



Resumen de Tesis:

Transcripción de música polifónica para piano basada en la resolución de grupos de notas y estados finitos

Silvana Gómez-Meire

Departamento de Informática. Escuela superior de Ingeniería Informática. Universidad de Vigo
sgmeire@uvigo.es

Resumen En ésta Tesis se desarrolla un modelo para la transcripción musical polifónica mediante una aproximación de estados finitos y la aplicación de reglas de decisión basadas en heurísticas para la resolución de grupos de notas. El sistema de transcripción desarrollado se implementa mediante un diagrama de estados finitos que representan grupos de notas, es decir, conjunto de notas que suena en el mismo tiempo. El algoritmo de análisis de notas se fundamenta en una serie de colas y un conjunto de reglas de decisión para determinar los grupos de notas, y hace uso de ecuaciones de estado para realizar las transiciones entre grupos de notas. El sistema desarrollado ha sido validado de forma exhaustiva, sobre una amplia variedad de grabaciones de audio, mediante la aplicación de una serie de métricas desarrolladas específicamente para la evaluación del éxito obtenido. Los resultados experimentales confirman la validez del sistema propuesto.

Palabras clave: transcripción musical, análisis y procesamiento de señales.

1 Introducción

La transcripción musical se define como el proceso mediante el cual, a partir de la audición de una pieza musical se reconstruye la secuencia de notas que forman la partitura [1]. Dicho de un modo más apropiado, consiste en obtener una representación simbólica de la pieza que contenga todos los aspectos musicales de la misma, es decir, además de la identificación de la nota, determinar el tono, el ritmo, y la duración de la misma.

Este proceso resulta extremadamente difícil cuando se trata de transcribir música polifónica, incluso para aquellas personas que poseen educación musical. Esta dificultad es independiente de la procedencia del sonido, sean varios instrumentos musicales que suenan al mismo tiempo o sea un instrumento polifónico, como es el caso del piano.

El principal objetivo de este trabajo es desarrollar un sistema que sea capaz de, a partir de una grabación de una pieza de piano¹ sin acompañamiento, transcribir la partitura. El sistema desarrollado se basa en el análisis de la señal en el espacio de frecuencias, a partir de aquí, se realiza la transcripción mediante el uso de estados finitos y la aplicación de reglas de decisión. Para ello, es necesario realizar un buen procesamiento inicial de la señal con el objetivo de obtener el mejor conjunto de picos de frecuencia posibles. Con este fin se diseña un filtro basado en el uso de técnicas bien conocidas como la media móvil y el ajuste no lineal, que logra optimizar la obtención de picos de la señal. Una vez conseguidos los mejores picos, se utiliza un método de agrupamiento de frecuencias que permite distinguir unas notas de otras.

El sistema de transcripción que se presenta se deriva, casi directamente, de la transición natural entre notas, y se implementa mediante un sencillo diagrama de estados finitos, donde cada estado representa un grupo de notas,

¹ Se ha escogido el piano por ser un instrumento polifónico por naturaleza

es decir, notas que suenan en el mismo momento. Se aplica la definición de ecuaciones de estado para realizar las transiciones entre grupos de notas.

En cuanto a la validación del modelo propuesto, se ha desarrollado un marco metodológico adecuado para evaluar los resultados obtenidos por el sistema propuesto para el problema de la transcripción musical polifónica. Se han desarrollado un conjunto de métricas que constituyen una función de medida de la eficacia del sistema.

Por último y a la vista de los resultados obtenidos tras la experimentación realizada, se concluye que el sistema propuesto representa un nuevo método de resolución del problema objeto de estudio y obtiene unos resultados equiparables a otros métodos desarrollados para a este fin.

2 Contribuciones

La principal aportación del trabajo desarrollado ha sido la creación de un sistema de transcripción basado en la transición de estados para la identificación de las notas presentes en un segmento musical. La figura 1 muestra una visión del sistema de alto nivel.

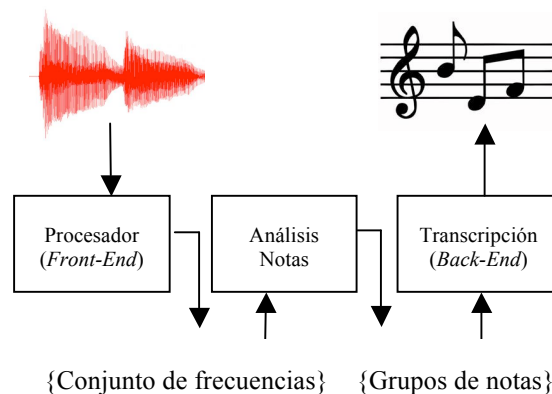


Figura 1. Arquitectura general del sistema

Para extraer las características de una señal de audio se requiere la realización de un análisis de la misma tanto en el dominio de tiempo como en el de frecuencia. Por consiguiente, la elección de una representación de tiempo-frecuencia apropiada determina el comportamiento del sistema de análisis. El objetivo de la primera etapa del sistema de transcripción desarrollado, denominado *front-end*, es transformar la señal original en el dominio de tiempo, al dominio de frecuencia y a partir de ella, extraer el mejor conjunto de frecuencias para determinar las notas. Para ello se ha conseguido realizar una óptima identificación de armónicos mediante la definición de una función umbral, formada por una media móvil y una función de *baseline*.

Otra aportación interesante de la investigación realizada es que se ha eliminado la necesidad de detectar el *onset* de las notas en el dominio de tiempo. La determinación de las notas se realiza enteramente a partir del análisis del espectro de frecuencias y esta tarea se lleva a cabo en la segunda etapa del sistema de transcripción. Para desarrollar el algoritmo de análisis de notas se han definido dos conceptos fundamentales. En primer, el concepto de *grupo de notas* se define como un conjunto de k notas $\{n_1, \dots, n_k\}$ que son tocadas en un tiempo determinado t_k con una duración Δt , de las cuales un subconjunto $j < k$ pueden haber concluido. En la figura 2 se puede ver un ejemplo de 3 grupos de notas definidos como $\{g^1, g^2, g^3\}$. En los dos primeros grupos de notas, ambas comienzan y terminan al mismo tiempo, mientras que en el tercer grupo la nota n_5 concluye antes de que lo hagan las demás.



Figura 2. Distintos grupos de notas

El otro concepto definido es el de *estados finitos*. Como se puede observar en la figura 3, un estado se corresponde con un grupo de notas bien definido del cuál decimos que está no concluido. Por lo tanto, la transición entre estados se corresponde con el momento en el que todas las notas pertenecientes a un grupo de notas determinado, ha concluido y se puede transcribir. Esta última etapa se denomina *back-end* e incluye todas las acciones necesarias para producir la partitura musical final. Para ello se hace uso de LilyPond [2], un sistema de tipografía con el que se pueden generar partituras musicales.

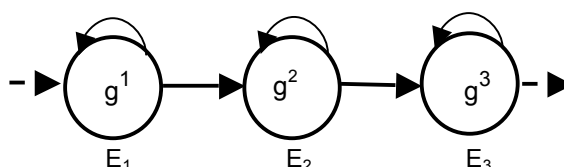


Figura 3. Diagrama de estados

Una de las bondades del sistema de transcripción radica en que la transición entre las notas se infiere directamente del espacio de frecuencias obtenido con el análisis de Fourier [3] y del gradiente de energía relativa de los armónicos asociados. Esto significa que el método se puede generalizar a otros instrumentos con sólo conocer el perfil del instrumento, es decir, las características que originan su sonido particular, de la misma manera que se ha hecho para el piano.

Además, cabe señalar otras aportaciones adicionales derivadas del propio desarrollo del modelo. A fin de evaluar el sistema, se han definido una serie de medidas para evaluar los resultados obtenidos. Esta metodología de evaluación se ha desarrollado para valorar la complejidad del segmento de audio que se desea transcribir y estimar el éxito alcanzado en el proceso de transcripción. La complejidad de la transcripción se medirá mediante una función diseñada *ad hoc* y a la que se denomina *función de complejidad*. En cuanto al comportamiento del sistema de transcripción, su evaluación se realiza mediante la utilización de una serie de estadísticos parciales y uno total denominado *índice de transcripción*, a fin de reflejar el éxito obtenido.

3 Estructura

El trabajo realizado se presenta organizado en siete capítulos, de los cuales el primero de ellos sirve para definir, en líneas generales, el problema que se pretende solucionar, establecer la hipótesis de partida y presentar la estructura organizativa de la memoria.

El capítulo dos describe cuestiones, tanto teóricas como prácticas, relacionadas con la producción del sonido del piano y con la naturaleza del análisis armónico de las señales de audio. En primer lugar se presentan una serie de conceptos musicales básicos que se van a manejar a lo largo de la memoria. En las siguientes secciones se da una visión general de la producción del sonido de los instrumentos que los caracteriza e identifica de modo inequívoco. Para finalizar, se discuten algunas cuestiones relacionadas con el proceso de grabación y formatos de audio que se manejan en esta investigación. El objetivo de este capítulo es dar una idea de la complejidad técnica que conllevan las tareas de reconocimiento de audio.

El capítulo tres realiza una revisión cronológica de trabajos previos desarrollados en el campo de la transcripción musical automática. El objetivo de este capítulo es indicar lo más claramente posible, los resultados obtenidos por otros grupos de investigación en este campo.

En el capítulo cuatro se aborda el *front-end* del sistema, la etapa de procesamiento inicial de la señal. Este capítulo explica en detalle el algoritmo fundamental para obtener el contenido armónico de una nota, así como los detalles del algoritmo de detección de picos que dará como resultado el mejor conjunto de frecuencias y amplitudes asociadas en el dominio de frecuencia, junto con la información del tiempo de las notas. En base a estos datos, la segunda etapa del sistema, el análisis de las notas, que se describe en el capítulo siguiente, es un problema independiente del procesamiento de la señal de audio.

En el capítulo cinco se describe el algoritmo de análisis de notas. En general, el problema se describe mediante una aproximación de espacio-estado, o lo que es lo mismo, a través de un autómata finito. La idea básica del algoritmo es considerar cada intervalo de tiempo como un quantum de tiempo durante el cual el sistema permanece en el mismo estado con el conjunto de notas actual, o cambia a un estado diferente con un nuevo conjunto de notas. Este algoritmo, aunque comparte algunos aspectos con otros trabajos de investigación, es una

contribución original al campo de la transcripción automática, siendo a la vez simple e intuitivo. En este capítulo también se describe el proceso de *back-end*, que tiene como objetivo traducir los grupos de notas resueltas a notación musical de partitura para piano. Esta traducción se realiza en base a reglas de decisión sencillas.

El capítulo seis se centra en la evaluación del modelo. Para determinar la eficacia de los algoritmos y del sistema de software se hace uso de un importante conjunto de grabaciones de audio de diferente calidad y complejidad. En este capítulo se describe un nuevo modelo para evaluar el éxito obtenido con los algoritmos. Para ello se han definido una serie métricas que constituyen una función de medida. Entre las funciones definidas se encuentran la función de complejidad, que cuantifica la dificultad de realizar la transcripción de un grupo de notas, y el índice de transcripción, que está formado por un conjunto de estadísticos que darán una medida de la efectividad del algoritmo en relación a la complejidad de los casos analizados.

Por último, el capítulo siete expone las conclusiones generales de la aplicación del modelo desarrollado al problema planteado. Se analizan los objetivos alcanzados y las líneas de trabajo futuro a desarrollar, tomando como punto de partida este trabajo de investigación, el sistema y el paquete de software que se ha desarrollado.

4 Resumen y conclusiones

En ese trabajo de tesis se ha abordado el problema de la transcripción musical polifónica. Para ello se ha desarrollado un sistema basado en la resolución de grupos de notas y estados finitos. La complejidad inherente del sonido originado por el piano, establece el marco de trabajo en el que se desarrolla la investigación. La no linealidad de este instrumento hace que la identificación de las notas se complique considerablemente, siendo insuficiente la realización de un análisis de la señal basado en la frecuencia fundamental de la nota.

Por consiguiente, se ha desarrollado un sistema en tres etapas. La primera etapa, denominada *front-end*, tiene como objetivo la discriminación de las notas y la determinación de su duración. Su objetivo es transformar la señal de audio original en el dominio de tiempo, al dominio frecuencia y obtener el conjunto de frecuencias correspondientes a cada nota. Esta etapa es de gran importancia para la posterior identificación de las notas, que se realizará en la segunda etapa del sistema, cuyo objetivo es el análisis de las notas. Para ello se ha desarrollado un algoritmo formado por un conjunto de funciones que realizan el análisis adecuado para determinar los grupos de notas junto con el tiempo asociado a cada una de ellas. Para ello se implementan una serie de colas que reflejan los diferentes estados del grupo de notas a analizar. Por último, en el última etapa del sistema se realiza el proceso de *back-end* que consiste en transcribir el conjunto de notas identificado a notación musical, obteniendo como resultado la partitura de la señal de audio analizada.

La validación del modelo se realiza mediante una métrica desarrollada para valorar la complejidad de la transcripción, así como una serie de parámetros estadísticos de utilidad para llevar a cabo la estimación de la efectividad de los algoritmos presentados.

References

- [1] Martin K.D., A blackboard system for automatic transcription of simple polyphonic music. *MIT Media Laboratory Perceptual Computing Section Technical Report No. 399*, 1996.
- [2] <http://www.lilypond.org/web>
- [3] Proakis, J.G., Manolakis, D.G. Tratamiento Digital de Señales. *Prentice Hall*, 1998.