2006

3D-audio object oriented coding

Guillaume Potard
University of Wollongong

3D-AUDIO OBJECT ORIENTED CODING

By

Guillaume Potard

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
UNIVERSITY OF WOLLONGONG
NORTHFIELDS AVE
WOLLONGONG NSW 2522
AUSTRALIA
SEPTEMBER 2006

UNIVERSITY OF WOLLONGONG
DEPARTMENT OF
SCHOOL OF ELECTRICAL, COMPUTER AND
TELECOMMUNICATIONS ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Faculty of Informatics for acceptance a thesis entitled "**3D-audio object oriented coding**" by **Guillaume Potard** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: September 2006

External Examiner: _____

Research Supervisor: _____
Ian Burnett

Examing Committee: _____
Peter Svensson

_____
Stephen Barrass

UNIVERSITY OF WOLLONGONG

Date: **September 2006**

Author: **Guillaume Potard**

Title: **3D-audio object oriented coding**

Department: **School of Electrical, Computer and Telecommunications Engineering**

Degree: **Ph.D.** Convocation: **October** Year: **2006**

Permission is herewith granted to University of Wollongong to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Signature of Author

I would to thank my supervisor, Ian Burnett, who guided me through the years of the PhD with useful advice and encouragement.

I then dedicate this thesis to my partner, Megan Sproats, for all her support, patience and love.

Thank you to my parents for helping me to study in Australia and a big hello to my brother.

Table of Contents

List of Tables

List of Figures

ABSTRACT

This thesis first presents a novel object-oriented scheme which provides for extensive description of time-varying 3D audio scenes using XML. The scheme, named XML3DAUDIO, provides a new format for encoding and describing 3D audio scenes in an object oriented manner. Its creation was motivated by the fact that other 3D audio scene description formats are either too simplistic (VRML) and lacking in realism, or are too complex (MPEG-4 Advanced AudioBIFS) and, as a result, have not yet been fully implemented in available decoders and scene authoring tools. This thesis shows that the scene graph model, used by VRML and MPEG-4 AudioBIFS, leads to complex and inefficient 3D audio scene descriptions. This complexity is a result of the aggregation, in the scene graph model, of the scene content data and the scene temporal data. The resulting 3D audio scene descriptions, are in turn, difficult to re-author and significantly increase the complexity of 3D audio scene renderers. In contrast, XML3DAUDIO follows a new scene orchestra and score approach which allows the separation of the scene content data from the scene temporal data; this simplifies 3D audio scene descriptions and allows simpler 3D audio scene renderer implementations. In addition, the separation of the temporal and content data permits easier modification and re-authoring of 3D audio scenes. It is shown that XML3DAUDIO can be used as a new format for 3D audio scene rendering or can alternatively be used as a meta-data scheme for annotating 3D audio content.

Rendering and perception of the apparent extent of sound sources in 3D audio displays is then considered. Although perceptually important, the extent of sound sources is one the least studied auditory percepts and is often neglected in 3D audio displays. This research aims to improve the realism of rendered 3D audio scenes by reproducing the multidimensional extent exhibited by some natural sound sources (eg a beach front, a swarm of insects, wind blowing in trees etc). Usually, such broad

sound sources are treated as point sound sources in 3D audio displays, resulting in unrealistic rendered 3D audio scenes. A technique is introduced whereby, using several uncorrelated sound sources, the apparent extent of a sound source can be controlled in arbitrary ways. A new hypothesis is presented suggesting that, by placing uncorrelated sound sources in particular patterns, sound sources with apparent shapes can be obtained. This hypothesis and the perception of vertical and horizontal sound source extent are then evaluated in several psychoacoustic experiments. Results showed that, using this technique, subjects could perceive the horizontal extent of sound sources with high precision, differentiate horizontally from vertically extended sound sources and could identify the apparent shapes of sound sources above statistical chance. In the latter case, however, the results show identification less than 50 % of the time, and then only when noise signals were used. Some of these psychoacoustic experiments were carried out for the MPEG standardisation body with a view to adding sound source extent description capabilities to the MPEG-4 AudioBIFS standard; the resulting modifications have become part of the new capabilities in version 3 of AudioBIFS.

Lastly, this thesis presents the implementation of a novel real-time 3D audio rendering system known as CHESS (Configurable Hemispheric Environment for Spatialised Sound). Using a new signal processing architecture and a novel 16-speaker array, CHESS demonstrates the viability of rendering 3D audio scenes described with the XML3DAUDIO scheme. CHESS implements all 3D audio signal processing tasks required to render a 3D audio scene from its textual description; the definition of these techniques and the architecture of CHESS is extensible and can thus be used as a basis model for the implementation of future object oriented 3D audio rendering systems.

Thus, overall, this thesis presents contributions in three interwoven domains of 3D audio: 3D audio scene description, spatial psychoacoustics and 3D audio scene rendering.