

Informe técnico CCADET-UNAM.

1 de junio de 2011.

Evocanto: programa de cómputo para analizar la voz cantada mediante técnicas de procesamiento digital de señales

Mtra. Gisela Gracida Olvera,
<gisgracida@msn.com>
Programa de Maestría y Doctorado en Música,
Universidad Nacional Autónoma de México.

Dr. Felipe Orduña Bustamante,
<felipe.orduna@ccadet.unam.mx>
Centro de Ciencias Aplicadas y desarrollo Tecnológico,
Universidad Nacional Autónoma de México.

Evocanto: programa de cómputo para analizar la voz cantada mediante técnicas de procesamiento digital de señales

Mtra. Gisela Gracida Olvera,
<gisgracida@msn.com>
Programa de Maestría y Doctorado en Música,
Universidad Nacional Autónoma de México.

Dr. Felipe Orduña Bustamante,
<felipe.orduna@ccadet.unam.mx>
Centro de Ciencias Aplicadas y desarrollo Tecnológico,
Universidad Nacional Autónoma de México.

Índice.

Resumen.....	2
1. Introducción.....	2
2. Parámetros del canto en relación con el análisis acústico de la voz cantada.....	3
3. Análisis de requerimientos.....	5
4. Diseño y desarrollo del programa.....	6
5. Pruebas técnicas y ejemplo de aplicación.....	8
6. Conclusiones.....	12
Bibliografía.....	12

Resumen

Se presenta el desarrollo de un programa de cómputo multiplataforma, con una interfaz gráfica amigable, orientado a estudiantes y maestros de canto como auxiliar didáctico en la enseñanza de la técnica vocal. El programa capta la señal de voz mediante un micrófono y una interfaz de audio y la somete a procedimientos de análisis espectral, mediante los cuales se determinan los elementos necesarios para estimar el funcionamiento del tracto vocal, así como características tímbricas de la voz cantada, tales como la entonación, el vibrato y la resonancia. El programa muestra representaciones gráficas del espectro y su análisis, lo que permite realizar una valoración objetiva. La interacción en tiempo real hace que el usuario pueda experimentar adecuaciones de la técnica vocal y observar inmediatamente los efectos manifestados en la señal acústica, por lo que se obtiene información adicional a las indicaciones técnicas y musicales del maestro de canto.

1. Introducción

Aunado al desarrollo de procedimientos para evaluar la calidad de la voz, surge la necesidad de proporcionar instrumentos auxiliares del análisis, que sean cada vez más accesibles a los diferentes tipos de demandas de usuarios. Es en la última década cuando ocurre un rápido incremento en la oferta de programas para estos fines, de los que sobresale la característica de utilizar las capacidades estándar de las computadoras personales y portátiles, sin tener que adquirir aparatos más especializados.

Cabe mencionar que la industria informática del canto ha sido una fuente inagotable de aplicaciones y programas que utilizan los principios de la investigación técnica y científica sobre la voz, pero que están hechos con una finalidad comercial y de entretenimiento, tales como diversos tipos de sistemas de "karaoke" y transformadores de distintas características vocales, que cambian el timbre o la entonación.

Sin embargo, han sido pocos los trabajos enfocados a la educación en el canto, por lo que el software que se presenta aquí es único por sus características, donde se logra un punto de conexión del análisis digital integral de la voz cantada, para su aplicación en la profesionalización del canto.

Modelo físico y análisis acústico de la señal de voz

La señal de voz es una onda acústica de presión sonora que se origina a partir de movimientos voluntarios del sistema físico de producción de la voz, que está conformado por los órganos anatómicos, cuya fisiología corresponde a los procesos de respiración, fonación y resonancia [Bunch, 1997].

La producción de la voz ha sido integrada en un modelo físico, ampliamente aceptado por la comunidad científica: la denominada *teoría de la fuente y el filtro*. Aunque su origen es anterior, el desarrollo matemático se debe al sueco Gunnar Fant (1960) [Rabiner, 1978]. La materia prima acústica se crea en la *fente*, o punto donde se genera el sonido básico que luego habrá de ser filtrado. Esta fuente se sitúa en las cuerdas vocales si se trata de sonidos sonoros como las vocales y consonantes sonoras. En el caso de las consonantes sordas, la fuente está en el punto bucal donde se produce el ruido consonántico.

Durante los periodos estables o estacionarios de los sonidos de tipo vocal, la onda generada repite el mismo ciclo a intervalos regulares de tiempo, por lo que es una onda periódica y cumple el Teorema de Fourier: está formada por una serie infinita de armónicos [Fant, 1997]. El primer armónico es la frecuencia fundamental F_0 , y corresponde al tono de la voz; el segundo tiene una frecuencia doble de la fundamental, el tercero, triple, y así sucesivamente.

El sonido generado en la fuente se propaga a través del tracto vocal y, de acuerdo con los principios de la *física de tubos* [Stanley, 1958], sufre un proceso complejo de *filtrado*, donde unas frecuencias se refuerzan y otras se atenúan. El tracto vocal es el área que inicia desde la parte superior de la laringe y termina en los labios. Los órganos articulatorios, que son los labios, la mandíbula, la lengua y el velo del paladar, permiten concentrar la energía en determinadas frecuencias, actuando como resonadores. Estas frecuencias que se potencian o atenúan, reciben el nombre de formantes. Las frecuencias formantes más significativas del tracto vocal, en cuanto a la acústica de la voz hablada y cantada, son las primeras cinco (F_1 , F_2 , F_3 , F_4 y F_5). Dependiendo de la relación que se establezca entre las diversas formantes, es que se percibe un determinado timbre de voz o calidad vocal. La frecuencia F_0 , que corresponde a la frecuencia fundamental de la voz, no se asocia con una formante, ya que no se origina como una resonancia del tracto vocal [Stanley, 1958].

Las dos primeras formantes (F_1 y F_2) permiten la identificación de las vocales. Las formantes agudas (F_3 , F_4 , F_5) determinan el color de la voz.

En las vocales, la primera formante (F_1) controla la amplitud del sonido y depende de la forma de la cavidad faríngea: cuanto más estrecha, mayor frecuencia y viceversa. La segunda formante (F_2) controla la inteligibilidad del sonido y se supedita a la posición de la lengua: si se eleva en la parte anterior, la frecuencia subirá en relación directa con la altura y la anterioridad alcanzadas; si es en la parte posterior, descenderá en relación inversa con la altura. La tercera formante (F_3) está relacionada con la acción de los labios. Su valor en frecuencia es más alto si éstos están estirados y más bajo si están redondeados.

Las formantes F_4 y F_5 varían con la anchura y longitud del tracto vocal; cuanto más corto y estrecho el tracto, más agudas estas formantes.

2. Parámetros del canto en relación con el análisis acústico de la voz cantada

Frecuencia fundamental

La frecuencia fundamental F_0 corresponde a la frecuencia en la cual vibran las cuerdas vocales. Las propiedades de longitud, masa y elasticidad de las cuerdas vocales, son controladas por la musculatura laríngea. El acortamiento de la cuerda vocal provoca la disminución de la tensión y aumenta la masa de la superficie vibrátil, por lo que la frecuencia fundamental será disminuida. La elongación de la cuerda vocal produce un aumento en la frecuencia fundamental, al tensar la cuerda vocal y disminuir la masa y la superficie de contacto.

La fuente de voz humana puede operar en diferentes modos de oscilación, que son los llamados *registros* vocales. De acuerdo a Titze (2002), estos registros dependen de una interacción entre las resonancias subglotales y las vibraciones de las cuerdas vocales. Según Johan Sundberg (1987), la definición más común de registro es: "un rango de frecuencias fonadas en el cual todos los tonos son percibidos como producidos de una forma semejante y con un timbre semejante". Varios autores como Van den Berg (1968), Large (1973) y Hollien (1971) describen la existencia de tres o cuatro registros básicos o principales: el **glottal fry** (registro de frote glotal o registro de pulso); el registro **modal** que incluye los denominados "registro medio" y "de cabeza"; el **falsete** y un cuarto registro, el registro **flauta** o silbido, en la zona de agudos.

Rigurosamente, un cantante bajo es capaz de cantar aproximadamente a 80 Hz (tono E₂) y hasta 330 Hz (E₄), un tenor entre 123 Hz (C₃) y 520 Hz (C₅), una contralto entre 175 Hz (F₃) y 700 Hz (F₅), y una soprano entre 260 Hz (C₄) y 1300 Hz (E₆) [Sundberg, 1987].

Formante del cantante

Un cantante bien entrenado puede lograr conformar su tracto vocal para hacer que las formantes F3, F4 y F5 formen un grupo aglutinado estrechamente en frecuencia. A esta agrupación de formantes se le llama "formante del cantante". Como las separaciones de frecuencia entre ellas están así disminuidas, sus niveles individuales incrementan y de ahí se obtiene un alto pico espectral entre 2500 y 3000 Hz. Esto es útil al cantar con acompañamiento de una orquesta.

Por un lado, la formante del cantante se debe a un factor perceptual, ya que el oído humano goza de especial percepción a estas frecuencias y, por otro lado, destacan por encima del enorme sonido de una orquesta, pues la mayoría de los instrumentos no las generan de manera intensa. La existencia de una distancia constante entre armónicos, permite que el oído humano *reconstruya* las frecuencias enmascaradas y *perciba* la frecuencia fundamental.

La frecuencia central de la formante del cantante varía ligeramente entre las diferentes clasificaciones de voz. El centro de frecuencia de la formante del cantante de bajos, barítonos y tenores está alrededor de 2.4, 2.6 y 2.8 kHz, respectivamente. Estas pequeñas diferencias son bastante relevantes para los timbres típicos de voz de estas tesituras. Su origen es debido probablemente a diferencias en la longitud del tracto vocal; los bajos tienden a tener un tracto vocal más largo que los barítonos, quienes a su vez tienen un tracto vocal más largo que los tenores [Sundberg, 1987].

Es difícil que la formante del cantante aparezca en registros agudos de voz, como las sopranos. Esto se explica por el hecho de que cantan en una alta F0, tienen parciales extensamente espaciados en el espectro y el grupo de F3, F4 y F5, que es el que produce una formante del cantante, es bastante estrecho en frecuencia. Por lo tanto, la formante del cantante es una característica común de cantantes de registros graves (bajos, barítonos, tenores, incluso contraltos y mezzosopranos [Benade, 1990]) y entrenados profesionalmente.

Sintonía formante - armónico

En el canto femenino de registro agudo, como es el caso de las voces de sopranos, se puede obtener un equivalente de formante del cantante, al lograr la llamada "voz resonante" que se obtiene abriendo mucho la mandíbula de forma que se sintoniza F1 o F2 con F0. Posteriormente, cada uno de los demás armónicos actúa como una formante, es decir, ocurre la denominada *sintonización de formantes* con armónicos. Eso explica que la inteligibilidad de las vocales sea menor a altas F0, como en el caso de la soprano, pues va variando la disposición de la lengua y la mandíbula para poder acercar F1 a F0. El resto de formantes sintoniza con los armónicos de la fuente; pero ya no guardan las mismas relaciones que en el habla. Este fenómeno puede ser no privativo de las sopranos, pues es probable que sea una tendencia natural de los cantantes profesionales el sintonizar las frecuencias de los formantes con las frecuencias de los armónicos [Sundberg, 1987], [Benade, 1990].

Vibrato

La modulación (variación cíclica sistemática de un parámetro) de frecuencia y/o amplitud de la voz del cantante, constituye el *vibrato*. El mecanismo primario es la modulación de la frecuencia fundamental y secundariamente se ve afectada la amplitud. La frecuencia fundamental y sus armónicos varían cíclicamente, mientras que las formantes no experimentan cambios. La variación

sincrónica con F0, hace que los armónicos se acerquen o alejen de las formantes, viéndose más o menos reforzados, de forma que la modulación en amplitud sobreviene pasivamente cuando coinciden con las zonas de resonancia de la función de transferencia del tracto vocal [Seashore, 1932].

Cuando existe únicamente modulación de amplitud, en el canto se le denomina *trémolo*. Es considerado un defecto, pero ocurre en raros casos. Si la modulación de la frecuencia fundamental es producida únicamente en la glotis, el vibrato tendrá una frecuencia entre 6.5 y 8.0 Hz, que en inglés se denomina *bleat*. Es común que sea producido por cantantes sin entrenamiento o por cantantes jóvenes en los primeros años de su formación. También es posible que el vibrato se genere desde el diafragma (por falta de control o tono muscular) y que tenga una frecuencia menor a 5 Hz. Este tipo de vibrato frecuentemente afecta la claridad del canto. Es uno de los defectos en voces envejecidas. En inglés, se denomina *wobble*. El sistema vocal de un cantante bien entrenado, posee una tendencia natural de producir un vibrato cuya frecuencia se encuentra entre 5 y 7 Hz [Sundberg, 2000], [Titze et al., 2002].

3. Análisis de requerimientos

El programa *Evocanto* está dirigido principalmente a estudiantes de canto profesional y a profesores de la misma área que deseen incorporarlo en sus clases o en su metodología de enseñanza. El objetivo más importante del programa *Evocanto* es realizar un análisis comprensible de la voz cantada, siendo posible evaluar ciertas cualidades vocales. Para lo cual debe ser fácilmente manipulable por el usuario y servir como herramienta complementaria en el estudio y comprensión de la técnica vocal.

Una vez activado el funcionamiento del programa, y con el uso de un micrófono para captar la señal de voz, el usuario cantará una nota, con una vocal, sosteniéndola durante algunos segundos. Al mismo tiempo, en el programa *Evocanto* serán llevadas a cabo los siguientes procesos:

- Análisis acústico de la voz cantada, que consiste en la obtención del espectro de la señal de voz y, a partir de él, la detección de armónicos y la envolvente espectral; la determinación de la frecuencia fundamental F0 y de las frecuencias formantes F1, F2,...; así como la amplitud o nivel de energía
- Verificación de la existencia de formante del cantante y del grado de sintonización de formantes con armónicos.
- Evaluación de entonación y vibrato, proporcionando la información correspondiente a la nota musical cantada y a la cantidad o frecuencia de vibrato.
- Representación del tracto vocal, mediante una imagen.

Es deseable que el manejo de *Evocanto* sea cómodo e interesante. Esto se logra al existir interactividad entre el usuario y el programa, por medio de la interfaz gráfica [Gándara, 1994]. Por ello, *Evocanto*, además de poder ser usado con facilidad operacional, proporcionará una gran interactividad visual, para que, así, el usuario pueda verificar lo que está ocurriendo con su voz en el momento de cantar. Esto proporcionará al usuario la facultad de experimentar cambios en la postura corporal, en la colocación de la voz, los movimientos de labios y mandíbula; así como ajustes en la entonación y el vibrato de la nota cantada.

Herramientas de programación

De acuerdo con los requerimientos del programa, para desarrollarlo, surgió la necesidad de encontrar un lenguaje de programación flexible y eficiente, en el cual se posibilite el manejo de señales de audio y análisis espectral, y que cuente con herramientas para desarrollo de la interfaz gráfica. De esa manera, fueron seleccionadas las siguientes bibliotecas multiplataforma de código libre, desarrolladas en C++:

- SndObj, para la adquisición de audio y análisis espectral, y
- WxWidgets, para la interfaz gráfica

Sin embargo se optó por no utilizar directamente el lenguaje C++, sino el lenguaje Python [Van Rossum, 2003]. Porque ofrece más ventajas y se puede utilizar de modo interactivo, lo que facilita hacer experimentos durante el desarrollo del programa. Además de que posee una sintaxis clara y sencilla.

Debido a las características de dichas herramientas, *Evocanto* posee las siguientes ventajas:

- Es multiplataforma: puede ejecutarse bajo cualquier sistema operativo de uso común.
- Presenta una apariencia visual nativa y ofrece un alto rendimiento gráfico.
- Combina eficiencia con sencillez de uso.

4. Diseño y desarrollo del programa.

El programa *Evocanto* está conformado por dos procesos principales: digitalización de la señal de audio y análisis espectral de la señal digitalizada. Más adelante se ilustran los dos procesos en forma gráfica.

Conformación de la señal de entrada.

La configuración típica de la entrada de audio consta de un micrófono y una tarjeta de sonido. Puede ser utilizada la propia tarjeta de la computadora y un micrófono convencional para computadora. Sin embargo, esta configuración tiene el inconveniente de captar ruido proveniente de la misma computadora, que interfiere en el buen funcionamiento del programa, por lo que es recomendable utilizar una interfaz de audio externa y, de esta manera, poder utilizar también un micrófono de mayor calidad, con mejor captación de voz.

Para digitalizar y obtener el espectro de la señal de entrada, *Evocanto* utiliza varias clases (objetos computacionales) de la biblioteca SndObj [Lazzarini, 2000]. Los objetos de la clase SndRTIO y SndIn, dan acceso a la interfaz de audio definida por omisión en el sistema operativo, con referencia específica a la señal de entrada de audio número 1. Un objeto de la clase HammingWindow se utiliza para definir una ventana de ponderación que se aplicará a cada bloque de muestras de la señal de entrada de audio. Finalmente, un objeto de la clase FFT implementa la *Transformada Rápida de Fourier*. Los objetos indicados previamente se enlazan en una secuencia o hilo de procesamiento, utilizando un objeto de la clase SndThread, e intervienen en la clase FFT para transformar cada bloque de muestras de la señal de tiempo, en un espectro de frecuencia.

La señal de entrada es determinada por el tamaño de bloque que se va tomando de la tarjeta de audio, luego es *ventaneada*, transformada por la FFT y escalada. Este proceso ocurre a intervalos regulares, determinados por el tamaño especificado para el tamaño del bloque de muestras de la entrada de audio. La salida proporciona un vector con datos de la FFT, conteniendo un par real-imaginario para cada punto de frecuencia sobre el lado positivo del espectro.

La máxima frecuencia que se muestra en el espectro de *Evocanto*, se estableció en 5000 hertz, debido a que este límite es suficiente para incluir las formantes superiores más relevantes en el análisis de la voz.

Cada componente del espectro está conformado por un número complejo representado por un punto con una coordenada x (parte real) y una coordenada y (parte imaginaria). Su correspondiente magnitud cuadrática, normalizada al tamaño de bloque, es calculada mediante la fórmula:

$$S^2 = \frac{x^2 + y^2}{N}$$

donde S es la magnitud y N el tamaño de bloque de la señal.

Esta magnitud cuadrática en unidades lineales; es decir, no logarítmicas como en el caso de los decibels, se utiliza posteriormente para calcular la frecuencia de vibrato. Se convierte a decibels (NPS: nivel de presión sonora), mediante la fórmula:

$$NPS = 10 \log_{10}(S^2)$$

para poder conformar el espectro en unidades logarítmicas (decibels). En esta aplicación (*Evocanto*) no es necesario que el NPS se calibre en unidades físicas de presión sonora.

Análisis espectral de la señal de voz cantada.

Una vez obtenido el espectro, se detectan los picos espectrales, los cuales corresponden a máximos en el rango de frecuencia y con un umbral de magnitud. Estos se utilizan para realizar una estimación de la frecuencia fundamental F_0 , de manera aproximada, identificando la serie

armónica que mejor explica los picos espectrales observados. A continuación se mejora esa estimación inicial calculando la frecuencia promedio, ponderada por las magnitudes espectrales. Es decir, se obtiene el centroide de frecuencia [Potter, 1950], con la siguiente fórmula:

$$F0 = \frac{\sum w_i f_i}{\sum w_i}$$

donde f_i es la frecuencia del i -ésimo componente; w_i es un factor que indica la cantidad de dB que el i -ésimo componente está debajo del máximo, y equivale a la relación S_i/S_0 , entre amplitudes, expresada en decibels.

La frecuencia fundamental calculada se obtiene en hertz y entonces se calcula su equivalencia en términos de notas musicales, así como la diferencia en cents c , en relación con el La_5 de 440 Hz, basando los cálculos en la siguiente fórmula:

$$c = 1200 \log_2 \left(\frac{F0}{440 \text{ Hz}} \right)$$

Para determinar la nota, el índice de octava y la desviación en cents, se procede como en el siguiente ejemplo:

Si $F0 = 600$ Hz, entonces $c = 537$ cents por arriba del La_5 de 440 Hz. Este valor en cents se expresa como un múltiplo de 100, en este caso 500 cents (cinco semitonos), y un residuo de 37 cents. Por lo tanto, la nota resultante está cinco semitonos arriba del La_5 , con una desviación de 37 cents. Esto corresponde a la nota Re_6 con una desviación de +37 cents. En este caso, el índice de octava se incrementa de 5 a 6.

Un procedimiento general que equivale al ejemplo mostrado, se encuentra programado en el código de *Evocanto*.

La frecuencia de vibrato de $F0$, es determinada utilizando un método indirecto, se toma en cuenta la anchura espectral como un valor aproximado del vibrato, sin distinguir modulaciones de amplitud o modulaciones de frecuencia. La anchura espectral se calcula mediante la fórmula [Cordourier, 2009]:

$$\Delta f = \sqrt{\frac{\sum_{n=N_1}^{N_2} [f_n - f_c]^2 \cdot S(f_n)}{\sum_{n=N_1}^{N_2} S(f_n)}}$$

Se ubican los armónicos, primero en forma aproximada, tomando múltiplos de la frecuencia fundamental $F0$. Posteriormente se mejora la aproximación, determinando los máximos picos espectrales más cercanos a estos múltiplos.

Para localizar los picos de las frecuencias formantes $F1$, $F2$, ..., se utiliza un proceso de detección de máximos y mínimos en las componentes armónicas del espectro. Se comparan progresivamente las amplitudes de los armónicos para determinar si la envolvente espectral sube o baja. En los puntos de máxima amplitud se localizan las formantes. Se estableció un valor mínimo de 200 Hz, para considerar una formante.

El centroide de frecuencias de armónicos para $F1$ (similarmente para $F2$, $F3$, etc.), se obtiene con la siguiente fórmula:

$$F1 = \frac{\sum_{i=-1}^{i=1} f_i \cdot S(f_i)}{\sum_{i=-1}^{i=1} S(f_i)}$$

f_i corresponde a la frecuencia y $S(f_i)$ a la magnitud, de cada armónico.

La sintonía de las formantes $F1$ y $F2$ con armónicos, se establece de la siguiente manera, para $F1$:

$$\Delta F1 = (F1 - F_i) / F0 * 100\%$$

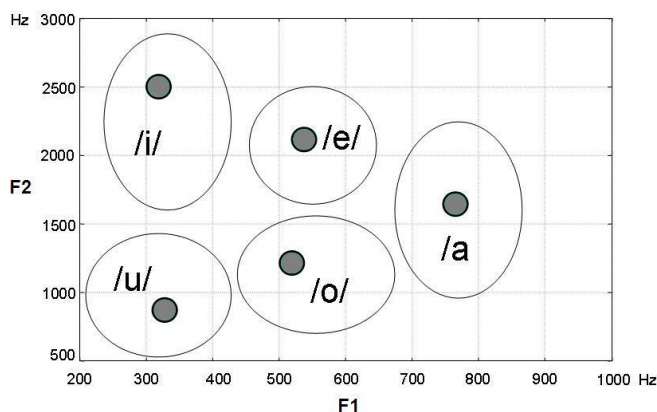
Con una fórmula similar para $F2$.

Es decir, equivale a la diferencia de frecuencias entre la formante correspondiente y el armónico alto F_i más cercano a $F1$. Se multiplica por 100% para calcular la sintonía formante-armónico

como el porcentaje de desviación o desajuste, entre -50% y +50%; los signos negativo (-) o positivo (+), indican, respectivamente, que la formante está debajo o encima del armónico.

El criterio utilizado para determinar si existe formante del cantante, es constituido por dos valores importantes. El primero es la frecuencia fundamental máxima, establecida en *Evocanto* en 300 Hz. El segundo es la tercera formante F3, que debe ser mayor que 2400 Hz.

Con base en los parámetros de las formantes F1 y F2, se establecen los criterios de articulación vocal. F1 indica el grado de apertura de la boca y F2 el grado de retracción del cuerpo de la lengua. En *Evocanto* se establecen las regiones vocálicas, de acuerdo a la siguiente gráfica de vocales para el idioma español:



Sin embargo, en *Evocanto*, la relación entre las formantes F1 y F2 es utilizada como una aproximación para determinar el desempeño del tracto vocal y no se indica la vocal correspondiente. Esto se debe a que intervienen diversos factores adicionales, incluso perceptuales, para distinguir y establecer con precisión cualquier vocal [Potter, 1950], [Ladefoged, 1989], [Neary, 1989].

5. Pruebas técnicas y ejemplo de aplicación

Se instaló *Evocanto* en una computadora tipo laptop, marca Toshiba, de procesador Intel Celeron, con sistema operativo Windows XP. También se utilizó una interfaz de audio marca Edirol, modelo UA-25, 24 bits 96kHz, conectada a la computadora a través del puerto USB; y un micrófono direccional dinámico super-cardioide AKG Acoustics modelo D880M (especial para canto) con conector XLR de tres polos, conectado a una entrada balanceada de la interfaz de audio, con un cable XLR balanceado.

Se comprobó el funcionamiento de *Evocanto*, al ser utilizado por diferentes cantantes. Para llevar a cabo las pruebas, se le pidió a cada cantante entonar al micrófono una nota del registro bajo, una del registro medio y una del registro agudo, con las cinco vocales, sosteniendo cada una por aproximadamente cinco segundos.

Después, se les pidió a los cantantes entonar algunas notas de distintos registros y con diferentes vocales, sosteniendo cada una por cinco segundos aproximadamente, sin vibrato y con vibrato exagerado.

A continuación se ilustran algunos de los resultados obtenidos de la voz de un cantante profesional con tesitura de barítono.

La figura 1 muestra la interfaz gráfica de *Evocanto* al analizar las cinco vocales, cantadas por el barítono. Se puede apreciar la diferencia en la envolvente espectral de cada vocal; en la distribución, magnitud y relación de las formantes, principalmente F1 y F2, que son las que determinan la configuración del tracto vocal, como se muestra en los esquemas correspondientes; y en las siguientes formantes, a partir de F3, se distingue la aparición de la formante del cantante. También presenta una medida de vibrato intermedia, propia de cantantes profesionales.

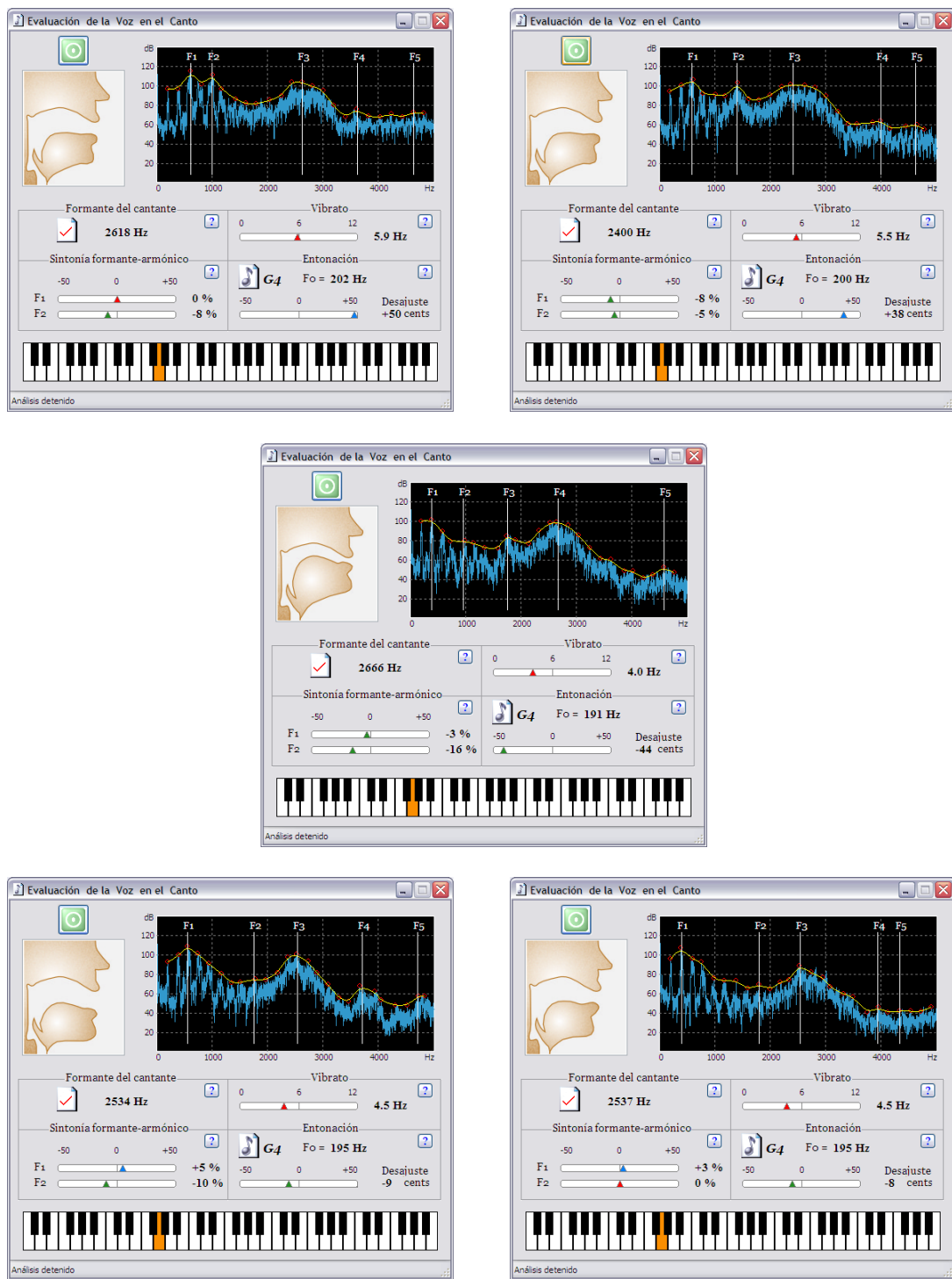


Figura 1. Vocales /a/, /e/, /i/, /o/, y /u/.

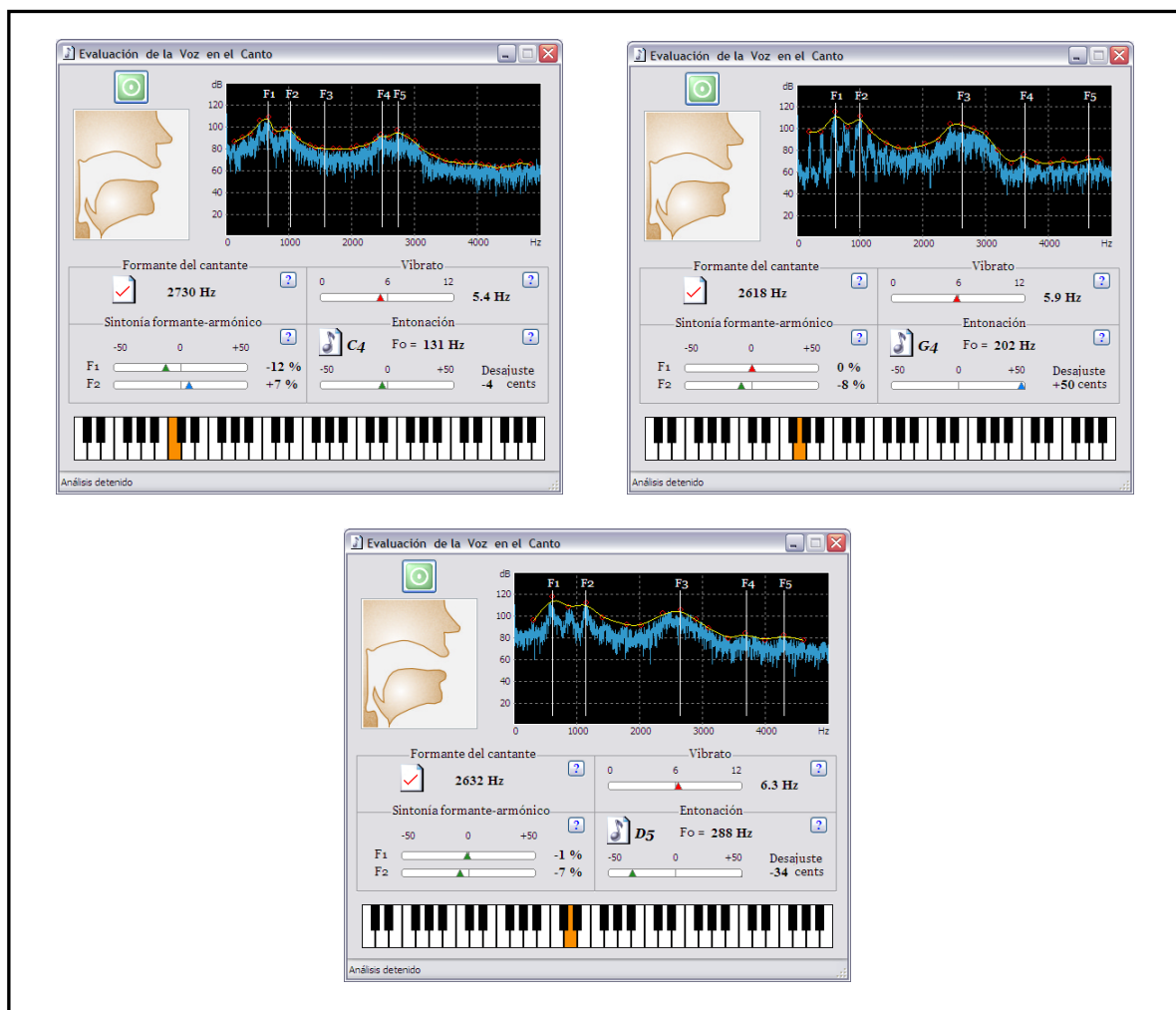


Figura 2. Vocal /a/ en los registros grave, medio y agudo.

En la figura 2 se muestran imágenes de Evocanto, que corresponden a la vocal /a/, cantada en los registros grave, medio y agudo del cantante barítono. Las diferencias determinadas en los espectros, muestran cambios en la amplitud; sin embargo la configuración del tracto vocal es semejante en los tres casos, por tratarse de la misma vocal. Se observa que, conforme se dirige la voz al registro agudo, se va extendiendo la distribución de los armónicos, es decir, se presentan más espaciados entre sí.

La figura 3 presenta imágenes de Evocanto, con la diferencia en la cantidad de vibrato de la vocal /o/, cantadas por el mismo barítono, en una nota del registro medio de la voz. Se aprecian cambios en el espectro, debido a esta diferencia. Cabe destacar que, para cantantes profesionales, con un vibrato natural bien desarrollado, se presenta cierta dificultad al intentar resaltar este tipo de diferencia entre poco y mucho vibrato. Cuando se intenta disminuir el vibrato, en este caso, disminuye también la energía o amplitud.

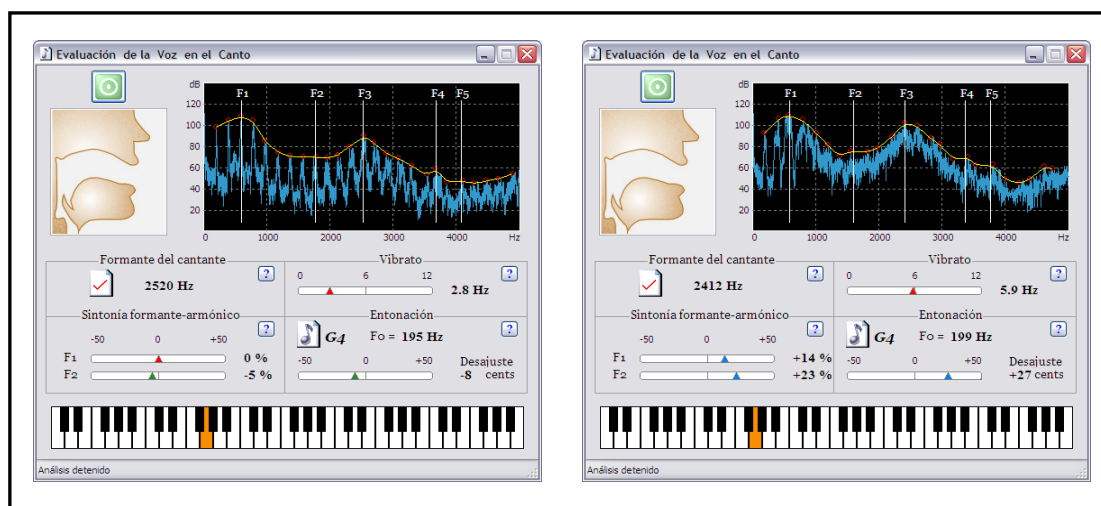


Figura 3. Diferencias en vibrato (poco y medio).

Pruebas de opinión

Adicionalmente, se elaboró un cuestionario para obtener la opinión de los usuarios de *Evocanto*, de acuerdo a los siguientes criterios de desempeño del programa:

1. Distribución de los elementos de la interfaz gráfica
2. Información que ofrece acerca de la voz cantada
3. Facilidad de uso
4. Tiempo de respuesta del análisis
5. Claridad de los conceptos relativos a la voz cantada
6. Evaluación como herramienta para el estudio del canto
7. Como herramienta para mejorar las habilidades vocales
8. Como herramienta para mejorar las habilidades auditivas
9. Como herramienta didáctica para enseñanza en grupo
10. Como ayuda para estudiar en forma independiente

Para cada uno de los criterios se presentó la siguiente escala de calificación:

Excelente = 4 puntos; Buena = 3 puntos; Suficiente = 2 puntos; Pobre = 1 punto; Mala = 0 puntos. El cuestionario se aplicó a distintos usuarios, cantantes profesionales, principiantes y avanzados, después de haber utilizado *Evocanto* en una sesión de capacitación y pruebas. La estadística de los resultados obtenidos arrojó que en opinión de los usuarios, el desempeño del programa en general, es bueno, como se observa en la siguiente tabla.

CRITERIO	CALIFICACIÓN PROMEDIO	PORCENTAJE DE ACEPTACIÓN
Distribución de los elementos de la interfaz gráfica	3.75	94%
Información que ofrece acerca de la voz cantada	3.75	94%
Facilidad de uso	3.75	94%
Tiempo de respuesta del análisis	3.25	81%
Claridad de los conceptos relativos a la voz cantada	2.62	66%
Evaluación como herramienta para el estudio del canto	3.62	90%
Como herramienta para mejorar las habilidades vocales	3.37	84%
Como herramienta para mejorar las habilidades auditivas	2.87	72%
Como herramienta didáctica para enseñanza en grupo	3.37	84%
Como ayuda para estudiar en forma independiente	3.62	90%

6. Conclusiones

Aportaciones

Una de las aportaciones más importantes de este trabajo, es la de proponer y materializar la idea de elaborar un programa de cómputo que analiza la voz cantada y que está dirigido a cantantes profesionales (alumnos y profesores) como una herramienta de trabajo para ayudar en el desarrollo de la técnica vocal.

A través del lenguaje de programación Python y las bibliotecas SndObj y WxWidgets, de código libre, se generó *Evocanto* como un programa multiplataforma, de interfaz gráfica amigable hacia el usuario. Mediante este programa el usuario puede observar lo que ocurre con su voz en el momento mismo de cantar, lo que permite orientarlo hacia un adecuado balance entre la funcionalidad de las estructuras anatómicas, particularmente del sistema vocal, y los aspectos estético-perceptivos implicados en los parámetros acústicos de la voz.

El análisis acústico de la voz cantada, que se realiza en el programa *Evocanto*, consiste en la obtención del espectro de la señal de voz para detectar elementos característicos, como son: armónicos, envolvente espectral, la frecuencia fundamental F0 y las frecuencias formantes F1, F2,... y la amplitud. Con estos elementos se logran indicar los siguientes parámetros de evaluación del análisis vocal:

- Verificación de la existencia de formante del cantante y del grado de sintonización de formantes con armónicos.
- Evaluación de entonación y vibrato, proporcionando la información correspondiente a la nota musical cantada y a la cantidad o frecuencia de vibrato.
- Noción del desempeño del tracto vocal, mediante una imagen.

Para lograr que *Evocanto* realice sus funciones con eficiencia y la mayor exactitud posible, fueron implementados algoritmos particulares para el análisis del espectro en la obtención de la frecuencia tonal F0, las frecuencias formantes F1, F2,... y la cantidad de vibrato. La implementación de dichos algoritmos es una aportación original de este trabajo. Asimismo, la forma en que se presentan los resultados en *Evocanto*, es una aportación original, que se realizó con la finalidad de mantener una coherente y funcional interfaz gráfica.

Se evaluó *Evocanto* con distintos tipos de voz y usuarios, aplicando pruebas de opinión, que dieron un resultado aceptable y satisfactorio acerca del desempeño del programa.

Sugerencias de trabajo a futuro

Las ideas que pueden llevarse a cabo en trabajos posteriores, basados en *Evocanto*, o como continuación de este trabajo, son:

- Utilizar *Evocanto* en un curso escolar de canto profesional o técnica vocal.
- Integración de ejercicios vocales de acuerdo a un método progresivo.
- Adaptar la interfaz gráfica del programa *Evocanto* de acuerdo al nivel de los usuarios cantantes (principiantes, intermedios, avanzados) y/o a las clasificaciones (tesituras) de voz.
- Mejorar los esquemas de configuración del tracto vocal.
- Perfeccionar los algoritmos de la detección de frecuencia tonal y de formantes.
- Determinación de la vocal que se está cantando.
- Crear e incorporar nuevos algoritmos para la medición de otros aspectos del vibrato y del timbre de la voz cantada.

Bibliografía

1. Bunch, Meribeth. *Dynamics of the singing voice*. Springer. Viena, 1997.
2. Cordourier, Héctor; Orduña, Felipe. *Active control of periodic fan noise in laptops: spectral width requirements in a delayed buffer implementation*. En: Journal of applied research and technology. Vol. 7. No. 2. August 2009. CCADET, UNAM. México.
3. Fant, Gunnar. *Acoustical Analysis of speech*. En: CROCKER, Malcom (editor). *Enciclopedia of Acoustics*. John Wiley, 1997.
4. Gándara, Manuel. *La interfaz en el usuario: una introducción para educadores*. En: Usos educativos de la computadora. CISE-UNAM. México, 1994.

5. Lazzarini, Victor. *The Sound Object Library*. Cambridge University Press. Cambridge, 2000.
6. Ladefoged, Peter. *A note on "Information conveyed by vowels"*. En: The Journal of the Acoustical Society of America, Vol. 85, no. 5 (pp. 2223 – 2224). May 1989.
7. Lindblom, Björn; Sundberg, Johan. *The human voice in speech and singing*. Cap. 16 en Rossing, Thomas D. (editor) *Springer handbook of acoustics*. Springer, 2007.
8. Nearey, Terrance M. *Static, dynamic, and relational properties in vowel perception*. En: The Journal of the Acoustical Society of America, Vol. 85, no. 5 (pp. 2088 – 2113) May 1989.
9. Potter, R; Steinberg, J. *Toward the Specification of Speech*. En: The Journal of the Acoustical Society of America, Vol. 22, no. 6. November 1950. Bell Telephone Laboratories. Murray Hill, New Jersey.
10. Rabiner, Lawrence; Schafer, Ronald. *Digital processing of speech signals* Prentice Hall. New Jersey, 1978.
11. Seashore, C. *The vibrato*. Iowa City: University of Iowa, 1932.
12. Stanley, Douglas. *The science of voice: an application of the laws of acoustics*. Fischer. New York, 1958.
13. Sundberg, Johan. *The science of the singing voice*. Northern Illinois University. Illinois, 1987.
14. Sundberg, Johan. *Where does the sound come from?* En: A. Potter (Ed.), *The Cambridge Companion to Singing* (pp. 231-247). Cambridge: Cambridge University Press, 2000.
15. Titze, I., Story, B., Smith, M., & Long, R. *A reflex resonance model of vocal vibrato*. Journal of the Acoustical Society of America, No. 111 (pp. 2272-2282), 2002.
16. Van Rossum, Guido; Drake, Fred. *The Python Language Reference Manual*. NetworkTheory. Bristol, 2003.