termina la etapa de sustain.
11. **CURVA DE RELEASE:** Indica si la curva de la etapa de release es lineal o logarítmica.
12. **VELOCITY-DEPENDENT:** Valor binario que señala si la intensidad de la envolvente depende del velocity, o si se usará el velocity como trigger únicamente. Esta es más una herramienta para el intérprete que para el modelo.
13. **REVERSE:** Valor binario que indica si se desea reproducir la envolvente en retrógrado. Esto permite generar sonidos reproducidos en reversa.
14. **FRECUENCIA:** Representa la altura del sonido en términos de frecuencia. Puede variar entre 20Hz y 20kHz.
15. **TIPO DE OSCILADOR:** es un número del 0 al 4, que permite variar de manera continua entre múltiples tipos de forma de onda. 0 es sinusoidal, 1 es triangular, 2 es diente de sierra, 3 es cuadrada y 4 es ruido blanco. Si el valor es intermedio a los previamente indicados, entonces es una mezcla de las dos formas de onda más

cercanas. Por ejemplo 0.5 es la combinación de sinusoidal y triangular en magnitudes iguales.

16. **TIEMPO DE GLISSANDO:** Representa el tiempo que se demora el instrumento en pasar de una altura a otra. Si es 0 significa que no hay glissando.
17. **TASA DE FM:** Valor que representa la relación entre la frecuencia portadora y moduladora en la etapa de FM.
18. **PROFUNDIDAD DE FM:** Valor que representa la cantidad de modulación en la etapa de FM. 0 representa sin FM.
19. **TIPO DE OSCILADOR FM:** Este valor define la forma de onda a usar como moduladora del FM, y se comporta como el parámetro “Tipo de oscilador”.
20. **COMPRESOR:** Valor de 0 a 5 que representa una cantidad específica de compresión. Esto se logra haciendo uso del objeto *compressor* de Pure Data, que lamentablemente no especifica lo que representa exactamente cada valor.
21. **FILTRO PASA ALTOS:** Representa la frecuencia de corte del filtro pasa altos. 0 significa sin filtro.
22. **FILTRO PASA BAJOS:** Representa la frecuencia de corte del filtro pasa bajos. 20000 es equivalente a sin filtro.
23. **TIEMPO DE DELAY:** El tiempo en milisegundos entre cada repetición del delay.
24. **CANTIDAD DE DELAY:** Valor que representa el porcentaje de nivel de salida del delay en comparación al nivel de entrada. 100% representa mismo nivel de entrada.
25. **DAMPING DE REVERBERACIÓN:** Valor que representa el porcentaje de amortiguación de sonido en la etapa de reverberación.
26. **TAMAÑO DE CUARTO:** Valor de 0 a 1 que representa el tamaño de cuarto en la reverberación.
27. **DRY/WET:** Relación entre sonido reverberado y sonido limpio.
28. **VOLUMEN:** Valor de 0 a 1 que representa el nivel de volumen a la salida del instrumento.
29. **PANEO:** Valor de -1 a 1 que representa la posición estéreo del instrumento.

Con esto podemos representar todo el estado de un instrumento en 29 valores individuales. Es importante destacar que esta definición fue realizada previamente a las composiciones, y algunos de estos parámetros no fueron manipulados en las bases de datos. Esto significa que se podría definir un número de parámetros menor a este, pero con la intención de mantener una consistencia con la idea de ofrecer la mayor variedad posible se mantendrán dentro de la definición del instrumento.

Habiendo desarrollado, y encapsulado toda la idea de un instrumento bajo estos parámetros, se puede especificar el tipo de sonido a producir asignando valores a cada parámetro. Como ejemplo:

Bombo

```

kick
127 1 25 1 0.2 150 1 0 0 100 1 0 0 60 0 0 0 0 0 0 20000 0
0 0.5 0.8 0.1 1 0
pd instrument

```

Este ejemplo demuestra qué parámetros usar para generar sonidos percusivos bajos como el bombo de una batería.

Flauta

```

flauta
127 1 120 1 0.8 350 0 2000 0 2000 0 0 0 800 0.6 0 0.005 1 0
1 0 800 0 0 0.5 0.8 0.3 0.4 0
pd instrument

```

Acá se puede ver la versatilidad del instrumento, ya que sin haber modificado la arquitectura se pudo obtener un sonido contrastante al bombo modificando los parámetros. En este caso, se hace uso del FM con una tasa tan baja que actúa como un LFO, generando un efecto de vibrato en la nota.

Sonido dinámico

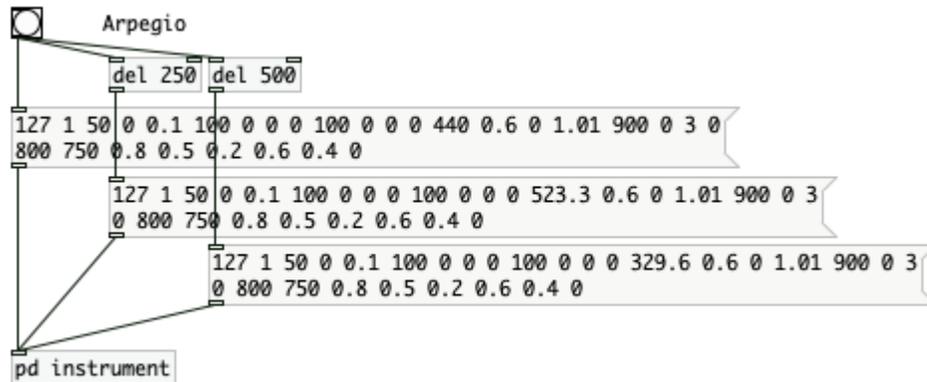
```

Estridente
del 20
127 1 1200 0 0.2 1500 0 0 0 1000 0 0 0 400 1 0 0.3 100 1 3
0 1000 0 0 0.5 0.8 0.8 0.14 0
127 1 1200 0 0.2 1500 0 0 0 1000 0 0 0 800 1 500 0.3 100 1
3 0 1000 0 0 0.5 0.8 0.8 0.14 0
pd instrument

```

Se pueden encadenar mensajes para generar sonoridades más interesantes. En este caso, hacemos uso del glissando para cambiar de altura de manera gradual, y haciendo uso de la síntesis FM, del compresor y de los filtros se obtiene un timbre interesante.

Arpeggio



El encadenamiento de mensajes también permite generar secuencias, lo cual abre la posibilidad a construir melodías en un solo instrumento. Acá se hace uso del delay para mantener el arpeggio por más tiempo.

La única limitante del instrumento es que es una sola fuente sonora, lo que significa que por diseño es monofónico. Para poder aumentar la textura del entorno sonoro, es por lo tanto necesario usar más de un instrumento. Priorizando la variedad sonora, se ha decidido manejar 6 instrumentos independientes, lo que puede ser usado como 4 voces para armonía y melodía, y dos para percusión o efectos. De esta forma se diseñó el entorno sonoro, el cual es usado tanto por la inteligencia artificial como para la construcción de las composiciones.

COMPOSICIONES

La intención de las composiciones es generar ejemplos de cómo la música se relaciona con la danza, al igual que demostrar cómo se manejan los parámetros del entorno sonoro para que el algoritmo pueda luego manipularlos. Estos dos requerimientos implican que las composiciones deben representar el movimiento, y deben hacer uso único del entorno sonoro para desarrollarse.

Como es mencionado en el texto principal, se usarán los esfuerzo-acciones de Laban como medio para parametrizar la música. Esta parametrización implica relacionar características del sonido a los 3 aspectos que definen cada esfuerzo-acción. Dicha parametrización es puramente subjetiva, y fue desarrollada para este proyecto con esa intención. La aproximación a este problema consistió en observar cada esfuerzo-acción, y con base a cada uno imaginarse qué tipos de sonidos acompañarían mejor el movimiento. Una vez establecido esto, se observaban qué características musicales estaban más relacionadas con cada dimensión del movimiento y según esto establecer la relación danza-música.

Esta tabla demuestra las conclusiones a las cuales se llegó de acuerdo al proceso anterior:

ESFUERZO- ACCIÓN	TIEMPO	PESO	ESPACIO	ALTURA	TIMBRE	TEXTURA	GLISSANDO	VIBRATO	LONGITUD
Flotar	Sostenido	Ligero	Indirecto	Aguda	Redondo	Delgada	Con	Con	Larga
Deslizar	Sostenido	Ligero	Directo	Aguda	Redondo	Delgada	Con	Con	Largo
Retorcer	Sostenido	Enérgico	Indirecto	Grave	Estridente	Densa	Sin	Con	Larga
Presionar	Sostenido	Enérgico	Directo	Grave	Estridente	Delgada	Sin	Sin	Larga
Sacudir	Súbito	Ligero	Indirecto	Aguda	Redondo	Densa	Sin	Con	Corta
Tocar	Súbito	Ligero	Directo	Aguda	Redondo	Densa	Sin	Con	Corta
Atacar	Súbito	Enérgico	Indirecto	Grave	Estridente	Delgada	Con	Sin	Corta
Golpear	Súbito	Enérgico	Directo	Grave	Estridente	Densa	Sin	Con	Corta

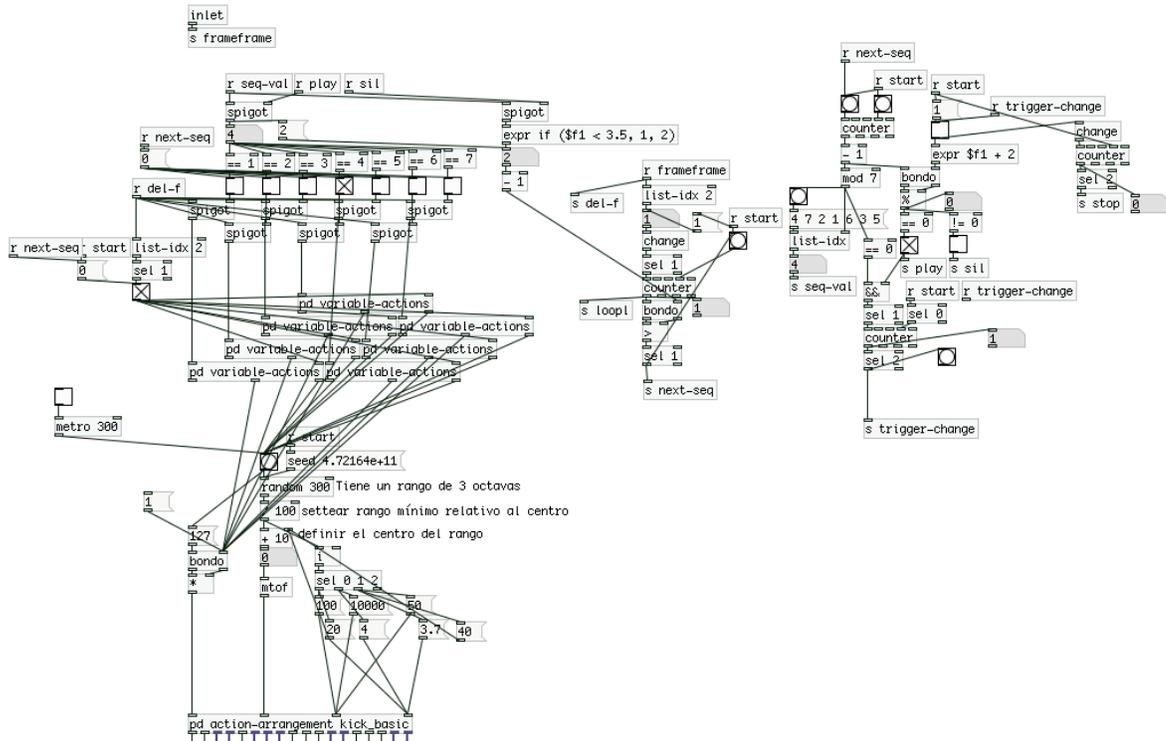
Observando la tabla anterior, se llega a la conclusión que la altura y el timbre están correlacionados al peso del movimiento, mientras que la duración del sonido está relacionado con el tiempo del movimiento. Las otras características no representan una relación directa, pero también fueron consideradas para el momento de componer.

Esta parametrización funciona como una guía básica para construir los ejemplos musicales, mas no representan una estructura rígida, ni completa de las composiciones. Esto significa que se puede reforzar el carácter de cada esfuerzo-acción a partir de otras características sonoras y musicales más específicas, y que se pueden sacrificar ciertas relaciones en pro de la musicalidad de las piezas.

Las implicaciones del tiempo súbito

Además de las consideraciones anteriores, es necesario que la relación demostrada en la danza y la música mantenga una correspondencia directa. Para lograr esto, cada acción musical debe ser representada por algún movimiento corporal. Esto representa un reto característico al momento de hablar del tiempo súbito, ya que este abarca tanto la temporalidad de cada movimiento, como de toda la secuencia de movimientos. Un tiempo súbito implica movimientos erráticos, que aparecen de manera impredecible en la danza. Esto implica que para mantener una coherencia directa, la música debe estar construida de manera que sea impredecible y errática.

Para lograr dicho cometido, se diseñó un algoritmo basado en el concepto de Serialismo para evitar dar mucha relevancia a algún patrón rítmico durante las piezas.



El algoritmo se basa en una serie o secuencia de números base del 1 al 7. A cada número se le define un patrón rítmico diferente al resto, de tal forma que cuando el algoritmo decida reproducir un número específico, se ejecutará dicha secuencia rítmica. Con la intención de complicar la recordación de la pieza, el algoritmo alterna entre ejecutar el ritmo o agregar silencio a la pieza, cuya duración también depende del valor de la serie que se esté ejecutando en el momento. La pieza termina una vez se hayan ejecutado todos los bloques rítmicos. Por último, las alturas de las notas son escogidas a través de un algoritmo pseudoaleatorio, el cual recibe como secuencia semilla la serie de la pieza. Esto, si bien no implica que las alturas no se repetirán, representará un nivel alto de incertidumbre.

Otro reto con las piezas de característica súbita es la recordación por parte del bailarín. Considerando que estas composiciones buscan impredecibilidad, el bailarín tiene muy pocas herramientas para recordar toda la obra y ejecutarla. Por esta razón, se tuvo que reducir el nivel de incertidumbre en las composiciones, asegurándose que las piezas tuvieran una métrica estable, y que los bloques rítmicos fueran de una duración máxima de dos compases. Además, para facilitar la interpretación de cada obra al momento de grabar, la pieza se reproducía dos veces de manera instantánea: una pista que era la que los bailarines debían interpretar al tiempo, y la misma pista reproducida con dos compases de

anticipación. Esto generaba un efecto de pregunta-respuesta, en el que el bailarín primero escuchaba la secuencia rítmica para inmediatamente después ejecutarla encima de la pista real.

Descripción de cada composición:

Flotar

Sostenido, Ligero, Indirecto.

Este esfuerzo-acción fue el primero trabajado, y requirió de varias iteraciones para llegar a esta composición, ya fuera por un mal manejo de la parametrización, o por resultados de longitud inaceptables para la captura de los datos. Esta composición consta de 3 voces sonando homorrítmicamente. La estructura de la pieza busca ser predecible, a través de repeticiones de frases, y de un ritmo base constante. Es a partir de esto, al igual que el uso de ataques largos, que se representa la temporalidad sostenida del movimiento. Por otro lado, el uso de notas agudas y timbres redondos permite ejemplificar el peso ligero que define este esfuerzo-acción. Por último, el vibrato, y el uso exagerado de reverberación, reflejan la espacialidad indirecta de la obra.

Deslizar

Sostenido, Ligero, Directo.

En esta composición se hace uso exagerado del glissando para representar el movimiento de deslizamiento. Por otro lado, al igual que con *Flotar*, el carácter sostenido se representa con notas largas, dirección clara y una secuencia rítmica constante. Debido a que su peso es ligero, también se hace uso de notas agudas durante toda la pieza. La característica directa fue representada a partir de la ausencia de vibrato y de reverberación. En esta pieza se hizo uso de delays, que ofusca el carácter directo. Aún así, fue utilizado debido a preferencias estéticas.

Retorcer

Sostenido, Enérgico, Indirecto.

Esta composición demuestra el carácter pesado y enérgico a través de su atmósfera oscura, timbres estridentes y alturas bajas. En esta, a diferencia del resto, entra el acompañamiento primero, para luego ser desarrollada con una melodía. El uso de glissandos exagerados, delays y reverberación ejemplifican el carácter indirecto de la obra, mientras que la

duración larga de las notas, sus entradas graduales y la estructura predecible de la pieza caracterizan la temporalidad sostenida.

Presionar

Sostenido, Enérgico, Directo.

El carácter enérgico de este esfuerzo-acción diferencia esta composición de las anteriores a través del uso constante de sonidos bajos y ricos armónicamente. Al igual que la anterior composición, se hace uso de más de una voz para generar armonías, aunque estas se mantienen en unísono. La característica de espacialidad directa se ve mucho mejor refleja en esta pieza, debido a la ausencia de reverberaciones y delays. Por último, el carácter de temporalidad se ve altamente demostrado aquí con el manejo de envolventes largas, haciendo que los sonidos entren y salgan de manera gradual.

Tocar

Súbito, Ligero, Directo.

Esta pieza demuestra el uso del sistema súbito desarrollado para las piezas con este carácter temporal. La temporalidad se ve además reflejada por la duración de las notas, que es bastante corta. Por otro lado, alturas agudas y timbres redondos demuestran el peso del esfuerzo acción, mientras que la ausencia de efectos y de glissando representan la espacialidad. Aquí se puede escuchar el algoritmo original para la generación de estas piezas, que fue diseñado para generar obras de una longitud cercana al minuto. Esta longitud fue luego recortada debido a que era poco práctico al momento de grabar la interpretación del bailarín. Para nivelar esta diferencia de tiempos, se grabaron más veces las piezas súbitas con cada bailarín, comparado a las piezas sostenidas.

Atacar

Súbito, Enérgico, Indirecto.

El esfuerzo-acción “*Atacar*” suele verse ejemplificado a través de los movimientos rápidos realizados para cortar maleza con un machete. Es por esta razón, que esta busca reflejar el sonido del viento al verse perturbado por tales movimientos. Esto implicó un manejo de ruido en la obra, el cual fue filtrado y reverberado, para representar el peso enérgico y la espacialidad indirecta.

Golpear

Súbito, Enérgico, Directo.

Las características de *Golpear* pueden ser fácilmente reflejadas a través de sonidos percutidos, principalmente sonidos de tambores bajos como el bombo. Es debido a esto que se diseñó esta obra a partir de dichos sonidos. Se agregó, además, el sonido de hi-hat, con la intención de generar mayor variedad sonora a este esfuerzo-acción.

Composiciones anteriores

Antes de desarrollar todas las obras, se realizó una prueba del concepto con dos esfuerzo-acciones opuestas: *Flotar* y *Golpear*. Estas composiciones fueron realizadas con el objetivo de realizar las primeras pruebas del sistema de aprendizaje, y diagnosticar la dificultad del proyecto, al igual que el rendimiento que cada etapa. Estas piezas son claramente distintas a las anteriores, debido a que fueron desarrolladas sin una parametrización totalmente definida, la duración no era apta para captura y entrenamiento, y se desconocía la relevancia del carácter súbito en la relación danza-música.