6

Musical Instruments and Sound Synthesis

Music is different from speech in that its role is not so much to convey linguistic and conceptual content – although it may have this role also – as it is to evoke an aesthetic and emotional experiences. Both speech and music are sort of 'utility' sounds that are experienced mostly with a 'positive' attitude, although in some circumstances they may become annoying or disturbing: noise.

In addition to speech, *music* is the second major form of acoustic communication between humans. The psychoacoustics of intervals and melodies in music will be discussed later in this book in Section 11.6. We begin with the discussion of the formation of sounds in acoustical and electric musical instruments in this section.

Acoustic instruments generate a large set of different sounds (Fletcher and Rossing, 1998). If electroacoustic and electric instruments are also accounted for, the discriminable set of different instrument sounds is, in practice, infinite. However, most music and speech is composed of a limited set of types and structures of sounds, where certain basic properties hold. For example, woodwind or plucked string instruments form distinguishable families of musical instruments, where the principle of sound generation and the structure of the sound signal are similar within each family. We will discuss shortly some basic properties of acoustic and electric instruments and their sounds.

6.1 Acoustic Instruments

6.1.1 Types of Musical Instruments

There exists a vast variety of more or less different musical instruments. In this book we are mainly interested in the sounds they produce and radiate to the environment. To deepen our interest, a basic understanding of their working principles is required.

von Hornbostel and Sachs (1961) classified musical instruments into four different categories based on how they produce sound:

Communication Acoustics: An Introduction to Speech, Audio, and Psychoacoustics, First Edition.
Ville Pulkki and Matti Karjalainen.

- *Idiophones* are instruments in which the body of the instrument is the main vibrating unit, and the instrument does not contain such elements as the string in a guitar, the membrane in a drum, or the column of air in a trumpet. Such instruments include, for example, the xylophone, the church bell, and a rattle.
- *Membranophones* have a membrane as their main vibrating unit and this also radiates the sound. Typical examples are different drums, which are often struck directly.
- *Chordophones* have a string as their main vibrating unit. In most cases the vibration is transduced into vibrations of the body of the instrument, which serves as the main source of acoustic radiation. Typical examples are the guitar, violin, and harp. The strings may be plucked, struck, or bowed.
- *Aerophones* are the class of instruments that do not have strings, membranes, or other vibrating bodies. However, the source of excitation may vibrate, like the reed in reed instruments or the player's lips in brass instruments. The vibration occurs typically in an air column. Examples are the clarinet, flute, church organ, and trumpet.

The classification of some instruments is the cause of some debate. For instance, the banjo has strings, but also a membrane, and typically it is classified as a membranophone.

6.1.2 Resonators in Instruments

Most instruments have one or more resonating bodies which modify and radiate the sound (Fletcher and Rossing, 1998). Let us use the body of the guitar as an example of this. The resonator of the guitar is formed by the *top plate* together with the *sound hole*, the *back plate*, and the air cavity between them, as shown in Figure 6.1.

The top and bottom plates and the air cavity have different resonances which all radiate sound together. Figure 6.2 shows an example of the acoustic response of a guitar measured with a microphone at a distance of 1 m in front of the sound hole in an anechoic chamber. The excitation was an impulse-like stroke with a sharp object to the *bridge*. The plot shows the lowest frequencies (<1 kHz) of the response, and the lowest modes are clearly visible. The lowest

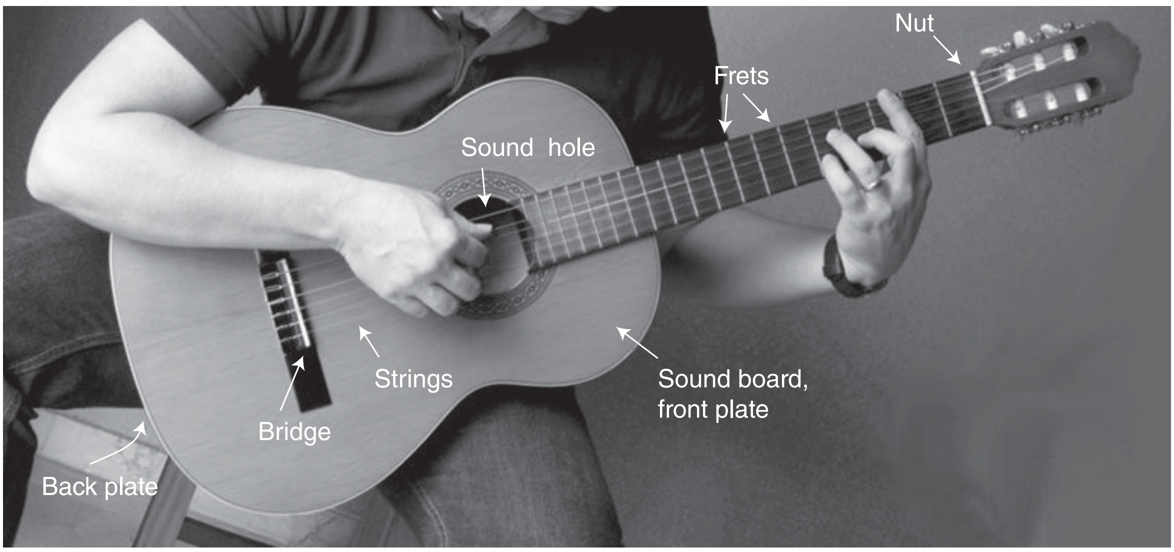

Figure 6.1 The classical acoustic guitar with the names of the most important parts labelled.

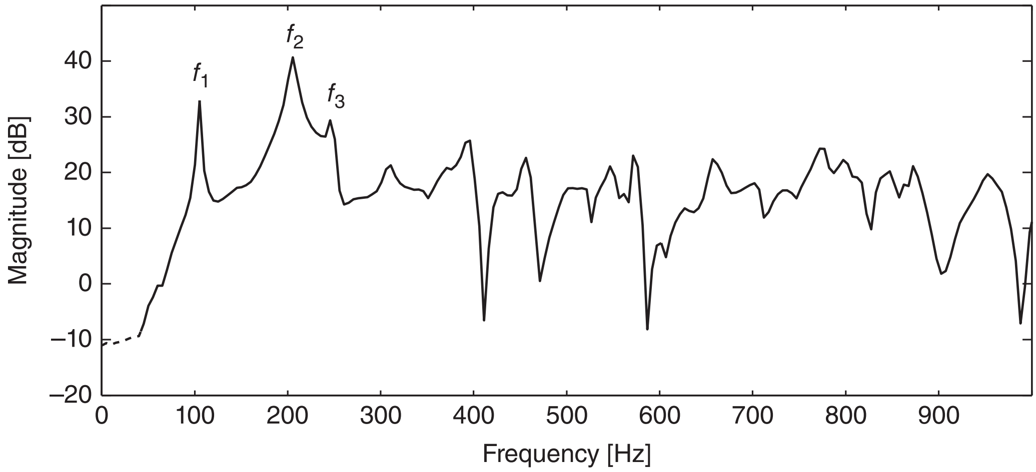

Figure 6.2 The magnitude spectrum of the guitar body at frequencies below 1 kHz. The lowest resonant frequencies are marked as f_1, f_2, and f_3.

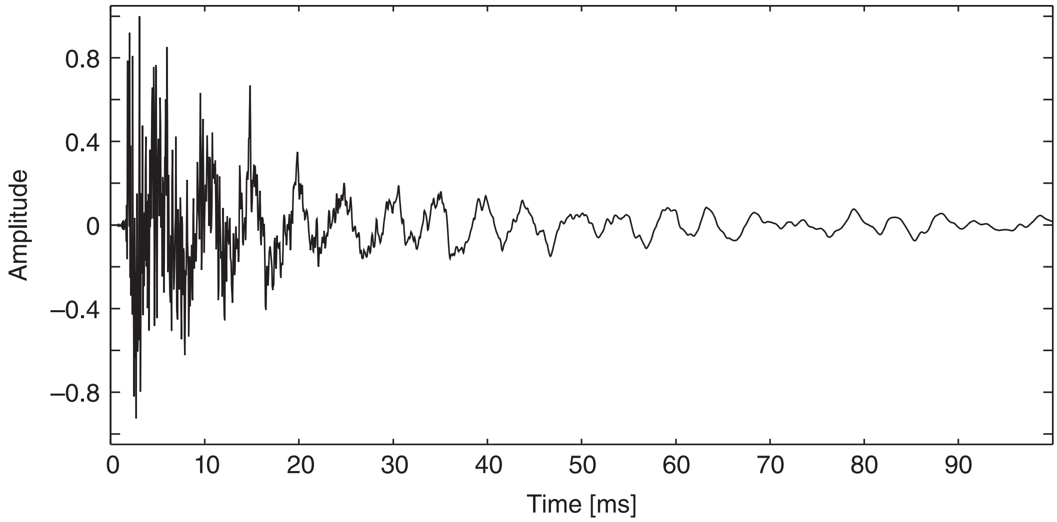

Figure 6.3 The impulse response of the guitar body.

resonance f_1 is the lowest resonance of the air cavity with a frequency of about 100 Hz. The next resonance f_2 is located just above 200 Hz, which corresponds to the in-phase movement of the front and back plates. When we get to the higher frequencies, the resonance structure gets more and more complex, due to the presence of more and more complex resonances in both plates and in the air cavity.

Each resonator colours the spectrum of the sound travelling through it, changing the spectral content of the sound. It is also important to know how the sound changes in the temporal domain. As will be shown in the following chapters, hearing can resolve temporal changes in the positions of the signal components if they change at least every few milliseconds, although with rapidly-changing signals, like a sawtooth wave, very small differences are audible. At the resonance frequencies, the build-up and decay time of the system response is usually larger than between the resonant frequencies. For example, the impulse response corresponding to the magnitude response of the body, shown in Figure 6.2, is so long that the ear perceives it as a sort of reverberation. The impulse response is shown in Figure 6.3. In principle, it is similar

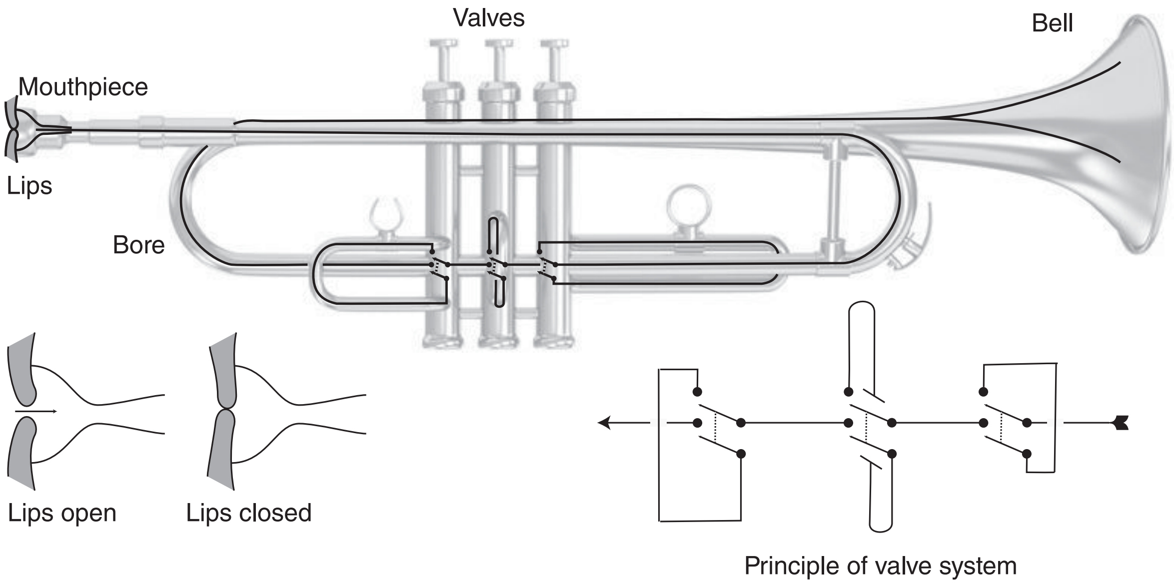

Figure 6.4 The trumpet with the names of its most important parts. The signal path through the valves is illustrated with lines. The movement of the lips of the player is shown schematically.

to the impulse response of a room, shown in Figure 2.20, but the modes are more sparsely distributed in frequency due to the smaller dimensions of the guitar body compared to those of the room.

In aerophones, the vibration occurs in an air column, which thus acts as a resonator in the system. An example of an aerophone is the trumpet, where the column is composed of the mouthpiece, the bore, and the bell, shown in Figure 6.4. The resonances of the trumpet remotely resemble the resonances of a tube closed at one end, as shown in Figure 2.12b on page 29. The acoustic effects of the mouthpiece and the bell, however, change the frequencies and amplitudes of the modes in a well-tuned manner to enable the musical use of the instrument. The interested reader is referred to Fletcher and Rossing (1998) for details on physics of the aerophones, as well as other musical instruments.

6.1.3 Sources of Excitation

In all cases, the vibrations in the instrument are caused by applying an external force to the system, such as by plucking, blowing, or striking (Fletcher and Rossing, 1998). Some examples are discussed here. When modelling musical instruments using signal processing, the excitation is thought to be the input signal to the system, and the rest of the system is assumed to act as a set of linear or non-linear filters.

Many aerophones resemble the human speech organs, where the glottal vibration is the source of excitation and the vocal and nasal tracts are the resonator. A basic difference exists though. In speech, the vibrations in the vocal folds are *not* coupled to the resonances of the tracts, in contrast to many instruments, where the vibrations in the excitation are strongly coupled and synchronized to the resonances in the air column. For example, in *brass instruments*, such as the trumpet, the lips of the player open and close in synchrony with one of the resonances of the bore (see the simplified illustration in Figure 6.4). In some *woodwind instruments*, such as the clarinet and the oboe, a reed or a pair of reeds function in a similar manner. In flutes, the player blows air sideways towards an opening, and the resonances of the air column cause the blowing stream to vibrate between the inside-hole direction and just-off-the-opening directions. These phenomena are non-linear from the viewpoint of signal processing, since the excitation is changed by the state of the system.

In plucked string instruments, the excitation is impulse-like. Ideally, the plucking imposes a pulse of acceleration on the position of plucking from where the wave starts to propagate in both directions of the string. Figure 6.1 shows the playing of the classical acoustic guitar, where the *string*s are plucked with fingers. When plucking with fingers, the temporal excitation of the string is smoother than an ideal impulse, causing a low-pass filtering effect in the sound. After the plucking, the string vibrates autonomously, the amplitude decaying with time. In this case, the modelling is relatively easy, because the vibrations of the string occurring after the excitation cannot affect the excitation itself.

In the piano, the strings are struck with a hammer covered with felt. The hammer gets its speed from pressing a key on the keyboard. The hammer hits the strings usually two to three times, depending on the force applied to the key. A similar effect is seen when the membrane of a drum is struck with a stick. This method of excitation is also non-linear, since the movements of the string or the membrane affect the excitation during the second and third hits.

In the violin, the bow and the string have a strong, non-linear interaction. When moving in one direction, the bow sticks to the string and moves it in the same direction until the tension of the string forces the string to retract and to stick to another position on the bow. The frequency of this repeated stick-and-slip effect depends on the resonances of the string. The movement of the bowing position of the string thus resembles the shape of the sawtooth wave (Equation (3.9) on page 46). From the point of view of modelling, this reminds us of the woodwind and brass instruments, where the excitation is continuous and non-linearly coupled and synchronized to the resonator.

6.1.4 Controlling the Frequency of Vibration

Many instruments are built such that a melody can be presented with them; that is to say, sounds with different frequencies can be produced. This section presents a very short overview of the basic methods to control the frequency, or the pitch, of the sound. The reader interested in more details is referred to Fletcher and Rossing (1998).

Chordophones typically include a number of strings tuned to different pitches. For example, the guitar has six strings, as shown in Figure 6.1. In some instruments, one string is used to produce only one pitch, and no variations are possible, as in the piano. In some other string instruments, the length of the vibrating part of the string can be changed by some means. In the guitar, the string is pressed with a finger against a *fret*, as shown in the figure, and the distance between the fret and the *bridge* defines the pitch of the sound. Yet another method to vary the pitch is the dynamic changing of the tension of the string.

In *aerophones*, the excitation and/or the acoustic properties of the air column define the pitch. In some cases, the effective length of the column is defined simply by the length of the bore and only one pitch is produced, such as in church organ pipes. In woodwinds, the effective length of the column is changed by opening and closing holes in the bore, which changes the pitch. In brass instruments, the vibration of the lips of the player is synchronized to one of the resonances of the instrument, which allows, in principle, the production of the harmonic series of frequencies above the lowest resonance. Additionally, in many of the brass instruments, the length of the column can be changed using valve mechanisms or sliding bores.

The trumpet utilizes valves, as shown in Figure 6.4. When the eight combinations of valve positions are combined with lip synchronization to different modes of the bore, a wide variety of pitches can be played with the instrument. Added to the control of the length of the air column, trained brass or woodwind players can also control the pitch slightly with the tension of their lips.

The pitch produced, if any, by *idiophones* is, in most cases, not controllable. In *membranophones* the pitch is defined by the properties of the membrane and the enclosure. In most cases it cannot be controlled, although exceptions exist.

6.1.5 Combining the Excitation and Resonant Structures

The final sound of the instrument is the sum of the excitation, the effect of the resonators and radiation properties. Also, the acoustic characteristics of the room where the instrument is played have a prominent effect on the sound.

For instruments that can be modelled as linear systems, the system can be expressed in the frequency domain to consist of the excitation $X(\mathrm{j}\omega)$ and the combined transfer function of the partial systems with transfer functions $H_i(\mathrm{j}\omega)$, which together produce the output $Y(\mathrm{j}\omega)$, as

$$Y(\mathrm{j}\omega) = X(\mathrm{j}\omega)\prod_i H_i(\mathrm{j}\omega), \tag{6.1}$$

if the partial systems are in *cascade*, that is the signal flows through them successively. In this manner, the guitar sound is formed by the excitation due to the plucking of the string and by the effects of the partial systems, which are the resonator formed by the string, the resonator formed by the body, and the radiation transfer function.

Unfortunately, many instruments are non-linear and such modelling is not possible. For example, in the trumpet, the air pressure affects the vibration of the lips in a complex manner that cannot be modelled by simple convolution. Also, the modelling of the dynamic control of the pitch brings an additional non-linear component to the system. This means that it is easier to model separated sounds with a constant pitch than an instrument whose physical state changes all the time, or, in technical terms, the frequencies of the resonances of vibrating structures change due to the playing of the instrument. Nevertheless, these effects can be numerically modelled in the time domain, which, however, may result in complicated models (Bilbao, 2009; Smith, 2010; Välimäki *et al.*, 2006).

6.2 Sound Synthesis in Music

The synthesis of musical sounds has been of interest for decades, and many musical instruments are based on such methods (Moore, 1990; Roads, 1996). The first systems utilized simple signal synthesis which did not have a counterpart in the world of acoustic instruments. One of the most influential early electronic synthesizers was manufactured by Robert Moog in the mid-1960s, and it included switchable oscillators, filters, and other modules for processing. Digital sampling systems became available in the late 1970s, and synthesizers with computational models of the physics of the instruments have been available commercially since the mid-1990s. The history of musical sound synthesis is thus different from the history of modelling of speech production, where the first models were based on the physics of speech organs, after which sampling methods were adopted.

6.2.1 Envelope of Sounds

The attributes of natural sounds typically change with time. Only periodic or noise-like sounds may sound the same for a longer time. Our hearing is very sensitive to temporal changes in sound, and a simple method to present the temporal variations in sound is to plot the instantaneous amplitude against time, which is called the amplitude envelope of sound. Another, slightly more complex method is to plot the spectrogram of the signal.

The envelopes of the sounds of musical instruments can often be divided into different phases in time. The *attack* represents the phase when the amplitude of sound increases. After this, the behaviour of the sound depends on the type of the source. In string instruments, the amplitude first decreases rapidly, after which a shallower decrease is reached. In electric instruments, the attack is often followed by *decay*, a decrease in amplitude; *sustain*, a constant amplitude; and *release*, a decrease of the amplitude until silence. This sequence is sometimes called the *ADSR-sequence*.

The amplitude envelope is not sufficient to represent the behaviour of different spectral parts of the sound produced by many musical instruments. The spectrogram (see Section 3.2.6) can be used to visualize instrument sounds. Figure 6.5 shows the magnitude envelopes of the harmonics of a kantele, a Finnish plucked string instrument. The plot shows the typical characteristics of the kantele sound; it has strong decay at higher harmonic partials, and it also has strong amplitude modulation at the harmonics, which appears with different modulation frequencies for different harmonics.

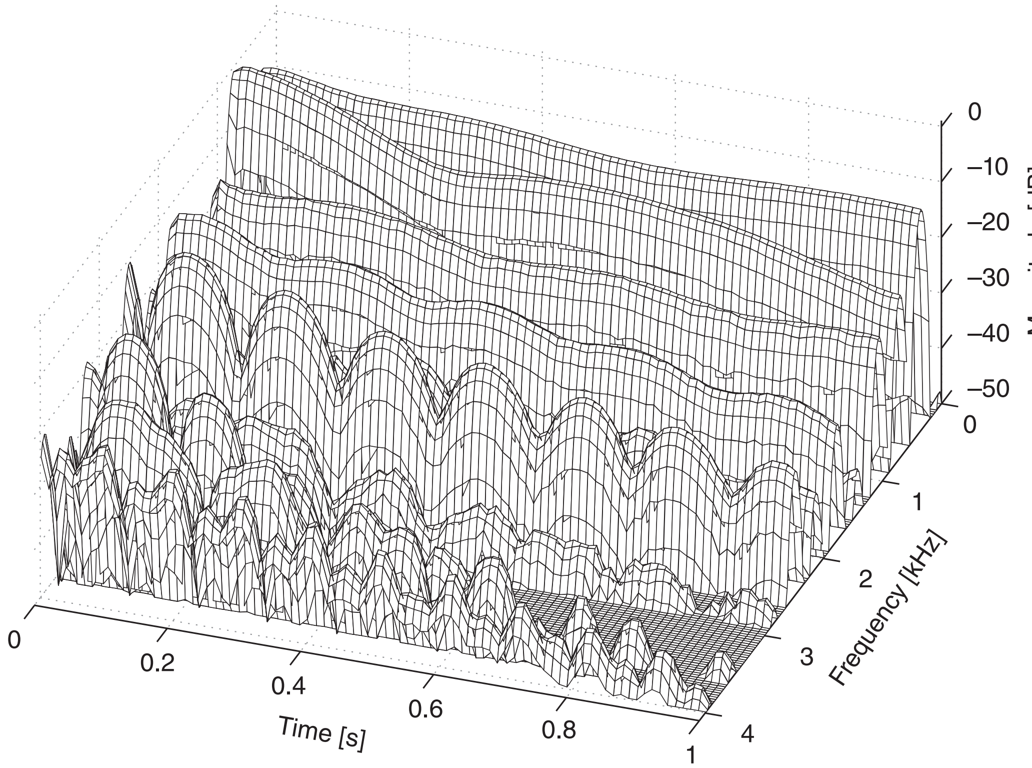

Figure 6.5 The amplitude envelopes of the harmonics of a single plucked string of the kantele.

Musical instruments with continuous excitation typically show slower attack times and steady sustain levels. For example, the sound of the flute starts from noise caused by blowing, and the actual vibration due to the resonance of the air column grows slower.

6.2.2 Synthesis Methods

The main methods used to synthesize musical sounds can be classified into the following categories (Rossing *et al.*, 2001):

- *Sampling*. In sampling, the sound of each note of an instrument is recorded and stored in the memory of a computer (see Section 3.3.1 on page 56) (Roads, 1996). After this, an interface, such as a musical keyboard, is used to control the playback of the notes. The system requires a considerable amount of memory, although the quality of sound can be very high. The memory requirement can be eased by repeating, or looping, certain parts of the sample using pitch shifting to reduce the number of samples in the memory and data compression algorithms (Borin *et al.*, 1997). However, modification of sampled sounds requires more sophisticated methods than the modification of sounds generated with other synthesis methods.
- *Additive synthesis*. In additive synthesis, each harmonic of a tone complex is synthesized separately and then added together (Grey and Moorer, 1977). In principle, the method can generate any sound desired. It is computationally relatively demanding, as musical sounds typically have a large number of harmonics. The system also allows the control of the level and phase of each harmonic arbitrarily, which makes possible the synthesis of any sound. The control data have to be obtained from some other process, for example using a set of heuristic rules, or they can also be analysed from sounds of musical instruments. In Serra and Smith (1990), the method was elaborated to *spectral modelling synthesis*, where a sound is partly represented by a set of sinusoids, and the remaining sound is modelled with white noise processed by a time-varying filter.
- *Subtractive synthesis* is computationally less demanding than additive synthesis. A spectrally rich sound is first synthesized, and filters are used to modify the spectrum to the desired form. The initial sound is typically a sound that can be easily synthesized with electronic circuits or digital computers, such as a sawtooth signal, a square or a triangular wave, impulses or pulses, or noise. The filtering can also be time-dependent. For example, in early synthesizers, the brass sounds were generated by synthesizing a sawtooth wave passed through a low-pass filter whose cutoff frequency and gain depended on the frequency of the sawtooth wave (Risset and Mathews, 1969). This system thus resembles the source–filter model, as discussed in speech synthesis. The early electronic synthesizers implementing subtractive synthesis produced a distinctive sound due to the many non-linear analogue effects, which can also be emulated with digital computers (Välimäki and Huovilainen, 2006).
- *Non-linear synthesis*. In non-linear synthesis, the system has an input–output relationship that produces new frequency components which are meaningful in musical sound synthesis.

  Perhaps the best-known example of non-linear synthesis is *FM synthesis* (Chowning, 1977), where a parameter of the carrier oscillator is modulated with another oscillator(s). Its basic form uses simple frequency modulation, such as $y(t) = \sin[\omega_c t + \sin(\omega_m)t]$, to form a spectrum with peaks at frequencies $\omega_c \pm n\omega_m$. If $\omega_c = \omega_m$, a harmonic spectrum is produced with the fundamental frequency $f_0 = 2\pi\omega_m$, which can be used in music sound synthesis. Different variations of FM synthesis have also been adopted, producing different sound colours.

- *Computational models of the physics of musical instruments*. In such systems, different parts of the physical instrument are modelled in the signal domain with functionally similar DSP blocks (Bilbao, 2009; Jaffe and Smith, 1983; Smith, 2010; Välimäki *et al.*, 2006). This approach is interesting in the context of this book, and an example of modelling plucked string instruments is given in the following section.
- *Computational models of the physics of natural sound sources*. 'Natural sources' refers here to non-musical and non-speech sources, such as sounds of walking in different terrains, the sound of a falling tree, thunder, and so on. Such synthesis methods have applications in computer games and other virtual environments. Different statistical and other methods are used to synthesize such sounds (Cook, 2007; Farnell, 2010).

6.2.3 Synthesis of Plucked String Instruments with a One-Dimensional Physical Model

The model of a string instrument is presented here as an example of a computational model of the physics of a musical instrument. The string in string instruments and the air column in brass and woodwind instruments are one-dimensional resonators, as already reviewed in Chapter 2. Figure 6.6 shows a computational model of such a vibrating string, called the *digital waveguide* (Smith, 2010), where each travelling direction is implemented as a digital delay line. When the transversal vibration arrives at the rigid end point of the line, in the ideal condition, it is fully reflected and the transversal polarity of vibration is changed. In real instruments, some damping of the vibration occurs both in the string and at the point of termination. In Figure 6.6, all of the damping is modelled as occurring at the ends, and the string itself is thought to be lossless. This allows the implementation of a computationally very effective method of synthesis. The excitation is brought to both lines, and the response of the string is captured from another position on the lines. The position may, for example, correspond to the position of the pick-up of an electric guitar or the bridge of an acoustic guitar.

Figure 6.7 shows the plot of a sound generated by the model of a string shown in Figure 6.6. The result is simply a decaying train of impulses. The parameters of the loop filter naturally affect the result; if the filter is set to be a low-pass filter, the decay of high frequencies is much faster than the low frequencies. With certain parameter values the model of the string sounds like a guitar not an acoustic guitar, but an electric guitar. This happens because the model does not include the model of the body of the guitar. The model of the body can be implemented simply by convolving the output of the string model with a measured impulse response of the guitar body shown in Figure 6.3. When the parameters of the model are tuned well, the resulting sounds are very realistic (Laurson *et al.*, 2001).

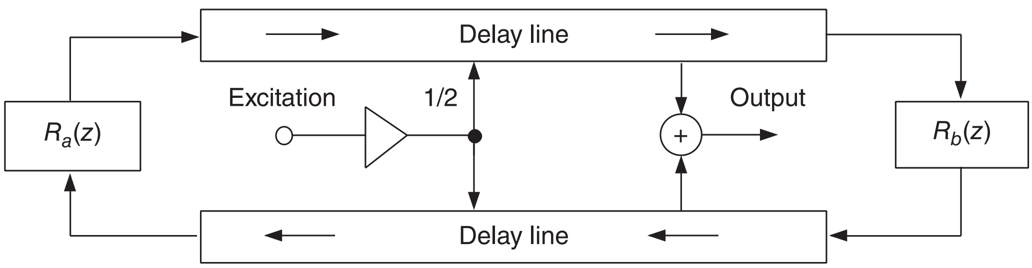

Figure 6.6 The digital waveguide, used here as a computational model of a string in the guitar.

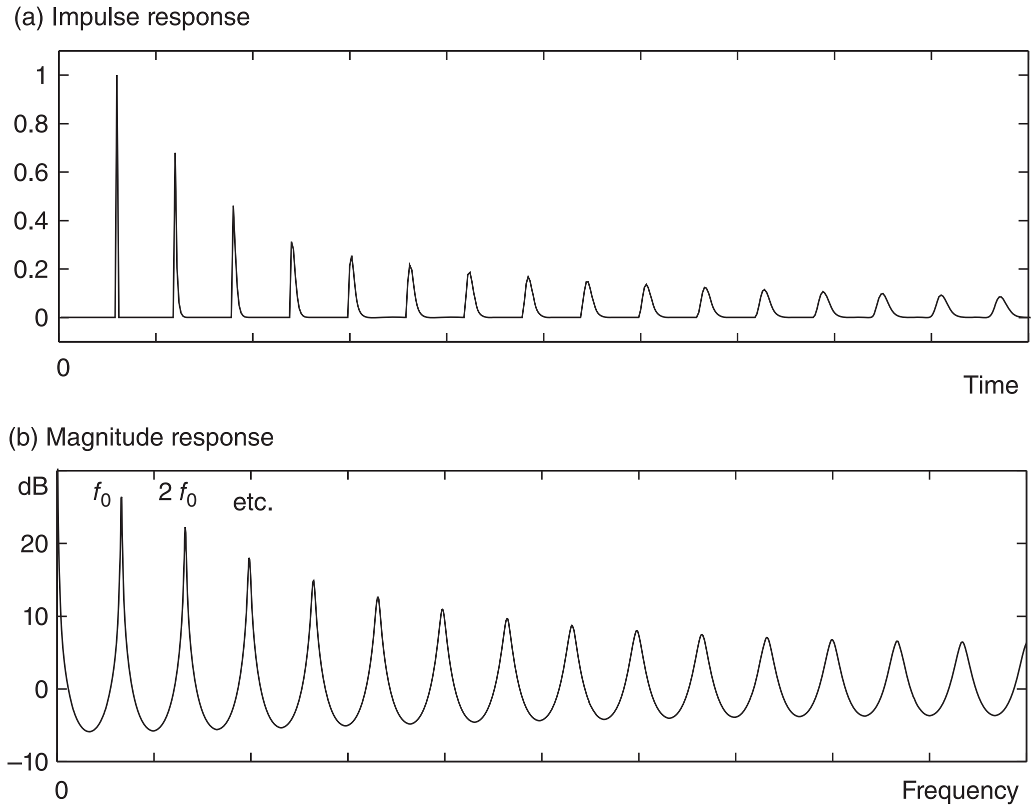

Figure 6.7 The output of the computational model of the guitar string implemented as the model shown in Figure 6.6.

Summary

This chapter has looked at music as a communication signal; how it is produced by humans, acoustic instruments, and technical systems. The modelling aspect has been taken to characterize the general principles that are, in most cases, common to human voice production, acoustic generation of sound, and to electronic and computer means in technical devices.

Further Reading

Musical acoustics has been studied widely, and interested readers should consider books such as (Bilbao, 2009; Fletcher and Rossing, 1998; Moore, 1990; Roads, 1996; Roederer, 1975; Rossing *et al.*, 2001; Smith, 2010).

References

Bilbao, S. (2009) *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics*. John Wiley & Sons.

Borin, G., De Poli, G., and Sarti, A. (1997) Musical signal synthesis. In Roads, C., Pope, S.T., Piccialli, A., and De Poli, G. (eds) *Musical Signal Processing*. Routledge, pp. 5–30.

Chowning, J.M. (1977) The synthesis of complex audio spectra by means of frequency modulation. *Computer Mus. J.*, **1**(2), pp. 46–54.

Cook, P.R. (2007) *Real Sound Synthesis for Interactive Applications*. AK Peters.

Farnell, A. (2010) *Designing Sound*. MIT Press.

Fletcher, N.H. and Rossing, T.D. (1998) *The Physics of Musical Instruments*. Springer.

Grey, J.M. and Moorer, J.A. (1977) Perceptual evaluations of synthesized musical instrument tones. *J. Acoust. Soc. Am.*, **62**, 454–462.

Jaffe, D.A. and Smith, J.O. (1983) Extensions of the Karplus–Strong plucked-string algorithm. *Computer Mus. J.*, **7**(2), 56–69.

Laurson, M., Erkut, C., Välimäki, V., and Kuuskankare, M. (2001) Methods for modeling realistic playing in acoustic guitar synthesis. *Computer Mus. J.*, **25**(3), 38–49.

Moore, F.R. (1990) *Elements of Computer Music*. Prentice-Hall.

Risset, J.C. and Mathews, M.V. (1969) Analysis of musical-instrument tones. *Physics Today*, **22**, 23–30.

Roads, C. (1996) *The Computer Music Tutorial*. MIT Press.

Roederer, J.G. (1975) *The Physics and Psychophysics of Music: An Introduction*. Springer.

Rossing, T.D., Moore, F.R., and Wheeler, P.A. (2001) *The Science of Sound*, 3rd edn. Addison-Wesley.

Serra, X. and Smith, J. (1990) Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Mus. J.*, **14**(4), 12–24.

Smith, J.O. (2010) *Physical Audio Signal Processing: For Virtual Musical Instruments and Digital Audio Effects*. W3K Publishing.

Välimäki, V. and Huovilainen, A. (2006) Oscillator and filter algorithms for virtual analog synthesis. *Computer Mus. J.*, **30**(2), 19–31.

Välimäki, V., Pakarinen, J., Erkut, C., and Karjalainen, M. (2006) Discrete-time modelling of musical instruments. *Reports on Progress in Physics*, **69**(1), 1–78.

von Hornbostel, E.M. and Sachs, C. (1961) Classification of musical instruments: Translated from the original German by Anthony Baines and Klaus Wachsmann. *The Galpin Society Journal*, **14**, 3–29.