

Monitor tecnológico para el estudio del solfeo

Danny Ramsey Lasso Chaves

Trabajo dirigido

Tutores

Juan Manuel López López PhD

Omar Baracaldo



**Universidad del Rosario
Escuela Colombiana de Ingeniería Julio Garavito
Programa de Ingeniería Biomédica
Bogotá
Mayo de 2019**

AGRADECIMIENTOS

Quiero agradecer a mi familia por apoyarme en cada momento durante este proceso, a mi madre, a mi padre y a mi hermana Hanna Isabella que siempre estuvieron para mí en los buenos y en los malos momentos. También al señor Juan Manuel López por acompañarme durante el camino a lo largo de mucho tiempo, por su dedicación, paciencia y por sus conocimientos para la realización de este proyecto. De igual modo quiero agradecerle al señor Omar Baracaldo quien con su experiencia apostó por ser mi tutor brindándome muchos conocimientos y siendo una guía para lograr el resultado que se obtuvo, no habría sido posible sin él.

Quiero agradecerles a mis amigos, a los que estuvieron siempre dándome ánimos para seguir adelante y no rendirme, a los que se ofrecieron a colaborar con este proyecto y los que estaban para escucharme cuando lo necesitaba. Agradezco a Cristian Castillo por ser un amigo y un guía, sin sus consejos seguramente no estaría escribiendo este agradecimiento a mis personas especiales hoy.

Quiero agradecerle de forma especial a los profesores que cultivaron su conocimiento en mí, que me brindaron las herramientas para ser un futuro ingeniero íntegro y seguro de sí mismo, por sus exigencias y motivación.

CONTENIDO	
INTRODUCCION.....	4
OBJETIVOS	5
General.....	5
Específicos	5
METODOLOGIA.....	6
Materiales.....	6
Organización de tareas y cronograma de actividades.....	6
Implementación del software	7
Recepción y procesamiento de datos	8
Algoritmos de detección de tonalidad	11
Autocorrelación	12
MIDI	15
Lectura y procesamiento MIDI desde Matlab.....	16
Integración de la información MIDI y la señal medida.....	18
Decisión y realimentación.....	20
Protocolo Experimental	23
RESULTADOS	24
Diseño de aplicación.....	24
Pruebas con sujetos y evidencias.....	30
DISCUSIÓN.....	38
RECOMENDACIONES Y TRABAJOS FUTUROS.....	40
CONCLUSIONES.....	43
REFERENCIAS	44

INTRODUCCION

La formación de un músico se basa en adquirir competencias teóricas y prácticas que no se limitan al dominio de un solo instrumento musical. Los cantantes en particular deben desarrollar competencias en teoría, lectura musical y solfeo que requieren un correcto desarrollo auditivo puesto que de este depende su capacidad de memorizar cada sonido de forma correcta y así mismo de reproducirlo a nivel vocal de forma precisa. El desarrollo auditivo ha sido el foco en la formación musical[1], tradicionalmente métodos implementados para dominar el solfeo consisten en interpretar diferentes discursos musicales de forma repetida para lograr la mayor perfección musical a nivel vocal[1].

problema del solfeo consiste en la capacidad del estudiante para comprender, desde su percepción si se está cantando dentro de un rango de frecuencias donde las alturas se sienten correctas y si se está iniciando cada nota en el momento preciso para dar la sensación rítmica correspondiente al texto musical que se está leyendo.

Históricamente, la enseñanza del solfeo se basa en un modelo “maestro-aprendiz” que permite al estudiante obtener una realimentación confiable basada en la experiencia y conocimientos del tutor.

Callaghan y su equipo [2] mencionan la importancia que tiene la realimentación externa por medio de modelos visuales que hacen parte de la inteligencia corporal-cinética, los cuales pueden ser aprovechados para contribuir al desarrollo del estudio del solfeo de manera activa.

En la discusión de esta investigación se explorarán detalladamente aquellos aplicativos similares con el fin de determinar si existen trabajos que realicen un análisis que se asemeje al realizado en este proyecto y de este modo, garantizar un proyecto viable comercialmente.

Con el marco presentado anteriormente, este proyecto propone el desarrollo de una herramienta tecnológica con base en las herramientas de la ingeniería que, desde el área del procesamiento digital de señales biomédicas apoye a estudiantes de programas de música de educación superior. La herramienta permitiría a los estudiantes conocer y mejorar su afinación y su precisión rítmica mediante el desarrollo de ejercicios propuestos por un profesional con un flujo enfocado al aprendizaje de diferentes habilidades que presentan diferentes grados de dificultad como se realizaría en la academia. Se explorarán las diferentes herramientas audiovisuales que esta investigación ofrece al estudiante y como se genera una retroalimentación y evaluación en tiempo real que le ayudará a ajustar su percepción y control de la afinación y su sensación de las duraciones musicales acompañado de los resultados posteriores al entrenamiento con la herramienta.

En los siguientes capítulos se definen los objetivos a partir de la problemática planteada, el planeamiento de las actividades y técnicas implementadas para el desarrollo de un software y se hará especial énfasis en cómo se realizó la adquisición y procesamiento de los datos para lograr la meta propuesta.

El sonido está en todas partes, hay que aprender a convertir nuestro entorno en arte al pensar y ver la forma de crear y analizar sonido para ayudar a sanar.

OBJETIVOS

General

Diseñar una herramienta audiovisual para el estudio de solfeo, dirigida a estudiantes de un programa de música de educación superior, enfocada al mejoramiento de la afinación y la precisión rítmica.

Específicos

- a. Alimentar el sistema con señales de voz y partituras en formato digital.
- b. Obtener la afinación y la precisión rítmica de señales de voz
- c. Incorporar al sistema una interfaz gráfica
- d. Realizar una evaluación técnica del sistema.
- e. Validación del sistema con los usuarios de manera remota.

METODOLOGIA

Materiales

Computadora con Matlab 2018b o posterior

Organización de tareas y cronograma de actividades

A continuación, se presenta las tareas consideradas para la realización del proyecto:

- 1 Revisión bibliográfica que arroje claridad respecto a métodos de adquisición, procesamiento y segmentación de señales de voz enfocadas al canto, procesamiento y comprensión de protocolo MIDI y extracción de características, modelos de aprendizaje del solfeo y representación matemática musical.
- 2 Adquisición de señales de voz.
 - Escoger el protocolo necesario para adquirir señales de voz en términos de frecuencia de muestreo.
 - Filtrar espectro fuera del rango de voz para el canto y sonidos de alta frecuencia que aparecen al pronunciar ciertas letras.
 - Procesar la señal fisiológica en tiempo real y extracción del contorno de la voz a partir de la frecuencia fundamental temporal.
- 3 Procesamiento de archivos MIDI con la información de la partitura a interpretar.
 - Implementación de algoritmos de extracción de características frecuenciales y temporales.
 - Construcción de una señal a partir de características que permitan generar análisis cuantitativos al comparar con la señal de voz procesada.
- 4 Procesar y analizar las señales de voz en conjunto con la información MIDI, para identificar la precisión del tono y la precisión rítmica.
 - Re muestrear y comparar las señales adquiridas y generadas para la posterior implementación de un algoritmo de evaluación basado en métodos estadísticos.
 - Almacenar los resultados de calificación del tono y del ritmo.
- 5 Diseñar interfaz gráfica de usuario
 - Implementar herramientas basadas en métodos de estudio del solfeo.
 - Visualización de curvas de contorno de voz y objetivo que permitan una fácil abstracción de información relevante para el usuario.
 - Selección de paleta de colores y diseño gráfico de componentes.
 - Programación de asistente que guíe al usuario hacia el correcto uso de la aplicación de manera autónoma.
- 6 Pruebas piloto
- 7 Evaluar el sistema
 - Evaluar el sistema con los estudiantes de música de los Andes durante un mes.
- 8 Correcciones finales

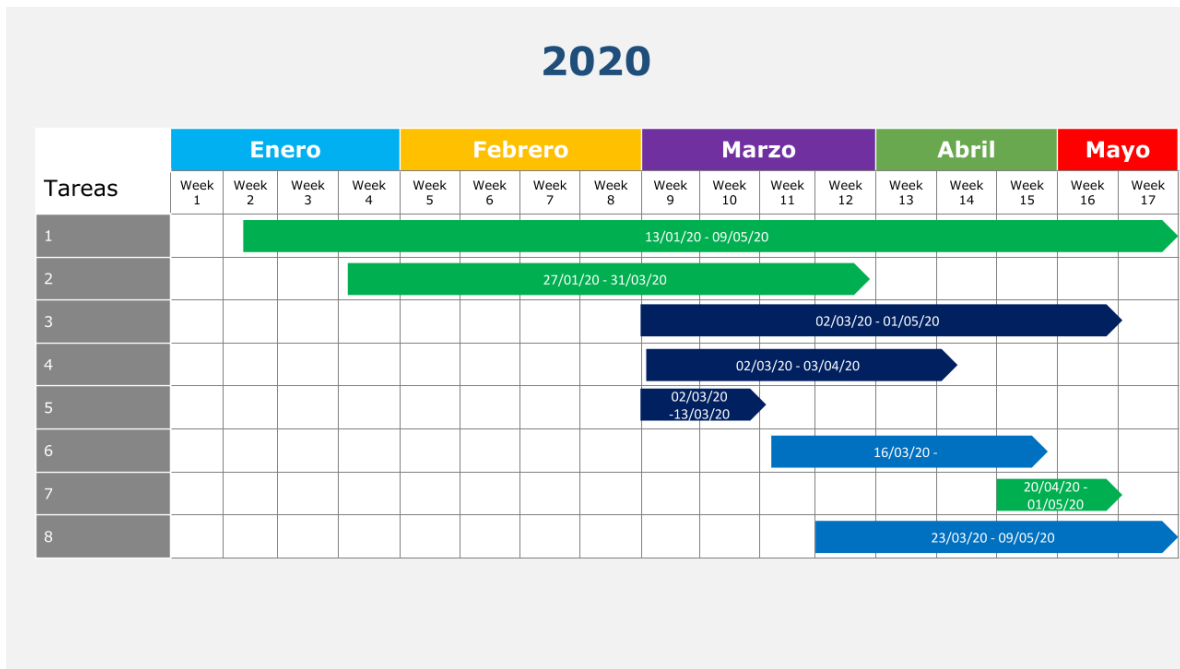


Figura. 1. Diagrama de Gantt del proyecto

El desarrollo de las actividades anteriormente propuestas tiene como propósito lograr una herramienta audiovisual que permita a los estudiantes mejorar sus habilidades de solfeo de forma intuitiva y que brinde las herramientas básicas que se usan comúnmente para el estudio acopladas con un software que orienten el aprendizaje de acuerdo con los lineamientos establecidos por un cronograma típico de aprendizaje de cursos de formación musical. Las limitaciones de este proyecto se encuentran principalmente en la dificultad de realizar experimentos de manera controlada enfocados al análisis de variables cuantitativas para determinar la eficacia del software en el mejoramiento de las habilidades. Esto debido a la dificultad del trabajo con personas bajo circunstancias inesperadas producto de la coyuntura actual a causa de la pandemia.

Implementación del software

Abordar el problema descrito y lograr que un estudiante de música logre generar dicha percepción de manera independiente al realizar un ejercicio de solfeo, requirió la aplicación de diferentes técnicas implementadas en el campo del procesamiento digital de señales biomédicas aplicadas a la voz para el canto. Diferentes Investigaciones han tenido como objeto de estudio este tipo de señal, algunos de los campos que abordan la problemática desde la Ingeniería Biomédica son la biomecánica, los sistemas de control y el procesamiento de señales biomédicas[2][3].

Una aplicación que involucre los métodos de adquisición de información digital y analógica realizada correctamente como una solución que permita segmentar un mismo grupo de características y permita realizar una comparación desde el punto de vista matemático

basado en la relación que tienen ambas señales, impulso al desarrollo de un producto con un funcionamiento relativamente simple cuyo proceso de manipulación por parte del usuario se resume en la siguiente figura.

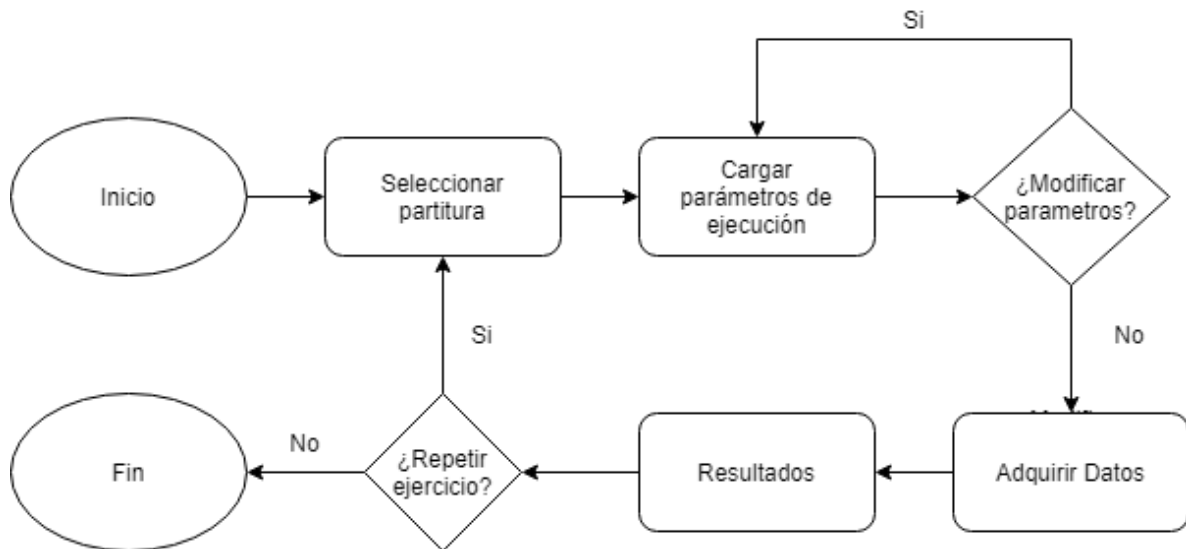


Figura. 2. Diagrama de flujo de la implementación del software.

Recepción y procesamiento de datos

Existen diferentes lenguajes de programación orientados a objetos que ofrecen herramientas avanzadas en el análisis y tratamiento de datos las cuales están dotadas de diferentes paquetes de funciones y comandos simples.

Algunos de los lenguajes más populares en la industria hoy en día son aquellos que permiten realizar toma de decisiones y manejar datos en grandes cantidades de manera estadística.

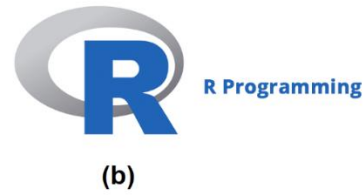


Figura. 3. Matlab(a), R Programming(b), Python(c) y Octave(d), son algunos de los lenguajes más ampliamente utilizados por estudiantes para investigación y estadísticas.

Existen numerosas ventajas y desventajas dependiendo de la aplicación deseada que podemos describir de los diferentes lenguajes. Python por su parte ofrece paquetes de uso libre como Pandas, Numpy, Scipy, entre otros, es popularmente usado tanto en aprendizaje general como en desarrollo de aplicaciones y es un lenguaje de procesamiento rápido.

R es un lenguaje para ejecutar análisis estadísticos y graficar resultados de salida, es un software de distribución libre y puede ejecutarse en cualquier sistema operativo. Sin embargo, no es adecuado para presentar datos de sencilla interpretación para los usuarios.[4]

Por otra parte, Matlab es el entorno que más se utiliza no solo como un entorno comercial numérico, sino también como lenguaje de programación. Este lenguaje tiene una librería estándar que incluye herramientas para álgebra matricial, procesamiento de datos y generación de curvas. Cuenta con numerosas Toolboxes que permiten incrementar las funciones básicas. Una de las grandes ventajas de Matlab sobre los otros lenguajes presentados, radica en la posibilidad de procesar y generar datos de manera sencilla de leer, esto aplica para el preanálisis y el realizado posteriormente para entregar al usuario final[4].

Teniendo en cuenta las ventajas y desventajas de los distintos programas analizados, fue decidido que el desarrollo de este software se llevó a cabo en el lenguaje de programación basado en operaciones matriciales Matlab. Este sistema de cómputo numérico ofrece un entorno de desarrollo integrado con un lenguaje de programación[5] propio que permite a sus usuarios desarrollar diferentes herramientas tecnológicas basadas en la ciencia que además cuenta con la capacidad de generar y procesar señales analógicas y digitales. Este

lenguaje ofrece soporte para diferentes problemáticas que aparecen a los usuarios a la hora de diseñar aplicaciones. Esto resulta apropiado para lograr un menor tiempo de desarrollo y mejor calidad de producto. Adicionalmente, Matlab cuenta con una colección de funciones que como se mencionó anteriormente, están diseñadas en su mismo entorno. Estas funciones llamadas Toolboxes, pueden ser usadas para desarrollar proyectos de mayor complejidad debido a que se presentan como mejoras al entorno básico ofrecido al usar Matlab. Para este proyecto en particular, se utilizaron dos toolboxes, la desarrollada por Ken Schutte[6] que incluye funciones básicas para sintetizar y leer archivos en formato MIDI permitiendo extraer información importante, sin embargo se hablara a mayor detalle sobre ella más adelante. Otra de las toolbox utilizadas es la Data “acquisition Toolbox” que provee al usuario distintas apps y funciones para configurar la lectura de datos mediante hardware resultando así muy útil para el proyecto puesto que uno de los objetivos responde a poder capturar señales de voz y procesarlas en tiempo real[7].

De este modo se realizó un código para automatizar la adquisición de las señales de voz provenientes del micrófono incorporado en la computadora y su procesamiento en tiempo real.

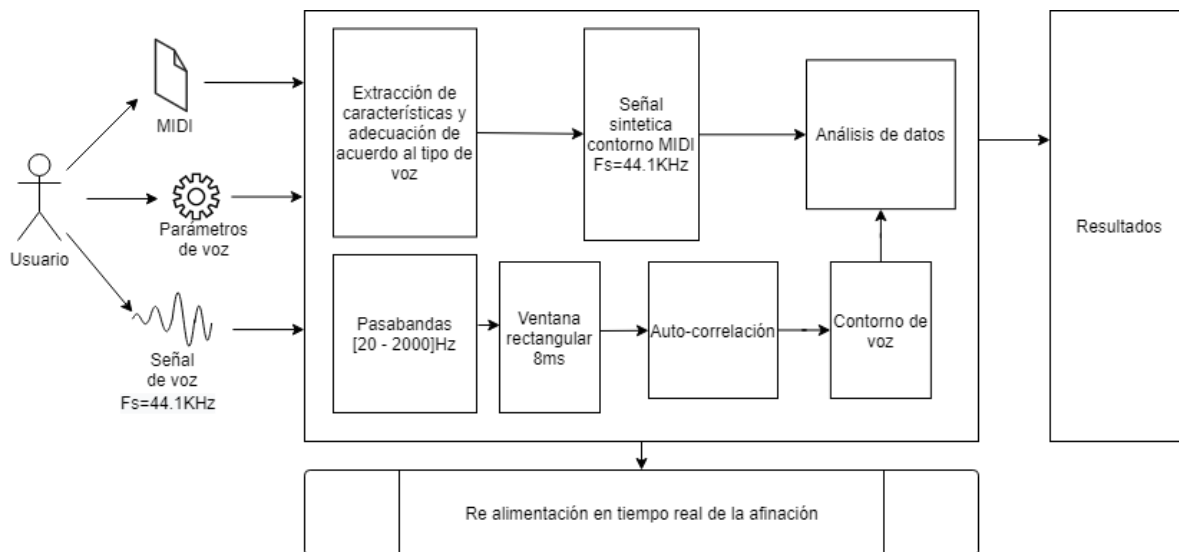


Figura. 4. Diagrama de flujo del algoritmo implementado para el procesamiento de los datos recibidos por la interfaz.

El ancho de banda de la señal de voz durante el canto responde a la naturaleza mecánica de oscilación de las cuerdas vocales, estas vibraciones varían de acuerdo con el sexo y se presentan en rangos diferentes para cada individuo. La literatura sugiere que el rango de la voz humana durante el canto se encuentre en un rango de entre los 65Hz y los 20KHz [8], por lo que según dicta el teorema de muestreo de Whittaker-Nyquist y teniendo en cuenta la máxima frecuencia de la señal, debemos muestrear a al menos el doble de esta

frecuencia y así garantizar que la información digitalizada contenga las características que deseamos en el ancho de banda establecido.

El objetivo del procesamiento en tiempo real se ha descrito por algunos autores como el proveer soluciones que puedan procesar cadenas de datos continuos de manera infinita involucrando los datos ya obtenidos y los que se están adquiriendo en vivo de manera rápida e interactiva[9].

En este orden de ideas, se debe garantizar una adquisición y procesamiento de la señal de manera continua a una frecuencia de muestreo que nos permita generar una realimentación en un lapso de tiempo menor a 200 milisegundos[10], tiempo menor al que le toma a la información viajar desde los ojos hasta niveles corticales más altos. Los cambios en estímulos visuales debajo de este rango son imperceptibles para la visión rápida y generan la sensación de tiempo real.

Con base en lo anterior se escogió una frecuencia de muestreo de 44.1Khz.

Algoritmos de detección de tonalidad

Evaluar la trayectoria de la frecuencia fundamental f_0 ha revelado en distintas investigaciones que las características del canto son más dinámicas y complejas que las de la voz hablada. Por ejemplo, durante el canto la frecuencia fundamental sufre diferentes fluctuaciones como el vibrato y el overshoot [11]. Estas deben detectarse a lo largo del tiempo, almacenarse e interpretarse de tal forma que nos permitan realizar un correcto análisis a la hora de discernir si una nota es correcta o no. El vibrato es una técnica muy comúnmente usada durante la voz para el canto y existe un rango aceptable descrito en diferentes estudios en música sobre solfeo[12] que sugieren que el vibrato aplicado correctamente debe tener una amplitud en frecuencia de entre 30 a 50 por ciento de un semitono. Se definió entonces un umbral del 30% de un semitono por arriba o debajo de la nota objetivo dentro del cual la nota es correcta en términos de altura o frecuencia.

Otro ítem importante para tener en cuenta para realizar una correcta segmentación de la información proveniente de la señal fisiológica fue el filtrado. Tomoyasu Nakano y su equipo, demostraron en sus investigaciones[12], que aplicar un filtro pasa bajas con una frecuencia de corte 5Hz, ayuda a eliminar dichas fluctuaciones y a preparar la voz del canto. En este proyecto se implementó un filtro pasa bandas digital en el rango de los 65 y los 1397Hz con el fin de analizar únicamente el espectro de la voz para el canto y limpiar la señal de ruido ambiental y particularmente de sonidos de siseo que se encuentran comúnmente en el espectro de los [4 - 7]KHz.

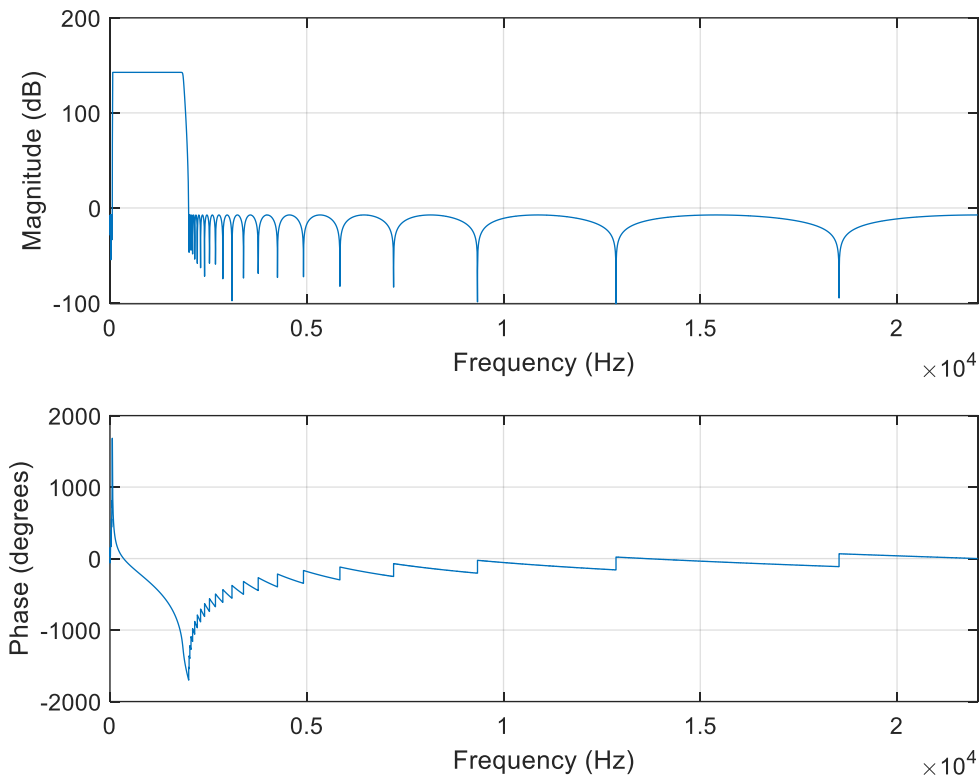


Figura. 5 se visualiza la respuesta en frecuencia de un filtro de orden 4 tipo IIR con frecuencia de corte inferior de 4Khz y frecuencia de corte superior de 7Khz a una frecuencia de muestreo de 44.1Khz

Una vez la señal esta lo más limpia de artefactos posible, se procede a extraer la frecuencia fundamental producida por el usuario al cantar. Para esto, se debe construir señales deterministas representativas a partir de una señal estocásticamente periódica. Una función que tiene la propiedad de detectar periodicidad en señales de estas características es la autocorrelación[13].

Autocorrelación

La secuencia de autocorrelación de una señal periódica posee las mismas características cíclicas que la señal original, es por tal motivo que la autocorrelación es una herramienta clave para la detección de la frecuencia fundamental representada como la presencia de ciclos y que permite también determinar las duraciones de dichos ciclos.

La autocorrelación, se realiza directamente sobre la curva de una manera bastante sencilla por lo que no requiere mucho tiempo al ordenador procesar este método. Implementando las funciones tic() y toc() se midió el tiempo que tarda dicho procesamiento y filtrado el cual arroja el siguiente tiempo: "Elapsed time is 0.004181 seconds" que corresponde a un tiempo de procesamiento de 4.181 milisegundos. Una muy importante consideración para aplicar la función de autocorrelación, es la de tener en cuenta el tamaño de la ventana a analizar

(window size), ya que lo recomendable es que el segmento para el análisis contenga al menos 2 o 3 periodos de la frecuencia a analizar[14]. Lo anterior se puede explicar debido a que la curva característica de la función de autocorrelación en señales periódicas se presenta como una serie de picos que indican el periodo entre cada uno de los periodos contenidos en la señal de interés los cuales varían entre los

Calcular la autocorrelación de una señal consiste en encontrar la correlación de una señal con una versión de sí misma con un desplazamiento en el tiempo. Matemáticamente, esto se puede expresar como:

$$\varphi_l(m) = \frac{1}{N} \sum_{n=0}^{N-1} [x(n + 1)w(n)][x(n + 1 + m)w(n + m)]$$

$$-0 \leq m \leq M_0 - 1$$

Ecuación. 1. Definición de la función de autocorrelación.

Donde $w(n)$ representa la ventana apropiada para el análisis, N es la longitud de la sección analizada, N' es el número de muestras usadas para computar $\varphi_1(m)$, M_0 es el número de puntos de correlación a computarse y l es el índice de la primera muestra de la ventana.

Al computar la función de correlación para comparar una señal, contra la versión de sí misma desplazada obtendremos la periodicidad de procesos estocásticos y también el periodo de oscilación como se mencionó anteriormente. Observaciones heurísticas señalan que la autocorrelación será máxima cuando la señal es comparada consigo misma sin desplazamiento alguno. Por otra parte, cuando un proceso es periódico, su autocorrelación genera una señal característica de la cual es posible calcular el periodo entre eventos encontrando así la frecuencia con que oscilan las cuerdas vocales en determinado instante.

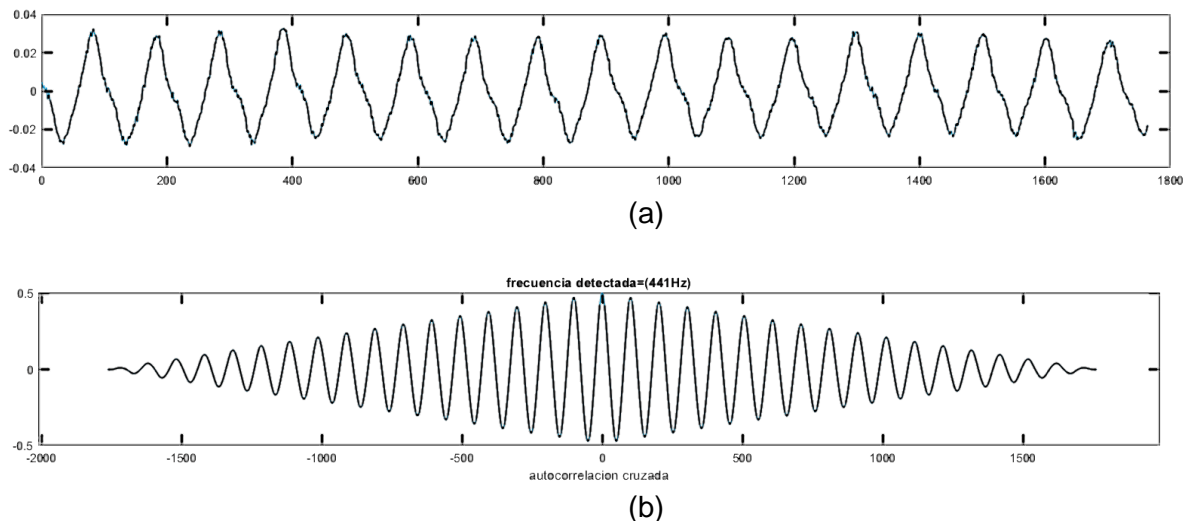


Figura. 6. Fragmento de voz a analizar (a) y señal de autocorrelación generada cuando una persona canta la nota “la4” a una frecuencia de 440Hz.

Encontrar la trayectoria de la frecuencia fundamental de la voz al cantar se resume en ser capaces de determinar periodos cortos y largos mediante la localización de picos y determinación de tiempo promedio entre ellos. Con el fin de lograr encontrar dichos picos, se tuvo en cuenta la frecuencia de las notas y en consecuencia se definieron ventanas que permitieran determinar tanto las notas de frecuencias altas como bajas. El análisis puede realizarse con ventanas de 5 a 20 ms para notas de alta frecuencia y para frecuencias bajas la ventana debe ser más larga alrededor de los 20 a 50ms [9]. Teniendo en cuenta esto, se escogió tomar ventanas de 40 milisegundos.



Figura. 7. Lógica implementada para calcular la autocorrelación en ventanas de 40 milisegundos

Los 40 milisegundos sugeridos se calculan teniendo en cuenta que la frecuencia de muestreo “fs.” a la que se adquirió la señal de voz fue de 44.1KHz, por lo que el tamaño de la ventana sería de 1764 muestras que, en términos de tiempo, representarían un total de 0.039 segundos de información procesada. Esto quiere decir que cada 40 milisegundos seremos capaces de extraer una medición de la frecuencia fundamental y tendremos todavía unos 60 milisegundos para generar información visual que genere la percepción de tiempo real como lo sugiere la literatura[10].

Tras haber calculado la correlación en la ventana de tiempo adecuado, se procedió a eliminar el pico máximo que aparece en 0 que simplemente nos dice que en ese punto la correlación es máxima, lo cual tiene todo el sentido puesto que estamos comparando la señal contra si misma sin ningún desplazamiento lo cual, genera un match perfecto. Tras eliminar el pico máximo en la señal obtenida, se procedió a calcular el valor absoluto de la señal con el fin de eliminar los valores negativos obteniendo la siguiente señal.

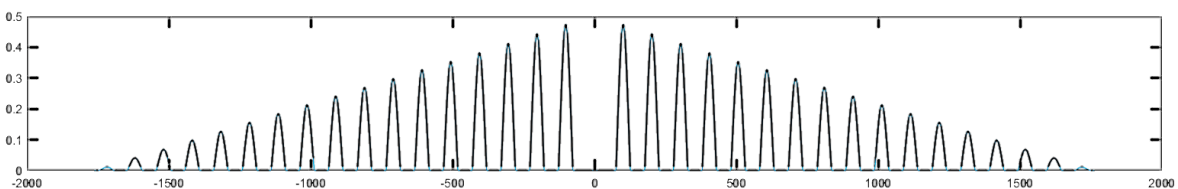


Figura. 8. Valor absoluto de la correlación tras eliminar el pico de la mitad y aplicar valor absoluto.

A continuación, se determinó la localización del siguiente pico máximo y con ella, se procedió entonces a calcular el número de muestras a la que se encuentra el siguiente pico y finalmente se calcula la diferencia en muestras entre el primer pico encontrado y el segundo para después dividir este valor por la frecuencia de muestreo y así determinar la frecuencia fundamental de la porción de señal medida en Hz.

Mediante la extracción de esta característica durante la adquisición de datos en pequeñas ventanas de tiempo, podemos realizar una realimentación visual para que el usuario sepa la frecuencia a la que está cantando de modo que el resultado se le entregue en notación musical conocida.

La información obtenida durante el registro de la voz se almacena entonces en un vector que nos muestra las notas interpretadas en cada instante medido en muestras donde la frecuencia de muestreo es conocida y por ende es fácil identificar las frecuencias en cualquier instante de tiempo. Esto último será la clave para generar los resultados al comparar las características extraídas de la voz contra las características extraídas de la partitura en formato MIDI. Se explicará dicha extracción más adelante.

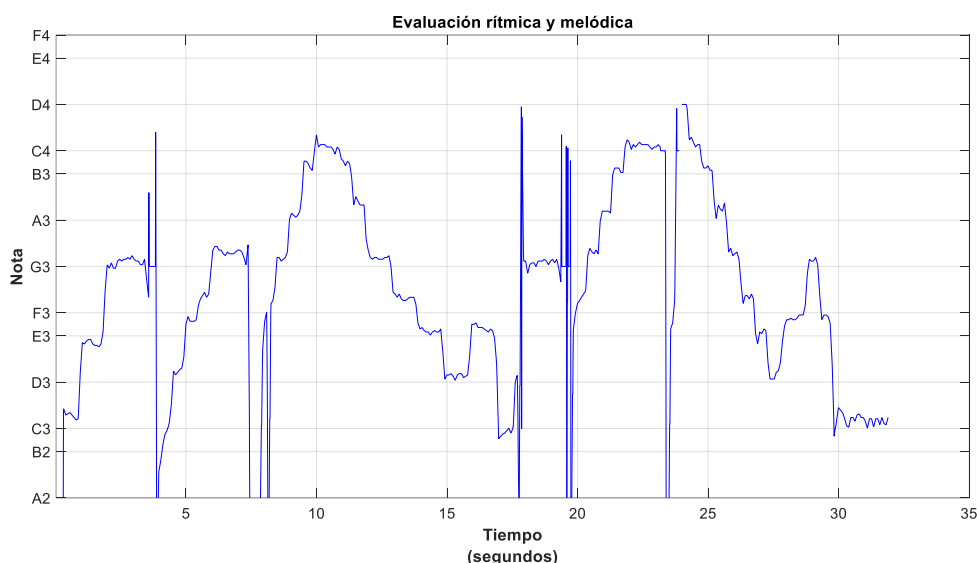


Figura. 9. Señal obtenida de realizar el seguimiento de la frecuencia fundamental a lo largo del tiempo. Se puede observar que la frecuencia de muestreo es diferente a la establecida para adquirir la señal.

MIDI

El formato de audio para comunicar instrumentos musicales MIDI (Musical Instrument Digital Interface) es el protocolo binario más ampliamente utilizado ya que permite conectar diferentes equipos musicales como sintetizadores, paneles de control y computadoras. Este protocolo tiene sus orígenes a principios de los años 80s cuando empresas que manufacturaban equipos musicales buscaron la forma de universalizar los métodos de comunicación de un equipo a otro sin necesidad de usar secuenciadores[15].

Los archivos en formato MIDI en su nivel más alto contienen dos elementos llamados eventos. Existen dos componentes en cada evento, estos son el tiempo MIDI y el mensaje MIDI. El tiempo MIDI consiste en la duración entre el tiempo que transcurre entre un evento

y otro[15]. El tiempo es un diferencial de tiempo que especifica la duración entre dos eventos y ocurre como se explica a continuación:

ΔT	Tiempo transcurrido hasta inicio del evento
$t_1 = \Delta T_1 = T_1 - 0$	$T_1 = t_1$
$t_2 = \Delta T_2 = T_2 - T_1$	$T_2 = t_1 + t_2$
$t_3 = \Delta T_3 = T_3 - T_2$	$T_3 = t_1 + t_2 + t_3$
.	.
.	.
.	.
$t_n = \Delta T_n = T_n - T_{n-1}$	$T_n = t_1 + t_2 + \dots + t_n$

Tabla. 1. Tiempo transcurrido en diferentes eventos consecutivos.

Por otra parte, el mensaje MIDI puede contener distintas instrucciones como encender y apagar una nota, la velocidad, volumen y otras características que no se detallaran en esta investigación. Los más importantes mensajes son el comando de nota encendida “Note-on” y nota apagada “Note-off” que se traducen en iniciar o parar la reproducción de una nota en particular. Es importante mencionar que en este evento existen dos modos de reproducción de la nota en ejecución, los cuales son el modo monofónico y polifónico. El modo polifónico es una restricción para las partituras con las que trabajaremos en la investigación, pues este consiste en reproducir múltiples notas al tiempo y esto no es algo habitual en el solfeo de partituras de las cuales hablaremos en otro apartado de este documento.

Por lo anterior, para poder realizar un análisis de notas consecutivas se trabajó únicamente con archivos MIDI codificados según el método monofónico, el cual recibe el comando “Note-on” a la vez que termina la ejecución de una nota previa al igual que una persona leyendo una partitura.

Dichos mensajes de encendido y apagado conllevan información sobre parámetros como la frecuencia del tono y la velocidad (la intensidad con la que se presiona la nota)

La implementación de un algoritmo que sea capaz de extraer información a partir de un archivo MIDI es una tarea que puede resultar exhaustiva debido al volumen de información técnica contenida en el entendimiento del formato, razón por la cual se implementó para el desarrollo de este proyecto la Toolbox mencionada anteriormente.

Lectura y procesamiento MIDI desde Matlab

El interés de manipular archivos MIDI en esta investigación, radica principalmente en lograr un sistema que sea capaz de leer la información sobre las notas y la duración de estas, con el fin de generar bloques de datos con características secuenciales de las notas relacionadas con: el tiempo de ejecución respecto a un tiempo definido y frecuenciales.

Dichas características deben extraerse de manera tal que permitan compararse con las mismas características extraídas de la voz por medio de métodos que se explicarán más adelante.

Al leer la información de los archivos MIDI en Matlab se pudo observar diferentes mensajes, dentro de los cuales se destaca “ticks_per_quarter_note”, que nos ofrece información clave sobre cómo vamos a procesar los datos.

```
4/2, clock ticks and notated 32nd notes=24/8
C Major
microsec per quarter note: 735294
ctrl=UNKNOWN CONTROLLER value=0
instr=0
ctrl=Volume value=100
```

Figura. 10. Información obtenida al leer un archivos MIDI.

El correcto análisis de piezas musicales conlleva el entendimiento profundo de un aspecto fundamental del patronaje o articulación temporal del sonido en la comunicación musical. Dicho esto, es necesario entender que el tempo es una taza de eventos en un entorno que ocurren con una frecuencia que puede ser lenta o rápida. En música, el tempo es la frecuencia con la que un oyente percibe latidos que ocurren en lapsos de tiempo iguales[16].

Estos latidos en lenguaje musical están notados y cada beat se identifica como una negra o cuarto de nota. Fue necesario extraer de los archivos MIDI la información referente a este componente pues esto nos permite estandarizar el tempo en las diferentes partituras para poder generar resultados con condiciones controladas.

1.0000	0	60.0000	80.0000	0	0.6970	11.0000	12.0000
1.0000	0	64.0000	80.0000	0.7353	1.4323	13.0000	14.0000
1.0000	0	67.0000	80.0000	1.4706	2.8661	15.0000	16.0000
1.0000	0	60.0000	80.0000	2.9412	3.2889	17.0000	18.0000
1.0000	0	62.0000	80.0000	3.3088	3.6566	19.0000	20.0000

Figura. 11. Información obtenida al ejecutar la función “midInfo()” un archivo MIDI.

El siguiente paso consistió en extraer la información en matrices de datos que contienen información sobre cada nota musical y el tiempo en el que ocurre cada una exclusivamente. Para ello se hizo uso de la función “midInfo()” la cual es capaz de retornar una matriz de Nx8 (siendo N el número de eventos en que se reproduce y detiene una nota en específico) como se observa en la figura anterior. Las columnas indican las siguientes características:

1. Número de pista
2. Número de Canal
3. Nota musical codificada en número MIDI
4. Velocidad de pulsación

5. Tiempo de inicio en segundos
6. Tiempo de finalización en segundos
7. Número de mensaje "Note-on"
8. Número de mensaje "Note-off"

Se extrajeron únicamente las características correspondientes al número MIDI que identifica la nota musical y los tiempos de inicio y fin. Correspondientes a las columnas 3,5 y 6. Es importante recalcar que las columnas 5 y 6 contienen unidades medidas en segundos por lo que para estandarizar un tiempo implica convertir el valor obtenido en "ticks_per_quarter_note" en segundos para poder realizar el cálculo. Realizado esto, podemos dividir las columnas de tiempo entre el valor de una negra para poder así generalizar un tempo de 60bpm para todas las partituras.

Número MID	Tiempo inicio	Tiempo parada
60	0	0.9479
64	1	1.9479
67	2	3.8979
60	4	4.4729
62	4.5	4.9729

Tabla. 2. Características extraídas y procesadas a un tempo de 60pbm

Una vez procesada esta información, seremos capaces de tener una referencia contra la cual comparar la voz, pero para ello, a partir de la información recopilada, se va a generar una señal homologa a la que de manera ideal una persona podría crear si lograra un desempeño perfecto del ejercicio.

Integración de la información MIDI y la señal medida

A partir de las características extraídas del archivo MIDI, se creó un vector de tiempo con base en las estampas de tiempo de la señal muestreada, este vector de tiempo se rellenó con las notas correspondientes en los instantes indicados según la matriz de características extraídas para poder generar así una señal sintética ideal con la misma frecuencia de muestreo que la señal de voz adquirida y el mismo número de muestras. De este modo se logró obtener una señal homologa que permite realizar una comparación directa entre ambas señales.

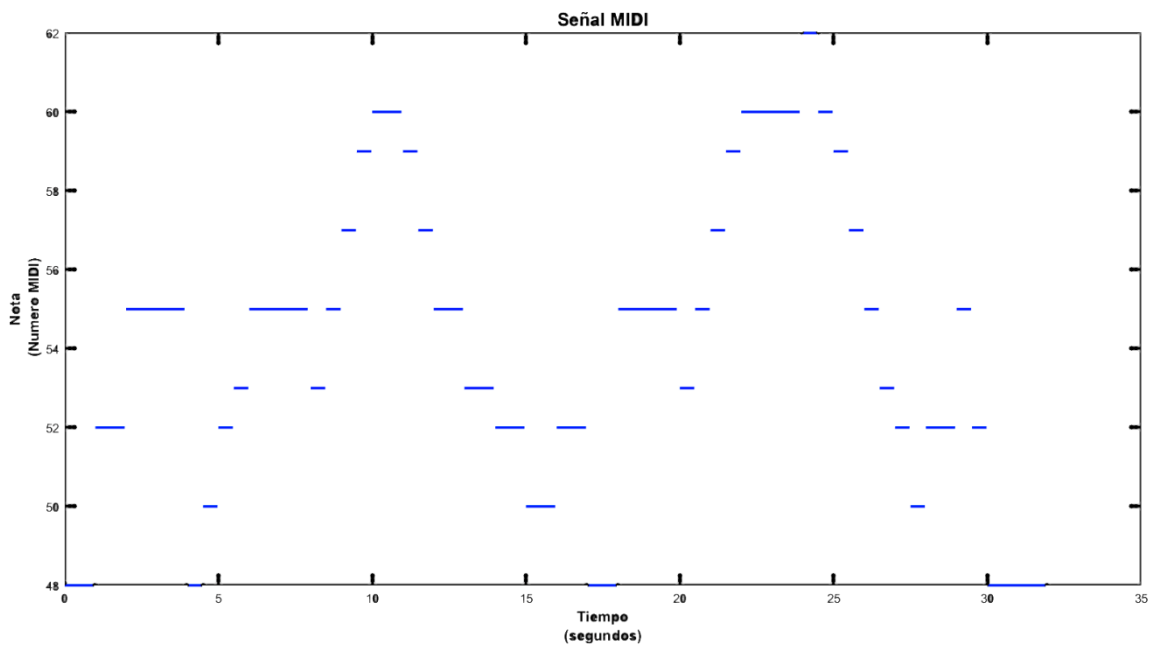


Figura. 12. Señal MIDI generada a partir del procesamiento de los datos recopilados.

Si superponemos la señal del contorno de la voz y la señal sintética generada a partir del archivo MIDI, obtendremos finalmente la curva como la que se muestra a continuación

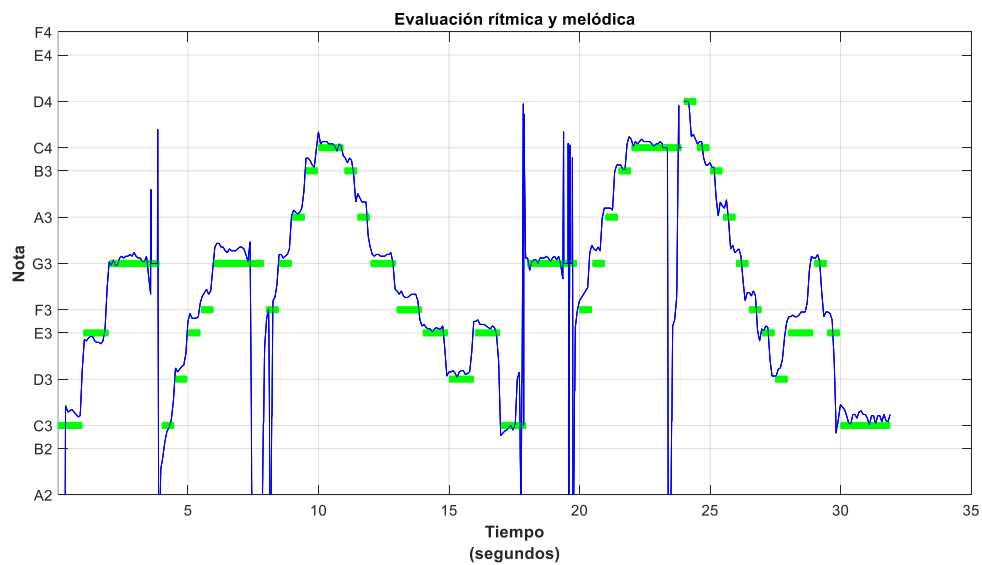


Figura. 14. Curva generada al superponer la señal del contorno de la voz y la señal sintética a partir del archivo MIDI con los respectivos ejes vertical y horizontal expresados en notas musicales y tiempo respectivamente.

Decisión y realimentación

Ahora se analizará los criterios de evaluación que se tuvieron en cuenta para el análisis de la información.

Para poder realizar la comparación de características nota a nota, fue necesario generar vectores que contengan los índices de las muestras correspondientes al principio y el fin de las notas en muestras del vector que contiene la información MIDI.

```
Notaactual=metrica(i,1);%componentes para analizar datos nota a nota
t0=metrica(i,2);      %componentes para analizar datos nota a nota
tf=metrica(i,3);      %componentes para analizar datos nota a nota
                    %i corresponde al numero de la nota actual

tseg=round(t*100); % t ms-> seg
t0=round(t0*100);
tf=round(tf*100);
% find index position of t0 and tf in the tseg vector
indice0(i) = find(tseg(1,:)== t0);
indicef(i) = find(tseg(1,:)== tf);
%extraemos la porcion de la senal que corresponde a esa ventana para
%hacer comparacion tiempo frecuencia nota a nota
segment=pitch(indice0(i):indicef(i));
reference=NoteSignal(indice0(i):indicef(i));
error=reference-segment; %el error entre voz y MIDI
```

Figura. 14. Algoritmo utilizado para encontrar los índices de inicio final de cada nota (t0 y tf)

Al encontrar los índices de cada nota podemos calcular el error en escala de notas MIDI que hay en el conjunto de muestras en cada nota. Esto fue muy útil para determinar cómo abordar el problema del vibrato y del overshoot.

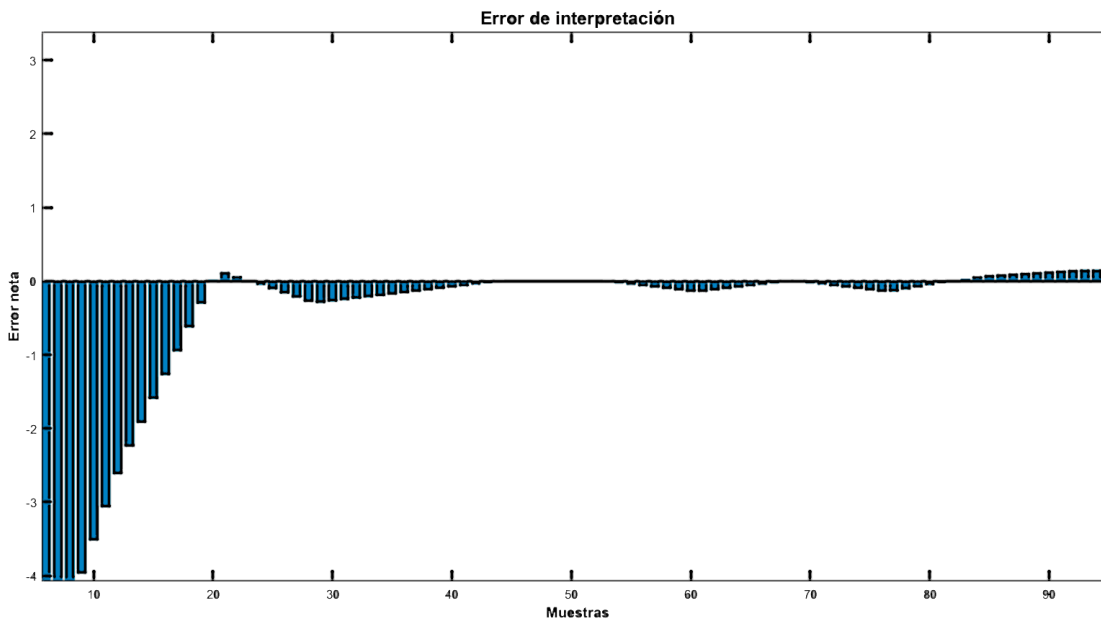


Figura. 15. Ejemplo del error en una nota con presencia de vibrato y overshoot

Con base en distintos artículos [17], [18], podemos definir el vibrato como una fluctuación periódica usualmente sinusoidal en la afinación de una nota con una frecuencia de modulación de entre 4 y 7 Hz. El overshoot por su parte se define como una deflexión que excede la nota objetivo justo antes del cambio de nota. Por lo tanto, se tomó como una nota acertada si la frecuencia medida no sobrepasaba estos rangos. En términos de notas, el vibrato bien realizado no debe entonces sobrepasar 1 semitono y se toman como correctas las notas que entren en este rango.

```

% componente melodica: se busca que el error no supere 1 semitono (vibrato)
inrange = find( abs(error(1,:))<=0.5); %threshold in tones
IR = isempty(inrange);
if IR==1
    accuracymel(i)=3;
elseif length(inrange)>=((3*length(error))/4)
    accuracymel(i)=1;
else
    accuracymel(i)=2;
end
Npositiv=length(find(error(1,*)>0));
Nnegativ=length(find(error(1,*)<0));
if Npositiv > Nnegativ
    signo(i)=1;
else
    signo(i)=-1;
end

```

Figura. 16. Algoritmo implementado para la detección de notas que cumplan con la condición de vibrato para generar un análisis melódico

En el algoritmo mostrado anteriormente, se tienen en cuenta dos aspectos para generar un análisis de la componente melódica. El primero es que los valores que se encuentran dentro del rango aceptable representen al menos tres cuartas partes de la duración de la nota para garantizar que se sostuvo la nota un tiempo considerable representando esto un acierto en la nota, y segundo que, de no cumplirse la primera condición, dependiendo de la magnitud del error, el sistema clasificara dicho error como aceptable o insuficiente tanto por lograr una nota por arriba del objetivo o por debajo.

El overshoot es una característica importante del canto puesto que a partir de ella podemos detectar los cambios de nota definidos como tiempos de ataque. Ya que el overshoot durante el canto ocurre justo después del cambio de nota, podemos usar esto como una referencia para determinar el tiempo de ataque y poder establecer un rango en el cual se considera que la nota se cantó a tiempo, tarde o muy pronto. Detectar estas deflexiones en la nota se resume en detectar el primer cruce por cero en el que el error pasa de un valor positivo a uno negativo o, al contrario.

```

% componente ritmica: Zero crossing
zerocross = find( error(1,:) >=-0.5 & error(1,:) <=0.5 );
TF = isempty(zerocross);
if TF==1;
    accuracy(i)=3;
    Tatakaue(i)=-1;
else
    rango=length(error)/3;
    if zerocross(1) <= rango
        accuracy(i)=1;
        Tatakaue(i)=0;%
    elseif (zerocross(1) >= rango & zerocross(1) <= 2*rango )
        accuracy(i)=2;
        Tatakaue(i)=1;
    else
        accuracy(i)=3;
        Tatakaue(i)=-1;
    end
end

```

Figura. 17. Algoritmo implementado para la detección de tiempo de ataque basado en el overshoot para generar un análisis rítmico.

Del mismo modo en el que se clasificaron las muestras de errores en la componente melódica, en la componente rítmica también se clasificó el error de acuerdo con el overshoot señalando si este ocurrió durante algún momento en el con la etiqueta de un tiempo de ataque tardío.

Protocolo Experimental

Para llevar a cabo la prueba de la herramienta auditiva, se llevarán a cabo las siguientes prácticas con el fin de obtener unos resultados confiables.

Esquema de sesiones:

Se tienen preparadas 8 sesiones en las cuales se evaluarán diferentes habilidades de los estudiantes.

Las sesiones se definen a continuación:

- Grados conjuntos
- Acorde tónica
- Acorde subdominante
- Acorde Dominante

En la primera etapa de cada sesión, se realizará un trabajo de calentamiento y preparación de la voz previo al trabajo con las partituras.

Posterior a la etapa de preparación, se le sugiere al participante que, con ayuda de un metrónomo, solfeé la partitura a 40 y a 52 bpm para mejorar sus resultados finales.

Finalmente, el participante, debe hacer uso de la herramienta auditiva disponible en internet para registrar su voz (Se requiere el uso de audífonos) haciendo la lectura de la partitura correspondiente a la sesión a trabajar a una velocidad standard de 60BPM la cual debe guardarse al finalizar el ejercicio y enviarse por correo para poder generar un historial de cada estudiante. Con el fin de encontrar evolución en el progreso, se repetirá el ejercicio transcurrido una vez pasado un día como mínimo entre ambas sesiones.

La fase de validación tendrá una duración de 8 días máximo.

- Se debe tener en cuenta que, para la recepción de los archivos generados por los estudiantes, se manejaran dos horarios definidos. Los estudiantes podrán enviar los resultados antes de las 5 de la tarde.
- Los estudiantes deben usar audífonos en el audio pues solo deberá escucharse la voz de la persona y no el metrónomo por que podría afectar la medición.
- Si el estudiante logra superar una calificación mínima de 4.0 en cada intento, podrá avanzar a la siguiente sesión que contendrá un ejercicio diferente y continuar con el proceso.

La realimentación se dará a los estudiantes de manera automática al finalizar la sesión de cada ejercicio. La aplicación le entregara al estudiante un puntaje en las componentes rítmica y melódica donde se le entregara la siguiente información:

- Errores melódicos graves en rojo, aceptables en amarillo y notas correctas en verde.
- Errores rítmicos graves en rojo, aceptables en amarillo y notas correctas en verde.
- Comentarios automáticos generados por el sistema

Para realizar las pruebas se dispondrá de la colaboración de 10 estudiantes del programa de música de la Universidad de los Andes a quienes se les enviará de manera virtual un consentimiento de tratamiento de datos.

RESULTADOS

Diseño de aplicación

Tras realizar el proceso de adquisición del software en internet e instalación, lo primero que el usuario encuentra al ejecutar la aplicación, es un sencillo menú donde este podrá encontrar el logo de la aplicación y un par de botones desde los cuales podrá cargar un archivo de trabajo e iniciar con su ejercicio.

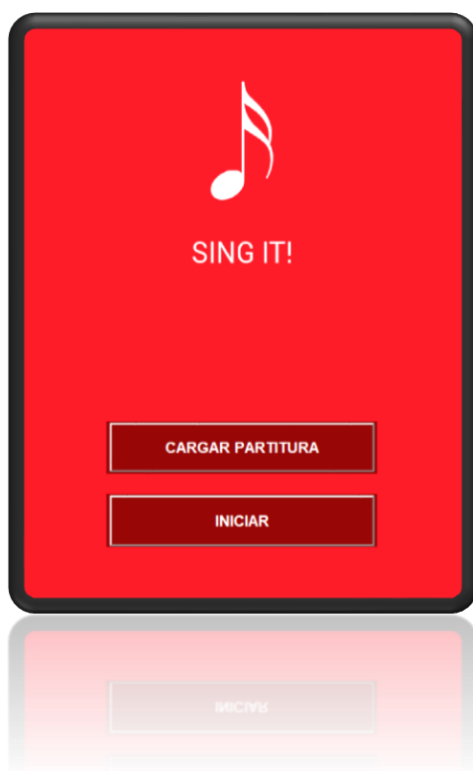


Figura. 18. Ventana que aparece al iniciar la interfaz.

Los archivos que el usuario podrá utilizar para estudiar están contenidos en el directorio de descarga, y son especiales puesto que contienen nombres específicos con algunas etiquetas usadas para decodificar información referente a la métrica de la partitura y relacionar otros archivos necesarios para la interpretación de datos y generación de señales. Las etiquetas incluyen al final del nombre un número de dos dígitos que representa el patrón rítmico que está directamente relacionado con la métrica musical que tiene relación con el tiempo e indica la cantidad de latidos que hay por cada división vertical en el pentagrama musical.



Figura. 19. Ejemplo de patrones rítmicos típicos usados en notación musical.

Estas etiquetas le indican al sistema el patrón auditivo que debe marcarse auditivamente y que servirá como guía para el correcto desarrollo de una sesión de solfeo por parte del músico. Una correcta sincronización de la generación de las marcas de sonido y el tiempo de adquisición de señales fue necesario para el correcto funcionamiento de los análisis realizados por el sistema.



Figura. 20. Ventana de análisis de voz de la aplicación "SingIt"

En la ventana que se genera cuando oprimimos iniciar en el menú principal, podemos observar 4 paneles situados en los extremos y en el centro. Hacia el lado izquierdo encontramos las herramientas que permitirán al usuario ajustar la tonalidad de la partitura MIDI que viene definida de acuerdo con el origen del archivo para que esta vaya de acuerdo con su tipo de voz. En el ámbito científico, se han logrado identificar diferentes modos de

oscilación en la voz llamados registros vocales y son diferentes en hombres y en mujeres[19].

Hombres	Mujeres
Pulse o vocal fry 148-174Hz	Modal o de pecho 293-349Hz
Modal o de pecho 293-349Hz	Falsetto 587-659Hz
Loft o Falsetto 493Hz	Cabeza 987Hz

Tabla. 3. Tipos de registro en la voz cantada.

De acuerdo con el tipo de voz de cada persona es necesario realizar modificaciones a la tonalidad de las partituras, para ello, el usuario deberá escoger la combinación de tonos y semitonos deseados que se ajusten al tipo de voz y finalmente pulsar aplicar. Para verificar si el ajuste es correcto, el usuario también tiene la opción de escuchar la tonalidad.

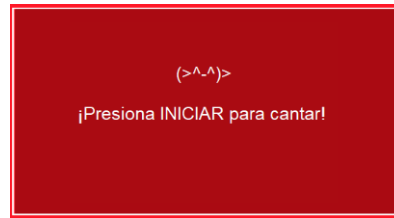
En el panel de la derecha encontraremos el botón de resultados el cual nos conducirá hacia otra ventana una vez se haya concluido el ejercicio, encontramos el botón de guardado que le permite al usuario almacenar la sesión después de realizar el ejercicio y finalmente está el botón de regreso, que retorna al usuario al menú principal donde puede seleccionar un ejercicio diferente si lo desea.

La parte inferior de la ventana contiene el panel donde se mostrará al usuario la partitura correspondiente al ejercicio que está llevando a cabo con información sobre el compás y la clave.

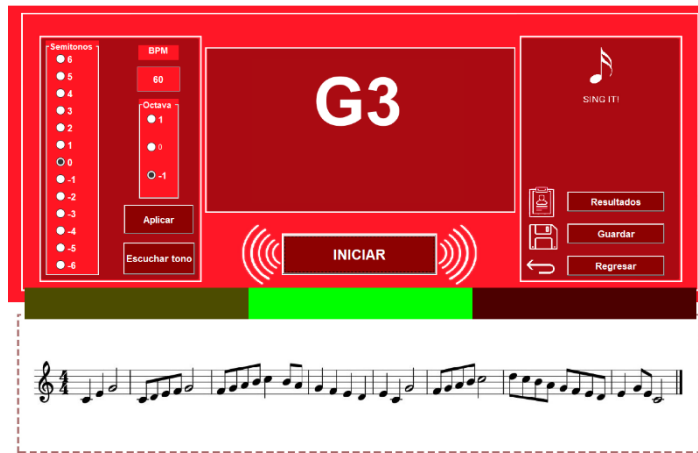
Finalmente se encontró el panel del medio que contiene el botón de inicio y una pantalla interactiva donde el usuario podrá recibir disantos mensajes dependiendo de las acciones que este tome. Esto se realizó con el fin de guiar al usuario a través de la interfaz y sea el mismo quien logre ejecutarla, navegar y generar progresos. En esta pantalla también se entrega información sobre la nota que está siendo ejecutada y un panel auxiliar que se despliega durante la ejecución entrega la información en tiempo real sobre si las notas que se están interpretando cumplen las condiciones frecuenciales establecidas por la información de la partitura suministrada por el archivo MIDI y por las ventanas permitidas de error de acuerdo con la literatura.



(a)



(b)



(c)

Figura. 21. Mensajes de usuario al presionar opciones incorrectas, (a) al presionar resultados sin haber realizado el ejercicio, (b) al presionar guardar sin haber realizado el ejercicio, (c) realimentación en tiempo real que señala que el usuario que se está realizando la ejecución correcta de la nota Sol 3.

Para finalizar, los datos recopilados en la ventana de análisis se almacenan en variables temporales y se espera a que el usuario las guarde de forma manual para que pueda llevar así un registro de sus sesiones previas y posteriormente, proceda al siguiente menú donde podrá generar la visualización de los resultados obtenidos.

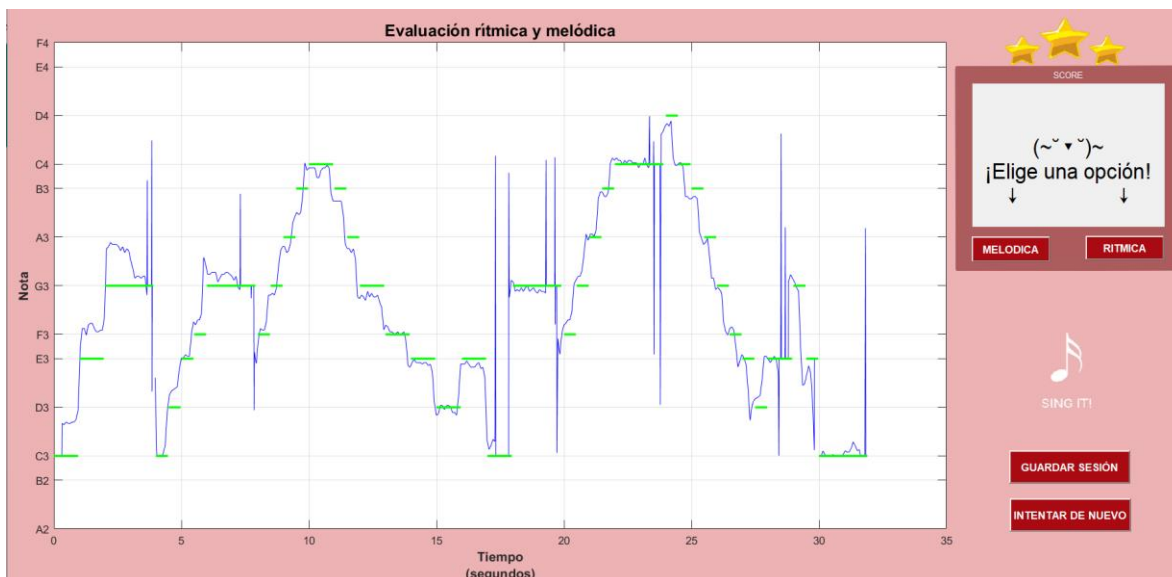


Figura. 22. Ventana generada tras oprimir la opción “Resultados” en la ventana de análisis. Se muestran dos graficas superpuestas que corresponden al seguimiento de la frecuencia fundamental(azul) y al resultado esperado obtenido a partir de la información MIDI.

En esta última ventana se generará toda la información que ayudará al usuario a reconocer si existen aspectos en su interpretación por mejorar, encontramos también una pequeña consola que de manera didáctica guía al usuario para que el mismo explore su desempeño en la componente que este elija. Los criterios de análisis que se tuvieron para cuantificar el desempeño de la persona se discutirán a detalle en la sección de decisión y realimentación.

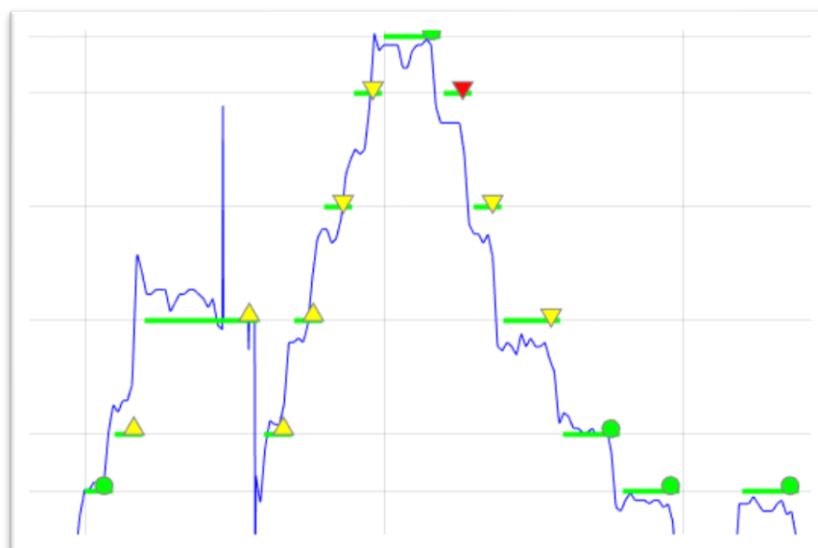


Figura. 23. Porción de la gráfica tras selecciona explorar la componente melódica.

La realimentación de la componente melódica se muestra si el usuario oprime el botón indicado, este le entrega de vuelta una serie de marcas con colores que especifican lo siguiente:

- **Verde:** Significa que se interpretó correctamente la nota
- **Amarillo:** Significa que la nota fue interpretada de manera aceptable pero que podría ser mejor. Esta marca puede aparecer en forma de triángulo apuntando hacia arriba o hacia abajo, lo cual señala además hacia que frecuencias o notas en la escala musical tiende a desplazarse su frecuencia fundamental durante dicha nota.
- **Rojo:** Significa que el resultado durante la interpretación de esta nota se aleja demasiado del objetivo puesto que la frecuencia fundamental registrada en ese instante se acerca más a una nota distinta a la deseada. Al igual que en las marcas amarillas, aparecen direcciones señalando que el error se asemeja más a una nota más alta en la escala o más baja.

Dichas marcas tienen como objetivo ayudar a interpretar las gráficas que pueden resultar difíciles de entender en un principio y lograr que el usuario pueda recibir una información profunda sobre su desempeño y así poder alcanzar el objetivo planteado.

Adicionalmente se entrega el número de notas acertadas sobre el total de notas y un valor numérico que refleja su desempeño.

Finalmente, si el usuario pulsa el botón indicado para la componente rítmica, este recibirá la misma información que en la componente melódica, pero basándose en criterios que están enfocados en medir los tiempos de interpretación de las notas. Las marcas generadas aquí respetan el mismo código de colores, pero se basan en el análisis del tiempo así:

- **Verde:** Significa que se interpretó la nota a tiempo
- **Amarillo:** Significa que la nota fue interpretada en un rango aceptable pero que podría ser mejor. Esta marca puede aparecer en forma de triángulo apuntando hacia la izquierda o hacia la derecha, lo cual señala si el error fue haber iniciado la nota antes de tiempo o demasiado tarde respectivamente.
- **Rojo:** Significa que la interpretación de esta nota se realizó demasiado pronto o demasiado tarde de lo esperado. Al igual que en las marcas amarillas, aparecen direcciones señalando el tipo de error.

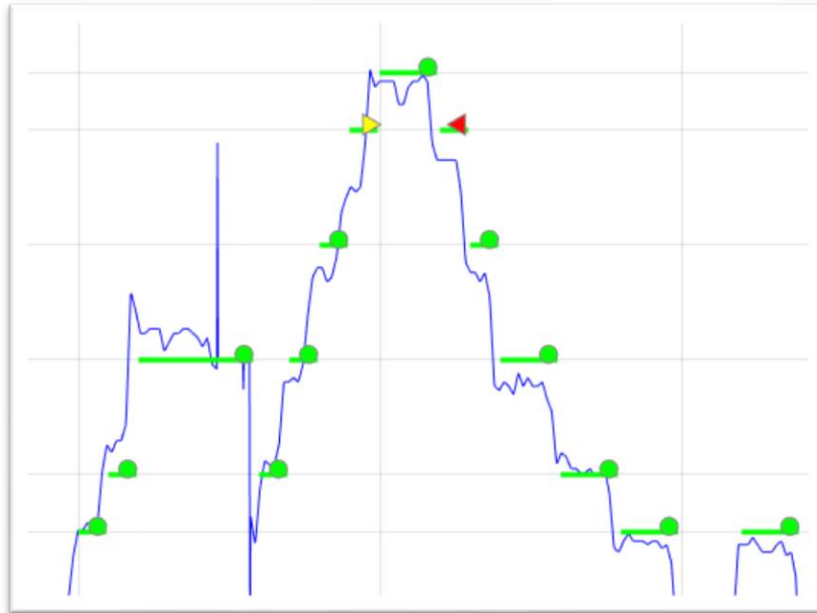


Figura. 24. Porción de la gráfica tras seleccionar explorar la componente rítmica

Pruebas con sujetos y evidencias

Definida la cantidad de voluntarios que utilizarían la aplicación, fue posible evidenciar el comportamiento esperado por parte de los usuarios al interactuar con el software, pues ellos lograron realizar diferentes ejercicios propuestos de manera satisfactoria. Los resultados sugieren que a pesar de que existen diferencias entre los tipos de voz que manejan los voluntarios, estos lograron configurar la herramienta para que esta se adapte a las características más adecuadas para ellos. Adicionalmente, se logró evidenciar diferencias significativas entre aquellos con mayor experticia en el solfeo respecto a aquellos que poseen un nivel básico en el tema. Otra observación realizada fue la presencia de artefactos debido a que algunos no siguieron la orden de utilizar audífonos causando que el metrónomo que lleva el tempo durante el ejercicio se registrara en la señal causando serias implicaciones en el desempeño del software.

A continuación, realizaremos la comparación de los resultados obtenidos en los diferentes ejercicios durante la interpretación de las partituras, se mostrarán resultados de 4 sujetos diferentes por cada prueba para un total de 8 participantes.

Ejercicio 1: Acorde subdominante

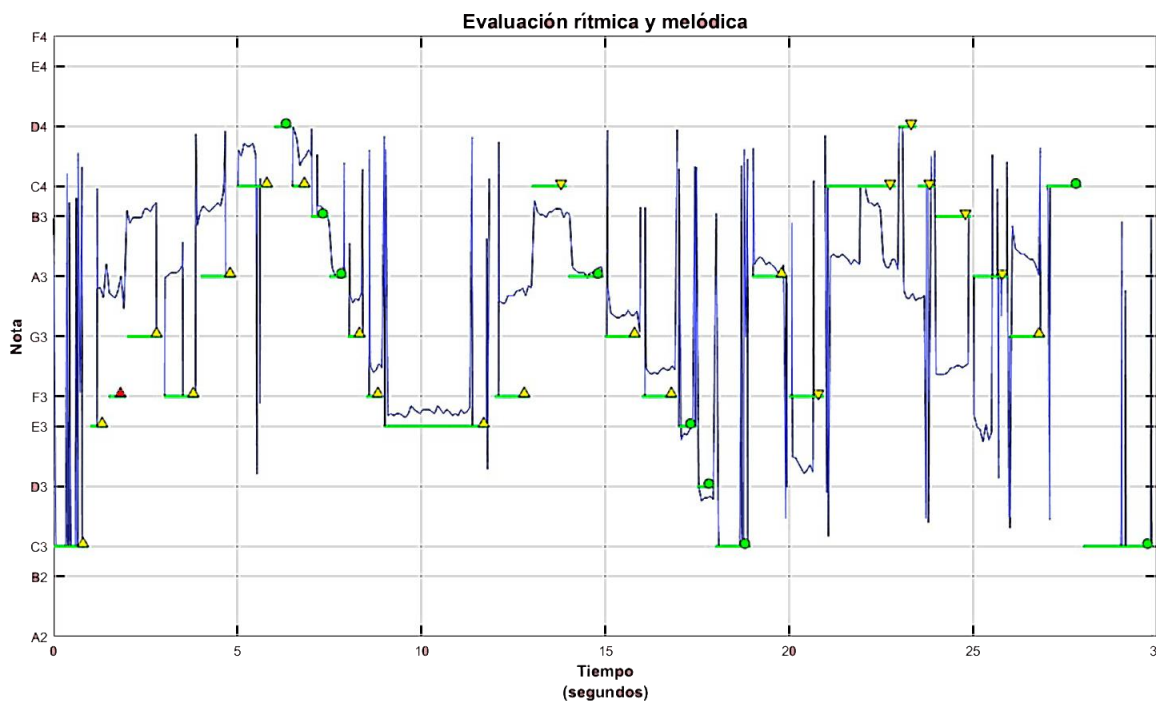


Figura. 25. Resultados ejercicio “acorde subdominante” del participante 1

En la figura anterior, observamos los resultados del ejercicio de uno de los voluntarios que afirmó haber realizado su ejercicio bajo condiciones de ruido, sin audífonos y además mencionó no ser muy diestro en el solfeo. A pesar de la presencia de ruido, el sistema es capaz de identificar la precisión melódica determinada por las alturas de las notas en verde. Las marcas de acierto también señalan correctamente si el error fue por debajo o por encima de lo esperado y nos permiten, además, evidenciar que efectivamente este participante no logra alcanzar algunas notas, pero que está muy cerca de lograrlo. El sistema arroja para este estudiante en la componente melódica una calificación de 1.33 de 5. Esto tiene sentido puesto que, aunque el estudiante logra entonar cerca al objetivo en algunas ocasiones, en otras está lejos de las notas y es justamente la evaluación por medio del error acumulado en cada nota, lo que ayuda al sistema a calificar de manera precisa de acuerdo con el desempeño logrado.

La cantidad de artefactos en esta señal causan errores en la evaluación de la componente rítmica ya que el ruido afecta principalmente los momentos en que el estudiante no está cantando y aparece en forma de líneas horizontales al principio y al final de cada nota.

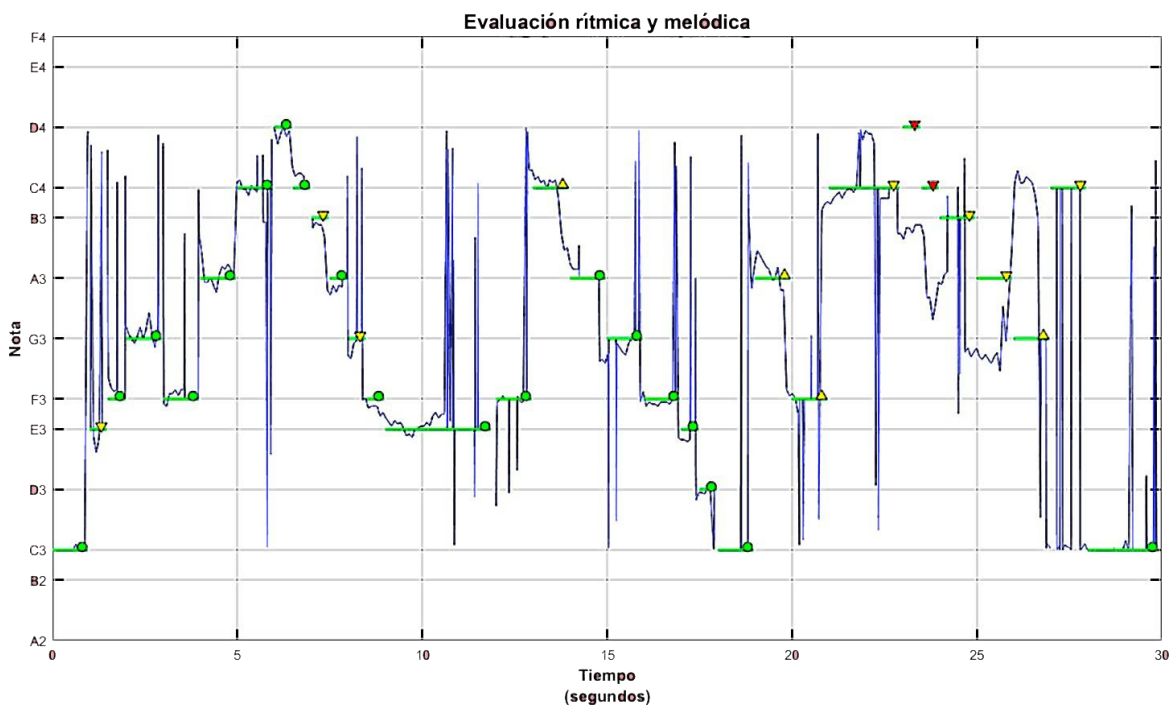


Figura. 26. Resultados del participante número 2

El participante numero dos muestra un buen desempeño durante cierta parte del ejercicio y eventualmente se perciben diferentes artefactos entre los cuales podemos identificar un error durante la ejecución del ejercicio que se presenta en el segundo 22 donde podemos apreciar que el estudiante leyó la mitad de una nota y prosiguió el ejercicio sin caer en cuenta de ello provocando así, un desfase de un tiempo en las notas siguientes. Se observa que el desempeño del estudiante fue bueno antes de dicho error y que el sistema es capaz de calificar correctamente su desempeño en la componente melódica. La calificación arrojada por el sistema en la componente melódica fue de 2.97 sobre 5 puesto que, aunque acierta bastantes notas, en varias otras aún puede mejorar mucho. Por otra parte, en la componente rítmica obtuvo un puntaje de 4.38 de 5. Podemos concluir que, si el estudiante no se hubiese confundido al final, podría haber logrado una calificación excelente en la componente rítmica y que, por otra parte, aún puede mejorar la estabilidad de sus notas.

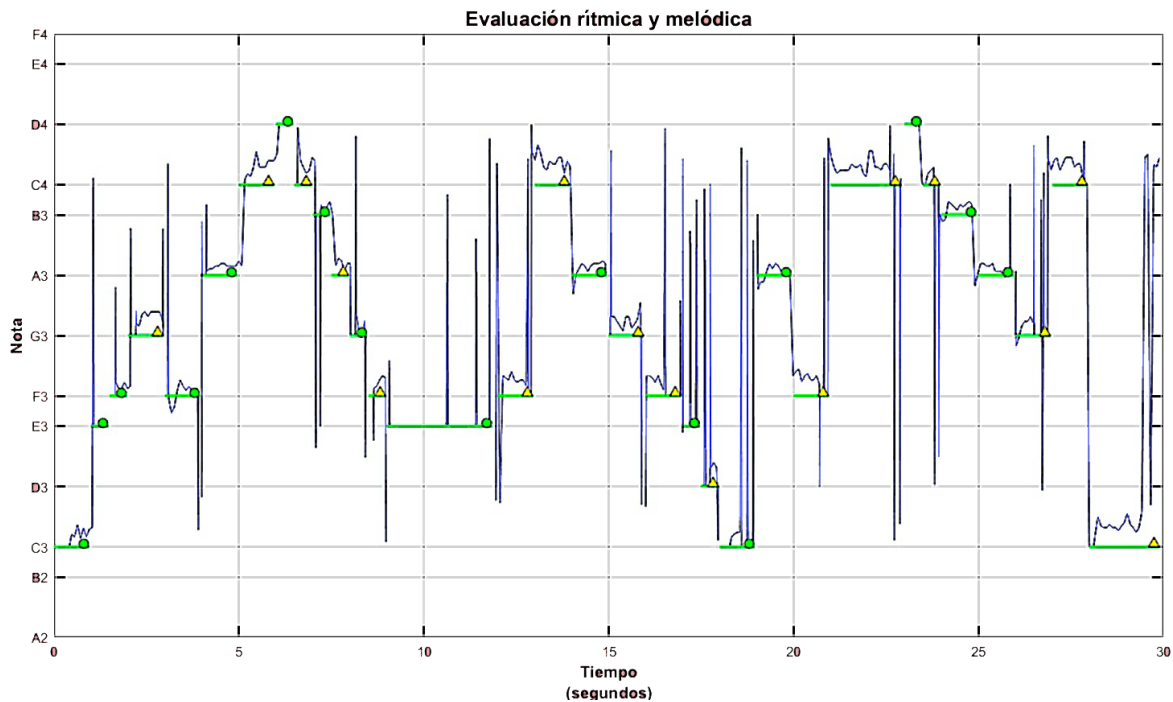


Figura. 27. Resultados del participante número 3

Se observa en la anterior figura que el participante número 3 tiene un mayor dominio de su voz pues es capaz de mantener las alturas de las notas constantes cerca de los umbrales definidos y también muestra una buena precisión rítmica. Podría pensarse que aunque la señal presenta menos ruido y se encuentra muy cerca de las alturas esperadas, este debería obtener una calificación mayor, sin embargo, podemos observar como el sistema califica correctamente la precisión melódica de cada una de las notas donde refleja que el estudiante puede mejorar aún más su desempeño si logra disminuir la tonalidad un poco en cada nota, puesto que, en promedio, sus notas se encuentran ligeramente más arriba de lo esperado, pero jamás comete errores graves.

En la componente melódica, el sistema cuantifica los aciertos y en base a estos genera una calificación de 2.5 sobre 5, la calificación en la componente rítmica fue de 5 sobre 5 y visualmente se puede validar al observar que, aunque el estudiante no tuvo una afinación perfecta, sus tiempos de ataque fueron muy precisos. Esto nos permite concluir que este estudiante posee una sólida habilidad rítmica y buen control de la estabilidad de su voz, con oportunidad de mejora en la percepción de las alturas.

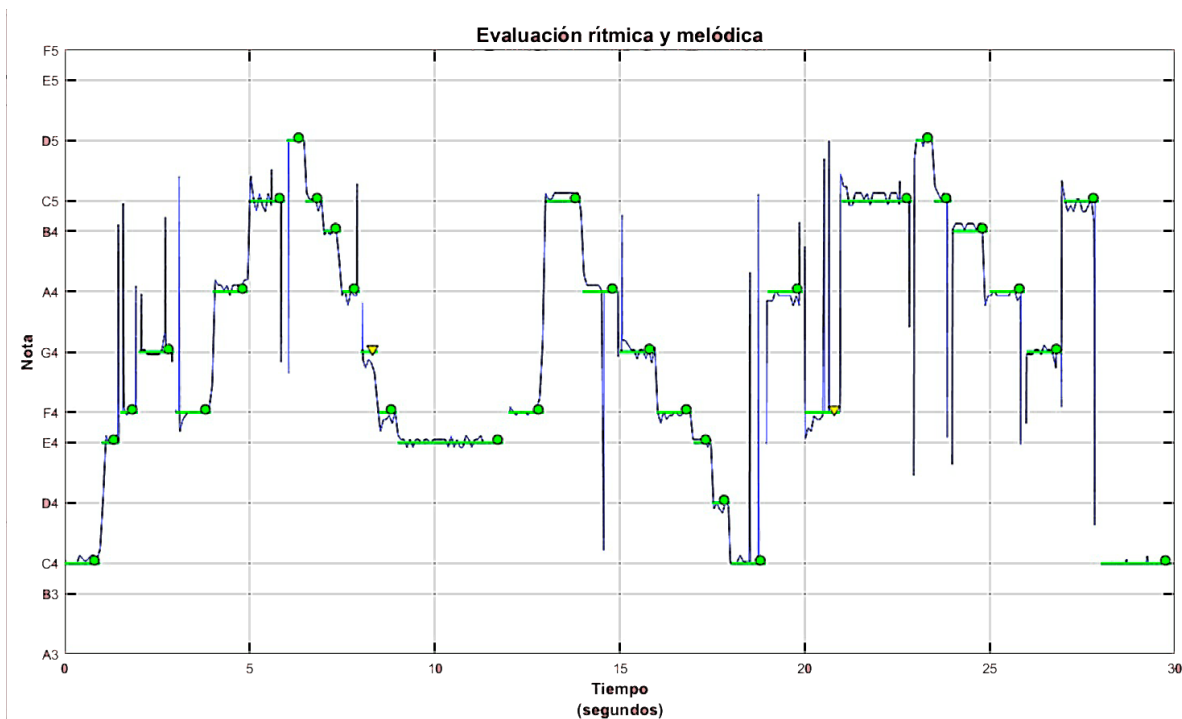


Figura. 28. Resultados del participante número 4

Por su parte, el participante número 4, estudiante de música de último semestre y con bastante experiencia en el solfeo logro obtener un desempeño excelente. Se puede visualizar a simple vista como el estudiante logra unos tiempos de ataque perfectos y también logra las alturas de las notas que además mantiene casi a la perfección durante el tiempo requerido. Este caso es de particular interés puesto que permite visualizar el alcance de detalle de la herramienta. Podemos observar que solo hay 2 notas que no son perfectas, pero si aceptables. Podríamos decir que, dada la corta duración de las notas, resulta para el estudiante un mayor grado de exigencia el poder entonar las alturas correctas durante un muy corto periodo. Este tipo de detalles resultan muy difíciles de percibir y mucho más aún de comunicar de manera efectiva a los estudiantes puesto que por la naturaleza de la fisiología humana, es casi imposible lograr este tipo de resolución y resultaría impensable poder generar resultados cuantificados que un estudiante pueda tomar como verdaderos sin la ayuda de una herramienta como esta.

La calificación generada por el sistema en la componente melódica fue de 4.69 sobre 5 y en la componente rítmica de 5 sobre 5.

Al estudiar estos resultados del primer ejercicio, fue posible observar que existen comportamientos repetitivos en la ejecución de la lectura musical, por ejemplo, el participante 3 lograba sostener notas con habilidad, pero no demostraba la precisión melódica del participante 4. En promedio sus errores se encontraron por encima del rango establecido. De igual modo, el participante 2 mostró una tendencia mayor a cantar por debajo del rango establecido. Esto hace pensar que se pueden incluir más modelos de análisis que ofrezcan la posibilidad de generar información automática en aspectos más

profundos mediante el análisis de estos patrones. En futuras versiones del software se podrían incluir análisis estadísticos, se discutirá esto en el siguiente capítulo.

Ejercicio 1: Acorde dominante

Con base en los resultados obtenidos anteriormente, se realizó una revisión de la etapa de filtrado con el fin de mejorar los resultados obtenidos en entornos de ruido y se reclutó a otros 4 participantes que realizaron comentarios positivos respecto al aplicativo. A continuación, podemos observar los resultados:

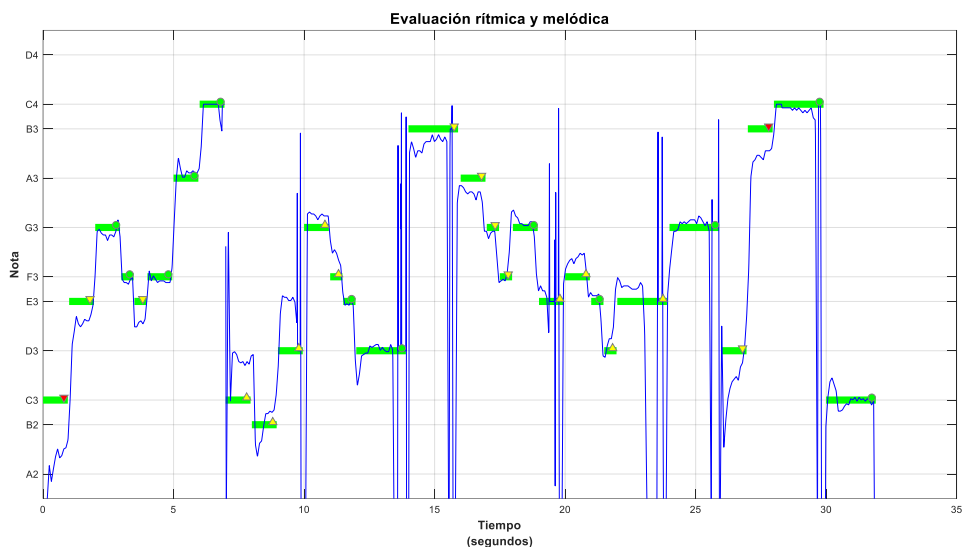


Figura. 29. Resultados del participante número 5, se observa la realimentación de la componente melódica.

Observamos que el anterior estudiante tuvo un buen desempeño pero que sin embargo está bastante lejos de una técnica adecuada debido a la inestabilidad de las notas. Los resultados cuantitativos obtenidos para dicha sesión fueron las siguientes:

- Melódica: acertó 12 de 30 notas y el resultado de dicha componente fue de 2 sobre 5.
- Rítmica: acertó 23 de 30 notas y una calificación de 3.8 sobre 5

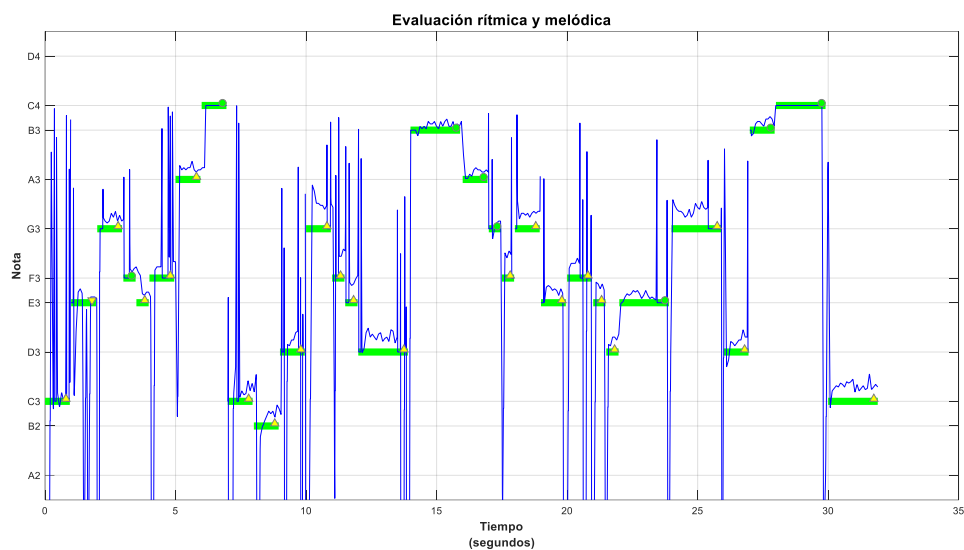


Figura. 30. Resultados del participante número 6, se observa la realimentación de la componente melódica.

A pesar de que se recomienda el uso de audífonos, existen otros factores que pueden afectar el desempeño del software como el zapateo durante la interpretación de la pieza que se usa de manera involuntaria por algunos usuarios para seguir el ritmo. Los resultados para este participante se muestran a continuación:

- Melódica: acertó 8 de 30 notas y el resultado de dicha componente fue de 1.33 sobre 5.
- Rítmica: acertó 28 de 30 notas y una calificación de 4.67 sobre 5

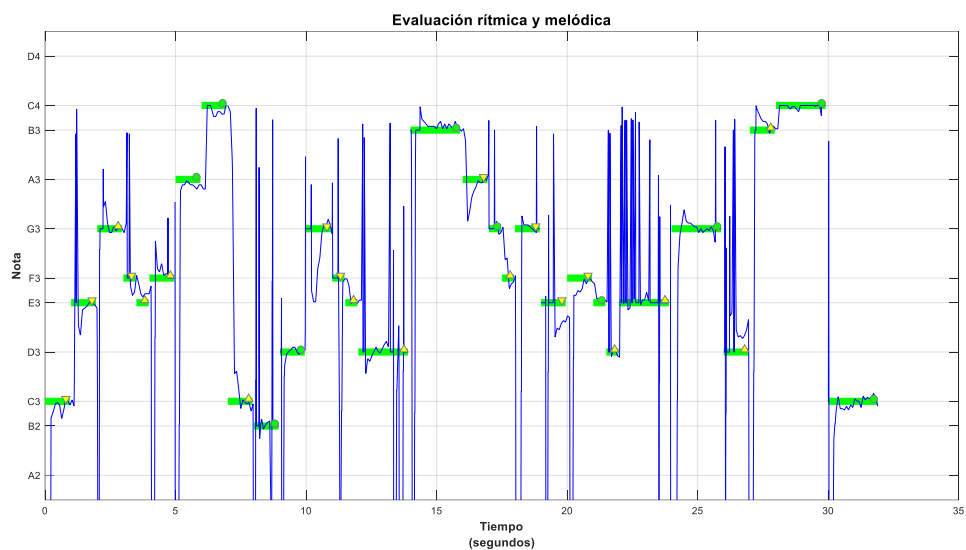


Figura. 31. Resultados del participante número 6, se observa la realimentación de la componente melódica.

Los resultados para este estudiante fueron los siguientes:

- Melódica: acertó 9 de 30 notas y el resultado de dicha componente fue de 1.67 sobre 5.
- Rítmica: acertó 26 de 30 notas y una calificación de 4.33 sobre 5

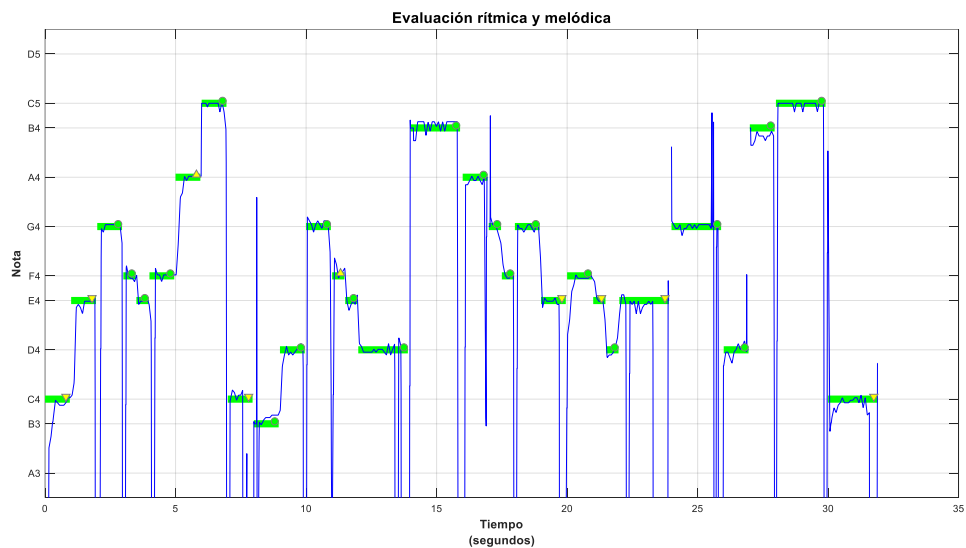


Figura. 32. Resultados del participante número 6, se observa la realimentación de la componente melódica.

Al igual que en el ejercicio anterior, se pueden observar muchas diferencias entre estudiantes con mayor experiencia en el solfeo. En particular también podemos encontrar una gran diferencia al realizar los ejercicios en un entorno silencioso respecto a otros estudiantes quienes realizaron el ejercicio en lugares con ruido ambiental como voces de fondo. La calificación para esta estudiante se puede apreciar a continuación

- Melódica: acertó 21 de 30 notas y el resultado de dicha componente fue de 3.5 sobre 5.
- Rítmica: acertó 29 de 30 notas y una calificación de 4.83 sobre 5

De este modo podemos evidenciar las diferencias en las habilidades de los participantes durante el solfeo y que las abstracciones cuantitativas tienen relación con los resultados observados.

DISCUSIÓN

Se realizó una revisión de aplicativos similares con el fin de determinar si existen trabajos que realicen un análisis que se asemeje al realizado en este proyecto.

Dentro de las aplicaciones que se exploraron para realizar una comparación, se tuvieron en cuenta principalmente a aquellas son más populares y con mejores calificaciones, así mismo también buscamos herramientas computacionales que cumplieran la características de comparación deseadas como evaluar al estudiante desde puntos de vista rítmicos, melódicos y que ofrezcan una realimentación en tiempo real. A continuación, encontramos las aplicaciones y software encontrados.

- **Music trainer:** Esta aplicación tiene como objetivo principal, ayudar a mejorar habilidades de lectura de solfeo, cuenta con 3 niveles de dificultad, permite trabajar en clave de sol y en clave de fa, escuchar los sonidos de las partituras y su plataforma soporta el español, inglés y alemán. El aprendizaje se basa en el trabajo que realiza el usuario al seleccionar las notas que van apareciendo en pantalla en el mismo sentido en que se leen en una partitura y selecciona entre múltiples opciones la correcta permitiéndole así reconocer los símbolos y también los sonidos que se van emitiendo a medida que avanza el ejercicio.
- **SolfaRead:** Es una aplicación enfocada especialmente a pianistas y permite seleccionar ejercicios en clave de sol, clave de fa y simultáneos. Posee niveles de dificultad diferentes y su entrenamiento se basa en la lectura adelantada donde el estudiante escoge la nota siguiente de forma manual entre diferentes opciones. Esta aplicación genera un score con valores relacionados a número de notas leídas en determinado tiempo y la precisión en la lectura.
- **DoReMiNotas:** Es una aplicación diseñada para leer partituras y que permite al usuario personalizar sus ejercicios en parámetros como las notas que desea entrenar, el tiempo que desea manejar, entre otras características musicales. Los

resultados no pueden guardarse, pero pueden ser enviados vía email para que un maestro los pueda revisar.

- **DoSolFaLite:** Una aplicación bastante completa que tiene 40 niveles de dificultad y tiene un diseño similar a un juego, posee una interfaz simple e intuitiva y también funciona mediante calificación de resultados de una lectura adelantada por selección de las notas que van apareciendo. Las notas suenan a medida que vas avanzando y permite trabajar en clave de sol y de fa.
- **Solfeador:** Es la aplicación con más popularidad y mejor calificación entre las antes mencionadas, posee una interfaz intuitiva y sencilla de manejar, permite trabajar en clave de sol y clave de fa, mismo funcionamiento de lectura adelantada manual, la notación musical puede modificarse de acuerdo a lo que el usuario prefiera, las notas suenan a medida que el usuario avanza en el ejercicio y ofrece resultados como promedio de notas por minuto o tiempo de acierto de las notas de manera cuantitativa.

Por otra parte, se realizó una búsqueda más profunda de aplicaciones con origen basado en investigaciones científicas y la aplicación encontrada fue Pentagram. Esta aplicación es una herramienta muy útil durante la lectura de partituras puesto que permite al usuario mostrar la nota que está tocando, como se llama y como se escribe en un pentagrama musical en tiempo real. Dentro de las características de personalización, encontramos la posibilidad de cambiar la clave de la partitura, la octava, semitonos, diferentes sonidos, grabar el audio de la sesión y permite reproducir partituras MIDI. Igual que las anteriores aplicaciones mencionadas, esta tampoco entrega una realimentación basada en un análisis profundo de la partitura a trabajar puesto que es más bien una herramienta de apoyo al aprendizaje que de monitoreo de este.

Los rasgos más llamativos que encontramos en las aplicaciones discutidas fue que un factor importante para su popularidad se basa en sus interfaces simples y fáciles de usar las cuales cuentan con diseños llamativos y distintas funciones personalizables. Sin embargo, muchas de las sugerencias que se realizan para poder mejorar en componentes melódicos y rítmicos específicamente son el uso de afinadores cromáticos para poder monitorizar que las notas sean acertadas al momento de interpretarse y el uso del metrónomo durante la práctica del solfeo.

Basándonos en esta investigación, se realizó la siguiente tabla de comparación en la que se comparan aplicaciones que prometen mejorar las habilidades de lectura de solfeo. Los productos por comparar son una aplicación cualquiera y la que tiene mejor puntuación de la playstore, Pentagram la aplicación con estudios de respaldo y SingIt.

Características	Singlt	Pentagrom	Solfeador	DoSolFaLite
Niveles de dificultad	5	NO	3	VARIOS
Idiomas	ESP	NO	VARIOS	NO
Soporta archivos MIDI	SI	SI	NO	NO
Historial	MAIL	NO	NO	NO
Cifrado musical	AMERICANO	EUROPEO	VARIADO	EUROPEO
Puntaje	MELODICO/RITMICO	NO	NºNOTAS/TIEMPO	NºNOTAS/TIEMPO
Realimentación	DURANTE*/POS	DURANTE	DURANTE	DURANTE
Soporta Partituras	SI	NO	NO	NO
Feedback auditivo	SI	SI	SI	SI

*Durante: Realimentación basada en análisis de voz y no en selección de notas manual.

Tabla. 4. Comparación entre diferentes aplicaciones disponibles en el mercado y la desarrollada en este proyecto.

Otra característica importante para el desarrollo de la aplicación fue la selección de una paleta de colores agradables que junto con una interfaz gráfica simple e intuitiva permita que los estudiantes de música que la utilicen se sientan cómodos y lleguen a recomendarla.

RECOMENDACIONES Y TRABAJOS FUTUROS

La proyección a corto plazo de este proyecto es el de encontrar un lugar en las aulas de música donde se pueda utilizar por estudiantes y profesores para fomentar un correcto desarrollo en el aprendizaje siguiendo una metodología definida. que permita además a los estudiantes realizar estudios de manera independiente en sus casas.

Se espera añadir más adelante otras herramientas al sistema tales como la opción de modificar el tempo al que se desea realizar la lectura, habilitar la opción de cargar resultados para poder visualizar sesiones previas y la capacidad de realizar la ejecución del ejercicio y mostrar los resultados en la misma ventana. Adicionalmente se podría implementar la opción que permita exportar resultados en diferentes formatos de imágenes que muestren los resultados de las componentes rítmica y melódica, así como sus respectiva realimentación numérica. Esto le daría la libertad al usuario de poder compartir resultados con otras personas que no utilicen la aplicación. Esto podría fomentar la difusión del uso entre los estudiantes.

Sería muy interesante permitir al estudiante manipular la señal resultante permitiendo realizar acercamientos en áreas de interés y poder observar información de la partitura proveniente de los metadatos extraídos de los archivos MIDI asociados tales como señalar la nota en la pieza grafica de la partitura, y mostrar información como el porcentaje de acierto de cada nota que puede ser fácilmente extraído del algoritmo diseñado.

El proyecto no termina allí, pues a mediano plazo pueden implementarse mejoras progresivas al software que ayuden a incrementar su popularidad junto con la implementación y desarrollo de un aplicativo para dispositivos móviles en diferentes sistemas operativos con un costo asociado.

Finalmente, y teniendo en cuenta las posibilidades de mejora y la limitada cantidad de herramientas que existen de este tipo actualmente en el mercado, en un futuro se podría pensar en desarrollar una aplicación muy competitiva que sea líder en análisis musical y que ofrezca diferentes herramientas para profesores y estudiantes, que sirva como canal entre los mismos permitiendo análisis e interpretación veloz y que pueda ser coordinada simultáneamente por profesor y alumno. Esto abre el camino a desarrollar un modelo de negocio que permita generar ingresos a docentes que quieran ofrecer servicios online de clases apoyadas con esta herramienta. Se podría pensar en aplicaciones que manejan diferentes módulos que comparten información, como por ejemplo en el transporte público existe un usuario que utiliza una aplicación para ser el conductor y otro que la usa como usuario de un vehículo. En este orden de ideas se ofrecerían servicios complementarios como proponer ejercicios basados en la experiencia del maestro que conlleven más rápidamente al aprendizaje del solfeo.

Encontrar una herramienta para uso conjunto de los estudiantes y profesores de cursos de solfeo, que les permitiera extraer información precisa sobre la afinación y tiempos de ataque, motivó al desarrollo inicial de este proyecto. Se plantearon inicialmente objetivos enfocados al desarrollo de una plataforma que sería para uso presencial de los estudiantes en la universidad de los Andes, con posibilidades de llegar a convertirse en una aplicación

móvil en trabajos futuros. En este orden de ideas, tanto el desarrollo de pruebas como recopilación de datos para culminar la investigación, se realizaría en conjunto con los estudiantes. Sin embargo, la actual coyuntura obligó una modificación de objetivos en un punto ya avanzado de la investigación y las pruebas no se pudieron llevar a cabo en su totalidad en el escenario deseado. Esto afectó la situación ideal prevista para realizar las pruebas por lo que en su momento, la limitante más importante para el proyecto radicó en que a pesar de que los estudiantes de la Universidad de los Andes cuentan con licencia de Matlab en el campus universitario, les fue imposible el desplazamiento por consecuencia de la Enfermedad por Coronavirus 2019 (COVID-19) que se declaró por la OMS como una emergencia de salud global entre el 30 de enero y el 11 de marzo, periodo durante el cual se venían desarrollando pruebas piloto. La aplicación se adaptó y abrió las ventanas para pensar en generar una herramienta disponible para estudiantes en cualquier parte del mundo.

Basándonos en el hecho de que no existe una aplicación en el mercado que realice un análisis de ejercicios comunes entre los estudiantes de música, que aproveche las herramientas de la ingeniería biomédica y el procesamiento de señales para contribuir al arte y al desarrollo de la educación, se pensó en las siguientes ideas que pueden complementar el trabajo de esta investigación.

Como primera observación, para futuros trabajos, la aplicación se migraría a un lenguaje que sea adecuado para generar un aplicativo móvil que conserve el fundamento y las herramientas ya logradas, pero que incluya ciertas mejoras en la interacción con el usuario, el almacenamiento y recopilación de datos mediante conexión a la red e inclusión que posea más niveles de dificultad y ejercicios con un flujo basado en evidencia científica en el ámbito educativo desarrollado por profesionales en la materia.

Un plus que significaría el desarrollo de una herramienta más especializada sería el de aprovechar características como el vibrato que permiten extraer aún más información. El vibrato en el canto presenta estabilidad en la frecuencia oscilatoria y esta refleja el estilo y la habilidad de un cantante. La literatura sugiere que diferentes estilos de vibrato dan a la voz una percepción de belleza y calidad[20]. Una continuación de este proyecto también permitiría analizar la amplitud, duración y frecuencia de estas fluctuaciones con el fin de entregar valores cuantificados de dichas características.

En la interfaz se implementarían algunas mejoras como la de permitir que el estudiante modifique la casilla de texto que muestra el tempo para que la ejecución pueda ser evaluada a diferentes velocidades añadiendo esto una herramienta más a las existentes. Esto también podría ser logrado mediante un botón que pueda calcular el tempo deseado a partir de medir la cantidad de veces que se oprime el botón en una cantidad de tiempo. De igual modo, se modificaría la barra que indica la precisión de la afinación en la ventana de monitoreo aumentando su resolución y poder así detectar con más precisión la frecuencia visualmente.

Un entorno de desarrollo especializado como Unity permitiría incrementar aún más la fluidez de procesos que requieren llevar a cabo diferentes tareas, tal es el caso de la implementación de una barra que guíe al estudiante a lo largo de la partitura a medida que esta se reproduce. Se puede pensar también en implementar más modos en los que se permita analizar diferentes instrumentos musicales además de la voz de manera individual o conjunta pues se puede adaptar la herramienta auditiva para recibir información en diferentes anchos de banda mediante modificaciones en los filtros para ampliar así el espectro de frecuencias que pueden ser analizados selectivamente.

Se pueden realizar análisis detallados de ejecuciones de otros instrumentos musicales en piezas musicales cada vez más avanzadas para usuarios mediante un trabajo más dedicado en análisis y síntesis de partituras a partir del archivo MIDI que requería de un estudio más profundo del tema.

CONCLUSIONES

Este proyecto nació debido a las dificultades que poseen muchos profesores de música al dar una realimentación a los estudiantes puesto que muchas veces estos no están de acuerdo con los comentarios del maestro. Esta herramienta va más allá de permitir realizar un análisis profundo del estudio del solfeo puesto que no solo resulta útil como herramienta para que docentes puedan obtener números basados en el desempeño de sus estudiantes, sino que también puedan apoyarse de los resultados gráficos para ofrecer así una realimentación no solo basada en la experiencia sino también en resultados numéricos y gráficos obtenidos a partir de técnicas de ingeniería permitiendo así, una mayor confianza y aceptación por parte de los estudiantes a la hora de recibir comentarios. Debido a esto podemos concluir que los resultados obtenidos son positivos puesto que se logró generar una herramienta audiovisual para el estudio del solfeo dirigida a estudiantes de un programa de música de educación superior que permite guiar al usuario hacia el mejoramiento de sus habilidades melódicas y rítmicas de manera remota.

REFERENCIAS

- [1] J. Salas, "ESTRATEGIAS DE APRENDIZAJE EN EL ESTUDIO DEL SOLFEO ENTONADO EN EL CURSO DE LENGUAJE Y PERCEPCIÓN MUSICAL," vol. 33, p. 21, 2018.
- [2] P. Gomez and E. Belmonte, "Biomechanical Evaluation of the Singing Voice," Dec. 2013.
- [3] D. Howard, H. Daffern, and J. Brereton, "Quantitative voice quality analyses of a soprano singing early music in three different performance styles," *Biomed. Signal Process. Control - BIOMED SIGNAL PROCESS CONTROL*, vol. 7, Jan. 2012, doi: 10.1016/j.bspc.2011.05.013.
- [4] "(PDF) MatLab vs. Python vs. R," *ResearchGate*.
https://www.researchgate.net/publication/328175547_MatLab_vs_Python_vs_R (accessed Apr. 26, 2020).
- [5] "Language Fundamentals - MATLAB & Simulink."
<https://www.mathworks.com/help/matlab/language-fundamentals.html> (accessed Apr. 21, 2020).
- [6] K. Schutte, *kts/matlab-midi*. 2020.
- [7] "Data Acquisition Toolbox User's Guide." .
- [8] E. J. Humphrey *et al.*, "An Introduction to Signal Processing for Singing-Voice Analysis: High Notes in the Effort to Automate the Understanding of Vocals in Music," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 82–94, Jan. 2019, doi: 10.1109/MSP.2018.2875133.
- [9] M. A. Talhaoui, "Real-time Data Stream Processing - Challenges and Perspectives," *Int. J. Comput. Sci. Issues*, vol. 14, Aug. 2018, doi: 10.20943/01201705.612.
- [10] R. Rensink, "The Modeling and Control of Visual Perception," in *Integrated Models of Cognitive Systems*, 2007, pp. 132–148.
- [11] T. Saitou, M. Unoki, and M. Akagi, "Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis," *Speech Commun.*, vol. 46, pp. 405–417, Jul. 2005, doi: 10.1016/j.specom.2005.01.010.
- [12] T. Nakano, M. Goto, and Y. Hiraga, "An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features," p. 4, 2006.
- [13] D. Moratal, "Biomedical Signal and Image Processing, 2nd Edition [Book Reviews]," *Pulse IEEE*, vol. 5, pp. 76–76, Mar. 2014, doi: 10.1109/MPUL.2013.2296811.
- [14] L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 25, no. 1, pp. 24–33, Feb. 1977, doi: 10.1109/TASSP.1977.1162905.
- [15] H. de Oliveira and R. Oliveira, "Understanding MIDI: A Painless Tutorial on Midi Format," May 2017.
- [16] J. McAuley, "Tempo and Rhythm," in *Music Perception*, 2010, pp. 165–199.
- [17] S. W. Lee and M. Dong, "Singing Voice Synthesis: Singer-Dependent Vibrato Modeling and Coherent Processing of Spectral Envelope.," presented at the Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, Jan. 2011, pp. 2001–2004.
- [18] T. Saitou and M. Goto, "Acoustic and perceptual effects of vocal training in amateur male singing.," Jan. 2009, pp. 832–835.
- [19] M. I. Uzcanga, S. Fernandez, M. Marqués, L. Sarrasqueta, and R. Urrutia, "Voz cantada.," *Rev. Med. Vol 50 N° 3 2006 Pags 49-55*, vol. 50, Jan. 2006.
- [20] "(PDF) Vibrato training in singers," *ResearchGate*.
https://www.researchgate.net/publication/262463866_Vibrato_training_in_singers (accessed May 02, 2020).