The SoundFont® 2.0 File Format

A White Paper

by Dave Rossum
Joint E-mu/Creative Technology Center

Introduction

In 1993, E-mu Systems realized the importance of establishing a single universal standard for downloadable sounds for sample based musical instruments. The sudden growth of the multimedia audio market had made such a standard necessary. E-mu's experience as a leader in sample based music synthesis led us to devise the SoundFont® file format standard as a solution.

The SoundFont file format was originally introduced with the Creative Technology Sound Blaster AWE32 product using the EMU8000 synthesizer engine. Since that introduction, E-mu and Creative have made evolutionary improvements in the SoundFont file format standard. Our resulting experience with the issues have given us the confidence to announce public disclosure of the SoundFont file format in its revision 2.0 embodiment.

A Brief History of Music Synthesis

The electronic music synthesizer was invented simultaneously by a number of individuals in the early 1960's, most notably Robert Moog and Donald Buchla. The synthesizers of the 1960's and 1970's were primarily analog, although by the late 70's computer control was becoming popular.

With the advances in consumer electronics made possible by VLSI and digital signal processing (DSP), it became practical in the early 1980's to replace the fixed single cycle waveforms used in the sound producing oscillators of synthesizers with digitized waveforms. This development forked into two paths. The professional music community followed the line of "sample based music synthesizers," notably the Emulator line from E-mu Systems. These instruments contained large memories which reproduced an entire recording of a natural sound, transposed over the keyboard range and appropriately modulated by envelopes, filters and amplifiers. The low cost personal computer community instead followed the "wavetable" approach, using tiny memories and creating timbre changes on synthetic or computed sound by dynamically altering the stored waveform.

During the 1980's, another relatively low cost music synthesis technique using frequency modulation (FM) became popular first with the professional music community, later transferring to the PC. While FM was a low cost and highly versatile synthesis technology, it could not match the realism of sample based synthesis, and ultimately it was displaced by sample based approaches in professional studios.

During the same time frame, the Musical Instrument Digital Interface (MIDI) standard was devised and accepted throughout the professional music community as a standard for the real-time control of musical instrument performances. MIDI has since become a standard in the PC multimedia industry as well.

The professional sample based synthesizers expanded in their capabilities in the early 1990's, to include still more DSP. The declining cost of memory brought to the wavetable approach the ability to use sampled sounds, and soon wavetable technology and sample sound synthesis became synonymous. In the mid '90s wavetable synthesis became inexpensive enough to incorporate in mass market products. These wavetable synthesizer chips allow very good quality music synthesis at popular prices, and are currently available from a variety of vendors. While many of these chips operate from samples or wavetables stored in read only memory (ROM), a few allow the downloading of arbitrary samples into RAM memory.

What Is the SoundFont File Format?

A SoundFont compatible file is, as the name implies, the audio equivalent of a character font. SoundFont compatible files are designed to present the information required to produce wavetable based musical instrument banks in a relatively implementation-independent format. They are also designed to present this information is a manner that is relatively compact and appropriately hierarchical.

The Musical Instrument Digital Interface (MIDI) language has become a standard in the PC industry for the representation of musical scores. MIDI allows for each line of a musical score to control a different instrument, called a preset. The General MIDI extension of the MIDI standard establishes a set of 128 presets corresponding to a number of commonly used musical instruments.

While General MIDI provides composers with a fixed set of instruments, it neither guarantees the nature or quality of the sounds those instruments produce, nor does it provide any method of obtaining any further variety in the basic sounds available. Various musical instrument manufacturers have produced extensions of General MIDI to allow for more variations on the set of presets. It should be clear, however, that the ultimate flexibility can only be obtained by the use of downloadable digital audio files for the basic samples.

The SoundFont file format differs from previous digital audio file formats in that files contain not only the digital audio data representing the musical instrument samples themselves, but also the synthesis information required to articulate this digital audio. A SoundFont compatible file represents a set of musical keyboards, each of which is associated with a MIDI preset. Each MIDI "preset" or keyboard of sound causes the digital audio playback of an appropriate sample contained within the SoundFont file. When this sound is triggered by the MIDI key-on command, it is also appropriately in a manner controlled by the MIDI parameters of note number, velocity, and the applicable continuous controllers. Much of the uniqueness of the SoundFont file format rests in the manner in which this articulation data is handled.

SoundFont compatible files are formatted using the "chuck" concepts of the standard Resource Interchange File Format (RIFF) used in the PC industry. Use of this standard format shell provides an easily understood hierarchical level to the SoundFont file format.

Issues in a Universal Synthesizer Data Format

The General MIDI standard was an attempt to define the available instruments in a MIDI composition in such a way that composers could produce songs and have a reasonable expectation that the music would be acceptably reproduced on a variety of synthesis platforms. Clearly this was an ambitious goal; from the two operator FM synthesis chips of the early PC synthesizers, through sampled sound and "wavetable" synthesizers and even "physical modeling" synthesis, a tremendous variety of technology and capability is spanned. The fact that many composers are disappointed in the results of the General MIDI standard is not surprising.

The task attempted by the SoundFont format is relatively simpler, but still by no means trivial. A SoundFont compatible file represents information to be loaded by a specific type of synthesizer technology - the sampled sound or modern "wavetable" synthesizer. Like General MIDI, the SoundFont format assumes only minimal basic capabilities of the synthesizer, but supports enhancements in an upwardly compatible manner. Most of the issues in the design of SoundFont compatible files are based around determining a format which can appropriately encapsulate minimal capabilities in a machine independent format, and yet allow for greater complexity as it becomes available.

Even something as seemingly straightforward as presenting the sample data itself is not a trivial issue. What resolution or word size(s) should be supported? Should data compression be employed, and if so, what method should be used? Are there any standards that must be followed by the samples themselves such that they can be reproduced with optimal fidelity on a variety of synthesis hardware platforms?