Abstract: Basic principles of digital waveguide modeling of musical instruments are presented in a tutorial introduction intended for graduate students in electrical engineering with a solid background in signal processing and acoustics. The vibrating string is taken as the principal illustrative example, but the formulation is unified with that for acoustic tubes. Modeling lossy stiff strings using delay lines and relatively low-order digital filters is described. Various choices of wave variables are discussed, including velocity waves, force waves, and root-power waves. Signal scattering at an impedance discontinuity is derived for an arbitrary number of waveguides intersecting at a junction. Various computational forms are discussed, including the Kelly-Lochbaum, one-multiply, and normalized scattering junctions. A relatively new three-multiply normalized scattering junction is derived using a two-multiply transformer to normalize a one-multiply scattering junction. Conditions for strict passivity of the model are discussed. Use of commutativity of linear, time-invariant elements to greatly reduce computational cost is described. Applications are summarized, and models of the clarinet and bowed-string are described in some detail. The reed-bore and bow-string interactions are modeled as nonlinear scattering junctions attached to the bore/string acoustic waveguide.
10.1 INTRODUCTION

Music synthesizers are beginning to utilize physical models of musical instruments in their sound-generating algorithms. The thrust of this trend is to obtain maximum expressivity and sound quality by providing all of the responsiveness of natural musical instruments. This is happening at a time when most synthesizers are based on “sampling” of the acoustic waveform. While sample-playback instruments sound great on the notes that were recorded, they tend to lack the expressive range of natural instruments.

Another potential application for physical models of sound production is in audio compression. Compression ratios can be enormous when coding parameters of a physical model of the sound source. High quality audio compression techniques such as used in MPEG are presently based on psychoacoustically motivated spectral models which yield up to an order of magnitude of “transparent” compression (see the chapter by Brandenburg or [Bosi et al., 1996a]). Physical models, on the other hand, can achieve much higher compression ratios for specific sounds. By combining model-based and spectral-based compression techniques, large average compression ratios can be achieved at very high quality levels.

The Musical Instrument Digital Interface (MIDI) format provides an example of the profound compression ratios possible for certain sounds by encoding only synthesizer control parameters. For example, two or three bytes of MIDI data can specify an entire musical note. In future audio compression standards, a compressed audio stream will be able to switch among a variety of compression formats. When arbitrary decompression algorithms can be included in the compressed-audio data stream, model-based compression will be fully enabled. In terms of existing standards, for example, one could extend MIDI to provide for MPEG-2 audio segments and “instrument definitions” (synthesis algorithms) written in Java (performance issues aside for the moment). In this context, instrument definitions serve a role analogous to “outline fonts” in a page description language such as PostScript, while sampled audio segments are more like “bit-map fonts.” General, self-defining, instrument-based, audio synthesis scripts have been in use since the 1960s when the Music V program for computer music was developed [Mathews, 1969].

10.1.1 Antecedents in Speech Modeling

The original Kelly-Lochbaum (KL) speech model employed a ladder-filter with delay elements in physically meaningful locations, allowing it to be interpreted as a discrete-time, traveling-wave model of the vocal tract [Kelly and Lochbaum, 1962]. Assuming a reflecting termination at the lips, the KL model can be transformed via elementary manipulations to modern ladder/lattice filters [Smith, 1986b]. The early work of Kelly and Lochbaum appears to have been followed by two related lines of development: articulatory speech synthesis and linear-predictive coding (LPC) of speech.