Abstract

Multichannel acoustic signal processing has undergone major development in recent years due to the increased complexity of current audio processing applications. People want to collaborate through communication with the feeling of being together and sharing the same environment, what is considered as Immersive Audio Schemes. In this phenomenon, several acoustic effects are involved: 3D spatial sound, room compensation, crosstalk cancelation, sound source localization, among others. However, high computing capacity is required to achieve any of these effects in a real large-scale system, what represents a considerable limitation for real-time applications.

The increase of the computational capacity has been historically linked to the number of transistors in a chip. However, nowadays the improvements in the computational capacity are mainly given by increasing the number of processing units, i.e expanding parallelism in computing. This is the case of the Graphics Processing Units (GPUs), that own now thousands of computing cores. GPUs were traditionally related to graphic or image applications, but new releases in the GPU programming environments, CUDA or OpenCL, allowed that most applications were computationally accelerated in fields beyond graphics. This thesis aims to demonstrate that GPUs are totally valid tools to carry out audio applications that require high computational resources. To this end, different applications in the field of audio processing are studied and performed using GPUs. This manuscript also analyzes and solves possible limitations in each GPU-based implementation both from the acoustic point of view as from the computational point of view. In this document, we have addressed the following problems:

Most of audio applications are based on massive filtering. Thus, the first implementation to undertake is a fundamental operation in the audio processing: the convolution. It has been first developed as a computational kernel and afterwards used for an application that combines multiples convolutions concurrently: generalized crosstalk cancellation and equalization. The proposed implementation can successfully manage two different and common situations: size of buffers that are much larger than the size of the filters and size of buffers that are much smaller than the size of the filters.

Two spatial audio applications that use the GPU as a co-processor have

been developed from the massive multichannel filtering. First application deals with binaural audio. Its main feature is that this application is able to synthesize sound sources in spatial positions that are not included in the database of HRTF and to generate smoothly movements of sound sources. Both features were designed after different tests (objective and subjective). The performance regarding number of sound source that could be rendered in real time was assessed on GPUs with different GPU architectures. A similar performance is measured in a Wave Field Synthesis system (second spatial audio application) that is composed of 96 loudspeakers. The proposed GPU-based implementation is able to reduce the room effects during the sound source rendering.

A well-known approach for sound source localization in noisy and reverberant environments is also addressed on a multi-GPU system. This is the case of the Steered Response Power with Phase Transform (SRP-PHAT) algorithm. Since localization accuracy can be improved by using high-resolution spatial grids and a high number of microphones, accurate acoustic localization systems require high computational power. The solutions implemented in this thesis are evaluated both from localization and from computational performance points of view, taking into account different acoustic environments, and always from a real-time implementation perspective.

Finally, This manuscript addresses also massive multichannel filtering when the filters present an Infinite Impulse Response (IIR). Two cases are analyzed in this manuscript: 1) IIR filters composed of multiple second-order sections, and 2) IIR filters that presents an allpass response. Both cases are used to develop and accelerate two different applications: 1) to execute multiple Equalizations in a WFS system, and 2) to reduce the dynamic range in an audio signal.

Keywords: Multichannel filtering, Spatial Sound, Sound Source localization, Graphics Processing Units.