UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería

Reconocimiento de intérpretes en canciones

Trabajo de graduación presentado por
Luis Antonio Monteros Méndez
para optar al grado académico de Licenciado en Ingeniería Electrónica

Guatemala
2013

Reconocimiento de intérpretes en canciones

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería

Reconocimiento de intérpretes en canciones

Trabajo de graduación presentado por
Luis Antonio Monteros Méndez
para optar al grado académico de Licenciado en Ingeniería Electrónica

Guatemala
2013

Vo. Bo. :

(f) 

Ing. Carlos Esquit

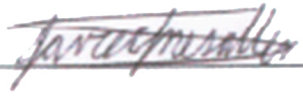
Tribunal Examinador:

(f) 

Ing. Carlos Esquit

(f) 

Ing.Julio Vásquez

(f) 

Ing. Javier Mesalles

Fecha de aprobación: Guatemala, 29 de noviembre de 2013.

PREFACIO

Este trabajo de graduación está destinado para todas las personas interesadas en conocer métodos para la detección de voz por medio del área de procesamiento de señales. Para ser más específicos, este trabajo es un reconocedor de cantantes, es decir, las pruebas realizadas se harán tratando de reconocer la voz sobre canciones, tratando de encontrar al intérprete.

La elaboración de este trabajo de graduación en modalidad de trabajo profesional, surgió del interés personal de explorar el procesamiento de señales sobre los archivos musicales, que en los últimos años han crecido en número gracias al auge del internet, principalmente. Este incremento en archivos también ha hecho que existan hoy en día algunas aplicaciones muy populares para móviles como "Shazam" o "SoundHound" que logran reconocer la pista que se está escuchando por medio del micrófono del teléfono.

Durante el transcurso de mi licenciatura siempre mostré un interés mayor por los proyectos que incluían la música como parte de estos. De esta manera, el trabajo de graduación que se presenta en las siguientes páginas busca entender de una mejor manera la estructura de los archivos musicales y qué trabajos se han hecho sobre ellos para poder llegar a reconocer quién es el intérprete que aparece en esas mismas pistas grabadas.

La idea de profundizar sobre el reconocimiento del cantante en una pista surge debido a la problemática que afronta este tema: Ninguna persona ha sido capaz de generar algún algoritmo que resuelva el problema a un 100%. En este trabajo no se pretende encontrar ese algoritmo que solucione el problema, sino más bien intenta además de introducirme en el tema, saber en qué consiste, lograr implementar un algoritmo de los muchos que existen y analizar sus resultados tomando en cuenta el género musical del archivo, para poder llegar a ideas claras de qué debiera ser el trabajo futuro sobre el algoritmo escogido.

Por otra parte, agradezco al Ing. Carlos Esquit y al Ing. José Quan por la guía que me dieron durante la realización de este trabajo. Sus consejos fueron fundamentales para mantenerme en el camino que el trabajo debía tomar y no tomar decisiones equivocadas que solamente me hicieran desviarme de mis objetivos. De igual forma, me gustaría mencionar al PhD Szabolcs Blazsek que me asistió sobre todo en la comprensión de los modelos estadísticos (GMM) que se usan en el algoritmo desarrollado. Por último, me gustaría mucho agradecer al departamento de Ingeniería Electrónica que me prestaron su equipo en todo momento para poder desarrollar el trabajo de una manera más eficiente durante el tiempo que tomó el proceso.

ÍNDICE

LISTA DE TABLAS

LISTA DE FIGURAS

RESUMEN

Este estudio presenta una técnica para la identificación de un cantante en una grabación musical. En este trabajo se presenta un algoritmo ya usado anteriormente, el cual se busca implementar. Se seleccionará un solo algoritmo, en base a trabajos realizados anteriormente por otros expertos; y sobre este mismo algoritmo se buscará saber si es más adecuado para cierto género musical o no existe tal relación. La validación de esta hipótesis se confirmará a través de evaluaciones que se harán con diferentes pistas musicales de diferentes géneros pero con el mismo intérprete.

El reconocimiento de cantantes es aún un tema que no se ha resuelto al 100%: no existe un algoritmo ya pre-establecido que no cometa ningún error. Esto se debe a muchos problemas que se tienen con la homogenización del proceso: No es lo mismo analizar un archivo .mp3 con un bitrate de 240kbps que un .wav y un bitrate de 320kbps. Además, los instrumentos que acompañan una canción no pueden ser filtrados para ser eliminados del todo y quedarse únicamente con la voz, para posteriormente compararla y reconocerla. Además, al mismo algoritmo se le pueden hacer diferentes parametrizaciones, que tienen como consecuencia, diferentes resultados al momento de intentar encontrar al intérprete original.

En este trabajo de graduación se encontró que la música pop resulta ser el género que funciona mejor con este algoritmo bajo ciertos parámetros, pero no es algo marcado que siempre suceda y el porcentaje de aciertos puede cambiar drásticamente con pequeños cambios en los parámetros que se detallan en este trabajo en secciones posteriores. Esto hace que no se pueda afirmar que el algoritmo trabaja mejor para cierto género bajo cualquier circunstancia.

I. INTRODUCCIÓN

El reconocimiento del habla es un problema que ha sido tratado durante las últimas décadas de forma intensa, y durante todo este tiempo se han diseñado diferentes algoritmos que buscan crear una metodología ya fija que resuelva este problema. Fue en los años ’60 cuando se realizaron los primeros estudios, en los cuales se hizo notorio que la variabilidad de patrones que se hacía en una palabra pronunciada por una misma persona era evidente. Clasificaciones estadísticas clásicas como densidades de probabilidad estáticas arrojaron tasas de error demasiado altas para ser consideradas, incluso. Con el tiempo, se realizaron técnicas de programación dinámica y cuantización vectorial, que lograron mejorar las tasas. Pero al mismo tiempo, estudios en el área de biología y psicología lograron incluir conocimiento en el modelado de la fuente de voz y en la extracción de características relevantes de ésta.

En los ’80 fue cuando los Modelos Ocultos de Markov marcaron la tendencia dentro del campo. Lo novedoso acerca de este modelo, es que permitía al algoritmo de Expectación Maximización (EM) adecuarse de forma consistente para modelar las observaciones. Primeramente este modelo fue utilizado para reconocer personas en base a alguna palabra, pero luego evolucionó para reconocer también por medio de frases y aún ahora continúa siendo usado. (Los Modelos Ocultos de Markov se basan principalmente en encontrar patrones de secuencias y probabilidad de encontrar a la misma al evaluar un vector de entrada). Pero son los ambientes ruidosos o muy melódicos los que hacen que estos algoritmos fallen. En lugares con mucho ruido, no hay forma de “limpiar” la voz antes de entrar al micrófono y al tratarse de sonidos sin un patrón, hace prácticamente imposible modelarlo. Por otro lado, en ambientes muy melódicos, donde hay más voces u otros instrumentos acompañando a la voz, es muy difícil lograr aislar la voz sin eliminarla en algunas partes o dejando espacios donde ésta no aparece, y es en esta última parte donde está el enfoque de este trabajo.

A pesar de los increíbles avances que se han dado en el reconocimiento del habla de una persona, lograr esto en una canción, donde aparecen una gran cantidad de instrumentos e incluso otras voces además de la principal, ha sido más difícil aún y por ello sigue siendo un tema de estudio que aún no encuentra un punto donde haya ya alguna forma de solucionar el problema sin cometer errores. Los estudios que se han hecho son habitualmente “controlados”, donde ya hay una base de datos de antemano y donde la música con la que se hacen las pruebas se conoce también desde antes. Esto último hace sentido ya que es prácticamente imposible tener almacenada la voz de todos los cantantes en una sola base de datos. En este trabajo, como se mencionó previamente, se busca encontrar algún algoritmo ya creado para luego implementarlo a través de algún software que los equipos de cómputo puedan ejecutar. Pero el principal objetivo para este trabajo es el encontrar si el algoritmo del que se está hablando logra un mayor porcentaje de aciertos para algún género musical en especial o no. Esto lograría contribuir en el campo, ya que de ser cierto que funciona mejor para cierta clase de canciones, podría establecerse ya como la mejor forma de identificar a algún artista según el género en el que cante; además, en los algoritmos encontrados, ninguno separa por medio de géneros su estudio.

Al no tener todas las canciones de todos los artistas, se formará una base de datos de 10 artistas para las pruebas. Uno de estos artistas de género femenino. Se

buscan hacer las pruebas si el algoritmo tiene mejor desempeño para cierto género con canciones que se buscaron en línea en donde la voz sea de la misma persona, pero el acompañamiento musical se haya visto alterado y ahora la canción tenga un ritmo/género distinto. Esto es algo poco común por la complejidad que tiene hacerlo, así que el número de canciones para medir el desempeño estará delimitado por la cantidad de canciones que se encuentren que cumplan estas características.

II. OBJETIVOS

A. Generales

Analizar, comprender detalladamente, e implementar un algoritmo ya diseñado anteriormente por otros ingenieros, que busque reconocer al intérprete de una canción.

B. Específicos

- Crear un software con la base de datos, capaz de tomar archivos digitales para realizar las pruebas. En este software se programarán los algoritmos y se desplegarán resultados.
- Explorar un algoritmo previamente utilizado en reconocimiento de cantantes y a partir de éste, determinar con qué género musical se obtiene un mayor porcentaje de efectividad al hacer la identificación.
- Implementar al sistema el algoritmo utilizado para reconocer el momento a partir del cual, el cantante aparece en la grabación.
- Presentar una tabla con los resultados obtenidos y encontrar si existe un género con menor porcentaje de error que el resto.

III. JUSTIFICACIÓN

En las últimas tres décadas, el estudio de la voz humana para facilitar la comunicación entre máquinas y humanos ha sido una tarea de investigación de gran importancia. Una gran cantidad de investigaciones se han hecho para poder lograr hacer un reconocimiento de palabras y así mismo de la persona que las produce. Siguiendo con estos estudios, este trabajo trata con el problema de lograr reconocer un cantante en una canción particular. El reconocimiento del cantante puede ser visto como uno de los más complejos, ya que la voz de éste va acompañado por frases, ritmos y melodías, e incluso emociones, al igual que está superimpuesta por un acompañamiento musical la mayor parte del tiempo.

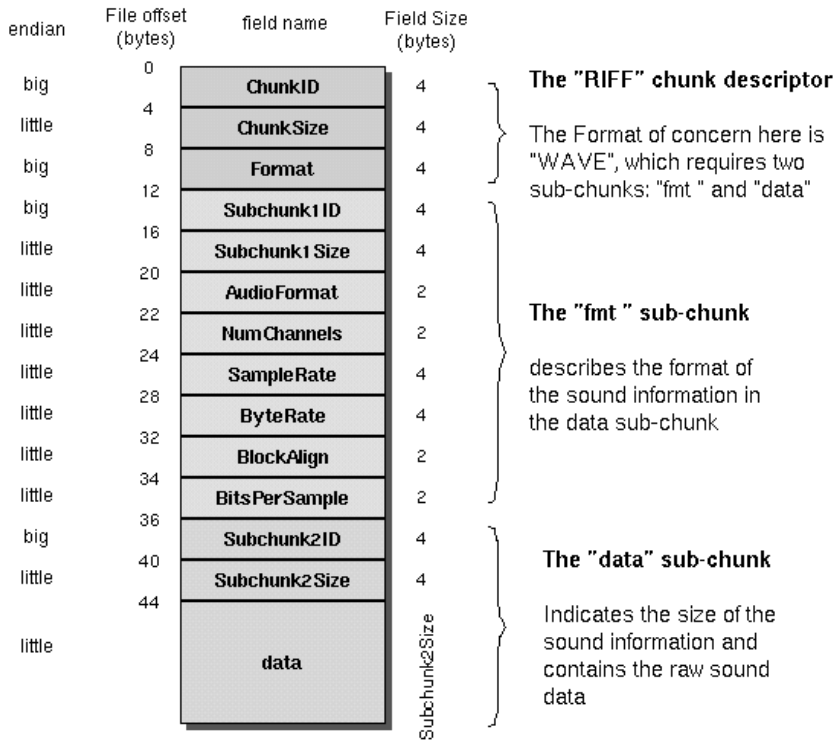
La identificación de un cantante es un campo emergente, que crece debido a la rápida proliferación de música popular en el Internet. En contraste con clasificaciones anteriores basadas en género ya sea femenino/masculino u otras características, este reconocimiento puede ser usado para distinguir entre una canción original o un cover de otra banda, o de igual forma, para ayudar a compañías a encontrar posibles versiones no originales de canciones previamente registradas.

IV. MARCO TEÓRICO

En esta sección se presenta toda la teoría necesaria detrás del trabajo de graduación:

- Archivo de audio .wav: El archivo usado para realizar las pruebas en el algoritmo, tenían extensión .wav, que es un subconjunto de los archivos RIFF, que a su vez forma parte de un grupo del archivos multimedia de Microsoft. La extensión ".wav" toma su nombre del nombre del tipo de archivo: Waveform Audio File Format. Los archivos RIFF comienzan con un encabezado seguido por una secuencia de datos en forma de trozos o chunks, como se conocen en inglés. Los archivos .wav o WAVE, como también se les conoce, consisten de dos sub-trozos: Uno "fmt" que especifica el formato de los datos y uno subsecuente que contiene en sí los datos. Los archivos .wav son uno de los más simples para archivos de audio, ya que contiene únicamente los bits para representar la señal de audio. La siguiente imagen muestra cómo se componen los archivos .wav (14):

Figura 1: Ejemplo de cómo se compone un archivo .wav



- Filtro de Pre-Énfasis: Los filtros de pre-énfasis son filtros que se usan como técnicas de reducción de ruido; en el cual una señal se aplica a la entrada del filtro, el cual incrementa la magnitud de la señal para las frecuencias altas, por lo que estos filtros son implementados como filtros pasa-altas. (16)
  La ecuación de diferencias que implementa el filtro de forma digital es:

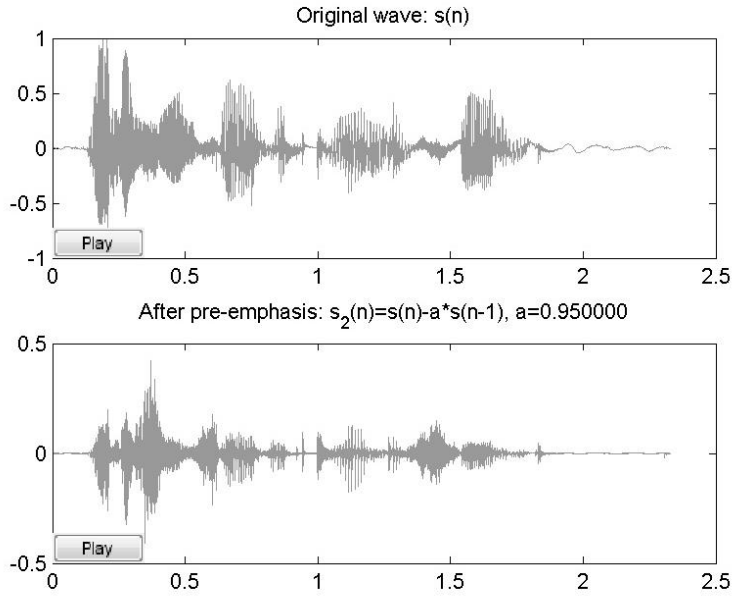
$$s_2(n)=s(n)-a*s(n-1) \quad (1)$$

En donde s_2 es la señal de salida del filtro y el valor del parámetro "a" oscila generalmente entre 0.9 y 1.
La transformada-z del filtro es:

$$H(z)=1-a*z^{-1} \quad (2)$$

Para esta aplicación en específico, este filtro busca compensar la parte de altas frecuencias que se atenúan naturalmente cuando los humanos las producen. De igual forma, este filtro logra amplificar los formantes que aparecen en las altas frecuencias. La siguiente figura demuestra lo que el filtro le hace a una señal de audio que fue usada como entrada al mismo (15):

Figura 2: Ejemplo de una señal de audio filtrada por un filtro pre-énfasis.



- Coeficientes Cepstrales de la Escala Mel (Mel Frequency Cepstral Coefficients o MFCC): Los MFCC son coeficientes que se usan comúnmente al hacer el reconocimiento de voz. Estos coeficientes logran de manera muy precisa representar la forma del tramo de la voz que se manifiesta en el espectro de la señal. Los MFCC fueron introducidos por Davis y Mermelstein en los años '80, y son un trabajo en "estado del arte". Antes de la aparición de los MFCC, los coeficientes de predicción lineal (LPCs) y los coeficientes de predicción lineal cepstral (LPCCs) eran los coeficientes que se usaban para el reconocimiento de voz.
  Los pasos para obtener los MFCC son generalmente:
  - Partir la señal en bloques
  - Para cada bloque, calcular su espectro y luego su magnitud.
  - Aplicar un banco de filtros mel a la magnitud del espectro y sumar la energía de cada filtro.
  - Aplicar el logaritmo a todos las energías del banco.
  - Aplicar la Transformada Discreta de Coseno a las energías.

  Este es el proceso que logra que en vectores de coeficientes la Transformada Discreta de Coseno logre representar las características de la voz del artista. Adicionalmente, se pueden remover coeficientes o agregar otros como la energía o la

derivada del vector final. Generalmente se toman los coeficientes número 2 al 13 y se descarta el resto, esto por propiedades de la Transformada Discreta de Coseno. Al terminar este proceso, los coeficientes resultantes son los llamados cepstros, que le dan el nombre al algoritmo (19).

- Ventana de Hamming: La ventana de Hamming es una función de ventana especial que logra que exista continuidad en los primeros y últimos puntos de cada bloque que representan la señal. Las funciones de ventana tienen la característica que tienen un valor de 0 a cualquier fuera de cierto rango que el diseñador elija. Por ejemplo, una función que es una constante dentro de un intervalo y cero fuera de éste, es llamada "Ventana Rectangular", debido a su representación gráfica. En otras palabras, cuando otra función o secuencia de datos es multiplicada por la función de ventana, el producto es cero fuera del intervalo que delimita la ventana; todo lo que resta es la parte en donde se traslaparon. Además de la de Hamming y la Rectangular, existen otras ventanas como muestra la Figura 3. (17)

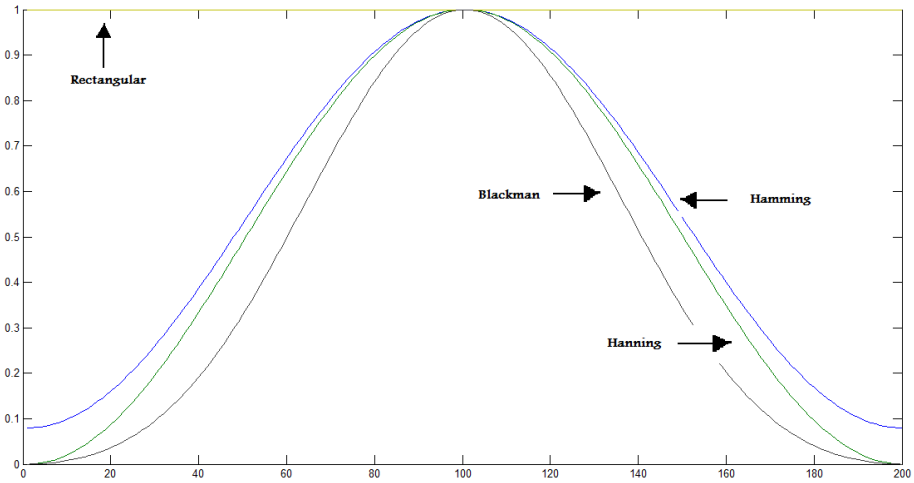
  La ventana de Hamming fue propuesta por Richard W. Hamming. La ventana es optimizada para minimizar los lóbulos laterales. La función para esta ventana es:

$$w(n) = \alpha - \beta \cos\left(\frac{2\pi n}{N-1}\right) \quad (2)$$

Generalmente se usan como coeficientes:

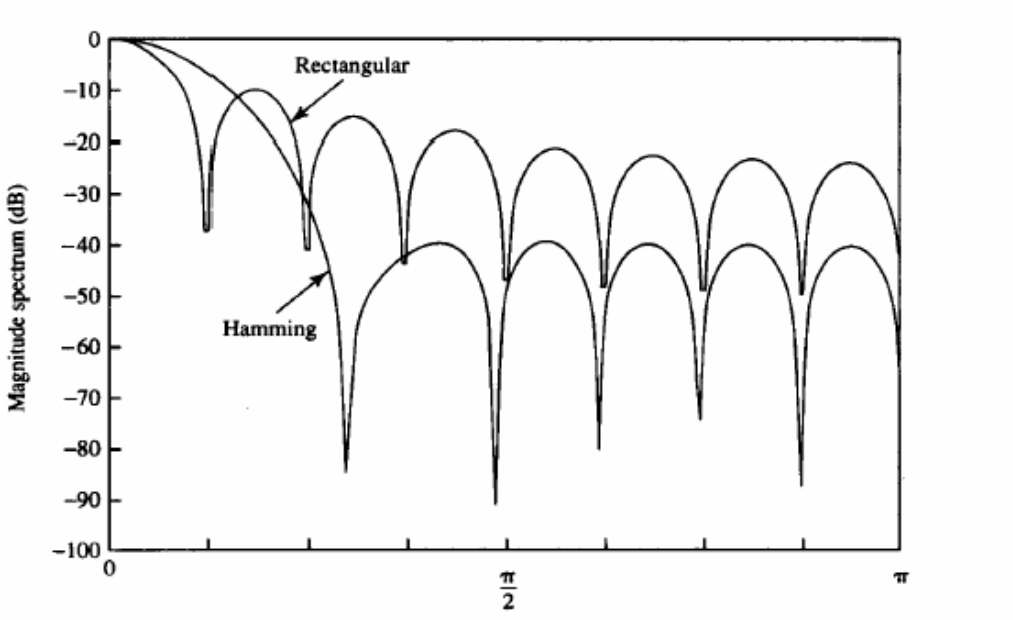
$$\alpha = 0.54, \ \beta = 1 - \alpha = 0.46, \quad (3)$$

Figura 3: Gráficas de diferentes tipos de ventana.



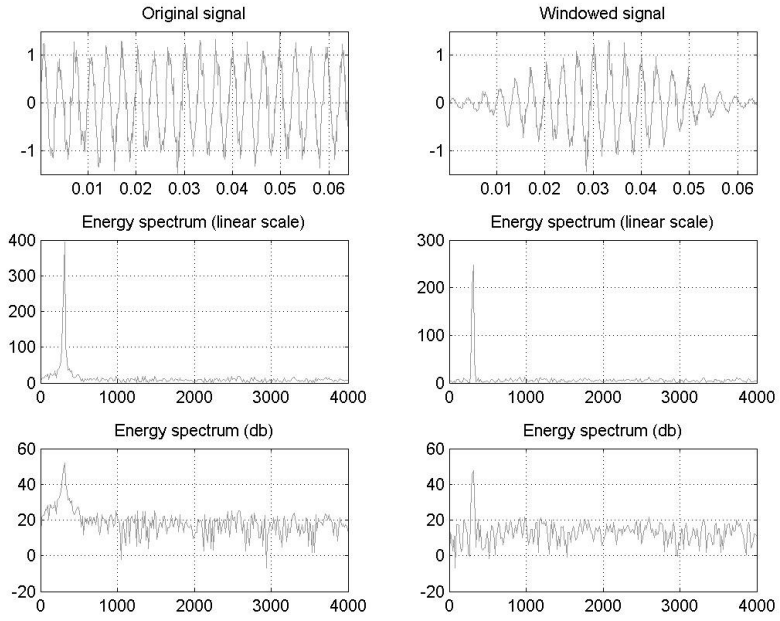
Entre las razones por la cual se eligió la ventana de Hamming sobre el resto, es que el efecto convolutivo con el espectro en frecuencia de la señal de voz será mucho menor. Esto en el dominio de frecuencia se puede interpretar con un lóbulo principal estrecho y la amplitud de los lóbulos siguientes los menor posible, como muestra la siguiente figura.

Figura 4: Gráfica en el dominio de frecuencia de la ventana de Hamming y la ventana Rectangular



En el tema de reconocimiento de voz, las ventanas más comunes a utilizar son las ya mencionadas ventana Rectangular y la de Hamming. Si se ve la gráfica anterior, se puede ver la ventaja que conlleva el uso de la ventana de Hamming. Sin la ventana de Hamming en el algoritmo, las discontinuidades que se dan al final de cada bloque en el que se divide el archivo de audio, pueden provocar que existan picos al momento de graficar el espectro y pueden causar confusiones o malinterpretaciones de los resultados. Los efectos de la ventana de Hamming se pueden ver en la siguiente imagen:

Figura 5: Ejemplo del efecto de una ventana de Hamming sobre una señal



- Transformada Rápida de Fourier (FFT): Al tratarse de reconocer la voz de un artista este trabajo, se busca que el algoritmo logre imitar el funcionamiento del oído humano. Lo que éste hace, es que dentro del oído, la cóclea vibra en diferentes lugares dependiendo de la frecuencia que escucha. Dependiendo de este lugar, diferentes nervios se activan para mandar señales al cerebro que dicen cuáles frecuencias están presentes. El proceso que realiza el algoritmo es similar ya que busca qué frecuencias produce cada artista para luego ser reconocido.

  Para saber las frecuencias que sí están presentes en cada bloque de datos es necesario realizar un análisis espectral que muestra cómo los diferentes timbres que aparecen en señales que contienen voz corresponden a diferentes distribuciones de energía sobre las frecuencias. Para esto se usa la FFT para obtener la magnitud en frecuencia de cada bloque a analizar. En este algoritmo, cuando se realiza la FFT, se asume que la señal dentro de cada bloque es periódica y continua. Esto no es necesariamente cierto para todos los casos, aún así, se puede realizar la FFT siempre y cuando antes la señal haya sido modificada por algún método para evitar que las discontinuidades en los extremos afecten y generen efectos indeseados (En este caso, la multiplicación por la ventana de Hamming hace ese trabajo). (17)

La FFT es un método para calcular la Transformada Discreta de Fourier (DFT). La FFT logra reducir el número de operaciones necesarias para *N* puntos de $2*N^2$ a $2*N*lg*N$, donde lg es el algoritmo de base 2. Si la función a transformar no es armónica, la respuesta de la FFT será parecida a una función sinusoidal.

La DFT tiene sus orígenes en la Transformada de Fourier Continua, esta última es definida para una secuencia de *N* muestras f(n), con índices n = 0, 1, …, N-1 como:

$$F(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n)\, e^{-j2\pi kn/N} \quad (4)$$

F(k) es comúnmente llamada “Coeficientes de Fourier” o “Armónicos”. De igual forma, la secuencia *f(n)* puede ser calculada teniendo *F(k)* usando la Transformada Discreta de Fourier Inversa (IDFT), de la siguiente manera:

$$f(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F(k)\, e^{+j2\pi nk/N} \quad (5)$$

En la mayoría de los casos, *f(n)* y *F(k)* son complejos. Igualmente, estas secuencias son conocidas como datos en “dominio de tiempo” y en “dominio de frecuencia” respectivamente. Aunque se establece que *n* y *k* tienen como rango desde 0 hasta N-1, la definición de la Transformada de Fourier tiene un período N:

$$F(k+N) = F(k) \qquad f(n+N) = f(n) \quad (6)$$

De esta forma, solo es necesario calcular los valores entre el rango de 0 a N-1 para obtener la representación total de *f(n)* y *F(k)*. Para la computación de esta Transformada de una forma más rápida, los algoritmos FFT e IFFT ignoran los factores de escalamiento y calculan lo siguiente:

$$FFT_N(k, f) = \sum_{n=0}^{N-1} f(n)\, e^{-j2\pi kn/N} = \sqrt{N} F(k) \quad (6)$$

$$IFFT_N(n, F) = \sum_{k=0}^{N-1} F(k)\, e^{+j2\pi nk/N} = \sqrt{N} f(n) \quad (7)$$

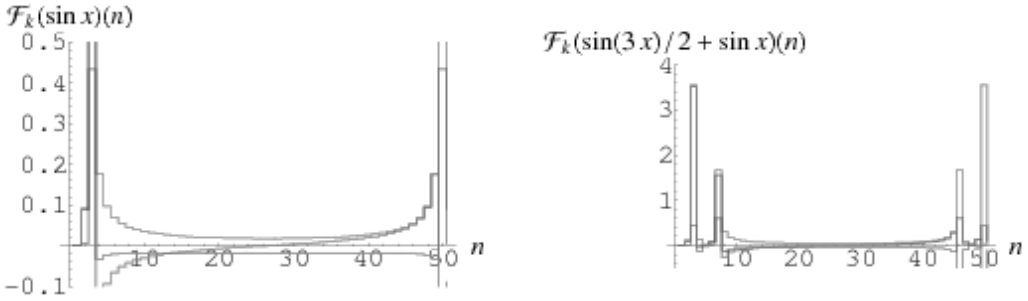
Algunos ejemplos de la DFT, se muestran a continuación:

- 1 punto:
  $[F(0)] = [1][f(0)]$
- 2 puntos:
  $$\begin{bmatrix} F(0) \\ F(1) \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \end{bmatrix}$$
- 4 puntos:
  $$\begin{bmatrix} F(0) \\ F(1) \\ F(2) \\ F(3) \end{bmatrix} = \frac{1}{\sqrt{4}} \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -j & -1 & +j \\ +1 & -1 & +1 & -1 \\ +1 & +j & -1 & -j \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{bmatrix}$$

Es importante notar que cada una de las matrices multiplicadoras puede ser invertida al conjugar los elementos. Esto es exactamente lo que se esperaría, dado que la única diferencia entre la DFT e IDFT es el signo del argumento en el exponencial. (15)

Las siguientes imágenes muestra la parte real (línea clara), la parte imaginaria (línea oscura) y el módulo complejo (línea intermedia) de la DFT de las funciones $f(x) = \sin x$ (izquierda) y $f(x) = \sin x + \sin(3x)/2$ (derecha) sampleada 50 veces en dos períodos. En la imagen en la izquierda, los picos simétricos a los lados son los componentes de frecuencia positivos y negativos de la onda sinusoidal. De igual forma, en la figura derecha, hay dos picos, con valores más altos de color verde que corresponden a los componentes en frecuencias más fuertes para las frecuencias bajas de *sin(x)* y picos más pequeños correspondientes a componentes de frecuencias más altos. (21)

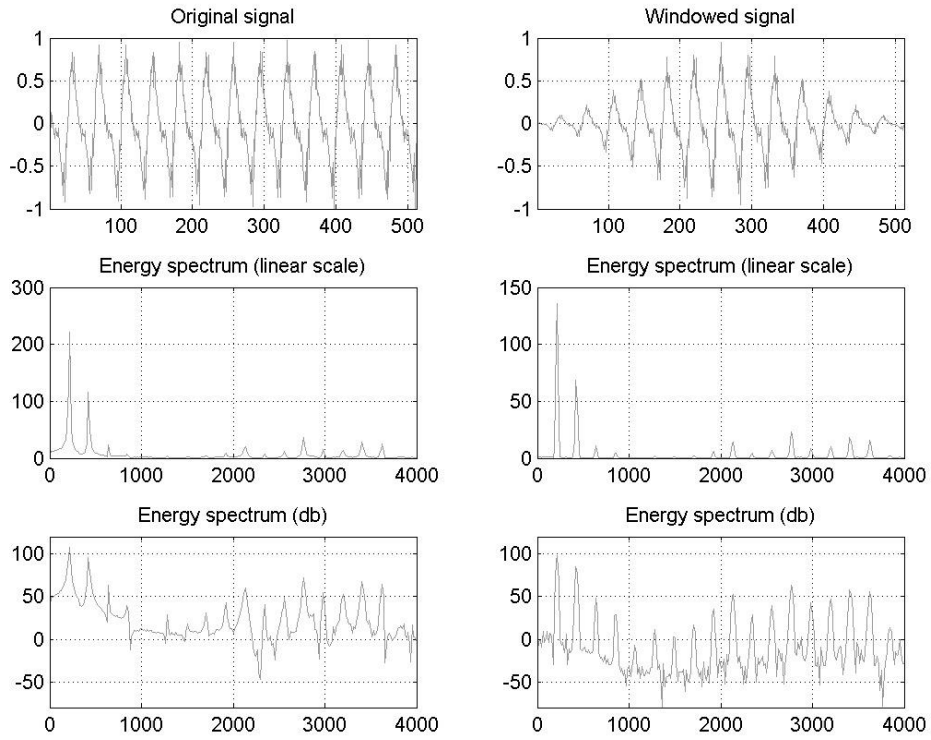
Figura 6: Representación gráfica de la DFT para diferentes funciones



La transformada discreta de Fourier es un caso especial de la Transformada-Z.

La siguiente figura ilustra el trabajo tanto de la ventana de Hamming como de la FFT aplicadas:

Figura 7: Ejemplo de una señal con su multiplicación por la ventana de Hamming y transformada a través de la FFT



- Filtros-Mel y escala Mel: La cóclea al momento de intentar distinguir entre frecuencias muy parecidas, las interpreta como si fueran la misma; y este efecto se vuelve aún más pronunciado cuando se trata de frecuencias relativamente altas. Por esto, se busca de otra forma saber cuánta energía hay entre ciertos rangos de frecuencia. Este lo realiza el "Banco de Filtros-Mel": El primer filtro es muy "delgado" y da una idea de cuánta idea existe para las frecuencias cercanas a 0Hz. Conforme las frecuencias se van incrementando, los filtros se vuelven más anchos ya que no se necesita saber tanto acerca de las variaciones. En el presente trabajo solo se busca saber cuánta energía existe en cada rango. La escala Mel es la que dice exactamente cuánto distanciar cada uno de los filtros y qué tan anchos sean. De igual forma, para fines de comparar a cada artista, lo que se busca es el envolvente de la respuesta en frecuencia, en lugar de la frecuencia en sí; los filtros Mel, que son triangulares, logran realizar esta tarea. Estos filtros triangulares logran también reducir el número de coeficientes que describen el segmento. (19)

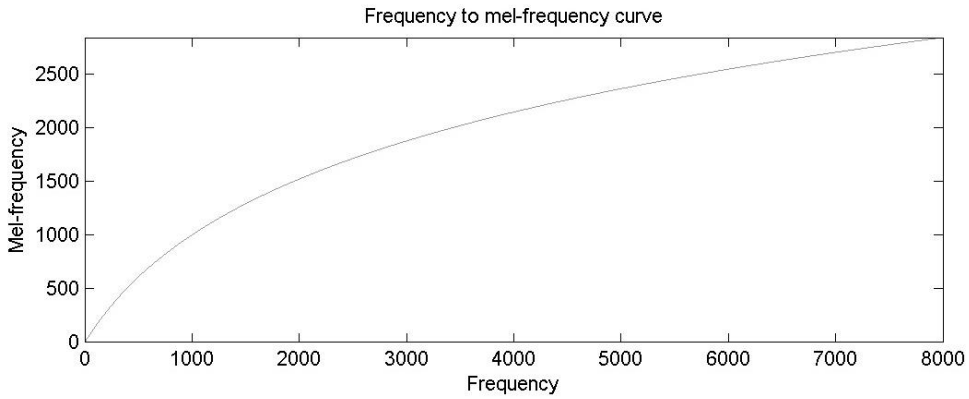
  Al realizar este proceso, como se mencionó anteriormente, se multiplica la magnitud de la respuesta en frecuencia por un conjunto de 20 filtros pasa-banda triangulares para obtener la energía en cada filtro triangular. Las posiciones iniciales de estos filtros están dados por la siguiente ecuación: (17)

$$mel(f)=1125*ln(1+f/700) \quad (8)$$

  Esta escala se construye equiparando un tono de 1000Hz y a 40 dBs por encima del umbral de audición del oyente, con un tono de 1000 mels. Alrededor de los 500Hz, los intervalos de frecuencia espaciados exponencialmente son percibidos

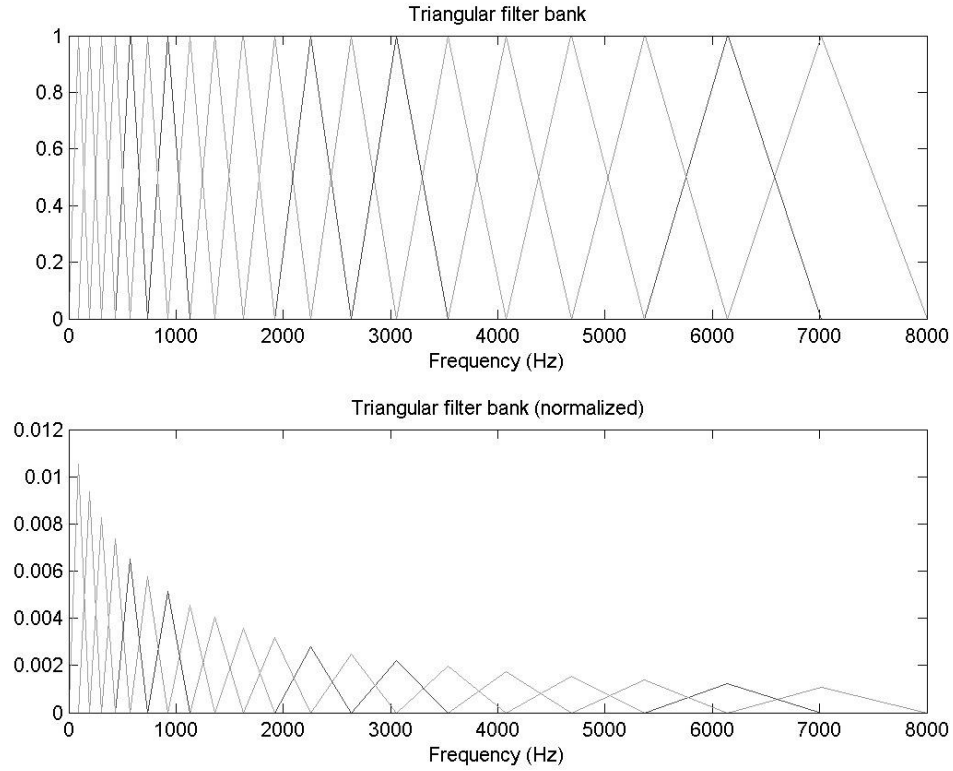
como si estuvieran espaciados linealmente. Por tanto, cuatro octavas en escala línea de frecuencias en hercios se comprimen alrededor de dos octavas en escala mel. Esta es la razón por la que el número de coeficientes decrece, como se dijo previamente.

Figura 8: Gráfica de la relación entre frecuencia lineal y frecuencia "Mel"



En práctica, existen dos formas de implementar el banco de filtros, se muestran a continuación:

Figura 9: Modelos que se pueden usar para la implementación de filtros Mel.



Para este trabajo, se programó el primer banco de la Figura 9, debido a la simplicidad de la implementación. Los coeficientes resultantes representan el

envolvente espectral de la señal. Lo que hace a esta técnica relevante es justamente el hecho que computa valores de energía para frecuencias que son críticas para el oído.

- Transformada Discreta de Coseno (DCT): La DCT es una transformación que forma parte de las transformadas sinusoidales unitarias. Estas transformadas son reales, ortogonales y relativamente rápidas de calcular con ayuda de otros algoritmos. Éstas tienen una gran relevancia cuando se necesita compresión de datos. Dentro de la familia de transformadas unitarias sinusoidales se encuentran muchas otras transformadas aparte de la DCT y DST (Discrete Sine Transform).

  La familia de transformadas trigonométricas discretas consiste en 8 versiones de la DCT. Cada una de estas es par o impar, y de tipo I, II, III o IV. Considerando los 4 tipos de DCT par, se tienen las siguientes características:

  - DCT-I: Definida para el orden N+1.

$$X_k = \frac{1}{2}(x_0 + (-1)^k x_{N-1}) + \sum_{n=1}^{N-2} x_n \cos\left[\frac{\pi}{N-1} nk\right] \qquad k = 0, \ldots, N-1. \tag{9}$$

  - DCT-II: Excelente compactación de energía, es la mejor aproximación para una transformada óptima.

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right] \qquad k = 0, \ldots, N-1. \tag{10}$$

  - DCT-III: Es la transformada inversa de la DCT-II.

$$X_k = \frac{1}{2}x_0 + \sum_{n=1}^{N-1} x_n \cos\left[\frac{\pi}{N} n\left(k + \frac{1}{2}\right)\right] \qquad k = 0, \ldots, N-1. \tag{11}$$

  - DCT-IV: Se logra implementar rápidamente.

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)\left(k + \frac{1}{2}\right)\right] \qquad k = 0, \ldots, N-1. \tag{12}$$

  Para todas las expresiones anteriores, existen N números reales x_0, ..., x_{N-1} que son transformados a N números reales X_0, ..., X_{N-1} .

  Aplicando esta transformada al trabajo, la DCT logra la de-correlación de los datos procesados; por lo que se pueden usar matrices con covarianza diagonal para modelar los resultados después del proceso. Esta transformada entonces modifica una señal en el dominio de frecuencia a uno parecido al de tiempo, que es llamado dominio “quefrency” (20).

- Modelos de Mezclas Gaussianas (GMM, Gaussian Mixture Models): Un modelo de mezclas, es un modelo probabilístico que se usa para representar subconjuntos dentro de un conjunto total, sin necesidad de que una serie de datos observados identifique al subconjunto al que pertenece cada uno de ellos. Formalmente, un modelo de mezclas corresponde a la mezcla de distribuciones que representa a su vez la distribución de probabilidad en el conjunto total. Este modelo hace que sea difícil producir información acerca del subconjunto, sino solamente se pueden realizar inferencias estadísticas acerca de éste.

  Para lograr implementar estos modelos, generalmente se usan supuestos acerca de las características de las observaciones individuales, que en esos casos pueden ser vistos estos pasos como una técnica de aprendizaje no supervisada.

  Un modelo de mezclas de dimensión finita es un modelo jerárquico que consiste en los siguiente elementos:

- *N* variables aleatorias que corresponden a las observaciones hechas, cada una se asume que está distribuida acorde a un número de *K* componentes; con cada componente perteneciendo a la misma familia de distribuciones (Normal, Zipfian u otras) pero con diferentes parámetros.
- *N* variables latentes aleatorias que especifican la identidad del componente de la mezcla de cada observación.
- Un conjunto de *K* pesos para cada mezcla, cada uno con valor entre 0 y 1; y donde la suma total es igual a 1.
- Un conjunto de *K* parámetros, cada uno especificando el parámetro del componente correspondiente. En muchos casos, cada "parámetro" es en realidad un conjunto de parámetros. Por ejemplo, en una observación distribuido según una mezcla de *V*-dimensiones, tendrá un vector de *V* probabilidades, todas debiendo sumar 1.

Matemáticamente, los parámetros básicos de un modelo de mezclas, en este Gaussianas, pueden ser descritos como:

K = Número de componentes en la mezcla
N = Número de observaciones
$\theta_{i=1\ldots K}$ = Parámetro de distribución de la observación asociada al componente i
$\phi_{i=1\ldots K}$ = Peso de la mezcla, para cada componente particular i
$\boldsymbol{\phi}$ = Vector K-dimensional compuesto por las $\phi_{i=1\ldots K}$ individuales que suman 1
$z_{i=1\ldots K}$ = Componente de la iésima observación
$x_{i=1\ldots K}$ = iésima observación
$F(x|\theta) =$ Distribución de probabilidad de la observación, parametrizada con θ

Los modelos de mezcla paramétricos son comúnmente usados cuando se conoce la distribución *Y* y se pueden obtener observaciones de *X*, pero se desean determinar los valores de medias y varianzas de ella. Estas situaciones generalmente ocurren en estudio en los que se obtienen muestras de una población que consiste en distantas sub-poblaciones.

Comúnmente, el modelado de mezcla de probabilidades es un problema que tiene el inconveniente de no tener todos los datos a disposición. Una forma de entender esto, es asumir que los datos que se están considerando tienen una especie de membresía en una de las distribuciones que están sirviendo para modelar los datos. Al empezar, esta membresía es desconocida. Por esto se realiza un proceso de estimación, el cual busca los parámetros correctos para la función del modelo que se elige. Existen varias formas de resolver este problema, con la mayoría enfocándose en métodos llamados de Expectación-Maximización (EM) o Estimación Máxima a Posteriori (MAP). Generalmente estos métodos consideran de forma separada la pregunta de estimación de parámetros y la del sistema de identificación, esto es para hacer una distinción entre la determinación del número y forma funcional de los componentes dentro de una mezcla y la estimación de los parámetros correspondientes.

- Expectación Maximización: Es una de las técnicas más populares usadas para determinar los parámetros de una mezcla con un número de componentes dado a priori. Es decir, si se etiqueta con *Z* al conjunto del cuál se buscan sus parámetros, se puede decir que *Z=(X,Y),* donde *X* es el conjunto de datos del que ya se dispone, o

conjunto visible; e Y es el conjunto de datos desconocidos o datos ocultos. Por medio del algoritmo, se hace un cierto número de iteraciones que llevan una máxima verosimilitud. Éste fue desarrollado en 1977 por Dempster, Laird y Rubin. Esta es una forma particular de implementar la estimación de Máxima Verosimilitud para este problema. Las ecuaciones presentes en este algoritmo son:

$$\mu_s^{(j+1)} = \frac{\sum_{t=1}^{N} h_s^{(j)}(t) x^{(t)}}{\sum_{t=1}^{N} h_s^{(j)}(t)} \quad (13)$$

$$\Sigma_s^{(j+1)} = \frac{\sum_{t=1}^{N} h_s^{(j)}(t)[x^{(t)} - \mu_s^{(j)}][x^{(t)} - \mu_s^{(j)}]^\top}{\sum_{t=1}^{N} h_s^{(j)}(t)} \quad (14)$$

$$h_s^{(j)}(t) = \frac{w_s^{(j)} p_s(x^{(t)}; \mu_s^{(j)}, \Sigma_s^{(j)})}{\sum_{i=1}^{n} w_i^{(j)} p_i(x^{(t)}; \mu_i^{(j)}, \Sigma_i^{(j)})}. \quad (15)$$

Teniendo en cuenta estas ecuaciones y basándose en la estimación para estos parámetros, la probabilidad condicional para una observación *x(t)* que fue generada de un estado s es determinada por cada *t = 1, ..., N; N* siendo el tamaño de la muestra. Los parámetros luego se actualizan para que los nuevos componentes de peso correspondan al promedio de la probabilidad condicional y cada componente de media y varianza es actualizado de igual forma. Dempster demostró que cada iteración en el algoritmo EM, no hará decrecer la verosimilitud, una propiedad que no sucede con otras técnicas de maximización.

El algoritmo, como su nombre lo indica, tiene dos pasos: El de expectación y el de maximización. El primero de estos, trata de adivinar la membresía parcial de cada punto en cada distribución computando los valores esperados por las variables de membresía de estos puntos. Es decir, para cada punto x_j y distribución Y_i, la membresía está dada por:

$$y_{i,j} = \frac{a_i f_Y(x_j; \theta_i)}{f_X(x_j)}. \quad (16)$$

Seguidamente, se realiza el paso de maximización, en el que se vuelven a computar los nuevos parámetros de distribución.

Los coeficientes de la mezcla a_i son las medias de la membresía sobre la *N* cantidad de datos.

$$a_i = \frac{1}{N} \sum_{j=1}^{N} y_{i,j} \quad (17)$$

Los parámetros θ_i también son calculados usando los datos x_j que han sido medidos usando los valores de la membresía. Por ejemplo, si θ es una media, entonces:

$$\mu_i = \frac{\sum_j y_{i,j} x_j}{\sum_j y_{i,j}}. \quad (18)$$

Con estos nuevos valores, se repite el paso de expectación para tener nuevos valores de membresía. El procedimiento sigue hasta que se converge a un valor específico o hasta cierto número de iteraciones.

- Decisión de máxima verosimilitud: El teorema de Bayes es el medio estadístico que permite la clasificación de los intérpretes por medio del modelado de los GMM. Este método es usado en diferentes aplicaciones y se basa en computar la probabilidad a posteriori de que dado cierto conjunto de datos *X*, este conjunto pertenezca a la distribución de alguna hipótesis formulada. Para esto, se usa la siguiente ecuación:

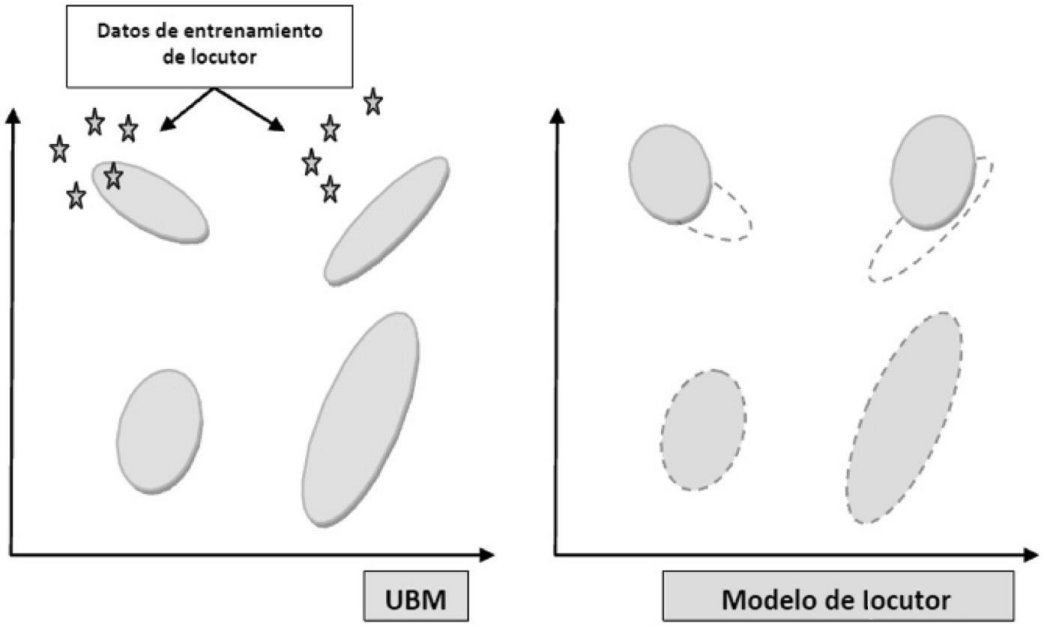
$$p(w_i|\mathrm{X}) = \frac{p(X|w_i)p(w_i)}{p(X)} \quad (19)$$

Donde: $p(w_i|x)$ es la probabilidad a posteriori de que agregando el conjunto de datos *X*, este último pertenezca a la clase i.
$p(X|w_i)$ es la probabilidad de que dada la clase w_i, el valor de la variable aleatoria sea *X*.
$p(w_i)$ es la probabilidad para cada una de las potenciales clases.
$p(X)$ es la probabilidad total.

- GMM-UBM: El modelo usado para la identificación de un artista en la pista puede variar según la información que se haya logrado obtener para modelar la voz de cada uno de ellos. En este trabajo se usa el UBM, que viene de las siglas Universal Background Model. El UBM ha demostrado ser una de las mejores opciones cuando se trabaja con GMM debido a que logra caracterizar a la población de manera fiable para el sistema y fue creado en 1995 por D.A. Reynolds.

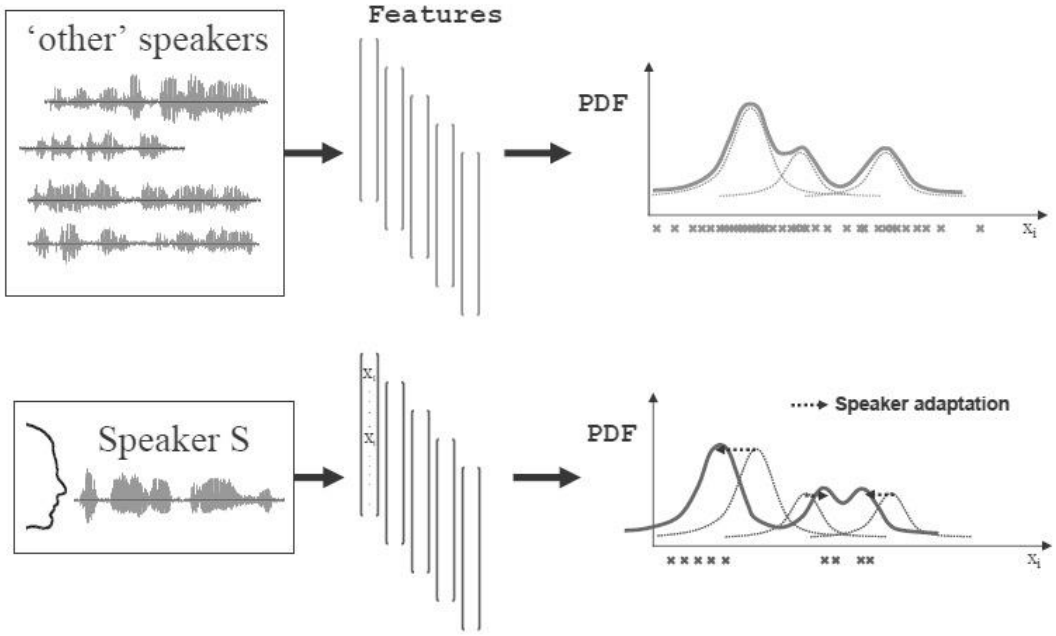
Este método consiste en tener un GMM que modele a todos los participantes de la base de datos para poder una representación de todos los posibles locutores en un solo conjunto. Este GMM universal existe tanto para las secciones con voz como para los que no tienen. Se crearán de igual forma, GMM para cada artista, pero estos GMM saldrán del GMM universal. Ahora bien, dado un conjunto de vectores de entrada, el GMM universal se verá modificado con estos datos de cada artista para cambiar sus parámetros, como lo muestra la figura 10. Entonces, para que hagan un puntaje al momento de pruebas, se resta el puntaje del GMM del presunto artista y el del GMM universal. Luego, se toma el valor más grande que representa el artista con mayor probabilidad de ser el que aparecía en la grabación.

Figura 10: Explicación gráfica del UBM



Partiendo de la idea de que vamos a representar a todos los artistas en este conjunto, es necesario que acá se introduzca un gran número de coeficientes para lograr describir a cada artista de una forma apropiada. Otro ejemplo de cómo se generan los GMM a partir de un UBM aparece en la siguiente figura. En ésta, cabe mencionar que los vectores en color azul son de entrenamiento y sirven solamente para modificar y no para realizar puntajes.

Figura 11: Ejemplo de la modificación de un GMM según cierto conjunto de vectores de entrenamiento.



V. ANTECEDENTES

La voz que se usa para cantar, además de ser el instrumento musical más antiguo, es también uno de los más complejos desde un punto de vista acústico. La investigación en la percepción del canto no está tan avanzada como en el reconocimiento de palabras. En esta sección se presentan algunos estudios previamente realizados por otras entidades o personas en reconocimiento del intérprete en canciones.

Chou y Gu (7) han utilizado GMM's para detectar regiones en la señal de audio que contienen voz. Los vectores característicos usados para el GMM incluyen modulación de energía en 4Hz, coeficientes armónicos de 4Hz y MFCC.

Berenzweig y Ellis (7) han usado un clasificador de reconocedor de palabras para distinguir segmentos vocales del acompañamiento. Se ha mencionado que aunque la voz cantada no es lo mismo que la voz en el habla normal, comparten ciertos atributos como la estructura de los formantes (pico de intensidad en un espectro de sonido) y características fónicas. Así estos modelos, analizando únicamente la voz, pueden responder diferente a los otros instrumentos. Tres conjuntos característicos han sido explorados: características de probabilidad posterior, valores estadísticos derivados como clasificadores de entropía y promedios de estos valores.

Kim y Whitman (2) han desarrollado un sistema para identificación de cantantes en grabaciones de música popular usando códigos característicos para las voces. Como un primero paso, un algoritmo es usado para extraer automáticamente segmentos vocales. Una vez estos segmentos han sido identificados, son presentados a un sistema de identificación con una base de datos previa. Para detectar la voz en el audio, primero se filtra la señal por un filtro pasa banda, que deja pasar las frecuencias comprendidas entre 200 a 2000 Hz. Esto se hace con un filtro IIR Chebychev. Para filtrar los otros instrumentos presentes en estas frecuencias, un filtro inverso comb es aplicado para obtener la frecuencia fundamental a la que la señal es mayormente atenuada. La armonicidad ha sido definida como la relación de la señal total de energía a la señal más atenuada armónicamente. Al aplicar un "threshold" a la armonicidad contra un valor fijo, se obtiene un detector de armónicos. La hipótesis detrás de esto es que la mayoría de estas secciones corresponden a regiones de voz cantada basado en su alta armonicidad comparado con los otros instrumentos.

Otro sistema fue propuesto por Zhang (12). Este es un proceso de dos pasos que comprende una fase de alimentar al sistema con un modelo estadístico creado para cada voz de cada cantante, y una fase de comparación donde se detecta el punto donde comienza la voz y se toman datos desde ese punto. Los datos de audio que se extrajeron se comparan con los modelos ya existentes para lograr la identificación del cantante. La detección de la voz se logra por medio de comparaciones de energía, average zero-crossing rate (ZCR, promedio de intersecciones por cero), coeficientes armónicos y flujo espectral computado en intervalos definidos que luego son comparados contra un conjunto de thresholds predeterminados.

Un sistema para agrupar música popular basado en las características de la voz ha sido propuesto por Tsai (9). Él aplica métodos para separar regiones que contienen voz de las que no la tienen, modelar características de la voz y agrupamientos basados en las características del cantante. La detección de la voz en la señal, se hace en dos etapas. En la primera, un clasificador estadístico con modelos parametrizados es usado para realizar transcripciones de la voz. Dos GMM son usados para esto, uno dedicado a la parte con voz y la otra a la que no la tiene. En la segunda etapa, el reconocedor toma como entrada el vector característico extraído de la grabación y produce una salida, que es la probabilidad para el GMM vocal y no vocal. El vector característico analizado fue MFCC.

Otro sistema para la detección de cantantes fue presentada por Tsai y Wang (8), éste se enfoca en pistas con múltiples cantantes en ella. Varios métodos se presentan para separar regiones con voz y sin ella, para modelar características vocales y distinguir entre un cantante de otro en un momento particular en el que pueden estar varios cantantes. La separación entre regiones con voz de las que no tienen se logra usando un clasificador estocástico que consiste de un procesamiento de señales para extraer vectores característicos basados en su cepstro, seguidos por otro procesador estadístico que realiza el modelado y el reconocimiento. Para operar, se hace una base de datos y luego, se hacen las pruebas. En la parte de la base de datos, las regiones con voces se agregan manualmente diferenciándolas de las regiones sin voz (que también se agregan). El primer GMM es formado usando las regiones con voces de un cantante particular. El segundo y tercer GMM se generan con las regiones sin voz de la canción. Al momento de hacer el reconocimiento y comparación, el clasificador toma como entrada el vector característico extraído y calcula qué tan parecido es con los GMMs anteriores.

Bartsch (7) ha propuesto un sistema para identificación automática de cantantes en música popular. Un sistema de separación conocido como PESCE ha sido diseñado para resolver dos problemas: detección de la voz del cantante y la extracción de ésta. Este sistema es un algoritmo de estimación sobre la frecuencia fundamental aplicado a música polifónica. Se toma una señal corta de audio como entrada, y produce estimaciones de la frecuencia fundamental que están presentes en la señal. PESCE asume que la voz contiene una modulación en frecuencia significante mientras los otros instrumentos contienen una especie de constantes en su frecuencia. Con esto, las fuentes de voz son aquellas que exhiben modulación en frecuencia. Si no existen señales generadas a partir de la voz, PESCE no produce ninguna señal en la salida. La estimación de la frecuencia fundamental permitirá extraer variaciones de amplitud en el tiempo para la señal de una distribución frecuencia en tiempo, como un espectrograma.

New y Wang (7) han propuesto un modelo estadístico para clasificar segmentos de audio musical separando regiones con voz usando HMM (Hidden Markov Model) como clasificador. La extracción se basa en un proceso que usa LFPC (Log Frequency Power Coefficients) para proveer una idea de cómo se da la energía a lo largo de las sub-bandas. También en este sistema se toman en consideración tanto el tempo como la estructura de la canción basados en las variaciones observadas en las características de la señal. Igualmente, en contraste del HMM convencional que usa un modelo para cada clase, el método aquí usa una técnica multi-modal para mayor precisión comparado con el método común.

Nwe y otros (7) han mejorado el modelo anteriormente mencionado incorporando una atenuación armónica a la señal de entrada usando las frecuencias del tono de la de la canción.

Maddage y otros (7) han adoptado un TICFT (Twice-iterated Composite Fourier Transform) para detectar donde aparece y termina la voz. El TICFT es computado sobre cada frame donde la magnitud del espectro de la primera transformada de Fourier es la entrada de una transformada rápida de Fourier. La parte de la voz se separa de las regiones sin voz usando un threshold lineal que se aplica sobre la energía de la segunda FFT. Reglas heurísticas basados en los acordes y patrones de la canción también se aplican para mejorar el reconocimiento del cantante.

Tzanetakis (10) ha propuesto un sistema semi-automático para resolver el problema localizando segmentos donde aparece la voz. En este sistema, una parte aleatoria de la canción es manualmente usada como entrada y la información tomada es usada para inferir el cantante de la canción entera. Por esto, un clasificador diferente se aplica a cada canción usando la información tomada de la canción. El conjunto característico usado consiste de un promedio y una derivación standard del centro, debilitamientos y flujo y el promedio de energía relativa en las sub-bandas que ocupan el cuarto más bajo y el segundo cuarto más bajo del ancho de banda. La mejor generalización de desempeño se obtuvo usando regresión lógica como clasificador y “neural networks”.

VI. DISEÑO EXPERIMENTAL

- Búsqueda bibliográfica.
- Seleccionar el algoritmo a implementar según el conocimiento que se tenga.
- Crear software que tome datos de archivos de audio.
- Programar el algoritmo para el tratamiento de los archivos de audio.
- Crear la base de datos.
- Experimentar con algunos parámetros del algoritmo para saber con cuál de estos responde mejor.

La figura siguiente muestra en diagrama de bloques los pasos que se realizaron para el diseño experimental:

Figura 12: Diagrama de bloques del diseño experimental

La búsqueda bibliográfica que se realizó como primer paso del Diseño Experimental consistió en investigar sobre los algoritmos que se han usado anteriormente en el reconocimiento de la voz y saber en qué exactamente consistía

cada uno de ellos. Los algoritmos encontrados son los que se describieron en la sección anterior "Antecedentes". Todos estos fueron encontrados como publicaciones en formato .pdf en internet y en ellos los autores explicaban lo que habían realizado como experimento sin dar detalle acerca de los parámetros o equipo usado, sino simplemente una visión general de sus resultados y el algoritmo que usaron para ello.

Luego se procedió a escoger el algoritmo a implementar y para ello se realizó una tabla comparativa con todos los algoritmos encontrados en la sección previa y se tomaron diferentes cualidades de cada uno de ellos: el nivel de conocimiento necesario para entender el algoritmo, su popularidad, la documentación disponible para lograr entender el algoritmo y el porcentaje de error que habían arrojado estas implementaciones.

Tabla 1: Tabla comparativa de algoritmos a implementar

Algoritmo	**Nivel de conocimiento necesario para entender el algoritmo***	**Popularidad****	**Documentación disponible*****	**Porcentaje de error en los experimentos con estos algoritmos**
Chu/Gu	Alto	Media	Poca	70%
Berenzweig	Alto	Baja	Poca	No disponible
Kim/Whitman	Justo	Baja	Suficiente	78%
Zhang	Justo	Baja	Suficiente	80%
Tsai	Justo	Alta	Suficiente	71%
Tsai/Wang	Justo	Alta	Poca	70%
Illinois	Justo	Alta	Suficiente	No disponible
Maddage	Alto	Baja	Poca	No disponible
Fujihara/Goto	Alto	Media	Poca	88%
Shen	Alto	Baja	Poca	76%
Chien/Wang	Alto	Baja	Poca	71%
Khine/Nwe	Alto	Baja	Poca	84%
Bartsch	Alto	Baja	Poca	No disponible

* En la columna de "Nivel de Conocimiento Necesario para Entender el Algoritmo" aparecen "Alto" y "Justo" para definir si es posible con el conocimiento hasta el momento, lograr hacer la implementación; siendo el primero un nivel fuera del entendimiento para lograr realizarlo en el tiempo dado; y el segundo, un nivel apropiado para conseguir este objetivo.

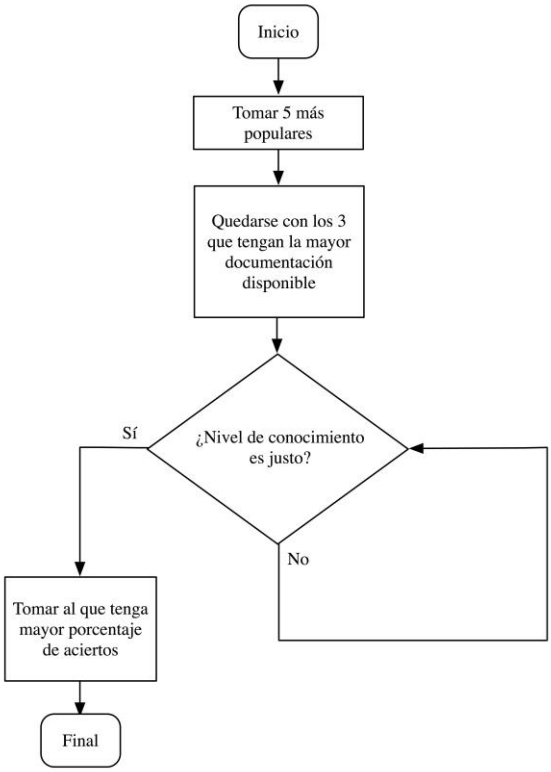
** La popularidad de los algoritmos se midió según qué tanto se ha usado este algoritmo para implementaciones donde se busca resolver el problema. Una popularidad baja se usa para decir que ese algoritmo solo se ha usado para ese experimento en específico y nada más. Una popularidad media refleja un algoritmo

que se ha usado un par de veces con variaciones pequeñas en sus parámetros. Un algoritmo con popularidad alta ha sido usado varias veces y se ha profundizado en su estudio para encontrar resultados haciendo variaciones en el mismo.

*** La Documentación Disponible se distingue de la Popularidad, en que un algoritmo puede ser muy popular, pero se ha publicado poco acerca de cómo se implementa, por lo que puede faltar información para lograr implementarlo. La Documentación Disponible se midió según el número de veces que se encontraba el algoritmo: “Poca” se usa para decir que el algoritmo se encontró en un máximo de 2 publicaciones. “Suficiente” quiere decir que el algoritmo aparece al menos en 3 publicaciones.

A continuación, se detalla el procedimiento a seguir para la elección del algoritmo a implementar. Del lado derecho aparece en diagrama de bloques el procedimiento que se realizó, y del lado izquierdo, los algoritmos que iban siendo escogidos en cada fase.

Figura 13: Diagrama de flujo para representar el proceso de selección del algoritmo a implementar



5 más populares:
- Chu/Gu
- Tsai
- Tsai/Whang
- Fujihara/Goto
- Illinois

3 con mayor documentación:
- Tsai
- Illinois
- Tsai/Whang

Nivel de conocimiento necesario:
- Tsai

Una vez elegido el algoritmo se procedió a buscar alguna forma de llegar a reproducirlo para tratar de encontrar si éste trabaja mejor con cierto género musical o no, como era el principal objetivo de este trabajo.

El algoritmo se programó en el lenguaje Python, en su versión 2.7. Esto se hizo debido al consejo del Ing. Javier Mesalles, que ya había trabajado con archivos

musicales en el curso de procesamiento señales, particularmente con archivos .wav, de igual forma. El programa final cuenta con una interfaz gráfica para seleccionar y procesar el archivo, una sección para desplegar los resultados y un tutorial. Estos detalles se explican en la sección de "Anexos", en la que se incluye el manual de usuario respectivo.

En la implementación del algoritmo existen varios parámetros que pueden modificarse según lo desee el diseñador, debido a la estructura del mismo. Por esto, se realizaron varios experimentos tratando de encontrar un nivel cada vez más alto de aciertos al momento de reconocer el artista. Los parámetros que se pueden modificar son: La cantidad de canciones en la base de datos para cada artista, el threshold para separar extractos vocales de los no vocales, el número de bloques a comparar de la canción que se busca reconocer, el número de componentes del GMM, cambiar el número de coeficientes que se usan para las comparaciones, cambiar proporciones de datos de los GMM universales. En este trabajo, debido a la limitante del tiempo, solo se hicieron modificaciones al threshold, cantidad de canciones en la base de datos, número de componentes en GMM y número de coeficientes en las comparaciones.

Los elementos que componen la implementación del algoritmo son los que se presentan a continuación:

- Para programar y lograr ejecutar el programa final, es necesario tener instalado en una computadora (personal o de escritorio, sin importancia) Python en su versión 2.7. La razón por la que se eligió esta versión y tal vez no una más reciente, fue que esta es una versión ya "consolidada"; la mayoría de librerías están actualizadas para funcionar sin "bugs" y prácticamente cualquier librería está disponible para trabajar sin problemas con ésta. Además, se instalaron 3 librerías extra para lograr usar ciertos métodos y clases que forman parte del algoritmo. Estas son: Numpy, Scipy y Scikit Learn. Todas para la versión de Python 2.7, pero es necesario mencionar que el archivo con el cual se instalaron depende del procesador de la computadora en el cual se correrá el programa, las únicas opciones son 32 o 64bits. Numpy y Scipy son librerías para el manejo numérico de los datos, mientras que Scikit Learn se encarga de realizar los GMM leyendo la base de datos. Estas librerías son gratuitas y están disponibles para su descarga en línea.
- Todos los archivos que forman parte de la base de datos y que se usaron para las pruebas, están sampleados a 22.050KHz. Esto se hace para eliminar las altas frecuencias que puedan estar presentes en el archivo (siguiendo el teorema de Nyquist si se sampleó a 22.050KHz, la frecuencia máxima presente debe ser 11.025KHz; además la voz, que es lo que se busca comparar, generalmente aparece entre 400 a 3400Hz) y para mantener un estándar sobre todos los archivos a procesar para evitar problemas en alguna malinterpretación de resultados. Igualmente, todas las canciones tienen 16 bits por muestra. La otra razón para esto, es que el algoritmo original usa estos valores.

  Por otro lado, todos estos archivos musicales tienen extensión .wav, debido a que estos archivos son los más simples y es más fácil procesarlos con librerías de Python disponibles. Hoy en día, la mayoría de archivos musicales son .mp3 debido al poco espacio que ocupan en memoria (un archivo .mp3 de 3Mb es cerca 10Mb en formato .wav), por lo que es muy complicado encontrar

canciones específicas en el formato .wav. Lo que se hizo, primeramente, fue conseguir los archivos de audio para probar y generar la base de datos; todas se encontraron en Youtube.com., y se descargaron como .mp3 usando una página en línea que descarga cualquier audio de cualquier video en Youtube solamente sabiendo el link donde se encuentra. Posteriormente, se usó el software "Audacity", el cual cambiaba la tasa de muestreo y generaba el archivo .wav que se describió anteriormente.

- Existen 10 artistas en la base de datos, es decir, cualquier canción que se use para reconocer su intérprete, tendrá solamente 10 posibles soluciones (artistas) que no pueden ser modificadas. Los 10 artistas son: Metallica, Oasis, Red Hot Chili Peppers, U2, Coldplay, Carly Jepsen, Nirvana, The Beatles, Maroon 5 y The Killers. Estos artistas fueron elegidos debido a que se conocían de anticipación y su popularidad hacía que encontrar versiones pirata de sus canciones no fuera tan complicado. La base de datos, como se describe posteriormente, fue variando en tamaño para observar si existía algún tipo de cambio en el porcentaje de aciertos del artista. Esta base de datos la conforman únicamente temas originales de los artistas.
- Todas las canciones, por otro lado, que se usaron para hacer las pruebas son canciones no originales que fueron modificadas por otro software y como se mencionó previamente, fueron encontradas en Youtube.com ya alteradas. Estas canciones sí contienen al artista original, pero con variaciones en el ritmo que lo acompaña. Ya que el objetivo central de este experimento era saber si el algoritmo original funcionaba mejor para cierto género musical, se buscó que estas canciones tuvieran diversos géneros; los que se encontraron son los siguientes: Rock, pop, acústico, tropical, electrónica, bossa nova, jazz, metal, balada. No todos estos géneros están presentes para todos los artistas. El encontrar canciones con voz original pero diferentes géneros fue un trabajo arduo debido a que es realmente atípico realizar esto; comúnmente las versiones pirata de una canción contienen una voz diferente a la del artista original.
- El programa en su versión final no llega a ocupar más de 18MB de espacio en disco, pero es necesario mencionar que realiza una gran cantidad de operaciones para poder analizar el archivo de audio. La computadora en la que se realizaron la mayoría de pruebas, una de 64bits, con 4GB de RAM y 1TB de disco duro, presentaba problemas al momento de procesar archivos con más de 10MB de tamaño, por lo que se sugiere revisar el tamaño de la pista de audio antes de proceder a realizar el reconocimiento de la canción. De igual forma, se recomienda tener más de 100GB de espacio en el disco duro si se va a utilizar el sistema operativo de MacOS, donde también se realizaron pruebas para verificar el correcto funcionamiento del programa.

Todos los experimentos que se muestran a continuación fueron realizados en una computadora Dell con las siguientes características:

- Sistema operativo Windows 7 Professional
- Service Pack 1
- Modelo Precision T7500
- Procesador Intel®Xeon® CPU, E5620 @ 2.40GHz
- Memoria instalada de 4GB RAM
- Sistema operativo de 64 bits

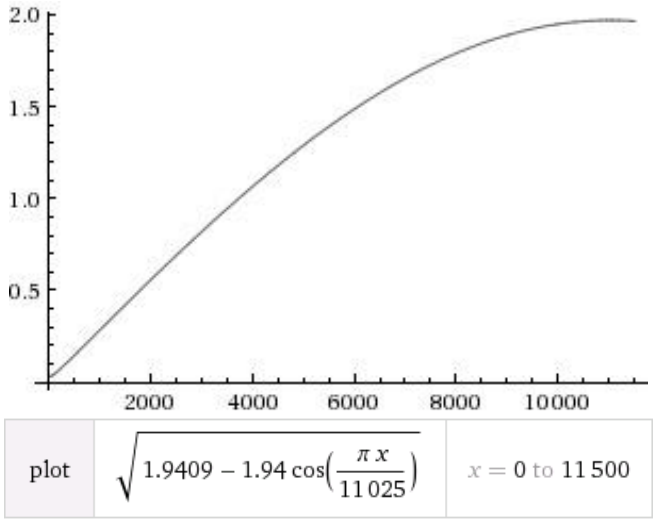
El proceso del experimento es el siguiente:

De la interfaz gráfica, se selecciona un archivo en el cual se desconoce al artista que la interpreta y es el que se busca identificar. Internamente, lo que hace el programa es lo siguiente: Toma el archivo y como primer paso, lo pasa por un filtro de pre-énfasis para evitar que las frecuencias altas de la voz se atenúen, además con esto se evitan problemas de precisión finita dentro del programa y se logra amplificar la zona que está arriba de 1KHz en el espectro, en donde según los diseñadores del algoritmo están los aspectos más importantes de la voz. La ecuación de diferencias del filtro digital de pre-enfásis es la siguiente:

$$H(z) = 1 - 0.97 * z^{-1} \quad (19)$$

El valor de 0.97 se eligió debido a que se busca atenuar más las frecuencias bajas (un valor más bajo tiene el efecto contrario) además que es el que más se ha usado según la bibliografía consultada, por lo que para esta aplicación que tiene a la voz involucrada, parece ser un valor adecuado. El filtro tiene una gráfica de respuesta en frecuencia, que es la siguiente:

Figura 14: Gráfica de la respuesta en frecuencia del filtro de pre-énfasis implementado.



Para graficar esta función, se hizo la sustitución $z = e^{(j2\pi f)/fs}$ siendo f_s la frecuencia de muestreo. Posteriormente se procedió a computar el módulo de dicha función para terminar con la función que finalmente aparece en la figura 11. A partir de los 5512.25Hz, ya se ha superado la pérdida de 3dB.

La variabilidad de la voz es una dificultad que se debe solventar, ya que al analizar una señal de voz de principio a fin, se tienen demasiadas variaciones en muy pocos segundos y el tratamiento de la señal de esta forma no es viable. La voz, como se sabe, es una señal que varía en el tiempo, y aunque ésta sea originada por órganos humanos cuyos cambios de posición no pueden realizarse de manera instantánea, causa que en ciertos períodos de tiempo en los que los órganos se encuentran en una posición fija la señal emitida pueda ser considerada estacionaria. Para solucionar este

problema, se propone realizar el segmentado de la señal en intervalos relativamente cortos. Tal y como lo propuso Tsai en el algoritmo MFCC+GMM, este segmentado se hace cada 32 milisegundos. Se hace un traslape de 10 milisegundos entre cada bloque; es decir, el siguiente bloque empieza a partir de los 22 milisegundos y termina en los 54 milisegundos, el tercer bloque comienza a los 44 milisegundos y termina a los 76 milisegundos, y así sucesivamente. El traslape se hace para no perder información y que los datos que estaban al final de un bloque ahora sean los centrales en el siguiente. Al ser estos archivos muestreados a 22.050KHz, cada 32 milisegundos están representados por 705.6 bits, que es el largo de cada vector de información a analizar; al ser arreglos los que se usan el programa, se aproximó a 706 bits. El traslape respectivo hace que cada bloque empiece cada 22 milisegundos, que equivalen a 485.1 bits, pero se aproximó a 486 con el fin de agrandar el vector y no perder información. Estos vectores de los cuales se está hablando son los que se usarán para analizar cada parte de la canción y reconocer las partes que tienen voz y las que no. Por esto, la cantidad de vectores es: el número de bits que conforman la canción dividido 486.

Seguidamente, a cada vector se le multiplica por una ventana de Hamming, que busca que los extremos en cada bloque hagan efectos en el espectro que puedan deducir que ciertas frecuencias están presentes cuando no es así. Esto se hace para cada uno de los vectores de entrada, estén traslapados o no.

Para continuar, a cada vector de 706 bits, se le aplica la FFT para encontrar su respuesta en frecuencia. En esta parte cabe recalcar que es la que más problemas causa debido a la falta de memoria en el equipo de cómputo que realiza esta operación. Si el archivo a procesar es de 13MB, aparecerá este error. La FFT devolverá un nuevo vector de largo 354, ya que se usó la FFT truncada para esta transformación. Este vector contendrá números complejos también, por lo que para el análisis posterior esto se elimina al obtener el módulo del número respectivo.

Posteriormente, se multiplica esta señal por el banco de filtros Mel para saber la cantidad de energía en cada filtro y se suman los diferentes valores. A estos mismos valores se les calcula su logaritmo base 10 para finalmente, aplicar la DCT a cada vector. El vector resultante tiene entonces, 20 coeficientes al finalizar este proceso.

En la siguiente sección del programa, se realizan los GMM, que serán los encargados de modelar las características de voz de cada artista y las respectivas comparaciones para poder encontrar al artista que canta en la pista que se procesó. Primero, se realizaron 20 GMMs con música instrumental para cada uno para modelar cada dimensión (coeficientes) de la sección anterior. Para esto, se procesaron de antemano las canciones y se separaron las secciones que contenían voz de las que no, para armar estos GMMs. De igual forma, se generaron 20 GMMs pero con información acerca de la voz de cada artista de la base de datos. Una vez hecho esto, se usó el método GMM-UBM para diferenciar a cada artista de la base de datos. Cuando solamente se tenía una canción para cada artista en la base de datos, la parte universal del GMM conformaba el 62.5% de los datos y el resto, del artista que se buscaba modelar. Por otro lado, cuando se incrementó la base de datos, la parte universal era el 85.71% de los datos, el resto era propio del artista. Estos valores se tomaron de manera aleatoria ya que no se encontró bibliografía al respecto. Para la primera parte de las pruebas se usó una canción solamente en la base de datos y ésta

aportaba 6400 coeficientes disponibles para realizar todos los GMM. En la segunda parte, se agrandó a una base de datos más "completa" agregando 3 canciones por artista y teniendo 54000 coeficientes para modelar las características de cada uno de éstos. En otras palabras, en la primera base de datos, se tenían 10 segundos de voz y de música instrumental por artista; en la segunda base, el número de segundos se expandió a 86.4.

Al momento de pedirle al programa que busque identificar al artista presente en la canción, los coeficientes son evaluados para verificar primeramente, si existe voz dentro del bloque procesado, esto lo logra verificando si el valor de puntaje es mayor al umbral que se le indicó. En caso de ser cierto esto, por la propiedad de decisión de máxima verosimilitud del GMM se decide cuál es el artista que es el más parecido, esto lo logra hacer al comparar los 10 GMM para cada artista en sus 20 coeficientes y realizando puntajes que son los que se comparan finalmente. Al artista que se le otorgue el mayor puntaje, ese es el que apunta el programa que es el intérprete que se buscaba.

A continuación se muestra un diagrama flujo para mostrar cómo se dió la implementación del algoritmo escogido (MFCC + GMM).

Figura 15: Primera parte del algoritmo implementado

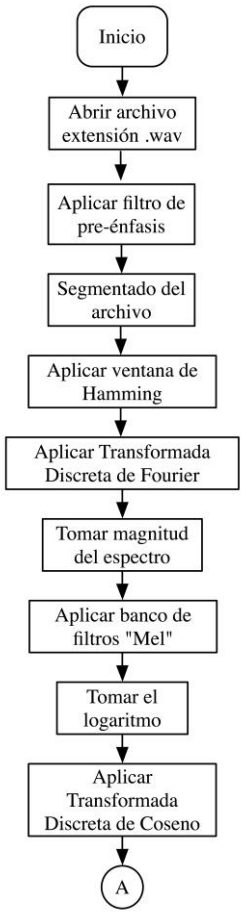
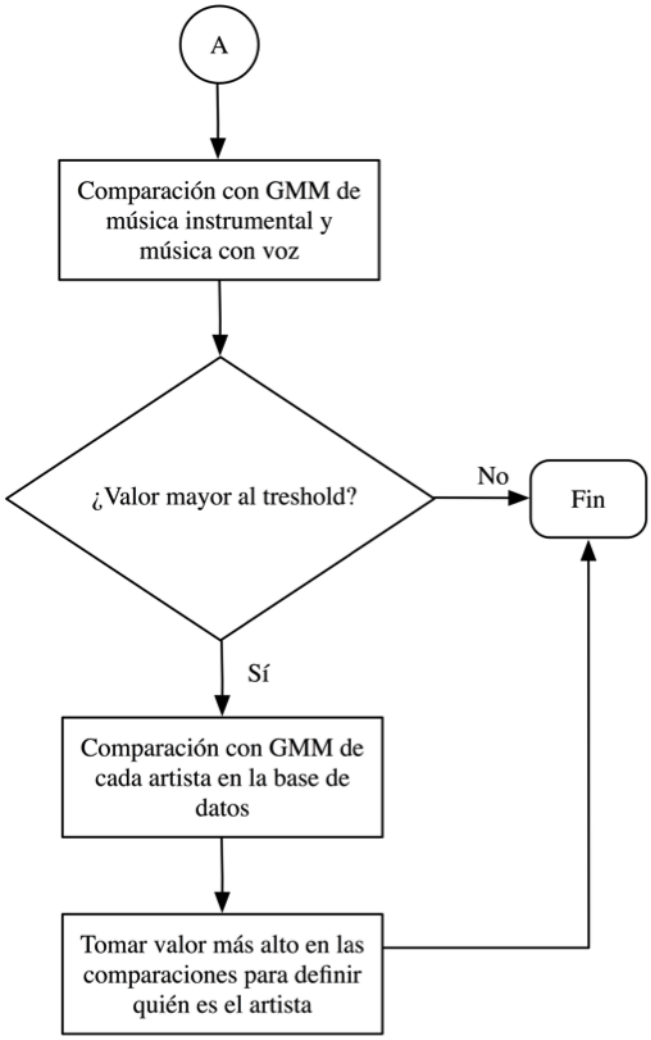


Figura 16: Segunda parte del algoritmo que se llegó a implementar



VII. RESULTADOS

Como se indicó anteriormente, existen muchos parámetros que pueden ser modificados para el programa. Acá se presentan los resultados según los parámetros establecidos:

- 1 canción para modelar el GMM de cada artista
- 2000 filas de vectores de coeficientes de DCT a procesar
- Umbral = 0.39
- 2 componentes en cada GMM
- Comparación de los 20 coeficientes del vector de cada bloque

Tabla 2: Resultados de la efectividad en la primera prueba del experimento

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Maroon 5	Rock	Maroon 5	Sí
Maroon 5	Rock	Coldplay	No
U2	Rock	Oasis	No
Coldplay	Pop	Coldplay	Sí
The Killers	Rock	Red Hot Chili Peppers	No
U2	Pop	Oasis	No
The Beatles	Acústico	Oasis	No
Maroon 5	Pop	Oasis	No
U2	Rock	The Killers	No
The Beatles	Pop	U2	No
Nirvana	Rock	Nirvana	Sí
The Beatles	Rap	Red Hot Chili Peppers	No
Red Hot Chili Peppers	Bossa Nova	U2	No
Coldplay	Balada	U2	No
Coldplay	Pop	Coldplay	Sí
Red Hot Chili Peppers	Acústico	Carly Rae	No
Carly Rae	Bossa Nova	Carly Rae	Sí
Carly Rae	Electrónica	Carly Rae	Sí
Carly Rae	Rock	Carly Rae	Sí
Carly Rae	Tropical	Carly Rae	Sí
Metallica	Metal	Carly Rae	No
Coldplay	Electrónica	Coldplay	Sí
Carly Rae	Bossa Nova	Coldplay	No
Maroon 5	Tropical	Coldplay	No
Carly Rae	Tropical	Coldplay	No
Maroon 5	Metal	Coldplay	No
The Beatles	Electrónica	Red Hot Chili Peppers	No
Oasis	Metal	Metallica	No
Nirvana	Balada	Coldplay	No
Red Hot Chili Peppers	Tropical	Maroon 5	No
U2	Balada	Maroon 5	No

Continuación Tabla 2

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Carly Rae	Tropical	Maroon 5	No
Red Hot Chili Peppers	Jazz	Maroon 5	No
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Electrónica	Maroon 5	Sí
Metallica	Electrónica	Metallica	Sí
The Beatles	Acústico	Metallica	No
The Beatles	Jazz	Metallica	No
Carly Rae	Electrónica	Nirvana	No
Maroon 5	Balada	Metallica	No
The Beatles	Acústico	Oasis	No
Maroon 5	Tropical	Oasis	No
Metallica	Balada	U2	No
Maroon 5	Tropical	Red Hot Chili Peppers	No
The Beatles	Acústico	Red Hot Chili Peppers	No
Red Hot Chili Peppers	Jazz	Red Hot Chili Peppers	Sí
Oasis	Tropical	Maroon 5	No
The Killers	Balada	The Killers	Sí
Nirvana	Tropical	Nirvana	Sí
The Beatles	Tropical	The Beatles	Sí
U2	Acústico	The Beatles	No
Metallica	Electrónica	The Beatles	No
Carly Rae	Tropical	The Beatles	No
The Killers	Electrónica	The Killers	Sí
Red Hot Chili Peppers	Acústico	The Killers	No
Coldplay	Acústico	Nirvana	No
Carly Rae	Acústico	Maroon 5	No
The Beatles	Rock	The Beatles	Sí
Red Hot Chili Peppers	Acústico	U2	No
Carly Rae	Electrónica	U2	No
U2	Metal	The Killers	No
Oasis	Electrónica	Oasis	Sí
Coldplay	Acústico	Coldplay	Sí
Carly Rae	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
Oasis	Pop	Oasis	Sí
Carly	Pop	U2	No
The Killers	Pop	The Killers	Sí
Metallica	Rock	Nirvana	No
Maroon 5	Pop	Maroon 5	Sí
Coldplay	Balada	Coldplay	Sí
The Beatles	Pop	Carly Rae	No
Maroon 5	Rock	The Beatles	No
The Beatles	Rock	Red Hot Chili Peppers	No

Se hizo una segunda tanda de pruebas con los siguiente parámetros:

- 1 canción para modelar el GMM de cada artista
- 2000 vectores de coeficientes de DCT a procesar
- Umbral = 0.39
- 1 componentes en cada GMM
- Comparación de los 20 coeficientes del vector de cada bloque

Tabla 3: Resultados de la segunda corrida

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Coldplay	Rock	Maroon 5	No
U2	Rock	Coldplay	No
The Killers	Rock	Oasis	No
Coldplay	Pop	Coldplay	Sí
Nirvana	Rock	Red Hot Chili Peppers	No
U2	Pop	Oasis	No
Oasis	Acústico	Oasis	Sí
U2	Pop	Oasis	No
U2	Rock	The Killers	No
The Beatles	Pop	U2	No
Nirvana	Rock	Nirvana	Sí
The Beatles	Rap	Red Hot Chili Peppers	No
Coldplay	Bossa Nova	U2	No
Coldplay	Balada	U2	No
Coldplay	Pop	Coldplay	Sí
Red Hot Chili Peppers	Acústico	Carly Rae	No
Carly Rae	Bossa Nova	Carly Rae	Sí
Red Hot Chili Peppers	Electrónica	Carly Rae	No
Carly Rae	Rock	Carly Rae	Sí
Carly Rae	Tropical	Carly Rae	Sí
Red Hot Chili Peppers	Metal	Carly Rae	No
Carly Rae	Electrónica	Coldplay	No
Carly Rae	Bossa Nova	Coldplay	No
The Killers	Tropical	Coldplay	No
Carly Rae	Tropical	Coldplay	No
Oasis	Metal	Coldplay	No
The Beatles	Electrónica	Red Hot Chili Peppers	No
Carly Rae	Metal	Metallica	No
Nirvana	Balada	Coldplay	No
Red Hot Chili Peppers	Tropical	Maroon 5	No
Red Hot Chili Peppers	Balada	Maroon 5	No
U2	Tropical	Maroon 5	No
Red Hot Chili Peppers	Jazz	Maroon 5	No
Maroon 5	Tropical	Maroon 5	Sí
Carly Rae	Electrónica	Maroon 5	No
Red Hot Chili Peppers	Electrónica	Metallica	No

Continuación Tabla 3

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Red Hot Chili Peppers	Acústico	Metallica	No
The Beatles	Jazz	Metallica	No
Carly Rae	Electrónica	Nirvana	No
Maroon 5	Balada	Metallica	No
Red Hot Chili Peppers	Acústico	Oasis	No
Maroon 5	Tropical	Oasis	No
Carly Rae	Balada	U2	No
The Beatles	Tropical	Red Hot Chili Peppers	No
The Beatles	Acústico	Red Hot Chili Peppers	No
The Beatles	Jazz	Red Hot Chili Peppers	No
Carly	Tropical	Maroon 5	No
Maroon 5	Balada	The Killers	No
The Beatles	Tropical	Nirvana	No
The Beatles	Tropical	The Beatles	Sí
The Beatles	Acústico	The Beatles	Sí
The Beatles	Electrónica	The Beatles	Sí
Maroon 5	Tropical	The Beatles	No
The Beatles	Electrónica	The Killers	No
The Beatles	Acústico	The Killers	No
Maroon 5	Acústico	Nirvana	No
Carly Rae	Acústico	Maroon 5	No
Carly Rae	Rock	The Beatles	No
The Beatles	Acústico	U2	No
Carly Rae	Electrónica	U2	No
U2	Metal	The Killers	No
Carly Rae	Electrónica	Oasis	No
Coldplay	Acústico	Coldplay	Sí
Carly Rae	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
Oasis	Pop	Oasis	Sí
Carly	Pop	U2	No
The Killers	Pop	The Killers	Sí
Metallica	Rock	Nirvana	No
Maroon 5	Pop	Maroon 5	Sí
The Beatles	Pop	Carly Rae	No
Maroon 5	Rock	The Beatles	No
The Beatles	Rock	Red Hot Chili Peppers	No
Coldplay	Balada	Coldplay	Sí

Una tercera vuelta de pruebas se realizó con los siguiente parámetros:

- 1 canción para modelar el GMM de cada artista
- 2000 vectores de coeficientes de DCT a procesar
- Umbral = 0.39
- 4 componentes en cada GMM
- Comparación de los 20 coeficientes del vector de cada bloque

Tabla 4: Resultados de las pruebas en la tercera corrida

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
The Beatles	Rock	Maroon 5	No
Metallica	Rock	Coldplay	No
The Beatles	Rock	Oasis	No
Maroon 5	Pop	Coldplay	No
Nirvana	Rock	Red Hot Chili Peppers	No
U2	Pop	Oasis	No
Oasis	Acústico	Oasis	Sí
Maroon 5	Pop	Oasis	No
U2	Rock	The Killers	No
The Beatles	Pop	U2	No
Nirvana	Rock	Nirvana	Sí
Red Hot Chili Peppers	Rap	Red Hot Chili Peppers	Sí
Carly Rae	Bossa Nova	U2	No
Carly Rae	Balada	U2	No
Coldplay	Pop	Coldplay	Sí
Carly Rae	Acústico	Carly Rae	Sí
Carly Rae	Bossa Nova	Carly Rae	Sí
Carly Rae	Electrónica	Carly Rae	Sí
Maroon 5	Rock	Carly Rae	No
Carly Rae	Tropical	Carly Rae	Sí
Red Hot Chili Peppers	Metal	Carly Rae	No
Metallica	Electrónica	Coldplay	No
The Beatles	Bossa Nova	Coldplay	No
Maroon 5	Tropical	Coldplay	No
Carly Rae	Tropical	Coldplay	No
Maroon 5	Metal	Coldplay	No
The Beatles	Electrónica	Red Hot Chili Peppers	No
Metallica	Metal	Metallica	Sí
Nirvana	Balada	Coldplay	No
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Balada	Maroon 5	Sí
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Jazz	Maroon 5	Sí
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Electrónica	Maroon 5	Sí
Red Hot Chili Peppers	Electrónica	Metallica	No

Continuación Tabla 4

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Oasis	Acústico	Metallica	No
Red Hot Chili Peppers	Jazz	Metallica	No
Carly Rae	Electrónica	Nirvana	No
Maroon 5	Balada	Metallica	No
Red Hot Chili Peppers	Acústico	Oasis	No
Maroon 5	Tropical	Oasis	No
Maroon 5	Balada	U2	No
Maroon 5	Tropical	Red Hot Chili Peppers	No
The Killers	Acústico	Red Hot Chili Peppers	No
Oasis	Jazz	Red Hot Chili Peppers	No
Coldplay	Tropical	Maroon 5	No
Maroon 5	Balada	The Killers	No
The Beatles	Tropical	Nirvana	No
The Beatles	Tropical	The Beatles	Sí
Red Hot Chili Peppers	Acústico	The Beatles	No
The Beatles	Electrónica	The Beatles	Sí
Metallica	Tropical	The Beatles	No
The Beatles	Electrónica	The Killers	No
The Beatles	Acústico	The Killers	No
Metallica	Acústico	Nirvana	No
Metallica	Acústico	Maroon 5	No
Metallica	Rock	The Beatles	No
Nirvana	Acústico	U2	No
Coldplay	Electrónica	U2	No
U2	Metal	The Killers	No
The Beatles	Electrónica	Oasis	No
Coldplay	Acústico	Coldplay	Sí
Carly Rae	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
Oasis	Pop	Oasis	Sí
U2	Pop	U2	Sí
The Killers	Pop	The Killers	Sí
Metallica	Rock	Nirvana	No
Metallica	Pop	Maroon 5	No
Metallica	Pop	Carly Rae	No
The Beatles	Rock	The Beatles	Sí
The Beatles	Rock	Red Hot Chili Peppers	No
Metallica	Balada	Coldplay	No

La cuarta tanda de pruebas se realizó con los siguiente parámetros:

- 3 canciones para modelar el GMM de cada artista
- 2000 vectores de coeficientes de DCT a procesar
- Umbral = -1.6
- 2 componentes en cada GMM
- Comparación de solamente los 12 coeficientes más importantes del vector de cada bloque

Tabla 5: Resultados de las pruebas con la cuarta corrida

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
The Beatles	Rock	Maroon 5	No
Coldplay	Rock	Coldplay	Sí
The Killers	Rock	Oasis	No
Coldplay	Pop	Coldplay	Sí
Red Hot Chili Peppers	Rock	Red Hot Chili Peppers	Sí
Oasis	Pop	Oasis	Sí
The Killers	Acústico	Oasis	No
Oasis	Pop	Oasis	Sí
Oasis	Rock	The Killers	No
U2	Pop	U2	Sí
Red Hot Chili Peppers	Rock	Nirvana	No
Red Hot Chili Peppers	Rap	Red Hot Chili Peppers	Sí
Nirvana	Bossa Nova	U2	No
Carly Rae	Balada	U2	No
Oasis	Pop	Coldplay	No
Nirvana	Acústico	Carly Rae	No
Red Hot Chili Peppers	Bossa Nova	Carly Rae	No
Maroon	Electrónica	Carly Rae	No
Coldplay	Rock	Carly Rae	No
The Killers	Tropical	Carly Rae	No
Metallica	Metal	Carly Rae	No
Carly Rae	Electrónica	Coldplay	No
Red Hot Chili Peppers	Bossa Nova	Coldplay	No
Nirvana	Tropical	Coldplay	No
Metallica	Tropical	Coldplay	No
The Beatles	Metal	Coldplay	No
Red Hot Chili Peppers	Electrónica	Red Hot Chili Peppers	Sí
Carly Rae	Metal	Metallica	No
Nirvana	Balada	Coldplay	No
U2	Tropical	Maroon 5	No
Nirvana	Balada	Maroon 5	No
Oasis	Tropical	Maroon 5	No
Red Hot Chili Peppers	Jazz	Maroon 5	No
Nirvana	Tropical	Maroon 5	No
Red Hot Chili Peppers	Electrónica	Maroon 5	No

Continuación Tabla 5

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Metallica	Electrónica	Metallica	Sí
The Killers	Acústico	Metallica	No
U2	Jazz	Metallica	No
Metallica	Electrónica	Nirvana	No
Red Hot Chili Peppers	Balada	Metallica	No
The Killers	Acústico	Oasis	No
The Beatles	Tropical	Oasis	No
Red Hot Chili Peppers	Balada	U2	No
Red Hot Chili Peppers	Tropical	Red Hot Chili Peppers	Sí
Red Hot Chili Peppers	Acústico	Red Hot Chili Peppers	Sí
Red Hot Chili Peppers	Jazz	Red Hot Chili Peppers	Sí
Nirvana	Tropical	Maroon 5	No
Oasis	Balada	The Killers	No
Maroon 5	Tropical	Nirvana	No
Red Hot Chili Peppers	Tropical	The Beatles	No
Red Hot Chili Peppers	Acústico	The Beatles	No
The Beatles	Electrónica	The Beatles	Sí
Carly Rae	Tropical	The Beatles	No
Nirvana	Electrónica	The Killers	No
Red Hot Chili Peppers	Acústico	The Killers	No
Nirvana	Acústico	Nirvana	Sí
Coldplay	Acústico	Maroon 5	No
Oasis	Rock	The Beatles	No
Nirvana	Acústico	U2	No
U2	Electrónica	U2	Sí
Oasis	Metal	The Killers	No
The Beatles	Electrónica	Oasis	No
Coldplay	Acústico	Coldplay	Sí
Red Hot Chili Peppers	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
The Beatles	Pop	Oasis	No
The Beatles	Pop	U2	No
The Killers	Pop	The Killers	Sí
Red Hot Chili Peppers	Rock	Nirvana	No
Nirvana	Pop	Maroon 5	No
Metallica	Pop	Carly Rae	No
The Beatles	Rock	The Beatles	Sí
Red Hot Chili Peppers	Rock	Red Hot Chili Peppers	Sí
Coldplay	Balada	Coldplay	Sí

La quinta tanda de pruebas se realizó con la siguiente configuración:

- 3 canciones para modelar el GMM de cada artista
- 2000 vectores de coeficientes de DCT a procesar
- Umbral = -1.6
- 8 componentes en cada GMM
- Comparación de solamente los 12 coeficientes más importantes del vector de cada bloque

Tabla 6: Resultados de las pruebas con la quinta corrida

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
The Beatles	Rock	Maroon 5	No
Oasis	Rock	Coldplay	No
The Killers	Rock	Oasis	No
Coldplay	Pop	Coldplay	Sí
Red Hot Chili Peppers	Rock	Red Hot Chili Peppers	Sí
Oasis	Pop	Oasis	Sí
The Killers	Acústico	Oasis	No
Oasis	Pop	Oasis	Sí
Oasis	Rock	The Killers	No
U2	Pop	U2	Sí
The Killers	Rock	Nirvana	No
Red Hot Chili Peppers	Rap	Red Hot Chili Peppers	Sí
Nirvana	Bossa Nova	U2	No
Carly Rae	Balada	U2	No
Maroon 5	Pop	Coldplay	No
Nirvana	Acústico	Carly Rae	No
Red Hot Chili Peppers	Bossa Nova	Carly Rae	No
Maroon 5	Electrónica	Carly Rae	No
Coldplay	Rock	Carly Rae	No
The Killers	Tropical	Carly Rae	No
Metallica	Metal	Carly Rae	No
Carly Rae	Electrónica	Coldplay	No
Red Hot Chili Peppers	Bossa Nova	Coldplay	No
Nirvana	Tropical	Coldplay	No
Metallica	Tropical	Coldplay	No
The Beatles	Metal	Coldplay	No
Red Hot Chili Peppers	Electrónica	Red Hot Chili Peppers	Sí
Carly Rae	Metal	Metallica	No
Nirvana	Balada	Coldplay	No
U2	Tropical	Maroon 5	No
Nirvana	Balada	Maroon 5	No
Oasis	Tropical	Maroon 5	No
Red Hot Chili Peppers	Jazz	Maroon 5	No
Nirvana	Tropical	Maroon 5	No
Red Hot Chili Peppers	Electrónica	Maroon 5	No

Continuación Tabla 6

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Metallica	Electrónica	Metallica	Sí
The Killers	Acústico	Metallica	No
U2	Jazz	Metallica	No
Maroon 5	Electrónica	Nirvana	No
Nirvana	Balada	Metallica	No
Nirvana	Acústico	Oasis	No
Oasis	Tropical	Oasis	Sí
Oasis	Balada	U2	No
Metallica	Tropical	Red Hot Chili Peppers	No
Maroon 5	Acústico	Red Hot Chili Peppers	No
Maroon 5	Jazz	Red Hot Chili Peppers	No
Maroon 5	Tropical	Maroon 5	Sí
Metallica	Balada	The Killers	No
Maroon 5	Tropical	Nirvana	No
Maroon 5	Tropical	The Beatles	No
Maroon 5	Acústico	The Beatles	No
Metallica	Electrónica	The Beatles	No
Maroon 5	Tropical	The Beatles	No
Carly Rae	Electrónica	The Killers	No
Metallica	Acústico	The Killers	No
Nirvana	Acústico	Nirvana	Sí
Maroon 5	Acústico	Maroon 5	Sí
Oasis	Rock	The Beatles	No
Maroon 5	Acústico	U2	No
U2	Electrónica	U2	Sí
Oasis	Metal	The Killers	No
Maroon 5	Electrónica	Oasis	No
Coldplay	Acústico	Coldplay	Sí
Oasis	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
The Beatles	Pop	Oasis	No
Oasis	Pop	U2	No
The Killers	Pop	The Killers	Sí
Red Hot Chili Peppers	Rock	Nirvana	No
Nirvana	Pop	Maroon 5	No
Metallica	Pop	Carly Rae	No
The Beatles	Rock	The Beatles	Sí
Red Hot Chili Peppers	Rock	Red Hot Chili Peppers	Sí
Coldplay	Balada	Coldplay	Sí

La sexta tanda de pruebas se realizó con la siguiente configuración en los parámetros:

- 3 canciones para modelar el GMM de cada artista
- 2000 vectores de coeficientes de DCT a procesar
- Umbral = -1.6
- 16 componentes en cada GMM
- Comparación de solamente los 12 coeficientes más importantes del vector de cada bloque

Tabla 7: Resultados de las pruebas con la sexta corrida

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Maroon 5	Rock	Maroon 5	Sí
The Killers	Rock	Coldplay	No
Maroon 5	Rock	Oasis	No
Coldplay	Pop	Coldplay	Sí
Maroon 5	Rock	Red Hot Chili Peppers	No
Metallica	Pop	Oasis	No
Oasis	Acústico	Oasis	Sí
U2	Pop	Oasis	No
Coldplay	Rock	The Killers	No
Nirvana	Pop	U2	No
Maroon 5	Rock	Nirvana	No
Oasis	Rap	Red Hot Chili Peppers	No
Nirvana	Bossa Nova	U2	No
The Killers	Balada	U2	No
Maroon 5	Pop	Coldplay	No
U2	Acústico	Carly Rae	No
The Killers	Bossa Nova	Carly Rae	No
Maroon 5	Electrónica	Carly Rae	No
Metallica	Rock	Carly Rae	No
Maroon 5	Tropical	Carly Rae	No
Coldplay	Metal	Carly Rae	No
Carly Rae	Electrónica	Coldplay	No
Nirvana	Bossa Nova	Coldplay	No
Metallica	Tropical	Coldplay	No
Nirvana	Tropical	Coldplay	No
Metallica	Metal	Coldplay	No
Maroon 5	Electrónica	Red Hot Chili Peppers	No
Carly Rae	Metal	Metallica	No
U2	Balada	Coldplay	No
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Balada	Maroon 5	Sí
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Jazz	Maroon 5	Sí

Continuación Tabla 7

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Electrónica	Maroon 5	Sí
Nirvana	Electrónica	Metallica	No
Metallica	Acústico	Metallica	Sí
Oasis	Jazz	Metallica	No
Maroon 5	Electrónica	Nirvana	No
Nirvana	Balada	Metallica	No
Metallica	Acústico	Oasis	No
Oasis	Tropical	Oasis	Sí
Metallica	Balada	U2	No
Maroon 5	Tropical	Red Hot Chili Peppers	No
Maroon 5	Acústico	Red Hot Chili Peppers	No
Maroon 5	Jazz	Red Hot Chili Peppers	No
Maroon 5	Tropical	Maroon 5	Sí
Metallica	Balada	The Killers	No
Maroon 5	Tropical	Nirvana	No
Maroon 5	Tropical	The Beatles	No
Maroon 5	Acústico	The Beatles	No
Metallica	Electrónica	The Beatles	No
Maroon 5	Tropical	The Beatles	No
Maroon 5	Electrónica	The Killers	No
Metallica	Acústico	The Killers	No
Nirvana	Acústico	Nirvana	Sí
Maroon 5	Acústico	Maroon 5	Sí
Oasis	Rock	The Beatles	No
Maroon 5	Acústico	U2	No
Maroon 5	Electrónica	U2	No
Maroon 5	Metal	The Killers	No
Maroon 5	Electrónica	Oasis	No
Maroon 5	Acústico	Coldplay	No
Metallica	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
The Beatles	Pop	Oasis	No
The Beatles	Pop	U2	No
Oasis	Pop	The Killers	No
Red Hot Chili Peppers	Rock	Nirvana	No
Nirvana	Pop	Maroon 5	No
Metallica	Pop	Carly Rae	No
U2	Rock	The Beatles	No
Red Hot Chili Peppers	Rock	Red Hot Chili Peppers	Sí
Coldplay	Balada	Coldplay	Sí

Una séptima tanda de pruebas se realizó con la siguiente configuración en los parámetros:

- 3 canciones para modelar el GMM de cada artista
- 2000 vectores de coeficientes de DCT a procesar
- Umbral = -1.6
- 64 componentes en cada GMM
- Comparación de solamente los 12 coeficientes más importantes del vector de cada bloque

Tabla 8: Resultados de las pruebas con la séptima corrida

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Maroon 5	Rock	Maroon 5	Sí
Carly Rae	Rock	Coldplay	No
Maroon 5	Rock	Oasis	No
Coldplay	Pop	Coldplay	Sí
Maroon 5	Rock	Red Hot Chili Peppers	No
Maroon 5	Pop	Oasis	No
Oasis	Acústico	Oasis	Sí
Red Hot Chili Peppers	Pop	Oasis	No
Oasis	Rock	The Killers	No
Maroon 5	Pop	U2	No
Nirvana	Rock	Nirvana	Sí
Coldplay	Rap	Red Hot Chili Peppers	No
Maroon 5	Bossa Nova	U2	No
Maroon 5	Balada	U2	No
Coldplay	Pop	Coldplay	Sí
U2	Acústico	Carly Rae	No
Carly Rae	Bossa Nova	Carly Rae	Sí
Coldplay	Electrónica	Carly Rae	No
Carly Rae	Rock	Carly Rae	Sí
Maroon 5	Tropical	Carly Rae	No
Carly Rae	Metal	Carly Rae	Sí
Oasis	Electrónica	Coldplay	No
Coldplay	Bossa Nova	Coldplay	Sí
Maroon 5	Tropical	Coldplay	No
Maroon 5	Tropical	Coldplay	No
Maroon 5	Metal	Coldplay	No
Carly Rae	Electrónica	Red Hot Chili Peppers	No
The Killers	Metal	Metallica	No
Carly Rae	Balada	Coldplay	No
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Balada	Maroon 5	Sí
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Jazz	Maroon 5	Sí
Maroon 5	Tropical	Maroon 5	Sí

Continuación Tabla 8

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Maroon 5	Electrónica	Maroon 5	Sí
Metallica	Electrónica	Metallica	Sí
Carly Rae	Acústico	Metallica	No
Maroon 5	Jazz	Metallica	No
Maroon 5	Electrónica	Nirvana	No
Metallica	Balada	Metallica	Sí
Coldplay	Acústico	Oasis	No
Oasis	Tropical	Oasis	Sí
Oasis	Balada	U2	No
Maroon 5	Tropical	Red Hot Chili Peppers	No
Maroon 5	Acústico	Red Hot Chili Peppers	No
Maroon 5	Jazz	Red Hot Chili Peppers	No
Maroon 5	Tropical	Maroon 5	Sí
Metallica	Balada	The Killers	No
Maroon 5	Tropical	Nirvana	No
Maroon 5	Tropical	The Beatles	No
Maroon 5	Acústico	The Beatles	No
Metallica	Electrónica	The Beatles	No
Maroon 5	Tropical	The Beatles	No
Carly Rae	Electrónica	The Killers	No
Metallica	Acústico	The Killers	No
Maroon 5	Acústico	Nirvana	No
Maroon 5	Acústico	Maroon 5	Sí
U2	Rock	The Beatles	No
Maroon 5	Acústico	U2	No
U2	Electrónica	U2	Sí
Maroon 5	Metal	The Killers	No
Maroon 5	Electrónica	Oasis	No
Coldplay	Acústico	Coldplay	Sí
Maroon 5	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
Maroon 5	Pop	Oasis	No
The Beatles	Pop	U2	No
Maroon 5	Pop	The Killers	No
Maroon 5	Rock	Nirvana	No
Maroon 5	Pop	Maroon 5	Sí
Maroon 5	Pop	Carly Rae	No
U2	Rock	The Beatles	No
Red Hot Chili Peppers	Rock	Red Hot Chili Peppers	Sí
Coldplay	Balada	Coldplay	Sí

La octava ronda de pruebas se realizó con la siguiente configuración en los parámetros:

- 3 canciones para modelar el GMM de cada artista
- 2000 vectores de coeficientes de DCT a procesar
- Umbral = -1.6
- 64 componentes en cada GMM
- Todos los coeficientes del vector de cada bloque (20).

Tabla 9: Resultados de las pruebas con la octava corrida

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Maroon 5	Rock	Maroon 5	Sí
The Killers	Rock	Coldplay	No
Maroon 5	Rock	Oasis	No
Coldplay	Pop	Coldplay	Sí
Maroon 5	Rock	Red Hot Chili Peppers	No
Maroon 5	Pop	Oasis	No
Oasis	Acústico	Oasis	Sí
U2	Pop	Oasis	No
Coldplay	Rock	The Killers	No
Maroon 5	Pop	U2	No
Maroon 5	Rock	Nirvana	No
Oasis	Rap	Red Hot Chili Peppers	No
Maroon 5	Bossa Nova	U2	No
Maroon 5	Balada	U2	No
Maroon 5	Pop	Coldplay	No
U2	Acústico	Carly Rae	No
Maroon 5	Bossa Nova	Carly Rae	No
Maroon 5	Electrónica	Carly Rae	No
Maroon 5	Rock	Carly Rae	No
Maroon 5	Tropical	Carly Rae	No
Coldplay	Metal	Carly Rae	No
Carly Rae	Electrónica	Coldplay	No
Nirvana	Bossa Nova	Coldplay	No
Metallica	Tropical	Coldplay	No
Nirvana	Tropical	Coldplay	No
Metallica	Metal	Coldplay	No
Maroon 5	Electrónica	Red Hot Chili Peppers	No
Carly Rae	Metal	Metallica	No
U2	Balada	Coldplay	No
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Balada	Maroon 5	Sí
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Jazz	Maroon 5	Sí
Maroon 5	Tropical	Maroon 5	Sí
Maroon 5	Electrónica	Maroon 5	Sí

Continuación Tabla 9

Artista según el programa	**Género**	**Artista original**	**¿Acierto?**
Maroon 5	Electrónica	Metallica	No
Maroon 5	Acústico	Metallica	No
Maroon 5	Jazz	Metallica	No
Maroon 5	Electrónica	Nirvana	No
Metallica	Balada	Metallica	Sí
Maroon 5	Acústico	Oasis	No
Maroon 5	Tropical	Oasis	No
Metallica	Balada	U2	No
Maroon 5	Tropical	Red Hot Chili Peppers	No
Maroon 5	Acústico	Red Hot Chili Peppers	No
Maroon 5	Jazz	Red Hot Chili Peppers	No
Maroon 5	Tropical	Maroon 5	Sí
Metallica	Balada	The Killers	No
Maroon 5	Tropical	Nirvana	No
Maroon 5	Tropical	The Beatles	No
Maroon 5	Acústico	The Beatles	No
Maroon 5	Electrónica	The Beatles	No
Maroon 5	Tropical	The Beatles	No
Maroon 5	Electrónica	The Killers	No
Maroon 5	Acústico	The Killers	No
Maroon 5	Acústico	Nirvana	No
Maroon 5	Acústico	Maroon 5	Sí
Oasis	Rock	The Beatles	No
Maroon 5	Acústico	U2	No
Maroon 5	Electrónica	U2	No
Maroon 5	Metal	The Killers	No
Maroon 5	Electrónica	Oasis	No
Maroon 5	Acústico	Coldplay	No
Metallica	Balada	The Beatles	No
Metallica	Metal	Metallica	Sí
The Beatles	Pop	Oasis	No
The Beatles	Pop	U2	No
Maroon 5	Pop	The Killers	No
Maroon 5	Rock	Nirvana	No
Maroon 5	Pop	Maroon 5	Sí
Maroon 5	Pop	Carly Rae	No
Maroon 5	Rock	The Beatles	No
Red Hot Chili Peppers	Rock	Red Hot Chili Peppers	Sí
Coldplay	Balada	Coldplay	Sí

Se hicieron 74 pruebas con canciones no originales sobre el algoritmo. Acerca de los géneros musicales representados se tienen 12 canciones de género tropical, 11 de rock, 10 de pop, 11 de acústico, 10 electrónica, 8 balada, 3 de bossa nova y jazz, 5 de metal y 1 de rap. Al ser tan pocas las canciones que se encontraron de bossa nova, jazz y rap, no se puede realmente asegurar la funcionalidad del algoritmo sobre estos estilos musicales. Ahora bien, respecto al primer experimento, se obtuvo un nivel de efectividad de 36.49% (27 canciones acertadas) repartido de la siguiente forma según género:

Tabla 10: Porcentajes de efectividad según género en el primer conjunto de corridas

Género	**Porcentaje de efectividad**
Tropical	33.33%
Rock	36.36%
Pop	50.00%
Acústico	9.09%
Electrónica	60.00%
Balada	25.00%
Bossa Nova	33.33%
Jazz	33.33%
Metal	20.00%
Rap	0.00%
Total	*33.78%*

Es recomendable primero decir que el porcentaje de efectividad sobre todas las pistas es relativamente bajo. Es verdad que aún no se tiene un algoritmo que trabaje al 100%, pero trabajos anteriores por otros doctores o ingenieros han alcanzado hasta un 80%. En este primer experimento, se tiene la limitante que a pesar que se tienen 12800 coeficientes para modelar a cada artista, todos estos coeficientes vienen de la misma canción, lo que hace que los resultados no encuentren un óptimo, debido a que en otra canción puede que canten en otra tonalidad o la música de fondo sea tan diferente que se identifique a otro cantante. Acerca del número de componentes para el GMM, se estableció un procedimiento de prueba y error, esto se debe a que en la mayoría de bibliografía consultada, no existe un número fijo que maximice el número de aciertos, por el contrario, este parámetro siempre lo varían para encontrar el valor después de las pruebas.

Se usaron los primeros 2000 coeficientes para comparar, esto se debe a que se está tomando el primer minuto de música. Comúnmente en este minuto ya se encuentran partes con voz y sin la misma. Asimismo, el tiempo para procesar e identificar la canción se reduce a cerca de 4 minutos. Igualmente se ha encontrado que el analizar un minuto de la canción o más no causa que el algoritmo deje de funcionar e incluso ya se han hecho pruebas así.

El “threshold” o umbral que se usó salió del valor de comparaciones hechas de antemano al algoritmo. En ellas, se ingresó el archivo sabiendo si contenía voz o no. Se tomó el valor de puntaje respectivo, y se hizo un promedio para tener la idea de un

valor que fuera a discriminar entre el valor del GMM instrumental y el vocal. El promedio resulta de una prueba entre la resta de estos valores.

Como es posible ver en este primer experimento con el algoritmo, las diferencias entre porcentaje de efectividad para cada género no son mínimas, sino que estas diferencias son muy pronunciadas. El género con peor porcentaje es el acústico, que hace poco de sentido sabiendo que este género es conocido por ser un intérprete acompañado de instrumentos que producen su sonido usando la acústica y no medios eléctricos. Este género alcanzó un 9.09% de efectividad al identificar al cantante, que es un porcentaje muy bajo, especialmente si se compara al resto.

El género con mayor porcentaje fue el de electrónica. No resulta una sorpresa muy grande que éste sea el mejor identificado, ya que la música electrónica no cambia drásticamente de pista en pista: las percusiones son muy parecidas y el tempo tampoco es muy diferente. Además, en la música electrónica tiende a agregar sintetizadores con notas muy melódicas a la voz, lo que refuerza la idea de encontrar al intérprete basado en coeficientes parecidos.

A continuación aparecen varios géneros que tienen una efectividad que oscila entre 30 y 40%, estos son tropical, rock y pop. Sorprende que las baladas hayan ocupado una posición mucho mejor en el desempeño que otros géneros ya que como las canciones acústicas, logran resaltar la voz sin usar tantos aparatos electrónicos al momento de acompañar. De los otros géneros como bossa nova, jazz y metal, se hablará solamente acerca de la comparación entre experimentos ya que al tener solo 3 temas musicales, el porcentaje podría llegar a ser engañoso.

Si se eliminan todas las corridas del género acústico, la efectividad del reconocimiento total subiría hasta 41%, que continúa siendo uno no tan alto pero sí 5% mayor al que se obtuvo en esta ocasión.

Se cree que la principal razón por la que el resultado es relativamente bajo es la falta de coeficientes en la base de datos, que se podría agrandar con el inconveniente de que esto haría las pruebas más lentas y el código fuente del programa más extenso.

Para la segunda corrida de experimentos, solamente se hizo un cambio: en lugar que cada GMM tuviera 2 medias, se decidió que solo tuvieran 1. Esto se hizo pensando en que tal vez por el escaso número de coeficientes que formaban el GMM, esta sería una mejor representación. Los resultados arrojaron un peor porcentaje de aciertos: 22.97%. En algunos estudios anteriores, el agregar más componentes al GMM, lo hacía mejor, pero teniendo en cuenta de que sus bases de datos eran mucho más grandes que la que se presenta en este trabajo. Los resultados son los siguientes:

Tabla 11: Porcentajes de efectividad según género en la segunda corrida

Género	**Porcentaje de efectividad**
Tropical	25.00%
Rock	18.18%

Continuación Tabla 11

Género	**Porcentaje de efectividad**
Pop	50.00%
Acústico	27.27%
Electrónica	0.00%
Balada	12.50%
Bossa Nova	66.67%
Jazz	0.00%
Metal	20.00%
Rap	0.00%
Total	*22.97%*

Como se puede ver, el achicar el número de componentes en los GMM hizo que el porcentaje de aciertos decreciera por un valor considerable. Lo que más llama la atención de este caso, es que prácticamente todos los géneros decrecieron, pero esta vez, el mayor del primer experimento, el género de la electrónica, cayó de un 60% a un 0%. El género acústico que había causado cierta interrogante durante las primeras pruebas debido al bajo rendimiento, esta vez mejora pero sin ser algo sobresaliente. El mejor género esta vez termina siendo el pop, que al arrojar un valor respetable de 50% de aciertos, tuvo el mismo porcentaje de aciertos si se compara con la primera corrida. Los géneros de metal, jazz, bossa nova y rap no presentan cambios respecto al anterior experimento.

Para la tercera corrida se obtuvieron los siguientes resultados:

Tabla 12: Porcentajes de efectividad según género en la tercera corrida

Género	**Porcentaje de efectividad**
Tropical	41.67%
Rock	18.18%
Pop	50.00%
Acústico	27.27%
Electrónica	30.00%
Balada	12.50%
Bossa Nova	33.33%
Jazz	33.33%
Metal	40.00%
Rap	0.00%
Total	*31.08%*

Como se puede ver, si se compara la Tabla 12 con la Tabla 11, la mayoría de los géneros tienen un mejor desempeño, aunque es para resaltar lo que sucede con el género pop, que sigue estableciéndose como el mejor con un 50% de aciertos. Es llamativo el hecho que el género de electrónica, que en la primera corrida alcanzó un 70% de efectividad, esta vez volvió a quedarse corto con solamente 30%. También si

se compara con la primera corrida, el género rock cae de 36.36% a la mitad, que fue el mismo resultado que se obtuvo en la segunda corrida.

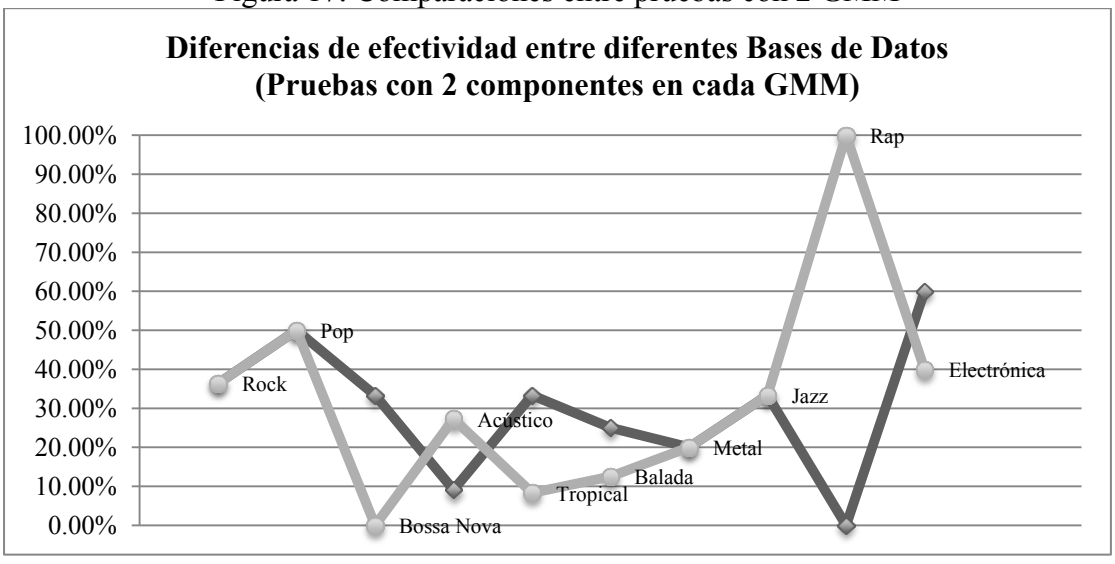
Para la cuarta tanda de pruebas se obtuvieron los siguientes resultados:

Tabla 13: Porcentajes de efectividad según género en la cuarta corrida

Género	**Porcentaje de efectividad**
Tropical	8.33%
Rock	36.36%
Pop	50.00%
Acústico	27.27%
Electrónica	40.00%
Balada	12.50%
Bossa Nova	0.00%
Jazz	33.33%
Metal	20.00%
Rap	100.00%
Total	*28.38%*

La Tabla 13 marca el comienzo de las tandas de corridas que se hicieron con lo que se llamó una base de datos "completa". Para las comparaciones a partir de esta tabla, los GMM, tanto universales como particulares, fueron generados a partir de 3 canciones que fueron procesadas previamente como se mencionó en la sección de "Diseño Experimental". Es ahora conveniente comparar la Tabla 13 con la Tabla 10 debido a que en ambas se usaron 2 GMM para modelar cada voz del artista. El umbral para esta corrida ha variado debido a que ha sido modificada la base de datos y además éste se encuentra experimentalmente. Para llegar a los resultados en Tabla 13 se usaron solamente 12 coeficientes por artista, mientras que en la Tabla 10 se usaron los 20 disponibles. A continuación, los resultados de esta comparación:

Figura 17: Comparaciones entre pruebas con 2 GMM



En esta gráfica se puede apreciar cómo la prueba con la base de datos de 1 canción (línea oscura) termina siendo superior en 4 géneros a la prueba con base de datos de 3 canciones (línea clara). A pesar de esto, se cree que es mucho mejor tener más datos para modelar al artista ya que se tienen modeladas más características de sus extractos vocales. El resultado en este caso, se cree que es solamente un caso particular y no una situación general. En total, la primera prueba tuvo 4 aciertos más que la segunda.

En cuanto los resultados propios de esta ronda de pruebas, se puede ver nuevamente que el género pop vuelve a ser el mejor, nuevamente con 50% de efectividad (el género rap tiene 100% porque fue un acierto de un intento y por lo tanto no se le da relevancia en este caso). Electrónica vuelve a repuntar con 40% seguido en porcentaje de aciertos por rock, jazz y acústico. Estos valores no son tan fáciles de comparar con los demás experimentos a excepción del primero ya que la base de datos cambió significativamente; pero servirá de base para las comparaciones con las tablas que se presentan desde este punto.

Durante la quinta ronda de pruebas se obtuvieron los siguientes resultados:

Tabla 14: Porcentajes de efectividad según género en la quinta corrida

Género	**Porcentaje de efectividad**
Tropical	16.67%
Rock	27.27%
Pop	50.00%
Acústico	27.27%
Electrónica	30.00%
Balada	12.50%
Bossa Nova	0.00%
Jazz	33.33%
Metal	20.00%
Rap	100.00%
Total	*25.68%*

La diferencia entre la cuarta y quinta ronda de pruebas fue únicamente la cantidad de componentes que formaban cada GMM, se usaron 2 y 8 para cada una, respectivamente. En la literatura encontrada relacionada con este problema, un aumento de componentes en GMM, generalmente agranda el porcentaje de aciertos del algoritmo.

Es necesario mencionar que si comparamos estas dos tandas de pruebas, resultan ser las más parecidas, no al ver el porcentaje, sino al ver las Tablas 5 y 6, en las que prácticamente no se encuentran cambios significativos. Por otro lado, el porcentaje de efectividad, decae con 8 componentes.

Resultados de la sexta tanda de pruebas:

Tabla 15: Porcentajes de efectividad según género en la sexta corrida

Género	**Porcentaje de efectividad**
Tropical	41.67%
Rock	18.18%
Pop	10.00%
Acústico	36.36%
Electrónica	10.00%
Balada	25.00%
Bossa Nova	0.00%
Jazz	33.33%
Metal	20.00%
Rap	0.00%
Total	*22.97%*

Al doblar la cantidad de componentes de todos los GMM, como se hizo en este caso, llegando a 16, la efectividad vuelve a ser menor por un valor considerablemente bajo. Llama la atención cómo el género pop que se había mantenido en 50% en todas las pruebas, decrece a 10%; mientras que el género tropical tiene un repunte de más de 25% si comparamos las dos últimas pruebas realizadas. Con esto, puede verse claro que el número de componentes GMM para los distintos géneros, no pareciera tener una tendencia, sino que los valores tienden a ir de más a menos y luego al contrario.

Resultados de la séptima ronda de pruebas:

Tabla 16: Porcentajes de efectividad según género en la séptima corrida

Género	**Porcentaje de efectividad**
Tropical	41.67%
Rock	36.36%
Pop	30.00%
Acústico	27.27%
Electrónica	30.00%
Balada	37.50%
Bossa Nova	66.67%
Jazz	33.33%
Metal	40.00%
Rap	0.00%
Total	*35.14%*

La séptima ronda se realizó con 64 componentes en todos los GMM y resultó la mejor en cuanto a porcentaje de efectividad, alcanzó un 35.14%. Si se compara con la corrida anterior, en la que modelaron los artistas con 16 componentes, todos los géneros tuvieron un aumento en efectividad excepto el género acústico. Sorprende que el género pop que había sido el que mejor se había estado desempeñando, vuelva a tener un resultado relativamente bajo con solamente 30% cuando ya había alcanzado el 50%.

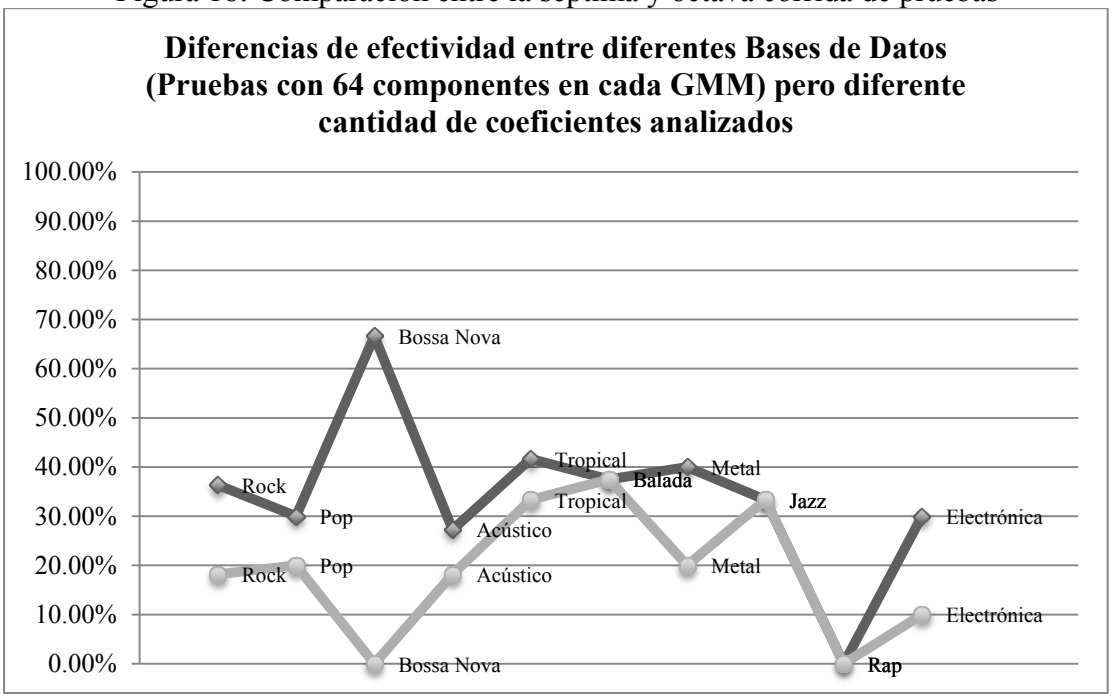
Resultados de la octava ronda de pruebas:

Tabla 17: Porcentajes de efectividad según género en la octava corrida

Género	**Porcentaje de efectividad**
Tropical	33.33%
Rock	18.18%
Pop	20.00%
Acústico	18.18%
Electrónica	10.00%
Balada	37.50%
Bossa Nova	0.00%
Jazz	33.33%
Metal	20.00%
Rap	0.00%
Total	*21.62%*

Al haber encontrado la configuración con mayor porcentaje de efectividad, como sucedió con la Tabla 15, se procedió a modificar la cantidad de coeficientes analizados, para saber si se podía llegar a tener una mejor configuración en el sistema. Al analizar los 20 coeficientes, se dio una baja en el porcentaje de efectividad bastante pronunciada y ningún género mejoró en su rendimiento; con esto se demuestra una vez la sensibilidad del algoritmo a pequeños cambios. La comparación entre Tabla 16 y 17 se ve a continuación:

Figura 18: Comparación entre la séptima y octava corrida de pruebas



Por otra parte, el promedio de las corridas se muestra en la siguiente tabla, para un análisis general:

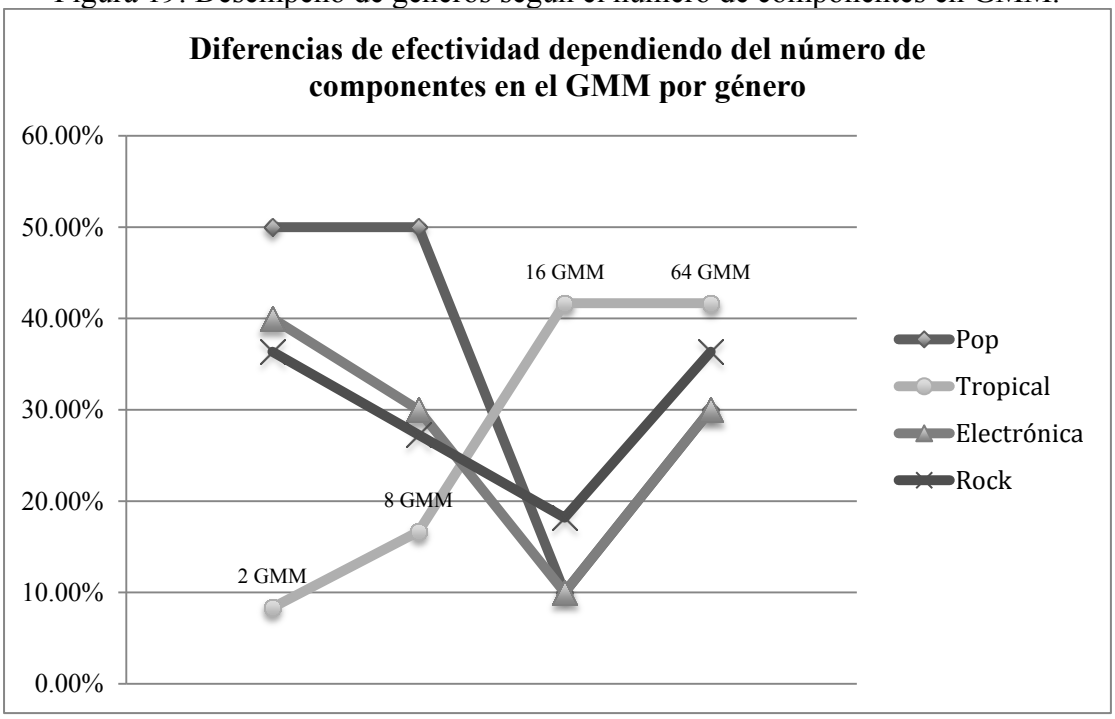
Tabla 18: Porcentajes de efectividad durante todos los experimentos

Género	**Porcentaje de efectividad**
Tropical	30.21%
Rock	26.14%
Pop	38.75%
Acústico	25.00%
Electrónica	36.67%
Balada	26.25%
Bossa Nova	21.88%
Jazz	25.00%
Metal	25.00%
Rap	25.00%
Total	*27.70%*

A pesar del pésimo resultado que se obtuvo cuando solo se usó una media, el género de pop fue el que mejor se desempeñó, junto con la electrónica, que tuvo un desempeño más irregular sobre los 8 experimentos. Bossa nova fue el peor. En tercer lugar aparece el tropical, que ni siquiera alcanzó un 50% de efectividad en cualquier corrida (ahora sí se pueden analizar todos los géneros porque ya hay suficientes pruebas para saber acerca de su efectividad). Aunque el por qué de que el género acústico haya sido uno de los que menos aciertos haya tenido sigue siendo una incógnita, o probablemente el género no tenga nada que ver con el desempeño del algoritmo, ya que se esperaba que el acústico fuera uno de los géneros más fáciles para reconocer con el algoritmo por sus características; o al menos, uno de los mejores y no el segundo peor.

Seguidamente aparece un gráfico en el que se puede ver el desempeño de 4 géneros según el número de componentes en el GMM de cada artista. Como es posible ver, no existe relación clara entre el desempeño de éstos y el número que lo compone. Solamente se puede ver que los extremos parecieran ser las mejores opciones según los porcentajes de efectividad. Solamente el género tropical muestra una tendencia clara de crecimiento, por ello, es difícil creer que existe relación alguna entre el desempeño de cierto género y los componentes del GMM.

Figura 19: Desempeño de géneros según el número de componentes en GMM.



A continuación se muestra una tabla con la desviación estándar de cada uno de los géneros. La variable que se analizó fue claramente el porcentaje de efectividad. Se hizo para saber si algún género era más volátil que otro al cambio de los parámetros del algoritmo. Para el análisis mostrado en la Tabla 19 se tomaron en cuenta todas las corridas.

Tabla 19: Desviación estándar de la efectividad de cada género durante todos los experimentos

Género	**Desviación estándar**
Tropical	12.55%
Rock	9.01%
Pop	16.42%
Acústico	8.06%
Electrónica	19.23%
Balada	11.08%
Bossa Nova	29.55%
Jazz	15.43%
Metal	9.26%
Rap	46.29%
Promedio	*17.69%*

La desviación estándar de un grupo de valores representa qué tan dispersos están estos mismos de la media. Un valor bajo en la Tabla 18 para cada género significaría que su valor se mantiene relativamente estable sin importar los cambios que se le hagan al algoritmo en las pruebas. El género acústico resultó con 8.06% de desviación estándar, el más estable; a pesar de esto, resultó de igual forma uno de los más bajos en porcentaje de aciertos, lo que hace pensar que ni siquiera tuvo una buena

tanda en las pruebas. El promedio se situó en 17.69% que es un valor alto y con el cual queda claro que el algoritmo resultó sumamente sensible a los cambios, que no es algo que sorprenda debido a que incluso, se cambió la base de datos una vez. Los géneros con mejor desempeño, pop y electrónica, tuvieron una desviación cercana al promedio.

Finalmente, se muestra una tabla donde se muestran los porcentajes de efectividad pero con la variante que acá se muestra el porcentaje según el artista, esto para ver si algún artista logró ser mejor identificado en estas corridas o hay alguno que haya hecho que el porcentaje total haya decrecido.

Tabla 20: Porcentajes de efectividad durante los cinco primeros experimentos según el artista evaluado

Artista	**Corrida 1**	**Corrida 2**	**Corrida 3**	**Corrida 4**	**Corrida 5**
Maroon 5	40.00%	20.00%	50.00%	0.00%	20.00%
U2	0.00%	0.00%	14.29%	28.57%	28.57%
Coldplay	45.45%	18.18%	18.18%	36.36%	27.27%
Oasis	25.00%	12.50%	25.00%	25.00%	37.50%
Red Hot Chili Peppers	14.29%	0.00%	14.29%	14.29%	57.14%
The Killers	50.00%	16.67%	16.67%	16.67%	16.67%
Nirvana	20.00%	0.00%	20.00%	20.00%	20.00%
Carly Rae	57.14%	42.86%	57.14%	0.00%	0.00%
Metallica	33.33%	16.67%	33.33%	33.33%	33.33%
The Beatles	28.57%	42.86%	42.86%	28.57%	14.29%

Tabla 21: Porcentajes de efectividad durante los últimos experimentos según el artista evaluado y su promedio

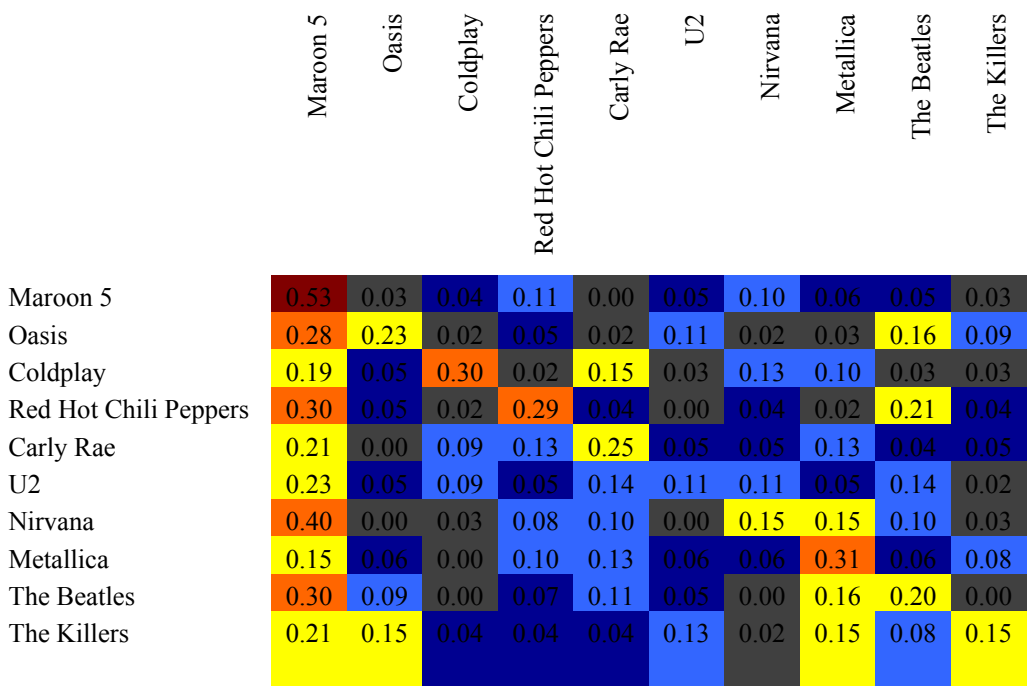
Artista	**Corrida 6**	**Corrida 7**	**Corrida 8**	**Promedio**
Maroon 5	90.00%	100.00%	100.00%	52.50%
U2	0.00%	14.29%	0.00%	10.71%
Coldplay	18.18%	36.36%	18.18%	29.55%
Oasis	25.00%	25.00%	12.50%	23.44%
Red Hot Chili Peppers	14.29%	14.29%	14.29%	28.57%
The Killers	0.00%	0.00%	0.00%	14.58%
Nirvana	20.00%	20.00%	0.00%	15.00%
Carly Rae	0.00%	42.86%	0.00%	25.00%
Metallica	33.33%	33.33%	33.33%	31.25%
The Beatles	0.00%	0.00%	0.00%	19.64%

Según la Tabla 20, el artista que se le complica menos al algoritmo al momento de identificar es Maroon 5. Este no es un resultado esperado, ya que se esperaría que Carly Rae siendo la única artista de género femenino en la base de datos, tuviera un registro muy diferente al del resto al momento de cantar y donde todos el resto son hombres. El segundo lugar lo ocupan Metallica y luego Coldplay. Estos dos artistas que han obtenido los dos primeros lugares son los únicos que obtuvieron arriba de un 30% de efectividad. Por otro lado, los que peor desempeño obtuvieron

son U2 y The Killers, los cuales no llegaron ni a un 15% de efectividad durante las 8 rondas de experimentos; incluso U2 no llegó ni a un 30% en ninguna de las corridas individuales.

Finalmente, se muestra la matriz de confusión, que se ha analizado con todas las pruebas realizadas. Esta matriz puede servir de guía para futuras modificaciones de la base de datos, ya que es aquí donde se puede ver que tan desviados están los resultados por artista y a quién apuntan los errores.

Figura 20: Matriz de confusión dados los resultados obtenidos con la experimentación del algoritmo



VIII. CONCLUSIONES

1. El algoritmo implementado funciona mejor de manera particular para el género electrónico con dos medias, donde alcanzó un 60% de efectividad según la Tabla 10 del trabajo escrito.

2. Para estos experimentos en concreto, el género pop y el electrónico son los que mejor se reconocen. Ahora bien, sigue sin darse una respuesta sobre un nivel general, debido a la variabilidad de la efectividad para canciones del mismo género que se observó a lo largo de los experimentos.

3. El número de componentes para los GMM que modelan la voz del artista modifica el porcentaje de aciertos del programa y no existe una relación entre estos dos valores.

4. El bossa nova es el género que más se le complica al algoritmo en estas pruebas para identificar al artista con solamente un 21.21% de efectividad sobre las ocho corridas realizadas como experimentos.

5. Para una mejor implementación del algoritmo, se deben cambiar varios parámetros y dejar otros fijos para llegar a un óptimo, esto surge de la idea de las diferencias entre las tablas del trabajo de escrito en el que se muestran muy diferentes resultados en cada tanda de pruebas.

6. El algoritmo MFCC+GMM todavía no es un proceso terminado y sus resultados no son 100% confiables, en este trabajo se logró un 27.70% de efectividad sobre todas las pruebas.

IX. RECOMENDACIONES

Se recomienda a futuros estudiantes y/o profesionales que deseen indagar en este tema lo siguiente:

- Continuar con pruebas estandarizadas para seguir aumentando el porcentaje de efectividad del algoritmo. (Variación de parámetros).
- Optimizar la base de datos y procesos del algoritmo para que las pruebas sean más rápidas y no existan posibles errores al correr el programa.
- Hacer modificaciones en el algoritmo en sí, con mezclas de otros algoritmos ya existentes y hacer las respectivas comparaciones.
- Agregar más GMM para seguir los estudios y encontrar mejoras en el porcentaje de aciertos del programa.

X. BIBLIOGRAFÍA

A. Referencias bibliográficas:

1. Camarena Ibarrola, J. *El Algoritmo E-M.* 34 páginas.
2. Kim,Y. and Whitman, B. *Singer identification in popular music recordings using voice coding features.* 6 págs.
3. Kulis. B. y Jordan. M. 2012. *Revisiting k-means: New Algorithms via Bayesian Nonparametrics.* 8 págs.
4. Marven, C. y Ewers, G., 1996. *A Simple Approach to Digital Signal Processing*. New York, NY. John Wiley & Sons, Inc. 236 págs.
5. Oppenheim, A. y Schafer, R., 1975. *Digital Signal Processing*. Englewood Cliffs, New Jersey. Prentice-Hall, Inc. 585 págs.
6. Reynolds, Douglas. «Gaussian Mixture Models». MIT. 5 págs.
7. Shenoy. A., Wu, Y. y Wang, Y. *Singing voice detection for karaoke application.* 11 págs.
8. Tsai, W. y Wang, H., *Automatic Detection and tracking of target singer in multi-singer music recordings.* 35 págs.
9. Tsai, W., Liao, S. y Lai, C. *Automatic Identification of Simultaneous Singers in Duet Recordings.* 6 págs.
10. Universidad Nacional Mayor de San Marcos. *Procesamiento Digital de Senales, Reconocimiento de Voz, Redes Neuronales Artificiales.* Lima. 204 págs.
11. University of Bridgeport. *Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal.* Connecticut. Electrical Engineering Department. 7 págs.
12. Zhang, T. *System and Method for Automatic Singer Identification*. Hewlett-Packard. 15 págs.

B. Referencias en Internet:

13. Abyssmedia. 2012. *What is WAV audio file format?* http://www.abyssmedia.com/formats/wav-format.shtml [27 de enero de 2,012]
14. Center for Computer Research in Music and Acoustics. *WAVE PCM SoundFile Format.* Disponible en: http://soundfile.sapp.org/doc/WaveFormat
15. Engineering Productivity Tools. *Definition of DFT and Inverse DFT (IDFT).* Disponible en: http://www.engineeringproductivitytools.com/stuff/T0001/PT01.HTM
16. Hydrogen Audio. *Pre-Emphasis.* Disponible en: http://wiki.hydrogenaudio.org/index.php?title=Pre-emphasis
17. Multimedia Information Retrieval LAB. *MFCC.* Disponble en: http://neural.cs.nthu.edu.tw/jang/books/audiosignalprocessing/speechfeaturemfcc.asp?title=12-2%20mfcc
18. National Semiconductor. 2010. *Application Note 779 A Basic Introduction to Filters - Active, Passive, and Switched Capacitor.* http://www.ti.com/lit/an/snoa224a/snoa224a.pdf [26 de enero de 2,012]
19. Practical Cryptography. *Mel Frequency Cepstral Coefficient.* Disponible en: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

20. The University of Texas, Arlington. *Discrete Cosine Transforms*. Disponible en: http://www-ee.uta.edu/dip/Courses/EE5355/Discrete%20class%201.pdf
21. Wolfram Mathworld. *Discrete Fourier Transform*. Disponible en: http://mathworld.wolfram.com/DiscreteFourierTransform.html

XI. ANEXOS

Complemento a Tabla 1. En esta tabla se muestran características de los algoritmos encontrados y que fueron puestos en comparación para una posible implementación.

Tabla 22: Tabla con las características de todos los algoritmos encontrados y sobre los cuales se hizo la comparación para saber cuál implementar.

ALGORITMO	CARACTERÍSTICAS
Chu/Gu	GMM, Modulación de energía, Coeficientes armónicos, MFCC, Comparaciones de Energía.
Berenzweig	Canales neurales multicapa, 13 coeficientes PLP, deltas and doble deltas.
Kim/Whitman	Filtro pasa banda (200-2000Hz), Filtro comb inverso, detector ármonico.
Zhang	Modelos estadísticos para representar todos los artistas
Tsai	GMM, MFCC. Ambos para regiones con y sin voz.
Tsai/Wang	Clasificador estocástico que consiste de un procesador front-end para extraer vectores característicos en cepstrums. GMM para regiones con y sin voz.
Illinois	FFT, Hamming, MFCC, K-Means.
Maddage	Máquinas de soporte vectorial, GMM, coeficientes cepstrales.
Fujihara/Goto	Estimación de la frecuencia fundamental, extracción de la estructura armónica, re-sintetización de audio, LPC Coeficientes cepstrales.
Shen	Máquinas de soporte vectorial, características timbrales de la voz, características vocales, estudio en instrumentos usados.
Chien/Wang	Filtros adaptivos, máquinas de soporte vectorial, descomposición sinusoidal, GMM
Khine/New	Ventana de Hamming, Coeficientes cepstrales frecuenciales para cada frecuencia, filtros sobre la escala Mel
Bartsch	Estimación de la frecuencia fundamental de la canción, espectrogramas.

La siguiente tabla se usó para encontrar un posible algoritmo que diferenciara las secciones en donde aparece la voz en una canción y dónde no está presente. Esto debido a que algunos algoritmos que aparecen en la Tabla 1, no cuentan con esta habilidad y se proponía utilizar uno de estos como complemento si el algoritmo que fuera el escogido no tuviera cómo reconocer las secciones en que sí aparecía la voz.

Tabla 23: Tabla con algoritmos encontrados que pueden diferenciar entre secciones en la canción con y sin voz.

Algoritmo	**Características**	**Dificultad de Implementarlo***	**Documentación Disponible****
Tzatenakis	Clasificador estadístico que usa media, varianzas y correlaciones en las sub-bandas de frecuencia.	Media	Baja
Maddage	Máquinas de soporte vectorial con reglas heurísticas.	Alta	Baja
Nwe/Wang	Modelos Escondidos de Markov, coeficientes frecuenciales logarítmicos, toma en cuenta tempo y distribuciones en las sub-bandas de frecuencia.	Alta	Baja
Zhang	Tasa de cruces por cero y coeficientes armónicos.	Baja	Media

* La Dificultad de Implementarlo se mide en Alta, Media y Baja; según el conocimiento adquirido hasta antes de la implementación. Alta hace referencia a un algoritmo que no va a poder ser implementado por falta de entendimiento de cómo funciona éste. Media se usa para describir un algoritmo que un poco más de estudio se podría llegar a implementar. Finalmente, Baja se refiere a un algoritmo que se puede proceder a implementar y que ya se tiene el conocimiento para lograr hacerlo.

** La Documentación Disponible describe qué tantos documentos se han encontrado para lograr entender el algoritmo a fondo. Si solo se encontró 1 documento, aparece la palabra Baja en esta columna; por otro lado, si el número de documentos es mayor a 1, aparece Media.

Manual de Usuario del Software Programado:

Requerimientos del Sistema:

- Sistemas Operativos: Microsoft XP o posterior. MacOS 10.3 o posterior.
- Procesador: Intel Core 2, 1.06GHz o más rápido.
- Memoria RAM: 4GB o más.
- Disco duro: 100GB de espacio disponibles.
- Python 2.7 instalado. Con SciPy, NumPy, Scikit Learn instalados para la versión correspondiente.

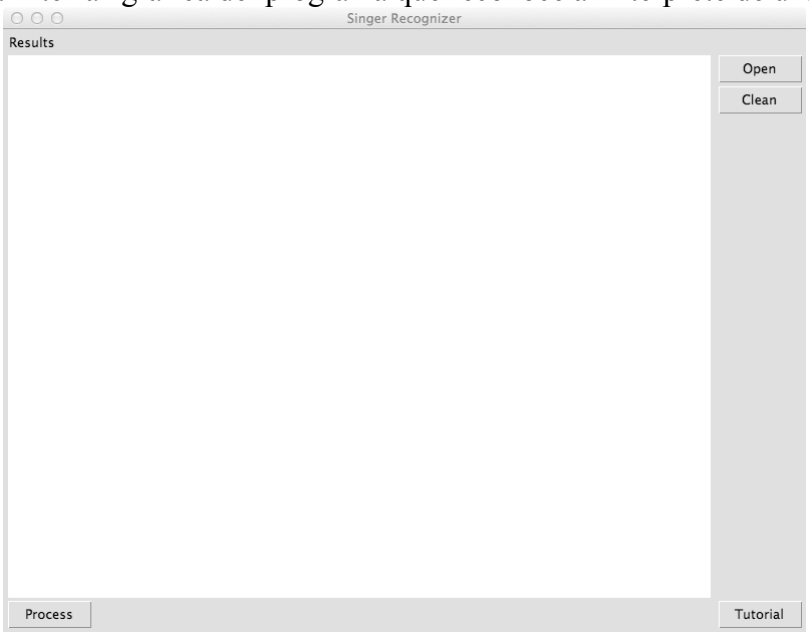
Para correr el programa en Windows:

Se guarda el archivo .py en una carpeta local. Por ejemplo, C:\Python27\somefile.py. Luego se abre la línea de comandos y se busca la carpeta de instalación de Python 2.7, generalmente se encuentra en C:\Python27. Al llegar a esta carpeta a través de la línea de comandos, se escribe los siguiente: C:\Python27\Python.exe C:\Python27\somefile.py y aparecerá la interfaz gráfica.

Para correr el programa en MacOS:

Se guarda el archivo .py en una carpeta local. Por ejemplo, Downloads\somefile.py. Se procede a abrir la terminal, que se encuentra dentro de la carpeta de Utilities en Applications. Una vez abierta, se busca la carpeta donde se guardó. Al estar ahí, se escribe: python somefile.py y aparecerá la interfaz gráfica.

Figura 21: Interfaz gráfica del programa que reconoce al intérprete de una canción.



Luego de que aparezca la interfaz gráfica, se encuentra una parte blanca donde aparecerá texto para desplegar ciertos resultados y a su alrededor 4 botones: Open, Clean, Process y Tutorial.

En caso de no dominar al programa, se recomienda hacer click en el botón de Tutorial para recibir otra guía sobre cómo usar el software. El botón de Open se encarga de abrir una ventana para buscar el archivo que se desea procesar. Este archivo no sufrirá cambios una vez haya sido procesado. En la ventana que aparecerá, solo se mostrarán los archivos con extensión .wav que se encuentren en esa carpeta, debido a que son los únicos que este software puede manejar. Los archivos .wav deben tener una tasa de muestreo de 22.050KHz y 16 bits por muestra.

Una vez se haya seleccionado el archivo a procesar (este proceso tarda alrededor de 2 minutos) aparecerá una figura que desplegará una gráfica que representa los bits de la canción de entrada, esto se muestra solamente para que el usuario sepa de qué canción se trata el proceso que se está realizando. Inmediatamente, aparecerá en el área blanca de texto un mensaje para indicar que el proceso ha terminado. Este mensaje despliega los primeros vectores que representan la voz del cantante.

Figura 22: Gráfica de la canción que se le entregó al programa para procesar.

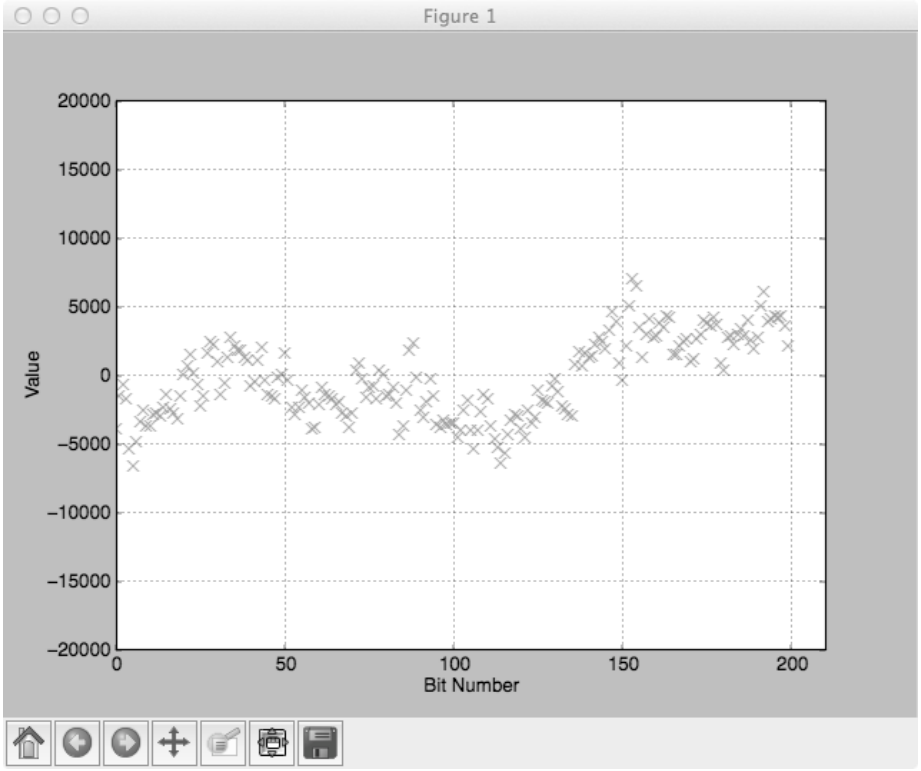
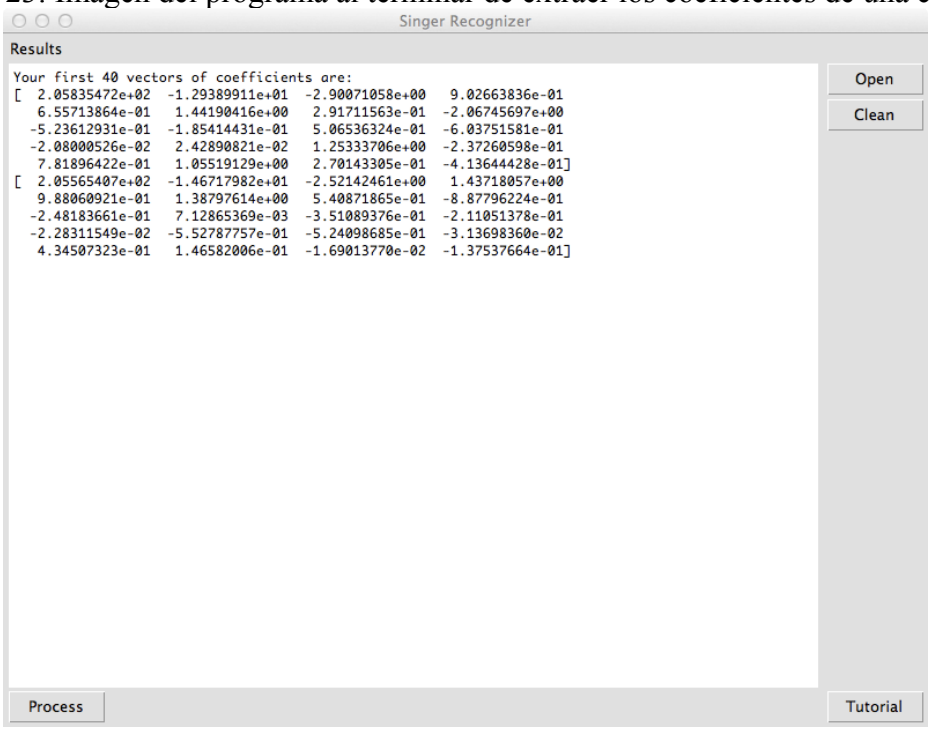
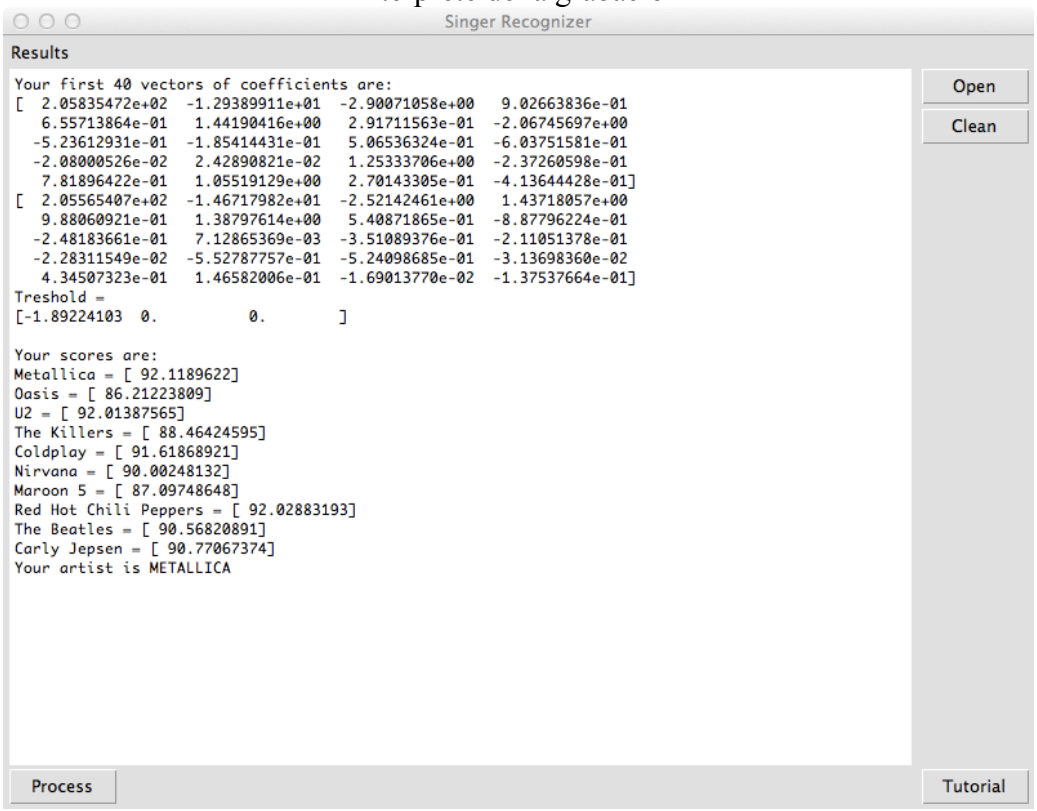


Figura 23: Imagen del programa al terminar de extraer los coeficientes de una canción



Posteriormente, se debe elegir el botón de Process, que lo que hará es hacer todas las comparaciones respectivas para poder decidir cuál artista es el que aparece en la grabación que se procesó anteriormente. Si no se ha seleccionado ningún archivo anteriormente, aparecerá un mensaje indicando que éste debió ser el primer paso. Aparecerá en el área blanca el resultado del puntaje de cada uno de los artistas seguido de la decisión que tomó el programa al hacer todas las comparaciones. Este proceso tarda alrededor de 2 a 3 minutos, dependiendo de la computadora en la que se esté ejecutando el programa.

Figura 24: Ejemplo de la muestra de resultados del programa al tratar de reconocer al intérprete de la grabación



Una vez terminado esto, el programa está listo para seguir procesando archivos y comparaciones. Si al usuario no le interesa ya los resultados anteriores, puede presionar en el botón Clean, que hará que en su pantalla se borren el texto anterior y quede completamente en blanco nuevamente. El programa no tiene límite de número de canciones que puede procesar. Solamente se sugiere que los archivos que se le ingresen no sean más grandes de 10MB para evitar que en algunas ocasiones genere error por falta de memoria.

XII. GLOSARIO

- Extensión .wav: Extensión para un archivo musical, creado por Microsoft en los años ’90. También conocido como .wave

- Python: Lenguaje de programación creado por Python Software Foundation. Trabaja con archivos extensión .py, .pyw, .pyc, .pyd, .pyo.

- Cepstro: O cepstrum, es el resultado de aplicarle la transformada de Fourier a una señal y luego cambiarla a una escala logarítmica.

- Espectro de frecuencias: Es un gráfico que representa la intensidad como variable dependiente y la frecuencia como independiente. Se usa generalmente para verificar qué frecuencias están presentes en cierta señal.

- Bit: Acrónimo de Binary digiT. Puede tomar el valor de 1 o 0 y es un dígito del sistema binario.

- Ruido: Perturbaciones eléctricas que deforman la señal procesada o transmitida y cambian su forma original.

- Atenuación: Pérdida de potencia de una señal en un medio por el cual se ha transmitido.

- Muestrear: También conocido como samplear, es el proceso de convertir una señal en una secuencia numérica.

- Cóclea: Estructura en forma de tubo enrollado en espiral situado en el oído interno. En su interior se encuentra el sentido de audición para los mamíferos.

- Armónicos: Son una serie de variaciones que se sitúan en un rango de frecuencias de emisión.

- Umbral (Threshold): Valor fijado que marca un límite para una discrimación, generalmente.