Algoritmos de estimación tiempo-frecuencia para la transcripción de melodías monofónicas empleando protocolo MIDI

***Abstract*—This work propose a simple and effective methodology to detect and estimate musical tones, in real time, from mid-register musical instrument executions. Digital signal processing not only involves time-frequency estimation algorithms to establish the executed tones but the capability to generate automatically the coding needed to adapt the information to MIDI protocol, a widely accepted standard in music industry.**

***Resumen*—El presente trabajo propone una metodología simple y eficaz para la detección y estimación de notas musicales provenientes de la ejecución, en tiempo real, de instrumentos musicales de registro medio estándar. El procesamiento digital involucra algoritmos de estimación tiempo-frecuencia para la determinación de las notas ejecutadas como así también la capacidad de generar en forma autónoma la codificación necesaria para adecuar la información obtenida al protocolo MIDI, estándar ampliamente aceptado en la industria musical.**

I. Introducción

En la actualidad, un gran número de músicos, tanto profesionales como aficionados, no poseen la capacidad de escribir sus obras.

Con la llegada de los sintetizadores, se ha desarrollado un método unificado para la digitalización de información musical, el protocolo MIDI. Mediante éste, se provee información de tono y duración, pudiéndose utilizar esta información desde el procesamiento a través de estudios virtuales hasta la obtención de un archivo estándar, el cual puede ser leído por editores de partituras.

Hasta el momento, son pocas las herramientas que proveen al músico que ejecuta instrumentos netamente analógicos una practicidad similar al protocolo MIDI.

Este trabajo propone como solución a las problemáticas planteadas, la interpretación digital de la información de instrumentos analógicos, siendo los casos de estudio particulares: la guitarra y el piano.

II. Marco Teórico

A. Nociones de teoría musical

A lo largo del presente trabajo, se buscará obtener una partitura en base a la interpretación de una melodía. Dada esta premisa, se utilizarán conceptos básicos acerca del lenguaje musical.

A continuación, se listan los aspectos más representativos dentro de la problemática [1].

1) Tempo: Indica la velocidad a la que se interpreta la melodía. Se mide en *negras*, beats o pulsos por minuto (BPM o PPM), siendo la *negra* lo que se denomina unidad de tiempo.

2) Compás: Es un conjunto de unidades de tiempo o subconjunto de éste, en donde sólo la primera unidad de tiempo está acentuada.

3) Figura: Indica la duración de cada nota en función del tempo. Cada tipo de figura se representa por medio de un símbolo diferente.

B. Transformada discreta de Fourier (DFT)

La trasformada discreta de Fourier tiene por objetivo la conversión de información contenida en dominio temporal al dominio de las frecuencias complejas. Es equivalente a la descomposición de una señal en tiempo discreto en una suma finita de señales sinusoidales de tiempo discreto [2], que conforman una base ortogonal en el espacio de frecuencias, de forma análoga a como un vector dado puede descomponerse en distintas bases ortogonales en un espacio de $\mathbb{R}^n$.

Sea una señal en tiempo discreto $x[n]$ de la cual se desea conocer su espectro en frecuencia, se aplica la ***DFT*** según la siguiente ecuación:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-ik(\frac{2\pi}{N})n} \qquad k = 0, 1, ... \ M \tag{1}$$

donde N es el número de muestras de la señal $x[n]$ y M la cantidad de puntos (bins) del espectro a calcular. Se define $N = M$ por periodicidad de la transformación [3].

La misma puede ser expresada de forma matricial como

$$X[k] = x[n] \cdot W_N^{kn} \tag{2}$$

con $x[n] \ \epsilon \ \mathbb{R}^{1xN}$ y $W_N^{kn} \ \epsilon \ \mathbb{C}^{NxM}$, siendo esta última la matriz base de la trasformación según la ecuación

$$W_N^{kn} = e^{-ik(\frac{2\pi}{N})n} \tag{3}$$

Analizando esta última expresión se puede demostrar que, producto de la simetría de la matriz base de la transformación, puede simplificarse el cálculo dando lugar a un algoritmo que reduce el número de operaciones intervinientes. Dicho algoritmo se conoce como *Transformada Rápida de Fourier o FFT* [4] por sus siglas en ingles.

Ambas implementaciones presentan los mismos inconvenientes para el desarrollo de la metodología de detección. Estos son limitaciones conocidas e inherentes a los propios algoritmos. A saber:

1) Resolución Espectral: Para una determinada frecuencia de muestreo fs, impuesta por el sistema de procesamiento, y un determinado número de muestras N de la señal en estudio $x[n]$, limitada también por la capacidad de almacenamiento de información del sistema, existirá una restricción en la representación espectral, por la cual no se podrá representar una frecuencia arbitraria sino sólo aquellas que son múltiplos de lo que se conoce como *resolución espectral*, la cual está impuesta por la razón entre estos dos parámetros previamente mencionados, es decir

$$\Delta f = \frac{f_s}{N} \tag{4}$$

esto es, cada valor del vector transformado representa el contenido espectral en cada frecuencia múltiplo de fs/N [3]. De aceptarse esta resolución espectral nominal como útil, se deberá entrar en una relación de compromiso donde la resolución espectral se incremente si

- Se incrementa el numero de muestras.
- Se reduce la frecuencia de muestreo.

Para los intereses del enfoque propuesto, donde se requiere hallar la componente fundamental del espectro en frecuencia, resulta una facilidad conocer la distribución de las notas musicales para la afinación estándar 12-TET A440 según la tabla I.

TABLA I
ESCALA DE NOTAS MUSICALES

	oc2	**oc3**	**oc4**	**oc5**	**oc6**	**oc7**
do	65,41	130,81	261,63	523,25	1046,50	2093,00
do#	69,30	138,59	277,18	554,37	1108,73	2217,46
re	73,42	146,83	293,66	587,33	1174,66	2349,32
re#	77,78	155,56	311,13	622,25	1244,51	2489,02
mi	82,41	164,81	329,63	659,26	1318,51	2637,02
fa	87,31	174,61	349,23	698,46	1396,91	2793,83
fa#	92,50	185,00	369,99	739,99	1479,98	2959,96
sol	98,00	196,00	392,00	783,99	1567,98	3135,96
sol#	103,83	207,65	415,30	830,61	1661,22	3322,44
la	110,00	220,00	440,00	880,00	1760,00	3520,00
la#	116,54	233,08	466,16	932,33	1864,66	3729,31
si	123,47	246,94	493,88	987,77	1975,53	3951,07

Se propone como instrumento de estudio inicial la guitarra de encordado estándar en cualquier fabricación, ya sea clásica, acústica o eléctrica. La misma tiene un registro de 4 octavas que se disponen entre *Mi* de la segunda octava y *Mi* de la sexta octava, es decir entre 82,41 Hz a 1318,51 Hz.

2) Resolución temporal: Considerando una frecuencia de muestreo fs de 22.050 Hz, y teniendo en cuenta que la distancia mínima entre las frecuencias de interés es de $4,9Hz$, se obtiene

$$N = \frac{22050Hz}{4.9Hz/muestra} = 4.500\ muestras \tag{5}$$

las cuales representan $0,204$ segundos de grabación. Por ende, se pueden observar dos condiciones heredadas:

- Se requiere espacio de memoria suficiente para alojar dicha cantidad de muestras.
- La resolución temporal por bloque resulta deficiente, puesto que es de interés la duración de cada nota ejecuta y ésta, por lo general, no está circunscrita al tiempo de grabación calculado.

Como consecuencia de esto, es necesario reducir la cantidad de muestras por bloques, lo cual trae aparejado la imposibilidad de representar todas las frecuencias de interés, mas aun, aquellas frecuencias que no puedan ser representadas de forma exacta, se representarán a través de una distribución energética en todo el espectro debido al efecto *leakage* [4].

C. Transformada de Q constante

Al igual que la transformada de Fourier, ésta también puede interpretarse en términos de un banco de filtros, sólo que a diferencia de ella, éstos se encuentran espaciados geométricamente $\Delta_k^{cq} = f_{k+1} - f_k$ donde cada $f_k = f_0 \cdot 2^{k/b}$ con b igual al número de filtros por octava y f_0 es la mínima frecuencia de interés [8].

La transformación está definida por:

$$X^{cq}[k] = \frac{1}{N_k} \sum_{n=0}^{N_k} x[n] \cdot e^{-j2\pi \cdot n \cdot \frac{Q}{N_k}} \tag{6}$$

donde $N_k = Q \cdot f_s / f_k$ representa el tamaño de la ventana de observación de la señal y se encuentra definido para $k < K$, con $K = b \cdot log_2\left(f_{max}/f_0\right)$, que representa el número máximo de frecuencias admisibles por la transformación a Q constante. Esta transformación presenta los siguientes inconvenientes:

- No conserva las propiedades de periodicidad y simetría de la transformada discreta de Fourier, con lo cual su costo computacional resulta mucho más elevado.
- N_k no es costante para las frecuencias de la base, y en particular es muy elevado para las frecuencias bajas.

D. Protocolo MIDI

El protocolo MIDI es ampliamente utilizado en instrumentos digitales como interfaz de control en estudios virtuales de edición de audio. El protocolo se limita a transmitir sólo información de las notas ejecutadas y los eventos que caracterizan su ejecución. A su vez, el archivo MIDI puede ser utilizado por editores de partituras tanto libres como pagos. para su posterior edición representando éste la partitura de la ejecución que le dio origen.

Si bien es sumamente amplio el conjunto de datos que ofrece el protocolo, en este trabajo, así como en numerosas aplicaciones, se hace uso de cuatro valores característicos:

- Identificador de nota
- Intensidad de la nota
- Notificación de comienzo
- Canal de transmisión

III. Desarrollo de la metodología de trabajo

En esta sección se detallarán los procedimientos intervinientes a efectos de lograr la correcta escritura de la partitura en función del audio adquirido. A su vez se expondrán los resultados parciales obtenidos.

A. Estimación temporal de cada nota

Dado que la finalidad de este trabajo es conocer no sólo la frecuencia de las notas ejecutadas, sino también su duración, se deberá estimar el comienzo y el fin de cada nota, para lo cual se utilizará como herramienta principal la estimación de la energía de cada bloque de señal.

Se sabe que para una señal discreta $x[n]$ y de longitud finita N, su energía se define como

$$E = \sum_{n=0}^{N-1} |x[n]|^2 \tag{7}$$

Será preciso también poder establecer la presencia o no de una nota ejecutada. En su ausencia será inevitable detectar el nivel de ruido medio propio del sistema y método de adquisición, el cual, por obvias razones, deberá ser ignorado. Para ello se propone establecer un rango dinámico η según:

$$\eta = E_0 \cdot k \tag{8}$$

siendo k una constante de escalamiento empírica y E_0 la energía de un bloque de N muestras en ausencia de señal de interés, que se determina previo al inicio de la ejecución de las notas.

Una vez obtenido esto, se procede a la clasificación de eventos sonoros según el estimador

$$\Delta\widehat{E} = \frac{E_a - E_p}{E_a + E_p} = \frac{\frac{E_a}{E_p} - 1}{\frac{E_a}{E_p} + 1} \qquad |\Delta\widehat{E}| \leq 1 \tag{9}$$

siendo E_a la energía del bloque actual y E_p la energía del bloque previo, ambos calculados según (7).

El comienzo y fin de una nota se encuentra ligado a la variación de energía de los sucesivos bloques de señal respecto al anterior. Cuando la diferencia energética entre los mismos es positiva, indica el comienzo de una nota, mientras que cuando su diferencia es negativa, comprenden la extinción de la nota, y los de diferencia energética cercana a cero representan el sostenimiento de la nota ejecutada o de un silencio. La ecuación (9) a su vez, normaliza dicha diferencia, lo cual comprende versatilidad para el calculo en procesadores digitales donde los valores que se adoptan estarán comprendidos entre valores fijos (Ej. codificación Q1.15).

Se comprueba empíricamente que la relación

$$\frac{E_a}{E_p} \geqslant 1.5 \tag{10}$$

responde al comienzo de una nota, lo que lleva a que $\Delta\widehat{E} = 0.2$. Es propicio decir que esta relación no tiene en cuenta el ruido, por lo que se debe implementar un método para reducirlo o un nivel mínimo de ruido para desestimarlo, como el propuesto según 8.

Teniendo en cuenta la relación (10) y el nivel de ruido propuesto por (8), se hallan tres estados posibles:

1) $\Delta\widehat{E} \geqslant 0.2$ y $E_a \geqslant \eta$, se trata del comienzo de una nota.
2) Si se detecto un comienzo de nota previamente y $E_a \geqslant \eta$, la nota persiste.
3) $E_a \leqslant \eta$, la nota se ha extinguido.

B. Adaptación de la transformada discreta de Fourier y la transformada a Q constante para tonos de interés

Visto los inconvenientes del estudio espectral planteados dentro de las técnicas tradicionales, se propone a continuación una adaptación a la transformada discreta de Fourier (Ec. 2) y a la transformación a Q constante (Ec. 6), la cual consiste en modificar la base de la trasformación.

Siendo conocida la distribución tonal estándar para señales provenientes de instrumentos musicales expresada en la tabla I se propone calcular $W_{m,n}$ según la ecuación

$$W_{m,n} = e^{-j\cdot 2\pi\cdot P_0\cdot P^m\cdot n\cdot Ts} \quad m = [0, 1, ..., M-1] \tag{11}$$

siendo P_0 la frecuencia de la nota tomada como base, es decir, la primera de interés, para este caso $65.41Hz$, P es la relación de adyacencia entre las frecuencias, en este caso $\sqrt[12]{2}$, Ts el periodo de muestreo elegido, y M la cantidad de elementos en el espacio transformado.

Teniendo en cuenta la expresión del calculo de la base de la transformada discreta Fourier

$$W_{m.n} = e^{-\frac{j2\pi}{N}mn} = e^{-j2\pi\frac{m\cdot fs}{N}n\cdot Ts}, \tag{12}$$

Se puede apreciar que $\frac{m\cdot fs}{N}$ define la resolución espectral y por consiguiente las frecuencias representables en la *DFT*. Si se compara la misma con 11, se puede observar que dicha resolución espectral fue sustituida por $P_0 \cdot P^m$, término que representa las frecuencias de la escala musical de la tabla I. Por tanto, el nuevo espacio transformado obtenido representa la energía de cada componente en función de la frecuencia al igual que en el espacio transformado de Fourier pero no en una escala lineal, sino exponencial.

En base a esto, se procede a determinar la constante M, la cual define la cantidad de elementos en el espacio transformado. La misma sólo queda limitada por la condición de *Nyquist* [3], entonces

$$\begin{aligned} P_0 \cdot P^{M-1} &\leq \frac{fs}{2} \\ P^{M-1} &\leq \frac{fs}{2\cdot P_0} \\ M-1 &\leq \log_P(\frac{fs}{2\cdot P_0}) \\ M &\leq \log_P(\frac{fs}{2\cdot P_0}) + 1, \quad M \in \mathbb{N} \end{aligned} \tag{13}$$

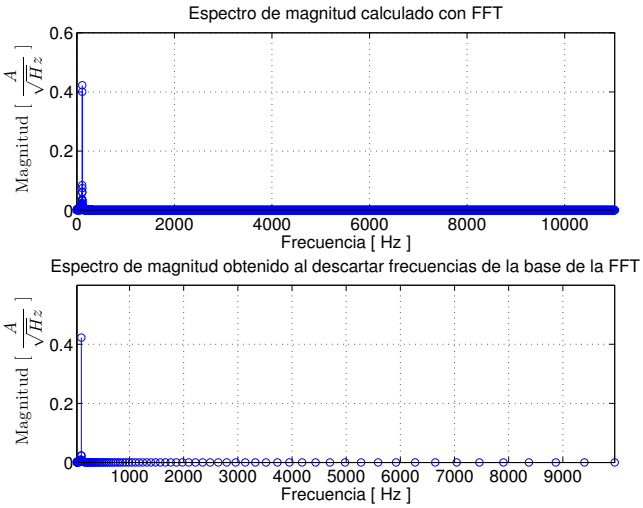

Fig. 1. Comparación energética de espectros.

A modo de comparación, podría considerarse esta herramienta como un ahorro de computo de la transformada de Fourier (FFT) para similares prestaciones, en la cual se descartarían aquellas frecuencias fuera de la escala musical. Si se considera la aplicación de la técnica Zero-padding [4] se llega a

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j \cdot 2\pi \cdot n \cdot Ts \cdot \frac{k \cdot fs}{\delta \cdot N}} \qquad (14)$$
$$k \in [0, M-1] \wedge M = \delta \cdot N$$

donde δ indica la cantidad de ceros por muestra necesarios. Para llegar a la hipotética resolución necesaria de $0,01Hz$ se halla que $\delta = 100$ para $fs = 22.050Hz$, lo cual conlleva gran cantidad de cómputos y memoria sumado a que es preciso un descarte posterior al cálculo espectral.

Del análisis de las ecuaciones, se concluye que:

- El cambio de base propuesto determina, del bloque a analizar, la potencia de las frecuencias vinculadas con las notas de la escala musical.
- El mismo permite cambiar la afinación de referencia P_0 y su relación de adyacencia P
- El tamaño del vector transformado M no varía con la cantidad de muestras N. Comparativamente, M disminuye de forma considerable, y por consiguiente, se reduce en la misma proporción la memoria necesaria para almacenar cada valor. En el ejemplo propuesto, se ahorra casi un 90% de memoria.
- La cantidad de cómputos requeridos disminuye respecto a la *DFT*.
- La energía contenida en el cambio de base es coincidente con las componentes de la transformada de Fourier con resolución acorde (Ver figura 1).

Dado que el piano es un caso de estudio del presente trabajo, se propone detectar hasta la segunda armónica de su nota más aguda, siendo ésta un si7 (ver Tabla I), por lo que, debido al teorema de Nyquist, se necesitaría por lo menos una frecuencia de muestreo de $16.000Hz$.

Además, se debe considerar que el sistema de adquisición se compone por un filtro pasa bajos con una cierta banda de transición, con lo que la frecuencia de muestreo debe ser superior a los 16 KHz propuestos para no atenuar las frecuencias de interés. Con esto en mente y en base al estándar de audio [6], se escogió una frecuencia de muestreo de 22.050Hz.

Si se supone una velocidad máxima de ejecución de 8 notas por segundo (0,125 segundos por nota), la cual es una velocidad elevada para un músico aprendiz, el número de muestras máximo para estimar la frecuencia de dicha nota es $N = 0,125fs = 2756\ muestras/segundo$, por lo que se propone que la transformada desarrollada conste de 2048 muestras.

Luego, la distancia mínima en frecuencia entre las notas que se buscan detectar se establece mediante la diferencia entre mi2 y fa2, esto es, $82,41Hz$ y $87,31Hz$ respectivamente, por lo que dicha resolución espectral será $\Delta f = fs/N = 4,9\ Hz$. Para evitar posibles casos en que la frecuencia de interés no queda explícitamente determinada por el bin mas cercano, se debe adoptar una resolución espectral mínima de $\Delta f/2 = 2,45Hz$.

En base a lo calculado anteriormente, para el caso de la FFT, la cantidad de muestras necesarias para obtener dicha resolución espectral será $N = fs/\Delta f = 9000\ muestras$. En este caso, es necesario una cantidad de muestras potencia de 2, y la más cercana a 9000 muestras es 2^{14} muestras.

A modo de comparación, la cantidad de operaciones que necesita la FFT en nuestro caso sería de $2^{14} \cdot 14 = 229.376$ y, para la transformada propuesta, es de $2048 \cdot 84 = 172.032$, con lo que supone, además, un ahorro en memoria y cantidad de operaciones.

C. Estimación espectral de la nota ejecutada

1) Selección del bloque a procesar: Generalmente, las notas de los instrumentos están formados por una parte percusiva y una parte armónica. La parte percusiva se caracteriza por tener alto contenido frecuencial y bajo contenido temporal, es decir, se proyecta en una gran cantidad de frecuencias pero con poca energía en cada componente debido a su rápida extinción temporal. Algunos instrumentos carecen de esta parte, por ejemplo, los instrumentos de cuerda frotada (violín, contrabajo) o los de viento (flauta, saxofón). La parte armónica se caracteriza por tener su energía concentrada en determinadas frecuencias, y su extinción es más gradual.

El bloque a elegir debe ser aquel cuya energía armónica sea máxima, y dado que la energía espectral de la parte percusiva de un instrumento es baja [7], se recomienda elegir el primer bloque de la nota para su análisis de frecuencia fundamental.

2) Estimación de frecuencias armónicas: El primer paso para el proceso de estimación realizada en cada bloque se basa en la búsqueda de el máximo absoluto del espectro, ya que, al tratarse de notas individuales, se asume que el mismo es el M-ésimo armónico de la señal original [5], es decir,

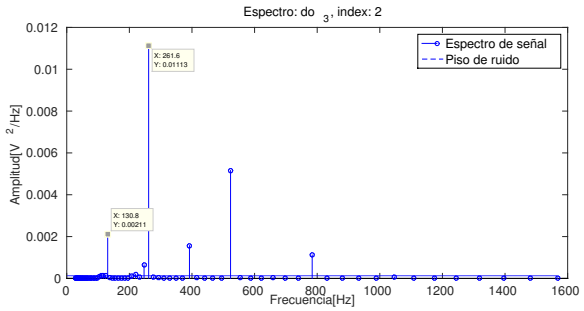

Fig. 2. Espectro correspondiente a la nota FA_3 de una guitarra, con su respectivo piso de ruido.

$$f_{max} = M \cdot f_0, \;\; M \in \mathbb{N} \tag{15}$$

El paso siguiente es calcular las M posibles frecuencias fundamentales, que a partir de 15 será

$$f_{fundamentales}[i] = \frac{f_{max}}{i}, \tag{16}$$

siendo $i = 1, 2, ..., M$.

Luego, se calcula la media del espectro, siendo el mismo considerado como umbral o piso de ruido, es decir, los picos debajo de dicho nivel son descartados. Un ejemplo de este piso de ruido puede observarse en fig. 2, calculado a través de la media aritmética escalada en función del sistema, esto es

$$\eta_f = media * \lambda \tag{17}$$

Finalmente, se compara cada uno de los máximos superiores al piso de ruido con las posibles frecuencias fundamentales obtenidas en 16, siendo la frecuencia más baja coincidente la frecuencia fundamental de la señal en cuestión [5].

D. Generación MIDI

Una vez obtenido la información de las notas en cuanto a tiempo y frecuencia, y considerando como objetivo principal la escritura de una partitura, se confecciona un archivo MIDI el cual contendrá los valores que simbolizan los datos estimados y el usuario deberá informar el *tempo* y *compás* de ejecución.

Una vez logrado el archivo final, se podrá editar en caso de existir algun error de ejecución o eficacia en la estimación realizada.

En la figura 3 se esquematiza el diagrama en bloques referido a la generación del archivo MIDI.

IV. Resultados

En condiciones controladas, se realizaron grabaciones en las cuales se convino previamente qué notas serán ejecutadas y qué duración tendrán, posteriormente se estudiaron dichos audios con la metodología presentada. Las mismas fueron adquiridas por interfaces de audio comerciales a saber:

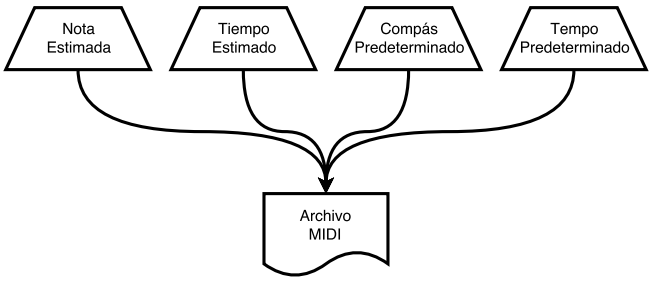

Fig. 3. Algoritmo para la generación de archivos MIDI.

- M-Audio M-track: resolución de 24 bits y fs=44100 Hz
- FocusRite Scarlett 2i:2 : resolución de 24 bits y fs=44100 Hz

Ambas presentaron niveles similares de ruido aditivo, sumado al propio de las técnicas de resampling para obtener la *fs* optima, de esta forma se logró imponer las mismas constantes $k = 100$ (ec. 8) y de escalamiento de media espectral en $\lambda = 0,4$ (ec. 17) para ambos sistemas de adquisición.

Se obtuvieron los siguientes resultados.

TABLA II
DATOS EXPERIMENTALES

Grabación	Instrumento	Nne	Nnd	Ncd	Fp	Fn
Cantata	Guitarra	73	71	71	0	2
	Piano	73	75	70	2	0
jijiji	Guitarra	47	47	47	0	0
	Piano	47	47	47	0	0
Mario Bros	Guitarra	24	25	24	1	0
	Piano	24	24	24	0	0
MiM2a4	Guitarra	15	15	13	0	0
	Piano	15	16	13	1	0
MiM3a5	Guitarra	15	15	13	0	0
	Piano	15	15	13	0	0
MiM5a6	Guitarra	8	8	7	0	0
	Piano	8	8	7	0	0

En la Tabla II se aprecian los valores obtenidos en base a interpretaciones en guitarra eléctrica y piano de la *Escala de Mi mayor* notada como *MiM2a4* donde los números indican las octavas ejecutadas. Dichas grabaciones comprenden todo el registro de la guitarra (Squier Stagemaster) y su interpretación en las respectivas octavas del piano (Medeli M20). Por otra parte, se obtuvieron grabaciones de tramos de 3 obras musicales de conocimiento publico:

- Cantata de puentes amarillos - Luis Alberto Spinetta [0:00 a 0:27]
- Jijiji - Patricio rey y sus redonditos de ricota [2:38 a 2:50]
- Super Mario Bros Theme - Koji Kondo [0:17 a 0:22]

Los valores expuestos corresponden a:

- Nne número de notas ejecutadas
- Nnd número de notas detectadas
- Nne número de notas correctamente detectadas
- Fp notas detectadas que no fueron ejecutadas
- Fn notas ejecutadas que no fueron detectadas

TABLA III
MÉTRICAS OBTENIDAS

Grabación	Instrumento	Precisión	Deteccion	Exactitud	Ebi
Cantata	Guitarra	1,00	0,97	0,97	99,85
	Piano	0,93	1,03	0,97	99,94
jijiji	Guitarra	1,00	1,00	1,00	99,76
	Piano	1,00	1,00	1,00	99,97
Mario Bros	Guitarra	0,96	1,04	0,96	99,99
	Piano	1,00	1,00	1,00	99,97
MiM2a4	Guitarra	0,87	1,00	1,00	99,94
	Piano	0,81	1,07	0,94	99,27
MiM3a5	Guitarra	0,87	1,00	1,00	99,98
	Piano	0,87	1,00	1,00	99,96
MiM5a6	Guitarra	0,88	1,00	1,00	99,95
	Piano	0,88	1,00	1,00	99,95

De esta forma se han obtenido las métricas de la tabla III entendiendo a cada una de ellas de la siguiente manera:

- $precision = \frac{N_{cd}}{N_{nd}}$. Noción de acierto respecto a la cantidad de notas y el tono detectado.
- $deteccion = \frac{N_{nd}}{N_{ne}}$. Noción de acierto respecto a la cantidad de notas ejecutadas y percibidas.
- $exactitud = \frac{N_{nd}}{N_{nd}+F_p+F_n}$. Noción de acierto respecto a la cantidad de notas ejecutadas y percibidas considerando aciertos efectivos y erróneamente aceptados.
- Accesoriamente se calculo la energía de la señal en la banda de interés (Ebi).

Deforma general los estadísticos planteados arrojan una media para cada caso de:

- $precision = 92\%$
- $deteccion = 98\%$
- $exactitud = 98\%$
- $Ebi = 99,87\%$

Finalmente se considera un promedio general de dichos valores a los efectos de alcanzar un valor representativo para los casos de estudio planteados, logrando una eficiencia general del 96%

V. Conclusiones

Los algoritmos propuestos presentan la ventaja de tener un bajo costo computacional, permitiendo implementarlos con procesadores económicos. Además, como se puede observar en los resultados obtenidos, los mismos tienen una gran tasa de acierto.

Si bien se tiene presente que en las interpretaciones a modo general se utilizan mayormente armonías que notas aisladas, se puede considerar al presente como base para desarrollar la detección genérica de forma óptima y precisa, así como para otros tipos de instrumentos que no han sido caso de estudio en este trabajo.

REFERENCIAS

[1] Otto Karolyi, *Introducción a la música*. Alianza Editorial, Madrid, España, 2006.

[2] Eduardo A. B. da Silva, Sergio L. Netto, Paulo S. R. Diniz. *Digital Signal Processing - 2nd Edition*. Cambridge University Press, Cambridge, United Kingdom, 2010.

[3] R. Armentano, D. O. Craiem. *Análisis de sistemas lineales*. Editorial CEIT, Buenos Aires, Argentina, 2011.

[4] Richard G. Lyons. *Understanding Digital Signal Processing*. Prentice Hall PTR, New Jersey, United States, 2004.

[5] Bozena Kostek. *Perception-Based Data Processing in Acoustics: Applications to Music Information Retrieval and Psychophysiology of Hearing*. Springer-Verlag, Berlin, The Netherlands, 2005.

[6] Christiane Rouseeau, Yvan Saint-Aubin. *Mathematics and Technology*. Springer, New York, United States, 2008.

[7] D. FitzGerald. Harmonic/percussive separation using median filtering. 2010.

[8] Judith C. Brown. Calculation of a constant Q spectral transform. 1991.