SÍNTESIS VOCAL POR FM LIBRE DE 'CLICKS'

Chris Chafe

Center for Computer Research in
Music and Acoustics, Stanford
University

RESUMEN

La Frecuencia Modulada (FM) y otras técnicas no lineales de modulación a frecuencia de audio, como la síntesis digital por Modelado de Onda [Waveshaping], la Amplitud Modulada (AM) y sus variantes, son técnicas bien conocidas para generar espectros sonoros complejos. Kleimola [1] aporta una descripción exhaustiva y actualizada de la familia entera. La síntesis de sonidos vocales y la de otros sonidos armónicamente estructurados que presentan formantes es problemática debido a que aparecen distorsiones cuando se intensifican los controles variantes en el tiempo.

Grandes desvíos de la altura o de los parámetros de los fonemas causan saltos en las aproximaciones enteras que se requieren para determinar la frecuencia central de los formantes. Cuando se intenta imitar la conducta vocal humana, con sus amplias excursiones prosódicas y expresivas, aparecen 'clicks' audibles. Una solución parcial a este problema yace oculta en algún código desde la década de los '80. Esto, combinado con un banco de osciladores de fase sincrónica descripto en Lazzarini y Timoney [2] produce componentes armónicos uniformes que aseguran espectros formantes exactos, libres de efectos secundarios no deseados [artifacts], aun bajo las condiciones dinámicas más extremas. Este artículo revisita la síntesis del canto y del habla usando la clásica estructura FM de múltiples portadoras y una única moduladora, derivada del trabajo pionero de Chowning [3]. El método revisado es implementado en Faust y es tan eficiente como su técnica predecesora. Los controles dinámicos llegan multiplexados a través de un flujo de datos [stream] a frecuencia de audio que se articula con los algoritmos escritos en Chuck, los cuales trabajan sincrónicamente a nivel de las muestras. Se puede ahora 'abusar' de la síntesis FM para el canto utilizando controles que varían drásticamente en el tiempo. Esto también ofrece potencialmente un medio eficiente para la codificación del habla mediante análisis-resíntesis con bajo ancho de banda. Se describen también en este artículo aplicaciones de la mencionada técnica en el terreno de la sonificación y de la música de concierto.

1. INTRODUCCIÓN

La síntesis vocal para el canto tiene una historia que arranca en los primeros años de la música por computadora. La canción *Daisy Bell (Bicycle Built for Two)* fue cantada por una computadora en 1961 en un arreglo de Max Mathews y Joan Miller, con síntesis vocal a cargo de John Kelly y Carol Lochbaum, cuando la síntesis digital de música tenía sólo cuatro años de historia. Ese fue un ejemplo temprano de codificación de la voz mediante síntesis/resíntesis que sirvió de base para la síntesis del canto. A lo largo de las décadas, la mayor parte de las técnicas de síntesis ha sido aplicada para emular la voz en el canto (aditiva, substractiva, modelado físico, FOF, etc.). Más de cincuenta años después la búsqueda continúa con compositores atraídos por sintetizadores vocales como el Vocaloid[1] de Yamaha, que permite explorar un mundo fascinante de personalidades musicales de cantantes que nunca existieron. Este artículo se une a una corriente que empezó con el trabajo de Chowning a fines de la década del 70' y comienzos de los '80, que involucró la síntesis vocal por FM y que virtualmente ha estado languideciendo desde su temprano uso en algunas pocas obras musicales.

La primera descripción del método de John Chowning para la síntesis por FM de la voz cantada es la que aparece en su artículo de 1980[3], antes de completar *Phone* en el IRCAM (1981). La cinta multicanal exhibe una amplio rango de voces que cantan y de transformaciones de timbres vocales con otros generados por FM, tales como sonidos de gongs. La técnica crea múltiples formantes con afinaciones independientes usando múltiples portadoras que comparten una misma moduladora. Dos formantes se usan para la voz de soprano cantando "eee" y tres formantes para su espectralmente rico bajo *profondissimo*. Una versión posterior añade un tercer formante al modelo de la soprano en la síntesis de la vocal "aah" [4]. El vibrato en la altura, que causa modulación espectral simultánea, es especialmente efectivo para hacer convincentes las vocales. "Es llamativo que el tono se funde y convierte en una percepción unitaria con el agregado de fluctuaciones

[1] Vocaloid3 usa un motor de síntesis concatenativa en el dominio de la frecuencia basada en secuencias de tres fonemas.

en la altura, ¡de este modo la envolvente espectral no hace una voz!" [3].

El método tiene un inconveniente intrínseco que limita cuan amplia puede ser la excursión del vibrato como también la transición de un fonema a fonemas adyacentes. Cuando se superan esos límites ocurren 'artefactos' evidentes que son causados por cambios discretos de la frecuencia central del formante. Las discontinuidades se perciben como 'clicks' y resultan de cambios en los números enteros que definen la razón portadora-moduladora *c:m* y que son necesarios para determinar la frecuencia central f_c del formante deseado para una altura dada f_p. El oscilador que hace de moduladora se establece siempre a f_p, de modo que m = 1.0. La razón c de la portadora es una *aproximación entera*, cuantización de la real razón f_c / f_p.

La síntesis formántica con FM es esencialmente contradictoria a la física. La naturaleza armónica de los sonidos vocales sólo permite razones en que la portadora $c \in \mathbb{N} \geq 1$ es un número de armónico. Cuando el sistema físico de producción sonora se basa un mecanismo excitación-resonancia, con afinación independiente de ambos elementos, la FM sólo puede aproximar las frecuencias resonantes cuando existe la restricción de producir espectros armónicos. El problema inherente es que estas aproximaciones son discontinuas en frecuencia. En la práctica, esto limita severamente la cantidad de desvío en la altura que se aplica, no solo al vibrato, sino a glissandos y portamentos, haciendo que estos desvíos no puedan ir más allá que un pequeño rango. En la forma original del método es imposible cambiar razones sin causar clicks como aquellos mostrados en la Figura 1.

Encontrar una solución fue necesario para usar la FM como técnica de síntesis en un proyecto de sonificación relacionado con señales cerebrales. El canto por FM ofrece ventajas en este campo que intenta modelar un coro que canta directamente desde la mente. La meta no es lo que uno primeramente podría imaginar, la de un conjunto de voces controladas mentalmente. Se trata en cambio de una tecnología para mostrar, auditivamente, las rápidas fluctuaciones de registros de electroencéfalogramas (EEG) y electrocorticografías (eCog). La síntesis del canto es atractiva en que puede aludir a imágenes de 'voces internas', pero resulta particularmente apta a causa de la facilidad con la cual los oyentes siguen patrones de transición tímbrica entre fonemas u otros sonidos de tipo vocal. El rango de datos encontrados en registros del cerebro (desde quietud hasta cuando se sufre un ataque) y el deseo de contar con una estrategia de mapeo muy flexible han sido las motivaciones para la presente investigación, destinada a resolver el problema de la discontinuidad. El trabajo concluido llegará al público como instalación de galería (que explora datos grabados) y como un dispositivo de monitoreo médico para detectar ataques (con la voz que canta controlada directamente por electrodos en tiempo real).

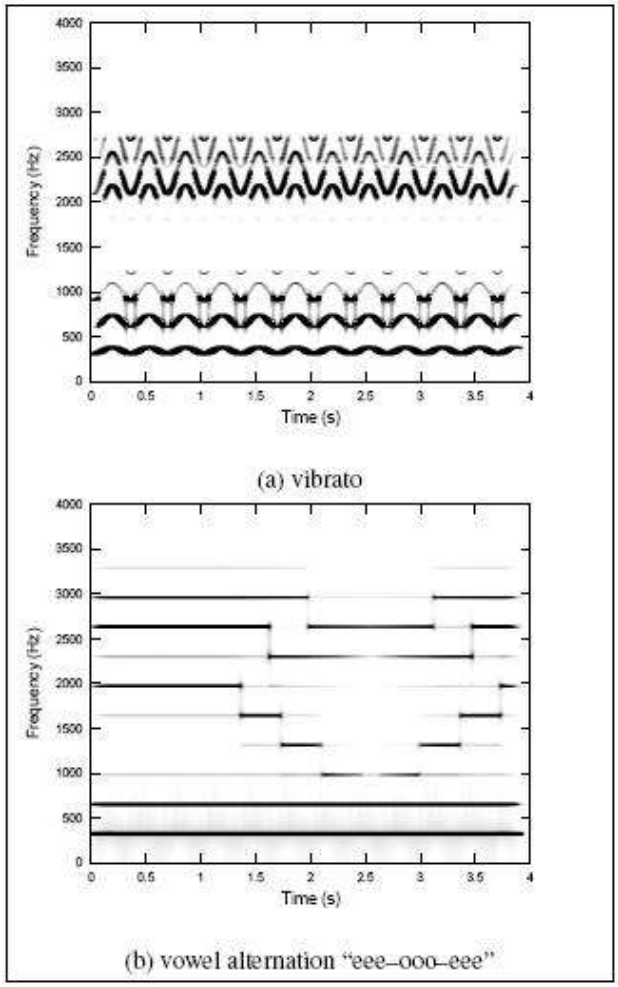

Figura 1. Los clicks ocurren siempre que las transiciones a una nueva frecuencia central de formante f_c obligan al oscilador que hace de portadora a cambiar su razón armónica. Los formantes vocales por FM usan una razón *c:m* donde $c \in \mathbb{N}_{\geq 1}$ y $m = f_p$, donde f_p es la altura deseada.

2. SOLUCIÓN TEMPRANA

En 1979 Marc Le Brun describió la síntesis digital por Modelado de Onda [Waveshaping] como un paradigma en el terreno de la síntesis por modulación no lineal [5]. La FM es un caso particular de síntesis por Waveshaping y Le Brun, al idear una solución para el problema de la discontinuidad en la Waveshaping, también lo resolvió para la FM. La solución de Le Brun permaneció inédita (hasta ahora) con una excepción: Bill Schottstaedt la ha preservado como un instrumento de síntesis en el proyecto Common Lisp Music (CLM) [6]. En un comentario del código se lee "**Vox**, un elaborado instrumento FM con múltiples portadoras, es el instrumento de voz escrito por Marc Le Brun y usado en *Colony* y otras piezas".

Vox evita las discontinuidades que surgen del cambio en la razón de números enteros implementando una solución basada en el 'fundido' [cross-fading]. Dos portadoras, que corresponden a números de armónico par e impar, se asignan a cada formante 'enmarcando' la frecuencia central del verdadero formante. Las asignaciones se hacen tomando los dos armónicos más cercanos, el armónico inferior inmediato $f_{lower} = \lfloor f_c / f_p \rfloor$ y el armónico superior inmediato $f_{upper} = \lceil f_c / f_p \rceil$. La asignación de armónicos a los osciladores individuales es dinámica y depende de si ellos tienen número par o impar. Cuando se requiere que un oscilador cambie su número de armónico, el otro estará aproximándose al verdadero objetivo f_c / f_p. Las amplitudes de los dos osciladores que actúan de portadoras suman 1, en una mezcla cuyas ganancias son complementarias y están linealmente determinadas por la proximidad al objetivo. La clave que hace que esto funcione radica en que el oscilador que está teniendo su frecuencia cambiada es silenciado. Un ventajoso efecto lateral es que se agudiza la precisión con que se sintetiza la frecuencia central del formante al cual se apunta.

El artículo de Le Brun describe "un marco conceptual unificado para un número de técnicas no lineales, incluyendo la síntesis por frecuencia modulada. Tanto la teoría como la práctica del método están desarrolladas considerablemente, empezando con formas simples aunque útiles y continuando con variaciones más ricas y complejas". Sin embargo, la solución por fundido [cross-fade] existía sólo en código de programación de esa misma época. Para detallar de manera precisa la historia, su primera implementación fue escrita en el compilador MUS10 (versión usada en el Laboratorio de Inteligencia Artificial de Stanford de los compiladores MusicN de los Laboratorios Bell). Más tarde fue llevado a CLM como **pqw-vox**, una "traducción desde MUS10 del instrumento de voz por waveshaping de Marc Le Brun (usando waveshaping por cuadratura de fase)". Hoy, tanto **pqw-vox** como la versión FM **vox** se encuentran traducidos a Scheme en el proyecto Snd [6] de Schottstaedt, como instrumentos definidos en el archivo *clm-ins.scm*[2].

Las implementaciones corrientes de la síntesis vocal por FM no han incorporado la solución del 'fundido' [cross-fade]. Lo más destacable hoy es el instrumento **FMVoice** incluido en el Synthesis Tool Kit (STK) [7]. La clase *FMvoices.cpp* puede descargarse libremente como parte del código fuente del STK y ha sido portada a varias plataformas, como ser Chuck [8] y Max / MSP / PeRColate [9]. Portando esta clase a Faust [10] y lidiando con el problema de la discontinuidad, fue que "redescubrí" para mí mismo la vieja solución de Le Brun. La misma solución del fundido aparece en Lazzarini y Timoney [2].

[2] Una precaución: algunas versiones de la implementación que pertenecen a esta familia señalan equivocadamente a osciladores que son portadoras como moduladoras y viceversa, la portadora es en realidad la moduladora.

3. NUEVO PROBLEMA

Lazzarini y Timoney describen también un método para generar formantes con osciladores de fase sincrónica. Su importancia será evidente. Después de tomar el código de Le Brun en mi propio trabajo y verificar que los clicks de las discontinuidades habían desaparecido (Fig. 2) advertí que el modo de arreglar eso introdujo un nuevo problema.

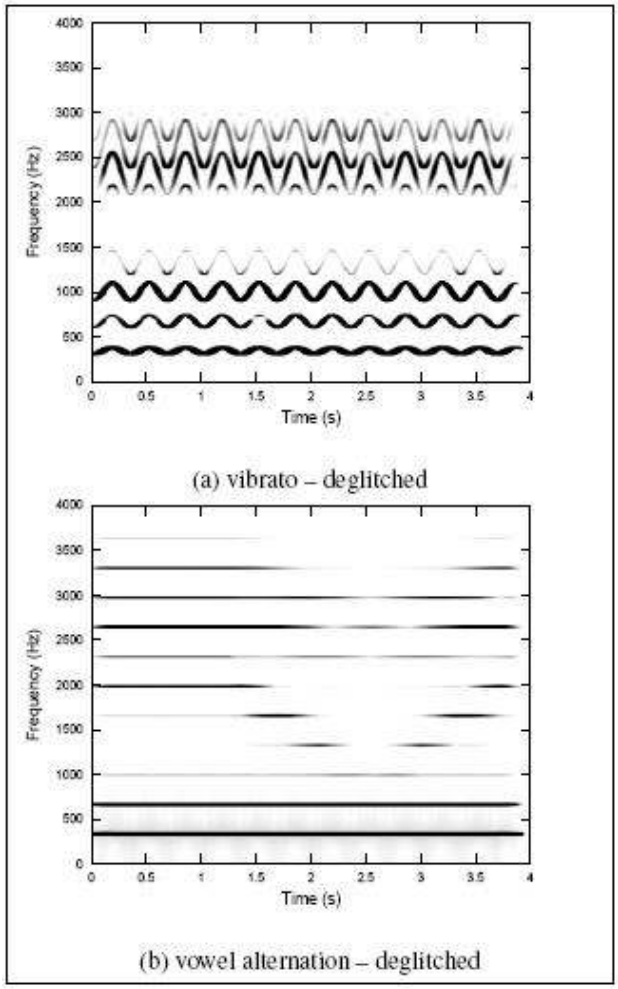

Figura 2. Resultado de aplicar la solución adoptada por Le Brun a la síntesis mostrada en la Fig. 1.

Se trataba nuevamente de un efecto secundario audible, no deseado [artifact], que atacaba al vibrato. No eran clicks, sino un nuevo tipo de efecto colateral no querido [artifact]. Donde un vibrato perfectamente periódico debería producir una modulación espectral perfectamente periódica, esto en realidad no ocurría. De un ciclo del vibrato al próximo se podía escuchar un patrón superpuesto de modulación

espectral. El problema se debe a la falta de una adecuada correspondencia entre las fases del par de formantes (con números de armónico par e impar) que se están mezclando para cada formante. Se trata del par 'fundido' [cross-faded] para combatir los clicks, problema al cual nos referiremos desde ahora como el "problema de primer orden".

La técnica del fundido supone que los pares de líneas espectrales coincidentes se sumarán aritméticamente. Sin embargo, esta suposición no toma en consideración a la fase. Surge un "problema de segundo orden" causado por la interferencia de fase entre líneas espectrales coincidentes. Estas son las líneas espectrales (correspondientes a las frecuencias de la portadora y a la de las bandas laterales) de los dos generadores de formante imbricados que completan la envolvente espectral del formante. Son espectros idénticos desplazados uno con respecto al otro por un número de armónico. Todas las fases están generadas en relación a sus respectivos osciladores portadores más que en relación al conjunto de frecuencias como un todo. Sin osciladores de fase sincrónica, estas fases son arbitrarias en el tiempo dado que son determinadas de modo independiente mediante los cambios de control.

Como se dijo en la sección 2, el 'artefacto' de primer orden se evidencia solo en condiciones dinámicas, cuando el 'blanco' respecto a la altura o el fonema deseado es cambiante. Similarmente, el efecto de segundo orden puede pasar inadvertido en condiciones estacionarias. Si las frecuencias de las portadoras no cambian, las fases de esas portadoras también serán constantes, como también lo será la mezcla espectral resultante. Sin embargo, la interferencia entre conjuntos de fases no relacionados puede tener un efecto que altere la envolvente espectral estática, que se percibe como un alejamiento de la cualidad a la que se debería apuntar para la porción estacionaria de una vocal. El problema se hace más evidente cuando las frecuencias portadoras están siendo cambiadas dinámicamente, especialmente si estos cambios están ocurriendo periódicamente. El vibrato es una buena manera de mostrar el problema. Distorsiones espectrales que podrían pasar desapercibidas bajo otras condiciones son más fáciles de oír con cambios de control repetitivos. El oído reconoce el efecto de distorsión como una especie de "isorritmo" espectral, o como un patrón 'alias' superpuesto. Un vibrato con un período determinado generará un patrón de distorsión con un período más largo, tal como se ve en la Figura 3. Si se estudia las regiones alrededor de los 700Hz y 1200Hz, se observarán patrones en que franjas de Moiré relacionadas con la fase se inscriben sobre la amplitud de los armónicos.

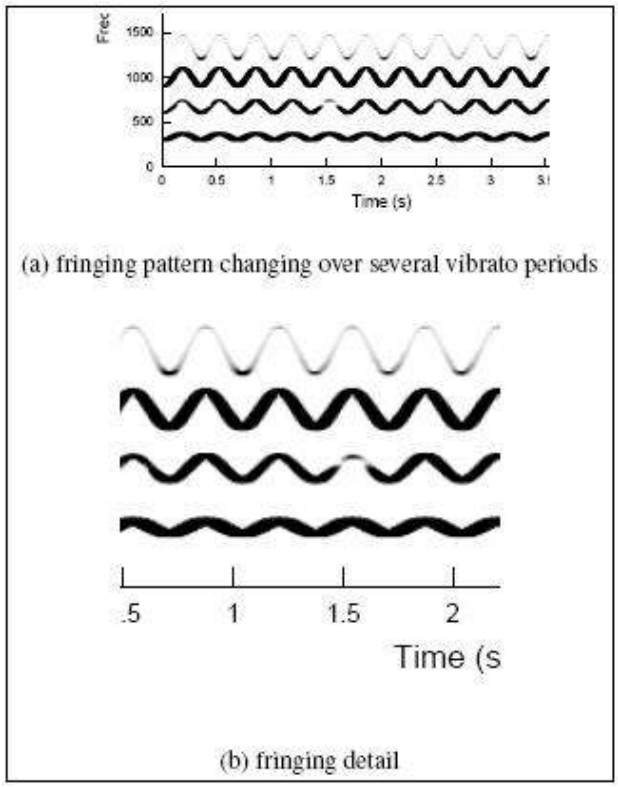

Figura 3. Franjas de Moiré relacionadas con la fase, ampliación de Fig. 2 (a). Cuando una de las dos portadoras en fundido [cross-faded carriers] redefine su frecuencia porque su razón armónica necesita cambiar, también redefine su fase respecto a la otra portadora. El resultado, visible en los espectrogramas, es el efecto de las franjas de Moiré.

3.1. Minimizando Franjas

Vale la pena mencionar un intento inicial por reducir el efecto audible de las franjas de fase, aun cuando no sea la solución que se adoptará en última instancia. Explota el hecho de que la interferencia es más notoria cuando las dos portadoras de la mezcla en fundido [cross-fade mix] tienden a participar en proporciones iguales en esa mezcla (la interacción es máxima en ese momento). Este es el punto donde las portadoras equidistan de la frecuencia central buscada. Por el contrario, la menor interferencia ocurre cuando una de ellas está lo más cerca posible del objetivo y la otra, esencialmente, silenciada. Sacando ventaja del caso en que un oscilador domina, se puede minimizar el efecto de las franjas expandiendo su tiempo en el ciclo activo [duty cycle] (relacionado con el vibrato).

Pruebas auditivas demuestran que la rampa del fundido [cross-fade ramp] puede hacerse no lineal y todavía enmascarar perfectamente la discontinuidad de primer orden. Usando la ley de potencias para la pendiente de la rampa, se reduce el efecto de las franjas porque se está menos tiempo en la porción del ciclo activo donde se dan

las proporciones problemáticas. Experimentos iniciales realizados bajo condiciones de vibrato periódico indican que se puede usar un exponente significativamente alto, por ejemplo, $f(x) = x^7$ y silenciar lo suficiente el efecto de primer orden como para evitar un click. Esto reduce en gran medida el tiempo empleado en el "modo de interferencia de fase" llegando casi a eliminar de este modo el efecto audible de segundo orden.

Este arreglo tiene sus contras. Algo del problema de las franjas persiste todavía durante la porción del ciclo activo donde la mezcla en fundido cruza brevemente la región de porciones iguales. Más significativo es, sin embargo, el hecho de que alterar la dinámica temporal de la mezcla, aleja a ésta del objetivo; es decir, de la mezcla que mejor aproxima la frecuencia central del formante que se busca.

3.2. Un camino mejor

La raíz del problema está en el empleo de osciladores independientes. Un camino para solucionar esto es emplear un banco de osciladores coordinados, tal como los osciladores de fase sincrónica descriptos en [2]. En este caso, la moduladora y todas las portadoras comparten un único fasor. En la presente implementación, el banco se puede construir con cualquier número de salidas armónicas y todas serán tomadas de un único fasor en común.

La familiar sinusoide muestreada genera la frecuencia (altura) fundamental de la señal para el banco a construir:

$$x(t) = Asin(\omega t + \varphi) \qquad (1)$$

donde A es la amplitud del oscilador y ω, en rad/s, es calculada como $2\pi f$, donde f es la frecuencia.

Expresada en pseudo-código, la Ec. (1) puede implementarse con la función módulo:

```
w = f / SR
mp = 0.0
for n = 0 to N
   y[n] = a * sin(2pi * mp)
   mp = (mp + w) mod 1.0
end
```

La constante *SR* es la frecuencia de muestreo [sample rate] y la variable *mp* la fase instantánea de la fundamental.

La clave para el paso siguiente es compartir la fase instantánea *mp* con cualquiera de los otros osciladores, donde *o* especifica el número de oscilador, cp_o su fase instantánea y h_o su número de armónico:

```
   cp[o] = (h[o] * mp) mod 1.0
   y[o][n] = a[o] * sin(2pi * cp[o])
```

y desde que estanos interesados en hacer FM

```
m[o] = y[n] * i[o]
cp[o] = (h[o] * mp) mod 1.0
y[o][n] = a[o] * sin(2pi * cp[o] +
m[o])
```

El pseudo-código de arriba implementa un simple par FM formado por una portadora independiente y una moduladora compartida que produce un formante centrado en el armónico *h*, de frecuencia *f* con índice de modulación *i*. El último coeficiente determina el ancho de banda y se usa habitualmente en un rango bajo (< 2.0). En la práctica, un banco de seis (o más) osciladores portadores de este tipo serán usados para generar un sonido vocal. Estos crearán fonemas de tres (o más) formantes representados por una distribución que varía en el tiempo de los coeficientes *h*, *a*, y *i*.

El método completo libre de clicks consta de la voz cantada por FM de Chowning + el algoritmo en cross-fade de Le Brun (Sec. 2) + el banco de osciladores de fase sincrónica de Lazzarini (Sec. 3). La Fig. 4 muestra un espectrograma con la solución completamente realizada (los apéndices presentan listados de código en Faust y en Chuck usados para generar el ejemplo). La síntesis clásica usando tablas de fonemas y el método de Chowning puede ahora extenderse a situaciones donde la conducta es arbitrariamente dinámica.

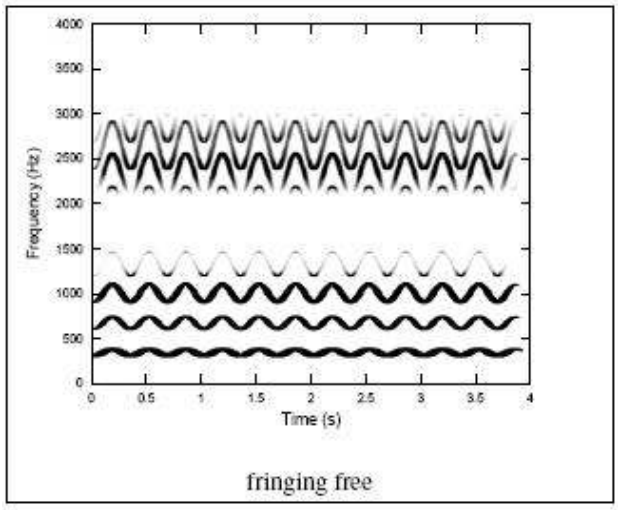

Figura 4. La realización del vibrato del ejemplo 2(a) con osciladores sincronizados en fase elimina el efecto de las franjas.

4. APLICACIONES

La técnica de la voz cantante ha sido utilizada en tres proyectos. Ejemplos sonoros para cada uno de ellos se pueden encontrar 'on-line' [11].

4.1. Convirtiendo Señales eCog a Música

La Electrocorticografía (eCog) registra la actividad eléctrica del cerebro directamente desde el interior del cráneo. El uso de sensores colocados en regiones donde se supone se origina la epilepsia ofrece, por un lado, un preciso control del diagnóstico y, por otro, señales de gran importancia para el estudio del cerebro mismo. En un corriente proyecto de sonificación los datos obtenidos por eCog son cantados por un coro digital. Cada cantante es una simulación vocal sintetizada con la técnica aquí presentada. Cuando los sensores han sido implantados por razones terapéuticas, un gran número de ellos está disponible (> 50) lo que permite que el coro pueda hacerse igualmente grande. El objetivo del proyecto es crear música directamente desde los sensores.

Existe una correspondencia entre las estructuras temporales de la música y la dinámica de la actividad cerebral registrada por eCog. El tiempo musical tiene sus notas, ritmos, frases y estructuras que van marcando etapas. Las señales cerebrales están estructuradas sobre las mismas escalas de tiempo. La traducción a música no requiere modificar la base temporal. De hecho, el presente enfoque evita 're-componer' o alterar los datos en forma alguna. Las personas que han contribuido con datos para este proyecto comparten nuestro interés en descubrir el potencial de la música como una forma diferente y nueva de aprehender la complejidad de la dinámica cerebral.

Los ataques que hemos escuchado tienen una progresión característica. Surgen con un 'aura' liviana, en rápida modulación, no distinta a un vibrato o tremolo super-rápido que da lugar a una fuerte marcha de pulsos, casi regular. Múltiples trenes de pulsos suenan uno contra el otro formando polirritmias. Esto va creciendo hasta un climax casi insoportable, el punto más alto del ataque. Cuando parece imposible que esto siga creciendo aun más, hay un cese abrupto que deja al descubierto un estado de vaciamiento, de nada. El paroxismo ha terminado y la música es tranquila, calma, acorde sostenido. El movimiento es reasumido después de este reposo, pero se está ahora en un nuevo mundo. Ondulaciones lentas y largas caracterizan la fase postictal creando un estado inquietante, casi nauseoso. Típicamente, esto puede durar 45 minutos hasta que la normal actividad cerebral se recobre.

El método para traducir señales cerebrales a música consiste en hacer que ellas modulen tonos sintetizados. La elección de la síntesis del canto establece una conexión con la "humanidad" de los datos. Nuestro coro de canales eCog "ejecuta" mediante cambios de altura, sonoridad, cualidades vocales y ubicación espacial.

4.2. Síntesis del Habla

La síntesis del habla, con sus amplias y variadas transiciones de alturas y fonemas, ofrece una auténtica prueba para la técnica de síntesis por formantes. Esa

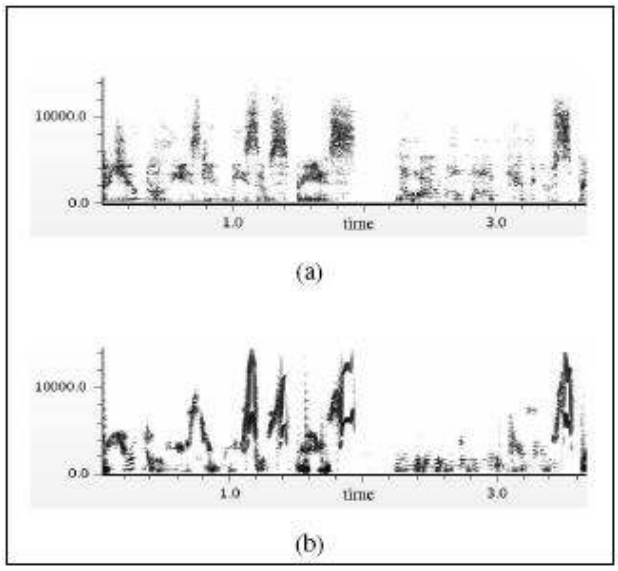

Figura 5. Análisis-resíntesis del diálogo: (a) es la entrada [input], discusión entre una hija adolescente y su madre, "can't you please give me some space", y "no, I will not give you some space", (b) Resíntesis por FM

prueba se llevó a cabo con una plataforma en "miniatura" para análisis-resíntesis, donde la síntesis es dirigida desde el canto y el lenguaje digitalizados. El analizador de trayectorias formánticas está escrito en Chuck y el sintetizador de formantes es una UGen (Unidad Generadora) de Chuck escrita en Faust. La parte del análisis emplea una FFT de ventana relativamente larga (4096 muestras) para detección precisa de formantes (a una frecuencia de muestreo de 48 kHz). La Fig. 5 compara, a modo de ejemplo, los espectrogramas de una entrada que consiste en un fragmento hablado y el de la salida resintetizada correspondiente obtenida con el método. En esta versión, la codificación de la señal consiste simplemente en registrar actualizaciones del parámetro del formante, las cuales son relativamente escasas (y que podrían optimizarse en gran medida). Los resultados son promisorios con vistas a desarrollar a partir de esto un codificador de voz humana basado en la FM. El ejemplo involucra a dos interlocutores diferentes que mantienen un diálogo acalorado. Sus voces e identidades se mantienen lo mismo que la inteligibilidad y expresividad de sus inflexiones. El análisis rastrea grupos de formantes de corta vida que, en el ejemplo, están limitados a cuatro a la vez (usando nueve osciladores en total).

4.3. Near the Inner Ear

La técnica de generación por formantes se aplicó recientemente en una composición para orquesta, música por computadora y video, estrenada en 2013 por la Stanford Symphony Orchestra bajo la conducción de Jindon Cai. La partitura, de Dohi Moon, se grabó y analizó

con un algoritmo de seguimiento de formantes [formant tracking algorithm]. También se obtuvo un análisis de trayectorias formánticas de una grabación de la Novena Sinfonía de Beethoven (ambas piezas se tocaron juntas al final de un ciclo completo dedicado a Beethoven). *Near the Inner Ear* incorporó fragmentos resintetizados de ambos conjuntos de datos obtenidos por análisis.

La composición usó resíntesis basada en formantes para generar una identidad tímbrica nueva cuya música está ligada a la escritura orquestal pero cuyo comportamiento es lo suficientemente diferente como para constituir una especie de 'alter ego' musical. La obra empieza con una sección de 90 segundos donde un intercambio antifonal explota tales contrastes. El instrumento de resíntesis vuelve en varios momentos de la pieza con gestos musicales que refuerzan el video acompañante de John Scott.

5. CONCLUSIÓN

Una meta hacia delante del sistema de análisis-resíntesis es crear una amplia base de datos de sonidos vocales adquiridos, con el fin de estructurar a partir de una tabla vastamente expandida una síntesis del canto basada en fonemas más compleja. Las aplicaciones arriba descriptas han demostrado un potencial para crear ricos reservorios de articulaciones e identidades tímbricas. El deseo es aprovechar una mayor variedad tímbrica en el trabajo de sonificación. Posiblemente, también adquirir rasgos vocales de los propios oyentes para usar en la síntesis controlada por EEG.

La resíntesis al servicio de proyectos musicales como *Near the Inner Ear* puede también modificar las estructuras acústicas con el fin de crear identidades instrumentales nuevas. El seguidor de formantes [formant tracker] que operó sobre el sonido orquestal en el proyecto *Near the Inner Ear*, infirió formantes donde ningún modelo hubiera predicho que éstos existirían. Experimentando con la resíntesis, se vio que el seguidor enfatizaría con frecuencia alturas de voces internas en las grabaciones analizadas. Más que imponer una altura fundamental f0, se permitió que las frecuencias formantes mismas pudieran llegar a ser f0. Los formantes que quedan crearían entonces una especie de "canto instrumental" cuyas resonancias se corresponden con la estructura de alturas de las grabaciones originales.

El presente método de síntesis puede usarse para sonidos no vocales cuya estructura acústica está también representada mediante resonancias de tipo formántico. Horner ha explorado la aproximación al timbre de una trompeta muestreada [timbre matching for a sampled trumpet] usando un algoritmo genético para hallar parámetros adecuados para formantes en FM [12].

Las mejoras detalladas en este artículo para la síntesis vocal por FM pueden extenderse a otros esquemas basados en la modulación a frecuencia de audio, particularmente a aquellos que también emplean una estructura de una sola moduladora y múltiples portadoras. Una síntesis vocal por AM libre de clicks, "pariente" de la aquí presentada, ha sido también implementada en Faust. La AM tiene la ventaja de que la predicción de la conducta dinámica de las bandas laterales es más simple (las bandas laterales AM están libres de la función de Bessel que determina las bandas laterales FM).

Reconocimientos

Muchas gracias a John Chowning por sus invenciones y estímulo, tanto técnico como musical. Bill Schottstaedt sigue trabajando en preservar para el futuro una enorme cantidad de instrumentos de síntesis y herramientas de análisis. Su proyecto Snd preserva y aporta algoritmos esenciales para la música por computadora sin los cuales mucho de este trabajo no hubiera sido posible.

6. REFERENCIAS

[1] J. Kleimola, "Nonlinear abstract sound synthesis algorithms", Ph.D. dissertation, Aalto University, Helsinki, Finland, 2013.

[2] V. Lazzarini and J. Timoney, "Theory and practice of modified frequency modulation synthesis," J. of Audio Eng. Soc., vol. 58, no. 6, pp. 459–471, 2010.

[3] J. Chowning, "Computer synthesis of the singing voice," en Sound Generation in Winds, Strings, Computers, J. Sundberg, Ed. Royal Swedish Academy of Music, 1980, pp. 4–13.

[4] ——, "Frequency modulation synthesis of the singing voice," en Current Directions in Computer Music Research, M. Mathews and J. Pierce, Eds. MIT Press, 1989, pp. 57–64.

[5] M. L. Brun, "Digital waveshaping synthesis," J. of Audio Eng. Soc., vol. 27, no. 4, pp. 250–266, 1979.

[6] "Scheme, Ruby, and Forth Functions included with Snd," última revisión 29 Mar. 2013. [Online]. Disponible: https://ccrma.stanford.edu/software/snd/snd/sndscm.html

[7] "FMVoices Class Reference, in The Synthesis ToolKit in C++," última revisión 29 Mar. 2013. [Online]. Disponible: https://ccrma.stanford.edu/software/stk/

[8] "Chuck : Strongly-timed, concurrent, and on-the-fly audio programming language," última revisión 29 Mar. 2013. [Online]. Disponible: http://chuck.cs.princeton.edu/

[9] "PeRColate, A collection of synthesis, signal processing, and image processing objects for Max/MSP," última revisión 29 Mar. 2013. [Online]. Disponible: http://music.columbia.edu/percolate/

[10] "FAUST (Functional Audio Stream)," última revisión Mar. 2013. [Online]. Disponible: http://faust.grame.fr/

[11] "Supporting online materials." [Online]. Disponible: http://ccrma.stanford.edu/_cc/vox/smac2013som/

[12] J. B. A. Horner and L. Haken, “Machine Tongues XVI: Genetic algorithms and their application to FM matching synthesis,” Computer Music J., vol. 17, no. 4, pp. 17–29, 1993.

7. PROGRAMA A

El programa Faust, FMVox.dsp [11], implementa un algoritmo para la síntesis de voz por FM con cuatro formantes que consisten en osciladores armónicos de fase uniforme leídos de una tabla [uniform phase table-lookup harmonic oscilators]. Un flujo de coeficientes multiplexados a frecuencia de audio controla la síntesis. Para cada formante se incluye un demultiplexador que extrae sus coeficientes del flujo de datos. Los formantes son realizados en forma paralela al proceso Faust que forma la salida como suma de sus señales. La unidad generadora resultante compilada por Faust puede tener un mayor número de formantes. Se puede crear una textura de múltiples voces usando múltiples flujos de control [control streams] independientes, que salen de unidades generadoras independientes. Usando esta arquitectura se puede distribuir las voces de un coro a través de múltiples subprocesos sincronizados a nivel de las muestras y/o múltiples núcleos.

```
declare name "FMVox";
import("filter.lib");
ts = 1 << 16;
fs = float(ts);
ts1 = ts+1;
ct = (+(1)~_ ) - 1;
fct = float(ct);
sc = fct*(2*PI)/fs:sin;
sm = fct*(2*PI)/fs:sin:/(2*PI);
dec(x) = x-floor(x);
pha(f) = f/float(SR):(+:dec) ~ _;
tbl(t,p)= s1+dec(f)*(s2-s1)
with {
f = p*fs;
i = int(f);
s1 = rdtable(ts1,t,i);
s2 = rdtable(ts1,t,i+1); };
fupho(f0,a,b,c) = (even+odd):*(a)
with {
cf = c/f0;
ci = floor(cf);
ci1 = ci+1;
isEven= if((fmod(ci,2)<1),1,0);
ef = if(isEven,ci,ci1);
of = if(isEven,ci1,ci);
frac = cf-ci;
comp = 1-frac;
oa = if(isEven,frac,comp);
ea = if(isEven,comp,frac);
ph = pha(f0);
m = tbl(sm,ph):*(b);
even = ea:*(tbl(sc,(dec(ef*ph))+m));
odd = oa:*(tbl(sc,(dec(of*ph))+m));};
frame(c) = (w ~ _ )
with {
rst(y)= if(c,-y,1);
w(x) = x+rst(x); };
demux(i,ctr,x) = coef
with {
trig = (ctr==i);
coef = (*(1-trig)+x*trig) ~ _;};
formant(f_num,ctlStream) = fsig
with {
ctr = frame(ctlStream<0);
co(i) = demux(i,ctr,ctlStream);
f0 = 1;
a = f0+1+f_num*3;
b = a+1;
c = a+2;
fsig = fupho(co(f0), co(a),
co(b), co(c)); };
nf = 4;
process = _<:par(i,nf,formant(i)):>_;
```

8. PROGRAMA B

El programa **FMVox** es usado por el programa Chuck FMVoxVib.dsp [11] para producir la Fig. 4. El ejemplo define una “shred master” que toca durante cuatro segundos. Prepara un gráfico DSP en que una instancia de FMVox recibe un flujo de datos de control sincrónico a nivel de las muestras y manda su salida al convertidor digital-analógico [dac]. La “shred master” usa el comando de Chuck “spork” para las “shreds” hijas creadas para el vibrato y el flujo de datos de control multiplexados. El arreglo de ‘floats’ data tiene los coeficientes para cuatro formantes de la vocal ‘aaa’, los cuales se describen mediantes sus amplitudes y frecuencias centrales. En este código de prueba, los anchos de banda de los formantes se establecen globalmente.

```
Step stream => FMVox fmv => dac;
4 => int nFormants;
1::ms => dur updateRate;
SinOsc vibLFO => blackhole;
vibLFO.freq(3);
vibLFO.gain(0.1);
Std.mtof(64) => float p => float f0;
fun void vibrato() {while (true){
((vibLFO.last()+1.0)*p) => f0;
1::ms => now;
```

```
}}
[ // "aaa"
[ 349.0, 0.0],[ 918.0,-10.0],
[2350.0,-17.0],[2731.0,-23.0]
] @=> float data[][];
fun void mux(float val) {
stream.next(val);
1::samp => now;
}
-1 => int startFrame;
95 => float db;
fun void muxStream() {
updateRate-14::samp => dur padTime;
while(true){
padTime => now;
mux(startFrame);
mux(f0);
for (0 => int f; f<nFormants; f++){
mux(Math.dbtorms(db+data[f][1]));
mux(0.2);
mux(data[f][0]);
}
}}
spork ~muxStream();
spork ~vibrato();
4::second => now;
```