



ESCUELA SUPERIOR DE INGENIERÍA

Ingeniería Técnica en Informática de
Sistemas

oFlute

Curso 2009-2010

José Tomás Tocino García
Cádiz, 21 de julio de 2011



ESCUELA SUPERIOR DE INGENIERÍA

Ingeniería Técnica en Informática de Sistemas

oFlute

DEPARTAMENTO: Lenguajes y Sistemas Informáticos.
DIRECTOR DEL PROYECTO: Manuel Palomo Duarte.
AUTOR DEL PROYECTO: José Tomás Tocino García.

Cádiz, 21 de julio de 2011
Fdo.: José Tomás Tocino García

Este documento se halla bajo la licencia FDL (Free Documentation License). Según estipula la licencia, se muestra aquí el aviso de copyright. Se ha usado la versión inglesa de la licencia, al ser la única reconocida oficialmente por la FSF (Free Software Foundation).

Copyright ©2010 José Tomás Tocino García.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

Agradecimientos

A Julian Raschke por crear y mantener Gosu.

Índice general

Índice general	9
Índice de figuras	11
Índice de cuadros	13
1. Introducción	15
1.1. Contexto y motivación	15
1.2. Objetivos	15
1.2.1. Funcionales	15
1.2.2. Transversales	16
1.3. Alcance	16
1.3.1. Limitaciones del proyecto	17
1.3.2. Licencia	17
1.4. Visión general	18
1.5. Glosario	18
1.5.1. Acrónimos	18
1.5.2. Definiciones	18
2. Desarrollo del calendario	19
2.1. Iteraciones	19
2.1.1. Primera iteración: conocimientos preliminares	19
2.1.2. Segunda iteración: analizador básico	19
2.1.3. Tercera iteración: interfaz gráfica de usuario	20
2.1.4. Cuarta iteración: motor de lecciones	20
2.1.5. Quinta iteración: motor de canciones	20
2.2. Diagrama de Gantt	20
2.3. Porcentajes de esfuerzo	20
3. Iteración 1: investigación preeliminar	21
3.1. Adquisición de conocimientos	21
3.1.1. El sonido	21
3.1.2. Descomposición de sonidos	23
3.1.3. Digitalización de sonidos	25
3.2. Estudio del software disponible	27
3.2.1. Aplicaciones comerciales	27

ÍNDICE GENERAL

3.2.2. Aplicaciones libres	28
3.3. Desarrollo con audio en GNU/Linux	29
3.3.1. Interfaces de bajo nivel, OSS y ALSA	30
4. Resumen	31
5. Conclusiones	33
6. Manual del usuario	35
7. Guía de ampliación	37
A. GNU Free Documentation License	39
1. APPLICABILITY AND DEFINITIONS	39
2. VERBATIM COPYING	41
3. COPYING IN QUANTITY	41
4. MODIFICATIONS	42
5. COMBINING DOCUMENTS	44
6. COLLECTIONS OF DOCUMENTS	44
7. AGGREGATION WITH INDEPENDENT WORKS	45
8. TRANSLATION	45
9. TERMINATION	45
10. FUTURE REVISIONS OF THIS LICENSE	46
Bibliografía	47

Índice de figuras

3.1. Rango de frecuencias de sonido	22
3.2. Componentes de una señal senoidal básica	22
3.3. Forma de ondas vs representación espectral	24

Índice de cuadros

1. Introducción

1.1. Contexto y motivación

Las nuevas tecnologías van filtrándose gradualmente en los centros educativos, y las técnicas de enseñanza se están adaptando a las opciones que ofrecen. El reparto de ordenadores portátiles a los alumnos andaluces de 5º y 6º de primaria, dentro del marco de la Escuela TIC 2.0, es buena muestra de ello.

Por otro lado, las nuevas generaciones están en plena simbiosis con las tecnologías de la información, cada vez más acostumbradas al empleo de dispositivos electrónicos, y su uso ya les es prácticamente instintivo. Por tanto, es beneficioso buscar nuevos métodos educativos que hagan uso de las nuevas tecnologías.

En la búsqueda de materias educativas en las que aplicar el uso de las nuevas tecnologías, la música, parte fundamental del programa curricular en la educación primaria, ofrece una gran variedad de aspectos que podrían desarrollarse utilizando tecnologías de la información. Es ahí donde este proyecto hace su aportación.

1.2. Objetivos

A la hora de definir los objetivos de un sistema, podemos agruparlos en dos tipos diferentes: **funcionales** y **transversales**. Los primeros se refieren a *qué* debe hacer la aplicación que vamos a desarrollar, e inciden directamente en la experiencia del usuario y de potenciales desarrolladores.

Por otro lado, los objetivos transversales son aquellos invisibles al usuario final, pero que de forma inherente actúan sobre el resultado final de la aplicación y sobre la experiencia de desarrollo de la misma.

1.2.1. Funcionales

- Crear un módulo de análisis del sonido en el dominio de la frecuencia para poder identificar las notas capturadas por el micrófono en tiempo real.

1. Introducción

- Crear una aplicación de usuario que identifique y muestre en pantalla las notas que toca el usuario en cada momento.
- Reutilizar el módulo de análisis en un juego en el que el usuario debe tocar correctamente las notas que aparecen en pantalla siguiendo un pentagrama.
- Incluir un sistema de lecciones multimedia individuales que sirvan al alumno de referencia y fuente de aprendizaje.
- Potenciar el uso de interfaces de usuario amigables, con un sistema avanzado de animaciones que proporcione un aspecto fluido y evite saltos bruscos entre secciones.

1.2.2. Transversales

- Obtener una base teórica sobre cómo se representa y caracteriza digitalmente el sonido.
- Conocer las bases del DSP (Digital Signal Processing), y su uso en aplicaciones de reconocimiento básico de sonidos, tales como sintonizadores y afinadores de instrumentos.
- Introducirme en la programación de audio en sistemas GNU/Linux.
- Entender las bases del análisis de sonidos en el dominio de la frecuencia.
- Utilizar un enfoque de análisis, diseño y codificación orientado a objetos, de una forma lo más clara y modular posible, para permitir ampliaciones y modificaciones sobre la aplicación por terceras personas.
- Hacer uso de herramientas básicas en el desarrollo de software, como son los **Sistemas de Control de Versiones** para llevar un control realista del desarrollo del software, así como hacer de las veces de sistema de copias de seguridad.

1.3. Alcance

oFlute se modela como una herramienta lúdico-educativa para alumnos que comiencen a aprender a usar la flauta dulce, proporcionando un entorno atractivo y ameno para el estudiante. Éstos tendrán la posibilidad de recorrer una serie de pequeñas lecciones sobre música en general, y el uso de la flauta dulce en particular.

Además, el usuario tendrá la posibilidad de comprobar sus conocimientos sobre el uso de la flauta practicando, gracias a las secciones de análisis de notas y de canciones, en las que la aplicación valorará la pericia del estudiante con la flauta.

1.3.1. Limitaciones del proyecto

El proyecto se limita al uso de la flauta dulce y no a otros instrumentos por la enorme variabilidad de timbre entre ellos, lo que supondría un enorme esfuerzo a la hora de generalizar el analizador de frecuencias.

El sistema de lecciones se basa en plantillas XML en las que es posible definir imágenes y texto para formar una pantalla de información. En un futuro se ampliará para incluir otros elementos multimedia así como lecciones con varias pantallas consecutivas.

Los sistemas de audio son una de las áreas en las que menos consenso hay entre plataformas informáticas, por lo que la transportabilidad de las aplicaciones suele ser compleja. El presente proyecto utiliza la API Simple de PulseAudio como subsistema de sonido, que es en teoría compatible con plataformas Win32, pero en la práctica su complejidad hace prácticamente inviable la portabilidad de la aplicación.

1.3.2. Licencia

El proyecto está publicado como software libre bajo la licencia GPL (General Public License) versión 2. El conjunto de bibliotecas y módulos utilizados tienen las siguientes licencias:

- A lo largo del proyecto se utilizan diferentes partes de las bibliotecas **Boost** [1], que utilizan la licencia *Boost Software License*¹. Se trata de una licencia de software libre, compatible con la GPL, y comparable en permisividad a las licencias BSD y MIT.
- **Gosu** [6], la biblioteca de desarrollo de videojuegos que ha proporcionado el subsistema gráfico, utiliza la licencia MIT (Massachusetts Institute of Technology). Cuando se compila en sistemas Windows, utiliza la biblioteca FMOD que es gratuita pero de código cerrado; en sistemas GNU/Linux, utiliza SDL_mixer, que utiliza la licencia LGPL (Lesser General Public License).
- **Kiss FFT** [4], la biblioteca utilizada para hacer el análisis de frecuencias, utiliza una licencia BSD (Berkeley Software Distribution).
- **PugiXML** [5], biblioteca de procesamiento de ficheros XML, se distribuye bajo al licencia MIT.
- **PulseAudio** [3] utiliza una licencia LGPL 2.1.

¹http://www.boost.org/LICENSE_1_0.txt

1.4. Visión general

Tras una revisión del calendario seguido, detallaremos a lo largo del resto de la memoria el proceso de análisis, diseño, codificación y pruebas que se siguió al realizar el proyecto.

Los manuales de usuario y de instalación se incluyen tras un resumen de los aspectos más destacables de proyecto y las conclusiones. En dicho manual, se hallan dos apartados dirigidos a la ampliación de la aplicación mediante la creación de nuevas lecciones y de nuevas canciones, respectivamente.

1.5. Glosario

1.5.1. Acrónimos

BSD Berkeley Software Distribution

DSP Digital Signal Processing

FDL Free Documentation License

FFT Fast Fourier Transform

FSF Free Software Foundation

GPL General Public License

LGPL Lesser General Public License

MIT Massachusetts Institute of Technology

1.5.2. Definiciones

Timbre Calidad de un sonido que permite distinguir la misma nota producida por dos instrumentos musicales u orígenes diferentes.

2. Desarrollo del calendario

El proyecto no se ha desarrollado siguiendo un calendario estricto, dado que era imposible cuantificar el tiempo que tomaría el adquirir las bases teóricas necesarias para poder afrontarlo con garantías. Su desarrollo se ha compaginado con los estudios del último curso de Ingeniería Técnica en Informática de Sistemas y las labores como becario en la Oficina de Software Libre y Conocimiento Abierto de la Universidad de Cádiz [2].

2.1. Iteraciones

Para la realización del proyecto se ha utilizado un modelo de desarrollo iterativo incremental. A continuación se detallan cada una de las etapas por las que ha ido pasando el proyecto.

2.1.1. Primera iteración: conocimientos preliminares

Antes de poder comenzar con el análisis y diseño del propio proyecto, era esencial adquirir una serie de conocimientos para poder afrontar su desarrollo con todas las garantías. Durante esta iteración, se llevaron a cabo labores de documentación y aprendizaje autodidacta con las que se asentaron los conocimientos necesarios.

Además, durante este periodo también se barajaron las diferentes posibilidades de implementación del sistema, así como las posibles herramientas y bibliotecas de terceros que pudieran ser de ayuda.

2.1.2. Segunda iteración: analizador básico

Una vez adquiridos los conocimientos teóricos necesarios, y decididas las técnicas y herramientas para llevar aquellos a la práctica, fue obvia la necesidad de empezar por diseñar un analizador de notas básico, que sería el corazón del programa. Del éxito del desarrollo temprano del módulo que se encargaría del análisis de sonidos dependería la viabilidad completa del proyecto.

2. Desarrollo del calendario

2.1.3. Tercera iteración: interfaz gráfica de usuario

Con el módulo de análisis desarrollado, *sólo* restaba desarrollar el resto de la aplicación alrededor del mismo. En esta tercera iteración se propusieron numerosos diseños para la interfaz gráfica de usuario y, una vez decantados por uno de ellos, comenzó el desarrollo de los elementos de la interfaz, haciendo énfasis en conseguir un aspecto dinámico y jovial.

2.1.4. Cuarta iteración: motor de lecciones

Uno de los subproductos de la aplicación es el motor de lecciones, que presenta una serie de unidades didácticas en formato multimedia, compuestas de imágenes y textos, con conceptos sobre música. En esta iteración se hizo un análisis de las posibilidades de este motor, concluyendo con el diseño y desarrollo de un mecanismo muy sencillo de ampliar y utilizar.

2.1.5. Quinta iteración: motor de canciones

La parte de mayor interactividad de la aplicación es el motor de canciones, en el que el usuario tiene la posibilidad de tocar una canción que aparece en pantalla, usando la flauta, mientras la aplicación valora en tiempo real su interpretación. Durante la quinta iteración se elaboró este sistema, encargado de listar y cargar las diferentes canciones, y puntuar al usuario según cómo lo haga.

2.2. Diagrama de Gantt

2.3. Porcentajes de esfuerzo

por hacer

3. Iteración 1: investigación preeliminar

3.1. Adquisición de conocimientos

Para poder enfrentarnos con garantías al desarrollo del proyecto fue necesario adquirir una base de conocimientos que nos permitiera entender los conceptos que se iban a usar y las herramientas para trabajar con ellos. Así, fuimos guiándonos por la intuición y, sobre todo, por las necesidades que iban surgiendo. Los conceptos que se presentan a continuación son básicos para comprender cómo funciona el módulo de análisis del proyecto.

3.1.1. El sonido

Un **sonido** es una vibración que se propaga por un medio elástico en forma de onda. Estas vibraciones se transmiten de forma longitudinal, esto es, en la misma dirección en la que se propaga la onda. El medio más común para la transmisión del sonido es el **aire**.

El sonido, en su forma más simple, se compone de una sola onda sinusoidal básica, con las características tradicionales: amplitud, frecuencia y fase. Una **onda sinusoidal** es aquella cuyos valores se calculan utilizando funciones seno.

Frecuencia y tono

La **frecuencia** mide el número de oscilaciones de la onda por unidad de tiempo. Por regla general, se utiliza el **hertzio** como unidad de medida de frecuencia, que indica la cantidad de repeticiones por segundo. La frecuencia determinará la **altura** del sonido, es decir, cómo de grave o agudo es. Los sonidos graves tienen una frecuencia baja, mientras que los sonidos agudos tienen una frecuencia alta.

A lo largo de los años se ha establecido un estándar de referencia que establece que la nota *la* que se encuentra encima del *do* central del piano debe sonar a 440 hertzios de frecuencia. Esta medida se utiliza a la hora de afinar los instrumentos, de modo que si al tocar la nota *la* se detecta un tono con una frecuencia de 440 hertzios, entonces el instrumento estará bien afinado.

3. Iteración 1: investigación preeliminar

El espectro audible por las personas lo conforman las **audiofrecuencias**, esto es, el conjunto de frecuencias que pueden ser percibidas por el oído humano.



Figura 3.1.: Rango de frecuencias de sonido

Un oído sano y joven es capaz de detectar sonidos a partir de los 20 hercios. Los sonidos por debajo de esa frecuencia se conocen como **infrasonidos**. Por otro lado, el límite auditivo en frecuencias altas varía mucho con la edad: un adolescente puede oír sonidos con frecuencias hasta los 18kHz, mientras que un adulto de edad media solo suele llegar a captar sonidos de hasta 13kHz. El límite genérico superior se establece en 20kHz, por encima de los cuales los sonidos se denominan **ultrasonidos**.

Amplitud

La **amplitud** representa la energía que transporta la onda. Cuando un instrumento u otro objeto genera una vibración, la amplitud es la cantidad de movimiento que esa vibración genera. Podría equipararse (de forma no estricta) a la intensidad del sonido: cuanto mayor sea la amplitud, más fuerte se oirá el sonido.

Fase

Por último, la **fase** (φ) indica el desplazamiento horizontal de la onda respecto del origen. Si la fase de una onda no es cero, entonces parecerá que está *desplazada* hacia la derecha, si la fase es positiva, y hacia la izquierda si la fase es negativa.

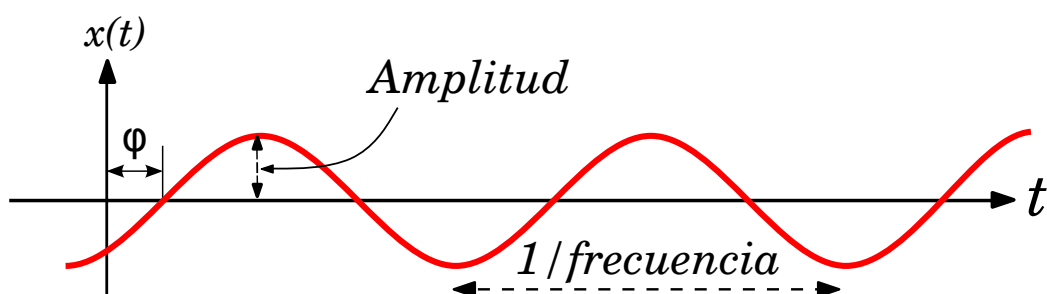


Figura 3.2.: Componentes de una señal senoidal básica

3.1.2. Descomposición de sonidos

Para desarrollar oFlute nos interesa conocer la altura de la nota que está tocando la flauta en un instante concreto. Para un tono puro, podríamos conocer la altura fijándonos en su frecuencia. El problema es que, en la naturaleza, **no existen** los tonos puros, sino que los sonidos se componen de multitud de tonos de diferentes amplitudes, frecuencias y fases.

Afortunadamente, la teoría dicta que cualquier tono complejo puede descomponerse como suma de tonos puros de distintas amplitudes, fases y frecuencias, llamados **parciales**. La menor de todas las frecuencias de los parciales se conoce como **frecuencia fundamental**, y es la que dicta la altura general del sonido – *general*, ya que aunque el resto de frecuencias puede corresponder a otras notas, es la altura de la frecuencia fundamental la que mayor relevancia tiene en el sonido.

Un subconjunto de esos parciales, conocidos como **armónicos**, tienen frecuencias múltiplos de la frecuencia fundamental. Estos armónicos sirven para enriquecer el sonido y, sobre todo, determinar el **timbre musical** del origen del sonido: dos instrumentos (o personas) pueden estar tocando la misma nota y emitir la misma frecuencia fundamental, pero será el conjunto total de armónicos el que nos ayude a distinguir qué instrumento está emitiendo el sonido.

Así pues, el objetivo es encontrar una forma de descomponer una señal (el sonido) en sus componentes y analizar sus frecuencias, buscando la frecuencia fundamental, que nos informará de la nota que se está tocando.

Representación gráfica de sonidos

La representación habitual de las señales se hace en el **dominio del tiempo**, es decir, podemos observar cómo la señal cambia a lo largo del tiempo, viendo el valor de su **amplitud** en cada instante. Por otro lado, la representación en el **dominio de la frecuencia** nos permite analizar una señal respecto a las frecuencias que la componen, dividiendo la señal en sus componentes.

En la figura 3.3 podemos comparar la representación de un sonido en el dominio del tiempo, en **forma de ondas**, tal y como aparecería en un osciloscopio, frente a su representación en **forma espectral**, en la que el eje vertical indica la frecuencia, y la intensidad del color indica la intensidad de esa componente frecuencial en el sonido.

Herramientas de descomposición de señales

La herramienta fundamental a la hora de descomponer una señal periódica, como puede ser un sonido, en sus parciales o armónicos es el **análisis armónico** o **análisis de Fourier**. Esta rama del análisis matemático estudia la representación de funciones o

3. Iteración 1: investigación preeliminar

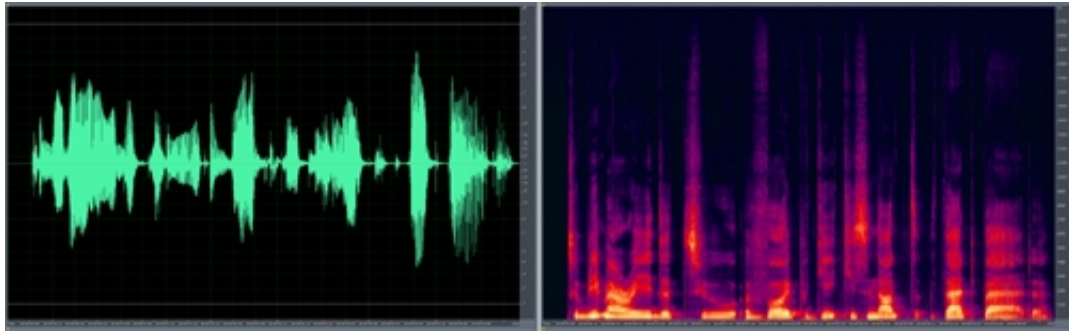


Figura 3.3.: Forma de ondas vs representación espectral

señales como superposición de ondas básicas, y hoy en día se aplica en innumerables campos de la ciencia, desde el procesamiento de señales para el reconocimiento de patrones, como es nuestro caso, a la neurociencia.

Una de las herramientas más conocidas de este área es la **transformada de Fourier**. Se trata de una aplicación matemática que descompone una función en su espectro de frecuencias a lo largo del dominio. Al aplicarla sobre una función f , se define de la siguiente manera:

$$g(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-i\xi x} dx$$

De cualquier modo, al estar tratando con un sistema digital como es una computadora, no es viable aplicar esta definición de la transformada de Fourier, ya que se basa en funciones continuas y derivables, y en nuestro caso dispondremos de datos discretos.

De ahí, aparece la **transformada discreta de Fourier** o **DFT**, que tiene el mismo uso que la transformada tradicional pero requiere que la función de entrada sea una secuencia discreta y de duración finita.

Existe un gran número de aproximaciones al cálculo de la transformada de Fourier, pero claramente el algoritmo más utilizado y eficiente es el **FFT**, **Fast Fourier Transform**. A pesar de imponer algunas limitaciones para mantener la eficiencia, el algoritmo FFT es la implementación que más habitualmente se encuentra en los chips DSP. Por regla general, computar la transformada de Fourier para N puntos usando FFT tardaría un tiempo $O(N \cdot \log(N))$, mientras que hacerlo utilizando la definición estándar de la DFT llevaría un tiempo $O(N^2)$.

A pesar de que fue el **DFT** el algoritmo que se utilizó finalmente en el proyecto, se estudiaron otras posibles herramientas para la detección de la frecuencia fundamental, como por ejemplo la **función de autocorrelación**, que también suele utilizarse en análisis de señales para encontrar patrones repetitivos, como señales enmascaradas por ruido. A pesar de ello, dada la poca bibliografía encontrada sobre estas técnicas secundarias y la conocida eficiencia de la transformada de Fourier, se decidió optar por la técnica más conocida.

3.1.3. Digitalización de sonidos

Antes de poder aplicar ninguna técnica sobre los sonidos, es necesario transformarlos de forma que el ordenador pueda trabajar con ellos.

Captación de sonidos

Lo más habitual a la hora de digitalizar un sonido es, primeramente, utilizar algún dispositivo que transforme las ondas sonoras en algo que pueda transmitirse al computador en forma de ondas eléctricas. Este dispositivo es el **micrófono**, en nuestro caso de tipo **electret**, que es el más utilizado en ordenadores personales, teléfonos móviles y demás dispositivos de consumo con requisitos de audio de media o baja fidelidad.

Estos micrófonos constan de una membrana que vibra libremente cuando capta cualquier onda acústica o de sonido, ya sea voz, música o ruidos, convirtiéndola en una señal eléctrica de baja frecuencia y de muy poca tensión o voltaje, semejante a la del sonido captado. Una vez que esta señal eléctrica llega a la tarjeta de sonido, comienza la siguiente parte del proceso.

Curiosamente, el proceso es el inverso del que ocurre en un altavoz. Es por eso que en el caso de algunos auriculares intrauditivos, como los que habitualmente acompañan a los reproductores MP3 de bolsillo, es posible utilizarlos como micrófonos de baja fidelidad. También es posible, aunque bastante más difícil, utilizar ciertos micrófonos de escritorio como altavoces improvisados, limitados a la reproducción de altas frecuencias.

Muestreo de la señal

El siguiente paso es el **muestreo** (o *sampling*) de la señal. El proceso consiste en medir la amplitud de la señal analógica en diferentes puntos, uniformemente espaciados, a lo largo del tiempo. El número de veces que se muestrea la señal por unidad de tiempo es conocido como **frecuencia de muestreo**, e influye directamente en la calidad de la digitalización del sonido.

La elección de la frecuencia de muestreo no suele ser trivial y tiene un impacto importante en el rendimiento y calidad del sistema, ya que el número de elementos a procesar es directamente proporcional a la frecuencia.

Otro factor importante es la clase de sonidos que vamos a digitalizar. Por regla general, los sonidos que se captan son los audibles por el oído humano. Tal y como se comentó en la sección anterior, estos sonidos son aquellos cuyas frecuencias se encuentran por debajo de los 20 kHz. Existe un teorema dictado por el ingeniero sueco **Harry Nyquist** que defiende que *“la frecuencia de muestreo mínima requerida para muestrear una señal debe ser igual al doble de la máxima frecuencia contenida en la señal”*. En

3. Iteración 1: investigación preeliminar

nuestro caso, como la máxima frecuencia audible es de 20 kHz, lo normal será utilizar una frecuencia de muestreo de 40 kHz. El estándar de CD, que normalmente se utiliza como base de muestreo en la mayoría de tarjetas de sonido, amplía la tasa un 10 % con objeto de contemplar el uso de filtros no ideales, quedando la frecuencia de muestreo en 44,1 kHz.

Cuantificación de las muestras

Una vez decidida la frecuencia de muestreo de la señal, será necesario acordar qué utilizaremos para representar sus niveles de amplitud de forma digital. Este proceso se conoce como **cuantificación**, y de él se desprenderá el número de bits de cada muestra. Cabe notar que tanto el muestreo como la cuantificación son procesos con **pérdidas**, ya que es imposible discretizar con total fidelidad un rango continuo de tensiones.

Existen diferentes métodos para decidir los niveles a los que se ajustarán las muestras. El más utilizado es el **PCM - modulación por impulsos codificados** en su variante *uniforme*, que utiliza una escala uniforme para digitalizar los valores de amplitud, a diferencia de la versión *no uniforme*, que utiliza escalas como la logarítmica.

La **resolución de cuantificación** (o *resolución digital*) más habitual es de 16 bits, que es la utilizada en los CDs de audio. Esto nos permite tener $2^{16} = 65536$ niveles distintos con los que cuantizar cada muestra.

Codificación de las muestras

El último paso antes de tener los datos listos para el procesamiento es la **codificación** en forma de bits. Aunque podría parecer un proceso trivial – convertir los valores digitales de las muestras en binario – existen multitud de parámetros que influyen a la hora de representar estas señales:

- **Tamaño de la muestra:** decidido en la cuantificación, la muestra puede tener tamaños desde los 8 a los 64 bits.
- **Orden de los bytes:** para muestras de más de un byte, es importante decidir el orden de los mismos – *endianness*. Popularmente, la mayoría de computadoras basadas en procesadores Intel utilizan *little endian* – esto es, se almacena primero los bytes de menos relevancia.
- **Signo:** cualquier señal en forma de ondas pasa constantemente por el origen, de forma que la amplitud toma valores positivos y negativos a cada momento. Puede parecer intuitivo usar un entero con signo para la representación, pero también es posible utilizar uno sin signo, de forma que el origen se represente como la mitad del rango, ahorrándonos así posibles complicaciones en la representación de números negativos.

- **Canales:** la mayoría de micrófonos de baja calidad producen sonido monoaural, de forma que solo es necesario utilizar un canal para su reproducción, a diferencia de los micrófonos estereofónicos que utilizan dos o más canales. En este aspecto, el flujo que se genera durante la digitalización de una señal estéreo es más complejo de procesar en tanto en cuanto los datos de cada canal vienen entrelazados en el flujo y, a veces, es difícil distinguirlos.

3.2. Estudio del software disponible

Existen algunas soluciones de software, juegos en la amplia mayoría de los casos, que explotan la idea del análisis de sonido en tiempo real como principal modo de interactuar con el usuario. En esta sección vamos a conocer algunas de estas soluciones y las ideas que adquirimos de su estudio.

3.2.1. Aplicaciones comerciales

SingStar

SingStar fue el primer videojuego en explotar el uso de un micrófono para que el usuario cantase y la aplicación reconociese el sonido. Apareció por primera vez en mayo de 2004 para sistemas PlayStation 2, y desde entonces han aparecido nada menos que 23 ediciones para este sistema y otras 6 para PlayStation 3.

La ventaja de SingStar es la que da ser el primero en explotar una idea atractiva, que rápidamente consiguió adeptos, principalmente entre el público más joven. Este éxito se vio fortalecido por la firma de una gran cantidad de contratos con discográficas a lo largo del tiempo, que permitió el lanzamiento de ediciones regulares con los *singles* más populares.

Las últimas ediciones de SingStar incluyen algoritmos avanzados que permiten, entre otras opciones, añadir efectos a las voces de los usuarios ó automáticamente limpiar las pista vocales de las canciones que los jugadores carguen mediante almacenamiento externo.

Lips

Lips fue la respuesta de Microsoft a SingStar para sus sistemas **Xbox 360**. El planteamiento es similar al de la versión de PlayStation, aunque incluye una serie de mejoras bastante atractivas.

Los micrófonos utilizados en Lips son inalámbricos e incluyen un sistema de detección de movimientos, de forma que es posible utilizarlos en secciones sin pista vocal

3. Iteración 1: investigación preeliminar

pero con percusión, siguiendo el ritmo a modo de maracas, o imitando movimientos que aparecen en pantalla.

Desde el principio, Lips ha permitido utilizar canciones de terceros mediante la conexión de un reproductor MP3. Esta opción solo estuvo disponible en sus competidores después de la aparición de Lips.

Además, Lips introdujo un sistema de juego colaborativo en forma de duetos, y competitivo, en el que los jugadores cantaban secciones consecutivas de una canción en busca de conseguir la mejor interpretación.

Apariciones menores en otros títulos

Aunque Lips y SingStar han sido los dos principales juegos del género, muchos otros juegos musicales han incluido pequeñas pruebas y minijuegos que han hecho uso de micrófonos. Por ejemplo, **DJ Hero**, **Guitar Hero**, **Band Hero** y **Def Jam Rapstar** permiten utilizar el micrófono para añadir acompañamiento vocal al juego. La ventaja es que en la mayoría de los casos, es posible utilizar los micrófonos de Lips y SingStar con estos juegos de terceros, evitando tener que adquirir más dispositivos.

3.2.2. Aplicaciones libres

UltraStar

UltraStar fue el primer clon libre de SingStar. Fue desarrollado por Patryk Cebula en 2007 y ha servido como base para diferentes forks posteriores. El juego permite a varias personas jugar a la vez mediante la conexión de varios micrófonos a una tarjeta de sonido, así como la adición de nuevas canciones de forma sencilla mediante ficheros de configuración en formato texto.

Aunque las versiones iniciales se liberaron bajo una licencia GNU GPL, desgraciadamente en la actualidad UltraStar se encuentra con licencia *freeware*, utilizando como excusa inválida que así “[...] se protegen los datos privados de los usuarios al ser enviados al servidor mediante SSL”. Realmente no existe razón para no utilizar software de código abierto con SSL.

UltraStar Deluxe

UltraStar Deluxe nació como una modificación básica de UltraStar, pero consiguió atraer la atención de muchos usuarios y desarrolladores, y finalmente se constituyó como un producto independiente. Los desarrolladores de UltraStar Deluxe decidieron trabajar

en varios aspectos que vieron mejorables respecto al UltraStar original. Primero, mejorar la fiabilidad del programa, arreglando numerosos bugs y aumentando el rendimiento. Segundo, trabajar la apariencia visual, basándose en gran medida en los efectos del SingStar de PlayStation 3. Finalmente, facilitar la expansibilidad del sistema, permitiendo un gran número de formatos para los ficheros de vídeo y audio, y creando un sistema de scripting basado en Lua para los modos de juego colaborativos.

Performous

Performous es uno de los juegos musicales open source más populares. Nació como una reescritura del UltraStar original, aunque posteriormente lo superó con creces. La fortaleza de Performous reside en su capacidad de reconocimiento de voces, basado en la *transformada rápida de Fourier (FFT)* y en una serie de algoritmos de post-procesamiento.

Performous ha evolucionado con el tiempo, naciendo como un juego de cante pero añadiendo características colectivas como Guitar Hero o Rock Band, permitiendo el uso de controladores adicionales, como guitarras o baterías electrónicas. Además, en las últimas versiones Performous incluye un modo de baile, muy similar a los clásicos DDR o StepMania.

3.3. Desarrollo con audio en GNU/Linux

El desarrollo de aplicaciones que realicen tareas de sonido en sistemas GNU/Linux es uno de los casos en los que más dificultades se encuentran. Tradicionalmente, el soporte del hardware de sonido en estos sistemas siempre ha sido de lo más básico, incluso limitándose, en ciertas ocasiones, a la reproducción de sonido, ignorando por completo la grabación. Afortunadamente, con el paso de los años el soporte ha ido mejorando gracias a la colaboración de los fabricantes y a la proliferación del sonido integrado en placa base.

A nivel de software, existen bastantes componentes diferentes, algunos alternativos y otros complementarios entre sí, que pueden conducir a confusiones. En otros sistemas operativos, como Windows o Mac OS, el programador cuenta con una interfaz de sonido común, que se encarga de la mezcla y de la comunicación de bajo nivel con la tarjeta de sonido. En GNU/Linux, dada su naturaleza modular, esa misma tarea se descompone en diferentes sistemas, por lo que una misma tarea puede realizarse de muchas formas distintas.

Buena muestra de ello es el artículo *Welcome To The Jungle* ¹, en el que el desarrollador de Adobe, Mike Melanson, hace un repaso sobre la *jungla* que supone la programación de audio en Linux. El artículo es antiguo y las cosas han mejorado desde

¹http://blogs.adobe.com/penguinswf/2007/05/welcome_to_the_jungle.html

3. Iteración 1: investigación preeliminar

entonces, pero aún así es muy fácil que los no iniciados se sientan abrumados por la cantidad de opciones disponibles.

3.3.1. Interfaces de bajo nivel, OSS y ALSA

El elemento de menor nivel en esta *escala* de componentes son las interfaces de hardware, que podrían equipararse al *driver de audio* de Windows.

OSS

Open Sound System (OSS) fue durante muchos años la interfaz de audio por defecto en todos los sistemas GNU/Linux. Está basada en el estándar UNIX para la comunicación con dispositivos mediante las funciones POSIX habituales (*open*, *read*, etc), lo que la hace relativamente sencilla de utilizar.

Desarrollada por Hannu Savolainen, sus comienzos fueron *open-source*. Posteriormente, el programador pasó a trabajar en una compañía que privatizó los nuevos cambios, aunque finalmente, en 2008, se volvió a liberar todo el código. Para entonces su mayor rival, ALSA, ya había tomado su lugar como API predeterminada en el kernel de Linux.

ALSA

Como alternativa a OSS surgió **Advanced Linux Sound Architecture** (ALSA), que acabó colocándose como la alternativa por defecto en todos los sistemas GNU/Linux a partir de la versión 2.6 del kernel.

Entre sus características, ALSA permite la síntesis de sonidos MIDI mediante hardware, soporte multiprocesador, configuración automática de tarjetas de sonido, etcétera. En gran parte, los objetivos de ALSA fueron las deficiencias de OSS en aquella época.

ALSA está estructurada en tres componente. La primera parte son los controladores en el kernel. La segunda parte es una API para los desarrolladores. Esta API es de muy bajo nivel, y es utilizada principalmente por middlewares y frameworks en lugar de por aplicaciones de usuario. Por último, el tercer componente es un mezclador que permite el multiplexado del sonido.

El desarrollo de sistemas relacionados con el sonido es uno de los que menos

4. Resumen

5. Conclusiones

6. Manual del usuario

7. Guía de ampliación

A. GNU Free Documentation License

Version 1.2, November 2002

Copyright ©2000,2001,2002 Free Software Foundation, Inc.

51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “**Document**”, below,

A. GNU Free Documentation License

refers to any such manual or work. Any member of the public is a licensee, and is addressed as “**you**”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “**Modified Version**” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “**Secondary Section**” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “**Invariant Sections**” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “**Cover Texts**” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “**Transparent**” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “**Opaque**”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing

tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The **“Title Page”** means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, **“Title Page”** means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

A section **“Entitled XYZ”** means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as **“Acknowledgements”**, **“Dedications”**, **“Endorsements”**, or **“History”**.) To **“Preserve the Title”** of such a section when you modify the Document means that it remains a section **“Entitled XYZ”** according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document’s license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of

A. *GNU Free Documentation License*

the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.

- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of

A. GNU Free Documentation License

peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under

A. GNU Free Documentation License

this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

Bibliografía

- [1] *Boost C++ libraries*. URL <http://www.boost.org>.

Boost es un conjunto de bibliotecas para C++ que ofrecen soluciones a todo tipo de problemas. Están escritas por los mejores desarrolladores de C++, y diez de estas bibliotecas formarán parte del nuevo estándar C++0x.

- [2] *Oficina de Software Libre y Conocimiento Abierto de la Universidad de Cádiz*. URL <http://osl.uca.es>.

Oficina de Software Libre y Conocimiento Abierto de la Universidad de Cádiz, en la que trabajé como becario durante el desarrollo del proyecto, realizando labores de organización y gestión de eventos, administración de software y asistencia técnica.

- [3] *PulseAudio*. URL <http://pulseaudio.org>.

PulseAudio es un servidor de sonido multiplataforma, compatible con sistemas GNU/Linux y Windows, y utilizado en algunas de las distribuciones más conocidas, como Ubuntu, Fedora, Mandriva, openSuse y Linux Mint.

- [4] Mark Borgerding. *Kiss FFT*. URL <http://sourceforge.net/projects/kisfft>.

Kiss FFT es una biblioteca para realizar Transformadas Rápidas de Fourier (FFT (Fast Fourier Transform)). Es una biblioteca muy pequeña, razonablemente eficiente y portable con capacidad para realizar operaciones en distintos formatos.

- [5] Arseny Kapoulkine. *Pugixml*. URL <http://code.google.com/p/pugixml>.

PugiXML es una biblioteca ligera para el procesamiento de archivos XML en C++. Tiene soporte completo Unicode, un parseador muy veloz y capacidad para usar consultas XPath.

- [6] Julian Raschke y otros. *Gosu*. URL <http://libgosu.org>.

Gosu es una biblioteca de desarrollo de videojuegos 2D para Ruby y C++, con aceleración gráfica por OpenGL y orientación a objetos.