# 1 Introduction

The theme of this work is computational modeling of acoustic tubes. The models are intended for use in sound synthesizers based on physical modeling. Such synthesizers can be used for producing realistic sounds of woodwind instruments or the human voice.

The technique used throughout this work is called *digital waveguide modeling*. The term 'digital waveguide' was proposed by Smith (CCRMA, Stanford University) who has had a prominent role in the development of the theory of physically based sound synthesis (Smith, 1987, 1992b, 1995a). The digital waveguide synthesis technique is well suited to modeling of one-dimensional resonators, such as a narrow acoustic tube, a vibrating string, or a thin bar.

## 1.1 Historical Notes

Methods similar to digital waveguide modeling have been used in digital sound synthesis for more than 30 years (Kelly and Lochbaum, 1962; Flanagan, 1972, pp. 272–276). They were first applied to *articulatory speech synthesis* which means production of human-like sounds by emulating the geometry and movements of articulators, such as the tongue, the jaws, and the larynx. Nevertheless, another technique referred to as formant synthesis, has become more popular in commercial speech synthesizers. For speech research, the articulatory approach is more attractive, since it is closely related to the physiology and physics of speech production.

At the end of 1970s and in the early 1980s, related techniques were used for modeling string instruments, especially the violin (McIntyre and Woodhouse, 1979; Smith, 1983). The theory of waveguide modeling began to develop after Karplus and Strong (1983) introduced their simple algorithm for synthesis of string instrument sounds. Their method—that computer music people have named the *Karplus–Strong algorithm*—was found to be a simplified version of a more elaborate physical model for a vibrating string (Smith, 1983).

Recently, physical modeling of musical instruments has become one of the most active areas of musical acoustics research. Some scientists work in this field primarily to understand the working principles of musical instruments (e.g., McIntyre and Woodhouse, 1979) while others aim at developing a model-based sound synthesizer (e.g., Jaffe and Smith, 1983; Karjalainen and Laine, 1991; Smith, 1992b, 1995b).

The methods for physical modeling can be divided into five categories (see, e.g., Välimäki and Takala, 1995):

1) source-filter modeling,

2) numerical solving of partial differential equations,

3) vibrating mass-spring networks,

4) modal synthesis, and

5) waveguide synthesis.

These approaches are briefly reviewed in the following.

The source-filter method is included here because it can, in some cases, be interpreted as a physical modeling technique. When the system is divided into the 'excitation' and 'resonator' parts according to its physics, there is no argument about this being a physical model. An example is the separated speech production model where the volume velocity waveform generated by the vibrating vocal folds acts as a source and the vocal tract is the filter. This approach has been used for synthesizing speech (Fant, 1960) and singing (see, e.g., Rodet, 1984 or Rodet *et al.*, 1984). In some musical instruments it is also easy to separate the source and the filter. For example, in many percussion instruments, the excitation is a short impulse that is not affected by the feedback from the resonator. However, most musical instruments are more complicated systems than just a combination of two subsystems. For instance in wind instruments a feedback from the resonator to the excitor is required, and in addition the interaction of the excitation mechanism and the feedback signal is nonlinear.

The first attempts to use a physical model for generating sounds of musical instruments were made by Hiller and Ruiz (1971). They started with the differential equation that governs the vibrations of a string and approximated this equation with finite differences. This technique is computationally very intensive. Even today real-time sound synthesis with this approach using affordable hardware is out of the question. This line of physical modeling has been continued by Chaigne, Askenfelt, and Jansson (1990). (See also Chaigne, 1992; Chaigne and Askenfelt, 1994)

Later in the seventies another approach to physical modeling was taken in Grenoble, France. A system called CORDIS simulates a musical instrument as a collection of point masses that have certain elasticity and frictional characteristics (Cadoz *et al.*, 1984; Florens and Cadoz, 1991). This approach is closely related to the so called finite element method (FEM) that is used in mechanical engineering to simulate the vibration of structures. The object is divided into a large number of pieces in space, each piece is connected to its neighbors with springs and microdampers, thus forming networks that imitate the vibrating object. At the beginning of eighties a special-purpose processor was built which enabled real-time sound synthesis based on a physical model (Cadoz *et al.*, 1984).

At IRCAM in Paris a third technique for physical modeling was developed (Adrien, 1991). It is called *modal synthesis* and is based on a representation of a vibrating structure as a collection of frequencies and damping coefficients of resonance modes, and coordinates that describe the mode shapes. When the instrument is excited at some point, this force excites some or all of the modes. An advantage of modal synthesis is that analysis tools exist and they are not too laborious to use. Adrien (1991) points out that modal analysis of a new object, say a violin body, only takes a couple of days. For some simpler structures, such as a string, the model data can be computed in an analytical form.

In modal synthesis, one of the problems to be solved is where to take the output of the model. At IRCAM a clever solution was found using the body of a real instrument as the 'loudspeaker': in the case of the violin, for example, this means that the simulated vibration of strings is fed to electrical shakers that are attached to a violin body, the strings of which have been carefully damped (Adrien, 1991). This approach has the very nice advantage that the radiational properties of the instrument need not be simulated. In virtual-reality applications, however, this practical trick is not amenable, since all the vibrating structures must create numerical data to be used by other parts of the virtual reality system. Recently, a new implementation of modal synthesis, called Modalys, has been developed at IRCAM (Eckel *et al*., 1995).

In the first half of this decade, *waveguide synthesis* (or, waveguide modeling) has turned out to be the most important of all the physical modeling methods (Smith, 1995b). This applies to both the academic and industrial communities: a majority of recent advances in the theory of physical modeling have concerned digital waveguide techniques, and all existing commercial physical modeling synthesizers utilize digital waveguides. This methodology is also used in this work and is reviewed in detail in Chapter 2.

The first commercial synthesizer to employ a physical model was the Yamaha VL-1 Virtual Acoustic Synthesizer that was introduced early in 1994. It was based on digital waveguide modeling techniques. Recently, other electronic instruments following the same guidelines have been released, and more are like to appear in the future. Physical modeling is considered the most important novel digital sound synthesis technique in the past decade. FM synthesis, the previous revolutionary technique, was introduced in 1973, and the first commercial product to employ it—the Yamaha DX-7—was introduced in 1983. Nowadays, FM synthesizers along with sampling keyboards, are the most widespread electronic musical instruments. Model-based synthesizers are bound to become popular because they sound and behave as acoustic instruments and thus offer more possibilities of expression than the earlier devices (Jaffe, 1995). A fascinating future prospect is to build hybrid synthesizers employing physical modeling combined with sampling and filtering techniques (Smith, 1995b).

Physical modeling is interesting also for research purposes. A physical model enables independent adjustment of individual parameters of an instrument, such as the mass or stiffness of a vibrating reed. This is typically very difficult with an acoustic instrument. Thus a model can help to more clearly understand how an instrument can and should be played and calibrated.

With a highly sophisticated model, it would be possible to design new acoustic musical instruments, and these could be played before they are actually built. Modification of existing instruments could also be tried out using such a system. Today this is still a dream, since modern physical models that produce sound in real time are greatly simplified.

The major strategy in the development of current waveguide synthesis models has been to include only the most crucial mechanisms of the instrument and eliminate the less significant ones. Properties of the human auditory system need to be considered, since it is the human ear that finally judges whether the synthetic sound is satisfactory. The rule of thumb is that those features of the instrument that do not bring any audible improvement to the sound are discarded. Learning to do this also means that one has understood the essential factors that make a musical instrument sound the way it does.

## 1.2 Overview of the Work

In this work, the objective is further development of waveguide models of acoustic tubes. Models of tube systems that are constructed of sections of cylindrical or conical tubes are treated. Consequently, the application that this study mainly aims at is model-based sound synthesis of woodwind instruments and speech. Some of the methods can as well be applied to waveguide models of brass instruments, strings, or bars.

In Chapter 2, the basic theory of digital waveguide modeling is summarized. The principles of modeling vibrating strings and tubes are discussed. Since waveguide modeling of strings is already quite well established, the theory of a complete synthesis model and an analysis technique are briefly described first. Thereafter, the classical lossless tube model constructed of concatenated tube sections of different diameters is examined. A step further is taken by considering modeling of a junction of three tubes and also a junction of two piecewise conical tubes.

At the end of Chapter 2, relations of waveguide synthesis to other signal processing techniques, especially to wave digital filters, are examined and former waveguide models for synthesis of speech and musical instruments are reviewed.

Chapter 3 deals with *fractional delay filters* that are important building blocks of waveguide models. Namely, the main task in waveguide models is to process time delays that it takes for a sound signal to propagate from one point of the system to another, e.g., the propagation time from one end of a tube to another. In digital systems, signals are represented as samples or sequences of numbers. Thus a delay is simply generated by storing the numbers in the computer's memory and retrieving them after the desired time. However, in uniformly sampled digital signal processing systems the delay time is always a multiple of the shortest possible time, called the *unit delay*. This delay is equal to the *sample interval* that is determined by the sampling rate of the system.

It is quickly discovered that time delays needed in a physical model are in general not multiples of any finite unit, but they can take any real value. For this reason, the concept of fractional delay is inevitable and is used for correcting the difference between the desired time delay and the nearest multiple of the sample interval. In mathematical terms, a fractional delay is produced by using *interpolation*.

Chapter 3 reviews several methods for designing digital interpolating filters. A choice that commonly has to be made in digital signal processing problems is whether to use recursive (IIR) or nonrecursive (FIR) digital filters. In Chapter 3 both approaches are reviewed, and the message to a DSP engineer is that in both cases the *maximally flat error criterion* yields an interpolating filter suitable for audio signal processing. In the case of recursive filters, an allpass filter with a maximally flat group delay approximation is a good choice while in the case of nonrecursive filters, a maximally flat frequency response approximation (taking into account both magnitude and phase) is preferred. The argument here is that the approximation error due to maximally flat approximation can be forced to be small at low frequencies which are in many audio applications perceptually more important than details at high frequencies.

In most fractional delay systems the delay must be changed during operation. For this reason it is important to consider time-varying fractional delay filters. Time-varying filters are not as simple to use as time-invariant (fixed) filters. Especially with recursive digital filters, *transients* will appear every time the coefficients of the filter are changed.

At the end of Chapter 3, a method to suppress these transients caused by changing the coefficients is introduced.

Chapter 4 discusses the possibilities that the use of fractional delays offer. In short, fractional delays are needed for two purposes in waveguide models. First, they are used for tuning the total length of the waveguide system. In musical instrument models this is equivalent to tuning. The fundamental frequency of a one-dimensional resonator is determined by the time it takes for a sound signal to travel from one end of it to the other. Obviously, this delay has to be very precisely modeled. Otherwise the waveguide synthesizer sounds out of tune. In vocal tract models, the length of the digital waveguide determines the formant frequencies.

Second, fractional delays are required for simulation of waveguides where some change in the medium takes place at an arbitrary point along the waveguide. For example, in a model of the human vocal tract the different vowels are generated by changing its profile. This is achieved by moving the tongue, jaws or other articulators. Hence the exact location where the change takes place is very significant because it determines how the voice sounds.

Chapter 4 first presents a new signal processing technique, *deinterpolation.* Using both interpolation and deinterpolation operations, a vocal tract model can simulate a change of profile at any given point. A new feature of the proposed model is that—thanks to fractional delays—it is possible to independently vary the length of each tube section.

Furthermore, tube systems with *side branches* are discussed. A practical example of an acoustic tube system with a side branch is the human vocal tract with both oral and nasal tracts included. Another example is a finger hole in the side of a woodwind bore. These systems can be modeled with digital waveguides employing fractional delay filters. Structures for both FIR and IIR fractional delay waveguide filters are presented. Errors due to the fractional delay approximation are examined.

Finally, Chapter 5 summarizes the results of this work and provides suggestions for future research.

## 1.3 Contributions of the Author

The principal innovation in this work is the *interpolated scattering junction* described in Chapter 4. The fractional delay filters are combined with the waveguide filter structure so that a new kind of discrete-time filter structure, that has been called a *fractional delay waveguide filter*, is formed. The author had an essential role in the development of this framework.

The author is responsible for the definition of the inverse operation to interpolation, which was named ‘deinterpolation’ in January 1993. This new concept simplifies the discussion of fractional delay waveguide systems. Especially in FIR fractional delay waveguide filters, deinterpolation helps in the systematic construction of new structures. However, its usefulness in other fields of signal processing still remains to be examined.

More new results developed by the author follow directly from the previous ones: several novel filter structures involving FIR and IIR-type fractional delay filters are described in Chapter 4. They all are physically meaningful since they are discrete-time simulations of a junction of two or three acoustic tubes. The acoustic tubes considered in this work are either cylindrical or conical.

In Chapter 2 of this work, modeling of vibrating strings is also discussed as an introduction to the topic of waveguide synthesis. The author has contributed to the development and implementation of the string models. He also developed and programmed the analysis methods described in Section 2.1.4 and 2.1.5.

The design methods for fractional delay filters described in Chapter 3 were first described in a report co-authored by this author (Laakso *et al.*, 1994). The author has made considerable contributions to this work describing the maximally flat design techniques for both FIR and allpass filters (see Sections 3.3 and 3.4 of this work). He also developed the Farrow structure for Lagrange interpolation (see Section 3.3.7).

Furthermore, the author has invented a new technique for *minimizing the transients* caused by time-varying filter coefficients in recursive filters (Section 3.5.5). This problem is relevant not only to this work but arises in every DSP application where filter coefficients need to be changed during the filtering operation. Here this technique is utilized with allpass fractional delay filters, but the same theory and algorithm are applicable to any recursive filter. The determination of the parameter $N_a$ (the advance time of the transient eliminator), based on the cumulated energy of the recursive filter's impulse response, is a result of cooperations with Prof. Laakso. He initialized this work but the author derived many of the results.

To conclude, in this work the author presents a new framework for simulation of physical systems. His main message is that it is not necessary to suffer from discretization of time delays inside a discrete-time system, such as a digital waveguide model, although its input and output signals are inevitably discretized in time. Consequently, discretization of the spatial variable in simulation of wave propagation is not obligatory. This limitation can be overcome using interpolation techniques described in this work.

## 1.4 Related Publications

Parts from the following publications which the author has contributed have been used in this work:

Timo I. Laakso, Vesa Välimäki, Matti Karjalainen, and Unto K. Laine, "Real-time implementation techniques for a continuously variable digital delay in modeling musical instruments," in *Proc. Int. Computer Music Conf. (ICMC'92)*, San Jose, California, October 14–18, 1992, pp. 140–141.

Vesa Välimäki, Matti Karjalainen, and Timo I. Laakso, "Fractional delay digital filters," in *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS'93)*, Chicago, Illinois, May 3–6, 1993, vol. 1, pp. 355–358.

Matti Karjalainen and Vesa Välimäki, "Model-based analysis/synthesis of the acoustic guitar," in *Proc. Stockholm Music Acoustics Conf. (SMAC'93)*, pp. 443–447, Stockholm, Sweden, July 28–August 1, 1993. Published in October 1994. Sound examples are included on the SMAC'93 CD.

Vesa Välimäki, Matti Karjalainen, and Timo I. Laakso, "Modeling of woodwind bores with finger holes," in *Proc. Int. Computer Music Conf. (ICMC'93)*, Tokyo, Japan, September 10–15, 1993, pp. 32–39.

Vesa Välimäki, Matti Karjalainen, and Timo Kuisma, “Articulatory control of a vocal tract model based on fractional delay waveguide filters,” in *Proc. IEEE Int. Symp. Speech, Image Processing and Neural Networks (ISSIPNN’94)*, Hong Kong, April 13–16, 1994, vol. 2, pp. 571–574.

Vesa Välimäki, Matti Karjalainen, and Timo Kuisma, “Articulatory speech synthesis based on fractional delay waveguide filters,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP’94)*, Adelaide, Australia, April 19–22, 1994, vol. 1, pp. 585–588.

Vesa Välimäki and Matti Karjalainen, “Digital waveguide modeling of wind instrument bores constructed of truncated cones,” in *Proc. Int. Computer Music Conf. (ICMC’94)*, Aarhus, Denmark, September 12–17, 1994, pp. 423–430.

Vesa Välimäki and Matti Karjalainen, “Improving the Kelly–Lochbaum vocal tract model using conical tube sections and fractional delay filtering techniques,” in *Proc. 1994 Int. Conf. Spoken Language Processing (ICSLP’94)*, Yokohama, Japan, September 18–22, 1994, vol. 2, pp. 615–618.

Timo I. Laakso, Vesa Välimäki, Matti Karjalainen, and Unto K. Laine, *Crushing the Delay—Tools for Fractional Delay Filter Design*. Report no. 35, Espoo, Finland, Helsinki University of Technology, Faculty of Electrical Engineering, Laboratory of Acoustics and Audio Signal Processing, 46 pages + figures + 2 tables, October 1994. A revised version entitled “Splitting the unit delay—tools for fractional delay filter design” has been accepted for publication in the *IEEE Signal Processing Magazine*, vol. 13, no. 1, January 1996.

Vesa Välimäki, Jyri Huopaniemi, Matti Karjalainen, and Zoltán Jánosy, “Physical modeling of plucked string instruments with application to real-time sound synthesis,” presented at the *98th AES Int. Convention 1995*, preprint no. 3956, 52 p., Paris, France, February 25–28, 1995. A revised version has been submitted to the *Journal of the Audio Engineering Society*, July 1995.

Vesa Välimäki, “A new filter implementation strategy for Lagrange interpolation,” in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS’95)*, Seattle, Washington, April 29–May 3, 1995, vol. 1, pp. 361–364.

Vesa Välimäki and Matti Karjalainen, “Implementation of fractional delay waveguide models using allpass filters,” in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP’95)*, Detroit, Michigan, May 9–12, 1995, vol. 2, pp. 1524–1527.

Vesa Välimäki, “Fractional delay waveguide filters for modeling acoustic tubes,” in *Proc. Second Int. Conf. Acoustics and Music Research (CIARM’95)*, Ferrara, Italy, May 19–21, 1995, pp. 41–46.

Vesa Välimäki, Timo I. Laakso, and Jonathan Mackenzie, “Elimination of transients in time-varying allpass fractional delay filters with application to digital waveguide modeling,” in *Proc. 1995 Int. Computer Music Conf. (ICMC’95)*, Banff, Canada, September 3–7, 1995, pp. 327–334.

Vesa Välimäki and Tapio Takala, “Virtual musical instruments—natural sound with physical models,” submitted to *Organised Sound*, vol. 1, no. 1, April 1996 (special issue on sounds and sources).

Parts of the present work have been published in the same or slightly modified form in

Vesa Välimäki, *Fractional Delay Waveguide Modeling of Acoustic Tubes*. Report no. 34, Espoo, Finland, Helsinki University of Technology, Faculty of Electrical Engineering, Laboratory of Acoustics and Audio Signal Processing, 133 pages, July 1994.