CRC Press
Taylor & Francis Group

Audio Effects

Theory, Implementation and Application

Joshua D. Reiss
Andrew P. McPherson

AUDIO EFFECTS

Theory, Implementation and Application

Joshua D. Reiss
Queen Mary University of London, United Kingdom

Andrew P. McPherson
Queen Mary University of London, United Kingdom

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20140911

International Standard Book Number-13: 978-1-4665-6028-4 (Hardback)

Library of Congress Cataloging-in-Publication Data

Reiss, Joshua D.
Audio effects : theory, implementation, and application / Joshua D. Reiss, Andrew P. McPherson.
pages cm
Includes bibliographical references and index.
ISBN 978-1-4665-6028-4 (hardback)
1. Computer sound processing. 2. Sound--Recording and reproducing--Digital techniques. 3. Signal processing--Digital techniques. I. McPherson, Andrew P. II. Title.

TK7881.4.R45 2014
621.382'2--dc23 2014033468

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

Audio effects are used in broadcasting, television, film, games, and music production. Where once they were used primarily to enhance a recording and correct artifacts in the production process, now they are used creatively and pervasively.

The aim of this book is to describe the theory behind the effects, explain how they can be implemented, and illustrate many ways in which they can be used. The concepts covered in this book have relevance to sound engineering, digital signal processing, acoustics, audio signal processing, music informatics, and related topics.

Both authors have taught courses on this subject. We are aware of excellent texts on the use of audio effects, especially for mixing and music production. We also know excellent reference material for audio signal processing and for audio effect research. But it was still challenging to find the right material that teaches the reader, from the ground up, how and why to create audio effects, and how they are used.

That is the purpose of this book. It provides students and researchers with knowledge of how to use the tools and the basics of how they work, as well as how to create them. It is primarily educational, and geared toward undergraduate and master's level students, though it can also serve as a reference for practitioners and researchers. It explains how sounds can be processed and modified by mathematical or computer algorithms. It teaches the theory and principles behind the full range of audio effects and provides the reader with an understanding of how to analyze, implement, and use them.

We chose not to shy away from giving the math and science behind the implementations and applications. Thus, it is one of the few resources for use in the classroom with a mathematical and technical approach to audio effects. It provides a detailed overview of audio effects and example questions to aid in learning and understanding. It has a special focus on programming and implementation with industry standards and provides source code for generating plug-in versions of many of the effects.

Chapter 1 begins by covering some fundamental concepts used often in later chapters. It also introduces the notation that we use throughout. Here, we describe some essential concepts from digital signal processing, thus allowing the subject matter to be mostly self-contained, without the reader needing to consult other texts.

In Chapter 2, we introduce delay lines and related effects such as delay, vibrato, chorus, and flanging. These are some of the most basic effects, and the concept of delay lines is useful for understanding implementations of the effects introduced in later sections.

Chapter 3 then covers filter fundamentals. We chose a quite general approach here and introduce techniques that allow the reader to construct a wide variety of high-order filters. Attention is also paid to some additional filters often used in other effects, such as the allpass filter and the exponential moving average.

In Chapter 4, we explore filters in more detail, covering effects that have filters as their essential components. These include the graphic and parametric equalizer, wah-wah, and phaser.

We then move on to nonlinear effects. Chapter 5 discusses modulation, focusing primarily on tremolo and ring modulation. Chapter 6 goes into detail on dynamics processing, especially the dynamic range compressor and the noise gate. Here, much emphasis is given to correct implementation and perceptual qualities of these effects. Chapter 7 then covers distortion effects. These are concerned with the sounds that result from highly nonlinear processing, beyond the dynamics processors of the previous chapter.

Having introduced the important signal processing concepts, we can now move on to the phase vocoder and introduce several effects that do their processing in the frequency domain. This is the focus of Chapter 8.

Up to this point, none of the effects has attempted to recreate how a natural sound might be perceived by a human listener in a real acoustic space. The next three chapters deal with spatial sound reproduction and spatial sound phenomena. Chapter 9 covers some of the main spatialization techniques, starting with panning and precedence, as can be used in stereo positioning, and then moves on to techniques requiring more and more channels, vector-based amplitude panning, ambisonics, and wave field synthesis. The final technique describes binaural sound reproduction using head-related transfer functions (HRTFs) for listening with headphones.

Chapter 10 covers the Doppler effect, which is a physical phenomenon. This short chapter gives both a general derivation and details of implementation as an audio effect based on delay lines. In Chapter 11, we move on to reverberation, describing both algorithmic and convolutional approaches. Though grouped together with the other chapters concerned with spatial sound, the reverberation approaches described here do not necessarily require the processing of two or more channels of audio.

Chapter 12 is about audio production. This is, of course, a very broad area, so we focus on the architecture of mixing consoles and digital audio workstations, and how the effects we have described may be used in these devices. We then discuss how to order and combine the audio effects in order to accomplish various production challenges.

Finally, Chapter 13 is about how to build the audio effects as software plug-ins. We focus on the C++ Virtual Studio Technology (VST) format, which is probably the most popular standard and available for most platforms and hosts. This chapter (and to some extent, Chapter 12) may be read at any point, or independently of the others. It makes reference to the effects discussed previously, but the chapter is focused on practical implementation. It complements the supplementary material, which includes source code that may be used to build VST plug-ins for a large number of effects described in the book.

The text has benefitted greatly from the comments of expert reviewers, most notably Dr. Pedro Duarte Pestana. We are also deeply indebted to Brecht De Man, who revised the audio effects source code, as well as contributed several implementations. This book would also not have been possible without all of the excellent work that has been done before. We are indebted to various people whose work is frequently cited throughout the text: Julius Smith, Roey Izhaki, Udo Zoelzer, Ville Pulkki, and Sophocles Orfanidis, to name just a few. The errors and omissions are ours, whereas the best explanations are found in the works of the cited authors.

About the Authors

Joshua D. Reiss, PhD, member of IEEE and AES, is a senior lecturer with the Centre for Digital Music in the School of Electronic Engineering and Computer Science at Queen Mary University of London. He has bachelor degrees in both physics and mathematics, and earned his PhD in physics from the Georgia Institute of Technology. He is a member of the Board of Governors of the Audio Engineering Society and cofounder of the company MixGenius. Dr. Reiss has published more than 100 scientific papers and serves on several steering and technical committees. He has investigated music retrieval systems, time scaling and pitch shifting techniques, polyphonic music transcription, loudspeaker design, automatic mixing for live sound, and digital audio effects. His primary focus of research, which ties together many of the above topics, is on the use of state-of-the-art signal processing techniques for professional sound engineering.

Andrew P. McPherson, PhD, joined Queen Mary University of London as lecturer in digital media in September 2011. He holds a PhD in music composition from the University of Pennsylvania and an M.Eng. in electrical engineering from the Massachusetts Institute of Technology. Prior to joining Queen Mary, he was a postdoc in the Music Entertainment Technology Laboratory at Drexel University, supported by a Computing Innovation Fellowship from the Computing Research Association and the National Science Foundation (NSF). Dr. McPherson's current research topics include electronic augmentation of the acoustic piano, new musical applications of multitouch sensing, quantitative studies of expressive performance technique, and embedded audio processing systems. He remains active as a composer of orchestral, chamber, and electronic music, with performances across the United States and Canada, including at the Tanglewood and Aspen music festivals.

Contents

1

Introduction and Fundamentals

In digital audio signal processing and digital audio effects, we are primarily concerned with systems that take a discrete, uniformly sampled audio signal, process it, and produce a discrete, uniformly sampled output audio signal. Therefore, we start by introducing some fundamental properties of sound that are used over and over again, then how we represent it as a digital signal, and then we move on to how we describe the systems that act on and modify such signals. This is not meant to give a detailed overview of digital signal processing, which would involve discussion of continuous time signals, infinite signals, and mathematical relationships. Rather, we intend to focus on just the type of signals and systems that are encountered in audio effects, and on the most useful properties and representations. Having said that, this is also intended to be self-contained. Very little prior knowledge is assumed, and it should not be necessary to refer to more detailed discussions in other texts in order to understand these concepts.

Understanding Sound and Digital Audio

Fundamentally, all audio is composed of waveforms. Vibrating objects create pressure waves in the air; when these waves reach our ears, we perceive them as sound. With the invention of the telephone in the 19th century, audio was first encoded as an electric signal, with the changes in electric voltage representing the changes in pressure over time. Until the late 20th century, electric recording and transmission was all analog: sound was represented by a continuous waveform over time.

In this book, we will work almost exclusively with digital audio. Rather than representing audio as a continuous voltage, as in analog, the waveform will be composed of discrete samples over time. These samples can be stored, processed, and ultimately reconstructed as sound we can hear. Digital audio systems generally begin with an *analog-to-digital converter* (ADC), which captures periodic snapshots of the electrical voltage on an audio transmission line and represents these snapshots as discrete numbers. By capturing the voltage many thousands of times per second, one can achieve a very close approximation of the original audio signal. This encoding method is known

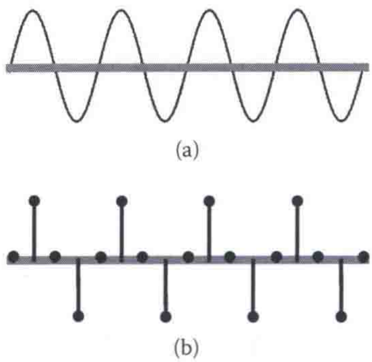

FIGURE 1.1
A continuous time signal (a), and its digital representation, found by sampling the signal uniformly in time (b).

as *pulse code modulation*, and is the encoding format used in the WAV and AIFF audio formats. Pulse code modulation is also one of the most popular forms of ADC, and certainly one of the simplest to explain.

Thus, a continuous time audio signal, such as captured from a microphone, is represented as a digital signal with uniform timing between samples (see Figure 1.1). But digital audio signals need not be derived from analog, nor even represent any physical sound. They can be completely synthetic, and generated using digital signal processing techniques. We will touch on this later in the text when discussing low-frequency oscillators (Chapter 2), phase vocoders (Chapter 8), and other concepts. It is important to note that unless additional information is stored, there is no distinction between those digital audio signals that were generated from conversion of analog signals and those that were generated from digital sound synthesis techniques (though, of course, real-world signals are likely to have more noise and more complex phenomena).

There are three important characteristics of almost any digital audio data: sample rate, bit depth, and number of channels.

Sample rate is the rate at which the samples are captured or played back. It is typically measured in Hertz (Hz), or cycles per second. In this case, one cycle represents one sample. An audio CD has a sample rate of 44,100 Hz, or 44.1 kHz. Higher sampling rates allow a digital recording to accurately record higher frequencies of sound, or to provide a safety margin in case of additional noise or artifacts introduced in the recording, processing, or playback; 48 kHz is often used in audiovisual production, and sample rates of 96 or 192 kHz are used in high-resolution audio, such as in DVD-Audio, or in professional audio production.

The *bit depth* specifies how many bits are used to represent each audio sample. The most common choices in audio are 16 bit and 24 bit. The bit depth also determines the theoretical dynamic range of the audio signal. In digital audio, amplitude is often expressed as a unitless number, representing a ratio between the current intensity and the highest (or lowest) possible intensity that can be represented. The maximum absolute value for this ratio is known as the *dynamic range*. In an ideal ADC, the dynamic range, in decibels (see below), is very roughly 6.02 times the number of bits. Thus, 16-bit audio could represent signals whose loudness ranges over 96 dB, e.g., from a quiet whisper to a loud rock concert.

The *number of channels* actually refers to the fact that audio content will often be composed of several different channels, each one representing its own signal. This is most often the case in stereo or surround sound, where each channel may represent the sound sent to each loudspeaker. Monaural audio, however, is typically encoded as a single channel. We will return to these concepts in Chapter 9.

Digital audio may be encoded with or without *data compression*. When data compression is used, sophisticated algorithms are used to encode and re-represent the data such that they take up much less space. Hence, a decoder must be used to convert the data back into time domain samples before playback. The compression can be either *lossless* (the decoded data are identical to the original data before compression) or *lossy*. Modern lossy audio compression techniques use knowledge of psychoacoustics to minimize the perceived degradation of audio that occurs when a substantial amount of the information contained in the original signal is discarded.

Data compression also introduces one more characteristic of audio data, the *bit rate*. This is the number of bits per unit of time. For lossless signals, this is simply the bit depth times the sample rate times the number of channels. For instance, CD audio would typically have a bit rate of 1,411.2 kbps (kilobits per second):

$$16\frac{\text{bits}}{\text{sample}}\cdot 44100\frac{\text{samples}}{\text{second}}\cdot 2\,(\#\,\text{channels}) = 1411200\frac{\text{bits}}{\text{second}} \tag{1.1}$$

For audio signals that have undergone lossy compression, the bit rate is usually greatly reduced. Most compression schemes, including mp3 and aac, transmit audio with a bit rate between 30 and 500 kbps.

It should be noted that there is a lot of fine detail regarding quantization, sampling, dynamic range, and lossy compression of audio data that has been omitted here. For the purpose of this text, it is sufficient to know the format and general meaning of these concepts, but the reader is also encouraged to refer to signal processing texts for more detailed discussion [1–5].

WHY 44.1 KHZ?

Perhaps the most popular sample rate used in digital audio, especially for music content, is 44.1 kHz, or 44,100 samples per second. The short answer as to why it is so popular is simple; it was the sample rate chosen for the Compact Disc and, thus, is the sample rate of much audio taken from CDs, and the default sample rate of much audio workstation software.

As to why it was chosen as the sample rate for the Compact Disc, the answer is a bit more interesting. In the 1970s, when digital recording was still in its infancy, many different sample rates were used, including 37kHz and 50 kHz in Soundstream's recordings [6]. In the late 70s, Philips and Sony collaborated on the Compact Disc, and there was much debate between the two companies regarding sample rate. In the end, 44.1 kHz was chosen for a number of reasons.

According to the Nyquist theorem, 44.1 kHz allows reproduction of all frequency content below 22.05 kHz. This covers all frequencies heard by a normal person. Though there is still debate about perception of high frequency content, it is generally agreed that few people can hear tones above 20 kHz.

This 44.1 kHz also allowed the creators of the CD format to fit at least 80 minutes of music (more than on a vinyl LP record) on a 120 millimeter disc, which was considered a strong selling point.

But 44,100 is a rather special number: $44{,}100 = 2^2 \times 3^2 \times 5^2 \times 7^2$, and hence, 44.1kHz is actually an easy number to work with for many calculations.

Working with Decibels

We often deal with quantities that can cover a very wide range of values, from very large to very small. The *decibel scale* is a useful way to represent such quantities. The *decibel* (dB) is a logarithmic representation of the ratio between two values. Typically, both values represent power, and hence, the decibel is unitless. One of these values is usually a reference, so that the decibel scale can represent absolute levels. The decibel representation of a level is then 10 times the logarithm to base 10 of the ratio of the two power quantities. Since power is usually the square of a magnitude, we can write a value in decibels in terms of the magnitudes or powers as

$$x_{dB} = 10\log_{10}\left(x^2/x_0^2\right) = 20\log_{10}\left(|x|/|x_0|\right) \tag{1.2}$$

If not specified, x_0 is usually assumed to be 1. So, for example, 1 million is 60 dB, and 0.001 is –30 dB. Whether a decibel or linear scale is used often depends just on which one best conveys the relevant information.