I-Sounds
Emotion-based Sound Generation for Virtual Environments

Ricardo Cruz

Contents

1 Introduction

Sound is an essential part of our perceptual world, and it is difficult to imagine our lives without them. We live surrounded by sounds and we use them to navigate and orientate ourselves in our daily life. This perceptual importance may pass unnoticed, for most of us since most of our sound processing processes are instinctive. However, the human being is also able to manipulate sounds and use them to achieve a purpose. Since the pre-historic era, humans' use sounds their behalf. With the evolution of mankind, this use become more refined and elaborate, and if initially sound was used to solve practical problems like, hunting or warning, it evolved for other uses like ceremonial practices. At some point in this evolution, man, started to combine sounds, manipulate pitch and duration and the result of this creative activity is music. We can say that music is a combination of sounds such as, an idea; emotion or message can be expressed and shared.

1.1 The I-Sounds project

The I-Sounds project explores the relation between sounds and emotions. How we can express emotional content trough sound. The objective is to generate simple sounds (we can even talk about simple melodies) that can "illustrate" an emotion or a particular mood. Starting in an emotional state, the input, the output will be a sound that somehow conveys and represents that emotional state. Emphasis will be put in rhythm and metre. Rhythmic constructions have an important role on music and sound perception. The properties of salience and kinesis will be the construction base for emotive sound. Melodic issues are also important; in this stage we will use some melodic dimensions like, mode.

The I-Sounds system should fulfil the properties of generality and extensibility. It should be general enough to be used in different contexts and in conjunction with other existent systems providing easy integration facilities. Extensibility is also an important issue. The architecture should provide extension mechanisms so it can be tuned to specific needs. Also the I-Sounds core should be extensible so the system main algorithms can be refined and improved.

1.2 Context

New multi-modal interactive systems require the integration of different media channels to provide an "improved" experience. The auditory channel is of course one those channels that can help in achieving that immersive experience.

The I-Sounds project appears in the context of an interactive system, more precisely an interactive narrative environment (the I-Shadows system).The aim of I-Shadows, is to create interactive narratives where the story emerges from the cooperation between computer controlled characters and user interaction. The user is able to manipulate its own shadows and by analysing this behaviour the computer is able to cooperate with him. The auditory channel can contribute for the intelligibility of the narrative. The I-Sounds project then appears, as the support for that auditory component. The I-Shadows project was the motivation and will be the testing ground for the applicability of I-Sounds under such environments.

Although the aim of I-Sounds is to achieve an abstraction level suitable for other applications and integration. If achieved, I-Sounds will be a general emotion-sound mapping system that can be used not only under narrative environments but also in other contexts.

1.3 The problem

The main problem is how to map emotions into sound. This problem can be subdivided into other two problems. The first one is the identification of the sound parameters and properties whose manipulation has emotional potential and how to release that potential; the second one is how we can express particular emotions such as, happiness or sadness, anger or indifference.

Not only we need to identify the meaningful sound parameters but also we need to study how they work in conjunction and how their values can express emotions.

Summarizing the problem that I-Sounds aim to solve; from the affective information to the final audible signal, the proposition of a mapping process.

2 Proposal

2.1 Theory context and related work

The theory research was compiled in a survey, produced in the first months of the project. This survey is a review of the related research areas and some related projects.

The document is divided into three main sections. The first section resumes the sound and music computing field with its main trends and open issues. Although a more theoretical view, and not centralized on the question of mapping emotions into sounds, this review provides the necessary background that can be used to situate the I-Sounds project inside the field. Also since this is a new area inside the research group, this summary provides and introduction to the field and a collection of important references as well. The second section of the survey relates to music, and musical issues, it presents a small overview over the related musical work and the presentation of some important theories that will form the musical base of the project. This section is divided into melodic and rhythmic aspects. Although this work will explore the rhythmic dimension, some melodic tools will also be used as a support. Finally, the third section discusses a practical application of the use of music to convey messages, ideas and also emotions. Film scoring born with the seventh art and since then it evolved into a specialized are in film production. Theories and models evolved into standards which provide a basis to the field. The survey focuses the functions that a film score can have in a movie.

The complete survey can be found in Anex A.

2.2 Architecture

2.3 iSounds General Architecture

Figure 1 presents a high level view of the iSounds architecture.

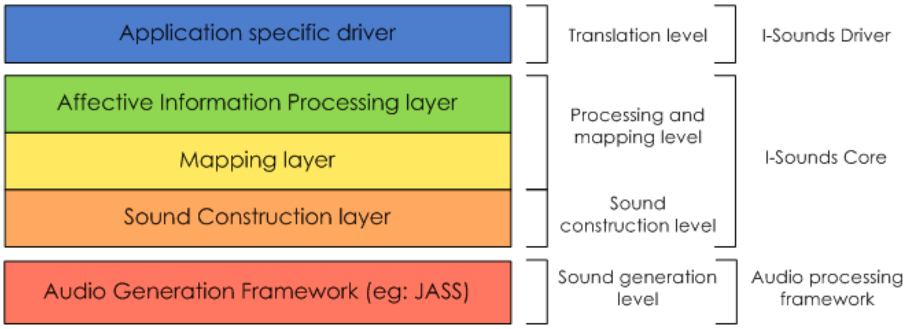

Figure 1: High-level view of the iSounds architecture

Translation Level. This level is the interface between the iSounds and other systems. It consists of a driver that must be written specifically for each system. This driver should make the translation between the iSounds and application domains. The iSounds system should be independent from other systems to guarantee generality and its ability to integrate with different systems. The iSounds system will provide an API for writing these drivers. It is the integrator or developer responsibility to write this module.

Processing and Mapping Level. In this level affective information is processed and then mapped into sounds. The processing stage may include, serialization, conflict resolving, packing and the computation of the emotional context. The mapping stage is supported by the mapping algorithm which takes the emotional context as input and quantifies it in terms of musical parameters such as, salience, kinesis, mode, etc.

Sound Construction Level. In the sonification level, musical parameters values (earlier quantified) will be used to compute a rhythmic and musical sequence. This musical sequence represents the emotional context. This algorithm will be the basis of all the system and can be thought as the core of iSounds. The virtual composer of iSounds is this level. Sounds will have a defined representation that can be used to generate audible signals or feed other system for further processing.

Sound Generation Level. This level will transform that intermediate representation of sounds into audible signals. It will send them to the system speakers and so they can be injected into the environment. This level will use an existent audio generation framework. The next section will discuss this in further detail.

2.3.1 Connection with I-Shadows.

Figure 2 presents the connection of the iSounds system with I-Shadows, which will be the testing architecture.

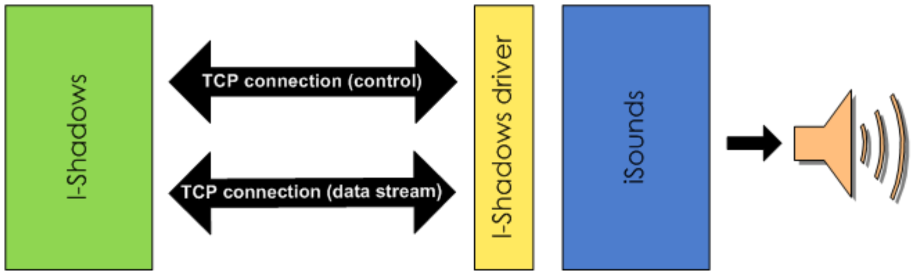

Figure 2: Architecture of the connection between the iSounds system and I-Shadows

2.4 Audio Framework Analysis

An audio generation framework is necessary in order to support sound generation. Several options were analysed but the initial set was reduced to three candidates, CLAM, JASS and the Synthesis Toolkit in C++. Each of these was deeper analysed.

The CLAM framework is developed in the Music Technology Group of the Pompeu Fabra University, Barcelona, Spain. It is an audio framework for research and application development. The framework provides all the tools necessary to build any kind of sound processing application. It provides tools for analysis, synthesis and transformation of audio signals and it establishes a conceptual model of the process. It has some GUI tools designed to help the use of the framework and to do rapid prototyping of sound synthesis systems. It is written in the C++ language and it is platform independent.

JASS stands for, Java Audio Synthesis System; it is a library written in pure JAVA for audio synthesis based on the concept of unit generators. JASS is developed by Kees van den Doel in the University of British Columbia, Vancouver, Canada. Specific unit generators are built by extending and implementing the foundation abstract classes and interfaces. The JASS library has a reduced number of classes and it offers simplicity and easy usability.

The Synthesis Toolkit in C++ is developed in collaboration between Gary Scavone from McGill Univeristy and Perry Cook from Princeton University. It is written in C++, and offers signal processing and algorithmic audio synthesis. The STK library is cross platform and it was designed, looking for real time control, ease of use and education purposes. STK is able to generate output in audio files or in real time. The RtMidi API provides real-time MIDI input/output and the RtAudio API provides support for real time input/output audio streaming. The STK library

Table 1 presents a comparison between the three frameworks:

After analysing the options and doing some testing, the chosen library was JASS. Its simplicity and its previous application in interactive systems are the main reasons, but also the fact that is built in JAVA which makes code portability completely transparent.

	CLAM	JASS	STK C++
Analysis Support	x		
Synthesis Support	x	x	x
Language	C++	JAVA	C++
Cross Platform	x	x	x
Real-time Support	x	x	x
License	GPL	Free	Free
Learning Curve	Longer	Shorter	Medium

Table 1: Audio framework comparison table

2.5 To do

Requirements identification This stage is not yet completed. A complete identification of requirements should be done before the final architecture design. This item is a little out of schedule.

High level architecture design This stage is currently under way. A preliminary version is presented in figure 1.

Detailed architecture Once the complete list of requirements is done, then the details of the architecture should be designed and specified. This document should be the guide to the code production.

Mapping strategy design The success of the iSounds system highly depends on the success of the mapping function between emotions and sound. This work will be developed in high cooperation with a musician and music professor. This stage includes some tests also.

Code production The code production stage should begin immediately, after the full detailed architecture.

Testing In the testing stage, the iSounds project will be tested together with the I-Shadows system. Changes to the algorithms can emerge here. The results of this stage will be important to validate the proposed solution.

Thesis production Thesis production will be made in parallel with the other project activities. The final month will be exclusively dedicated to this task.

2.6 Project schedule

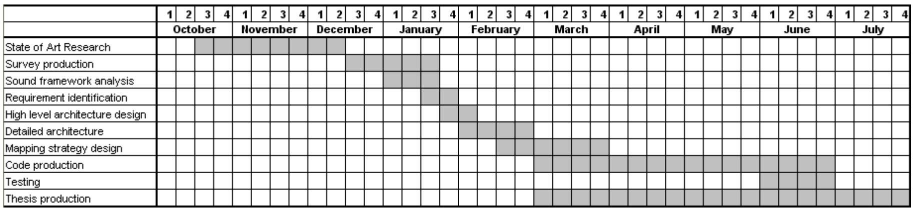

Figure 3: Project schedule