UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO

Facultad de Ingeniería Eléctrica

División de Estudios de Posgrado

EVALUACIÓN AUTOMÁTICA DE ESTUDIANTES DE MÚSICA

TESIS

Que para obtener el grado de

MAESTRO EN CIENCIAS EN INGENIERÍA ELÉCTRICA

Opción en Sistemas Computacionales

Presenta

Sonia Tonatzi Morales Pintor

Dr. José Antonio Camarena Ibarrola

Director de Tesis

Morelia Michoacán, Agosto 2015

Resumen

El procesamiento para la extracción de características de señales de audio musicales, tal como la determinación de cromagramas, se pueden utilizar para: estimar la tonalidad de una pieza musical, medir la similitud entre dos interpretaciones, clasificar interpretaciones musicales, entre otras aplicaciones. Por otro lado, el alineamiento de audio es un proceso mediante el cual se calcula la distancia entre los vectores característicos de dos señales de audio, para poder evaluar que tanto se parecen entre sí. Para la implementación de sistemas de alineamiento de audio, comúnmente es utilizado el algoritmo de doblado dinámico en el tiempo (DTW del inglés Dynamic Time Warping).

En este trabajo de investigación se pretende crear un sistema de evaluación automática de estudiantes de música. La evaluación requiere del procesamiento de señales digitales de audio para la determinación de espectrogramas y cromagramas, mediante el uso de la transformada de Fourier de tiempo corto (STFT del inglés short-time Fourier Transform). El uso de los cromagramas es más factible para la comparación de interpretaciones musicales ya que se construyen a partir de los 12 semitonos de la escala musical. Para el alineamiento de las interpretaciones musicales se propone el uso del algoritmo de DTW utilizando distancia euclidiana y distancia coseno. Para la evaluación se tendrá la interpretación del estudiante, que será comparada con la interpretación de referencia (interpretada por el profesor de música). Los alumnos podrán obtener una calificación de 0 a 10, esta dependerá de la distancia obtenida en el proceso de alineación de audio.

Los resultados obtenidos en este trabajo de investigación fueron satisfactorios, ya que al comparar los resultados dados por el sistema de evaluación de las interpretaciones de los estudiantes, con las calificaciones realizadas por los profesores de música, se obtuvieron resultados iguales o muy similares. Además, se pudo observar que los mejores resultados fueron obtenidos al representar los vectores característicos mediante cromagramas y utilizando la distancia coseno para medir las distancias dentro del algoritmo DTW. En conclusión se logró implementar un sistema que obtiene una evaluación confiable para estudiantes de música.

Palabras clave: evaluación, música, procesamiento de señales de audio, espectrogramas, cromagramas, alineación de audio.

Abstract

The processing for feature extraction of musical audio signals, such as determining chromagrams, can be used to estimate the tone of a track, measure the similarity between two interpretations, classify musical performances, among other applications. On the other hand, the audio alignment is a process by which the distance is calculated between the characteristic vectors of two audio signals in order to evaluate both resemble each other. To implement alignment of audio systems, is commonly used the DTW algorithm.

In this document is pretend to create a system for automatic evaluation of music students. Evaluation requires processing digital audio signals for determining chromagrams and spectrograms, by using the short-time Fourier Transform (STFT). The use of chromagrams is more feasible to compare musical performances because they are built from the 12 semitones of the musical scale. For the alignment of the musical performance it's proposed the DTW algorithm, using Euclidean distance and cosine distance. For the interpretation of student assessment, which will be compared with the performance of reference (played by the music teacher). Students may earn a grade of 0-10, this will depend on the distance obtained in the process of alignment of audio.

The results obtained in this document were satisfactory, because comparing the results given by the evaluation system of the students performances, with ratings by music teachers, same or very similar results were obtained. Furthermore, it was observed that the best results were obtained by representing feature vectors by using chromagrams and the cosine distance within the DTW algorithm to measure distances. In conclusion it was achieved implement a system to get a reliable assessment for students in music.

Keywords: evaluation, music, audio signal processing, spectrograms, chromagrams, audio alignment.

Contenido

Lista de Figuras

Lista de Tablas

Lista de Símbolos

HMM	Hidden Markov Model.
$MFCC$	Mel Frequency Cepstral Coefficients.
DTW	Dynamic Time Warping.
DFT	Discrete Fourier Transform.
FFT	Fast Fourier Transform.
$STFT$	short-time Fourier Transform.
EM	Expectation Maximization.
$IMUTUS$	Interactive Music Tuition System.
LCS	Longest Common Subsequence.
z	Frecuencia en Barks.
f	Frecuencia en Hertz.
f_c	Frecuencia central.
m	Mels.

Capítulo 1

Introducción

1.1. Planteamiento del problema

En la actualidad existen escuelas de música con una gran cantidad de alumnos inscritos en los diferentes programas que ofrecen, para cada uno de los instrumentos musicales, en los que el alumno desea especializarse. Por tal motivo, al momento de evaluar a cada uno de sus alumnos las escuelas requieren invertir cierta cantidad de tiempo, además requieren de la atención de varios profesores, esto debido a que se deben formar mesas de sinodales que evalúen al alumno con la finalidad de evitar que un sólo profesor decida la calificación del mismo y de esta manera, aunque exista la posibilidad de que algún profesor tenga preferencia por cierto alumno, la evaluación pueda ser más justa. Por lo tanto surge la necesidad de buscar alternativas para poder automatizar la evaluación de los estudiantes y de esta manera poder acelerar las evaluaciones, y que además los resultados puedan ser confiables.

En este trabajo de investigación se pretende crear un sistema que pueda evaluar a cada uno de los estudiantes inscritos a un programa de una escuela de música, en este caso a estudiantes de piano, así se podrán disminuir los costos de los tiempos de evaluación y se obtendrán resultados más confiables, ya que el sistema será imparcial al momento de evaluar. Para la implementación de este sistema de evaluación automática de estudiantes de música, se hicieron grabaciones a tres estudiantes y un profesor de piano del "Conservatorio

de las rosas", escuela de música de la ciudad de Morelia, se grabaron varias interpretaciones de una misma pieza musical; la interpretación del profesor, tomada como la interpretación de referencia y las interpretaciones de los estudiantes, tomadas como interpretaciones de prueba para las evaluaciones. Es importante aclarar que las interpretaciones se deben grabar y posteriormente evaluar, es decir, el estudiante no se evalúa al momento de interpretar una pieza musical, sino que primero se graban las interpretaciones de los estudiantes y del profesor, y posteriormente mediante el sistema implementado es comparada la interpretación del alumno contra la interpretación de referencia, y así el estudiante obtiene una calificación.

Para el desarrollo de este sistema es necesario utilizar técnicas de procesamiento digital de señales de audio para la extracción de características, también es preciso encontrar una técnica de alineamiento adecuada para que el sistema sea eficiente en el proceso de comparación entre las piezas interpretadas, y por lo tanto el resultado de evaluación del alumno sea confiable.

Para llevar a cabo el procesamiento de la señal de audio de la interpretación musical, se aplica la Transformada de Fourier de Tiempo Corto (STFT del inglés short-time Fourier Transform) a la señal, para lo cual primero se divide la señal en marcos a los que después se les aplica una ventana de Hamming. Posteriormente se aplica, a cada uno de los marcos, la Transformada rápida de fourier (FFT del inglés Fast Fourier Transform). La interpretación física correcta de estos parámetros en términos de unidades tales como segundos y Hertz depende de; la tasa de muestreo, el tamaño de la ventana y el tamaño de salto que sea utilizado en el cálculo de la STFT. La STFT puede ser visualizada por medio de una gráfica denominada espectrograma, en la que el eje horizontal representa el tiempo, el vertical la frecuencia y la amplitud de los coeficientes está representada por la intensidad del color o escala de grises de la imagen, también se puede obtener otra gráfica llamada cromagrama, por medio de la asignación de tonos mediante los valores croma a cada uno de los coeficientes espectrales obtenidos de la señal.

Finalmente mediante la técnica de alineamiento de doblado dinámico en el tiempo (DTW del inglés Dynamic Time Warping) se mide la similitud entre las dos interpretaciones musicales obteniendo la distancia entre las mismas, para poder entregar una calificación al alumno según la calidad de la interpretación. Se debe tener en cuenta que se toma

como criterio principal de evaluación la precisión en el sonido, es decir, la habilidad que el estudiante tiene para tocar las notas correctas según la pieza musical que este interpretando.

1.2. Antecedentes y estado del arte

La alineación de interpretaciones musicales es un campo en el que ya se lleva tiempo trabajando, esto surge por la necesidad de tratar el problema de reconocimiento de voz, donde ahora no sólo se quiere reconocer o alinear señales de voz, sino también piezas musicales.

La curiosidad de tratar el problema de reconocimiento de voz surge principalmente de la idea de poder interactuar con la computadora o algún dispositivo mediante el habla. Algunos de los primeros en interesarse en esta área fueron los laboratorios de AT&T y Bell a principios del año 1940, quienes desarrollaron un aparato primitivo que podía reconocer voz, de manera sencilla, puesto que se basaba básicamente en identificar una palabra con sólo ver su espectrograma. Sin embargo, se sabía que para obtener el éxito deseado de esta nueva tecnología todo dependería de la habilidad de percibir información verbal compleja con buena precisión, por lo que en 1960 los científicos se enfocaron en desarrollar sistemas de reconocimiento de voz más avanzados, fue entonces cuando desarrollaron un aparato que podía usar la conversación discreta, es decir el usuario debía dictar oraciones con pausas entre cada palabra, fue también por aquellos años cuando la escala de Bark es propuesta por Eberhard Zwicker [Zwicker61] quien además aportó otras investigaciones que ayudaron al avance de esta tecnología.

Sin embargo en el año de 1970 fue cuando realmente se desarrolló una tecnología de reconocimiento de voz que no requería que el usuario hiciera pausas entre palabras, en los años 80 esta se volvió práctica y sigue siendo utilizada en la actualidad. Como ejemplo de esto se puede ver [Davis80] el cual propone el reconocimiento de palabras en oraciones continuamente habladas mediante la extracción de características de la señal de los Coeficientes Cepstrales de Frecuencia de Mel (MFCC del inglés Mel Frequency Cepstral Coefficients).

En el año de 1989 Lawrence R. Rabiner expone un tutorial sobre Modelos Ocultos de Márkov (HMM del inglés Hidden Markov Model) [Rabiner89], en donde revisa los aspectos teóricos de este tipo de modelos estadísticos y muestra como se han aplicado a ciertos problemas de reconocimiento de voz. Los HMM fueron introducidos a finales de los 60 y principios de los 70, y en la actualidad siguen siendo utilizados en reconocimiento de voz. Además en los últimos años los HMM se han comenzado a utilizar en reconocimiento de acordes musicales.

En el trabajo de investigación [Orio01] se propone una nueva metodología para la alineación automática de interpretaciones musicales monofónicas y polifónicas basada en el uso de DTW, en donde se utilizan los picos espectrales de la señal de audio para calcular la distancia entre las interpretaciones, la metodología logra hacer frente a intepretaciones que se consideran difíciles de alinear, como la música polifónica, secuencias rápidas o música multi-instrumento.

Por otro lado, mientras que en el documento anterior utilizan DTW para la alineación de interpretaciones musicales, Sheh y Ellis [Sheh03] proponen un método de aprendizaje estadístico para el reconocimiento de acordes utilizando HMM los cuales se forman mediante el algoritmo de EM (del inglés Expectation Maximization), puesto que calcula los valores promedio y varianza de los vectores así como las probabilidades de transición para cada acorde, los acordes son tomados como valores ocultos y los modelos utilizan la secuencia de acordes como entrada. Con los parámetros ya definidos el modelo se puede utilizar ahora para determinar un etiquetado acorde para cada canción, el algoritmo de Viterbi [Gold00] se utiliza para alinear o reconocer estas etiquetas; en alineación forzada, las observaciones están alineadas a un HMM compuesto cuya transiciones se limitan a las dictadas por una secuencia de acordes específica, los resultados del reconocimiento no fueron tan satisfactorios, pudiendo ser la razón principal la falta de datos para el entrenamiento.

En el trabajo de investigación [Hu N.03] se propone la alineación de audio polifónico mediante archivos MIDI, para lo cual se hace uso de vectores cromas que en conjunto forman lo que se conoce como cromagrama, los cromas representan la energía espectral a un tono. Para la alineación del audio se utiliza DTW, este trabajo está estrechamente relacionado con el de Orio y Schwarz [Orio01] en el cual también utilizan el algoritmo de

DTW para alinear la música polifónica, pero a diferencia de este último no se utilizan los picos espectrales de la señal de audio para calcular las distancias, puesto que se construyen cromagramas de los audios, existen aún posibilidades de mejorar el rendimiento de este trabajo, sin embargo los resultados que se obtienen en este trabajo son muy satisfactorios.

En el documento [Fober04] se presenta el sistema IMUTUS (del inglés Interactive Music Tuition System), el cual es un proyecto europeo que consiste en un nuevo enfoque de aprendizaje musical, y que tiene como objetivo el desarrollo de una plataforma abierta para la formación de estudiantes de música, así como de dar conocimientos teóricos de música, demostrando que una computadora puede proporcionar retroalimentación que en conjunto con las enseñanzas del docente puede utilizarse para reflejar el rendimiento de los estudiantes. IMUTUS presenta de manera visual una evaluación para mejorar la eficiencia de la práctica, también propone un editor de partituras de música que son de gran ayuda a los estudiantes.

En [Schoonderwaldt05] se presentan algunos resultados del proyecto IMUTUS, este trabajo se enfoca en el ambiente de práctica del sistema IMUTUS, este ambiente integra herramientas para el análisis y evaluación automática de interpretaciones de estudiantes de música, además de esquemas de interacción que proporcionan un enfoque eficaz para el aprendizaje de la música, las actividades de prueba realizadas mostraron que el enfoque IMUTUS es muy bueno, debido a que la validación del mismo mostró que los estudiantes que lo utilizaron mostraron una mejora musical.

En el documento [Robine07] se presenta un sistema capaz de analizar la técnica o habilidad de los saxofonistas. Se proponen varios métodos para evaluar el desempeño de los mismos examinando varias características de audio, de manera que los resultados se acerquen a la evaluación de un profesor de música profesional, aunque en este trabajo se enfocan en el saxofón, los métodos utilizados pudieran aplicarse a otros instrumentos de viento. Para que pueda ser evaluada la técnica de un artista principalmente se califica la exactitud del tono (pitch en inglés), ritmo (duración), y el tempo (velocidad en música).

En [Percival07] se presenta un estudio de los trabajos recientes de enseñanza asistida por computadora de algún instrumento musical y formula varias preguntas a considerar en el desarrollo de futuros proyectos. En particular, se sugiere que el área de mayor necesidad

de ayuda de la computadora es mejorar la práctica diaria, a través de juegos y multimedia, y proporcionar un análisis objetivo de los resultados de los estudiantes. Muchos proyectos existentes intentan reemplazar a los maestros humanos, proporcionando lecciones durante la práctica diaria. Además se identifica que para mejorar la práctica individual diaria es necesario aumentar la motivación y eficiencia de ejercicios técnicos. Esta motivación puede tomar varias formas, por ejemplo los juegos de entretenimiento educativos de instrumentos musicales los cuales presentan algunos desafíos especiales, además se deben utilizar técnicas de visualización novedosas para proporcionar retroalimentación intuitiva para los estudiantes.

En [Manzo-Martinez13] se propone el análisis de las características de entropía por croma, utilizando la entropía de Shannon para estimar el nivel de información contenida en cada bin de croma de una señal de audio, y se aplica en los sistemas de reconocimiento de audio basados en audio-huellas dactilares. En los experimentos se utiliza un conjunto de interpretaciones de música polifónica. Además se igualan dos versiones, que tienen el mismo rendimiento, con la finalidad de demostrar que ambas interpretaciones producen casi el mismo conjunto de características a pesar de ser interpretadas por otro músico y/o diferentes instrumentos.

1.3. Objetivos de la tesis

Objetivo general

El objetivo de este trabajo de investigación es implementar un sistema que sea capaz de evaluar a un alumno de música en cualquiera de los instrumentos musicales que se imparten en los programas de una escuela de música. El programa entregará una calificación de 0 a 10, dependiendo el desempeño del alumno.

Objetivos particulares

- Implementar mediante el procesamiento digital de señales de audio un proceso de extracción de características útil para el desarrollo del sistema.

- Determinar los espectrogramas y cromagramas, mediante la extracción de características de cada una de las interpretaciones musicales, de manera que puedan compararse los resultados que se obtienen utilizando cada uno, y así utilizar la gráfica que devuelva los mejores resultados.

- Implementar algún algoritmo de alineamiento de series de tiempo o alguna técnica de procesamiento de cadenas, tales como; DTW, Levenshtein y LCS (del inglés Longest Common Subsequence).

- Comparar los resultados que se obtienen, al utilizar el algoritmo de DTW utilizando distancia euclidiana contra los resultados que se obtienen mediante DTW utilizando distancia coseno, con el fin de utilizar el algoritmo con los resultados óptimos en la etapa de alineamiento.

- Evaluar los resultados del sistema utilizando los diferentes esquemas implementados, mediante el calculo de la correlación, para obtener los mejores resultados.

- Evaluar la interpretación musical del estudiante con el proceso de extracción de características y la técnica de alineamiento que obtuvó los resultados óptimos.

1.4. Justificación

En los últimos años el uso de la tecnología ha impactado en todas las áreas de las actividades humanas, es por esto que es normal que para facilitar las tareas en cada una de estas áreas se haga uso de la automatización de las actividades mediante la implementación de algún sistema.

Por esta razón en este trabajo de investigación se propone la implementación de un sistema de evaluación automática de estudiantes de música, este sistema se puede ver como una caja negra, que como entrada recibe; la interpretación de referencia y la interpretación a ser evaluada, interpretada por el estudiante, estas interpretaciones se comparan y la salida del sistema entrega una calificación según el desempeño del estudiante. El desempeño del estudiante se evalúa según la exactitud de tono, es decir, el sistema califica que el estudiante haya tocado las notas correctas acorde a la interpretación de referencia.

Por lo tanto, la implementación de este sistema podría acelerar las evaluaciones de los estudiantes de una escuela de música, obteniendo resultados confiables en poco tiempo.

1.5. Descripción de capítulos

En el capítulo 2 se explica el procesamiento digital de la señal de audio para la extracción de características, ventana de hamming, la aplicación de la transformada de Fourier, el cálculo de la energía, hasta llegar a la obtención del espectrograma y cromagrama de la señal.

En el capítulo 3 se exponen las diferentes técnicas para la etapa de alineamiento o comparación, como es el algoritmo de DTW, la distancia de Levenshtein y la distancia LCS.

En el capítulo 4 se explica la implementación del programa para la evaluación automática de estudiantes de música, se exponen los diferentes esquemas utilizados y el proceso seguido para la implementación de estos, explicando la etapa de procesamiento de la señal de audio para la extracción de características y la etapa de alineamiento de audio.

En el capítulo 5 se muestran las pruebas realizadas, así como los resultados obtenidos y comparaciones entre estos, además se evalúa el rendimiento del sistema.

En el capítulo 6 finalmente se exponen las conclusiones y se explica el trabajo futuro que puede realizarse, así como la manera en que puede mejorarse el sistema implementado para obtener mejores resultados.

Capítulo 2

Caracterización de la señal de audio

Las señales de audio son una representación del sonido originado por la voz de una persona o algún instrumento musical. Normalmente se encuentran acotadas al rango de frecuencias audibles por el ser humano entre los 20 Hz y los 20 kHz aproximadamente. Una señal de audio se puede comprender mejor si primero se descompone en segmentos llamados tramas o marcos, lo que es mejor para las etapas de procesamiento posteriores.

2.1. Determinación del espectrograma

Un espectrograma resulta de calcular el espectro de cada una de las tramas de una señal, en este caso de una señal de audio de una interpretación musical, como resultado se obtiene una gráfica tridimensional, como se muestra en la Figura 2.1 [Smith07]. En un espectrograma el eje horizontal representa el tiempo, el eje vertical la frecuencia y la amplitud de las señales se representa mediante una escala de grises o colores, en donde la intensidad corresponde con la amplitud de los coeficientes.

El proceso para determinar un espectrograma se inicia con la obtención de segmentos cortos de la señal de audio, conocidos como marcos o tramas, esto para poder considerar a la señal en ese segmento como una señal estacionaria, y deben ser considerablemente cor-

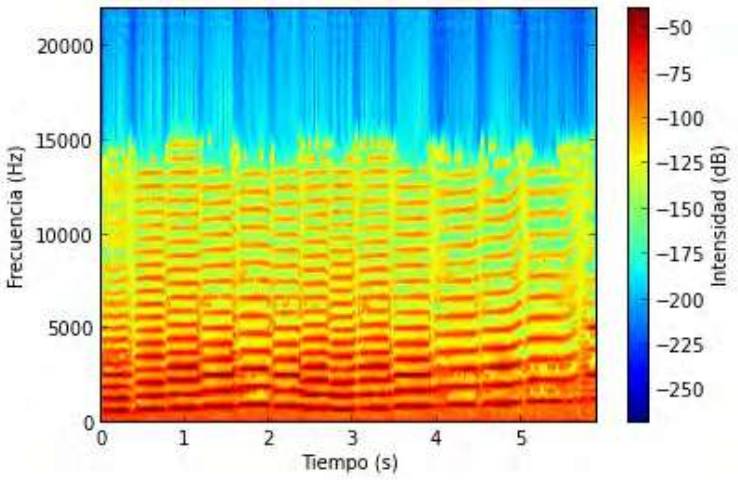


Figura 2.1: Espectrograma.

tos pero, lo suficientemente grandes como para que quepan dos periodos completos de la frecuencia más baja que estemos utilizando.

Puesto que cada uno de los marcos que se obtengan de la señal de audio, puede que estos comiencen o terminen con un valor diferente de cero y que estos valores sean muy distintos entre sí, la transformada de Fourier puede entender esta diferencia como una discontinuidad. Por lo tanto para poder evitar que, al momento de aplicar la transformada de Fourier, ocurran estos cambios bruscos en la señal se debe aplicar una ventana que desvanezca la señal en los extremos del marco, al utilizar esta ventana se evita el efecto llamado "leakage", normalmente traducido como escurrimiento. Existen varios tipos de ventanas como son; la ventana rectangular, la ventana de Welch, la ventana de Hann, la ventana de Hamming, entre otras.

La ventana más comúnmente utilizada es la de Hamming, la cual se muestra en la Figura 2.2 y está definida con la ecuación (2.1):

$$w_h[n] = 0.54 - 0.46cos\left(\frac{2\pi n}{N}\right) \tag{2.1}$$

donde n se refiere a la muestra, y N es al tamaño de ventana.

La transformada de Fourier, transforma una señal en el dominio del tiempo a una señal en el dominio de la frecuencia, revelando el espectro de componentes de frecuencia que

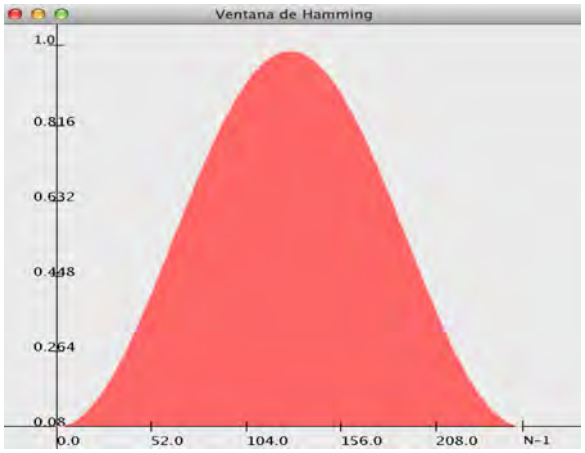


Figura 2.2: Ventana de Hamming.

componen la señal original. De manera informal y haciendo una analogía, se podría decir que se tienen dos panoramas distintos de un mismo lugar, puesto que desde un punto de observación se muestra la información del tiempo, pero si cambiamos el punto de observación la transformada de Fourier revela información sobre la frecuencia. Para poder obtener la información en frecuencia sin perder la información en el tiempo, Gabor introdujo en el año de 1946 [Gabor46] la transformada de Fourier modificada, actualmente conocida como la transformada de Fourier de tiempo corto (abreviada STFT del inglés short-time Fourier Transform). La STFT no sólo nos dice qué frecuencias están contenidas en la señal, sino también en que intervalos de tiempo aparecen estás frecuencias.

Después de aplicar una ventana de Hamming al marco de tiempo se procede a obtener la STFT, definida por la ecuación (2.2), posteriormente se recorre la ventana y se repite el proceso, es necesario que al avanzar los marcos exista un traslape entre marcos consecutivos, con la finalidad de evitar cambios bruscos por pérdida de información. El desplazar el marco ayuda a conocer como van cambiando los componentes de frecuencia de la señal en el tiempo, ya que como se explicó anteriormente se necesita tanto del dominio del tiempo como del dominio de la frecuencia, finalmente el resultado de este proceso será obtener una secuencia de espectros de frecuencia.

$$X[m,k] = \sum_{n=0}^{N-1} x[n+mH]w_h[n]e^{\frac{-j2\pi nk}{N}} \tag{2.2}$$

donde $X[m,k]$ denota el k-ésimo coeficiente de Fourier para el m-ésimo marco de tiempo, x es la señal discreta obtenida por las muestras equidistantes, tomadas de la señal continua, según la frecuencia de muestreo F_s dada en Hz, w_h es una ventana de tiempo discreto de longitud N y H es el tamaño de salto.

La STFT a menudo se visualiza por medio de un espectrograma, el cual se origina a partir del cálculo de la potencia, esta se obtiene elevando al cuadrado la magnitud $X[m,k]$, obtenida a partir de la STFT, por tanto un espectrograma está dado por la ecuación (2.3).

$$espectrograma = |X[m,k]|^2 \tag{2.3}$$

Para poder apreciar mejor un espectrograma es conveniente expresar la magnitud de los coeficientes en decibeles, es decir $20log_{10}(|X[m,k]|)$. Un ejemplo de un espectrograma se muestra en la Figura 2.1.

2.1.1. Resolución del espectrograma

La resolución fija es uno de los principales problemas de la STFT. La resolución dependerá del ancho de la función de ventana, puesto que una ventana angosta da una buena resolución en tiempo pero mala resolución en frecuencia, mientras que una ventana amplia da una buena resolución en frecuencia pero mala resolución en tiempo, usualmente son llamadas transformada de banda angosta y transformada de banda ancha, respectivamente. Para poder visualizar lo anterior, en la Figura 2.3 se muestra un ejemplo de la comparación entre dos espectrogramas [Flanagan J. L.66], el primero tiene un tamaño de ventana de $N = 256$ y el segundo tiene un tamaño de ventana de $N = 1024$.

Existe la manera de conocer la resolución en tiempo y la resolución en frecuencia de un espectrograma, cada coeficiente de Fourier $X[m,k]$ se asocia con la posición del tiempo dado en segundos, esto se muestra en la ecuación (2.4):

$$T_{coef} = \frac{mH}{F_s} \tag{2.4}$$

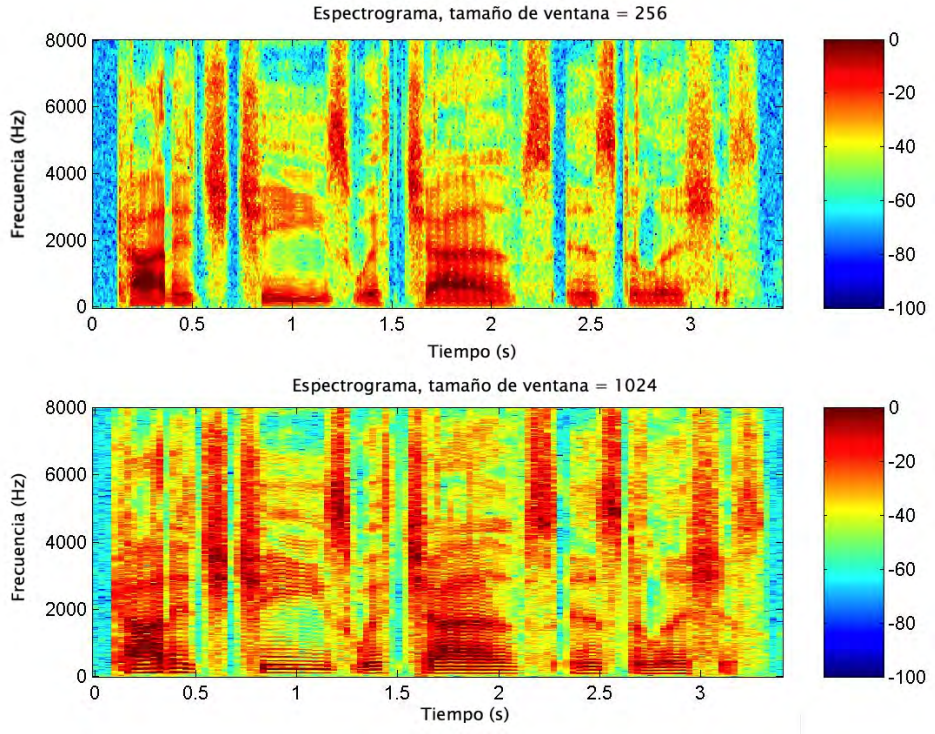


Figura 2.3: Ejemplo de resolución en espectrogramas

donde $H \in \mathbb{N}$ es un parámetro de tamaño de salto y F_s es la frecuencia de muestreo fija dada en Hertz (Hz).

Por otro lado la resolución en frecuencia dada en Hertz (Hz) se obtiene con la ecuación (2.5):

$$F_{coef} = \frac{kF_s}{N} \tag{2.5}$$

donde k es el k-ésimo coeficiente de Fourier y N es el tamaño de la longitud de la ventana.

Para poder entender esto mejor se considera el siguiente ejemplo, suponiendo una frecuencia de muestreo de $Fs = 44100$ Hz como la que se utiliza en una grabación de CD, una longitud de ventana de $N = 4096$ y un tamaño de salto de $H = N/2$, se obtiene lo siguiente:

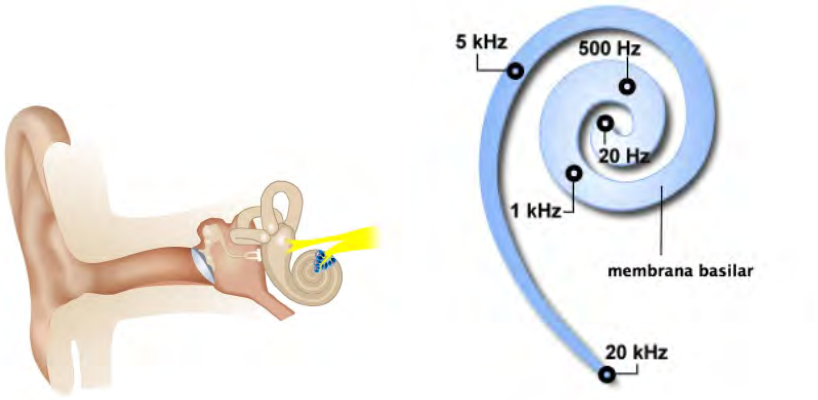
$$H = N/2 = 4096/2 = 2048$$

$$H/F_s = 2048/44100 \approx 46.4 \text{ ms}$$

$$F_s/N = 44100/4096 \approx 10.76 \text{ Hz}$$

2.2. Caracterización de la señal de audio mediante energía por banda de Bark

El oído humano percibe los sonidos mediante una descomposición en frecuencias realizada en la cóclea. La Figura 2.4(a) muestra como el oído se divide en; externo, medio e interno, en la Figura 2.4(b) se observa un diagrama de la cóclea, esta se localiza en el oído interno. Diferentes frecuencia acústicas activan diferentes localidades de la membrana basilar [Luo12]. El rango frecuencial que puede percibir el oído humano se encuentra entre los 20 Hz y 20 kHz, ya que el oído no es capaz de distinguir frecuencias muy altas o muy bajas. Otro dato importante es que el rango frecuencial de una locución es menor que el de una pieza musical, puesto que generalmente la voz oscila alrededor de los 60 y 7000 Hz, mientras que las piezas musicales son capaces de llenar todo el espectro auditivo, es decir de 20 Hz a 20 kHz.



(a) ***Sistema auditivo periférico*** **(b)** ***Frecuencias que se perciben en la cóclea***

Figura 2.4: Sistema auditivo y la cóclea.

Se sabe que el oído humano percibe mejor las frecuencias bajas que las altas. Asemejando la forma en que el oído trabaja se utiliza una escala que agrupa frecuencias

en bandas, llamada escala de Bark. La escala de Bark es una escala psicoacústica llamada así en honor a Heinrich Barkhausen quien realizó las primeras mediciones de la percepción de intensidad o sensación sonora. Esta escala describe una distancia fija a lo largo de la membrana basilar, fue propuesta por Eberhard Zwicker [Zwicker61] y diseñada para modelar el oído humano.

La escala de Bark define 25 bandas críticas, es decir, comprende un rango de Bark 0 a Bark 24, según estudios la membrana basilar tiene una distancia de 32 mm, lo cual indica que cada banda crítica representa aproximadamente 1.3 mm de distancia a lo largo de la membrana basilar, es decir, cada banda corresponde a un segmento de la cóclea. En la Tabla 2.1 se describen las 25 bandas críticas de Bark.

Tabla 2.1: Las 25 bandas críticas de la escala de Bark

Frecuencia inic (Barks)	Frecuencia inic (Hz)	Frecuencia final (Hz)
0	0	100
1	100	200
2	200	300
3	300	400
4	400	510
5	510	630
6	630	770
7	770	920
8	920	1080
9	1080	1270
10	1270	1480
11	1480	1720
12	1720	2000
13	2000	2320
14	2320	2700
15	2700	3150
16	3150	3700
17	3700	4400
18	4400	5300
19	5300	6400
20	6400	7700
21	7700	9500
22	9500	12000
23	12000	15500
24	15500	20000

Para convertir Hertz a Barks se hace uso de la Ecuación 2.6 [Traunmüller90]:

$$z = 13tan^{-1}\left(\frac{0.76f}{1000Hz}\right) + 3.5tan^{-1}\left(\frac{f}{7500Hz}\right)^2 \tag{2.6}$$

donde z es frecuencia en Barks y f es la frecuencia en Hertz.

Entonces utilizando esta escala de Bark se puede obtener el espectrograma de una canción, para lo cual no es necesario utilizar las 25 bandas de Bark. Por ejemplo con 18 bandas puede ser suficiente, puesto que generalmente, un audio no cubre todo el espectro auditivo.

El espectrograma obtenido tendría una resolución en tiempo que dependerá del ancho de los marcos así como el traslape que se tenga entre los mismos y una resolución en frecuencia que dependerá precisamente del número de bandas que se tomen en cuenta, si fueran 18 bandas entonces se tendrían por cada marco un vector de longitud 18, donde cada uno de los 18 valores corresponde a la energía obtenida al aplicar la transformada de Fourier de tiempo corto a ese marco.

2.3. Caracterización de la señal de audio mediante valores croma

La percepción humana del tono (en inglés llamado pitch) es periódica en el sentido de que un ser humano percibe de manera similar en color dos tonos distintos que pueden ser la misma nota musical, pero que difieren en una o más octavas (puesto que juegan un papel armónico similar).

En la música una octava está dividida en 12 tonos los cuales incluyen 7 notas más los bemoles y los sostenidos, es decir, están dados por el conjunto: Do, Do#, Re, Re#, Mi, Fa, Fa#, Sol, Sol#, La, La#, Si, siendo '#' una nota sostenida, tal como se observa en la Figura 2.5 [Armónica07]. Sin embargo, existe otra notación musical llamada notación anglosajona o notación americana, utilizada en Inglaterra y Estados Unidos. La notación americana es un sistema de notación musical alfabética.

El nombre de octava se debe a la distancia que existe entre una misma nota musical

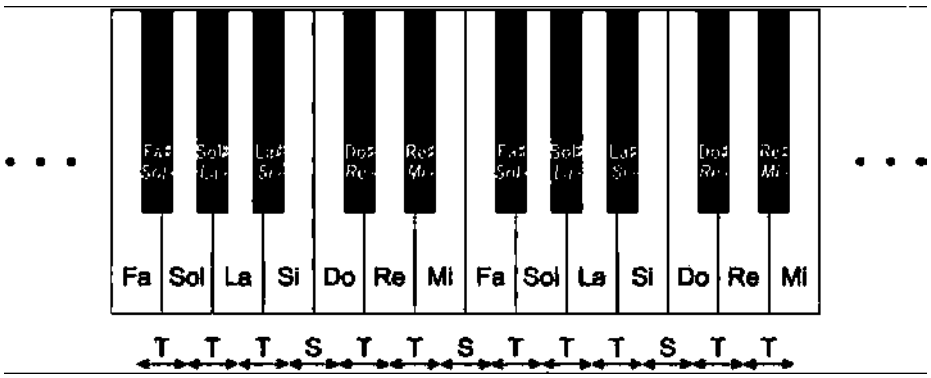


Figura 2.5: Tonos y semitonos de la escala musical.

pero perteneciente a diferente octava, por ejemplo **Do** - Re - Mi - Fa - Sol - La - Si - **do** para la notación musical latina y **C** - D - E- F- G- A - B - **C** para la notación musical americana. La representación de la notación musical en el pentagrama se puede observar en la Figura 2.6.

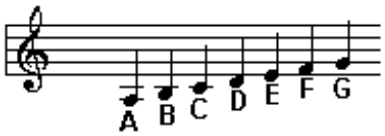


Figura 2.6: Representación de la escala musical en el pentagrama.

De manera formal una octava se define como el intervalo que separa dos sonidos cuyas frecuencias fundamentales tienen una relación de dos a uno, por ejemplo la nota A3 está una octava arriba de la nota A2, puesto A3 tiene una frecuencia fundamental de 220 Hz, mientras que A2 tiene una frecuencia fundamental de 110 Hz, es decir, la frecuencia de A3 es el doble de la frecuencia de A1. La Tabla 2.2 muestra algunas de las frecuencias centrales de la escala musical. La diferencia entre octavas, también puede verse en valores de tonos, por ejemplo, el tono (abreviado p del inglés pitch) $p = 60$ y $p = 72$ tienen una octava de diferencia, mientras que los tonos $p = 60$ y $p = 78$ están alejados dos octavas. La diferencia entre octavas puede calcularse con la Ecuación (2.7).

El número de octavas entre dos frecuencias puede calcularse mediante el uso de

logaritmos en base 2, por ejemplo, sabemos que el rango de frecuencias audibles por el oído humano es de 20 Hz a 20 kHz, por lo tanto, el número de octavas que abarca este rango es de:

$$log_2(\frac{20000}{20}) = 9.9657 \text{ octavas}$$

Tabla 2.2: Algunas frecuencias centrales de la escala musical.

nota	frecuencia	nota	frecuencia	nota	frecuencia
C2	65.406 Hz	C3	130.812 Hz	C4	261.625 Hz
C# 2	69.295 Hz	C# 3	138.591 Hz	C# 4	277.182 Hz
D2	73.416 Hz	D3	146.832 Hz	D4	293.664 Hz
D# 2	77.781 Hz	D# 3	155.563 Hz	D# 4	311.127 Hz
E2	82.406 Hz	E3	164.813 Hz	E4	329.627 Hz
F2	87.307 Hz	F3	174.614 Hz	F4	349.228 Hz
F# 2	92.498 Hz	F# 3	184.997 Hz	F# 4	369.994 Hz
G2	97.998 Hz	G3	195.997	G4	391.995 Hz
G# 2	103.826 Hz	G# 3	207.652 Hz	G# 4	415.304 Hz
A2	110 Hz	A3	220 Hz	A4	440 Hz
A# 2	116.540 Hz	A# 3	233.081 Hz	A# 4	466.163 Hz
B2	123.470 Hz	B3	246.941 Hz	B4	493.883 Hz

Un tono p tiene dos componentes: la altura de tono (pitch height) y un croma. La altura de tono se refiere al número de octava y el croma al tono respectivo según la escala musical. Entonces los cromas pueden definirse como cada uno de los doce tonos distintos de los que se compone la escala musical. Como ya se explicó anteriormente, los doce tonos están dados por el conjunto {C, C#, D, D#, E, F, F#, G, G#, A, A# y B}, al enumerar los valores cromas identificamos este conjunto como $c = \{0, 1, 2, 3, ..., 11\}$ donde $c = 0$ se refiere al croma C (Do, en la notación latina), $c = 1$ al croma C# (Do#, en la notación latina), y así sucesivamente.

Una clase de tono (en inglés pitch class) se define como el conjunto de todos los tonos que comparten el mismo croma. Por ejemplo, la clase tono que corresponde al croma $c = 0$, es decir, C consiste en el conjunto $p = \{0, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120\}$, que serían las notas {C-2, C-1, C0, C1, C2, C3, C4, C5, C6, C7, C8, C9, C10}, en la Figura 2.7(a) se muestra la representación en el teclado de este ejemplo y en la Figura 2.7(b) se muestra un teclado de piano estándar con sus valores de frecuencia por cada nota musical,

además se puede observar a que valor croma y octava pertenecen.

La idea principal de las características croma es agregar toda la información espectral que se refiere a una clase tono en un coeficiente.

A continuación se definen varias fórmulas para cálculos importantes. Supóngase que dos notas tienen frecuencias f_1 y f_2, entonces la fórmula para calcular el número de octavas entre $f1$ y f_2 está dado por la Ecuación (2.7):

$$n_o = |log_2(f_2/f_1)| \tag{2.7}$$

Ahora para dividir la octava en unidades más pequeñas, donde todos los semitonos tienen la misma relación de frecuencia de $2^{1/12}$, la conversión entre el nombre de la nota y la frecuencia es simple. Primero se necesita una nota de referencia y la frecuencia, por lo general se utiliza A4, que tiene una frecuencia central de 440 Hz, entonces para una nota que se encuentra n semitonos más abajo o más arriba en la escala, la frecuencia se obtendría con la Ecuación (2.8).

$$f_n = 2^{n/12} 440 \text{ Hz} \tag{2.8}$$

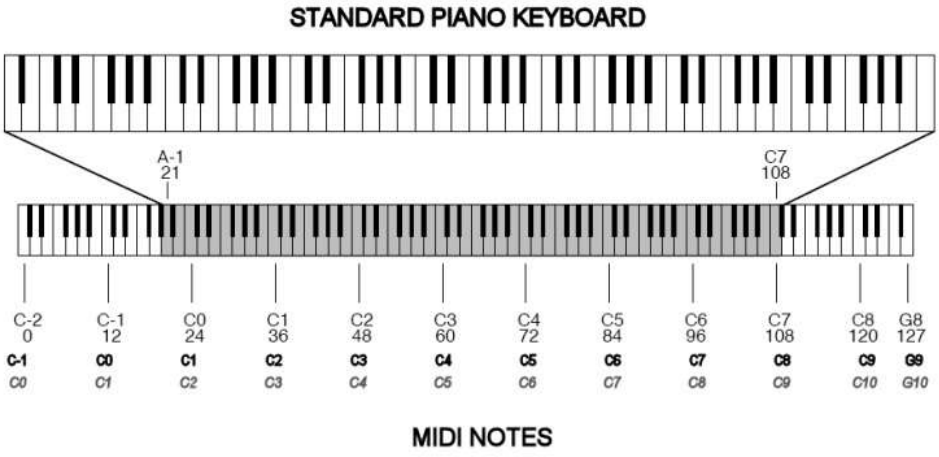
En la música electrónica el tono a menudo está dado por un número MIDI (m), para la nota A4 es de 69, y se incrementa en uno por cada semitono, así que esto nos da una sencilla conversión entre frecuencias y números MIDI (de nuevo utilizando 440 Hz como tono de A4) obtenemos la Ecuación (2.9) y la Ecuación (2.10) :

$$m = 12log_2(f_m/440Hz) + 69 \tag{2.9}$$

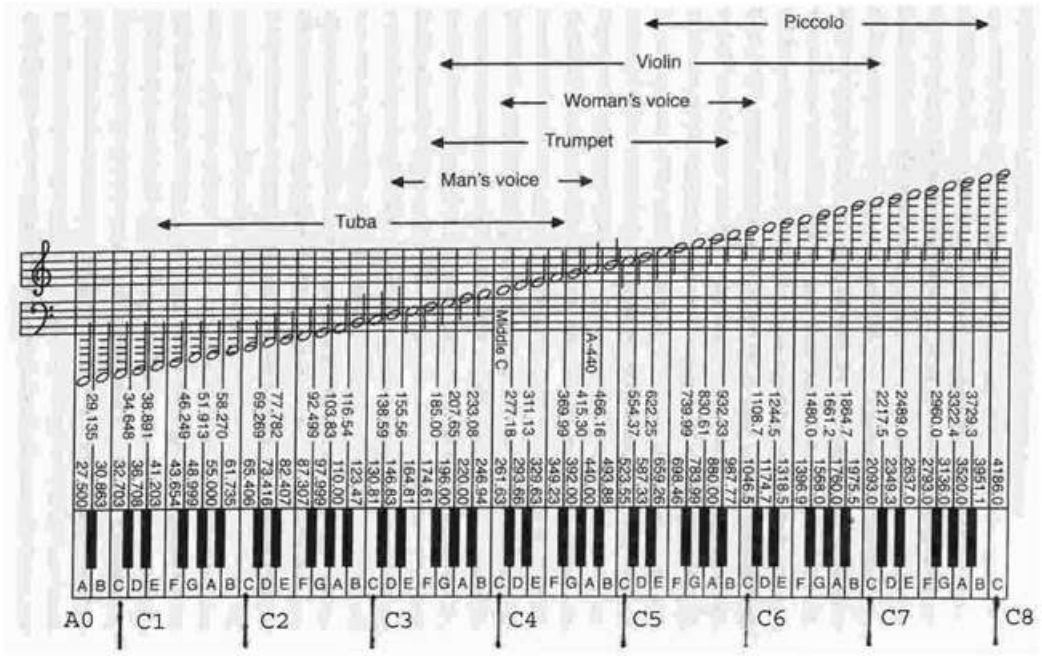
$$f_m = 2^{\frac{m-69}{12}} 440Hz \tag{2.10}$$

2.4. Cromagrama de entropía

Un cromagrama es la representación en el tiempo, de los coeficientes de energía de cada una de las bandas de un banco de filtros adaptado a las octavas de cada una de las doce notas conocidas como cromas.



(a) ***Representación de valores croma en el teclado de un piano***



(b) ***Valores de frecuencia en Hz por cada nota musical de un piano estándar***

Figura 2.7: Valores croma y frecuencias de las notas musicales en un piano.

El proceso para calcular el cromagrama a partir de un archivo de audio es el siguiente:

1. Primero se debe calcular la energía por cada nota de la escala musical.

2. Después se divide el espectrograma de la señal de audio con un detector de ritmo o pulsos, es decir, donde la canción muestra cambios notables, por ejemplo; en el tempo, en la potencia, etc. Los divisores obtenidos se conocen como *onsets*, también se puede optar por dividir el audio en fragmentos de duración fija. De manera más detallada la obtención de un cromagrama se puede hacer a partir del espectrograma obtenido. Como ya se ha mencionado el espectrograma es la representación de la magnitud al cuadrado de la STFT, es por esto que para la obtención del cromagrama, primero se derivan algunas características de la STFT para convertir el eje de frecuencias (dadas en Hz) en un eje correspondiente a tonos musicales. Como se mencionó en la sección anterior, podemos ver estos tonos como en un teclado o piano donde los tonos de la escala corresponden a cada una de las teclas de un piano, en esta escala cada octava se divide en doce unidades logarítmicamente espaciadas. Por ejemplo, en la notación MIDI, se tienen 128 tonos p, que se enumeran comenzando por 0 y terminando con 127. El tono MIDI $p = 69$ corresponde al tono A4 (utilizado generalmente como estándar para afinar instrumentos musicales), el cual tiene una frecuencia central de 440 Hz. En general la frecuencia central F_c de un tono donde $p \in [0 : 127]$ está dada por la Ecuación (2.11).

$$F_{pitch}(p) = 2^{(p-69)/12} 440Hz \tag{2.11}$$

La percepción logarítmica de la frecuencia motiva el uso de una representación tiempo-frecuencia con un eje de frecuencias logarítmicas etiquetadas por tonos de la escala musical. Entonces, para derivar una representación a partir de un espectrograma, la idea es asignar a cada coeficiente espectral $X[m, k]$ el tono p con frecuencia central más cercana a la frecuencia $F_{coef}(k)$, es decir, se define para cada tono $p \in [0 : 127]$ el conjunto $P(p) = \{k \in [0 : K] : F_{pitch}(p-0.5) \leq F_{coef}(k) < F_{pitch}(p+0.5)\}$, a partir de

esto obtenemos un espectrograma log-frecuencia Y_{LF} definido por la Ecuación (2.12),

$$Y_LF(m,p) = \sum_{k \in P(p)} |X(m,k)|^2 \tag{2.12}$$

Según esta definición el eje de frecuencias se divide logarítmicamente y es etiquetado linealmente de acuerdo a los tonos MIDI.

3. Una vez dividido el audio en pequeños fragmentos, se extrae la información de los doce semitonos (cromas) de la escala musical. Es decir, una representación croma puede ser derivada mediante la suma de todos los coeficientes de tono que pertenecen al mismo valor croma, mediante la Ecuación (2.13):

$$C(m,c) = \sum_{p \in [0:127] | p \ mod \ 12 = c} Y_{LF}(m,p) \tag{2.13}$$

para $c \in [0:11]$.

2.4.1. Entropía Cromática

La entropía cromática es una variante de la entropía espectral, pero en lugar de calcular la entropía directamente a partir del módulo de la FFT normalizado, se hace primero una asignación del espectro de potencia que posteriormente se divide en doce sub-bandas, donde la frecuencia central de cada banda coincide con cada uno de los doce semitonos de la escala musical mayor.

El proceso a seguir hasta llegar al cálculo de la entropía cromática se muestra en la Figura 2.8.

Figura 2.8: Proceso para la obtención de la Entropía Cromática

2.4.2. Perfiles de la clase de tono armónicos

Los Perfiles de la clase de tono armónicos (HPCP, del inglés Harmonic pitch class profile), es un vector de características extraídas de una señal de audio de una interpretación musical propuesto por Fujishima [Fujishima99] en el contexto de un sistema de reconocimiento.

Mediante el procesamiento de señales musicales las características HPCP pueden ser usadas para: estimar la tonalidad de una interpretación musical, medir la similitud entre dos interpretaciones musicales, y clasificar la música por compositor, género, etc. Este proceso se relaciona con el análisis de tiempo-frecuencia. La Figura 2.10 muestra el diagrama del procedimiento de extracción de características HPCP y la Figura 2.9 muestra un ejemplo de un Cromagrama.

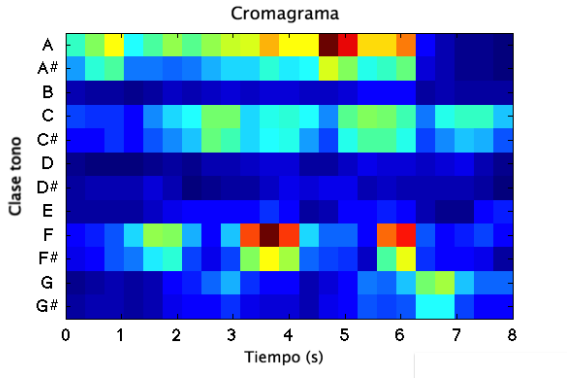


Figura 2.9: Ejemplo de un Cromagrama

2.5. Resumen del capítulo

En este capítulo se describen los fundamentos teóricos para la caracterización de la señal de audio mediante la determinación de espectrogramas y cromagramas, se presenta el proceso para la extracción de características que debe seguirse para la obtención de los mismos, además se presentan y analizan diferentes maneras de llevar a cabo este proceso. Al mismo tiempo que se dan a conocer conceptos básicos para la comprensión del mismo. Después de explicar el proceso que debe llevarse a cabo para la extracción de características

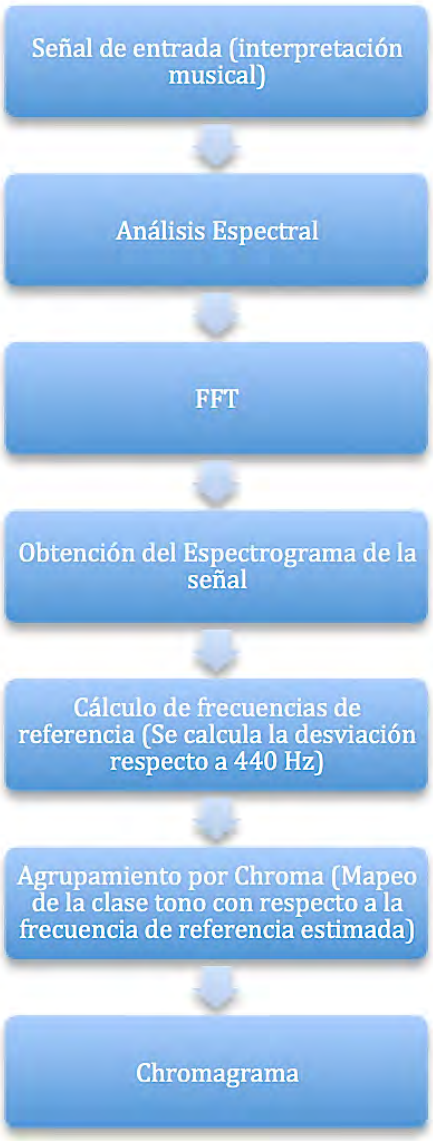


Figura 2.10: Diagrama del proceso para la extracción de características HPCP para la obtención del Chromagrama.

en el capítulo siguiente se puede proceder a explicar las técnicas de alineamiento, para posteriormente pasar a la explicación de la implementación del sistema de evaluación automática de audio.

Capítulo 3

Técnicas de alineamiento

3.1. Doblado dinámico en tiempo

El doblado dinámico en tiempo (DTW, del inglés Dynamic Time Warping) es un algoritmo utilizado en el análisis de series de tiempo, ya que permite medir la similitud entre dos series de tiempo que pueden variar en el tiempo o velocidad. Es por esto que en los últimos años el algoritmo de DTW ha sido utilizado para comparar secuencias de audio, video y gráficos, etc, puesto que este algoritmo puede analizar todos los datos que se puedan convertir en una secuencia lineal. Una de las aplicaciones mas conocidas ha sido la del reconocimiento automático de voz, también se ha utilizado para reconocimiento de firmas en línea, entre otras.

DTW es un método de alineación de secuencias y consiste en calcular una coincidencia óptima entre dos secuencias dadas. Alinear dos series de tiempo $A(i)$, donde $0 \leq i \leq N$, y $B(j)$, donde $0 \leq j \leq M$, es equivalente a encontrar una función $j = d(i)$ que mapea cada índice i de la serie A a un índice j de la serie B. Entonces la función d estará sujeta a las siguientes condiciones de frontera; $d(0) = 0$ y $d(N) = M$, y a ciertas restricciones locales, la más utilizada es la restricción que nos indica que si la trayectoria función óptima pasa por el punto (i, j), entonces paso por el punto $(i-1, j-1)$, $(i-1, j)$ o por el punto $(i, j-1)$ como se muestra en la Figura 3.1. Se impone una penalización de 2 cuando se elige $(i-1, j-1)$ y de 1 cuando se elige $(i-1, j)$ o $(i, j-1)$. Para hacer el

alineamiento temporal de las secuencias o series de tiempo, el eje temporal de la señal a evaluar se comprime y expande para alinear los vectores de características entre patrón y prueba, lo cual produce un camino de alineamiento como se muestra en la Figura 3.2.

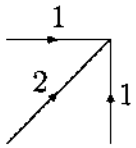


Figura 3.1: Restricción local.

Para la Figura 3.1 se tiene:

$$\left\{ \begin{array}{c} D(i-1,j)+d(i,j) \\ D(i-1,j-1)+2d(i,j) \\ D(i,j-1)+d(i,j) \end{array} \right\}$$

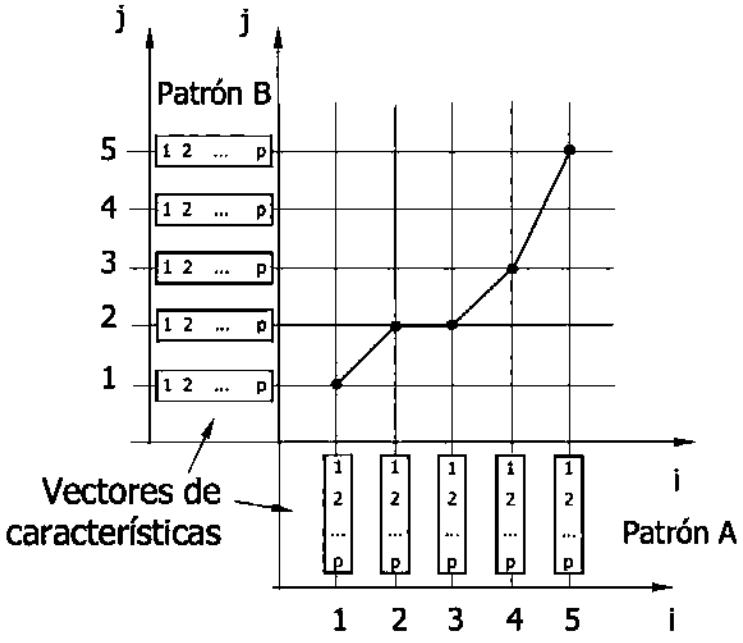


Figura 3.2: Ejemplo de un camino de alineamiento, si la función pasa por (i,j) se compara el vector 'i' de A con el 'j' de B.

Por lo tanto el Doblado Dinámico en Tiempo puede ser utilizado para determinar la distancia entre dos matrices que tengan el mismo número de renglones sin importar que su número de columnas sea diferente. Específicamente hablando de señales de audio, DTW se puede utilizar para obtener la distancia entre dos audios sin importar que la duración en tiempo entre los mismos sea diferente mientras sus vectores característicos tengan el mismo tamaño, para ello se debe poder comparar estos vectores y algunas de las distancias más utilizadas para este fin son: la distancia euclidiana y la distancia coseno.

3.1.1. Distancia Euclidiana

La distancia Euclidiana podemos obtenerla mediante el Teorema de Pitágoras, puesto que como ya se sabe la distancia más corta entre dos puntos es la línea recta, por lo tanto la fórmula para el cálculo de la distancia Euclidiana está dada por la Ecuación (3.1).

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \tag{3.1}$$

3.1.2. Distancia Coseno

Esta distancia es conveniente cuando se necesita que la comparación no dependa de la magnitud de los vectores, sino del ángulo que existe entre ellos, es por esto que es más factible utilizarla en audio, puesto que la magnitud de los vectores se obtiene determinando la energía contenida por cada banda de frecuencia y esta dependerá de algún factor como puede ser; el volumen con el que fue grabado, o en el caso de interpretaciones musicales podrá depender, de la escala o tonalidad en que haya sido grabada, al igual que pudiera afectar el volumen, y no dependerá tanto del tipo de sonido. La distancia coseno determina el ángulo menor que hay entre dos vectores que desean compararse, es por esto que el resultado que se obtenga dependerá de la orientación de los mismos y no de la magnitud, partiendo de la Ecuación 3.2 donde θ es el ángulo menor entre los vectores a y b.

$$a \cdot b = |a||b|cos\theta \tag{3.2}$$

entonces,

$$cos\theta = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}} \tag{3.3}$$

La Ecuación 3.3 devuelve en realidad una medida de similitud, la medida de similitud se acerca a cero mientras menos parecidos sean los vectores, siendo 1.0 el valor mayor que puede devolver. Por lo tanto mediante la ecuación anterior podemos obtener finalmente la distancia con la Ecuación 3.4.

$$d(a,b) = 1 - |cos\theta| = 1 - \left| \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}} \right| \tag{3.4}$$

A continuación se muestra el algoritmo de DTW, en donde la función $d()$ puede obtenerse con alguna de las distancias estudiadas en esta sección; la distancia Euclidiana o la distancia Coseno.

```
entrada: Cadena1, Cadena2.
salida   : DTW[n][m]
para i ← 0 hasta n hacer
|  DTW[i,0] ← i
fin
para j ← 0 hasta m hacer
|  DTW[0,j] ← j
fin
para i ← 1 hasta n hacer
|  para j ← 1 hasta m hacer
|  |  costo = d(s[i], t[j]);
|  |  DTW[i,j] = costo + minimum(
|  |  DTW[i-1,j] ;                          // inserción
|  |  DTW[i,j-1] ;                        // eliminación
|  |  DTW[i-1,j-1] ;                      // sustitución
|  |  )
|  fin
fin
devolver DTW[n][m];
```

Algoritmo 1: DTW

3.2. Técnicas de procesamiento de cadenas o secuencias biológicas (DNA)

3.2.1. Distancia de Levenshtein

La distancia de Levenshtein fue propuesta por el científico ruso Vladimir Levenshtein en [Levenshtein66]. Esta distancia se utiliza para la comparación de cadenas ya que se obtiene el número mínimo de operaciones que se requieren para transformar una cadena de caracteres en otra, las operaciones que pueden realizarse son: inserción, sustitución y eliminación de un caracter. Por ejemplo, suponga que se desea conocer la distancia entre las cadenas “hola” y “hello”. La distancia es 3, puesto que se requiere de 3 operaciones para

transformar la primer cadena en la segunda.

- Sustitución de 'o' por 'e': "hola" - "hela".
- Inserción de 'l' entre 'l' y 'a': "hela" - "hella".
- Sustitución de 'a' por 'o': "hella" - "hello".

A continuación se muestra el algoritmo para obtener la distancia de Levenshtein entre dos cadenas.

entrada: Cadena$_1$, Cadena$_2$.
salida : d[longitudCad1][longitudCad2]
para $i \leftarrow 0$ **hasta** *longitudCad1* **hacer**
| d[i,0] ← i
fin
para $j \leftarrow 0$ **hasta** *longitudCad2* **hacer**
| d[0,j] ← j
fin
para $i \leftarrow 1$ **hasta** *longitudCad1* **hacer**
| **para** $j \leftarrow 1$ **hasta** *longitudCad2* **hacer**
| | **si** *Cadena$_1$ [i] = Cadena$_2$ [j]* **entonces** costo = 0 ;
| | **de lo contrario** costo = 1 ;
| | d[i,j] = minimum(
| | d[i-1,j] + 1 ; `// eliminación`
| | d[i,j-1] + 1 ; `// inserción`
| | d[i-1,j-1] + 1 ; `// sustitución`
| |)
| **fin**
fin
devolver d[longitudCad1][longitudCad2];

Algoritmo 2: Algoritmo de Levenshtein

3.2.2. LCS

La distancia LCS (del inglés Longest Common Subsequence) consiste en encontrar la subsecuencia común más larga entre dos cadenas o secuencias de caracteres, por lo tanto dadas dos cadenas se debe encontrar la longitud de la cadena mas larga.

Si se tienen dos sequencias definidas como: $X = (x_1, x_2, ..., x_m)$ y $Y = (y_1, y_2, ..., y_n)$, entonces $LCS(X, Y)$ representa el conjunto de subsecuencias comunes más largas, este conjunto está dado por:

$$LCS(X,Y) = \begin{cases} 0 & \text{si } i = 0 \text{ o } j = 0 \\ LCS(X_{i-1}, Y_{j-1}) + 1 & \text{si } x_i = y_i \\ max(LCS(X_i, Y_{j-1}), LCS(X_{i-1}, Y_j)) & \text{si } x_i \neq y_i \end{cases}$$

Como ejemplo, supóngase que se tiene la cadena $X = AGCGTAG$ y la cadena $Y = GTCAGA$ y se quiere encontrar la subsecuencia común más larga, entonces se tiene:

- La cadena $X = AGCGTAG$ tiene longitud igual 7.
- La cadena $Y = GTCAGA$ tiene longitud igual a 6.
- Por lo tanto, el valor LCS es 4, ya que el tamaño de la subsecuencia común más larga (LCS = "GCGA") es 4.

El problema de la subsecuencia común más larga tiene importante aplicación en el área de bioinformática, ya que surge la necesidad de comparar secuencias de cadenas de ADN entre dos o más organismos.

3.3. Resumen del capítulo

En este capítulo se describren fundamentos teóricos para el alineamiento de audio, ya que se muestran distintas técnicas para la comparación de audio, desde el algoritmo de doblado dinámico en tiempo (DTW), hasta el uso de técnicas de comparación de cadenas

que también pueden ser aplicadas en la comparación de audio, tales como: distancia de Levenshtein y distancia LCS, además se explican algunos conceptos importantes sobre este proceso de alineamiento, después de explicar el proceso que se debe llevar a cabo para la caracterización de la señal de audio y posteriormente poder llevar a cabo el alineamiento de audio para obtener una medida de comparación entre las señales que desean compararse, se procede en el siguiente capítulo con la descripción de la implementación del sistema de evaluación de estudiantes de música.

Capítulo 4

Descripción del sistema implementado

En este capítulo se describe la implementación del sistema de evaluación automática de estudiantes de música, el cual consta básicamente de; una etapa de preprocesamiento, con el fin de enfatizar la frecuencia mayor de la señal y así mantener una relación constante, una de procesamiento de la señal de audio, para la extracción de las características de la señal, y otra etapa de alineación de audio para la comparación de las interpretación musicales.

4.1. Procesamiento de la señal de audio

Para este trabajo se obtuvo una colección de archivos WAV, cada uno de los cuales contiene la grabación de una pieza musical interpretada ya sea por el maestro o por alguno de los alumnos. Las piezas interpretadas por el maestro son las interpretaciones de referencia, ya que se consideran interpretaciones correctas, y las piezas interpretadas por los estudiantes son las interpretaciones de prueba que serán evaluadas. Para poder hacer una evaluación el sistema toma la interpretación de referencia y la compara con la interpretación del estudiante.

Para la etapa de preprocesamiento de la señal de audio se realizaron los siguientes

pasos:

1. Obtención de la señal audio, a partir de la interpretación musical del archivo WAV.

2. Eliminación de la componen CD, para lo cual se calcula el promedio de la señal y se resta a la misma.

3. Aplicación de un filtro de préenfasis, el cual está dado por la Ecuación (4.1)

$$y(n) = x(n) + 0.95x(n-1) \tag{4.1}$$

Para la etapa de la extracción de características mediante el procesamiento de la señal de audio se realizaron los siguientes pasos:

1. División de la señal de audio en marcos de cierta duración de tiempo en milisegundos (ms) con cierto traslape entre los mismos y posteriormente la aplicación de algún tipo de ventana. En este trabajo se utilizaron marcos de 92 ms aproximadamente con un traslape del %50 y se aplicó una ventana de Hamming a cada uno de los marcos.

2. Aplicación de la transformada de Fourier, para lo cual se utilizó la transformada rápida de Fourier (FFT del inglés Fast Fourier Transform), puesto que al implementar la STFT, en un programa de computadora, lo mejor es utilizar la FFT para el cálculo de la transformada de Fourier, ya que vuelve mas rápido el proceso de extracción de las características.

3. Determinación de espectogramas y cromagramas, a partir de la obtención de los vectores característicos.

La Figura 4.1 muestra un diagrama del proceso para la extracción de características, anteriormente descrito.

4.1.1. Determinación de los espectrogramas

Para poder determinar el espectrograma de una interpretación musical, primero la señal de audio que se obtiene a partir de los archivos WAV (muestreada a 44100 Hz) se

dividide en marcos o tramas, los cuales son de aproximadamente 92ms con un traslape de 50 %, a cada uno de los marcos se les aplica una ventana de hamming:

$$w_h(n) = 0.54 + 0.46cos\left(\frac{2\pi n}{N}\right) \tag{4.2}$$

donde n se refiere a la muestra, y N es al tamaño de ventana.

El tamaño de ventana es de $N = 4096$, es por ello que se puede calcular la duración en ms de cada marco mediante el siguiente cálculo:

$$t = N/Fs$$

Sabemos que $Fs = 44100$ Hz y que $N = 4096$ por tanto,

$$t = 4096/44100$$

$$t = 0.09287$$

$$t = 92.87ms$$

Y si se tiene un traslape del 50 % quiere decir que se tiene un vector de características cada 46 ms aproximadamente.

Posteriormente se aplica la STFT, por lo que tenemos:

$$X[m,k] = \sum_{n=0}^{N-1} x[n + mH]w_h[n]e^{\frac{-j2\pi nk}{N}} \tag{4.3}$$

donde $X[m,k]$ denota el k-ésimo coeficiente de Fourier para el m-ésimo marco de tiempo, x es la señal discreta obtenida por las muestras equidistantes, tomadas de la señal continua, según la frecuencia de muestreo F_s dada en Hz, w_h es una ventana de tiempo discreto de longitud N y H es el tamaño de salto.

La STFT puede obtenerse con el cálculo de la FFT para la extracción de los vectores característicos de cada uno de los marcos de la señal, de otra manera la extracción de características se puede volver un proceso muy lento.

Después se obtiene la potencia a partir de la magnitud cuadrada de la transformada rápida de Fourier, para cada uno de los marcos, es decir para el m-ésimo marco se tiene:

$$Potencia[k] = Parte_{real}(X[k])^2 + Parte_{imag}(X[k])^2 \tag{4.4}$$

donde k es el késimo coeficiente, además $k = 0, 1, 2, 3, ..., K$, donde $K = N/2$.

Una vez obtenida la potencia, se debe hacer la suma de los coeficientes de Fourier para cada una de las bandas de Bark, por lo tanto para poder obtener la frecuencia dada en Hertz de cada uno de los coeficientes se tiene la siguiente ecuación:

$$F_{coef}(k) = \frac{kF_s}{N} \tag{4.5}$$

donde F_s es la frecuencia de muestreo fija, para los archivos WAV $F_s = 44100$ Hz y N es la longitud de ventana que en este caso es de 4096. Al obtener la frecuencia en Hz de cada coeficiente se puede conocer a que banda de Bark corresponde, en este trabajo se utilizan 18 bandas de Bark con los siguientes valores en Hz: 0, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400. Finalmente al conocer a que banda de Bark corresponde cada uno de los coeficientes, se puede realizar la suma de las potencias obtenidas para cada k y así obtener la energía por bandas de Bark.

Por lo tanto el espectrograma está dado por la Ecuación (4.6):

$$espectrograma = |X[k]|^2 \tag{4.6}$$

Para poder apreciar mejor el espectrograma conviene expresar la magnitud de los coeficientes en decibeles, es decir:

$$20log_{10}(|X[k]|) \tag{4.7}$$

El algoritmo 3, recibe cada uno de los marcos de la señal de audio y devuelve el vector característico del mismo, que posteriormente servirá para la construcción del espectrograma. La Figura 4.2 muestra como ejemplo el espectrograma obtenido para el archivo chun c.wav.

Figura 4.1: Proceso para la extracción de características.

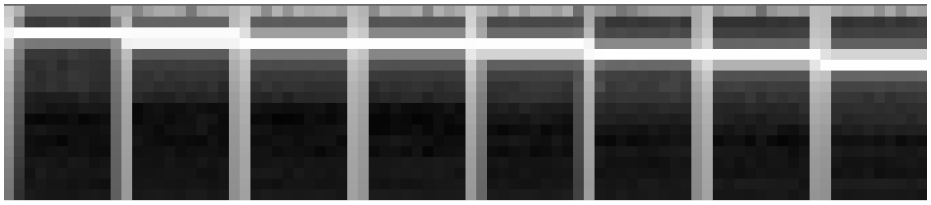


Figura 4.2: Ejemplo del espectrograma de la escala musical.

entrada: real[N] , imag[N] , sampleRate

salida : Esp[i]

N ← real.length;

Bark[ ] ← 0,100, 200, 300, 400, 510, 630, 770, ..., 4400 ;

fk ← 0.0;

para $i \leftarrow 0$ **hasta** *N/2* **hacer**
| P[i] ← real[i] real[i] + imag[i] imag[i] ;

fin

para $i \leftarrow 0$ **hasta** *18* **hacer**
| energia ← 0.0;
| **para** $k \leftarrow 0$ **hasta** *N/2* **hacer**
| | fk ← $k * sampleRate/N$;
| | **si** *fk* $\geq$ *Bark[i]* & *fk* $\leq$ *Bark[i+1]* **entonces** energia + = P[k] ;
| **fin**
| Esp[i] ← 20∗ log10(energia);

fin

devolver Esp[i];

Algoritmo 3: Determinación de vectores característicos del espectrograma.

4.1.2. Determinación de los cromagramas

Se puede decir que es una variante del proceso para la determinación del espectrograma, pero a diferencia de este último para la determinación de los cromagramas debe extraerse por cada marco un vector de entropía por chroma (d), por lo que en lugar de tener 25 sub-bandas de la escala de Bark se tienen 12 sub-bandas, cada una estas sub-bandas corresponden a cada uno de los 12 semitonos de la escala musical: C, C#, D, D#, E, F,

F#, G, G#, A, A#, B.

Para llevar a cabo este proceso se debe conocer la frecuencia de muestreo (F_s) que como ya sea ha mencionado es $Fs = 44100Hz$ para los archivos WAV, el tamaño de ventana que corresponde a la variable N en este caso de $N = 4096$. Con esto, al igual que en la determinación de los espectrogramas, se puede saber que se tienen marcos de aproximadamente 92 ms con un traslape del 50 %, lo cual quiere decir que aproximadamente cada 46 ms se tiene un vector croma. Además se debe tener una frecuencia inicial f_0 y una frecuencia máxima f_max, y un valor para K el cual depende de la frecuencia máxima que se tenga.

Se obtiene fk primero para la frecuencia inicial f_0 mediante la siguiente ecuación:

$$fk = f_0 * 2^{\frac{d}{b}} \tag{4.8}$$

Donde d tiene un valor de 0, 1, 2, 3,4... K, y $K = *blog_2(\frac{f_max}{f_0})$, y b tiene un valor de 12, debido a los 12 cromas que representan los semitonos de la escala musical.

Posteriormente de la Ecuación (4.7) se despeja k y se obtiene:

$$k = \frac{fk * N}{Fs}$$

De esta manera se puede calcular k y se obtiene el valor de la siguiente frecuencia fk hasta llegar a la frecuencia máxima, para poder obtener la sumatoria de los coeficientes espectrales por chroma y así obtener el vector chroma por cada uno de los marcos. El algoritmo 4 describe este proceso y la Figura 4.3 muestra el ejemplo de un chromagrama obtenido con el sistema implementado.

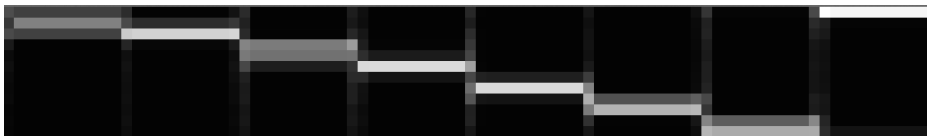


Figura 4.3: Ejemplo del chromagrama del archivo EscalaMusical.wav, el cual muestra las notas Do, Re, Mi, Fa, Sol, La, Si, Do, el eje vertical corresponde a los 12 chromas mientras que el horizontal al tiempo dado en ms.

entrada: real[N] , imag[N] , sampleRate

salida : Chromas[i]

N ← real.length;

b ← 12;

p ← 0;

f $_0$ ← 261;

f $_{max}$ ← 8372;

$f_{sup} \leftarrow f_0 * 2^{1.0/b}$;

k $_1$ ← f_0* N / sampleRate ;

k $_2$ ← f_{sup}* N / sampleRate ;

k $_{fin}$ ← f_{max}* N / sampleRate ;

mientras $k_2 \leq k_f in$ **hacer**

  para $k \leftarrow k_1$ **hasta** $k \leq k_2$ **hacer**
    Chromas[p % b] += (real[k] * real[k] + imag[k] * imag[k] ;

  fin

  p++;

  $k_1 \leftarrow k_2$

  $f_{sup} \leftarrow f_0$* 2 $^{(p+1.0)/b}$;

  k $_2$ ← f_{sup}* N / sampleRate ;

fin

devolver Chromas[i];

Algoritmo 4: Determinación de vectores característicos del Chromagrama.

4.2. Alineamiento de audio

Para este trabajo de investigación se utilizó el algoritmo de DTW para el proceso de alineamiento, para lo cual se tuvo que hacer la modificación de guardar los valores en un solo vector y no en una matriz, puesto que si se utilizaba la matriz completa, la memoria de la máquina se terminaba, esto debido a que los archivos de audio son muy grandes.

entrada: melodia$_1$[nr][nc] , melodia$_2$[nr][nc]

salida : distancia

nr ←melodia$_1$.length;

nc ←melodia$_2$.length;

per $i \leftarrow 1$ **a** $i \leq nr{-}1$ **fai**
| D[i] ← DistanciaCos(C(melodia$_1$, i),C(melodia$_2$, 0)) + D[i-1];

fine

per $j \leftarrow 1$ **a** $j \leq nc{-}1$ **fai**
| VectorAux[j] ← DistanciaCos(C(melodia$_1$, 0),C(melodia$_2$, j)) + D[j-1];

fine

per $i \leftarrow 1$ **a** $i \leq nr{-}1$ **fai**

per $j \leftarrow 1$ **a** $k \leq nc{-}1$ **fai**

aux ← D[i-1];

D[i-1] ← VectorAux[j];

dij ← DistanciaCos(C(melodia$_1$, i),C(melodia$_2$, j));

D[i] ← min(2.0 ∗ dij + aux, // 2.0 ∗ dij + di1j1

dij + D[i − 1], // dij + dij1

dij + D[i]); // dij + di1j

VectorAux[j] ← D[i];

fine

fine

distancia ← D[nr − 1]/(nr+nc);

devolver distancia;

Algoritmo 5: DTW para la comparación de dos interpretaciones musicales.

El algoritmo 5 describe el algoritmo de alineamiento de doblado dinámico en el tiempo utilizando para la comparación de dos interpretaciones musicales, de las cuales la interpretación del profesor es tomada como referencia y la interpretación realizada por el estudiante es la que se evalúa. Este algoritmo nos entrega una distancia al utilizar el método de la distancia Euclidiana (algoritmo 6) y otra distancia al utilizar el método de distancia Coseno (Algoritmo 7), los cuales posteriormente en el Capítulo 5 de Resultados serán comparados, además la distancia entregada por el Algoritmo 5 es la que nos servirá pa-

ra finalmente obtener la calificación correspondiente a la interpretación realizada por el alumno.

entrada: vector1[i], $vector_2$[i]

salida : distancia

N $\leftarrow$ vector1.*length*; *suma*$\leftarrow$ 0;

per $i \leftarrow 0$ **a** $i \leq N{-}1$ **fai**
| suma += (vector_1[i] − vector_2[i]) ∗ (vector_1[i] − vector_2[i]);

fine

distancia $\leftarrow$ Math.sqrt(suma);

devolver distancia;

Algoritmo 6: Distancia Euclidiana.

entrada: vector1[i], $vector_2$[i]

salida : distancia

N $\leftarrow$ vector1.*length*; $suma_1 \leftarrow 0.0$;

$\text{suma}_2 \leftarrow 0.0$;

per $i \leftarrow 0$ **a** $i \leq N{-}1$ **fai**
| mult += vector_1[i] ∗ vector_2[i]; suma_1 += vector_1[i] ∗ vector_1[i];
| suma_2 += vector_2[i] ∗ vector_2[i];

fine

$\text{suma}_1 \leftarrow$ Math.sqrt(suma_1);

$\text{suma}_2 \leftarrow$ Math.sqrt(suma_2);

distancia $\leftarrow$ 1.0 − Math.abs(mult / (suma_1∗ suma_2));

si *distancia* $\leq$ *0.0* **entonces** return 0.0; ;

en otro caso
| return distancia;

fin

devolver distancia;

Algoritmo 7: Distancia Coseno.

4.3. Resumen del capítulo

Este capítulo se expone la descripción del sistema propuesto, se explica a detalle la implementación de la etapa de procesamiento de la señal de audio para la extracción de

las características y determinación de los espectrogramas y chromagramas y se muestran los algoritmos desarrollados para este fin. También se hace la explicación de la etapa de alineamiento de audio para la obtención de la medida de distancia que permitirá comparar que tanto se parecen las interpretaciones de los estudiantes con las interpretaciones de los profesores, finalmente en el siguiente capítulo se muestran los resultados de los experimentos realizados con la colección de archivos obtenida de estudiantes y profesores de música, y las comparaciones realizadas entre los resultados obtenidos.

Capítulo 5

Resultados

En este capítulo se muestran los resultados obtenidos a partir de los experimentos realizados con el sistema de evaluación de estudiantes de música.

La colección de archivos WAV contiene; interpretaciones por maestros, que son las interpretaciones que sirvieron como referencia e interpretaciones por alumnos, que son las interpretaciones que se tomaron como las piezas musicales a ser evaluadas. Primero se obtuvo una evaluación por los mismos maestros de música para así poder comparar con los resultados obtenidos por el sistema implementado.

La Tabla 5.1 muestra los resultados según la evaluación de profesores, la primera columna es la interpretación de referencia, la segunda columna es la interpretación realizada por el alumnno y la tercera columna es la evaluación que puede ser una de las siguientes 5 opciones: excelente (10), bien (8-9), regular (6-7), mal (5), muy mal (menos de 5).

En este capítulo se muestran los espectrogramas y cromagramas obtenidos para cada pieza musical,y los resultados obtenidos mediante DTW con distancia Euclidiana y distancia Coseno.

Tabla 5.1: Resultados de la Evaluación real realizada por el profesor de música.

interpretación de referencia	interpretación alumno	desempeño
chun c	chun c2	excelente
chun c	chun c3	bien
chun c	chun c4	mal
extra c	extra c2	bien
extra c	extra c3	bien
extra c	extra c4	mal
extra c	extra p	regular incompleta
fugaVI c	fugaVI c2	excelente
fugaVI c	fugaVI c3	regular
fugaVI c	fugaVI p	mal incompleta
fur elise c	fur elise c2	bien (tempo: lento)
fur elise c	fur elise c3	bien (tempo: moderado)
Los changuitos c	Los changuitos c2	bien
Los changuitos c	Los changuitos c3	bien (volumen bajo)
microcosmos2 c	microcosmos2 c2	excelente
microcosmos2 c	microcosmos2 c3	muy mal
microcosmos2 c	microcosmos2 c4	bien
minuet bach c	minuet bach ca2	bien
minuet bach c	minuet bach ca3	regular
minuet bach c	minuet bach ca4	bien
minuet bach c	minuet bach ca5	mal (tempo:lento)
minuet bach c	minuet bach Gmayor	mal (G mayor)
pasos sobre la nieve c	pasos sobre la nieve c2	bien
pasos sobre la nieve$_c$	pasos sobre la nieve c3	muy mal
praeludiumI bach c	praeludiumI bach c2	bien
praeludiumI bach c	praeludiumI bach c3	bien
praeludiumVI bach c	praeludiumVI bach c2	excelente
praeludiumVI bach c	praeludiumVI bach c3	regular

5.1. Espectrogramas y cromagramas resultantes

Para cada una de las piezas musicales de la colección de archivos WAV, se obtuvieron tanto el espectrograma como el cromagrama mediante la extracción de los vectores de características al procesar las señales de audio obtenida de cada archivo wav.

Cada uno de los archivos wav fue grabado con una frecuencia de muestreo de 44100 Hz a 16 bits por muestra, y se utilizó un tamaño de marco de aproximadamente 92 ms con un traslape del 50 %, los resultados obtenidos se muestran en el Apéndice A.

Los vectores de características son de tamaño 18 para el espectrograma, puesto que se utilizaron 18 bandas de la escala de Bark, mientras que para el cromagrama son de tamaño 12, ya que se tienen 12 valores cromas, uno por cada semitono de la escala musical.

A continuación se muestran los espectrogramas y cromagramas obtenidos para dos de las piezas musicales utilizadas en este trabajo de investigación, esto con la finalidad de mostrar un ejemplo de los espectrogramas y cromagramas obtenidos, y presentar una comparación gráfica de los mismos.

Para el archivo chun c.wav, chun c1.wav, chun c2.wav, chun c3.wav y chun c4.wav se obtuvieron un total de 174, 162, 165 y 169 vectores respectivamente, los cuales son de longitud 18 para los espectrogramas y 12 para los cromagramas. La Figura 5.1 Muestra los espectogramas de las distintas interpretaciones de la pieza musical Chun mientras que la Figura 5.2 muestra los cromagramas de la misma.

Para el archivo Microcosmos2 c.wav, Microcosmos2 c2.wav., Microcosmos2 c3.wav y Microcosmos2 c4.wav se obtuvieron un total de 199, 197, 198 y 206 vectores respectivamente, los cuales son de longitud 18 para los espectrogramas y 12 para los cromagramas. La Figura 5.1 Muestra los espectogramas de las distintas interpretaciones de la pieza musical Chun mientras que la Figura 5.2 muestra los cromagramas de la misma.

(a) ***Espectrograma Chun c.wav***

(b) ***Espectrograma Chun c2.wav***

(c) ***Espectrograma Chun c3.wav***

(d) ***Espectrograma Chun c4.wav***

Figura 5.1: Espectrogramas de la pieza musical Chun

(a) *Chromagrama Chun c.wav*

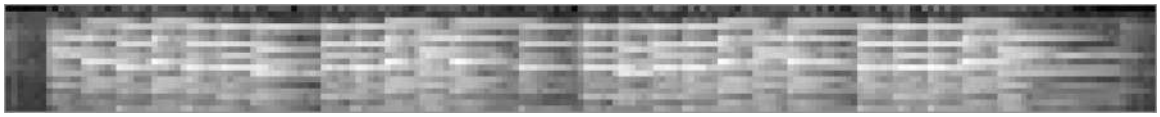
(b) *Chromagrama Chun c2.wav*

(c) *Chromagrama Chun c3.wav*

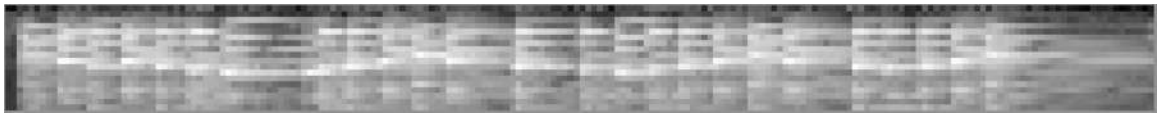
(d) *Chromagrama Chun c4.wav*

Figura 5.2: Chromagramas de la pieza musical Chun

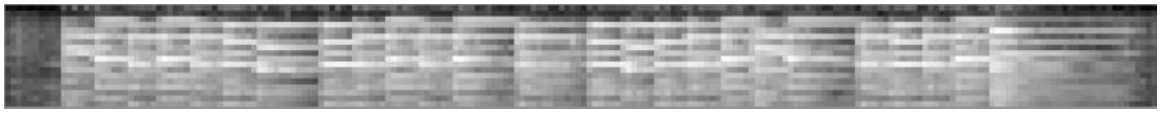
(a) *Espectrograma Microcosmos2 c.wav*



(b) *Espectrograma Microcosmos2 c2.wav*



(c) *Espectrograma Microcosmos2 c3.wav*

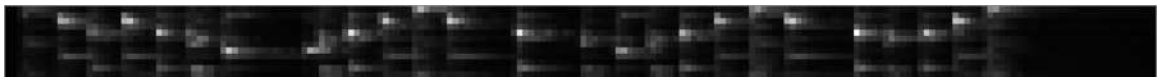


(d) *Espectrograma Microcosmos2 c4.wav*

Figura 5.3: Espectrogramas de la pieza musical Microcosmos2

(a) ***Chromagrama Microcosmos2 c.wav***

(b) ***Chromagrama Microcosmos2 c2.wav***



(c) ***Chromagrama Microcosmos2 c3.wav***

(d) ***Chromagrama Microcosmos2 c4.wav***

Figura 5.4: Chromagramas de la pieza musical Microcosmos2

5.2. Distancias obtenidas mediante el proceso de evaluación con DTW

A continuación se muestra la Tabla 5.2 con las distancias obtenidas mediante el proceso de alineamiento para lo cual se utilizó el algoritmo de DTW. En esta tabla se puede ver la comparación de las distancias obtenidas al utilizar espectrogramas contra las obtenidas al utilizar cromagramas, y las distancias obtenidas utilizando distancia euclidiana y distancia coseno, así como las distancias obtenidas de las evaluaciones realizadas por los profesores de música.

Tabla 5.2: Distancias obtenidas por el sistema implementado según el método utilizado para la Evaluación

		Distancia Euclidiana		Distancia Coseno	
interpretación de referencia	interpretación alumno	Comagrama	Espectrograma	Cromagrama	Espectrograma
chun c	chun c2	5.30532	0.043827	0.033127	7.43896E-4
chun c	chun c3	6.23874	0.045489	0.040026	7.76837E-4
chun c	chun c4	9.56110	0.0468498	0.085455	9.20865E-4
extra c	extra c2	6.16048	0.0463029	0.042439	0.001084
extra c	extra c3	6.30861	0.0461043	0.046933	0.001022
extra c	extra c4	8.98106	0.0516959	0.066865	0.001255
extra c	extra p	7.37823	0.0473277	0.049331	0.001030
fugaVI c	fugaVI c2	2.50754	0.0343383	0.017351	6.68010E-4
fugaVI c	fugaVI c3	3.86666	0.0441898	0.060108	0.001136
fugaVI c	fugaVI p	5.85737	0.0676642	0.136987	0.002489
fur elise c	fur elise c2	6.99040	0.0780035	0.071796	0.002446
fur elise c	fur elise c3	4.76281	0.1033444	0.080464	0.004743
Los changuitos c	Los changuitos c2	2.86068	0.0951468	0.049118	0.004887
Los changuitos c	Los changuitos c3	2.31739	0.1109374	0.061850	0.007181
microcosmos2 c	microcosmos2 c2	1.46528	0.0358030	0.01212	7.700100E-4
microcosmos2 c	microcosmos2 c3	4.03831	0.0810402	0.11352	0.0033553
microcosmos2 c	microcosmos2 c4	2.17728	0.0506871	0.05390	0.0016784
minuet bach c	minuet bach ca2	2.54849	0.0619231	0.052133	0.0021304
minuet bach c	minuet bach ca3	2.79525	0.0595723	0.046772	0.0017631
minuet bach c	minuet bachca4	3.10161	0.0578995	0.054706	0.0019675
minuet bach c	minuet bach ca5	4.72303	0.0732151	0.113516	0.0026664
minuet bach c	minuet bach Gmayor	4.79927	0.0857645	0.188152	0.0038098
pasos sobre la nieve c	pasos sobre la nieve c2	1.26283	0.0454651	0.042226	0.0013794
pasos sobre la nieve$_c$	pasos sobre la nieve c3	3.17248	0.0682060	0.115186	0.0024338
praeludiumI bach c	praeludiumI bach c2	3.90725	0.0449442	0.049192	0.0011519
praeludiumI bach c	praeludiumI bach c3	3.58925	0.0470638	0.050343	0.0011588
praeludiumVI bach c	praeludiumVI bach c2	4.40799	0.0375463	0.027161	7.3522831E-4
praeludiumVI bach c	praeludiumVI bach c3	6.02364	0.0506911	0.058050	0.0012462

La Tabla 5.3 muestra una comparación entre las evaluaciones realizadas por el profesor de música y las calificaciones entregadas por el sistema implementado cuando se usan chromagramas y DTW utilizando distancia coseno, la tabla 5.4 muestra la comparación

de los resultados entregados cuando el sistema utiliza espectrogramas y DTW utilizando distancia coseno y la Tabla 5.5 la comparación de los resultados entregados cuando el sistema utiliza chromagramas y DTW utilizando distancia euclidiana, recordando que las evaluaciones realizadas por el profesor pueden ser cualquiera de las siguientes opciones: excelente (10), bien (8-9), regular (6-7), mal (5), muy mal (menos de 5).

Tabla 5.3: Comparación entre los resultados obtenidos por el sistema implementado (cromagramas y DTW-coseno) y las evualuaciones realizadas por el profesor de música.

interpretación de referencia	interpretación alumno	evaluación profesor	evaluación sistema
chun c	chun c2	excelente	10
chun c	chun c3	bien	9
chun c	chun c4	mal	4
extra c	extra c2	bien	9
extra c	extra c3	bien	8
extra c	extra c4	mal	6
extra c	extra p	regular incompleta	8
fugaVI c	fugaVI c2	excelente	10
fugaVI c	fugaVI c3	regular	7
fugaVI c	fugaVI p	mal incompleta	0
fur elise c	fur elise c2	bien (tempo: lento)	6
fur elise c	fur elise c3	bien (tempo: moderado)	5
Los changuitos c	Los changuitos c2	bien	8
Los changuitos c	Los changuitos c3	bien (volumen bajo)	7
microcosmos2 c	microcosmos2 c2	excelente	10
microcosmos2 c	microcosmos2 c3	muy mal	2
microcosmos2 c	microcosmos2 c4	bien	8
minuet bach c	minuet bach ca2	bien	8
minuet bach c	minuet bach ca3	regular	8
minuet bach c	minuet bach ca4	bien	8
minuet bach c	minuet bach ca5	mal (tempo:lento)	2
minuet bach c	minuet bach Gmayor	mal (G mayor)	0
pasos sobre la nieve c	pasos sobre la nieve c2	bien	9
pasos sobre la nieve$_c$	pasos sobre la nieve c3	muy mal	1
praeludiumI bach c	praeludiumI bach c2	bien	8
praeludiumI bach c	praeludiumI bach c3	bien	8
praeludiumVI bach c	praeludiumVI bach c2	excelente	10
praeludiumVI bach c	praeludiumVI bach c3	regular	7

La correlación es una herramienta utilizada en probabilidad y estadística, dice que si se tienen dos variables A y B existe una correlación entre ellas si al aumentar los valors de A también aumentan los valores de B, la correlación indica la fuerza y la dirección

Tabla 5.4: Comparación entre los resultados obtenidos por el sistema implementado(espectrogramas y DTW-coseno) y las evualuaciones realizadas por el profesor de música.

interpretación de referencia	interpretación alumno	evaluación profesor	evaluación sistema
chun c	chun c2	excelente	10
chun c	chun c3	bien	10
chun c	chun c4	mal	10
extra c	extra c2	bien	9
extra c	extra c3	bien	9
extra c	extra c4	mal	7
extra c	extra p	regular incompleta	9
fugaVI c	fugaVI c2	excelente	10
fugaVI c	fugaVI c3	regular	8
fugaVI c	fugaVI p	mal incompleta	5
fur elise c	fur elise c2	bien (tempo: lento)	5
fur elise c	fur elise c3	bien (tempo: moderado)	1
Los changuitos c	Los changuitos c2	bien	1
Los changuitos c	Los changuitos c3	bien (volumen bajo)	0
microcosmos2 c	microcosmos2 c2	excelente	10
microcosmos2 c	microcosmos2 c3	muy mal	3
microcosmos2 c	microcosmos2 c4	bien	5
minuet bach c	minuet bach ca2	bien	5
minuet bach c	minuet bach ca3	regular	5
minuet bach c	minuet bach ca4	bien	5
minuet bach c	minuet bach ca5	mal (tempo:lento)	4
minuet bach c	minuet bach Gmayor	mal (G mayor)	2
pasos sobre la nieve c	pasos sobre la nieve c2	bien	6
pasos sobre la nieve$_c$	pasos sobre la nieve c3	muy mal	5
praeludiumI bach c	praeludiumI bach c2	bien	8
praeludiumI bach c	praeludiumI bach c3	bien	8
praeludiumVI bach c	praeludiumVI bach c2	excelente	10
praeludiumVI bach c	praeludiumVI bach c3	regular	7

de una relación lineal entre dos variables, por lo tanto se considera que dos variables están correlacionadas cuando los valores de una varían sistemáticamente con respecto a los valores de la otra.

Para aplicar la correlación a los resultados de este trabajo se tomo como variable X las calificaciones de los profesores que como ya se sabe pueden ser las siguientes: excelente (10), bien (8-9), regular (6-7), mal (5), muy mal (menos de 5) y como variable Y las calificaciones obtenidas por el sistema, además se cambio la calificación del profesor de 8-9 por su promedio 8.5 y de igual manera la de 6-7 por 6.5, para poder realizar los calculos, la Figura 5.5 muestra los resultados obtenidos de la correlación 0.907(cromagramas utilizando DTW - distancia coseno), la Figura 5.6 los resultados obtenidos de la correlación $r = 0.462$(espectrogramas utilizando DTW - distancia coseno) y la Figura 5.7 los resultados

Tabla 5.5: Comparación entre los resultados obtenidos por el sistema implementado(cromagramas y DTW-euclidiana) y las evualuaciones realizadas por el profesor de música.

interpretación de referencia	interpretación alumno	evaluación profesor	evaluación sistema
chun c	chun c2	excelente	10
chun c	chun c3	bien	9
chun c	chun c4	mal	5
extra c	extra c2	bien	9
extra c	extra c3	bien	9
extra c	extra c4	mal	6
extra c	extra p	regular incompleta	7
fugaVI c	fugaVI c2	excelente	10
fugaVI c	fugaVI c3	regular	9
fugaVI c	fugaVI p	mal incompleta	7
fur elise c	fur elise c2	bien (tempo: lento)	5
fur elise c	fur elise c3	bien (tempo: moderado)	4
Los changuitos c	Los changuitos c2	bien	10
Los changuitos c	Los changuitos c3	bien (volumen bajo)	9
microcosmos2 c	microcosmos2 c2	excelente	10
microcosmos2 c	microcosmos2 c3	muy mal	7
microcosmos2 c	microcosmos2 c4	bien	9
minuet bach c	minuet bach ca2	bien	8
minuet bach c	minuet bach ca3	regular	9
minuet bach c	minuet bach ca4	bien	7
minuet bach c	minuet bach ca5	mal (tempo:lento)	6
minuet bach c	minuet bach Gmayor	mal (G mayor)	5
pasos sobre la nieve c	pasos sobre la nieve c2	bien	10
pasos sobre la nieve$_c$	pasos sobre la nieve c3	muy mal	8
praeludiumI bach c	praeludiumI bach c2	bien	8
praeludiumI bach c	praeludiumI bach c3	bien	8
praeludiumVI bach c	praeludiumVI bach c2	excelente	8
praeludiumVI bach c	praeludiumVI bach c3	regular	6

obtenidos de la correlación $r = 0.643$(cromagramas utilizando DTW - distancia euclidiana), analizando las figuras se observa que los mejores resultados se obtuvieron utilizando cromagras y como algoritmo de alineamiento DTW mediante el cálculo de la distancia coseno.

calificación profesor	calificación sistema			
x	y	x^2	y^2	xy
10	10	100	100	100
8.5	9	72.25	81	76.5
5	5	25	25	25
8.5	9	72.25	81	76.5
8.5	8	72.25	64	68
5	6	25	36	30
6.5	8	42.25	64	52
10	10	100	100	100
6.5	7	42.25	49	45.5
2.5	0	6.25	0	0
8.5	8	72.25	64	68
10	10	100	100	100
2.5	2	6.25	4	5
8.5	8	72.25	64	68
8.5	8	72.25	64	68
6.5	8	42.25	64	52
8.5	8	72.25	64	68
5	2	25	4	10
5	0	25	0	0
8.5	9	72.25	81	76.5
2.5	2	6.25	4	5
8.5	8	72.25	64	68
8.5	8	72.25	64	68
10	10	100	100	100
6.5	7	42.25	49	45.5
Σx	Σy	Σx^2	Σy^2	Σxy
178.5	170	1410.25	1390	1375.5

82.31% de las variaciones de las 'y' se explican mediante la relación lineal con x

Sxx	
135.76	
Syy	
234	
Sxy	
161.7	
r	
0.907	
100r^2=100(0.921)^2=84.81	
82.31	

Figura 5.5: Correlación obtenida para los cromagramas utilizando DTW con distancia coseno.

calificación profesor	calificación sistema				
x	y	x^2	y^2	xy	
10	10	100	100	100	Sxx
8.5	10	72.25	100	85	135.76
5	10	25	100	50	Syy
8.5	9	72.25	81	76.5	215.04
8.5	9	72.25	81	76.5	Sxy
5	7	25	49	35	78.98
6.5	9	42.25	81	58.5	r
10	10	100	100	100	0.462
6.5	8	42.25	64	52	
2.5	5	6.25	25	12.5	100r^2=100(0.462)^2=21.37
8.5	0	72.25	0	0	21.37
10	10	100	100	100	
2.5	3	6.25	9	7.5	
8.5	5	72.25	25	42.5	
8.5	5	72.25	25	42.5	
6.5	5	42.25	25	32.5	
8.5	5	72.25	25	42.5	
5	4	25	16	20	
5	0	25	0	0	
8.5	6	72.25	36	51	
2.5	5	6.25	25	12.5	
8.5	8	72.25	64	68	
8.5	8	72.25	64	68	
10	10	100	100	100	
6.5	7	42.25	49	45.5	
Σx	Σy	Σx^2	Σy^2	Σxy	
178.5	168	1410.25	1344	1278.5	

21.37% de las variaciones de las 'y' se explican mediante la relación lineal con x

Figura 5.6: Correlación obtenida para los espectrogramas utilizando DTW con distancia coseno.

calificación profesor	calificación sistema			
x	y	x^2	y^2	xy
10	10	100	100	100
8.5	9	72.25	81	76.5
5	5	25	25	25
8.5	9	72.25	81	76.5
8.5	9	72.25	81	76.5
5	6	25	36	30
6.5	7	42.25	49	45.5
10	10	100	100	100
6.5	9	42.25	81	58.5
2.5	7	6.25	49	17.5
8.5	10	72.25	100	85
10	10	100	100	100
2.5	7	6.25	49	17.5
8.5	9	72.25	81	76.5
8.5	8	72.25	64	68
6.5	9	42.25	81	58.5
8.5	7	72.25	49	59.5
5	6	25	36	30
5	5	25	25	25
8.5	10	72.25	100	85
2.5	8	6.25	64	20
8.5	8	72.25	64	68
8.5	8	72.25	64	68
10	8	100	64	80
6.5	6	42.25	36	39
Σx	Σy	Σx^2	Σy^2	Σxy
178.5	200	1410.25	1660	1486

Sxx	
135.76	
Syy	
60	
Sxy	
58	
r	
0.643	
100r^2=100(0.643)^2=41.30	
41.30	

Figura 5.7: Correlación obtenida para los cromagramas utilizando DTW con distancia euclidiana.

5.3. Resumen del capítulo

En este capítulo se muestran los resultados obtenidos de los experimentos realizados con el sistema de evaluación automática de estudiantes de música, y se comparan los resultados obtenidos entre los diferentes esquemas utilizados, los cuales fueron: espectrogramas utilizando DTW con distancia Euclidiana, espectrogramas utilizando DTW con distancia Coseno, cromagramas utilizando DTW con distancia euclidiana y cromagramas utilizando DTW con distancia coseno, para evaluar la eficiencia del sistema se hizo el calculo de la correlación, finalmente en este capítulo se demuestra que los mejores resultados fueron obtenidos mediante el uso de cromagramas y utilizando el algoritmo DTW con distancia coseno. Una vez presentados los resultados el siguiente capítulo expone las conclusiones obtenidas a partir de los mismos y los posibles trabajos futuros.

Capítulo 6

Conclusiones y Trabajos Futuros

6.1. Conclusiones Generales

La evaluación de interpretaciones musicales es un problema que requiere de la extracción de los vectores de características de la señal de audio correspondiente a determinada pieza musical para poder compararse con la pieza musical de referencia mediante alguna técnica de alineamiento y así poder obtener una calificación.

En este trabajo de investigación se implementan dos maneras de presentar los vectores característicos de las interpretaciones musicales, los espectrogramas y los cromagramas, para lo cual se trabajó con 38 interpretaciones musicales, para cada una de las cuales se obtuvo tanto su espectrograma como su cromagrama.

Para poder realizar la alineación de las interpretaciones con los 38 espectrogramas y chromagramas que se obtuvieron, 9 espectrogramas y 9 cromagramas correspondían a las interpretaciones realizadas por profesores de música y fueron tomadas como interpretaciones de referencia, mientras que las 27 restantes son interpretaciones realizadas por estudiantes de música y son las que fueron evaluadas.

Para poder evaluar el rendimiento del sistema se compararon los resultados obtenidos por el sistema contra los resultados reales obtenidos por la evaluación de los profesores de música, además e obtuvo la correlación de los resultados más eficientes obtenidos por el sistema, los cuales fueron los obtenidos con el uso de cromagramas y como algoritmo de

alineamiento DTW con distancia Coseno.

6.2. Conclusiones Específicas

- A partir de los resultados obtenidos y presentados en el Capítulo 5, se puede observar que resulta más eficiente el utilizar cromagramas para la evaluación de las interpretaciones musicales. Esto se debe a que los cromagramas se construyen siguiendo la escala musical. Al momento de visualizar un cromagrama se puede observar como casi logran verse las notas musicales como en un pentagrama musical, mientras que los espectogramas son más usados para la alineación de voz como en reconocimiento de voz.

- En cuanto al alineamiento se pudo observar que los resultados más factibles se obtuvieron al utilizar el calculo de la distancia Coseno en el algoritmo de DTW, puesto que la distancia coseno no depende de la magnitud de los vectores y para la comparación en interpretaciones musicales es lo que se requiere, por tal motivo para obtener la calificación final de las interpretaciones los valores que se utilizaron fueron los obtenidos mediante DTW con distancia coseno.

- Al evaluar el sistema mediante el cálculo de la correlación se obtuvo que los resultados más eficientes fueron los obtenidos utilizando chromagramas y DTW mediante el uso de la distancia Coseno, se obtuvo un valor $r = 0.921$ lo cual indica que se tiene un porcentaje de 84.81 %, por tanto quiere decir que el 84.81 % de las variaciones de las 'y' se explican mediante la relación lineal con 'x'.

6.3. Trabajos Futuros

Algunos de los posibles trabajos futuros para este trabajo de investigación pudieran ser los siguientes:

1. Podrían utilizarse otros métodos para la alineación de audio como los mencionados en el Capítulo 3, como la Distancia de Levenshtein o los Modelos Ocultos de Markov, para ver si se obtienen mejores resultados que los obtenidos en este trabajo.

2. Se pudiera modificar este sistema para que las pruebas que realice pudieran hacerse en tiempo real, y que mientras el alumno toca el sistema pudiera ir evaluando e incluso obtener los errores en el momento que se equivoca, para que al alumno al terminar su interpretación pudiera obtener su calificación y conocer cuales fueron sus errores.

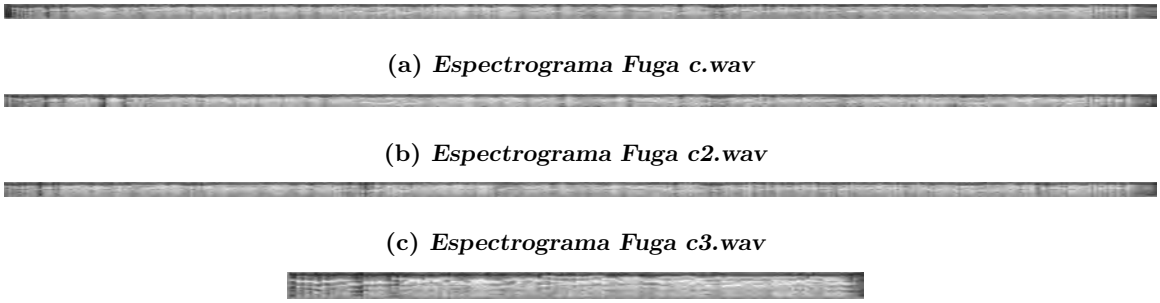
3. Este trabajo pudiera modificarse para poder evaluar no solo interpretaciones con un instrumento, sino interpretaciones con varios alumnos tocando distintos instrumentos al mismo tiempo.

Apéndice A

Espectrogramas y cromagramas

En este apéndice se presentan los espectrogramas y chromagramas de cada una de las interpretaciones musicales utilizadas durante los experimentos realizados en este trabajo de investigación.

A.1. Espectrogramas



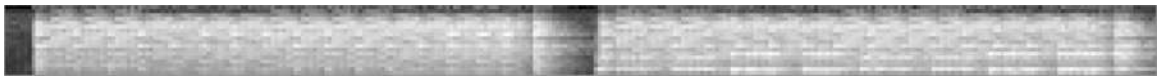
(a) ***Espectrograma Fuga c.wav***

(b) ***Espectrograma Fuga c2.wav***

(c) ***Espectrograma Fuga c3.wav***

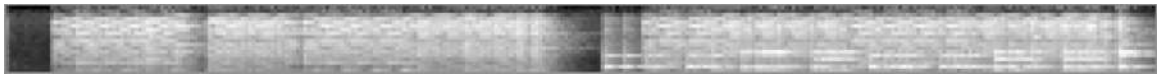
(d) ***Espectrograma Fuga p.wav***

Figura A.1: Espectrogramas de la interpretación musical Fuga VI



(a) ***Espectrograma Extra c.wav***

(b) ***Espectrograma Extra c2.wav***



(c) ***Espectrograma Extra c3.wav***

(d) ***Espectrograma Extra c4.wav***

(e) ***Espectrograma Extra p.wav***

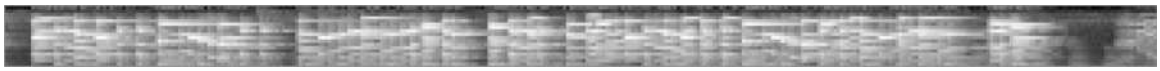
Figura A.2: Espectrogramas de la interpretación musical Extra

(a) ***Espectrograma LosChanguitos c.wav***

(b) ***Espectrograma LosChanguitos c2.wav***

Figura A.3: Espectrogramas de la interpretación musical Los Changuitos

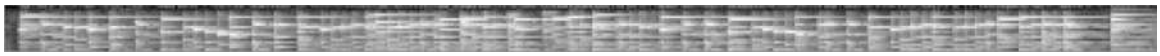
(a) ***Espectrograma minuet bach c.wav***



(b) ***Espectrograma minuet bach c2.wav***

(c) ***Espectrograma minuet bach c3.wav***

(d) ***Espectrograma minuet bach c4.wav***



(e) ***Espectrograma minuet bach c5.wav***

(f) ***Espectrograma minuet bach 6.wav***

Figura A.4: Espectrogramas de la interpretación musical Minuet Bach

(a) ***Espectrograma PraeludiumVI Bach c.wav***

(b) ***Espectrograma PraeludiumVI Bach c2.wav***

Figura A.5: Espectrogramas de la interpretación musical Praeludium VI

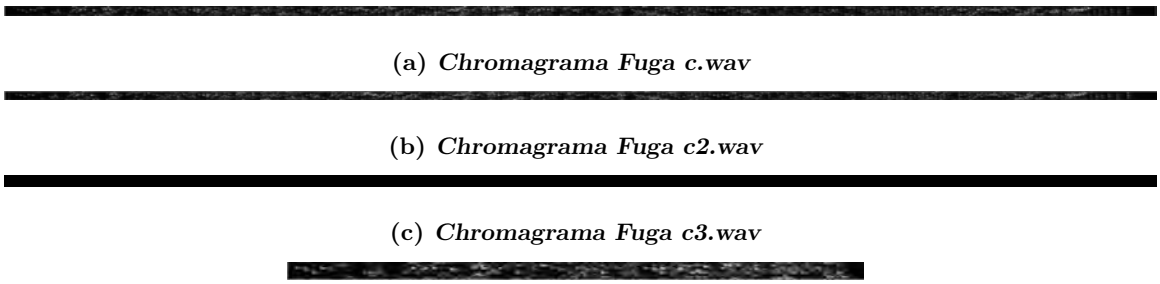
(a) ***Espectrograma PraeludiumI Bach c.wav***

(b) ***Espectrograma PraeludiumI Bach c2.wav***

(c) ***Espectrograma PraeludiumI Bach c3.wav***

Figura A.6: Espectrogramas de la interpretación musical Praeludium I

A.2. Cromagramas



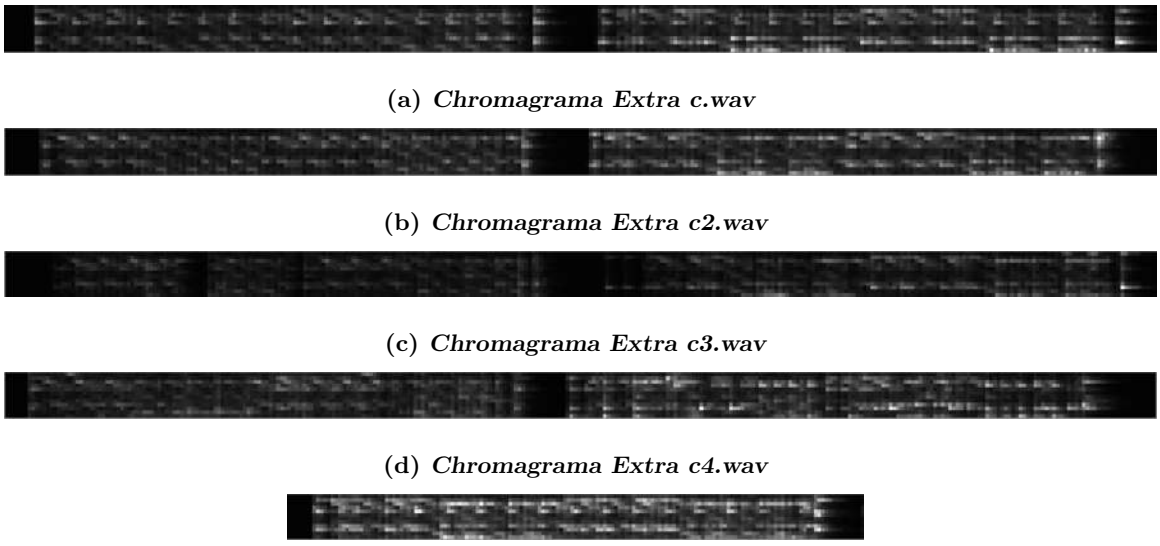
(a) ***Chromagrama Fuga c.wav***

(b) ***Chromagrama Fuga c2.wav***

(c) ***Chromagrama Fuga c3.wav***

(d) ***Chromagrama Fuga p.wav***

Figura A.7: Chromagramas de la interpretación musical Fuga VI



(a) ***Chromagrama Extra c.wav***

(b) ***Chromagrama Extra c2.wav***

(c) ***Chromagrama Extra c3.wav***

(d) ***Chromagrama Extra c4.wav***

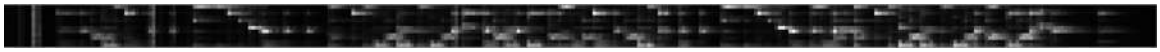
(e) ***Chromagrama Extra p.wav***

Figura A.8: Chromagramas de la interpretación musical Extra

(a) *Chromagrama LosChanguitos c.wav*

(b) *Chromagrama LosChanguitos c2.wav*

Figura A.9: Chromagramas de la interpretación musical Los Changuitos

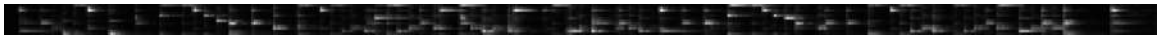


(a) *Chromagrama minuet bach c.wav*

(b) *Chromagrama minuet bach c2.wav*

(c) *Chromagrama minuet bach c3.wav*

(d) *Chromagrama minuet bach c4.wav*



(e) *Chromagrama minuet bach c5.wav*

(f) *Chromagrama minuet bach 6.wav*

Figura A.10: Chromagramas de la interpretación musical Minuet Bach

(a) *Chromagrama PraeludiumVI Bach c.wav*
PraeludiumVI Bach c2.wav]

Figura A.11: Chromagramas de la interpretación musical Praeludium VI

(a) ***Chromagrama PraeludiumI Bach c.wav***

(b) ***Chromagrama PraeludiumI Bach c2.wav***

(c) ***Chromagrama PraeludiumI Bach c3.wav***

Figura A.12: Chromagramas de la interpretación musical Praeludium I

Apéndice B

Partituras

En este apéndice se presentan las partituras de las interpretaciones musicales realizadas por los profesores y alumnos de música del Conservatorio de las Rosas de la ciudad de Morelia, Michoacán y que fueron utilizadas para los experimentos de este trabajo mostrados en el capítulo 5.

Figura B.1: Vals de la pulga, conocida como Los Changuitos.

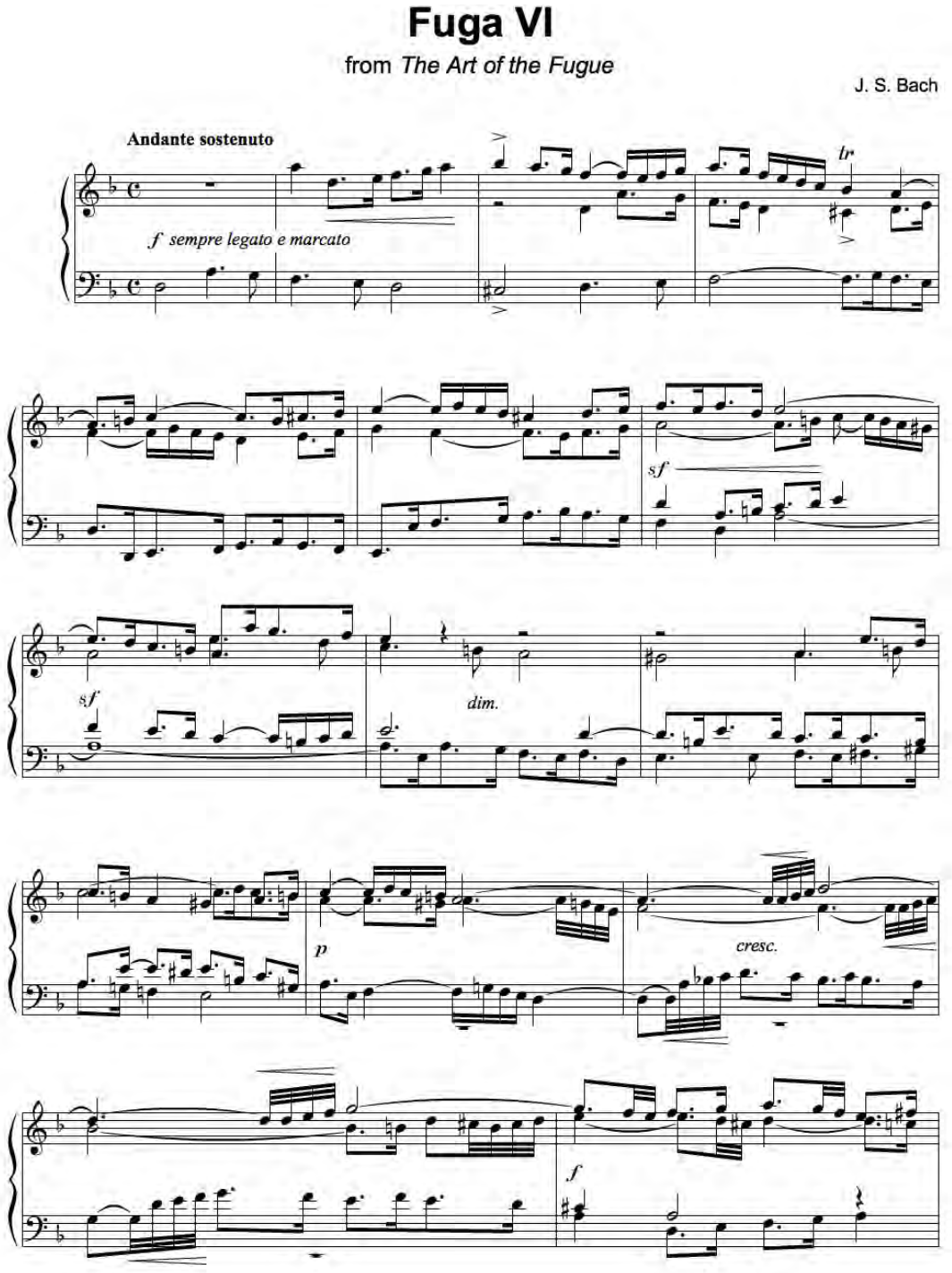


Figura B.2: Fuga VI, J. S. Bach

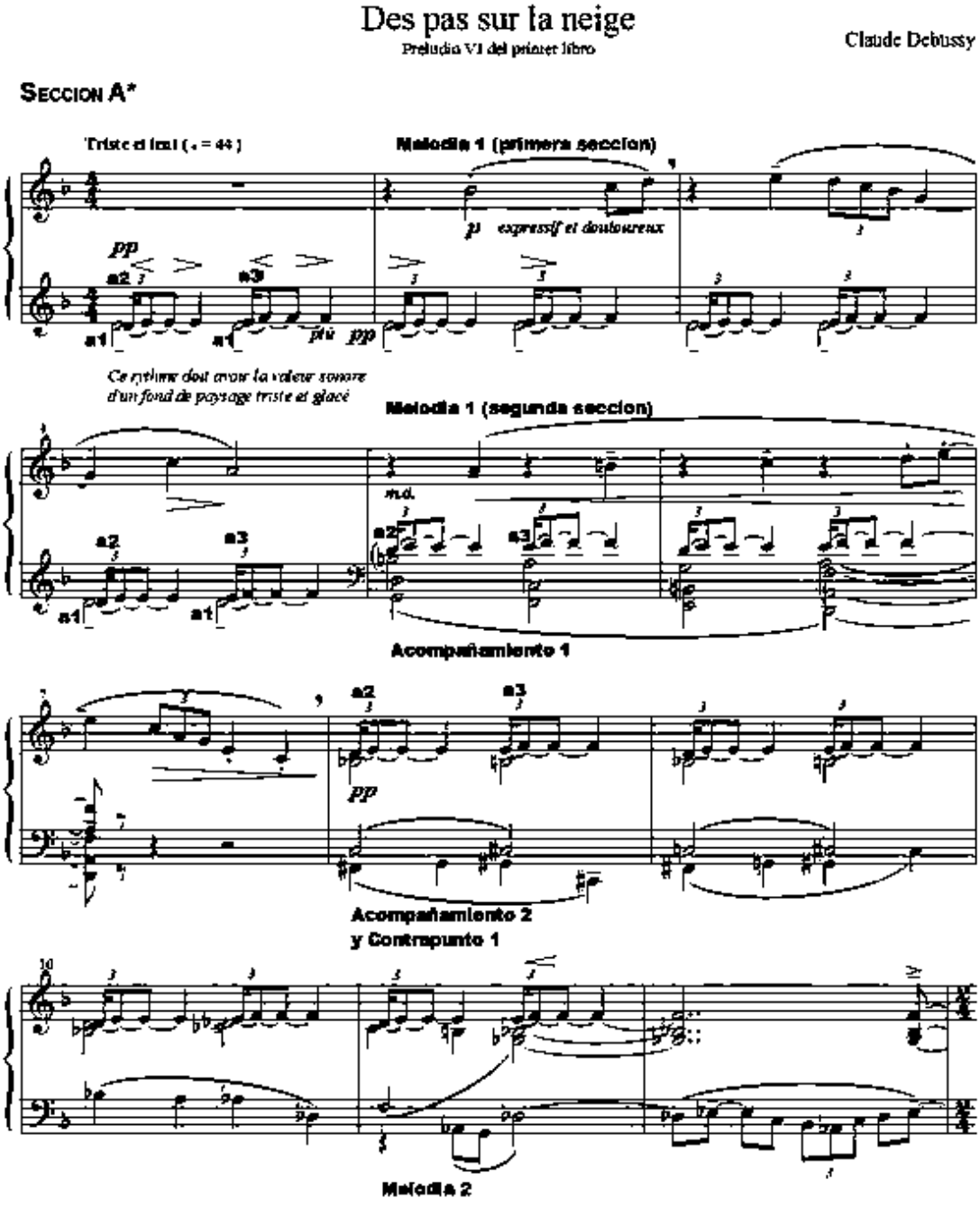


Figura B.3: Pasos sobre la Nieve, Preludio VI del primer libro.

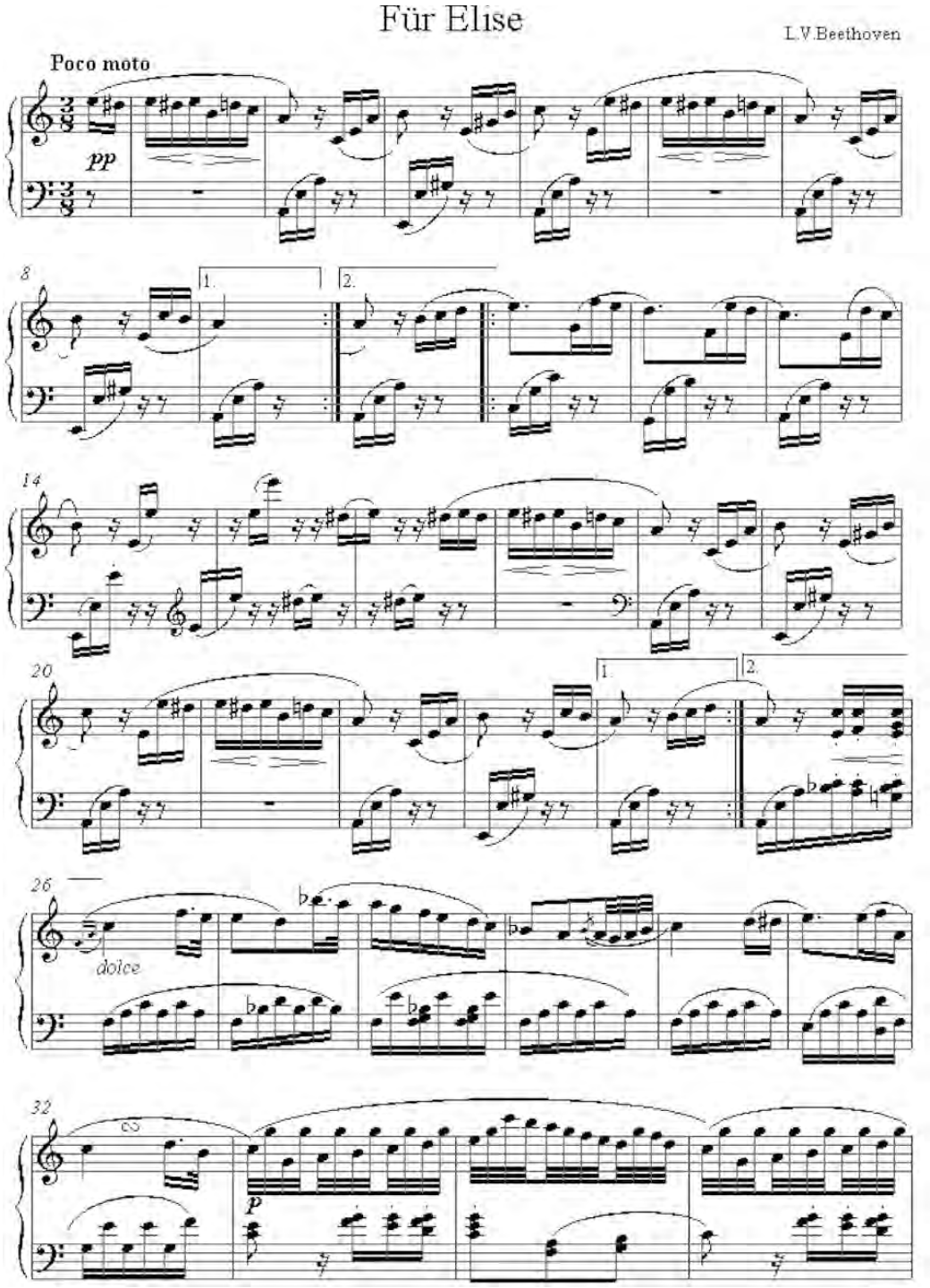


Figura B.4: Fur Elise, L. V. Beethoven.

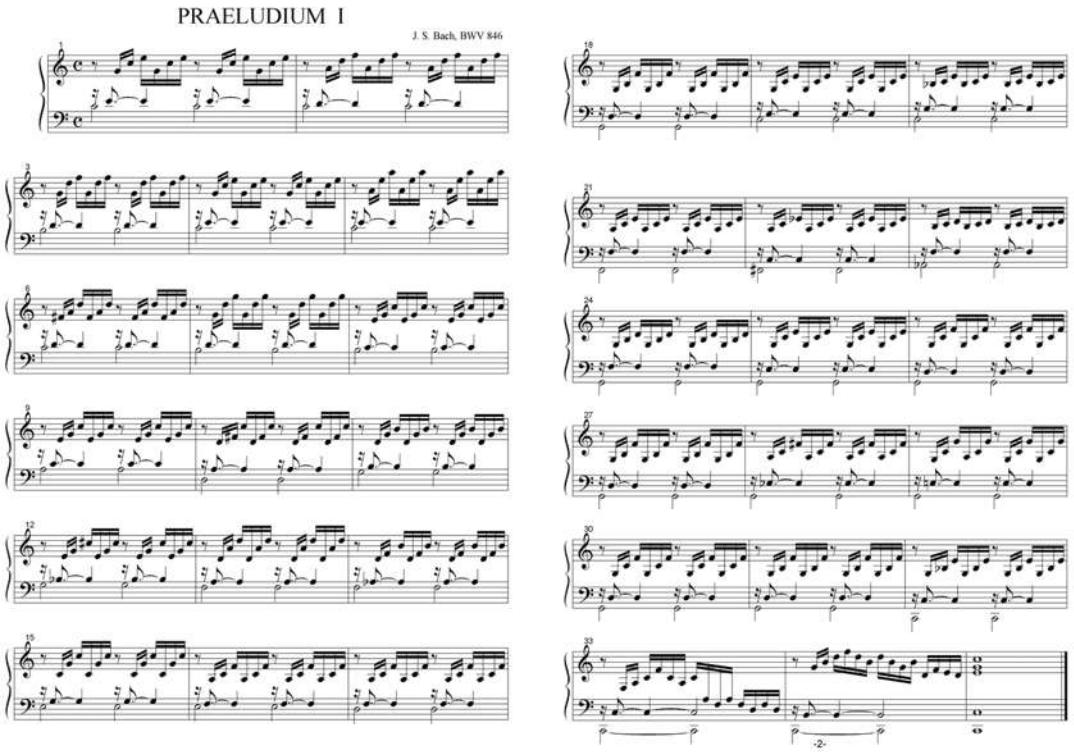


Figura B.5: Praeludium I, J. S. Bach.

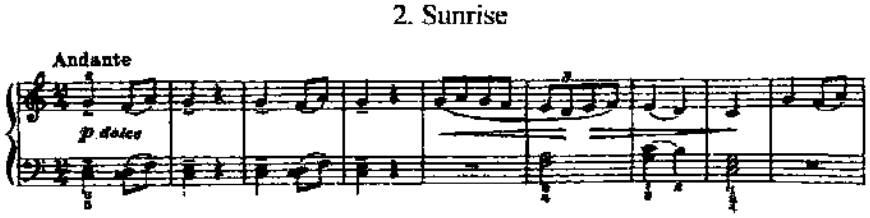


Figura B.6: Sunrise.

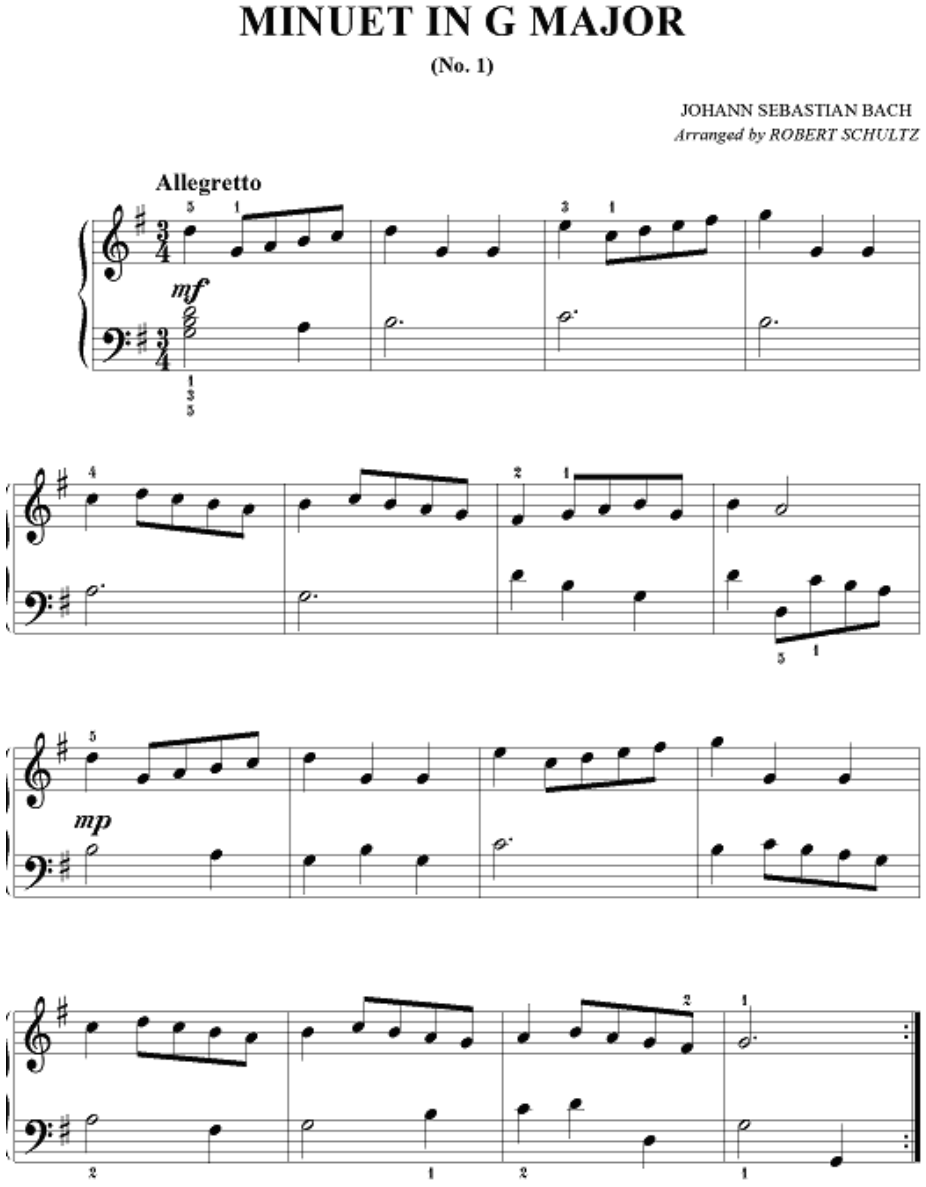


Figura B.7: Minuet G Major, Johann Sebastian Bach.

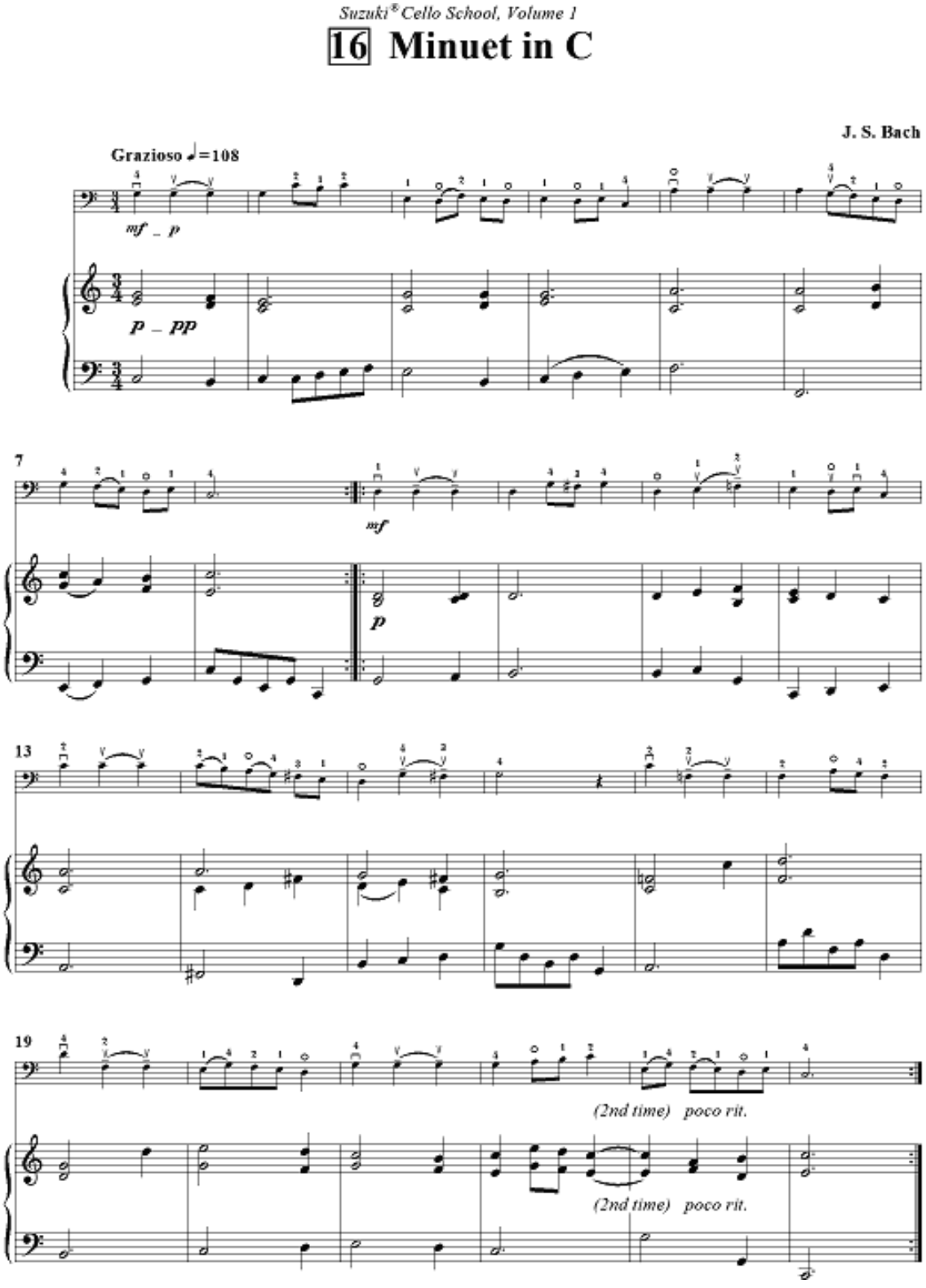


Figura B.8: Minuet C, Johann Sebastian Bach.

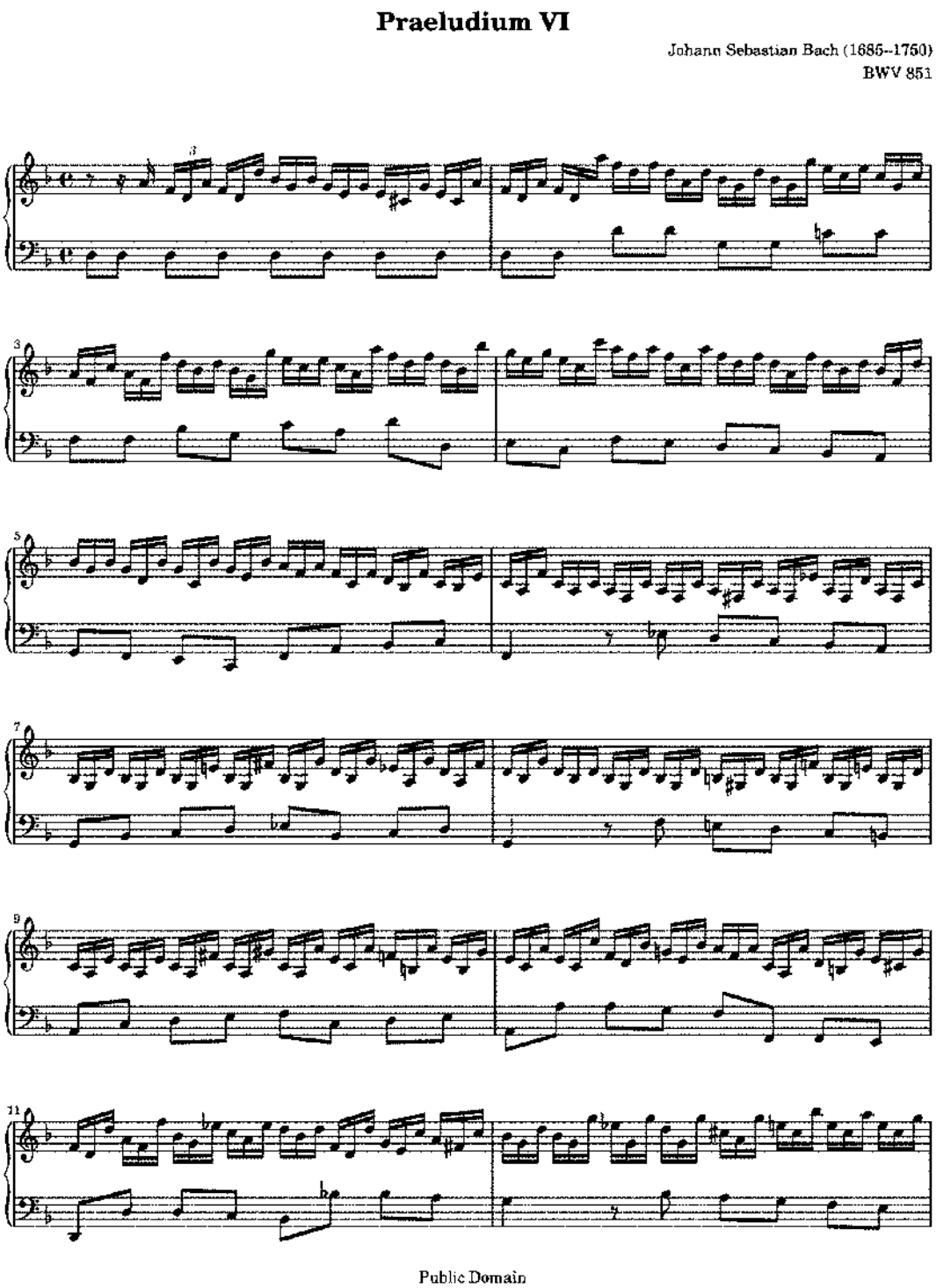


Figura B.9: Praeludium VI, J. S. Bach

Referencias

[Armónica07] Armónica, R. *Fundamentos de teoría de la música.* `http://www.ruedaarmonica.com/5.php`, 2007.

[Davis80] Davis, M. P., S. B. Comparison of parametric representations for monosyllabic word recognition in continuosly spoken setences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28:357–366, 1980.

[Flanagan J. L.66] Flanagan J. L., G. R. M. Phase vocoder. *Bell System Technical Journal*, págs. 1493–1509, 1966.

[Fober04] Fober, D., Letz, S., Orlarey, Y., Askenfelt, A., Falkenberg Hansen, K., y Schoonderwaldt, E. Imutus: An interactive music tuition system. *En the Sound and Music Computing Conference (SMC 04), October 20-22, 2004, IRCAM, Paris, France*, págs. 97–103. 2004.

[Fujishima99] Fujishima, T. Realtime chord recognition of musical sound: A system using common lisp music. *Proceedings of the International Computer Music Conference.*, págs. 464–467, 1999.

[Gabor46] Gabor, D. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[Gold00] Gold, B. y Nelson, N. Speech and audio signal processing: Processing and perception of speech and music. *New York, Wiley and Sons*, 2000.

[Hu N.03] Hu N., D. R. y. T. G. Polyphonic audio matching and align- ment for music retrieval. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.*, 2003.

[Levenshtein66] Levenshtein, V. Binary codes capable of correction deletions, insertions and reversals. *Soviet Physics Doklady.*, 10(8):707–710, 1966.

[Luo12] Luo, C. *Cochlear Implant*. `http://cnx.org/contents/17161975-9bc6`, 2012.

[Manzo-Martinez13] Manzo-Martinez, A. y Camarena-Ibarrola, A. An eigenvalues analysis with entropy-per-chroma feature. *En Power, Electronics and Computing (ROPEC), 2013 IEEE International Autumn Meeting on*, págs. 1–6. IEEE, 2013.

[Orio01] Orio, D., N. y Schwarz. Alignment of monophonic and polyphonic music to a score. *Proceedings of International Computer Music Conference (ICMC).*, págs. 155–158, 2001.

[Percival07] Percival, G., Wang, Y., y Tzanetakis, G. Effective use of multimedia for computer-assisted musical instrument tutoring. *En Proceedings of the International Workshop on Educational Multimedia and Multimedia Education*, Emme '07, págs. 67–76. ACM, New York, NY, USA, 2007. ISBN 978-1-59593-783-4. doi:10.1145/1290144.1290156.
URL `http://doi.acm.org/10.1145/1290144.1290156 [1]`

[Rabiner89] Rabiner, L. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE.*, 77:257–286, 1989.

[Robine07] Robine, M., Percival, G., y Lagrange, M. Analysis of saxophone performance for computer-assisted tutoring. *En Proceedings of the International Computer Music Conference (ICMC07)*, tomo 2, págs. 381–384. 2007.

[Schoonderwaldt05] Schoonderwaldt, E., Askenfelt, A., y Hansen, K. F. Design and implementation of automatic evaluation of recorder performance in imutus. *En In Proceedings of the International Computer Music Conference (ICMC*, págs. 97–103. 2005.

[Sheh03] Sheh, A. y Ellis, D. P. Chord segmentation and recognition using em-trained hidden markov models. *ISMIR 2003*, págs. 185–191, 2003.

[Smith07] Smith, J. O. *Mathematics of the Discrete Fourier Transform (DFT) with audio applications.* `https://ccrma.stanford.edu/ jos/mdft/mdft.html`, 2007.

[Traunmüller90] Traunmüller, H. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, 88(1):97–100, 1990.

[Zwicker61] Zwicker, E. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, (33 (2)):248, 1961.