

Resumen

El procesado de audio multicanal ha experimentado un gran desarrollo en los últimos años y como consecuencia se ha producido un aumento notable de la complejidad computacional en las nuevas aplicaciones. Actualmente, se pretende que la telecomunicación ofrezca una sensación de cercanía, compartiendo incluso el mismo entorno entre usuarios distantes. Es lo que llamamos: Esquemas de Audio Inmersivo. En este fenómeno intervienen varios efectos acústicos: sonido espacial 3D, compensación de salas, cancelación crosstalk, localización de fuentes sonoras, entre otros. Pero a su vez, para llevar a cabo cualquiera de estos efectos en un sistema real, se necesita una alta capacidad computacional, lo que representa una severa limitación cuando se trata de ejecutar dichas aplicaciones en tiempo real.

El aumento de la capacidad computacional ha ido históricamente unida al número de transistores en un chip. Actualmente, las mejoras en la capacidad computacional están íntimamente ligadas al número de unidades de proceso que tiene un computador, lo que permite un alto grado de paralelismo en computación. Este es el caso de las Unidades de Procesamiento Gráfico (GPUs, *Graphics Processing Units*), que poseen actualmente miles de núcleos computacionales. Las GPUs se han relacionado tradicionalmente con la computación gráfica o el tratamiento de imágenes, pero con la aparición de nuevos entornos de programación para GPUs (CUDA o OpenCL) muchas aplicaciones de otros campos científicos han podido ser aceleradas mediante su implementación en las GPUs. Esta tesis tiene como objetivo desarrollar aplicaciones de audio que necesiten gran cantidad de recursos computacionales, demostrando con ello que las GPUs son herramientas totalmente válidas para llevarlas a cabo. Para ello, se han implementado y evaluado sobre el entorno de programación CUDA diferentes aplicaciones del campo de procesado de señales de audio. También se han analizado y resuelto las posibles limitaciones surgidas durante el proceso de implementación, tanto desde el punto de vista acústico como desde el punto de vista computacional.

En la tesis se han abordado los siguientes problemas:

La primera operación a implementar en GPU era la operación fundamental en el procesado de audio: la convolución, ya que la mayoría de aplicaciones de audio multicanal están basadas en el filtrado masivo. En

principio, la convolución se ha desarrollado como un núcleo computacional, que posteriormente se ha usado para desarrollar una aplicación que combina múltiples convoluciones concurrentemente: cancelación *crosstalk* generalizada y ecualización. La implementación propuesta es capaz de gestionar dos situaciones comunes en el filtrado multicanal: buffers para muestras de audio de tamaños mayores que los tamaños de los filtros; y buffers para muestras de audio de tamaños menores que los tamaños de los filtros.

Se han desarrollado dos aplicaciones de audio espacial a partir del filtrado masivo multicanal que usan las GPUs como co-procesadores. La primera aplicación gira en torno al sonido binaural. Esta aplicación presenta dos características importantes: 1) es capaz de sintetizar fuentes sonoras en posiciones espaciales que no estén incluidas en las bases de datos de los filtros HRTFs, y 2) genera movimientos continuos entre diferentes posiciones. Estas características se han obtenido en la implementación después de diversas pruebas tanto objetivas como subjetivas. Posteriormente, se ha estudiado el máximo número de fuentes sonoras que pueden ser gestionadas por diferentes arquitectura GPU. El mismo estudio se ha llevado a cabo en un sistema de síntesis de onda *Wave Field Synthesis* (segunda aplicación de sonido espacial) compuesto por 96 altavoces. La implementación de este sistema en GPU es capaz de reducir los efectos de sala durante la reproducción.

Otro problema que se ha abordado en esta tesis es la localización de fuentes sonoras en entornos ruidosos y con mucha reverberación. Para este problema se ha propuesto una implementación basada en el algoritmo de localización *Steered Response Power with Phase Transform* (SRP-PHAT) en un sistema multi-GPU. La exactitud en la localización de las fuentes sonoras está íntimamente ligada a una malla espacial de puntos donde se busca la fuente, y al número de micrófonos utilizados en el algoritmo. En esta tesis, se han evaluado las capacidades de las GPUs cuando éstas implementan el algoritmo SRP-PHAT bajo condiciones de tiempo real, atendiendo a diferentes parámetros: tamaño de malla, número de micrófonos, reverberación en la sala, y relación señal a ruido.

Finalmente, esta tesis trata el problema del filtrado masivo multicanal cuando los filtros presentan una respuesta al impulso infinita (*Infinite Impulse Response*, IIR). Se han analizado dos casos particulares: 1) Filtros IIR compuestos de múltiples secciones paralelas de orden dos, y 2) Filtros IIR que presentan una respuesta plana en frecuencia (allpass filters). Ambas estructuras se usan para desarrollar y acelerar dos aplicaciones de audio

diferentes: 1) implementar múltiples ecualizaciones en un sistema WFS, y 2) reducir el margen dinámico en señales de audio.

Palabras Clave: Filtrado Masivo Multicanal, Síntesis de Campo de Ondas, Localización de Fuentes Sonoras, Sonido Espacial, Unidades de Procesamiento Gráfico.