

## Capítulo IV

### La captura gestual. Una posible implementación

Luis Tamagnini

#### I. Antecedentes.

La visión integradora de gesto musical que se plantea en el presente trabajo hace referencia a la importancia del gesto instrumental en los procesos de construcción de lenguaje que se manifiestan en la música instrumental, de cámara o sinfónica. Es esta inter-gestualidad kinética y sonora la que ha permitido a lo largo de la historia de la música el funcionamiento coordinado de estos organismos polifónicos.

Claro que, detrás de cada gesto instrumental individual existe un individuo, instrumentista que es quien comanda dicho gesto a partir de la interpretación de una partitura y de la interrelación gestual que establece con los demás instrumentistas.

Es obvio que esta interpretación, es exclusiva del intérprete y no puede ser realizada por una cinta o un gesto sonoro registrado en algún tipo de soporte fijo. Este hecho se manifiesta de forma explícita en la música electroacústica para medios mixtos sobre soporte, siendo esta una de sus principales características. En ella, el gesto instrumental esta subordinado al gesto sonoro fijo de la cinta y su flexibilidad es limitada.

Muchos desarrollos se han llevado adelante durante el siglo XX, en forma simultanea al propio devenir de la música electrónica y electroacústica, con el propósito de establecer una relación gestual mas profunda entre los instrumentos acústicos y los instrumentos electrónicos. Algunas de estas experiencias datan de las primeras décadas del siglo XX, como el **Theremin**, instrumento que permite captar un gesto kinético y transformarlo en una señal de corriente continua, de voltaje variable con la cual es posible controlar un oscilador u otro dispositivo analógico.

Durante los años 60 y 70 esta tecnología experimentó una notable expansión dando origen a toda una cantidad de instrumentos que pueden ser conectados o “patcheados” entre si y en donde se establecen relaciones de

control gestual a partir de algún tipo de interfaz como un teclado o un micrófono. Un caso bastante significativo de esta aplicación es el **vocoder**, instrumento cuyo sonido es el producto de una relación inter-gestual entre un sintetizador controlado (o no) por un teclado o un secuenciador y la señal acústica captada por un micrófono.

La aparición de las computadoras personales y de las tecnologías digitales, en los años 70 y 80, abrieron un nuevo y vasto horizonte, no solo para el diseño de nuevos instrumentos musicales electrónicos sino también para el desarrollo de nuevas formas de interconectarlos y controlarlos a partir del gesto instrumental.

Es sin duda el **MIDI (Musical Instruments Digital Interface)** una de las tecnologías mas influyentes de esos años. Este protocolo de intercambio de información codificada en forma digital tuvo, desde principios de la década del 80 y hasta nuestros días un impacto notable en la industria de los instrumentos electrónicos generando de manera inevitable, similar impacto en la praxis musical contemporánea. Existe en la actualidad una incontable cantidad de dispositivos que permiten la captura y codificación MIDI de casi cualquier tipo de gestualidad instrumental así como también del mero gesto kinético. Casos arquetípicos de estos dispositivos son los teclados, triggers de percusión, controladores de viento, conversores "pitch to midi" etc.

Algunos de estos dispositivos, como el conversor "pitch to midi" funcionan a través de la implementación en un **DSP (Digital Signal Procesor)** de un programa o algoritmo de análisis que permite extraer datos mas o menos precisos del gesto sonoro, como su nivel dinámico general o la distribución de esta energía en las diferentes bandas de frecuencia.

## **II. Panorama actual. Desarrollos desde finales de los 90 hasta la nuestros días.**

Casi simultáneos al surgimiento del MIDI son los esfuerzos por aplicar algoritmos de análisis directamente sobre computadoras personales. Son conocidos, entre otros, los desarrollos de **Miller Puckette** quien a mediados de los 80 escribió una aplicación originalmente llamada Patcher, piedra fundamental de sistemas actuales muy extendidos como **MAX/MSP** o **Pure Data**. Esta aplicación fué utilizada por primera vez en **Plutón**, una composición de **Philippe**

**Manoury** para piano y ordenador. En ella, se sincroniza el ordenador con el piano y se controla un hardware que procesa la señal de audio en tiempo real.

Hacia fines de los 90, el desarrollo de computadoras cada vez mas veloces e interfaces de audio mas potentes y accesibles, conjuntamente con el igualmente necesario desarrollo del software para su utilización, colocaron a la computadora en una posición de privilegio, como plataforma para el diseño, no sólo de nuevos instrumentos musicales electrónicos sino también de completos sistemas complejos donde se pueden establecer relaciones gestuales entre dispositivos, tanto dentro como fuera del ordenador. Un elemento clave en este campo es la capacidad de las computadoras modernas de realizar una representación digital de una señal de audio. Esta representación es realizada por un dispositivo llamado "conversor analógico digital" comúnmente presente en cualquier computadora y naturalmente en interfaces de audio dedicadas.

La información provista por estos dispositivos es fundamental para el análisis del gesto sonoro y también para la transformación del mismo a partir de la aplicación de algoritmos escritos para tal fin. Estos algoritmos pueden ser desarrollados en cualquier plataforma, programando directamente en lenguajes de aplicación general (todos los cuales proveen diversas librerías para ello) o utilizando plataformas dedicadas de alto nivel de abstracción, como las antes mencionadas **MAX/MSP** y **Pure Data** entre muchas otras de uso extendido, como **C-Sound**, **SuperCollider**, **Chuck**, **Processing**, **OpenMusic**, etc. Todas ellas permiten el acceso no solo a la información que proveen los conversores analógico digitales sino también a otros periféricos y dispositivos del sistema y el sistema operativo, todo lo cual permite el desarrollo de aplicaciones con un alto nivel de sofisticación y personal originalidad.

La digitalización de una señal analógica continua, como hemos visto antes, consiste en la representación de la misma mediante una secuencia de números o símbolos. Esto puede ser realizado en varios "dominios", siendo comúnmente utilizadas en audio digital las representaciones en el **dominio temporal** y en el **dominio frecuencial**. La representación en el dominio temporal muestra las variaciones de la señal de audio en función del tiempo y codifica estas variaciones en un sistema de variables discretas. A partir de esta representación temporal, la señal puede ser transformada, comúnmente mediante la aplicación de algún tipo

de filtro o retardo. Con esta técnica se desarrollan los algoritmos que implementan los efectos y procesadores elementales como delay, filtros, reverberación, etc.

La señal puede ser convertida desde el dominio temporal al dominio frecuencial mediante la aplicación de la **Transformada Rápida de Fourier (FFT)**. Mediante este algoritmo, se convierte la información de la señal en una serie de datos referidos a la frecuencia y fase de sus componentes individuales, permitiendo así reconstruir una representación espectral de la misma. Esta técnica permite la implementación de otro tipo de operaciones sobre la señal, por ejemplo la **convolución**, que consiste en la multiplicación de las componentes espectrales de dos sonidos, de uso muy extendido actualmente en síntesis y en simulación de espacios acústicos.

Otro aporte fundamental de esta técnica es que permite realizar un análisis de las componentes espectrales de un sonido y determinar de esta manera cual es su frecuencia fundamental o componente con más energía, lo cual es inicialmente exacto para sonidos de altura definida (o espectro armónico), y resulta aproximado (en diferentes grados) para otros tipos de espectros. Este análisis es determinante para la **detección de altura**, uno de los problemas referidos a la captura del gesto sonoro más estudiados en la última década.

### **III. Captura y mapeo del gesto instrumental en “Triple Canon”, para violín y procesamiento en tiempo real de audio y video.**

**Triple Canon** es una composición para violín y procesamiento en tiempo real de audio y vídeo. En ella se realiza una captura sonora del gesto instrumental, extrayéndose de éste una serie de datos que son luego utilizados para controlar parámetros que comandan en tiempo real, diferentes procesos de transformación de la señal del propio violín. Estos datos también son utilizados para comandar la reproducción y mezcla de señales de video generadas en el ordenador. La pieza fué compuesta durante el año 2010 por Luis Tamagnini y esta totalmente implementada en MAX/MSP/Jitter.

Es evidente que la captura del gesto instrumental así como el análisis de su información significativa está determinada por la naturaleza acústica del instrumento cuyo gesto intentamos captar. Existen actualmente escritos para MAX/MSP varios objetos que pueden ser utilizados para la captura del gesto

instrumental, tanto de instrumentos de altura determinada como para aquellos de altura indeterminada. En el caso de Triple Canon utilizamos el objeto **pitch~**, desarrollado por **Tristan Jehan**, el cual esta basado a su vez, en un desarrollo previo de **Miller Puckette**, otro objeto llamado **fiddle~**.

La señal acústica del violín es captada por un micrófono de condensador y enviada al conversor analógico digital, MAX/MSP accede a la señal digitalizada en forma directa. La misma es analizada por el algoritmo que podemos observar en la figura 1. El objeto **pitch~** recibe la señal digital y realiza un análisis FFT, el cual puede ser configurable mediante argumentos que determinan el funcionamiento del mismo, y nos entrega una serie de datos. La frecuencia fundamental en Hz, la amplitud general en dB, una lista de parciales, la detección del ataque del sonido, entre otros. La velocidad con la que se realiza este análisis está determinada por la configuración del objeto, y es inherente a la propia naturaleza de la aplicación de la FFT. En este caso usamos un tamaño de ventana de 4096 muestras, lo que determina que obtendremos un nuevo conjunto de datos cada 92 milisegundos aproximadamente.

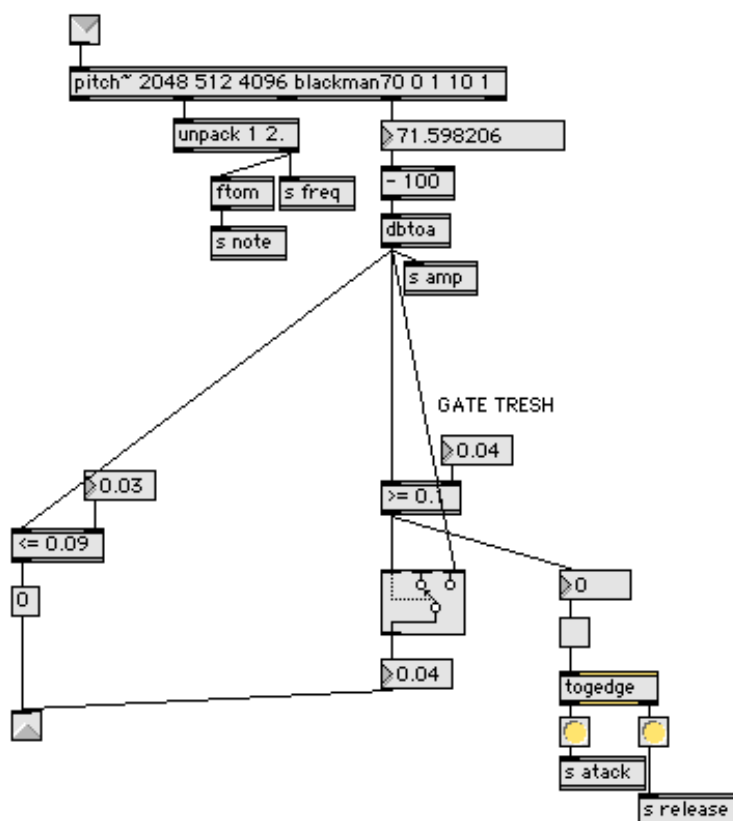


Figura 1.

Como se puede observar en la figura 1, obtenemos directamente del análisis del pitch~, una lectura de frecuencia en valor de punto flotante, (frecuencia fundamental) y una lectura de amplitud general en valor de punto flotante expresada en decibeles.

La lectura de amplitud es luego escalada a valor de amplitud lineal y utilizada para controlar una compuerta que permite filtrar datos generados por debajo de un cierto nivel dinámico. Hemos observado en la práctica que la detección de altura es mas eficiente durante el sostenimiento del sonido y es mas proclive a realizar falsas lecturas durante los transientes de ataque o extinción del mismo. La implementación de esta compuerta permite además el registro del ataque y extinción del sonido y su representación como eventos instantáneos. (trigger o disparo).

En la figura 2 se puede ver toda la información entregada por el pitch~ dispuesta para ser consultada durante la composición del trabajo. Estos parámetros significativos son, la lectura lineal de amplitud, la frecuencia fundamental y su expresión en altura musical temperada, indicadores para el ataque y la extinción, y una ventana que muestra un seguimiento de la envolvente de amplitud.

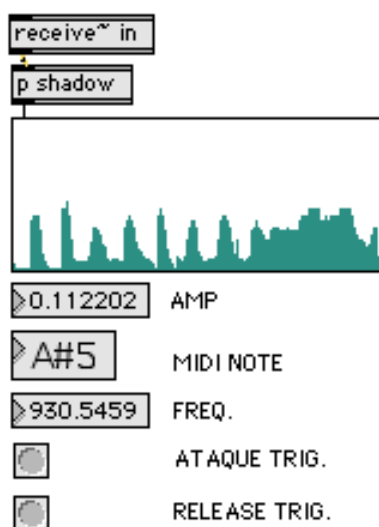


Figura 2.

Los datos obtenidos en el proceso de la captura gestual o parametrización pueden ser ahora directamente utilizados como parámetros de entrada para

controlar el funcionamiento de dispositivos de cualquier naturaleza, o ser mapeados para obtener variaciones paramétricas con distintas magnitudes de escala y resolución. Este último caso es aplicable cuando los parámetros de entrada de los dispositivos controlados operan con magnitudes o en rangos numéricos diferentes a los parámetros directamente obtenidos de la captura gestual.

En Triple Canon se presentan ambas situaciones, que a continuación se detallan en algunos breves ejemplos.

En la figura 3 se puede ver de que modo la lectura de frecuencia fundamental es utilizada en forma directa o previamente escalada por un valor de punto flotante para determinar la frecuencia de un oscilador sinusoidal cuya forma de onda es utilizada para realizar una modulación en anillo, es decir una modulación de la amplitud de la señal del violín. De esta manera podemos adecuar la frecuencia del modulador a la frecuencia de las notas ejecutadas por el instrumento.

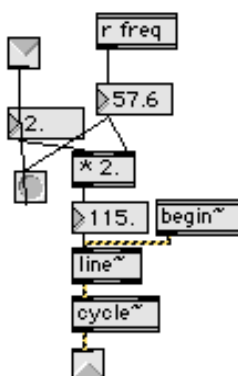


Figura 3.

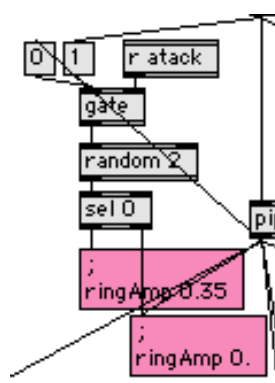


Figura 4

En la figura 4, la aplicación directa del trigger originado por el parámetro ataque sirve para disparar un evento aleatorio que determina la activación o no de un procesador.

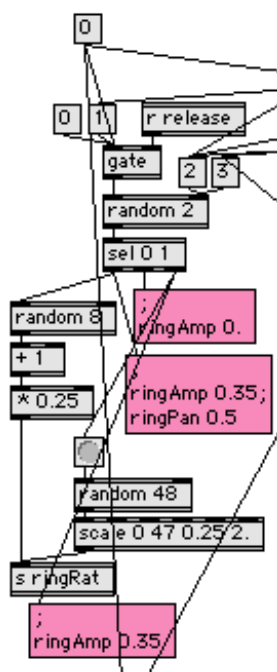


Figura 5.

En un sentido similar, puede verse en la figura 5 la utilización del parámetro release para el disparo de procesos aleatorios mientras que en la figura 6 se observa cómo el parámetro ataque puede ser utilizado para generar saltos de cuadro en uno o varios reproductores de video.

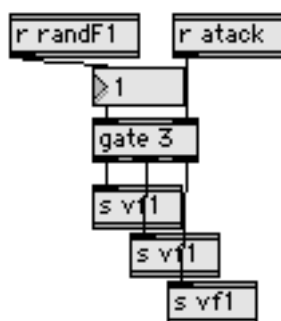


Figura 6.

Las figuras 7 y 8 muestran el mapeo de la lectura lineal de amplitud. En la figura 7, se utiliza para controlar la amplitud de un reproductor de audio mientras que en la figura 8 se observa un doble mapeo del mismo parámetro para realizar



un control independiente, diferente en cada caso, de la posición en la imagen estéreo de dos señales de audio.

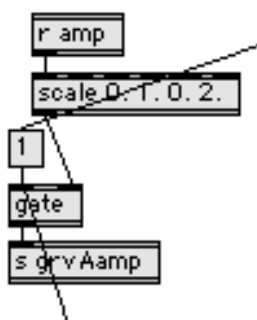


Figura 7

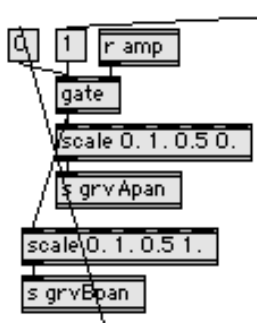


Figura 8.

Asimismo, los procesadores, también pueden generar parámetros continuos que son factibles de ser mapeados para generar otros parámetros. En este caso, una señal lineal de sincronización es generada por un reproductor de audio, ésta consiste en un valor de punto flotante que varía entre 0 y 1 o entre 1 y 0, dependiendo del sentido de la reproducción y que es utilizada para controlar la posición en la imagen estéreo de la señal de dicho reproductor (Figura 9).

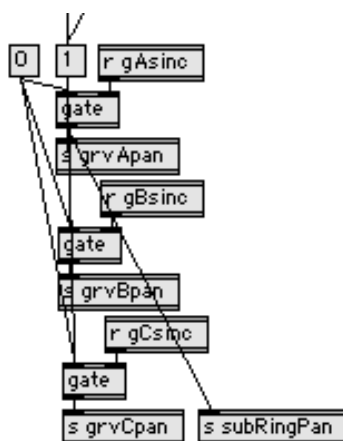


Figura 9.

Por último otra aplicación consiste en la utilización directa del valor de amplitud lineal de salida de ciertos procesadores para controlar la mezcla de las señales de video (Figura 10).

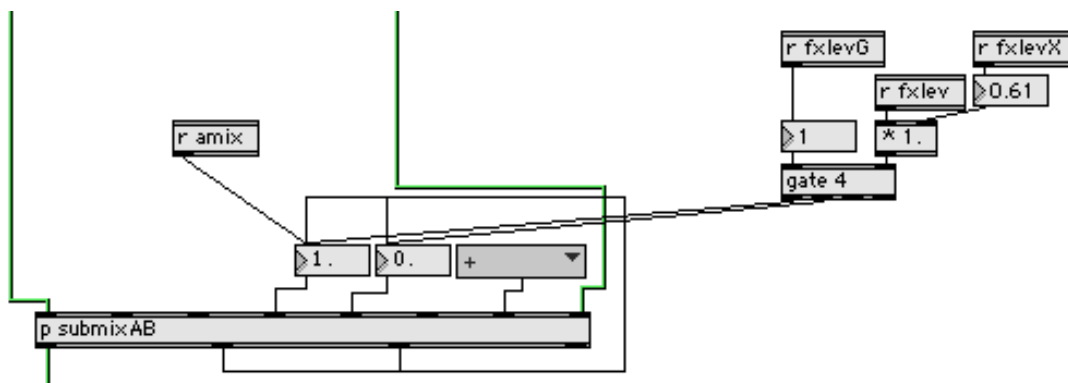


Figura 10.

#### IV. Bibliografía

Di Liscia, Oscar Pablo. Generación y procesamiento de sonido y música a través del programa C-sound. Universidad Nacional de Quilmes Editorial. Buenos Aires. 2004

Lluan, Claudio, Data, Gabriel y Tamagnini, Luis: A theoretical and aesthetics approach to the study and practice of mixed Electroacoustic Music: a pedagogical proposal. Electroacoustic Music Studies. <http://www.ems-network.org/ems09/proceedings.html>. Buenos Aires. 2009

Puckette, Miller y otros. Real-time audio análisis tools for Pd and MSP. reprinted from Proceedings. ICMC. 1998.