Análisis de audio

Santiago Vilanova Ángeles

PID_00184754

Universitat Oberta de Catalunya

Índice

Introducción

En este módulo haremos referencia a todo aquello relacionado con el diseño de interacción mediante el sonido. Gracias a dispositivos de captura de audio como los micrófonos, hoy en día incorporados a casi todos los aparatos electrónicos de comunicación y ocio (teléfonos, ordenadores, videoconsolas), podemos diseñar interacciones persona-ordenador sencillas y efectivas, con el canal de nuestra propia voz como interactuador o mediante el análisis de cualquier otra fuente sonora (instrumentos musicales, sonido de ambiente, etc.).

Pensamos, por ejemplo, en sistemas de iluminación domótica que permiten la activación de la luz mediante dos palmadas seguidas, en sistemas de visualización musical que posibilitan la generación automática de imágenes a partir del análisis frecuencial de la música, o en sistemas de análisis del ruido de ambiente para detectar el nivel de contaminación acústica.

Objetivos

1. Repasar conceptos fundamentales de la física ondulatoria, como la frecuencia y la amplitud.

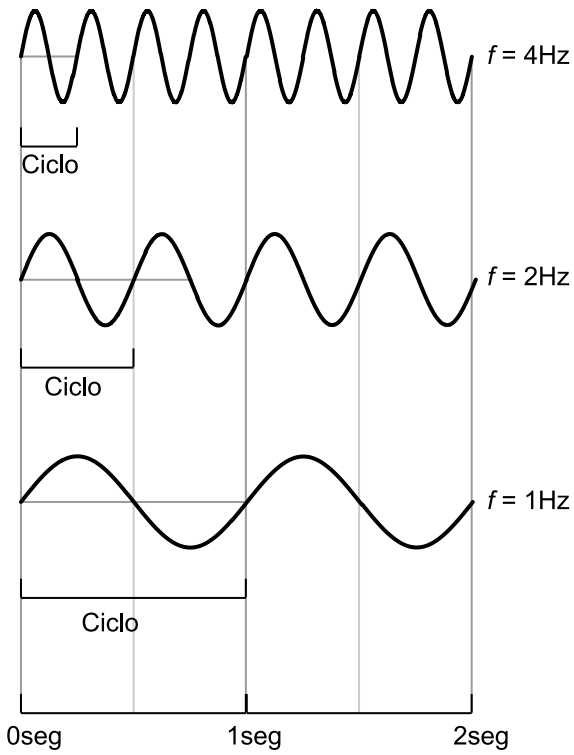
2. Sedimentar los conceptos y características fundamentales del audio digital.

3. Aprender los algoritmos más importantes relacionados con el análisis de audio.

4. Estudiar las herramientas existentes para la captura del sonido y su posterior digitalización.

1. Conceptos teóricos

1.1. Frecuencia

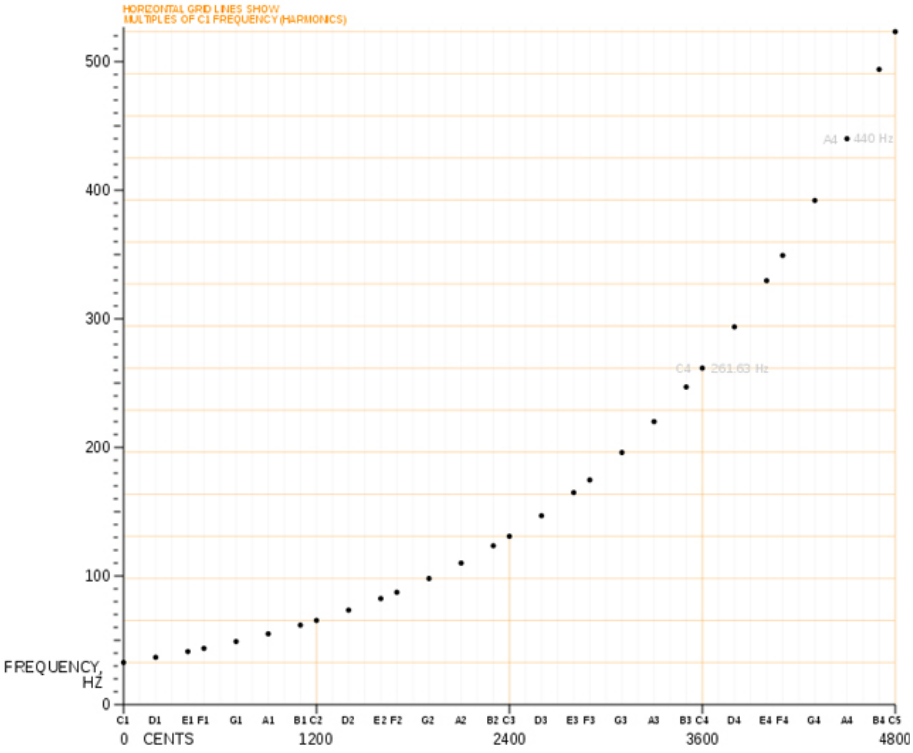


La frecuencia del sonido hace referencia a la cantidad de veces que vibra el aire que transmite ese sonido en un segundo. La unidad de medida de la frecuencia es el **hercio** (Hz). La medición de la onda puede empezar en cualquiera de sus puntos.

Para que el ser humano pueda oír un determinado sonido, su frecuencia debe estar comprendida entre los 20 y los 20.000 Hz. La mayoría de sistemas de captación convencional no superan este rango de frecuencias.

Los sonidos que percibimos como **agudos** tienen una frecuencia mayor que los sonidos **graves**. La frecuencia de los tonos agudos oscila entre los 2.000 y los 8.000 Hz, mientras que la de los graves varía entre los 20 y los 250 Hz. Los tonos **medios** tienen una frecuencia de oscilación de entre 500 y 1.000 Hz.

1.2. Frecuencia y tonos musicales



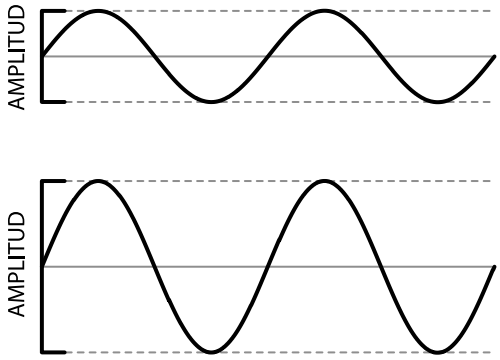
Fuente: Wikipedia

En la notación musical occidental, se toma como referencia la frecuencia de 440 Hz, que es asignada a la nota A4 (la de la 4.ª octava). Entre una nota y su octava superior, que es del doble de frecuencia, hay doce subdivisiones o semitonos.

La ratio numérica entre las frecuencias de dos semitonos sucesivos es exactamente la raíz duodécima de 2 (un factor de aproximadamente 1,05946).

Este sistema de subdivisión de la escala musical en doce tonos se denomina ***temperamento igual*** y, a pesar de que es el sistema de afinación más extendido en la música occidental, no es el único existente, de modo que pueden construirse sistemas de división infinitos de una escala musical, con distintas ratios de frecuencia entre notas correlativas y distinto número de subdivisiones.

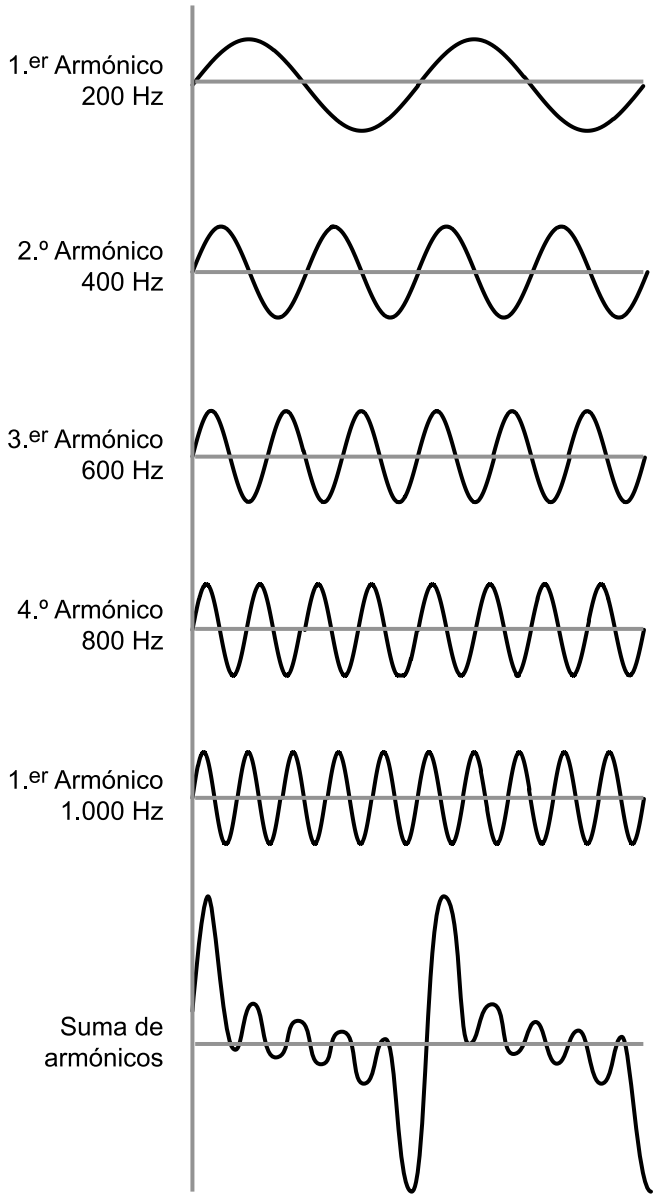
1.3. Amplitud



La amplitud de una onda sonora es el grado de movimiento de las moléculas de aire en la onda, y corresponde a la intensidad de la expansión y la compresión atmosférica que la acompañan. Cuanto más elevada es la amplitud de la onda, más intensamente golpean las moléculas el tímpano y más **fuerte** es el sonido percibido.

La amplitud de una onda sonora puede expresarse en unidades absolutas midiendo la distancia de desplazamiento de las moléculas del aire o la diferencia de presión atmosférica entre la compresión y la expansión. Habitualmente nos referiremos a la amplitud de una onda sonora empleando su medida en decibelios (dB), aunque también puede venir expresada en pascales y milibares.

1.4. Señales complejas: armónicos



La mayoría de osciladores, incluyendo la voz humana, un violín o una estrella cefeida, son más o menos periódicos y están formados por armónicos.

La mayoría de osciladores pasivos, como las cuerdas de una guitarra, la membrana de un tambor o una campana, oscilan de modo natural a diferentes frecuencias simultáneas, conocidas como parciales. Cuando el oscilador es largo y fino, como una cuerda de guitarra o la columna de aire en una trompeta, la mayoría de los parciales son múltiplos enteros de la frecuencia fundamental; estos parciales se denominan **armónicos**.

Los armónicos son las diferentes frecuencias presentes en el movimiento de un oscilador, múltiplos íntegros de la frecuencia fundamental.

Los parciales con frecuencias que no son múltiplos enteros de la frecuencia fundamental se denominan **inarmónicos**, y generalmente son percibidos como desagradables. El sonido de las campanas, por ejemplo, contiene muchos inarmónicos.

Percepción del oído humano

El oído humano no percibe los armónicos como notas separadas. De hecho, una nota musical formada por muchas frecuencias armónicamente relacionadas se percibe como un sonido único, cuya calidad o timbre es el resultado de la amplitud relativa de cada una de las frecuencias armónicas. El sonido de una nota de piano, por ejemplo, está formado por una combinación compleja de armónicos, aunque nuestro cerebro interpreta como definitorio de la altura tonal solo el armónico fundamental. El resto de armónicos presentes en el sonido del piano contribuyen a su carácter tímbrico y nos hacen percibir su sonido característico, diferenciable tímbricamente del sonido de la guitarra, por ejemplo.

1.5. Monofonía/polifonía

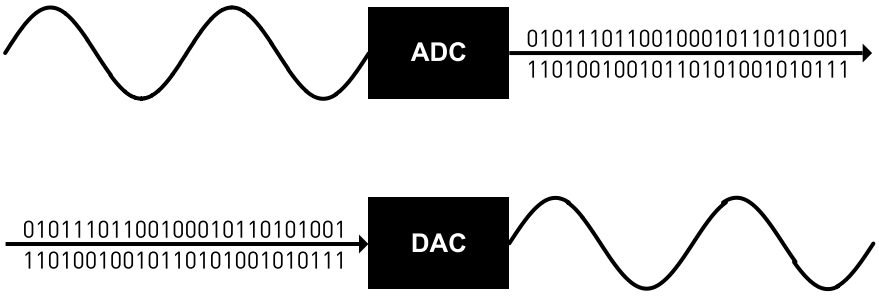
A la hora de plantear el diseño de nuestra interacción, hemos de tener muy en cuenta si vamos a trabajar con señales **polifónicas** o **monofónicas**. Una señal monofónica está formada por un único armónico fundamental (la voz humana o una trompeta) y es mucho más sencilla de analizar (como tono puro o *pure pitch*, como veremos más adelante) que una señal polifónica, que está compuesta por un conjunto de sonidos simultáneos a distintas frecuencias (una orquesta sinfónica, acordes de piano, rumor de la calle).

Aun así, como veremos más adelante, el análisis de señales mediante FFT nos puede ofrecer información muy útil, incluso en el caso de señales de audio polifónico.

Las características más importantes del sonido son la **frecuencia** (número de oscilaciones por segundo, medida en hercios) y la **amplitud** (intensidad de la onda sonora, medida en decibelios). Además, todos los sonidos están formados por un número de frecuencias que se producen simultáneamente a la frecuencia fundamental, que denominamos **armónicos** o **inarmónicos** dependiendo de si sus frecuencias son múltiplos enteros de la frecuencia fundamental o no lo son, y que aportan a cada sonido su carácter tímbrico.

1.6. ADC/DAC

ADC/DAC: convertidor analógico-digital / convertidor digital-analógico.



Dentro del dominio digital, los datos se almacenan en formato binario y con una resolución determinada por las capacidades de memoria y poder de computación de los procesadores. Del mismo modo que una imagen se almacena como una matriz de un número finito de píxeles, una onda sonora se almacena y se procesa como una lista finita de valores de amplitud en el tiempo.

La conversión de las vibraciones sonoras del mundo real, de "resolución infinita", en valores digitales comprensibles por un microprocesador se hacen gracias a un conjunto de componentes y microcontroladores electrónicos llamados **convertidores analógico-digital** (en inglés *analog to digital converters*, ADC).

Mediante transductores electroacústicos (micrófonos), se convierten las vibraciones en datos de audio analógico (cambios de voltaje), y los ADC se encargan de traducir estos valores de voltaje en señales binarias procesables por los ordenadores.

En el caso inverso, el de los **convertidores digital-analógico** (en inglés *digital to analog converters*, DAC), se convierten datos digitales en valores de voltaje que, cuando se aplican a un sistema de amplificador y altavoces, provocan el movimiento de las membranas de estos últimos, que a su vez provocan el movimiento del aire en frecuencias audibles.

1.7. *Sampling rate*

Frecuencia de muestreo

> La **frecuencia de muestreo**, normalmente medida en hercios (valores por segundo), es la resolución en el dominio temporal del que se compone una determinada muestra de audio.

Cuanto más alta sea la frecuencia de muestreo, más valores de amplitud por segundo tendrá la muestra y, por tanto, más resolución. Por poner varios ejemplos, la frecuencia de muestreo de una música comercial distribuida en CD o

MP3 suele ser de 44.100 Hz; la frecuencia de muestreo de la línea de telefonía fija es de 11.000 Hz, y las grabaciones profesionales de audio en estudio se pueden llegar a hacer a 192.000 Hz.

1.8. *Bit depth*

Bit depth: profundidad de bits.

Cada uno de los valores de amplitud (volumen) de una muestra de audio se entrega con un número entre –1 y 1, siendo 0 el valor neutro. Así pues, estamos hablando de números decimales.

Dependiendo de la profundidad de bits, estos números decimales entre –1 y 1 serán más o menos precisos.

El estándar actual está fijado en 16 bits para las grabaciones de audio en CD o MP3, aunque todavía encontramos audio a 8 bits en las videoconsolas portátiles y en pequeños dispositivos electrónicos. Las grabaciones de audio profesional suelen hacerse a 24 o 32 bits.

Así pues, la profundidad de bits define la resolución de los números decimales de los valores de amplitud de una onda sonora.

2. Las herramientas

2.1. Microfonía

La herramienta básica del diseño de interacciones mediante el sonido es el micrófono.

El **micrófono** es un transductor electroacústico. Su función es la de traducir las vibraciones debidas a la presión acústica ejercida sobre su cápsula por las ondas sonoras en energía eléctrica, lo que permite grabar sonidos de cualquier lugar o elemento.

Podemos hallar micrófonos de diferentes formas y sistemas de operación, de los que nos fijaremos en tres: direccionales, omnidireccionales y por contacto.

1) Direccionales

Los micrófonos unidireccionales o direccionales son aquellos micrófonos muy sensibles a una única dirección y relativamente sordos a las restantes. Su principal inconveniente es que no dan una respuesta constante: son más direccionales si se trata de frecuencias altas (agudos) que si son de frecuencias bajas (graves), puesto que la direccionalidad del sonido, como de todo tipo de ondas (ya sea mecánicas o electromagnéticas), depende de su frecuencia. Su principal ventaja es que permiten una captación localizada del sonido. Normalmente, se utilizan acoplados a jirafas de sonido.

2) Omnidireccionales

Los micrófonos omnidireccionales tienen un diagrama polar de 360° (la circunferencia completa).

Este tipo de micrófonos tienen una respuesta de sensibilidad constante, lo que significa que captan todos los sonidos, independientemente de la dirección desde la que lleguen.

Su principal inconveniente es que, al captarlo todo, captan tanto lo que queremos que capten como lo que no: ruido del entorno, reflexiones acústicas, etc.

3) De contacto

Toman el sonido porque están en contacto físico con el instrumento. Se utilizan también como sensores para disparar un sonido de un módulo o *sampler* por medio de un disparador o *trigger* MIDI.

2.2. Tarjetas de sonido

Estos sistemas de microfonía, en el caso del diseño de un sistema interactivo basado en software, se han de conectar a un ordenador para que procese los sonidos recogidos. Para conectar un micrófono al ordenador, usaremos una **tarjeta de audio**. Casi todos los ordenadores actuales llevan tarjetas de audio integradas y suelen incorporar un simple conector mini o *minijack* de 3,5 mm.

La mayoría de sistemas de microfonía profesional se conectan mediante XLR, mucho más estable y libre de ruidos, por lo que recomendamos que, en el caso de que se necesiten datos sonoros netos y fiables para el diseño interactivo, se usen tarjetas de audio externas con entradas y salidas con conectores balanceados y de gran formato.

2.3. Fuentes sonoras

Una decisión importante cuando diseñemos una interacción basada en audio es cuál será la fuente sonora que desencadenará las acciones diseñadas: la voz humana, un instrumento musical, palmadas...

Según cuál sea esta fuente, deberemos elegir el sistema técnico más adecuado: micrófonos de contacto, direccionales, inalámbricos...

3. Diseñando interacciones con sonido

A continuación, enumeraremos algunas de las técnicas que nos permiten interpretar datos para generar reacciones interactivas con el sonido. Como sabéis, una parte importante del diseño de una interacción es escoger qué información se recoge, cómo se analiza y cómo se definen reacciones al respeto. En este apartado, describiremos diversas formas de análisis del sonido obtenido. A partir de estas técnicas, podréis pensar más adelante en reacciones, ya sean en pantalla o mediante actuadores físicos (luces, motores, pistones, etc.).

3.1. Análisis de amplitud

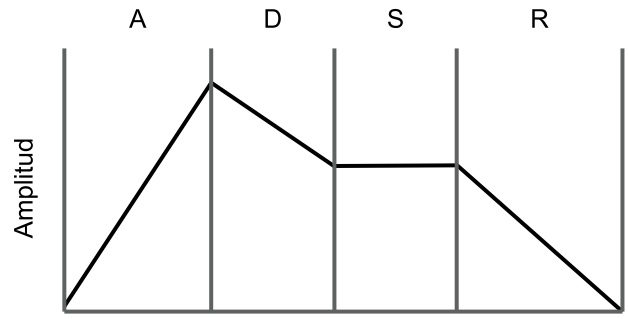
Por medio del análisis del volumen aparente de una muestra de sonido, podemos diseñar toda una serie de interacciones sencillas.

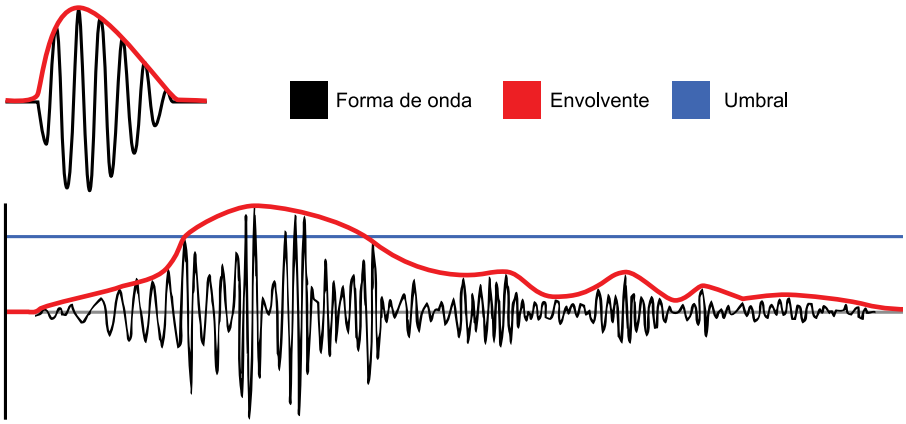
Hemos visto que, estrictamente, la amplitud de una muestra de audio cambia muchas veces en un pequeño fragmento de tiempo (la onda sonora audible más "lenta" tiene una frecuencia de 20 ciclos por segundo). Esto nos podría llevar a la conclusión de que, en realidad, la amplitud de una muestra cambia tantas veces por segundo que es imposible extraer de ella una información útil de cara al diseño de interacción.

1) Seguidores de envolvente

Para superar esta dificultad, se acostumbra a utilizar una técnica de "suavizado" de la onda sonora que permite extraer los datos de amplitud aparente. Esta técnica se conoce como **seguimiento del envolvente** (*envelope following*) y, por medio de la definición de los valores de ataque y caída de la curva de análisis, podemos obtener un muestreo suavizado de los valores de amplitud de la onda sonora.

Normalmente, las envolventes se pueden ajustar mediante cuatro parámetros: ataque, caída, sostenimiento, extinción (ADSR: *attack*, *decay, sustain, release*). Definiendo los tiempos de cada uno de estos parámetros, podemos generar envolventes de todo tipo (la envolvente de amplitud de una nota de piano, por ejemplo, tendría un ataque muy rápido, una caída y un sostenimiento muy breves, y una extinción bastante extensa).





A partir de estos datos que nos ofrece la curva de envolvente (ved imagen), podemos empezar a diseñar interacciones basadas en el volumen del sonido.

2) Definición de umbrales

Quizás el diseño de interacción más popular se base en la definición de un **umbral de amplitud** (*amplitude treshold*), definido en decibelios (dB), a partir del que se dispara un acontecimiento o *event*.

> Pongamos por caso que definimos un umbral alto y obligamos a los usuarios de nuestro diseño interactivo a emitir un sonido muy alto (gritando, por ejemplo) para que se dispare una acción determinada (diferente de un acontecimiento audiovisual, encender una luz, mover un motor, etc.).

3) Cálculo estadístico

Sin embargo, a partir del análisis estadístico y el uso de las matemáticas en general, podemos llegar a diseñar interacciones complejas que tengan en cuenta factores como el nivel de cambio de volumen a lo largo del tiempo.

> Por medio de algoritmos de programación, podríamos definir una interacción basada en el número de veces que se da una palmada, en la progresión de amplitudes (*in crescendo* o *in decrescendo*) o en la frecuencia de variación de amplitudes en el tiempo, por poner solo algunos ejemplos.

Ejemplo

Por ejemplo, para medir la expresividad de un instrumento musical o detectar alteraciones repentinas del espacio sonoro.

3.2. Análisis de frecuencia

Un algoritmo de detección de tono puro o *pure pitch* se encarga de estimar el tono o la frecuencia fundamental de una señal casi periódica o virtualmente periódica proveniente de una grabación digital.

Hay dos métodos populares de seguimiento de tono o *tracking pitch*: **ZCD** (*zero crossing detector*, 'detector de cruces por cero'), en el dominio del tiempo, y **FFT** (*fast fourier transforms*, 'transformada rápida de Fourier'), en el dominio de la frecuencia.

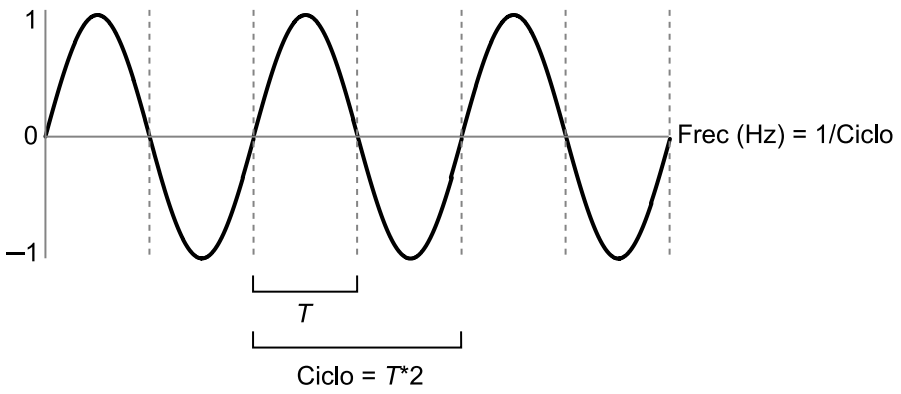
El ZCD es un algoritmo muy ligero que consume pocos recursos y que se puede implementar en la mayoría de sistemas (incluso en Arduino) con capacidad limitada de procesado de señales. Sin embargo, el ZCD funciona razonablemente bien en situaciones como el análisis de tonos telefónicos (DTMF) o en cualquier otro caso de análisis de señales sintéticas y periódicas, pero no es muy fiable cuando la señal que se ha de analizar es muy compleja o rica en armónicos, como señales de audio polifónico o procesamiento del habla.

El análisis por FFT es muy fiable y robusto, a pesar de que es muy exigente en cuanto a carga de procesador y no todos los dispositivos están preparados para hacer tareas de FFT en tiempo real.

1) ZCD

El algoritmo de cruce por cero se basa en la detección del cruce de la señal por el punto de amplitud cero y el cálculo del tiempo entre dos cruces sucesivos. Esta medida del tiempo entre los sucesivos cruces por cero nos da un dato fiable de la frecuencia de la onda sonora, que podemos aislar tal como sigue:

Tiempo entre ceros (en segundos) = T
Frecuencia en hercios = F
F = (1 / T) * 2



2) FFT

El análisis frecuencial mediante FFT (transformada rápida de Fourier) es muy versátil y nos puede dar una gran cantidad de datos útiles en el contexto del diseño de interacciones. *Grosso modo*, el método de Fourier se basa en la reducción de un fragmento (*window*) de la señal a sus componentes sinusoidales para facilitar el análisis matemático. Una vez hecha esta reducción sinusoidal, el algoritmo da información sobre las diferentes frecuencias presentes en la señal y sus valores de amplitud.

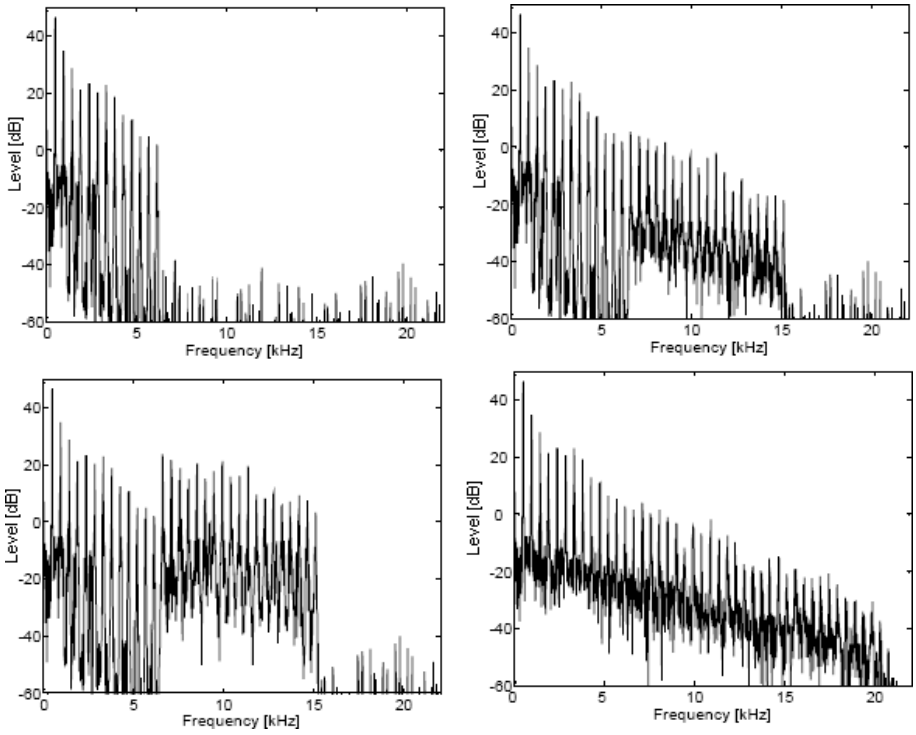
Este método, por tanto, además de darnos información sobre el tono fundamental de una señal de audio monofónico (una trompeta, por ejemplo), es capaz de facilitar una relación de energía para cada banda de frecuencias dentro del espectro sonoro.

Con esta información sobre cada banda del espectro, podemos diseñar un algoritmo de interacción sonora que mida la amplitud de tres bandas frecuenciales (graves, medios y agudos, por ejemplo) y asignar diferentes acontecimientos interactivos a la energía presente en cada una de esas bandas.

Ejemplo

Un algoritmo como este podría ser efectivo en una situación de visualización musical, por ejemplo.

Del mismo modo que con el análisis de amplitud, aplicando algoritmos matemáticos sobre la señal obtenida tras el análisis, podríamos diseñar interacciones más complejas: detección de escalas o patrones musicales, reconocimiento de bandas de formantes (fonemas), detección de estados de ánimos del habla poniendo en relación frecuencias y amplitudes y su evolución en el tiempo...



4. Más allá. Recursos y bibliografía específica

4.1. Otras estrategias de análisis

Reconocimiento del habla: http://en.wikipedia.org/wiki/Speech_recognition

4.2. Música visual

A lo largo de la historia del arte, podríamos encontrar numerosas analogías entre música y artes visuales gracias a artistas que incorporan un cierto sentido **sinestésico** a sus obras, trabajando en el análisis y la generación de pinturas o de música que evocan conceptos o sensaciones extravisuales o extramusicales. Casos de artistas visuales como Wassily Kandinsky son ilustrativos de esta corriente y representan una analogía clara con los métodos modernos de **visualización sonora** ejecutada mediante algoritmos de análisis de audio.

En esta línea, recomendamos que consultéis la obra de artistas como Oskar Fischinger, Norman McLaren o Walter Ruttman, puesto que su trabajo tiene mucho que ver con el contenido de este módulo y ofrece un punto de vista original acerca del análisis y la comprensión de la música como portadora de valores extramusicales.

http://en.wikipedia.org/wiki/Visual_music

4.3. Instituciones y centros de investigación

IRCAM: http://www.ircam.fr

MTG: http://mtg.upf.edu/

4.4. Herramientas especializadas de software

SuperCollider: http://www.audiosynth.com/

ChucK: http://chuck.cs.princeton.edu/

Csound: http://www.csounds.com/

Pure Data: http://puredata.info/

Max/MSP: http://cycling74.com/

Reaktor: http://www.native-instruments.com/

4.5. Bibliografía y recursos en línea

Miller Pucket, *The theory and technique of electronic music*

http://crca.ucsd.edu/~msp/techniques.htm

http://es.wikipedia.org/wiki/Transformada_r%C3%A1pida_de_Fourier

http://en.wikipedia.org/wiki/Pitch_detection_algorithm

http://en.wikipedia.org/wiki/Envelope_detector