

O ~ A  
Body Loops

Sebastián González Dixon  
Javier Chávarri Alvarez

Master en Artes Digitales  
Universidad Pompeu Fabra de Barcelona  
Noviembre 2009

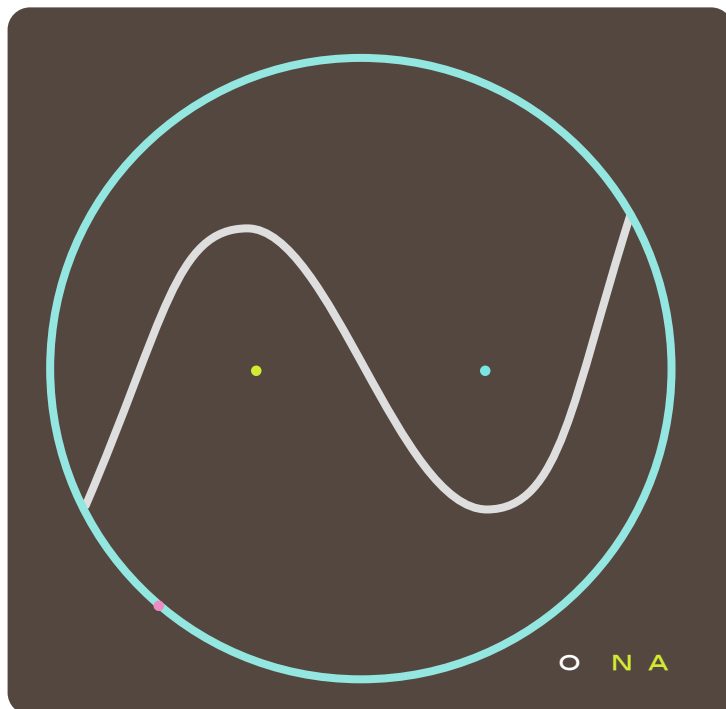
## Ficha del proyecto

Título: **O~A**

Autores: Sebastián González Dixon y Javier Chávarri Alvarez

Tutores: Carles Sora y Sergi Jordà

Descripción del proyecto: O~A es un sistema de captura de movimiento que rastrea las extremidades del cuerpo, sin usar marcadores ni dispositivos, para aplicaciones escénicas. Implementa un sampler de gestos, propio del sistema, que permite sintetizar, grabar, reproducir y controlar parámetros a través de expresiones corporales.



Web: <http://www.wearewaves.info/ona>

## Agradecimientos

Agradecemos a los tutores Carles y Sergi por sus recomendaciones y sugerencias a lo largo del proceso.

También al resto de profesores del MAD, que abrieron nuestras miras y han contribuido en el trazado de nuestros caminos.

Queremos agradecer especialmente a los directores del master Jose y Eugenio porque sin su apoyo no podríamos haber disfrutado de las experiencias de este año tal y como lo hemos hecho.

A Carmen, por saber lidiar en el tira y afloja entre los dos bandos.

Y por supuesto al resto de nuestros compañeros del máster, que han sido sin duda el mejor hallazgo del año.

Finalmente no pueden faltar en estos agradecimientos las personas que nos apoyan y soportan cada día: a nuestras familias, a Natalia y a Charlotte.

¡Gracias!

# Indice

1. Proceso de creación	5
1.1. Puntos de vista, perspectivas e intuiciones	5
1.1.1. Caminos abiertos para las creaciones digitales en torno a la música electrónica y las artes escénicas	
1.1.2. Diversos acercamientos al problema	
1.2. Somos ondas	7
1.2.1. Experiencias y antecedentes	
1.2.2. Motivaciones e inquietudes	
1.3. O~A	9
1.3.1. Generalidades del sistema O~A	
2. Proceso de producción	11
2.1. Elección de las herramientas	11
2.2. Puesta en marcha del sistema de visión	11
2.3. Macrocontrol y microcontrol	12
2.4. Descripción del software desarrollado. Arquitectura	13
2.4.1. Visión	
2.4.2. Audio	
2.4.3. Protocolo GDO (Gesture Data for O~A)	
2.5. Diagrama de montaje	20
3. Conclusiones	21
3.1. Autocrítica	21
3.2. Desarrollos futuros	22
4. Bibliografía y referencias	23
4.1. Papers	23
4.2. Web	24

# 1. Proceso de creación

## 1.1. Puntos de vista, perspectivas e intuiciones

### 1.1.1. Caminos abiertos para las creaciones digitales en torno a la música electrónica y las artes escénicas

Ya hace tiempo que la música y las artes visuales le sacan provecho a la computación para plasmar sus ideas. Por el contrario las artes performativas aún no la han explotado. Las incursiones de las artes escénicas en tecnología han sido a través de la escenografía. Existen, hace tiempo, diversos sistemas digitales que controlan la iluminación, los sistemas sonoros, los visuales e incluso mecanismos robóticos y vestuarios. La incorporación del cuerpo dentro del sistema digital ha permanecido distante e inmaculada.

En la otra mano está la música electrónica. Esta ha evolucionado mucho desde su aparición usando el medio digital de forma espléndida. Sin embargo carece de lo que a la anterior le sobra: evolucionar en la forma en la que se relaciona con el autor, el intérprete y el espectador.

La tercera pata, para no quedarnos cojos, es cuestionar el uso que le damos al medio. La expresividad de los artistas digitales está limitada por la precisión de la interfaz que estén usando. Esto ha brindado una estética particular a las creaciones digitales en cada era tecnológica. El pixel y el sample se han encargado de demarcar los límites del imaginario digital. A pesar de la gran capacidad de los ordenadores actuales, pareciera que la estética digital continuase añorando aquellos tiempos en los que era necesaria maquinaria pesada para calcular unos cuantos cientos de datos. Tal vez el hombre se acostumbró a que el sonido electrónico estuviese cuantizado y que las imágenes digitales estuviesen pixeladas.

Nos preguntamos entonces:

- Si es el cuerpo del performer es el que se está exponiendo en las artes escénicas, ¿por qué no apostar en el desarrollo de herramientas que lo integren a los entornos digitales sin traumatismos?
- Si la música se ha nutrido desde tiempos ancestrales de la gestualidad de los intérpretes, ¿cómo no otorgarle esta capacidad a su heredera digital?
- Si los sistemas digitales ya han sobrepasado en resolución a nuestra percepción del mundo analógico y continuo, ¿qué esperamos para usarlos en todo su potencial dentro de las artes digitales?

### 1.1.2. Diversos acercamientos al problema

Nos remontamos para empezar al Theremin, un instrumento electrónico pionero creado por Leon Thérémín hace casi un siglo, que proporciona una respuesta sonora de los gestos realizados cerca de su par de antenas. El Theremin estableció una ruptura en la concepción de instrumentos musicales, al ser el primer instrumento que “se toca sin tocarlo”. Este hecho podría parecer contradictorio, y llevar a pensar que es imposible llegar a un grado de virtuosismo a través de la práctica, ya que no existe una respuesta (o feedback) físico por parte del instrumento. Sin embargo, el Theremin ha ganado su puesto históricamente gracias a músicos como Clara Rockmore o Samuel Hoffman, y a día de hoy siguen naciendo nuevos intérpretes.

Más recientemente Atau Tanaka propone en su trabajo *Sensors\_Sonics\_Sights*, junto a Cecile Babiole, Laurent Dailleau, un sistema de sonido que consta de tres interfaces/instrumentos diferentes, entre ellos un Theremin, que leen gestos realizados con las manos de los intérpretes y los transforman en salidas sonoras y visuales. En este caso se logran buenos resultados en términos de mapeo entre gesto y output audiovisual, al mismo tiempo que los intérpretes captan la atención del público con sus acciones. Sin embargo los sistemas usados son sensores atados al computador, lo que reduce la movilidad de los músicos solo al movimiento de sus manos alrededor de los dispositivos. Por esto el sistema se limita, pues al igual que en la interpretación de cualquier instrumento, es de suma importancia el uso de todo el cuerpo para dar la intención al sonido final.



*Fig. 1: Atau Tanaka. Sensor\_sonics\_Sights*

Dispositivos como joysticks, tabletas gráficas, guantes de realidad virtual e incluso trajes exoesqueléticos como los usados por de Marcel.Í Antúnez han sido usados para contrarrestar la rigidez del teclado y el ratón. Sin embargo siempre han tenido limitaciones mecánicas y ergonómicas, por lo cual no han tenido una acogida universal como sistemas de registro de la expresividad.

Un trabajo muy representativo en este sentido es *Very Nervous System* de David Rokeby. Mediante el uso de cámaras, procesadores y sintetizadores, el sistema creado a finales de los años 80, planteaba la creación de música y/o sonido haciendo uso del propio cuerpo. Aquí Rokeby habla de la invisibilización de la interfaz, que es difuminada en una larga extensión de espacio, contrariamente a la mayoría de interfaces que están focalizadas y bien definidas.



*Fig. 2: David Rokeby. Very Nervous System*

Un desarrollo cercano y es el sistema iXKa, creado por el grupo de investigaciones en tecnología para las artes escénicas Konic THTR. El sistema iXKa es un dispositivo electrónico inalámbrico de captación de datos físicos que usa un sistema de reconocimiento de secuencias de movimiento creado por investigadores del IRCAM. Este dispositivo cuenta con un acelerómetro tridimensional (similar al mando de control de la video consola Nintendo Wii) que mide la dirección del movimiento relativa al dispositivo. El sistema de reconocimiento de secuencias, que recibe las señales de iXKa, relaciona sonidos con movimiento precisos. Una vez establecidas estas relaciones, los sonidos son reproducidos si se ejecuta este mismo movimiento, pudiendo así alterar la velocidad de ejecución del sonido a voluntad. Este sistema está siendo usado en obras escénicas, aún de forma experimental. El sistema iXKa tiene como ventaja el hecho de no ser afectado por la iluminación, como ocurre con los sistemas basados en visión artificial, lo que es provechoso en términos escenográficos. Sus grandes limitaciones son el tamaño, que aún no ha sido optimizado para ser cómodamente vestido, y el hecho de que solo registra el movimiento de una extremidad a la vez, según donde se vista.



*Fig. 3: Konic THTR, sistema iXKa*

Entre los muchos proyectos que se están gestando alrededor de este tema nos concentramos en el estudio de uno en particular: el proyecto WeAreWaves. WeAreWaves es una instalación sonora/interactiva realizada por los autores de este trabajo. Este proyecto es un antecedente muy cercano con el cual ganamos experiencia y pudimos sacar muchas conclusiones que llevaron a la formulación del proyecto O~A.

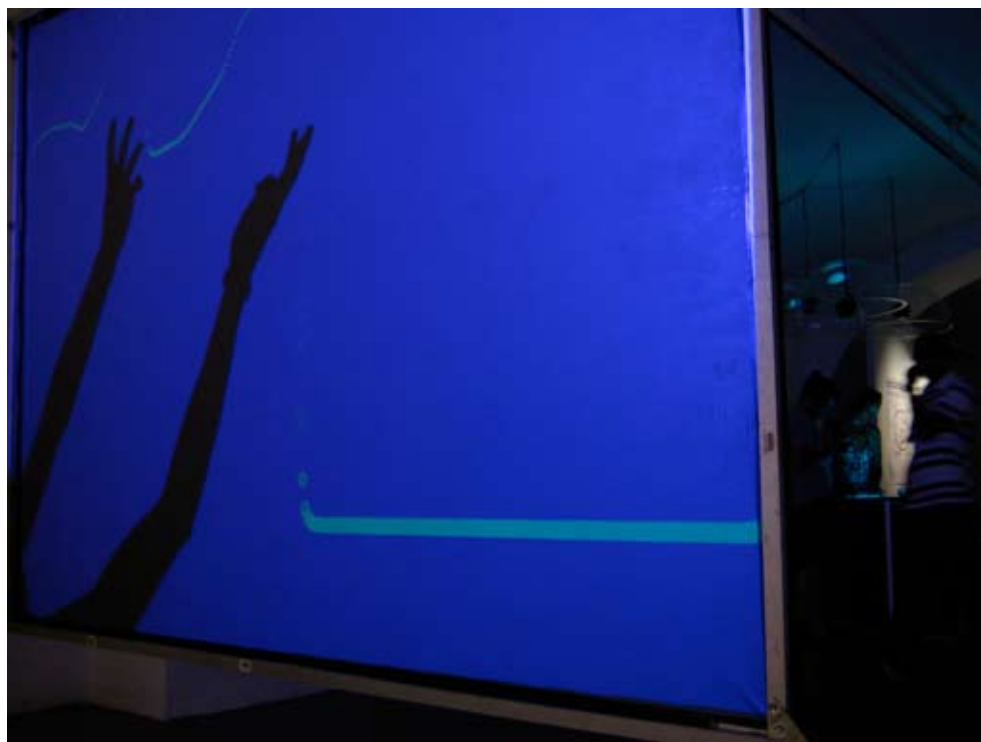
## 1.2. Somos ondas

### 1.2.1. Experiencias y antecedentes

El proyecto WeAreWaves es la piedra angular del proyecto O~A. WeAreWaves es un modelo sencillo y efectivo de instalación, que invita a los visitantes a explorar con sus cuerpos desde el momento en que descubren lo que se les propone. Desde su primer prototipo, realizado como un trabajo de clase durante el master de artes digitales, el proyecto demostró su potencial. WeAreWaves no surgió tras el resultado casual de un experimento ni por un error en el código. Su creación fue consecuentemente la unión de nuestras inquietudes y necesidades por oír el resultado de eso que nos estábamos preguntando: ¿Como sonaría el cuerpo si fuera una onda? Nuestras experiencias en programación, visión artificial, música y danza confluyeron y al cabo de un par de semanas teníamos la respuesta. Sabíamos que esta pregunta podía ser resuelta de diversas formas, para lo cual optamos por nuestra propia estrategia. En el caso de WeAreWaves resolvimos que fuese el contorno superior de la silueta el que modificara la onda. El mapeo entre imagen y sonido se planteó como una transformación directa. De esta forma el valor de

la amplitud de cada punto del sonido tomaría el valor de la altura de la silueta en ese lugar de la imagen y así se hizo. Como consecuencia teníamos una onda estacionaria que sería modificada en tiempo real por las formas corporales que estuvieran en el espacio. Si estas formas variasen, entonces, variaría el timbre del sonido producido. Si por el contrario no hubiese formas, no habría sonido alguno.

Al día de hoy la instalación ha sido expuesta en 6 festivales, en Barcelona y Cadaqués en España y Avellino en Italia. Esto ha representado la visita de más de mil personas, de todas las edades y en ambientes diversos tanto culturales, como familiares, pedagógicos y festivos. Estas oportunidades únicas nos han dado experiencia, proporcionado retroalimentación de las personas que lo han usado y brindado asombro por las situaciones infinitas que se producen alrededor de una propuesta semejante. Vimos muchas escenas interesantes, algunas incluso emocionantes, de personas que al entrar en el espacio, cuidadosamente adecuado para la interacción, transformaban su actitud cotidiana y empezaban a realizar movimientos que seguramente ni siquiera ellos mismos sabían que podían. Jugaban, saltaban e incluso interpretaban piezas de movimiento que nacieron y murieron con el momento. Estas vivencias nos han provocado muchas nuevas preguntas y necesidades, las cuales empezamos a abordar desarrollando el proyecto O~A.



*Fig. 4: WeAreWaves @ Sonar2009*

### 1.2.2. Motivaciones e inquietudes

Una inquietud primordial que surgió a raíz de las presentaciones de WeAreWaves fue con relación al tiempo y al ritmo. ¿Como podríamos incorporar el tiempo en las acciones que se desarrollan al interactuar con una interfaz corporal?

Nos inquietaba mucho la pérdida constante de momentos espléndidos que se daban en la instalación WeAreWaves, tanto a nivel sonoro como a nivel de expresión corporal del visitante. Además de simple, WeAreWaves es efímera.

En WeAreWaves el visitante afecta el espacio con su huella sonora particular. Las vibraciones que en los altavoces se reproducían tenían la forma de la persona. Esto ha resultado ser muy impactante para quienes se han interesado en conocer el sistema más allá de la simple experiencia. La idea de que los seres somos ondas que afectamos el espacio es en particular una de nuestras principales inspiraciones.

Además, a través de WeAreWaves estábamos proponiendo un método de síntesis de formas de onda en tiempo real, el cual podría ser sacado de las cuatro paredes de una instalación y llevado a la música. Imaginamos conciertos en los que los



bailarines producían los sonidos sintéticos que eran interpretados por los músicos (y en consecuencia que si no hubiesen bailarines entonces no habría sonido). A partir de este imaginario empezamos a pensar en un sistema que grabase los gestos de movimiento, así como los resultados sonoros que se producen tras una interacción con el sistema, para así no perderlos y poder usarlos a posteriori para composiciones musicales. Con esto podríamos acondicionar el sistema para que pudiese ser parte de un espectáculo en vivo.

Una observación sobre los sistemas gestuales de visión artificial usados para artes digitales, es que suelen carecer de métodos de grabación de las acciones. Tal vez no se contaba con la tecnología adecuada para abordar este problema anteriormente pero esto ya no es una excusa. La mayoría de sistemas interactivos leen al usuario y lo analizan en tiempo real, en un flujo continuo de datos derrochados. Surgió entonces la necesidad de recuperar estos datos para analizarlos y para darles usos posteriores.

Podrían añadirse otras muchas motivaciones. A continuación enumeraremos solo algunas cuantas:

1. La relación sonido - cuerpo
2. El gesto como vía de creación
3. La grabación del movimiento
4. La postura y el gesto como lenguaje de comunicación con la maquina
5. El cuerpo como interfaz
- 6 La interfaz natural
7. El loop corporal

## 1.3. O~A

### 1.3.1. Generalidades del sistema O~A

Queríamos una herramienta de creación e interpretación musical que funcionase a través de las formas y la cinética del cuerpo. El sistema O~A está diseñado para ser usado en escena por un intérprete especializado, al contrario de WeAreWaves que es una instalación abierta a todo tipo de usuario. Nos ubicamos en el límite entre la música y la danza. Para el público, una demostración de O~A debe ser entendida como un concierto o como una performance, indiferentemente. El intérprete será tanto músico como performer. No se determina que el sistema realice un tipo de música en particular, ni que deba ser usado mediante una técnica de movimiento predeterminada. La principal expectativa es imprimir las cualidades de movimiento en los rasgos tímbricos, rítmicos y energéticos del sonido generado.

Otro factor importante es la capacidad de composición musical del sistema. Con el dominio del eje temporal podemos enmarcar las acciones de movimiento dentro de un sistema métrico de compases, similares a los musicales. Esto hace que las acciones grabadas se ajusten a longitudes fijas para que mantengan una relación temporal entre sí a lo largo de la interpretación.

A grandes rasgos O~A es un “sampler” de gestos. Un sampler convencional es un sistema que graba y reproduce fragmentos de sonido. Con O~A, logramos grabar fragmentos de movimiento, representados por las curvas gestuales de diferentes partes del cuerpo.

La descripción del cuerpo se realiza ubicando las extremidades cabeza, manos y pies, visibles para la cámara, y la posición del centro de gravedad de la persona. Estas curvas son rastreadas en tiempo real y sus valores son almacenados en arreglos numéricos y archivos de texto. Al mismo tiempo estas curvas son

enviadas hacia módulos de sonido, que han sido mapeados con anterioridad, los cuales generan sonidos sintéticos o procesan sonidos grabados, modificando sus múltiples parámetros con la misma dinámica que tiene o tuvo el intérprete.

O~A es un sistema modular, que permite a los intérpretes escoger sus propios sistemas de sonido y mapearlos según sus necesidades e intereses. El esquema de funcionamiento se ha diseñado para que sea el propio intérprete quien realice todas las funciones de control una vez inmerso en el sistema. De esta forma puede decidir que función desea que el sistema realice a continuación incorporando estas acciones dentro de su interpretación performativa.

El sistema tiene entonces dos modos generales: uno de control y otro de exploración gestual. Es importante resaltar que no buscamos un sistema donde el cuerpo reemplace lo que bien podría hacerse más fácilmente con un botón. Además creemos que los sistemas donde se usa el cuerpo como interfaz, no deben sugerir al usuario que sea un botón o un joystick humano. Teniendo esto en cuenta, encontramos en el lenguaje corporal la mejor forma de comunicar al intérprete con el instrumento. Por esto establecimos las señales de control en un conjunto de posturas de fácil ejecución. Además las posturas forman parte del guión coreográfico que guía el espectáculo.

## 2. Proceso de producción

### 2.1. Elección de las herramientas

Para resolver el problema empezamos por desarrollar el sistema de visión artificial, el cual es la columna vertebral del sistema. Este fue planteado desde cero, con procedimientos que ya conocíamos teóricamente, pero que deseábamos implementar en el lenguaje de programación C++, con librerías de OpenCV, para alcanzar su máximo potencial. Para esto incursionamos en la plataforma de programación openFrameworks, la cual nos permite abordar problemas de alta complejidad en programación, a un más bajo nivel de la máquina, sin preocuparnos por la compilación de librerías y otros menesteres propios de la programación en C++.

En su momento, la etapa de visión artificial del sistema WeAreWaves fue desarrollada en la plataforma Processing. Esto llevaba a la máquina a trabajar casi al 100% de la capacidad de procesamiento en ordenadores de última generación. Esto nos limitaba para hacer cualquier tipo de implementación posterior, pues el sistema colapsaría.

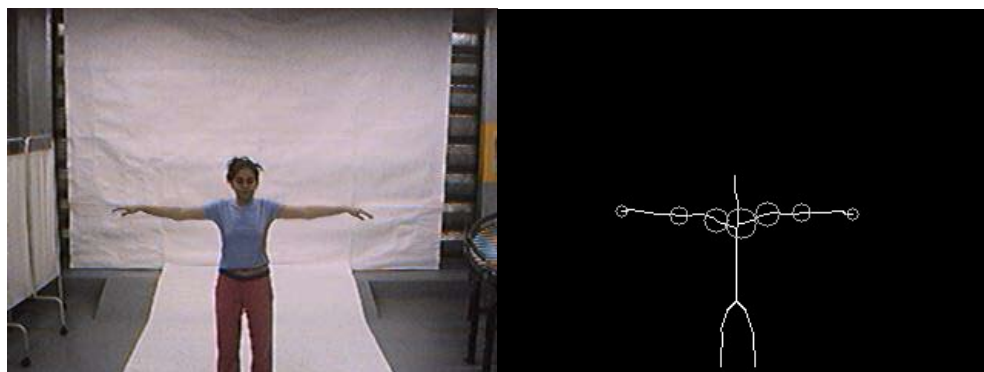
Mientras que Processing es un hijo de Java, openFrameworks es un gestor de proyectos de C++ directamente, el cual trae implementadas librerías básicas para temas de artes digitales, entre ellas OpenCV y OpenGL.

### 2.2. Puesta en marcha del sistema de visión

El sistema de visión nos tomó gran parte del tiempo de desarrollo. Debíamos analizar desde cero las lecturas de la cámara, para finalmente tener un sistema que rastrear la posición de las extremidades del cuerpo de una forma continua.

Empezamos por procesar imágenes estáticas en blanco y negro que dibujábamos con el ratón. Estas imágenes nos ayudaron a realizar los primeros experimentos. El primero fue realizar una versión de WeAreWaves que en lugar de tomar solo el contorno superior de la silueta, realizaba una lectura radial de toda la silueta. Se hizo midiendo las distancias desde el centro de gravedad del blob hasta su contorno y convirtiendo estos datos en una onda de sonido. Este experimento nos sirvió para dos cosas esenciales: implementar la comunicación entre OpenFrameworks y PureData y realizar un reconocimiento del lenguaje de programación que abordaríamos.

El siguiente paso fue implementar los algoritmos de esqueletización, los cuales habían sido programados previamente por uno de los autores del proyecto, como parte del sistema Vitruvia, ancestro del sistema O~A. Vitruvia es un sistema de visión artificial que reconoce un conjunto de 64 posturas de las extremidades superiores del cuerpo humano (la combinación de 8 posturas de cada brazo) y almacena las curvas de movimiento de las articulaciones de hombros y codos. El sistema Vitruvia fue realizado en el 2005, cuando la capacidad de procesamiento era casi 16 veces menor a la actual, con unos resultados bastante buenos. El sistema de esqueletización usado en Vitruvia implementa un algoritmo propio del proyecto, bautizado Esqueletización Cristal, el cual soluciona todos los inconvenientes que



*Fig. 5: Sistema VITRUVIA*

traen los algoritmos de esqueletización convencionales como ramificaciones incorrectas o quiebres en la estructura del esqueleto.

Una vez implementada la esqueletización surgió la necesidad de probarlo en vivo. No nos era posible montar un escenario para realizar pruebas reales. Afortunadamente durante el tiempo de desarrollo del proyecto tuvimos la oportunidad de realizar una muestra del sistema WeAreWaves. Esta fue una excelente ocasión en la cual presentamos en público los primeros experimentos del proyecto O~A. Un intérprete realizó varias demostraciones durante la jornada de exposición, de las cuales se realizaron vídeos con la misma cámara de captura del sistema. Estas grabaciones fueron fundamentales para poder desarrollar lo que es el sistema O~A en su fase actual.

A partir de los vídeos pudimos verificar el funcionamiento de la esqueletización en tiempo real y observar la concordancia entre las ramas del esqueleto y las extremidades del cuerpo visualizado. Seguidamente implementamos un sistema de detección de las extremidades, que aún simplemente estaban dibujadas en la imagen del esqueleto. Luego de detectadas debimos ordenar las detecciones de tal forma que en cada nueva captura se supiera cual extremidad era cual en capturas sucesivas. Con esto listo procedimos a dibujar las curvas de posiciones tanto en el eje X como en el eje Y.

Observamos el comportamiento de las curvas, de las extremidades, de las siluetas y del centro de gravedad; filtramos las señales obtenidas, promediamos valores, calculamos velocidades y aceleraciones. Analizamos las curvas y las imágenes una y otra vez hasta estar seguros de tener un sistema confiable, que nos estuviese retornando los datos esperados.

### 2.3. Macrocontrol y microcontrol

En términos de sonido, aunque suponíamos cómo deberían ser las señales obtenidas tras el rastreo de las extremidades, no alcanzábamos a dimensionar como podrían estas ser estas mapeadas para ser aplicadas a sintetizadores, filtros o secuenciadores. Aunque contábamos con algunos sintetizadores ya implementados, no teníamos una fuente de datos como la que introduciríamos

finalmente, por lo cual no podíamos saber a ciencia cierta como sonarían. Los experimentos de sonido fueron postergados múltiples veces por inconvenientes encontrados en el sistema de visión.

Finalmente, con el sistema de visión prácticamente terminado, fue posible vislumbrar la orientación exacta que podríamos darle al sistema de sonido. Los datos obtenidos son tratados en dos categorías: macrocontrol y microcontrol. El macrocontrol se encarga de manejar las funcionalidades del sistema, mientras que el microcontrol sería el encargado de los procesos sonoros. El macrocontrol esta regido por posturas estáticas y lecturas generales del cuerpo, mientras que el microcontrol toma los múltiples valores de la dinámica de movimiento de cada extremidad, o de elegir el sistema de partículas como entrada al sistema, lo que nos da una definición muy exhaustiva de los gestos del intérprete como entrada al sistema.

De esta forma podemos brindar al intérprete la libertad que se deseamos en sus movimientos de creación e incorporamos un lenguaje de comunicación entre el interprete y el sistema que se funde con la interpretación misma.

## 2.4. Descripción del software desarrollado. Arquitectura

La arquitectura de O~A ha sido diseñada de manera modular. Es decir, los datos obtenidos a través del sistema de visión han sido utilizados para la generación de sonido en este caso, pero se podría utilizar esta información a cualquier otro tipo de aplicación. Por ejemplo, para controlar vídeo en tiempo real de una manera compleja. Esto es posible gracias a la definición de un protocolo (GDO), explicado posteriormente.

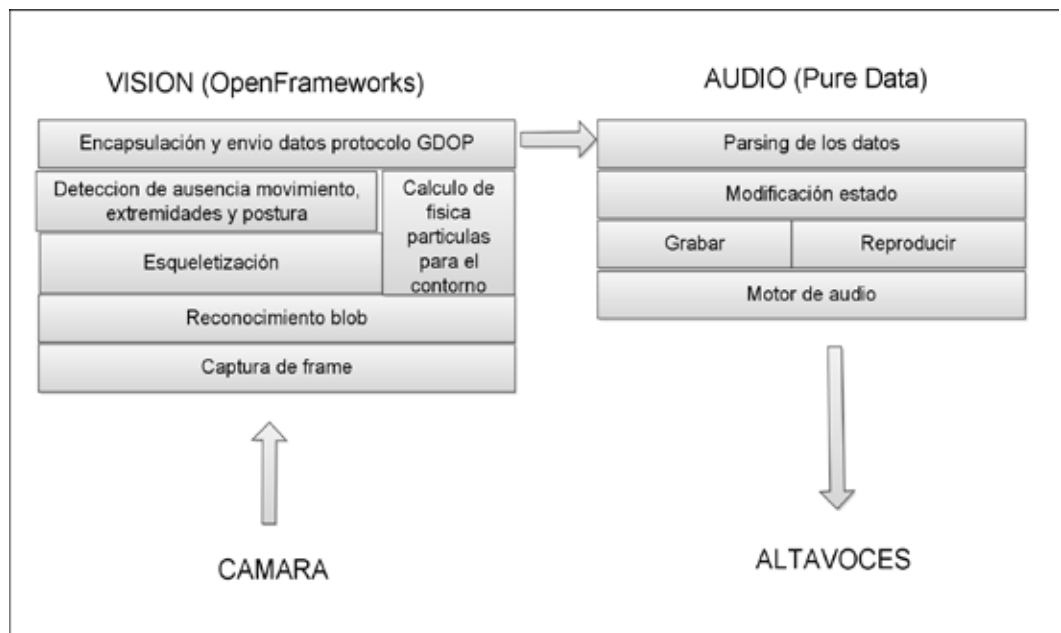


Fig. 6: Diagrama de la arquitectura de O~A

### 2.4.1. Visión

La parte de reconocimiento visual recorta la silueta del intérprete (*blob*), y procede a aplicar diversas operaciones sobre esta información:

#### Algoritmo de esqueletización

Nos da los puntos de extremidades, así como la identificación de posturas, y la detección de la ausencia de movimiento.

Para realizar la esqueletización se asume que la persona está representada por los píxeles blancos y el fondo por los píxeles negros en la imagen de entrada. Se realizan adelgazamiento sucesivos, encontrando los bordes según la evaluación de número de vecinos  $B(p)$  y conectividad  $X(p)$  de cada píxel para encontrar y borrar píxeles de la silueta en cada barrido. Seguidamente se implementa un par de condiciones, desarrolladas en este trabajo, que evitan borrar los píxeles que conforman el esqueleto. Para estas es necesario definir el vecindario de un píxel, numerando el vecino superior izquierdo como 1 y continuando en sentido de las manecillas del reloj. Se evalúa tanto en la silueta como en la imagen donde se marcan los candidatos de un ciclo.

P1	P2	P3
P8	Pi	P4
P7	P6	P5

C1	C2	C3
C8	Ci	C4
C7	C6	C5

Fig. 7: Vecinos en la imagen de a) la silueta y b) de la mascara de candidatos.

El adelgazamiento y las condiciones que mantienen unido el esqueleto forman un total de 4 condiciones, las cuales son:

1.  $2 < B(p) < 5$
  2.  $X(p)=1$
- Si  $B(p)=5$  no borre cuando:
3.  $(P3 + P4 + P5) = 0 \ \&\& \ C8 = 0$
  4.  $(P5 + P6 + P7) = 0 \ \&\& \ C2 = 0$

Es necesario para obtener buenos resultados filtrar la silueta para eliminar píxeles ruidosos en el contorno y en el interior de la silueta, puesto que estos píxeles suelen generar falsas ramificaciones en el esqueleto. Un buen método encontrado es realizar un filtrado morfológico en el que se realiza una iteración de apertura (opening), para eliminar los píxeles del contorno, seguida de una iteración de cerramiento (closing), que cierre los huecos producidos por ruido dentro de la silueta.

**Posturas reconocidas**

Las posturas que reconoce O~A son 3, aunque este número es fácilmente ampliable:

- La postura 0 es reconocida cuando el intérprete o bailarín se encoge hasta que ninguna extremidad es reconocida, y la figura casi se puede interpretar como una circunferencia.
- La postura 1 es captada al permanecer erguido, con los brazos pegados al tronco.
- La postura 2 es captada al abrir los brazos y situarlos de forma que ambos queden orientados de forma simétrica respecto al otro, es decir, con las manos a la misma altura.



Fig. 8: Posturas 1 y 2 reconocidas por el módulo de visión de O~A

En cualquier otro caso, la parte de visión enviará como postura el número extremidades detectadas+1. De este modo, si tenemos los brazos abiertos y las piernas juntas tendremos 4 extremidades (2 de los brazos, 1 de la cabeza, y 1 de las piernas), y O~A reconocerá esto como la postura 5 (= 4+1).

### **Reconocimiento del blob de la figura del intérprete**

El blob reconocido es para nosotros otra fuente valiosa de información gestual (o de microcontrol), y se ha diseñado un sistema para reconocerlo y tratar esta información de manera estable.

El principal problema encontrado a la hora de implementar esta funcionalidad es que los blobs cambian de un frame a otro continuamente de tamaño. En un momento dado la figura reconocida puede contener 230 puntos y al siguiente 231. Esta ligera modificación que en principio puede parecer nimia, finalmente resulta bastante difícil de tratar, sobre todo por aplicaciones terceras que pueden recibir estos datos, y nunca se sabrá a ciencia cierta el tamaño de los mismos.

Por tanto, procedimos a implementar un sistema de partículas, en el que cada partícula se adaptase a un punto determinado del blob. Calculamos aproximadamente la media de puntos que se usa para definir el blob, y el número máximo, y así llegamos a la conclusión de que para una captura de 320x240 pixels es más que suficiente con un sistema compuesto por 1000 partículas.

En esta adaptación del blob a las partículas, se procede a la asignación a cada partícula de un punto particular en el blob, empezando por el punto situado a ángulo 0 (más a la derecha según el punto de vista del espectador).

El sistema de partículas, basado en el sistema utilizado en WeAreWaves, incluye además parámetros de control y configuración del mismo:

- gravedad
- masa de cada partícula
- elasticidad
- amortiguación

Con lo cual, la figura que obtenemos puede ser estrictamente la del intérprete o una “deformación” modelada físicamente del mismo, para los usos que se crean convenientes.

### **Envío de los datos a través del protocolo GDO por puerto UDP**

Finalmente, todos estos datos son encapsulados en mensajes GDO y enviados a la aplicación de audio.

### **Implementación y monitorización**

El procesado de la imagen recogida y su análisis ha sido uno de los procesos más complejos de la implementación de O~A.

Por un lado, el análisis de las extremidades no es evidente ya que se trata de un número que no siempre es constante, y además el concepto de extremidad es difícil de definir en términos de visión artificial. Se han implementado diferentes sistemas de suavizado de los datos para así poder reconocer las extremidades de una forma persistente y más adaptada al movimiento real que se está llevando a cabo por el intérprete.

Así, para la depuración y análisis de los datos recogidos, se ha procedido a la implementación de una interfaz gráfica que nos permite ver en cualquier momento el comportamiento del sistema. Esto nos ha permitido recoger patrones de comportamiento del sistema, y de cómo reacciona ante distintos eventos.

Por un lado, se puede observar la imagen que está siendo captada tal cual. Al lado se ve el resultado del procesado de las partículas y su adaptación al blob (explicado en el apartado 2), así como la curva temporal del desplazamiento del centro de gravedad del blob, en el eje X y en el eje Y.

Por último, en esta primera línea tenemos la imagen que nos muestra el resultado de la aplicación del algoritmo de esqueletización, así como puntos de diferentes colores que representan las extremidades y el centro de gravedad en tiempo real.

El siguiente apartado en la interfaz de monitorización son los datos precisos de cada extremidad. Se muestran los siguientes parámetros asociados a cada punto: distancia al centro de gravedad, ángulo con el centro de gravedad, velocidad en el eje x, velocidad en el eje y, velocidad radial, velocidad angular, aceleración en el eje x, aceleración en el eje y, aceleración radial, aceleración angular y dirección de movimiento. El objetivo era tener el máximo de información posible que pudiese describir el gesto o movimientos del intérprete, y así, todos estos parámetros pueden ser utilizados y enviados al sistema de sonido.

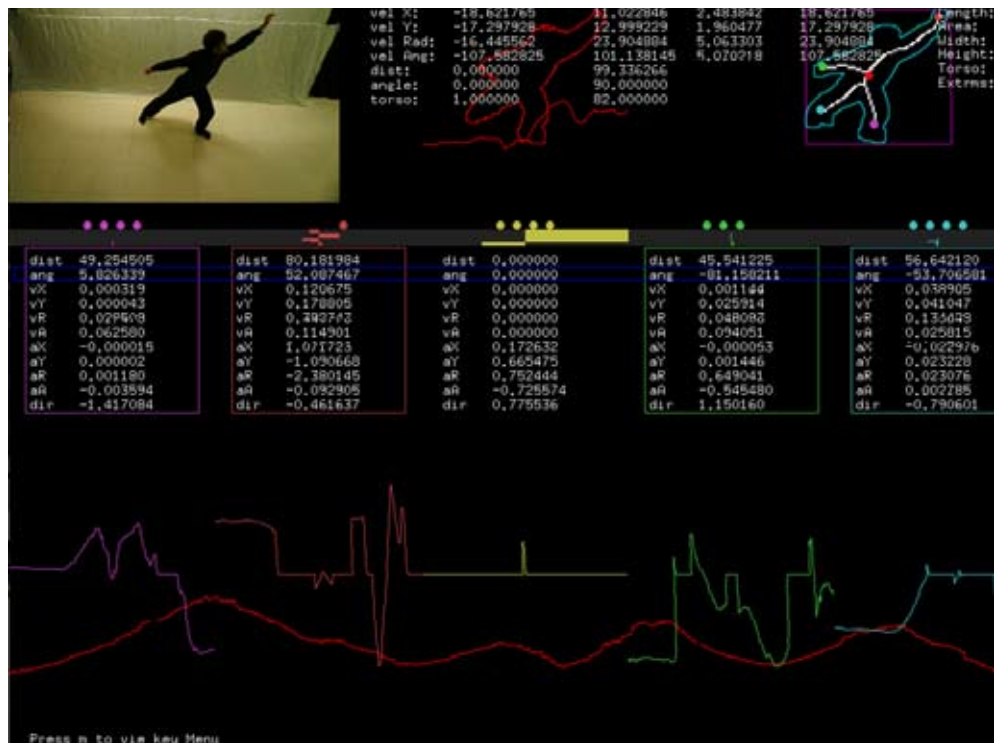


Fig. 9: Captura de pantalla del sistema de monitorización de O~A

Por último, se muestran las cinco curvas en el tiempo (los últimos 100 frames) de manera que se pueda tener una representación visual de la evolución temporal de las mismas

Y para acabar, se muestra la onda producida por el blob. Esta onda corresponde a las distancias desde el centro de gravedad de cada partícula del sistema de partículas, y la lectura se hace de forma radial en el sentido contrario a las agujas del reloj.

#### 2.4.2. Audio

En el caso de O~A, es el motor de audio quien gestiona el estado actual del sistema y las transiciones necesarias de un estado a otro. Hay que remarcar que estas transiciones entre estados vienen dadas por la ausencia de movimiento en el intérprete. En estos puntos de inflexión donde no hay movimiento alguno, las lecturas de la postura del intérprete nos darán cual es la acción a realizar según el punto de interacción donde nos encontremos.

#### La gestión de tiempo en O~A

Un elemento fundamental en el motor de audio de O~A es el tiempo. Se ha implementado un reloj que marca el inicio de cada compás 4/4, cuya velocidad debe ser configurada en bpm antes de la performance. Todas las grabaciones que se efectúen estarán sincronizadas con este reloj, así como las reproducciones. El reloj no marca los golpes de negra, sino los compases, disponiendo además de otro reloj para semicorcheas que es usado por el módulo de generación de secuencias. De esta manera, cuando se indica por parte del intérprete que debe comenzar una grabación mediante un gesto, la grabación no comenzará hasta que llegue el inicio del siguiente compás. Del mismo modo, cuando se indica que la grabación ha terminado, se elimina la parte sobrante del último compás



inacabado. Esta herramienta permite tener sincronizados los loops por parte del sistema, sin limitar aún así la longitud de los mismos, que será finalmente decisión del performer/intérprete.

### Estados

Ante la premisa de realizar un sistema de control basado en movimientos del cuerpo, hemos diseñado una máquina de estados que permite lanzar las distintas acciones en O~A.

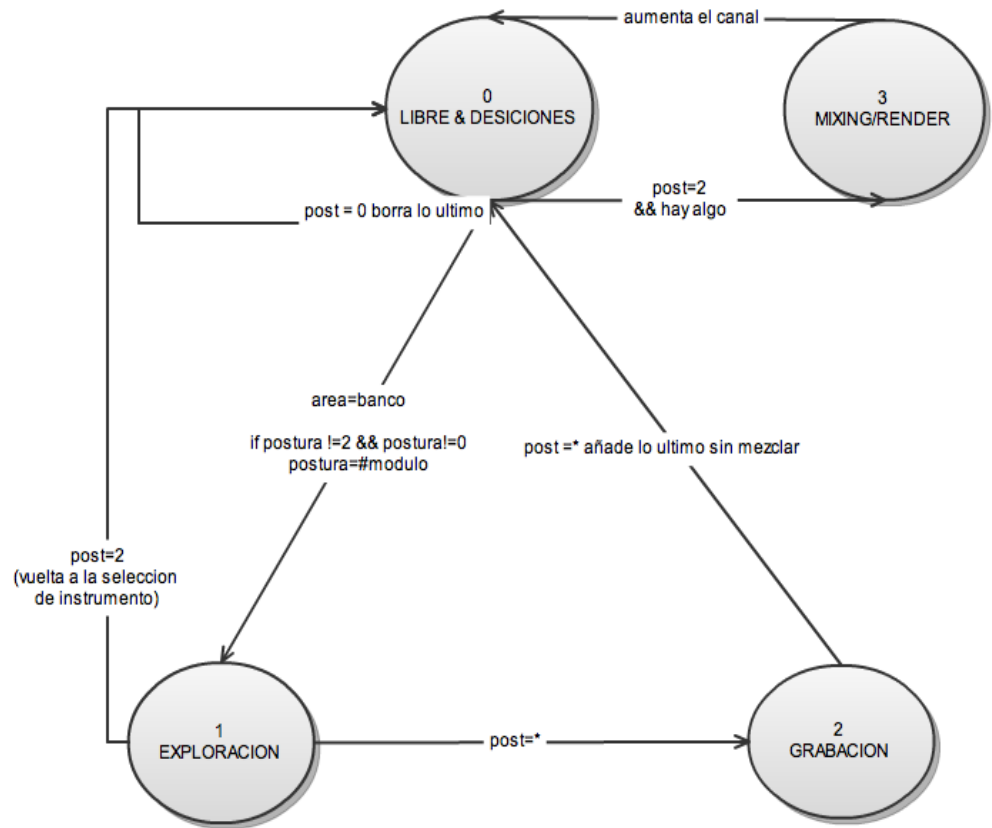


Fig. 10: Diagrama de estados de O~A

Descripción de los estados:

- El estado 0, “Libre/Decisiones”

Es el estado inicial. Al mismo tiempo, es el estado al que se llega tanto después de una grabación como de una mezcla. En este estado, cuando el sistema viene de arrancar, no se oye ningún sonido. Por otro lado, si ya se han llevado a cabo grabaciones, éstas serán reproducidas en loop. En cualquiera de los dos casos, los movimientos del intérprete no tienen ninguna repercusión sonora.

La parte de Decisiones viene cuando decidimos qué es lo que queremos hacer a continuación. En este estado, la postura de detención llevada a cabo por el intérprete tiene gran importancia ya que permite elegir:

a) Postura 0, si queremos borrar las repercusiones sonoras de nuestro último movimiento. Esta decisión nos llevará de nuevo al estado 0. (NOTA: La memoria disponible es de una sola posición, es decir, no podremos borrar más que la última capa sonora)

b) Postura 1, si queremos elegir un componente del motor de sonido y pasar a la fase de exploración (estado 1).

En este caso es importante también el área en la cual nos detenemos, ya que según si es a la izquierda, en el centro, o a la derecha, seleccionaremos los sintetizadores, efectos o secuenciadores, respectivamente.

c) Postura 2, si por el contrario queremos pasar al estado de mixing para acabar de mezclar el loop que hayamos creado. Para ello iremos al estado 3 (Mixing/Render).

- El estado 1, "Exploración"

En este estado se procede a una reproducción en tiempo real del sonido (o efecto) escogido por el intérprete en el estado 0, sin llevar a cabo ninguna grabación. Así, se pueden probar movimientos, golpes, velocidades, aceleraciones... Y nos permite ver qué repercusiones sonoras tienen estos movimientos. Hay que notar que este estado tiene sentido si se está utilizando O~A en fase experimental (como es el caso actual), y como banco de pruebas, pero se podría eliminar y pasar directamente al estado de grabación 2 en un entorno donde el intérprete ya esté habituado a la herramienta.

Del estado 1 se pueden producir dos tipos de transiciones según la postura del intérprete en el momento de detención:

a) Postura 2, se vuelve al estado 0 para seleccionar un nuevo instrumento

b) Posturas 1 ó 3, se pasa al estado de grabación (estado 2).

- El estado 2, "Grabación"

En este estado comienza la grabación del sonido producido o modificado por los movimientos del intérprete. Esta grabación comienza sincronizada con el reloj de compases (leer implementación), y termina de la misma manera. Así, se garantiza que los loops tendrán una duración en compases conocida, lo cual permite una gestión por parte del secuenciador. En el estado de Grabación, en el momento en que nos detengamos, se volverá al estado 0, sea cual sea la posición adoptada en ese momento. Esta decisión viene dada por la necesidad que podemos tener de querer grabar tal o cual sonido al principio o al final de nuestro loop. De esta manera, el diseño del diagrama de estados ha intentado respetar que el estado de grabación tenga una entrada y salida lo más abierta posible.

El estado de grabación siempre nos devuelve al estado 0 (Libre).

- El estado 3, "Mixing/Render"

En este estado, se nos permite hacer un control del volumen de mezcla con el que se añadirá la capa al conjunto de las 16 canales de grabación del sistema. Este estado ofrece además la posibilidad de borrar el resultado de todo lo que se haya grabado en el canal, si el volumen seleccionado es 0, o de renderizar el loop grabado en el canal activo y pasar al siguiente canal. En ambos casos, se vuelve al estado 0 (Libre) después.

### **Implementación en Pure Data**

El patch de audio consta de los siguientes elementos:

- Módulo de comunicación: gestiona la recepción y parsing de los datos que llegan desde openFrameworks, y realiza los envíos de las variables recibidas para que puedan ser utilizadas en el resto de módulos del patch.

- Módulo de control: gestiona las transiciones entre estados según las variables recibidas desde openFrameworks. También realiza el envío de las variables de control necesarias para activar las acciones a ejecutar en cada caso. Por último, el módulo de control también realiza la gestión temporal del patch, siendo quien envía las señales de reloj que permiten sincronizar la grabación y reproducción de loops.

- Módulo master: contiene 16 canales para la grabación y posterior reproducción de los loops. Estos canales contienen una tabla de longitud variable modificable, y ofrecen además la posibilidad de realizar renders a archivos wav de forma independiente. Es decir, tras una performance/concierto de O~A, obtendremos tantos archivos como canales se hayan grabado, pudiendo disponer así de la información sonora más desgranada para posibles ediciones o mezclas. También contiene una tabla extra, denominada tabla activa, que sirve como soporte para poder aplicar efectos o secuenciar el loop que está siendo modificado. Por último, el módulo máster realiza todas las operaciones necesarias según las recepciones de las variables de control recibidas desde el módulo de control: copia de tablas,

grabación, reproducción, render de archivos wav, mezcla y modificación de volúmenes.

- Módulo sound engine: el motor de sonido contiene todos los elementos que generan o modifican sonido en el patch. En el caso actual de O~A hay tres bancos de sonido diferenciados por su aplicación: banco de sintetizadores, banco de efectos y banco de secuenciadores.

### 2.4.3. Protocolo GDO (Gesture Data for O~A)

Desde el principio tuvimos claro que la gestión del vídeo y del sonido se haría de forma independiente usando plataformas diferentes. Por un lado, estábamos seguros de que OpenFrameworks era una plataforma idónea para la detección y tratamiento de la información visual. Al mismo tiempo, las librerías de sonido que han sido incorporadas hasta el momento en openFrameworks no nos daban la flexibilidad necesaria que buscábamos. Por lo tanto, a nivel de audio la elección sería Pure Data.

En un primer momento, intentamos conectar openFrameworks y Pure Data usando OSC, ya que existen implementaciones de este protocolo en ambas plataformas. Sin embargo, encontramos problemas desde el primer momento ya que el flujo de datos que proporcionan los frames de vídeo hacen necesario un mecanismo de comunicación muy orientado a la eficiencia. Como esta conexión se llevaría a cabo en red local, no estábamos tan interesados en la seguridad y tolerancia a fallos de estos envíos, y así nos decidimos a implementar nuestro propio protocolo de conexión entre las dos aplicaciones.

Gesture Data for O~A es un protocolo implementado directamente sobre la capa de transporte, y encima directamente del protocolo UDP. En este caso, no está orientado a conexión, y no se efectúa comprobación alguna en la recepción de paquetes.

Los datos son exclusivamente de tipo lista, y cada mensaje (o lista) debe estar encabezado por un identificador exclusivo que da nombre a la serie de datos que vendrá a continuación.

Ejemplo de mensaje GDO:

```
extremidad1          0.3 0.1222 0.265 1.7 5.2  
encabezado          lista de datos
```

Cada mensaje debe ir acompañado de un carácter nulo al final del mismo.

Los datos se envían como cadenas de bytes, con lo cual pueden ser parseados en destino a cualquier tipo de datos. Este parsing no lo garantiza el protocolo, con lo cual deberá ser llevado a cabo en cada implementación.

Las pruebas que hemos llevado a cabo han demostrado la robustez del protocolo GDO, que si bien no es adecuado a la hora de definir estructuras de datos complejas y configurables, si que da un resultado óptimo cuando se trata de enviar una gran cantidad de datos en tiempos muy cortos.

## 2.5. Diagrama de montaje

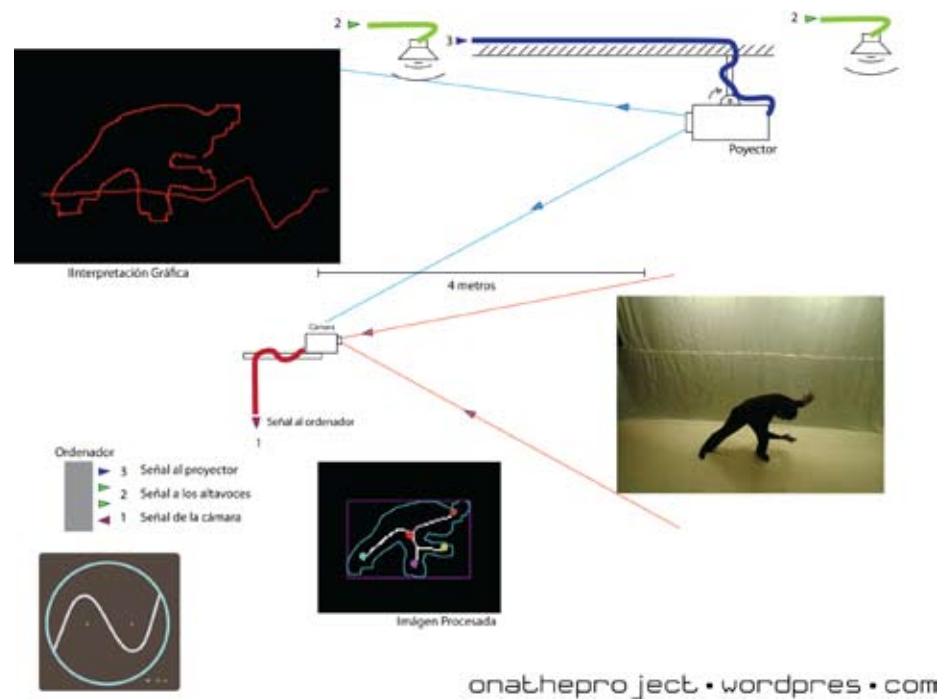


Fig. 11: Diagrama de montaje del sistema O~A

### 3. Conclusiones

Estamos convencidos de que hace falta humanizar la interfaz, el output/feedback y el espectáculo mismo si queremos dar un paso adelante en el camino de las artes digitales.

Encontramos la necesidad de desarrollar sistemas que no limiten al intérprete en su movilidad, y que registren con la mayor cantidad de información su dinámica de movimiento, su expresividad y en general sus formas corporales únicas.

Esperamos que el proyecto propuesto tenga una repercusión a medio plazo, tanto para la evolución de los sistemas de sonido como para muchas otras aplicaciones de artes digitales.

Hace falta mucho camino aún para tener un sistema tan completo como nos lo hemos planteado.

#### 3.1. Autocrítica

O~A es un proyecto ambicioso y complejo. Comprende ramas de muy diversa índole como la visión artificial, la generación musical en tiempo real o los protocolos de comunicación entre aplicaciones, así como la danza o la escritura escénica.

En esta primera aproximación, hemos sido capaces de llegar a tener un sistema estable de visión y comunicación, y un primer esbozo de lo que puede dar de sí la aplicación del mismo a la generación musical.

Aún así, nos hemos visto involucrados en el desarrollo de la parte de visión más tiempo del que habríamos esperado. Esto ha provocado retrasos en todas las tareas sucesivas.

Otra inconveniencia que hemos encontrado es la dificultad de adaptar los sintetizadores y demás módulos de generación sonora a una fuente de datos como la nuestra, que varía constantemente y que ofrece una cantidad de información muy elevada en comparación al input que suele tener un sistema musical normal. Esto ha provocado también dificultades a la hora de encontrar comportamientos sonoros interesantes, e interacciones complejas.

Si hay además una labor que en el diseño de un proyecto como este tiene una importancia fundamental es el ensayo del entorno con diversos intérpretes que deben dar un feedback fundamental a la hora de decidir por qué caminos debe seguir el mismo. Este feedback permite establecer desde cuestiones más técnicas como rangos de variables, sensibilidades o configuraciones hasta decisiones a mayor escala como aplicaciones, mapeos, comportamientos,... Lamentablemente, y debido a la ausencia del tiempo, y sobre todo de medios, necesarios para esta labor, nos hemos visto abocados a hacer pruebas la mayoría de las veces en entornos simulados o reducidos.

Aún a pesar de no haber podido someter a O~A a un testeo y puesta en marcha práctica exhaustiva, las pruebas que hemos llevado a cabo tienen resultados prometedores.

### 3.2. Desarrollos futuros

El futuro inmediato de O~A pasaría por establecer una serie de puesta en manos de la aplicación en talleres o workshops, de donde se podría obtener una respuesta de usuarios expertos que evaluaran el proyecto y concretasen los desarrollos futuros del mismo.

A más largo plazo, se debería formalizar el protocolo de gestos GDO, así como implementar una interfaz gráfica más configurable. Otra de las tareas a realizar en el futuro a largo plazo sería la evaluación de otras posibles configuraciones de O~A, utilizando otras fuentes de imagen o movimiento como origen de datos, por ejemplo desarrollando tracking tridimensional de las extremidades y recurriendo a sistemas de detección que no dependan del entorno. A día de hoy ya se han realizado pruebas utilizando el sistema como detección de gestos de las manos, y la adaptación a este tipo de gestualidad es impecable.

Por otro lado, se debería iniciar un estudio de las posibilidades de usar fuentes de gran flujo de datos como microcontrol de sistemas de generación sonora. En este campo, existen unas enormes posibilidades que se comienzan a descubrir hoy.

## 4. Bibliografía y referencias

### 4.1. Papers

- [1] Making Motion Musical, Tedd Winkler, International Computer Music Conference, 1995.
- [2] Contrôles Gestuels Bimanuels de Processus Sonores, Loïc Kessous, 2004.
- [3] Human Movement Tracking, Axel Mulder, School of Kinesiology, Simon Fraser University, 1994.
- [4] Virtual Musical Instruments, Accessing the Sound Synthesis Universe as a Performer, Axel Mulder, School of Kinesiology, Simon Fraser University, 1994.
- [5] Monalisa: “see the sound, hear the image”, Kazuhiro Jo, Norihisa Nagano.
- [6] Computer Vision for Artists and Designers: Pedagogic Tools and Techniques for Novice Programmers, Golan Levin.
- [7] Painterly Interfaces for Audiovisual Performance, Golan Levin.
- [8] A personal Chronology of Audiovisual Systems Research, Golan Levin.
- [9] Gestural Interactions for Multi-Parameter Audio Control and Audification, Thomas Hermann, Stella Paschalidou, Dirk Beckmann and Helge Ritter, Proceedings of the International Gesture Workshop, 2005
- [10] Correspondences between music and human body movement, E.Haga, 2008.
- [11] Action – Sound: Developing Methods and Tools to Study Music-related Body Movement, de A.R. Jensenius, 2007.
- [12] Mapping the musical commons, Josef Prögler.
- [13] VITRUVIA: Interacción hombre-máquina por medio de un sistema de visión artificial, Congreso Andescon - IEEE, Sebastian Gonzalez, 2006.

## 4.2. Web

[14] MOPRIM, Laboratory of Computer vision and Motion Capture

<http://www.cvmt.dk/projects/moprim/index.htm>

[15] ANVIL – The video annotation research tool, Michel Kipp

<http://www.anvil-software.de/>

[16] ConGAS – Gesture controlled audio systems

<http://www.cost287.org/>

[17] Workshop on Motion Capture for Music Performance, 2006

<http://www.idmil.org/mocap/mocap.html>

[18] The Musical Gestures Project / fourMs

<http://www.hf.uio.no/imv/forskning/forskningsprosjekter/musicalgestures/index.html>

<http://www.fourms.uio.no/>

[19] motion-e, Arizona State University.

<http://ame.asu.edu/motione/index.html>

[20] BIPED. Merce Cunningham y openended group

<http://www.openendedgroup.com/index.php/artworks/biped/>

[21] Didascalie.net

<http://www.didascalie.net>

[22] Tapemovie

<http://tapemovie.org/presentation>

[23] Jamoma

<http://www.jamoma.org/>

[24] Plateforme Virage

<http://www.plateforme-virage.org/>

[25] Theater der Klänge.

<http://www.theater-der-klaenge.de/>

[26] GDIF – Gesture Description Interchange Format

<http://www.gdif.org/>

[27] Speak. Alejandra Ceriani, Fabricio Costa y Fabian Kesler

<http://speakinteractive.blogspot.com/>

[28] BigEye – STEIM

<http://www.steim.org/steim/bigeye.html>

[29] Music via Motion

<http://www.leeds.ac.uk/icsrim/mvm/>



[30] EyeCon. Palindrome.

<http://eyecon.palindrome.de/>

[31] Mapping is not Music – IRCAM

<http://ftm.ircam.fr/index.php/MnM>

[32] SoftVNS – David Rokeby

<http://homepage.mac.com/davidrokeby/softVNS.html>

[33] Mudanx – Stan Wijnans

<http://www.mudanx.nl/index.html>

[34] DanceSpace, Flavia Sparacino.

<http://alumni.media.mit.edu/~flavia/resume.html>

[35] ShapeSynth, Andrew Glassner.

<http://www.glassner.com/andrew/cg/research/shapesynth/shapesynth.htm>

