

MuRET v1.0

Resumen de la aplicación

La digitalización de partituras, y otros documentos relacionados con la música, puede aportar varias ventajas, como un acceso más fácil para investigadores, profesionales de la música, musicólogos y el público en general, así como la conservación del patrimonio musical. Las partituras digitalizadas pueden encontrarse, reproducirse, analizarse y editarse utilizando herramientas informáticas especializadas, lo que permite nuevos descubrimientos e interpretaciones sobre la historia y la cultura musical. Para poder realizar este tipo de procesos, no obstante, se debe disponer del contenido musical que las imágenes describen en formato de partitura digital.

La aplicación online MuRET (Music Recognition, Encoding, and Transcription) está enfocada a solventar ese problema: desde una o varias imágenes que constituyen una composición musical, la herramienta permite obtener de manera interactiva, asistida por técnicas de inteligencia artificial (IA), la transcripción de la obra en una partitura digital.

Frente a otras herramientas de reconocimiento óptico de música (OMR - optical music recognition), MuRET permite ser adaptada fácilmente a través del etiquetado manual de obras para trabajar con distintos tipos de notación musical, soportes y disposiciones o composiciones gráficas.

Para obtener el resultado final, se deben realizar una serie de operaciones que describimos a continuación.

Los usuarios deben ser registrados por un administrador del sistema para trabajar en la herramienta. No se ofrece ningún proceso de autorregistro. Todas las obras se organizan en espacios de trabajo, colecciones y subcolecciones, cuyo acceso concede el administrador.

Un documento es la entidad central de un proyecto de transcripción. Tras cargarse como archivos de imagen individuales o dentro de un documento PDF, las imágenes de un documento que se va a transcribir se agrupan, como mínimo, en una sección predeterminada. Esto resulta útil cuando se trata de obras compuestas, como misas y óperas. Las imágenes y las secciones pueden borrarse, editarse y reordenarse. Las imágenes que contienen

portadas, o contenido no musical, pueden ocultarse para posteriores procesos de reconocimiento automático. Los usuarios pueden asignar a cada obra metadatos como el tipo de notación, el tipo de manuscrito (impreso o manuscrito), el compositor y el impresor.

Con el fin de ayudar al usuario en las tareas diarias de trabajo, se pueden marcar las etapas de trabajo en curso y de anotaciones de reconocimiento de imágenes hasta la transcripción final.

Una vez organizadas las imágenes, el primer paso para transcribir una obra, conocido como análisis del documento, consiste en segmentar cada imagen en componentes separados, donde se identifican una serie de regiones de distintos tipos como pentagramas y letras. Normalmente, cada imagen contiene una sola página, pero también es habitual contar con imágenes digitalizadas de libros en los que esas imágenes contienen varias páginas. Este proceso puede realizarse manualmente, dibujando cuadros delimitadores sobre la imagen y asignando un tipo de región a cada cuadro dibujado, o utilizando un clasificador automático basado en IA que identifique los distintos segmentos de la imagen. En MuRET, cuando una operación puede realizarse automáticamente, el usuario puede seleccionar y aplicar un clasificador, y corregir la salida si es necesario. Los clasificadores se ejecutan actualmente en el navegador del usuario.

La mayoría de las fuentes existentes susceptibles de ser procesadas son polifónicas, formadas por varias voces, instrumentos o partes. Existe una gran variedad de arreglos, como obras compuestas por partes distribuidas a lo largo de las páginas, libros de coro en los que una misma página contiene dos voces, conjuntos o partituras orquestales. El tipo de libro a transcribir puede ser de naturaleza totalmente distinta. Por ejemplo, puede ser una recopilación de canciones no relacionadas con ningún instrumento, como un real-book de jazz, o ser un catálogo que contenga listas de incipits. En algunos casos, el volumen a transcribir describe teoría musical como es el caso de los tratados de música, donde la mayor parte del contenido es textual con algunos ejemplos musicales ilustrativos. El proceso de tratamiento de las partes se realiza actualmente de forma manual en la herramienta.

En todos los casos, la implementación interna de todas esas situaciones se reduce al caso en que toda la imagen o página está vinculada a una parte, o

cuando cada región debe asignarse a una parte. El sistema también está preparado para tratar disposiciones de coral con dos pentagramas para cuatro voces. En ese caso, cada símbolo individual dentro de la región debe vincularse a cada parte. Para reducir el esfuerzo del usuario, el sistema ofrece ayudas para gestionar el conjunto de instrumentos y reutilizar las distintas disposiciones entre páginas.

Una vez identificados los distintos pentagramas y asignados a la parte a la que pertenecen, hay que reconocer y codificar el contenido musical dentro del recorte de imagen que corresponde a cada región. Existen varios enfoques que pueden seguirse para obtener una codificación a partir de la imagen, ya sea manualmente o aplicando un clasificador automático basado en IA. El primero consiste en trazar manualmente los símbolos gráficos “agnósticos” para que una IA, utilizando tanto el trazo como la imagen obtenida de la caja delimitadora que encierra el trazo, identifique cada símbolo entre un conjunto de posibles símbolos gráficos (es decir, símbolos gráficos sin un significado musical atribuido aún), y la posición vertical en el pentagrama como valor absoluto independientemente de la clave. Aunque utilizamos este concepto de símbolo agnóstico cuando identificamos glifos completos, podríamos adaptarlo fácilmente para reconocer primitivas (cabeza de nota, plica, etc.). Otra alternativa es simplemente marcar el cuadro delimitador que rodea a cada símbolo para utilizar sólo el recorte de la imagen como información de entrada para que la IA identifique automáticamente el tipo de símbolo agnóstico. Por último, otra opción es utilizar una IA que automáticamente recupere todos los símbolos que hay en el pentagrama, utilizando enfoques holísticos basados en redes neuronales. En cualquiera de las opciones mencionadas, se obtiene una secuencia de símbolos agnóstica, es decir, una secuencia de símbolos ordenados de izquierda a derecha, de arriba abajo.

La secuencia agnóstica debe convertirse a una codificación con significado musical que denotamos como codificación semántica. Esta traducción de agnóstico a semántico puede realizarse bien mediante un transductor de autómatas basado en reglas o bien mediante tecnologías de traducción basadas en enfoques de aprendizaje automático.

Una decisión clave a la hora de diseñar MuRET fue el método para almacenar esas secuencias semánticas. No hemos elegido ningún formato estándar como codificación interna, sino una representación ad-hoc extendida a partir

de los formatos Humdrum **kern** y **mens** en lo que denominamos **skm**. Posteriormente, al exportar la codificación final, se utilizan formatos estándar como Music Encoding Initiative (MEI) u otros.

Por último, una vez finalizados todos los pasos anteriores, el usuario puede seleccionar las imágenes que desea utilizar para generar una partitura final. Esta operación se realiza concatenando todos los pentagramas de las imágenes seleccionadas agrupados por las partes a las que pertenecen, exportándolos desde nuestro formato interno a MEI. Para compartir la transcripción con servicios externos, se puede exportar el MEI previo tal y como lo renderizamos. MuRET también es capaz de convertir a un formato MEI basado en partes incluyendo información gráfica en el elemento facsímil, lo que permite compartir el trabajo con herramientas especializadas como MP-Editor para realizar el procesamiento métrico en notación mensural.

Los clasificadores que apoyan los procesos automáticos utilizan modelos de aprendizaje automático, que necesitan datos de entrenamiento para ser construidos. El sistema permite descargar diferentes conjuntos de entrenamiento de colecciones y trabajos seleccionados. Estos datos de entrenamiento no son más que un conjunto de archivos JSON que contienen una exportación de los objetos de nuestro modelo de datos interno. Una vez entrenados offline, estos modelos pueden cargarse de nuevo en el sistema. Este flujo de trabajo incremental, es decir, corregir la salida de los diferentes clasificadores, descargar conjuntos de datos con datos corregidos, construir nuevos modelos y cargarlos para mejorar el rendimiento del sistema, está demostrando en nuestros proyectos de transcripción que es una forma adecuada de proceder.

Algunos modelos funcionan mejor que otros de forma teórica, pero las correcciones necesarias para fijar sus resultados suponen un mayor esfuerzo para el usuario. Esto puede medirse analizando las acciones que cada usuario individual realiza en la herramienta, que se registran y clasifican convenientemente en operaciones significativas, como la edición de regiones, símbolos, contenido semántico, gestión de piezas o uso de clasificadores. También se registran las marcas de tiempo de todas las operaciones, así como el documento, la región o el símbolo implicados en cada operación.