Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Elettronica ed Informatica

DOTTORATO DI RICERCA IN
INGEGNERIA INFORMATICA
ED ELETTRONICA INDUSTRIALI

CICLO XIV

COMPUTATIONAL ISSUES IN PHYSICALLY-BASED SOUND MODELS

Coordinatore: Ch.ma Prof.ssa Concettina Guerra
Supervisore: Ch.mo Prof. Giovanni De Poli

Dottorando: Federico Avanzini

Revised Version
January, 2001

Università degli Studi di Padova
Dipartimento di Elettronica ed Informatica
Via Gradenigo, 6/A 35131 Padova - Italia
<avanzini@dei.unipd.it>
www.dei.unipd.it/~avanzini

Prefazione

Nuove tecniche per la sintesi digitale del suono hanno rapidamente guadagnato una crescente popolarità nell'ultimo decennio. Tali metodi possono genericamente essere denominati "fisici", dal momento che gli algoritmi di sintesi vengono sviluppati attraverso la modellazione dei meccanismi fisici alla base della produzione sonora. Nello stesso tempo, periferiche per l'audio digitale ad alta qualità sono ormai divenute un equipaggiamento standard su ogni *personal computer*, e la potenza di calcolo oggi disponibile permette l'implementazione in tempo reale di algoritmi di sintesi complessi. Per analogia, l'approccio basato sulla modellizzazione fisica è stato adottato anche dalla comunità della *computer graphics*, per modellare la radiosità e la propagazione della luce. La ricerca nell'audio e nella grafica ha mostrato che tali modelli sono in grado di produrre risultati convincenti e permettono di controllare gli algoritmi di sintesi in modo fisicamente consistente.

Il suono non-verbale veicola molteplici flussi di informazione all'ascoltatore, e può essere usato per integrare e modulare l'informazione visiva. Al contrario del canale visivo, quello uditivo è sempre aperto e riveste un'importanza primaria nella percezione di eventi fisici nei casi in cui le indicazioni visive vengono a mancare o sono confuse. Sulla base di queste considerazioni, la ricerca nell'ambito dei sistemi multimediali sta dedicando una sempre maggiore importanza al suono, al fine di complementare la parte visiva e di fornire all'utente un ambiente realmente multimodale. Inoltre, vi sono casi in cui il *rendering* grafico può non essere possibile o non essere efficiente, laddove il *rendering* audio può essere usato con modesti costi computazionali.

L'approccio per modelli fisici trova importanti applicazioni nell'ambito dell'interazione uomo-macchina, dove opportuni sensori ed effettori permettono lo scambio di informazione tattile e di forza. In questo caso, l'informazione uditiva può aumentare sensibilmente la sensazione di presenza dell'utente e il realismo dell'interazione. In questo senso, la sintesi basata su modelli fisici presenta dei vantaggi rispetto ad altre tecniche (come il campionamento), per due ragioni principali: innanzitutto, la descrizione fisica degli algoritmi permette all'utente di interagire con gli oggetti sonori. Ad esempio, in un modello fisico di contatto il suono di frizione prodotto da una mano su di una superficie cambia a seconda della pressione esercitata dall'utente. Analogamente, il suono prodotto da un oggetto che viene percosso varia in funzione della velocità di impatto. In secondo luogo, i modelli fisici per l'audio e per la grafica possono in linea di principio venire facilmente sincronizzati. Questo permette di raggiungere un alto grado di coerenza tra gli eventi acustici e quelli visivi. Gli sforzi verso la costruzione di modelli unitari per la

sintesi audio-visiva vengono ricompensati in termini di semplificazione nello sviluppo di sistemi multimediali.

Lo studio di strumenti musicali virtuali può essere visto come un caso particolare di interazione uomo-macchina. Nell'usare uno strumento acustico od elettroacustico, lo strumentista interagisce con esso in maniera complessa ed esercita controllo attraverso lo scambio di informazione gestuale, tattile e di forza. Le tecniche adottate nei sintetizzatori commerciali sono perlopiù basate su metodi di sintesi per campionamento (ovvero, suoni registrati ed algoritmi di post-processing), le quali permettono limitate possibilità di manipolazione dei suoni. Di conseguenza, l'unica interfaccia per strumenti elettronici che ha avuto larga diffusione è la tastiera di pianoforte. Di nuovo, tale interfaccia fornisce all'utente un controllo sui suoni sintetizzati che è molto limitato. I modelli fisici di strumenti musicali ampliano di molto le possibilità di interazione con lo strumento virtuale. Ad esempio, la produzione del suono in un modello fisico di violino è controllata attraverso parametri quali velocità e pressione dell'archetto. Analogamente, i parametri di controllo in un modello di clarinetto sono la pressione esercitata dallo strumentista e parametri meccanici determinati dall'imboccatura. Lo sviluppo di modelli fisici di strumenti musicali che siano accurati ed efficienti incoraggia la progettazione di interfacce più sofisticate, che a loro volta forniscano all'utente un più vasto spazio di controllo.

Molte delle considerazioni appena fatte sono valide anche quando riferite alla ricerca nell'ambito della produzione e della sintesi della voce. Le tecniche di sintesi articolatoria sono basate su una descrizione della fisica dell'apparato fonatorio, in termini di pressione polmonare, vibrazioni delle corde vocali, e articolazione del tratto vocale. La ricerca nel campo della sintesi articolatoria si è evoluta per molti aspetti in parallelo con quella sul suono non-verbale, e lo scambio di informazione tra questi due ambiti è stato in generale piuttosto limitato. Ciò nonostante, in più di un caso le tecniche di modellizzazione che vengono adottate sono molto simili. Un vantaggio dell'approccio fisico rispetto ad altre tecniche (come ad esempio la sintesi concatenativa o metodi di analisi/sintesi basati su predizione lineare), è che permette di ottenere segnali acustici più realistici. Inoltre, i modelli possono in linea di principio essere controllati attraverso parametri fisici, quali pressione polmonare, tensione delle corde vocali, parametri articolatori del tratto vocale e della bocca. D'altro canto va sottolineato che quello del controllo è un problema ancora aperto, dal momento che non è banale mappare i parametri fisici direttamente in dimensioni di tipo percettivo (intensità, altezza, registro). Infine, analogamente a quanto detto per il suono non-verbale, i modelli di sintesi articolatoria possono essere sincronizzati con modelli grafici. È risaputo che l'uso di cosiddette "teste parlanti" (*talking heads*), in cui il parlato viene sintetizzato a livello sia audio che visivo, contribuiscono a migliorare sensibilmente la percezione del parlato.

Oltre alla sintesi, i modelli fisici per il suono verbale e non-verbale possono venire usati anche a fini di codifica. In particolare, *MPEG-4* ha implementato uno standard di codifica denominato Audio Strutturato (*Structured Audio – SA*). Se le tecniche tradizionali sono basate sulla codifica e la trasmissione del segnale acustico (ovvero la forma d'onda), lo standard SA codifica e trasmette la descrizione simbolica del suono stesso (ovvero il modello ed i parametri). Il principale vantaggio di tale approccio è che esso permette la

trasmissione a bassissimo *bit-rate*. I modelli fisici sono un esempio di rappresentazione altamente strutturata del suono. Inoltre i parametri hanno tipicamente un'interpretazione fisica e quindi variano abbastanza lentamente per essere usati a fini di codifica efficiente.

La ricerca riportata in questa tesi verte sullo sviluppo di metodi numerici accurati ed efficienti nella progettazione di modelli fisici per il suono. I modelli fisici vengono, in maniera naturale, sviluppati nel dominio del tempo continuo e sono descritti da sistemi di equazioni differenziali ordinarie ed equazioni alle derivate parziali. In una fase successiva, tali equazioni devono essere discretizzate. Al fine di minimizzare l'errore numerico introdotto dalla discretizzazione, di garantire la stabilità degli algoritmi, e di preservare quanto più possibile il comportamento dei sistemi a tempo continuo, vi è la necessità di tecniche numeriche accurate. Tali tecniche devono al tempo stesso produrre algoritmi efficienti, che possano essere implementati in tempo reale. Questi due requisiti di accuratezza ed efficienza comportano tipicamente la ricerca di compromessi. Ad esempio, metodi di discretizzazione impliciti ed iterativi garantiscono una sufficiente accuratezza ma hanno effetti sull'efficienza degli algoritmi risultanti. Analogamente, basse frequenze di campionamento sono preferibili a fini di efficienza, ma l'accuratezza viene deteriorata.

I primi due capitoli forniscono una rassegna della letteratura e presentano le tecniche usate nel resto della tesi. I capitoli successivi contengono risultati originali su diversi modelli fisici: l'ancia singola in strumenti a fiato, le corde vocali nell'apparato fonatorio umano, e la forza di contatto in collisioni tra due oggetti risonanti. Si mostra che tutti questi sistemi possono essere *(1)* modellizzati usando gli stessi approcci, *(2)* interpretati in termini di strutture e blocchi funzionali molto simili, e *(3)* discretizzati usando le stesse tecniche numeriche. Si dimostra inoltre che le tecniche qui adottate forniscono un metodo robusto per la soluzione numerica di modelli fisici non-lineari, producendo al tempo stesso strutture computazionali efficienti.

Il capitolo 1 discute in maggiore dettaglio molti dei punti toccati in questa prefazione. I modelli di sorgente vengono confrontati con altri paradigmi per la sintesi audio, successivamente si analizza l'uso di modelli fisici a fini di sintesi e di codifica.

Il capitolo 2 presenta le tecniche di modellizzazione ed i metodi numerici usati nel seguito della tesi. Si fornisce una breve rassegna della teoria dei modelli *waveguide* in una dimensione, e delle loro applicazioni per la modellizzazione di tubi acustici. Si discute l'approccio ad elementi concentrati, ed il loro uso nella modellizzazione di una larga classe di sistemi meccanici ed acustici. Si affronta infine il problema della discretizzazione. In particolare, viene discusso un metodo numerico che fornisce una soluzione efficiente per il calcolo di anelli computazionali privi di ritardo all'interno di algoritmi non-lineari.

Il capitolo 3 è incentrato su modelli di ancia singola. Si discute un modello ad elementi concentrati e si sviluppa uno schema di discretizzazione accurato ed efficiente. Si mostra che il comportamento dell'ancia digitale così ottenuta riproduce quello del sistema fisico. Vengono poi discusse le limitazioni dei modelli ad elementi concentrati attualmente in uso, e viene sviluppato un nuovo modello non-lineare che permette di prendere in considerazione gli effetti delle interazioni ancia-bocchino ed ancia-labbro, al prezzo di un lieve aumento della complessità del modello.

Il capitolo 4 verte su modelli di produzione della voce. L'attenzione è concentrata sulla modellizzazione delle corde vocali piuttosto che del tratto vocale. Viene discusso il modello glottale ad elementi concentrati di Ishizaka-Flanagan (IF), e si evidenziano le somiglianze strutturali tra questo modello ed i modelli di ancia singola. Vengono proposti due modelli glottali, i quali forniscono entrambi una descrizione semplificata del modello IF. Si mostra che i modelli possono essere implementati in maniera efficiente, preservando al tempo stesso le principali proprietà del modello IF.

Il capitolo 5 discute modelli di contatto per il *rendering* audio di eventi di collisione, e sviluppa un modello martello-risonatore. Si mostra che tale modello è strutturalmente simile a quelli descritti nei capitoli precedenti, e può essere discretizzato usando le stesse tecniche. Il sistema numerico così prodotto ha bassi costi computazionali e può essere implementato in tempo reale. Si esamina poi l'influenza dei parametri fisici sul comportamento del modello. In particolare, viene studiato il problema dell'integrazione nel modello delle proprietà relative al materiale.

Ringraziamenti

I miei sinceri ringraziamenti vanno a tutte le persone che hanno contribuito, con il loro sostegno ed i loro consigli, a questa tesi. In rigoroso ordine topologico da Padova agli Stati Uniti, queste persone sono

Giovanni De Poli, relatore di questa tesi, che mi ha accolto a Padova tre anni fa e da allora ha costantemente sostenuto il mio lavoro con il proprio entusiasmo, e mi ha guidato lungo il mio corso di dottorato con la propria esperienza.

Tutte le persone al CSC – Centro di Sonologia Computazionale all'Università di Padova. In molte occasioni Gianpaolo Borin mi ha grandemente aiutato –più di quanto lui pensi– a risolvere problemi spinosi e a decidere quali direzioni di ricerca valesse la pena esplorare. Ho trascorso molti pranzi e pomeriggi insieme a Carlo Drioli, discutendo di produzione della voce e di sintesi ed elaborazione del parlato.

Davide Rocchesso, all'Università di Verona. È una rara combinazione di preparazione scientifica e resistenza illimitata ai carichi di lavoro. Questa tesi è costellata delle sue idee, dei suoi suggerimenti e consigli. Federico Fontana (pure all'Università di Verona) mi ha aiutato a sviluppare molte idee sugli oggetti sonori, attraverso lunghe discussioni sulle forme, i materiali, la percezione.

Maarten van Walstijn, all'Università di Edimburgo, da cui ho imparato molto sulla simulazione numerica di strumenti a fiato. Nel corso della nostra collaborazione a distanza la mia efficienza nello scrivere *e-mail* è aumentata sensibilmente. Cosa più importante, Maarten mi ha insegnato che le ance singole sono complicate e vanno prese sul serio.

Tutte le persone al *Lab of Acoustics and Audio Signal Processing*, al Politecnico di Helsinki. I mesi passati presso il laboratorio mi hanno fornito una grande opportunità di imparare in molti campi dell'elaborazione del segnale audio, dell'acustica, dell'elettroacustica, dei rituali della sauna. Sono particolarmente grato al Prof. Matti Karjalainen, per il suo prezioso aiuto nell'indirizzare la mia ricerca e per le sue lezioni di psicoacustica; ed al Prof. Paavo Alku, con il quale ho avuto il piacere di lavorare e che mi ha svelato alcuni dei segreti della produzione della voce.

Il Prof. Mark Kahrs, all'Università di Pittsburgh, per le sue lezioni sulle applicazioni del *DSP* in elettroacustica, così come per le lunghe conversazioni su una vasta gamma di argomenti. Ha speso molte energie cercando di migliorare il mio inglese scritto, ed ha letto la maggior parte di questa tesi evidenziando i molti errori e fornendo una quantità di utili commenti.

Preface

In the last decade, new techniques for digital sound generation have rapidly gained popularity. These methods can be generically referred to as *physically-based*, since the synthesis algorithms are designed by modeling the physical mechanisms that underlie sound production. At the same time, high quality digital audio converters have become standard hardware equipment on personal computers, and the available computational power permits real-time implementation of these synthesis algorithms. By analogy, physically-based modeling approach has been also adopted by the computer graphics community, for modeling radiosity or light propagation. Research both in audio and in graphics has shown that such models can provide convincing results and physically consistent control over the synthesis algorithms.

Non-speech sound conveys a large amount of information to the listener, and can be used to augment and modulate visual information. As opposed to the visual channel, the auditory channel is always open and has a primary importance in the perception of physical events whenever visual cues are lacking or confusing. Based on these considerations, research in multimedia systems is devoting more and more importance to sound, in order to complement vision and to provide a multimodal surrounding to the user. Moreover, there are cases where graphic rendering is not possible or cost effective, whereas audio rendering can be used with little computational overhead.

An important application of physically-based sound models is in human-computer interaction, where tactile or force information can be exchanged through suitable sensors and effectors. In this case, auditory information can greatly augment the sense of presence of the user and the realism of the interaction. In this respect, physically-based synthesis is advantageous over other techniques (such as sampling) for two main reasons: first, the physical description of the sound algorithms allows the user to interact with the sound objects. As an example, in a physical model of contact the friction sound of the user's hand on a surface changes with the pressure exerted by the user. Likewise, the sound produced by a struck object varies with the impact velocity. Second, physically-based models for audio and graphics can (in principle) be easily synchronized. This way, a high degree of perceptual coherence of acoustic and visual events can be achieved. The efforts toward the development of models for joint audio-visual synthesis is rewarded in terms of simplification in the design of multimedia systems.

Virtual musical instruments can be regarded as a particular case of human-computer interaction. When using an acoustical or electro-acoustical instrument, the player interacts with it in a complex way and exerts control by exchanging gestual, tactile and force

information. Techniques used in commercial synthesizers are mostly based on wavetable methods (i.e., recorded and post-processed sounds), that allow little sound manipulation. Consequently, the only interface that has been widely used so far in commercial electronic instruments is the piano keyboard. Again, this interface provides the user only with very little control over the synthesized sounds. In this respect, physical models of musical instruments greatly improve the possibilities of interacting with the virtual instrument. As an example, sound production in a physical model of the violin is controlled by parameters such as bow velocity and pressure. Analogously, the control parameters in a clarinet model are the player's blowing pressure and mechanical parameters related to the player's embouchure. The design of accurate and efficient physical models of musical instruments encourages the development of more sophisticated interfaces, that in turn give the user access to a large control space.

Many of the above considerations also hold when discussing research in voice production and speech synthesis. Articulatory speech synthesizers produce speech signals through a physical description of the phonatory system in terms of lung pressure, vocal fold vibrations and vocal tract shape and articulation. The research in articulatory speech synthesis has to a large extent progressed in parallel with research in non-speech sound, with little exchange of information between these two research fields. However, the modeling techniques in the two fields are very similar in many cases. One advantage of the physical modeling approach with respect to other techniques (such as concatenative synthesis or LPC-based analysis/synthesis method) is that more realistic signals can be obtained. Moreover, the models can (in principle) be controlled using physical parameters, such as lung pressure, tension of the vocal folds, articulatory parameters of the vocal tract and the mouth. However, it must be stressed that the problem of control is still an open one, since finding direct mappings between the physical parameters and perceptual dimensions (loudness, pitch, register) is not a trivial task. Finally, analogously to non-speech sounds, articulatory models of speech synthesis can be synchronized with graphic articulatory models. It is known that the use of so called *talking heads*, where visual and audio speech signals are synthesized simultaneously, improve the perception of speech considerably.

In addition to synthesis, physically-based models of both speech and non-speech sound can also be used for coding purposes. In particular, MPEG-4 has implemented a standard named *Structured Audio* (SA) coding. While traditional lossless or perceptual coding techniques are based on the codification and transmission of the sound signal (i.e., the waveform), the SA standard codes and transmits the symbolic description of the sound (i.e., the model and its parameters). The main advantage of this approach is that ultra-low bit-rate transmission can be achieved. Physically-based models are a highly structured representation of sound. Moreover, the control parameters have a direct physical interpretation, and therefore vary slowly enough to be used for efficient coding.

The focus of this thesis is the development of accurate and efficient numerical methods for the design of physically-based sound models. Physical models are naturally developed in the continuous-time domain and are described through sets of ordinary and partial

differential equations. In a subsequent stage, these equations have to be discretized. In order to minimize the numerical error introduced in the discretization step, to guarantee stability of the numerical algorithms, and to preserve as closely as possible the behavior of the continuous systems, accurate techniques are needed. At the same time, the numerical techniques have to produce efficient algorithms, that can be implemented in real-time. These two demands of accuracy and efficiency often require making trade-offs. As an example, implicit and iterative discretization methods guarantee sufficient accuracy but effect the efficiency of the resulting algorithms. Likewise, low sampling rates are preferable for efficient implementations, but the accuracy deteriorates.

The first two chapters review the existing literature and present the general techniques used in the rest of the thesis. The remaining chapters contain original results on various physical models: single reed systems in wind instruments, vocal folds in the human phonatory system, and contact forces in impacts between two resonating objects. It is shown that all of these systems can be *(1)* modeled using the same approaches, *(2)* interpreted using very similar structures and functional blocks, and *(3)* discretized using the same numerical techniques. It is also shown that the techniques used here provide a robust method for the numerical solution of non-linear physical models, while resulting in efficient computational structures.

Chapter 1 discusses in more detail the topics already addressed in this preface. The source modeling approach is compared to other sound synthesis paradigms, then the use of physical models for synthesis and coding purposes is analyzed.

Chapter 2 presents all the modeling paradigms and the numerical techniques that are used in the remaining of the thesis. One-dimensional waveguide structures and their applications to the modeling of acoustic bores are reviewed in detail. Lumped elements are discussed, as well as their use in modeling a large class of mechanical and acoustic systems. Finally, the issue of discretization is addressed. In particular, a numerical method is reviewed that provides an efficient solution of delay-free computational loops in non-linear algorithms.

Chapter 3 discusses single reed modeling. A lumped model is reviewed, and an efficient and accurate discretization scheme is developed. It is shown that the behavior of the resulting digital reed closely resembles that of the physical system. The limitations of existing lumped models are discussed, and an improved non-linear model is formulated. In this new formulation, the interaction of the reed with the mouthpiece and the player's embouchure are taken into account, although at the expense of a slight increase of the model complexity.

Chapter 4 presents results on models of voice production. Attention is focused on vocal fold modeling rather than on the vocal tract. The Ishizaka-Flanagan (IF) lumped model of the glottis is reviewed, and the structural similarities between this model and single reed models are pointed out. Two glottal models are proposed, both providing a simplified description of the IF model. It is shown that the models can be implemented in an efficient way, while preserving the main properties of the IF model.

Chapter 5 discusses contact models for sound rendering of impacts and develops a hammer-resonator model. It is shown that this model is structurally similar to those

described in the previous chapters, and can be discretized with the same numerical techniques. The resulting numerical system has low computational costs and can be implemented in real-time. The influence of the physical parameters on the model behavior is also examined. More specifically, particular attention is devoted to the problem of embedding material properties into the model.

Acknowledgments

My sincere thanks go to all the people who contributed to this thesis with their support and advice. In strict topological order from Padova to the United States, these people are

Giovanni De Poli, my advisor, who welcomed me in Padova three years ago and since then has constantly supported my work with his enthusiasm, and has guided me through my doctoral studies with his experience.

All the people working at the CSC – Center of Computational Sonology in Padova. In many occasions Gianpaolo Borin has helped me greatly –more than he thinks– in solving tricky problems and in deciding what directions of research might be worth exploring. I have spent many lunches and afternoons with Carlo Drioli discussing of voice production and speech synthesis and processing.

Davide Rocchesso, at the University of Verona. He is a striking combination of scientific skill and unlimited resistance to work loads. This thesis is studded with his ideas, suggestions and remarks. Federico Fontana (also at the University of Verona) has helped me develop many ideas on sounding objects, through long discussions on shapes, materials, perception.

Maarten van Walstijn, at the University of Edimburgh, from whom I have learned so much about numerical simulations of wind instruments. During our distance-collaboration my efficiency in writing e-mails has improved considerably. More important, Maarten has taught me that single reeds are complicated and have to be taken seriously.

All the people at the Lab of Acoustics and Audio Signal Processing at the Helsinki University of Technology. The months spent at the lab gave me a great opportunity to learn in many fields of audio signal processing, acoustics, electroacoustics, sauna rituals. I am especially grateful to Prof. Matti Karjalainen, for his precious help in directing my research and for his wonderful lectures on psychoacoustics; and to Prof. Paavo Alku, with whom I had the pleasure to work, and who revealed me some of the secrets of voice production.

Prof. Mark Kahrs, at the University of Pittsburgh, for his inspiring lectures on DSP and electroacoustics, as well as the long conversations on a vast range of topics. He has generously spent time and energies trying to improve my english writing, and has proof-read most of this thesis pointing out the many mistakes and providing a number of helpful comments.

Contents

List of Figures

List of Tables

List of symbols

Symbol	*Quantity*	*Unit*
$t \in \mathbb{R}$	Continuous time	[s]
$n \in \mathbb{N}$	Discrete time	
$s \in \mathbb{C}$	Complex variable in the Laplace domain	[rad/s]
$z \in \mathbb{C}$	Complex variable in the Z domain	
ω	Continuous frequency	$\mathrm{Im}(s)$ [rad/s]
ω_d	Discrete frequency	$\arg(z)$ [rad]
F_s	Sampling rate	[Hz]
T_s	Sampling period	$1/F_s$ [s]
$s(t)$	Continuous-time signal	
$s(n)$	Discrete-time signal	
$S(s)$	Laplace-transformed signal	
$S(z)$	Z-transformed signal	
c	Sound speed in air	347 [m/s]
ρ_{air}	Air density	1.14 [Kg/m^3]

Chapter 1

Physically-based models

1.1 Introduction: sounds, sources

Sound source modeling, structured audio, object-based sound synthesis, synthetic audio, sounding objects, physically-based models. All of these terms are becoming more and more popular among the researchers working on audio signal processing and sound synthesis, and share a common underlying idea: sound representations that are merely based on a description of the sound waveform (i.e. the one-dimensional signal $p(t)$ corresponding to the time-varying acoustic pressure) do not contain information about the way the sound has been generated and processed by the surrounding environment before arriving to the listener's ear. In order to allow users to interact with the sound objects and to manipulate the sound attributes, there is the need for high-level, structured representations. One approach for developing a structured representation of sound is *physical modeling*, in which sound synthesis algorithms are designed based on a description of the physical phenomena involved in sound generation.

From the point of view of signal theory, a satisfactory numerical representation of sound is obtained by sampling the pressure signal at a sufficiently high rate. Since the human auditory system is largely insensitive to frequencies above 20 [kHz], any sampling rate higher than 40 [kHz] provides adequate numerical sound signals. Depending on the application, even lower bandwidths can be used: speech signals are traditionally coded using low sampling rates, since the focus is often on intelligibility rather than quality. Sampling in time the sound signal does not assume any underlying structure, or process, or generative model, in sound representation. The symbolic description is extremely poor, and as a consequence very little interaction with the sound representations is allowed. Although sophisticated signal processing techniques are available that provide meaningful modifications (e.g. pitch shift, time stretching), sampling is basically a *static*, low-level description of sound.

High level representations of acoustic signals are necessarily associated with some abstract paradigms that underlie sound production. In trying to develop a taxonomy of sound synthesis methods (see Table 1.1), a first distinction can be traced between *signal models* and *source models* [34, 124]. Any algorithm which is based on a description of the

Signal Models			***Source Models***
Temporal	*Spectral*	*Abstract*	
Wavetable	Additive	FM	Waveguides
Granular	Subtractive	Non-linear methods	Lumped models
Concatenative speech	SMS	Scanned	Finite differences
	LPC speech		Articulatory speech
			Cordis-Anima

Table 1.1: *A tentative taxonomy of sound/speech synthesis techniques.*

signal $p(t)$ (or its counterpart $P(s)$ in the frequency-domain) and makes no assumptions on the generation mechanisms belongs to the class of signal models. Additive synthesis is a good example: it exploits the Fourier theorem to describe a complex sound as a sum of sinusoids, each controlled in amplitude, frequency and possibly phase. Using the Short-Time Fourier Transform (*STFT*) as the main operational tool, additive synthesis can successfully model quasi-periodic sounds. As such, it is not suitable for synthesizing noise-like signals (such as a flute tone containing a great deal of "breath"). However Serra [122, 123] proposed an extended analysis/synthesis technique, named SMS for Spectral Modeling Synthesis, that includes a time-varying noise component. A major drawback of additive synthesis is its enormous number of control parameters: at least one amplitude and one pitch envelopes have to be specified for each partial. Moreover, the sound representation has not a strong *semantic* interpretation, since these parameters do not have a high-level meaning.

Analogous remarks hold for subtractive synthesis. This method is characterized by a spectrally rich source that is subsequently filtered, and is in a sense complementary to additive synthesis since it is well suited to model noisy and percussive sounds. Early analog synthesizers exploited this idea: the output of a periodic oscillator (saw-tooth, triangle, square, pulse waves), or even noise, was passed through a VCF (Voltage Controlled Filter), whose characteristics could be modified using simple envelope generators. Such a source-filter structure provides a more semantic description of sound than additive synthesis: in certain cases the two blocks can be given a physical interpretation in terms of an exciting action and a resonating object, respectively. However, in many cases this interpretation does not hold, and the control parameters in the model (e.g., the coefficients of the filter) do not have a high-level meaning. LPC based speech synthesis

is another example of subtractive synthesis; in this case, the broadband input signal can be interpreted as a glottal source signal, and the shaping filter represents the action of the vocal tract (see Sec. 1.3).

Similarly to additive and subtractive synthesis, other signal based techniques suffer from lack of semantic description. As an example, FM synthesis is a completely abstract representation, its main control parameters being a carrier and a modulating frequency. Musicians know that controlling a FM algorithm is far from being an intuitive task, and that small changes in the parameters can correspond to large timbral distances in the sounds.

Source models aim at describing the physical objects and interactions that have generated an acoustic event rather than the acoustic signal itself. This modeling approach often gives rise to rather complex descriptions, that can lead to computationally expensive numerical algorithms. Several modeling paradigms and techniques are available in the literature for deriving efficient implementations of such descriptions, including lumped/distributed modeling, waveguide structures, finite difference methods, and so on. The following sections and the next chapter describe in detail a few of these approaches. It is worth discussing another aspect, i.e. that of control. Since the synthesis algorithms are directly based on description of physical reality, the resulting control parameters have a straightforward physical interpretation: typical parameters in the models are associated with masses, hardness/softness characteristics, blowing pressures, lengths. In other words, the models provide a semantic sound representation and can in principle allow physically-based interaction.

This has important consequences for applications in human-computer interaction and, in general, multimedia communication. Increasingly, multimodal multimedia systems are integrating new sensors and haptic modalities with the well developed visual representations. In this context, realistic interaction with audio objects can significantly improve the coherence in joint audio-visual synthesis and the feeling of presence in a virtual environment. Section 1.4 discusses these issues in more detail.

Analogous remarks hold also for speech sounds: as an example, emerging research in Auditory Visual Speech Processing shows that accurate design of "talking heads", where visual and acoustic signals are synthesized simultaneously, can improve speech perception considerably.

1.2 Musical instruments

Research on the acoustics of musical instruments has undergone extremely fast development since the development of the digital computer [67, 68, 120, 90, 49], and many of the basic mechanisms that initiate self-sustained oscillations in musical instruments are now understood. Accurate physical models have been designed and simulated numerically, aimed at studying the functioning of various blocks of an instrument, and their mutual interactions.

As early as 1971, Hiller and Ruiz [67] used numerical simulations of the wave equation

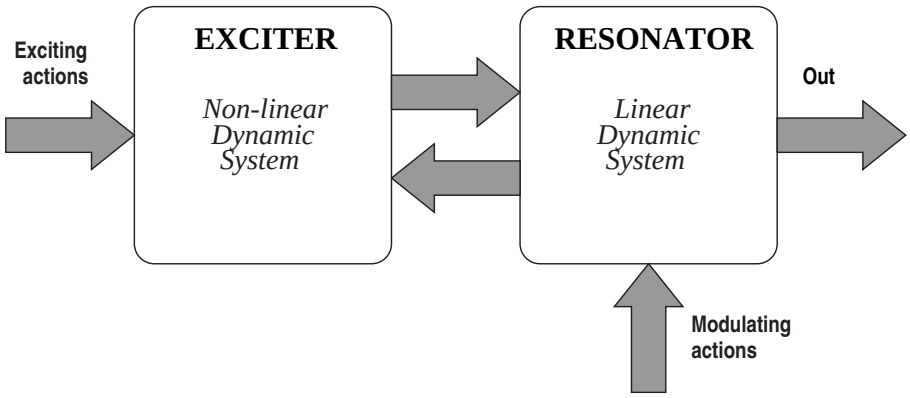

Figure 1.1: *Exciter-resonator interaction scheme for a musical instrument, to be compared to Fig. 1.3. Figure based on De Poli and Rocchesso [35].*

for sound synthesis applications: ≪ [...] This is a completely new approach to electronic sound synthesis insofar as the starting point is the physical description of the vibrating object [...] ≫. A decade later, Karplus and Strong [76] proposed a numerical algorithm for the synthesis of plucked string sounds: their string model can be regarded as the first step toward the development of digital waveguide structures. In the same year, McIntyre, Schumacher and Woodhouse [90] suggested the use of non-linear models of sound generation in wind and bowed-string instruments for real-time sound synthesis: ≪ [...] a fast minicomputer could produce results at a cycle rate in the audible range. The result would perhaps have some novelty: an electronic musical instrument based on a mathematical model of an acoustic instrument [...] ≫. Twenty years later, audio signal processing systems have improved substantially, and the –once– expensive algorithms described by these authors can now be easily implemented in real-time in less and less hardware [74, 157].

Musical oscillators are often strongly non-linear. A typical example is found in woodwind and brass instruments, where self-sustained oscillations in an acoustical bore can only be explained in terms of a non-linear, persistent excitation mechanism. More precisely, the valve (a single or double-reed, or the player's lips) at the bore termination acts as a non-linear element that injects energy into the system. A very similar description holds for bowed string instruments, where the bow is the exciting element. In other cases the instrument is non-linearly excited only for a limited amount of time: a struck string or bar interacts with the hammer or mallet through a non-linear contact force. Values for the contact time are typically a few milliseconds, and after this short excitation the system evolution is linear. There are also examples where non-linearities are negligible: plucked string instruments can be conveniently treated as linear systems (strings and instrument body), where the "pluck" is simply described as a non-equilibrium initial condition (i.e., it gives a string a non-zero displacement distribution and a null velocity distribution).

In all of these cases, the musical instrument can be schematized by means of two

main functional blocks [25, 35], as depicted in Fig. 1.1. The *resonator* is the part of the instrument where the oscillations actually take place. Depending on the instrument, this can be the acoustical bore, the string, the bar. It is therefore related to such sound attributes as pitch and spectral envelope, and in general to *sound quality*. The *exciter* controls the way energy is injected into the system, thus initiating and possibly sustaining the oscillations. It relates to properties of the transient attack, which is known to have a primary role in defining timbre and *sound identity*.

The interaction between the two blocks can be feedforward or feedback, depending on the instrument. Persistently excited instruments –such as winds– are described by a feedback structure, while for plucked string instruments a feedforward scheme can be assumed without significant loss in accuracy of the description. A very simple yet striking demonstration of the effectiveness of the exciter/resonator schematization is provided by mounting a clarinet mouthpiece on a flute [155]. The bore boundary conditions are changed from open-open to closed-open so that it can play one octave lower, and the resulting instrument is perceived as a bad sounding clarinet. In other words, the excitation mechanism defines sound identity ("it's a clarinet"), and the resonator merely controls sound quality ("it's a *bad* clarinet").

Outlining such functional blocks helps the modeling process; each of them can, to a certain extent, be modeled separately and with different representation strategies. Moreover, the block decomposition can be refined, i.e. both the exciter and the resonator can be described by simpler and more elementary constitutive elements. As an example, the resonating block of a wind instrument is made of a bore, a number of tone holes and a radiating bell, and each of these can be described by their own models. Both "white-box" and "black-box" approaches can be taken. The term white-box indicates that the block is modeled by further decompositions in finer physical elements. The black-box approach amounts to describe a given block according to its input-output behavior, without further assumptions on its internal structure. As an example, the radiating bell in a wind instrument is often modeled using a black-box approach: since the bell acts as a filtering element which reflects low frequencies and radiates high frequencies pressure waves, the modeling problem reduces to filter design.

As far as modeling paradigms are concerned, these are often grouped into two broad categories, namely *lumped* and *distributed* models. Lumped models are used when a physical system can be conveniently described in terms of ideal masses or rigid bodies, connected to each other with spring and dampers, and possibly non-linear elements. The resulting systems are naturally described in the time domain, in terms of Ordinary Differential Equations (*ODE*s).

Pressure-controlled valves, such as single, double or lip reeds, are typically described using the lumped modeling paradigm. Indeed, these systems are quite complicated: a clarinet reed is a non uniform bar clamped at one termination and free at the other one, and has many vibrational modes [130, 128]. Similarly, a lip reed is made of non-homogeneous tissue and exhibits horizontal and vertical modes of vibration. Nonetheless, these systems have been successfully modeled using lumped elements [120, 1], and it is widely understood that such a simplified mechanical description captures the basic

behavior of pressure controlled valves [48]. Similar remarks hold for hammers and mallets: during collision, they are deformed and subject to internal losses and non-linear restoring forces. However, interactions with strings and bars have been modeled and efficiently implemented in sound synthesis algorithms by assuming the hammer/mallet to be a lumped mass and deriving empirically the corresponding expression for the contact force [133, 23]. Chapter 2 discusses in detail structural and computational problems concerned with non-linear lumped models. Chapter 3 presents original results for lumped models of single reeds.

Distributed models, as opposed to lumped ones, are more often used for describing vibrating bodies or air volumes, where forces and matter depend on both time and space. These bodies are not easily decomposed into systems of discrete masses. One-, two- and three-dimensional resonators (such as strings, bars, acoustical bores, membranes, plates, rooms, etc.) are usually treated as distributed systems and mathematically described by means of Partial Differential Equations (*PDE*s). Many of these systems can be conveniently modeled in the frequency domain in terms of their input impedance. Schumacher [120] has used measured input impedance functions of a clarinet bore for relating the acoustic flow and the pressure at the bore entrance. Among the sound synthesis community, however, the most successful and popular approach for dealing with many distributed systems is in the time domain: *waveguide modeling*, developed mainly by Smith [125, 126, 127, 110].

Chapter 2 discusses waveguide models in detail. In its simplest form the method exploits the existence of an analytical solution to the D'Alembert wave equation, which can be seen as a superposition of traveling waves (rigidly translating waveforms). Such a solution can be simulated in the discrete space-temporal domain using delay lines, and the resulting numerical algorithms are extremely efficient and accurate. Moreover, physical phenomena such as frequency dependent losses and dispersion can be included in the models by incorporating low-pass and all-pass filters in the delay line scheme. Again, careful design of such filters allows for very accurate and relatively low-cost simulations. Some sound synthesis algorithms based on the waveguide approach have been successfully implemented on commercial integrated circuit.

Although waveguides are extremely successful in modeling nearly elastic mediums (where the D'Alembert equation or some of its generalizations hold), they are not equally good in dealing with systems where rigidity has a major role and bending forces are the main source of oscillation. As an example, oscillations in a bar are governed by the so called Euler-Bernoulli equation. No analytical general solution is given for this fourth order PDE, and no traveling-waves schematization can be assumed. In order to deal with such systems, *finite difference* or *finite elements* methods are the most suitable techniques. These time-domain techniques are sometimes referred to as "brute force" methods, since they are based on direct discretization of the PDEs and have high computational costs. On the other hand, when properly used they provide stable and very accurate numerical systems. Finite difference scheme have been used in the literature for modeling idiophones [28, 29] and single reed systems [130, 128].

Other approaches are available, though less popular, for dealing with distributed sys-

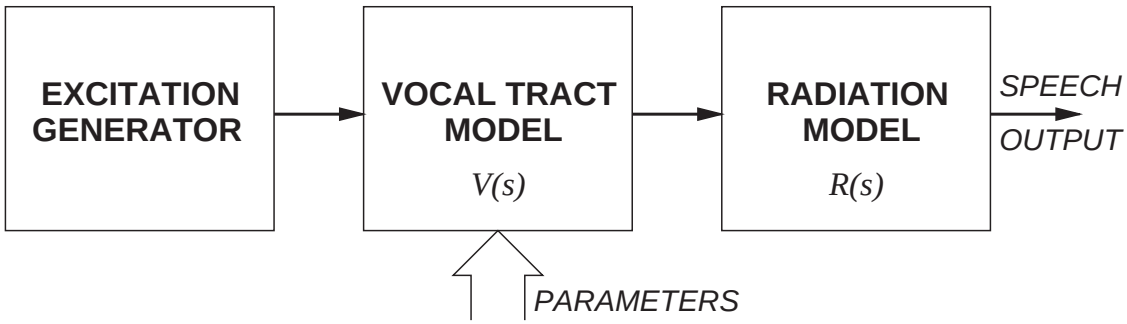

Figure 1.2: *Feedforward source-filter block scheme associated to LPC based speech synthesis.*

tems: cellular models decompose a resonating body into a multitude of interacting particles whose dynamics is discretized and quantized, thus giving rise to a cellular automaton. In the early nineties, Cadoz and his coworkers have introduced CORDIS-ANIMA systems [50], that describe vibrating bodies as a set of interconnected mass-spring-damper cells. Extremely high computational costs are a major drawback of this approach. Furthermore, no analytical tools are available for assessing stability properties of the discretized systems.

1.3 Voice production and speech synthesis

Dominant techniques in speech synthesis are commonly based on signal models. Concatenative synthesis is largely used for commercial applications. This technique is based on the concatenation of short recorded sounds, namely *dyphones*, through which words and entire expressions are constructed. Suitable processing stages are provided for changing the sound quality and to account for prosody. Many techniques of speech coding and synthesis use Linear Prediction Coefficients (LPC) [37]. The assumption underlying this method is that the phonatory system can be schematized as a feedforward source-filter model, as depicted in Fig. 1.2. According to such a schematization, the source block represents the airflow at the glottis, while the linear filter accounts for the combined effects of the vocal (and possibly nasal) tract and lip radiation.

The effect of the vocal tract is to "shape" the glottal flow waveform by providing formants (resonances). An all-pole model is therefore a good representation for such a filter. Lip radiation acts as a high-pass operation: at low frequencies it can be argued that radiated pressure is approximately the derivative of the flow, and this approximation is indeed used in most cases. The system can be further simplified by switching the two filters $V(s)$ and $R(s)$, thus obtaining the speech pressure signal by filtering the glottal flow derivative with the vocal tract filter. Under these assumptions, techniques based on linear prediction can be used for estimating the vocal tract all-pole filter $V(s)$, and the glottal flow derivative is obtained as the residue from inverse filtering of the radiated

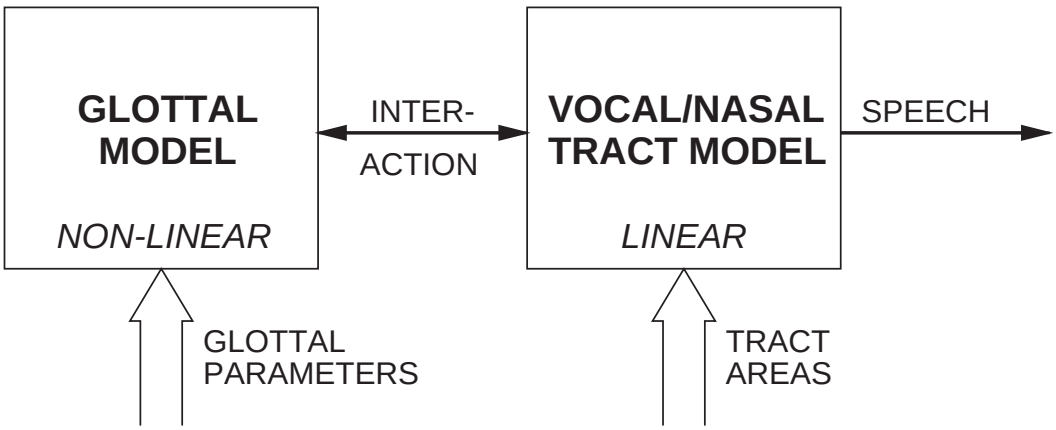

Figure 1.3: *General block scheme of an articulatory speech synthesizer, to be compared to Fig. 1.1. Figure based on Sondhi and Schroeter [129].*

pressure.

LPC synthesis is, to a certain extent, physically-based, since the main blocks depicted in Fig. 1.2 can be given a physical and physiological interpretation. However its major assumption, i.e. a feedforward interaction between glottal source and vocal tract, holds only as a first order approximation and is imprecise. In a real phonatory system, the vocal tract behaves as an acoustical air column, thus providing feedback to the vocal source through its input impedance. Detailed modeling has to take into account this acoustical interaction in order to allow for more natural sounding output.

Speech synthesizers that are based on acoustical models are commonly referred to as *articulatory synthesizers*. Figure 1.3 depicts a typical block scheme of an articulatory synthesizer, as proposed by Sondhi and Schroeter [129]. When compared to the scheme in Fig. 1.2, it appears to have a very similar structure. However in this case the two main blocks interact in a feedback configuration. On the other hand, this scheme exhibits a striking resemblance to the exciter-resonator decomposition outlined in Sec. 1.2 for musical instruments (see Fig. 1.1). Indeed, the modeling approaches adopted in articulatory speech synthesis are very similar to those used for musical instruments, as noticed by Sondhi and Schroeter [129].

The vocal tract is a non-uniform, time-varying resonator. As a first approximation, it is described by its cross-sectional area function $A(x,t)$ (x being the position along the tract). A widely accepted approach [99, 129, 119, 37] approximates this profile with small uniform (cylindrical) sections, that can be modeled separately and connected through junctions (chapter 2 provides a brief introduction to the Kelly-Lochbaum junction). Most of the modeling techniques are extremely similar to those outlined in Sec. 1.2 for musical resonators. Frequency domain approaches [129, 119] describe each cylindrical section using "chain matrices" that are ultimately related to input impedance at the section terminations. Time-domain methods make use of transmission lines for describing pressure wave propagation in the cylindrical section [99, 37], and are formally identical to

waveguide models.

Since the articulatory parameters are directly related to vocal tract geometries, in human voice production they vary slowly enough to be good candidates for efficient coding. Moreover, these parameters can be easily interpolated, because the interpolated values between two physical area functions are physically realistic. This is not true in general for LPC vocal tract filters: an interpolation between two realizable coefficient vectors may lead to unstable filters or to impossible vocal tract cross-sectional areas. Unlike musical instruments, however, vocal tract resonators are time varying: a major problem in articulatory models is concerned with the estimation of vocal tract areas $A(x,t)$. One approach to such an estimation is through X-ray motion pictures [43], but these measurements can only be obtained on a limited scale. On the other hand, direct estimation of $A(x,t)$ from the speech signal (radiated pressure) is not straightforward and has been studied extensively (see for instance [119]): the main problem of this approach is that the mapping from the acoustic parameters (e.g. a set of formants) to the articulatory ones is not one-to-one.

Several approaches are possible for providing excitation signals to an articulatory vocal tract. Parametric models, such as *Liljencrants and Fant* [44], are one option. These fit a given glottal flow waveform using piecewise analytical functions, and are therefore signal models. Alternatively, the vocal tract can be excited by a time varying section that represents the glottis, and driving this additional section using synthesized or measured glottal area signals. However a fully physical description has to account for the interaction between the glottal source and the resonating vocal tract. Time-domain, lumped element models have been developed that describe the vocal folds by means of one or more masses and viscoelastic elements [73, 131]. These exhibit clear resemblance to lip reed models for brass instruments [1, 69]. The first and most popular lumped model for the vocal folds has been developed by Ishizaka and Flanagan [73], and describes each vocal fold using two masses connected to the fold body and to each other with non-linear springs and dampers. In this way it is possible to account for phase differences between the upper and lower edges of the fold. The masses are driven by glottal pressure, which in turn depends on glottal area, and the complete system is described by a set of non-linear differential equations.

Chapter 4 provides more detailed description of the Ishizaka-Flanagan model and lumped glottal models, together with original results. Here it is worth mentioning that such lumped models can be easily coupled to an articulatory vocal tract and give rise to "natural" interaction effects, that cannot be obtained using simpler feedforward source-filter schemes. These effects include skewing of the glottal flow waveform when narrow constrictions of the vocal tract occur, presence of oscillatory ripples on the positive slope portion of the glottal waveform, dependence of pitch and open quotient on the load characteristics. However, a major drawback of these lumped physical models is that they are not easily controllable. The model parameters, such as mass and length of the vocal fold, spring stiffness, rest position, do not map directly into perceptual dimensions such as loudness, pitch, register, and so on. A possible strategy for attacking this problem consists in developing an intermediate mapping that relates laryngeal muscle activation

to lumped parameters [136]. Once such a physiologically realistic control space has been constructed, a set of rules that transform perceptual variables into muscle activation can be developed.

Articulatory speech synthesis can find an important application in the emerging field of Auditory Visual Speech Processing –AVSP– [89]. It is widely known that joint synthesis of visual (i.e. lip, jaw and tongue movements) and audio speech signals can improve speech perception significantly. Development of "talking heads" has recently become a very active research field, and indeed MPEG-4 already provides a standard for 3-D facial modeling [39]. One ultimate goal is to create a synthetic speaking face that is as realistic as a natural face in the movements of speech. So far, techniques based on concatenative synthesis have been mostly used in auditory visual speech models. However, articulatory speech synthesizers can be in principle naturally integrated in these models, since some of the articulatory parameters (e.g. jaw opening, tongue tip displacement) can be used to drive both the acoustic and the visual model.

1.4 Sound sources in virtual environments

Terminology related to applications of virtual environments and multimedia is sometimes confusing or even contradictory. The terms *Virtual Reality* and *Virtual Environment* are often used with the same meaning; "environment" may be preferable with respect to "reality", since reality is too complex to simulate and it is often better to simulate a generic environment.

A rather general scheme summarizing the main components of a virtual environment or a multimedia system is depicted in Fig. 1.4. Virtual environments can be characterized in terms of three main elements: a *scenario*, i.e. one or more host computers that provide algorithms for organizing and creating visual, auditory and possibly haptic output for one or more users. A set of *effectors*, i.e. hardware devices (helmet-mounted displays, headphones, force-feedback devices) that transduce output data from the scenario engine to the user. And a set of *sensors*, again hardware such as position/velocity trackers or data gloves, that allow for feedback from the user into the scenario. On the other hand, a commonly accepted characterization of a multimedia system is simply that of a real-time interactive software interface which includes multiple effectors. A multimedia system should provide advanced graphics, animation and/or video-based visual interface. And sound.

The development of visual interfaces within virtual environments is far beyond the level that audio and sound have reached. As an example, high quality two-channel digital sound has become standard equipment on personal computers only in recent years, while high resolution color graphic hardware and software have been available long time ago. In order to provide the feeling of presence in a virtual environment, however, a truly multimodal surrounding is required, which includes not only vision but also auditory and possibly haptic modalities. Sounds, in particular, *(1)* are characterized by a high semantic richness, *(2)* can be associated in a direct way to well defined physical phenomena, and

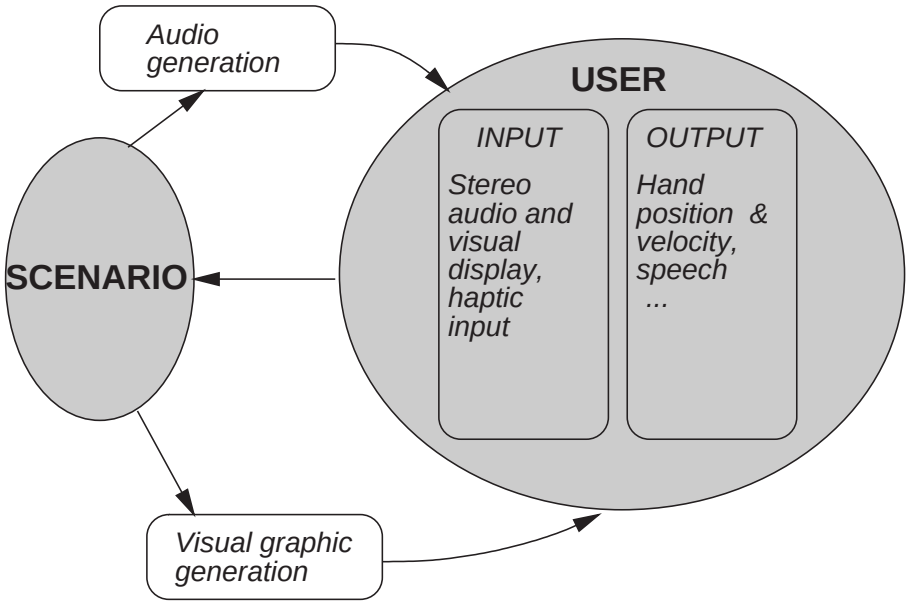

Figure 1.4: *Schematic representation of the main components in a virtual environment or multimedia system.*

(3) convey many information flows and can be directly related to visual information in order to modulate (or sometimes even override) visual cues.

It is worth mentioning here that MPEG-4 provides a set of tools that enable construction and concise transmission of audiovisual scenes composited from several component pieces of content such as video clips, computer graphics, recorded sound, and parametric sound synthesis. These tools are collected under the name BIFS (BInary Format for Scene description), and a specific subset controlling the compositing of sound scenes is called AudioBIFS. It provides a unified framework for sound scenes that use streaming audio, interactive and terminal-adaptive presentation, three-dimensional (3-D) spatialization, and/or dynamic download of custom signal-processing effects (see Sec. 1.5 for a concise introduction to BIFS and AudioBIFS).

Research in sonification of virtual environments makes use of simple (or increasingly complex) communication theory models in order to describe the listening experience of a user. Three main elements are involved in such models, namely a *source*, a *medium* and a *receiver* (or listener). Each of these can contain a number of transformations (both physical and perceptual) through which information is conveyed to the listener. The source represents one or more vibrating objects, that radiate sound into the environment. The medium involves the path by which sound propagates from the source to the listener: this includes the environmental context (e.g. reverberation from the enclosure and diffraction from physical objects in the scenario). The receiver element involves the listener's auditory system, from ear to high level neurological processing.

Figure 1.5 shows a basic example of a source-medium-receiver model, as proposed by

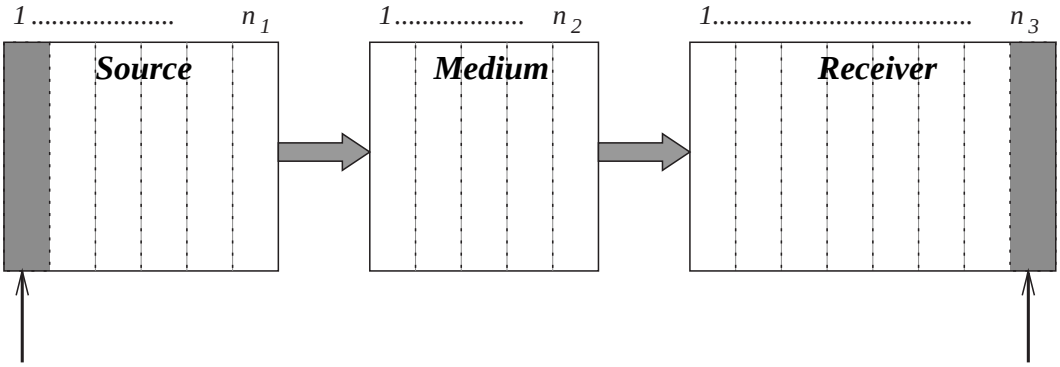

Figure 1.5: *A source-medium-receiver model, based on Begault [17]. Vertical numbered sections represent n_i processing stages associated to each block .*

Begault [17]. This is a perfectly suitable model in the context of sound spatialization, reverberation and 3-D sound processing [93, 57, 17], where the information to be conveyed is mainly concerned with spatial features such as perceived distance and position of the sound source and environmental effects. Processing of such information involves mainly the medium and the receiver blocks, and therefore a feedforward scheme can be reasonably assumed.

Far less attention has been devoted to the *source* block. In particular, the scheme of Fig. 1.5 totally neglects direct flow of information from the source to the listener. Indeed, sound signals convey information about a number of source features, such as shape, size, material, and even about the physical interaction that has originated the sounds themselves. Moreover, a human receiver is not only a listener but also a *player*, i.e. can interact with sound sources: thus a feedforward interaction scheme is not fully satisfactory.

Early research addressing the problem of source perception from the sound signals is to be found in the field of ecological psychology [100, 152, 153, 59]. The main idea in these works is that listeners typically tend to describe sounds in terms of the sound-producing events (Gaver [59] refers to this attitude as "everyday listening"). As a consequence, an *ecological acoustics* has to be developed, which identifies those sound features that convey information about physical properties of objects and events. Such properties are commonly grouped into two broad categories [59], namely *structural invariants* (specifying individual properties of objects such as size, shape, material) and *transformational invariants* (characterizing interactions, i.e. the way objects interact or are played). Several conjectures have been formulated about what acoustic cues are significant for conveying information on these invariants [152, 153].

Psychoacoustic experiments have provided experimental basis to the above mentioned conjectures [55, 84]. More recently some authors have tried to exploit these results for sound modeling purposes. An example application is that of simulated contact with objects in a virtual environment, where sound can provide important information. Klatzky

et al. [78] have shown that the acoustical cues proposed by Wildes and Richards [153] can be conveniently incorporated in a contact model in order to provide perception of material to the receiver. Analogous research has addressed the problem of shape perception [141]. However, the sound synthesis algorithms used in this works are based on additive synthesis techniques, and are therefore signal models. As already mentioned, one major drawback of these models is that they are do not provide physically meaningful control parameter. Conversely, the physical approach can in principle allow for algorithms whose control parameters are naturally associated with physical variables and are responsive to various kinds of interactions. Moreover, they can be easily synchronized with analogous graphic models [98], thus providing a high degree of coherence and consistency for the perception of objects and events [134].

The modeling paradigms used for developing physical models of musical instruments can also be adopted in this case. As an example, resonating objects are conveniently described using waveguide modeling. Recent work by Rocchesso and Dutilleux [102, 106, 104] has shown that 3-D waveguide structures can be successfully used for modeling resonating enclosures, and can convey shape information to the listener. Interesting links are also found with research in robotics. As an example, contact models for dynamic simulation of robotic systems have been studied extensively, and are typically derived using a lumped physical description (see [86]). Chapter 5 deals with physically-based contact models, and shows how simple models and computationally efficient structures can be used to obtain realistic contact sounds and natural interaction between the user and the virtual objects.

1.5 Structured audio coding

Sound models for virtual environments and multimedia systems, discussed in the previous section, face a series of constraints such as limited storage capacity, low cost and reduced channel bandwidth in network applications. Such constraints have created a demand for high-quality digital audio delivery at low bit rates, and considerable research has been devoted to the development of algorithms for efficient "hi-fi" audio coding. This section discusses the use of physically-based models for sound coding purposes; in particular, the MPEG-4 *Structured Audio* standard is reviewed.

The central goal of audio coding is to represent the signal with a minimum number of bits while achieving transparent signal reproduction, i.e., generating output audio that cannot be distinguished from the original input. Traditional *lossless* coders remove entropic or information-theoretic redundancy from a sampled signal. This redundancy arises from the fact that successive samples are not statistically independent and that some sample values occur more often than others. These techniques allow perfect reconstruction of the coded sounds.

In the last decade, a new class of algorithms have been proposed, based on *perceptual* coding. Exhaustive introductions to this topic are provided by Painter and Spanias [97] and by Brandenburg [26]. Suffice it to say that perceptual coders remove redundancies

that are created by over-specifying the sound format with respect to the human auditory system and by including features that cannot be perceived by a listener. In the coding process, input signals are segmented into quasi-stationary frames. Then, time-frequency analysis (usually matched to the analysis properties of the human auditory system) estimates the temporal and spectral components on each frame. In parallel, psychoacoustic analysis provides *masking* thresholds, and therefore estimates –for each point in the time-frequency plane– the amount of quantization distortion that can be introduced without causing any audible artifacts. Using this procedure a minimum number of bits can be allocated for the coding of each frame, and as early as 1992 the MPEG-1 (Moving Picture Expert Group) Audio provided standards for perceptual audio coding at various sampling rates and bit rates: the now so popular "mp3" technology (MPEG-1, Layer-3) is already ten years old.

Later generations of MPEG Audio have continued the work on perceptual coding. In particular, the so called Advanced Audio Coding (AAC) became a standard in 1997. AAC was able to provide the same quality of MPEG-1 Layer-2, with half the bit rate. More recently, MPEG-4 has implemented new *structured coding* methods [148, 115, 116]. The main idea underlying these methods is that most sound signals exhibit *structural redundancy*, i.e. they are more efficiently represented as processes –or models– than as waveforms. Coding the symbolic description of a sound, i.e. an underlying model together with its parameters, instead of the sound signal itself allows for ultra-low bit-rate transmission.

A trivial example of a non-structured representation is given by a sampled waveform coded in pulse code modulation (PCM) format: it is a very general and inefficient sound representation, and no contextual model is embodied in the parameter space. In contrast, a musical excerpt stored in MIDI format is a highly structured representation, but it is not general: only sounds that obey the pitch-onset-duration model (such as in piano music) can be encoded while subtly changing ones (such as wind or string instruments) can not be completely specified. Physical models are clearly another good example of structured representation of sound.

Table 1.2 summarizes the properties of various structured audio descriptions, including physical models, as proposed by Vercoe *et al.* [148]. It can be noticed that none of the representations are *general* –except for MPEG-4– i.e. each of them can be successfully applied exclusively to some specific kinds of sounds. On the other hand, all of these techniques are *synthesizable*, since they can be turned to sound without any additional information. Some of these sound formats are *semantic*, that is, their parameters have a clear high level meaning. Physical models have indeed a strong semantic interpretation, since their control parameters are directly related to the real controls. In contrast signal based representation, such as additive and subtractive synthesis, do not provide semantic descriptions, as already discussed in Sec. 1.1.

A very important property is that of *encodability*, i.e. the possibility of extracting the representation directly from an audio waveform. This is not an easy task in the case of physical models. A first obvious observation is that a given physical model is suitable for encoding purposes only with specific classes of sounds, i.e. those produced by the modeled

Method	General	Encodable	Synthesizable	Semantic	Accurate	Efficient
MIDI/Event List	No	No	Yes	Yes	No	Yes
Additive synth.	No	Yes	Yes	No	Somewhat	Somewhat
Subtractive synth.	No	No	Yes	No	Yes	Yes
LPC models	Speech only	Yes	Yes	Somewhat	Yes	Somewhat
Physical models	No	Difficult	Yes	Yes	Yes	Yes
Synth. Languages	No	No	Yes	Somewhat	Yes	Yes
MPEG-4 Scenes	Yes	Difficult	Yes	Yes	Yes	Yes

Table 1.2: *Properties of various structured representations of sound. Table based on Vercoe* et al. *[148].*

instrument. This is a disadvantage with respect to other structured representations, and implies that a underlying physical model must be assumed for each class of sounds to be encoded. A second difficulty comes from the fact that physical models are typically non-invertible, and thus no analytical approach can be adopted.

One of the earliest studies on this topic is due to Casey [27]. He showed that two layer neural networks can be used in order to reconstruct input parameters from the sound signal. In his work, however, the matching was performed on the time-domain waveform while spectral –and perceptual– features were neglected. More recent research by Serafin *et al.* [121] uses frequency-domain features for estimating the control parameters of a bowed string physical model. Their early results, obtained with training sets made of synthetic sounds, are promising.

The MPEG-4 Structured Audio Standard unifies all of the structured representations listed in Table 1.2 (and this is why MPEG-4 scene is classified as general). A sophisticated synthesis language, called *SAOL* for Structured Audio Orchestra Language, allows the description of synthesis and processing algorithms that can be designed using any of the above methods. It is based on older languages, such as Music V and Csound, and includes many new features such as more powerful functional abstraction, better score-based controllability, and so on. Algorithms written in SAOL can be transmitted as part of the bitstream, and a decoding terminal capable of understanding the language can synthesize sounds as required by the content stream. Standard MIDI commands can be used to drive the algorithms, and in addition the MPEG-4 specification provides a

second tool, *SASL* (Structured Audio Score Language). This control language is more complete than MIDI, in that it allows more complex control functions and mappings to be specified.

Structured Audio, as well as other sound and speech representations, is integrated into the MPEG-4 Binary Format for Scene Description, or *BIFS*. This tool enables the concise transmission of audiovisual scenes [117]. Most of the concepts in BIFS are taken from the popular Virtual Reality Modeling Language (*VRML*) standard, but AudioBIFS (the part of BIFS controlling the description of sound scenes) contains significant advances in quality and flexibility compared to VRML audio.

Like VRML, MPEG-4 BIFS encode the audiovisual material using a *scene graph* representation, in which the content is organized as a set of hierarchically related nodes. The BIFS scene graph is transmitted in the MPEG-4 bitstream, at the beginning of the content session, and is therefore part of the content like the media objects themselves. A node in the visual scene graph represents a visual object (like a cube or image), a property of an object (like the textural appearance of a face of a cube), or a transformation of a part of the scene (like a rotation or scaling operation). By connecting multiple nodes together, object based hierarchies can be formed.

The AudioBIFS scene graph relies on a slightly different concept: an *audio subgraph* represents a signal-flow graph describing various signal processing stages. Sounds flow in from MPEG-4 audio decoders at the bottom of the scene graph. A child node passes its processed output to one or more parent nodes. Through this chain of processing, sound streams eventually arrive at the top of the audio subgraph and the final result is presented to the user. In MPEG-4, a finished sound at the top of an audio subgraph is termed a *sound object*.

Summarizing, this section has discussed the use of physically-based sound synthesis algorithms for coding purposes. Research in this field is still in its first stages, and robust and general encoding techniques still need to be found.

Summary

The chapter has compared the physically-based approach to other traditional sound synthesis techniques. Specifically, the ability of the various techniques to provide a semantic representation of sound has been discussed.

Sections 1.2 to 1.4 have reviewed the use of physically-based synthesis algorithms for the development of virtual musical instruments, for voice and speech synthesis applications, and for the sonification of multimedia environments. In each of these application cases the discussion has focused on control. Since the synthesis algorithms describe the mechanisms of sound generation, the control parameters have a straightforward interpretation in terms of physical quantities, which helps in the design of effective control models.

Finally, Sec. 1.5 has reviewed recently developed coding techniques, that are based on the coding and the transmission of sound models instead of sound signals. These

techniques, generally referred to as structured audio coding, permit ultra-low bit-rate transmission. The possible use of physically-based algorithms for coding purposes has been discussed. It has been pointed out that one of the major open problems is how to encode the models, i.e. how to extract the parameters of the physical models from the sound waveforms.

Chapter 2

Structural and computational aspects

2.1 Distributed models: the waveguide approach

This section introduces the basic concepts of waveguide modeling. Discussion is focused on one-dimensional resonators, and gives the theoretical bases for understanding the simple waveguide structures that are used in the following chapters. More complete formulations are to be found in works by Smith [125, 126, 127]. Berners [20] provides a rigorous yet enjoyable discussion of waveguide modeling for acoustical resonators. The accurate design of waveguide structures for simulation of woodwind instruments is addressed by Scavone [114]. Bilbao [22] compares digital waveguide networks to other similar structures, namely multidimensional wave digital filters, and proposes a unified formulation of these techniques.

No attention is devoted here to higher dimensional waveguide structures, since these are not used in the remaining of the thesis. Considerable research has been carried out on this topic, see [143, 52, 113, 53].

2.1.1 One-dimensional wave propagation

The D'Alembert equation describes vibrational phenomena in an ideal *elastic* medium. The one-dimensional version of this equation is written as

$$\frac{\partial^2 y}{\partial x^2}(x,t) = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}(x,t). \tag{2.1}$$

This equation holds, for instance, in an ideal string of length L, linear mass density μ and tension T. In this case the variable $x \in [0, L]$ stands for position along string length and y stands for *transversal* displacement of the string. The constant c has the value $\sqrt{T/\mu}$ and has the dimensions [m/s] of a velocity. A full derivation of Eq. (2.1) for the ideal string can be found in many textbooks [18, 49]: roughly speaking, the two main assumptions are that (i) the infinitesimal string segment dx moves only in the vertical

direction, so that its acceleration can be computed using only the transverse component of the tension as the acting force; and (ii) the amplitude of the vibrations is very small. A fundamental property of Eq. (2.1) is that it describes *propagation* phenomena. Indeed, factoring the equation yields

$$\left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right) y = 0,$$

and from this factorization it is easily seen that generic solutions take the form

$$y(x,t) = y^+(ct - x) + y^-(ct + x). \tag{2.2}$$

The two functions $y^\pm$ describe waveforms that translate rigidly with velocity c, in the right-going and left-going directions, respectively. Their shape is determined by the boundary conditions (in space) and the initial conditions (in time). Another important result is that complex sinusoids $e^{j(\omega t \pm kx)}$, with $k = \omega/c$, are solutions of the D'Alembert equation. Depending on boundary conditions, only certain values are allowed for k. For a fixed string (i.e. $y(0,t) = y(L,t) = 0$) these turn out to be $k_n = n\pi/L$ with $n \in \mathbb{N}$, i.e. the only allowed frequencies form a harmonic series. Due to the Fourier theorem, the general solution is a linear combination of these sinusoids.

A n-dimensional generalization of Eq. (2.1) is found to be

$$\nabla^2 y(\boldsymbol{x}, t) = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}(\boldsymbol{x}, t), \tag{2.3}$$

where the symbol $\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$ stands for the n-dimensional Laplacian operator. With $n = 2$, Eq. (2.3) describes for instance mechanical vibrations in an ideal membrane, while $n = 3$ is well suited for describing acoustic disturbances in an air volume. In this latter case $\boldsymbol{x}$ represents Euclidean coordinates in space and y stands for the acoustic pressure p. As opposed to mechanical vibrations in a string or membrane, acoustic vibrations are *longitudinal* rather than transversal, i.e. the air particles are displaced in the same direction of the wave propagation. Again, simplifying assumptions have been made for deriving Eq. (2.3) in the acoustic case. Namely, disturbances are considered to be small so that the acoustic pressure p is related to density ρ via a linear relation: $p = B(\rho - \rho_{air})/\rho_{air}$, where B is the linearized adiabatic bulk modulus and ρ_{air} is the air equilibrium density. The constant c is then given the value $\sqrt{B/\rho_{air}}$, and again has the dimensions [m/s] of a velocity.

There are interesting cases where acoustic disturbances can be assumed to be one-dimensional up to a reasonable approximation. Propagation in a cylindrical tube of radius r_0 is an example: by exploiting boundary conditions and symmetries, and looking for harmonic solutions (those with time dependence $\exp(j\omega t)$), the acoustic pressure can be written in cylindrical coordinates as $p(r, \phi, z, t) = \exp(j\omega t) \cdot R(r)Z(z)$ and the equation is separable (see Fig. 2.1 for an illustration of cylindrical coordinates). This leads to the coupled spatial solutions [49]

$$R(r) = \mathcal{I}_0(\beta r), \qquad Z(z) = e^{\pm j\left(k^2 - \beta^2\right)^{1/2} z}, \tag{2.4}$$

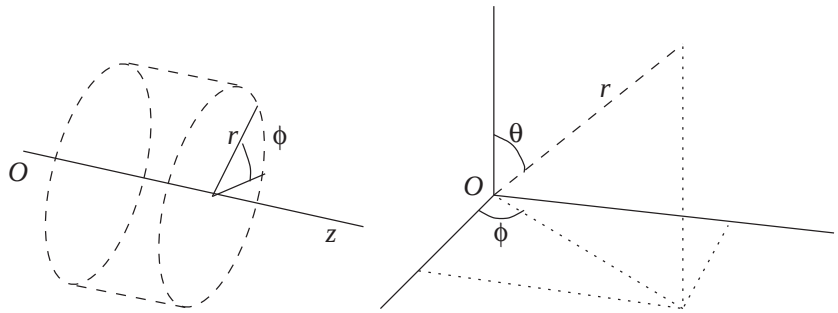

Figure 2.1: *Illustration of cylindrical and spherical coordinates.*

where $\mathcal{I}_0$ is the Bessel function of the first kind and zero order. The boundary condition on the cylindrical surface is $d/dr[\mathcal{I}_0(\beta r_0)] = 0$, or equivalently $\mathcal{I}_1(\beta r_0) = 0$. Therefore, only the β values for which $\mathcal{I}_1(\beta r_0) = 0$ are allowed. The first allowed value is obviously $\beta = 0$: this corresponds to the zero-th order modes with $Z(z) = \exp(\pm jkz)$ and $R(r) \equiv 0$, i.e. plane wave propagation along z. The next allowed value corresponds to the first zero of $\mathcal{I}_1$, i.e. $\beta r_0 = 3.83171$. If $r_0 = 8 \cdot 10^{-3}$ [m] (the approximate radius of a clarinet bore), then $\beta = 479$ [m^{-1}] and the first corresponding mode in the z direction has a cutoff frequency $f_c = \beta c/2\pi = 26.15$ [kHz]. Only frequencies higher than f_c do propagate, and they are well out of the range of human hearing. Therefore, for audio applications higher order non-planar modes can be neglected and one-dimensional wave propagation in the z direction can be conveniently described using Eq. (2.1).

Conical geometries are a second example where one-dimensional propagation can be approximately assumed. Again, by exploiting boundary conditions and symmetries and looking for harmonic solutions, pressure can be written in spherical coordinates as $p(r, \theta, t) = \exp(j\omega t) \cdot \Theta(\theta)R(r)$ and the equation is separable (see Fig. 2.1 for an illustration of spherical coordinates). Without going into details, analysis analogous to that outlined for cylindrical geometries shows that higher-order modes can also be neglected in this case, and propagation in the r direction is conveniently described with zero-th order modes. Since the Laplacian operator is expressed in spherical coordinates as $\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}$, the one-dimensional equation for spherical wave propagation is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)(r,t) = \frac{1}{c^2}\frac{\partial^2 R}{\partial t^2}(r,t). \tag{2.5}$$

Using the substitution $R = \tilde{R}/r$, it is easily seen that Eq. (2.5) reduces to the one dimensional D'Alembert equation (2.1). Therefore $\tilde{R}$ is the sum of two traveling waves $\tilde{R}^{\pm}$, and the general solution for the zero-th order radial modes is

$$R(r,t) = \frac{1}{r}[\tilde{R}^{+}(ct - r) + \tilde{R}^{-}(ct + r)]. \tag{2.6}$$

So far, only displacement y and acoustic pressure p have been considered in the wave equation. However, alternative wave variables can be used in strings and acoustical bores.

As an example, the force acting on a string section dx is defined as

$$f(x,t) = -T\frac{\partial y}{\partial x}(x,t) = -T\left[\frac{\partial y^+}{\partial x}(ct-x) + \frac{\partial y^-}{\partial x}(ct+x)\right] = \frac{T}{c}\dot{y}^+(ct-x) - \frac{T}{c}\dot{y}^-(ct+x).$$

Therefore, using this equation force waves $f^\pm$ can be defined as $f^\pm := \mp\frac{T}{c}\dot{y}^\pm$. On the other hand, the transversal velocity wave variable in the same string is given by

$$v(x,t) = \frac{\partial y}{\partial t}(x,t) = \dot{y}^+(ct-x) + \dot{y}^-(ct+x).$$

From this, velocity waves $v^\pm$ are defined as $v^\pm := \dot{y}^\pm$. The pair of force and velocity variables is sometimes referred to as *Kirchhoff variables*, in analogy with voltage and current in electrical systems (Sec. 2.2 provides a detailed discussion of Kirchhoff variables and analogies between electrical, mechanical and acoustic systems). From the previous equations it immediately follows that

$$f^\pm(ct \mp x) = \pm Z_0 v^\pm(ct \mp x), \qquad \text{with} \qquad Z_0 = T/c = \sqrt{T\mu}. \tag{2.7}$$

The quantity Z_0 takes the name of *wave* (or *characteristic*) *impedance* of the string, and its reciprocal $\Gamma_0 = Z_0^{-1}$ is termed *wave admittance*. Note that using Z_0 both the force f and the velocity v can be related to the force waves $f^\pm$. Namely, the following relations hold:

$$\begin{aligned} f &= f^+ + f^-, & v &= \frac{1}{Z_0}\left[f^+ - f^-\right], \\ f^+ &= \frac{f + Z_0 v}{2}, & f^- &= \frac{f - Z_0 v}{2}, \end{aligned} \tag{2.8}$$

that transform the pair (f, v) into the pair (f^+, f^-), and vice versa. Wave impedance can be defined also in a cylindrical bore. In this case the Kirchhoff variables are taken to be pressure p and flow u (volume velocity). These can be related through the wave impedance Z_0: $p^\pm(ct \pm x) = \pm Z_0 u^\pm(ct \pm x)$, where $Z_0 = \rho_{air}c/S$ and S is the constant cross-sectional area of the bore (see e.g. [49] for a rigorous justification of this formula). For conical geometries, the cross-section S is not constant and the definition of Z_0 has to be generalized. The wave impedance is then defined as a function $Z_0(s)$ such that the relations $P^\pm(r,s) = \pm Z_0(s)U^\pm(r,s)$ hold in the Laplace domain. It can be seen that $Z_0(s) = \rho_{air}c/S \cdot [rs/(rs+c)]$.

Summarizing, this section has shown that vibrational phenomena in many elastic media can be described as one-dimensional wave propagations. Furthermore, Kirchhoff and wave variables in these media are related through wave impedance. This results provide the basis for developing 1-D waveguide structures.

2.1.2 Waveguide structures

Waveguide models exploit the existence of the solution (2.2) to the D'Alembert equation and discretize this solution instead of the differential equation itself. This remark

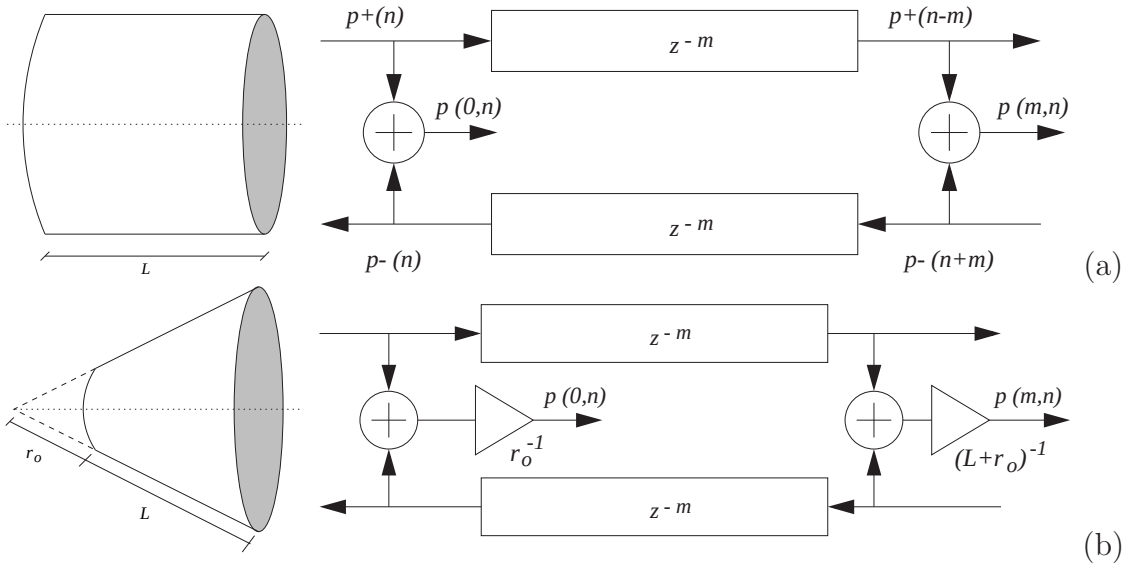

Figure 2.2: *Lossless waveguide sections with observation points at position* $x = 0$ *and* $x = mX_s = L$*; (a) cylindrical section; (b) conical section.*

explains to a large extent why waveguide structures are much more efficient than finite difference methods in simulating vibrations of elastic media.

Consider a pressure distribution $p = p^+ + p^-$ inside an ideal lossless cylindrical bore. If T_s is the sampling period, a suitable choice for the spatial sampling step is $X_s = cT_s$. Thus, a discretized version of p is obtained through the variable substitution $x \mapsto mX_s$ and $t \mapsto nT_s$ (with $m, n \in \mathbb{N}$), and leads to

$$p(mX_s, nT_s) = p^+(ncT_s - mX_s) + p^-(ncT_s + mX_s) = p^+[(n-m)cT_s] + p^-[(n+m)cT_s].$$

Removing the constant sampling steps yields:

$$p(m, n) = p^+(n - m) + p^-(n + m). \tag{2.9}$$

The term $p^+(n-m)$ in Eq. (2.9) can be thought of as the output from a digital delay line of length m, whose input is $p^+(n)$. Analogously, the term $p^-(n+m)$ can be thought of as the input of a digital delay line with the same lenght, whose output is $p^-(n)$. This remark leads to the definition of a *waveguide section* as a bidirectional delay line, as depicted in Fig. 2.2(a). Note that the horizontal direction of this structure has a straightforward physical interpretation: it corresponds to the position x along the axis of the cylindrical bore. In the example depicted in Fig. 2.2, two "observation points" have been chosen at $x = 0$ and $x = mX_s = L$. At these points, the pressure signal at time n is reconstructed by summing the corresponding pressure waves $p^{\pm}$. A very similar structure can be outlined for numerically simulating a pressure distribution in an ideal lossless conical bore [19]. In this case, propagation is described by the one-dimensional equation (2.5), whose general solution is given by Eq. (2.6). The conical waveguide is

therefore defined as in Fig. 2.2(b). Observation points can be chosen analogously to the cylindrical case.

As already mentioned, these waveguide sections describe ideal systems where the D'Alembert equation (2.1) or its spherical version (2.5) hold. Real systems exhibit more complex behavior. As an example, energy *dissipation* occurs in any real vibrating medium. In an acoustical bore this is due to air viscosity, thermal conduction and wall losses. Dissipation in a string comes from internal losses related to elastic properties of the material, energy transfer through terminations, and friction with air. For clarity, consider the pressure distribution in a cylindrical bore. In the simplest approximation, all of the dissipation phenomena can be incorporated in the D'Alembert equation by including an additional term proportional to the first time derivative:

$$\frac{\partial^2 p}{\partial t^2}(x,t) = c^2 \frac{\partial^2 p}{\partial x^2}(x,t) - \epsilon \, \frac{\partial p}{\partial t}(x,t). \tag{2.10}$$

In the limit of small ϵ, Eq. (2.10) still admits a traveling wave solution, which can be digitized with the same procedure described in the ideal case:

$$\begin{aligned} &p(x,t) = e^{-\frac{\epsilon x}{2c}} p^+(ct-x) + e^{\frac{\epsilon x}{2c}} p^-(ct+x), \qquad \text{then} \\ &p(m,n) = g^m p^+(n-m) + g^{-m} p^-(n+m), \qquad \text{with} \quad g = e^{-\frac{\epsilon T_s}{2}} < 1. \end{aligned} \tag{2.11}$$

Thus the traveling waves are exponentially damped along the propagation direction, and this phenomenon can be easily incorporated in the waveguide structure. This is shown in Fig. 2.3(a), where losses have been consolidated, or *lumped*, in a single multiplier cascaded to the delay line. The loss factor g^m summarizes the distributed losses occurring in the spatial interval $[0, mX_s]$. In most of real phenomena, however, losses increase with frequency. A better approximation of dissipation phenomena can account for this frequency dependence by substituting the constant factor g with a lowpass filter $G(z)$. Moreover, in order to avoid frequency dependent delay, $G(z)$ must be a zero-phase FIR filter. Alternatively, a linear-phase filter can be used; in this case the length of the delay line has to be reduced correspondingly, in order to obtain the desired overall delay. For a more detailed discussion of lossy waveguide structures see [110].

A second important phenomenon in natural wave propagation is that of *dispersion*. In a string, dispersion is introduced by string stiffness. This is usually modeled in the D'Alembert equation (2.1) by introducing an additional term proportional to the fourth spatial derivative:

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}(x,t) = \frac{\partial^2 p}{\partial x^2}(x,t) - \epsilon \, \frac{\partial^4 p}{\partial^4 x}(x,t), \tag{2.12}$$

where the dispersive correction term ϵ is proportional to the string Young's modulus. If ϵ is small, its first order effect is to increase the wave propagation speed with frequency:

$$c(\omega) = c_0 \left(1 + \frac{\epsilon \omega^2}{2c_0^2}\right), \tag{2.13}$$

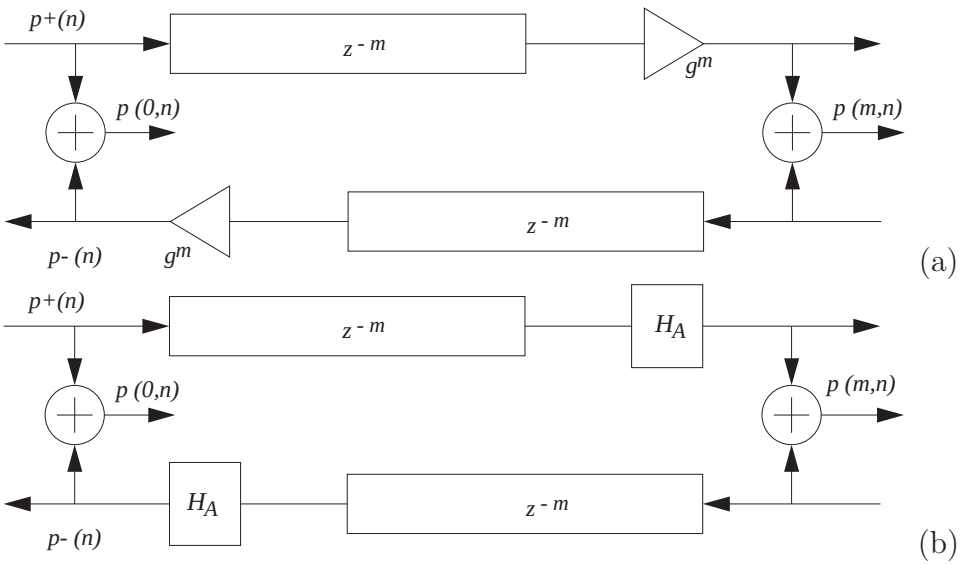

Figure 2.3: *Waveguide simulation of non-ideal media; (a) frequency independent dissipation; (b) dispersion.*

where c_0 is now the wave travel velocity in the absence of dispersion. Equation (2.13) states that a traveling wave is no longer a rigid shape that translate at constant speed. Instead, frequencies "disperse" as they propagate with different velocities. As a consequence, the frequencies ω_k of the allowed partials are not harmonic, instead they are stretched onto an inharmonic series according to the equation

$$\omega_k = k\omega_0 I_k, \quad \text{where} \quad I_k \approx \sqrt{1 + Bk^2},$$

and where $B = \pi^2\epsilon/L^2$. The quantity I_k is usually termed *index of inharmonicity*. Dispersion is particularly important in piano strings, where the lower tones exhibit significant inharmonicity [49].

Having a non-uniform wave velocity $c(\omega)$ implies that it is not possible to define a sampling step as $X_s = c_0 T_s$. Instead, it can be said that a component with frequency ω travels a distance $c_0 T_s$ in the time interval $c_0 T_s / c(\omega)$. As a consequence, the unitary delay z^{-1} has to be substituted with the all-pass filter $H_a(z) = z^{-c_0/c(\omega)}$, which has a unitary magnitude response but non-constant phase delay. Similarly to dissipative low-pass filters, m all-pass delays can be *lumped* in a single filter H_a^m. Moreover, the linear and non-linear parts of the phase response can be treated separately. H_a^m can thus be written as $H_a^m(z) = z^{-m} \cdot H_A(z)$, where $H_A(z)$ is another all-pass filter approximating the non-linear part of the phase response. In summary, a dispersive resonator is modeled as in Fig. 2.3(b). For a detailed discussion of dispersion filter design see Rocchesso and Scalcon [109].

One last improvement to the basic waveguide structure of Fig. 2.2 is provided by *fractional delay lines*. It is easily verified that with a sampling rate $F_s = 44.1$ [kHz] and with a wave velocity $c = 347$ [m/s] (sound velocity in air at 20 C°), the resulting spatial

step is $X_s = 7.8 \cdot 10^{-3}$ [m]. This distance produces perceivable pitch variations in a wind instrument. It is therefore necessary to design fractional delays in order to provide fine tuning of the length of a waveguide section. Without going into details, this can be ideally achieved by including an additional filter in the structure, with flat magnitude response (that does not affect the overall magnitude response of the waveguide structure) and linear phase response (that adds the desired fractional delay). Both interpolation filters (FIR) and all-pass filters (IIR) can be used for approximating such characteristics. More detailed discussions are provided by Välimäki [139], by Laakso *et al.* [79] and by Rocchesso [109, 103].

2.1.3 Junctions and networks

The last section has introduced the main concepts of waveguide modeling for a signal propagating in a *uniform* medium. When discontinuities are encountered, the wave impedance changes and signal *scattering* occurs, i.e. a traveling wave is partially reflected and partially transmitted. Examples of non-uniform media are a cylindrical bore where the cross-sectional area changes abruptly, or a string where the value of the linear mass density jumps changes discontinuously. In order to model these discontinuities, appropriate junctions have to be developed, that connect two (or more) waveguide sections. Terminations can be regarded as special cases of junctions, as discussed in the following paragraphs.

Consider two cylindrical bores, with cross-sectional areas $S_{1,2}$ and wave admittances $\Gamma_{1,2} = Z_{1,2}^{-1} = S_{1,2}/\rho_{air}c$, connected to each other. Analysis of this problem leads to the derivation of the well known *Kelly-Lochbaum* junction. First of all, physical constraints have to be imposed on the Kirchhoff variables p, u at the junction, namely pressures $p_{1,2}$ must have the same value p_J and the flows $u_{1,2}$ from the two sides must sum to zero:

$$u_1 + u_2 = 0, \qquad p_1 = p_2 = p_J. \tag{2.14}$$

Using the Kirchhoff analogy $p \leftrightarrow v$ (voltage) and $u \leftrightarrow i$ (current), Eqs. (2.14) can be regarded to describe a parallel junction. If pressure wave variables are introduced as in Eq. (2.8) (with p^+ and p^- denoting incoming and outgoing waves, respectively), and the junction pressure p_J is used, then the relation $p_l^- = p_J - p_l^+$ (for $l = 1, 2$) holds. Substitution in the first of Eqs. (2.14) yields

$$\begin{aligned} 0 &= (u_1^+ + u_1^-) + (u_2^+ + u_2^-) = \Gamma_1(p_1^+ - p_1^-) + \Gamma_2(p_2^+ - p_2^-) = \\ &= \Gamma_1(2p_1^+ - p_J) + \Gamma_2(2p_2^+ - p_J). \end{aligned}$$

From this, the junction pressure p_J can be expressed in terms of the incoming pressure waves $p_{1,2}^+$ as

$$p_J = 2\frac{\Gamma_1 p_1^+ + \Gamma_2 p_2^+}{\Gamma_1 + \Gamma_2}.$$

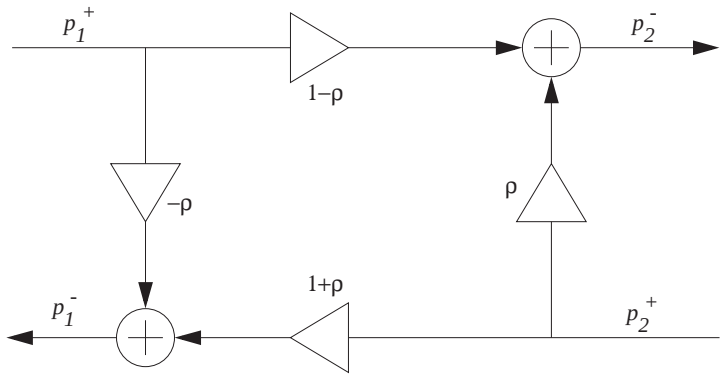

Figure 2.4: *Kelly-Lochbaum junction for two cylindrical bores with different areas.*

Using this latter expression, the outgoing pressure waves $p_{1,2}^-$ can be written as

$$\begin{aligned} p_1^- &= p_J - p_1^+ = -\frac{\Gamma_2 - \Gamma_1}{\Gamma_2 + \Gamma_1} p_1^+ + \frac{2\Gamma_2}{\Gamma_2 + \Gamma_1} p_2^+, \\ p_2^- &= p_J - p_2^+ = \frac{2\Gamma_1}{\Gamma_2 + \Gamma_1} p_1^+ + \frac{\Gamma_2 - \Gamma_1}{\Gamma_2 + \Gamma_1} p_2^+. \end{aligned} \tag{2.15}$$

If the *reflection coefficient* ρ is defined as

$$\rho := \frac{\Gamma_2 - \Gamma_1}{\Gamma_2 + \Gamma_1},$$

then Eqs. (2.15) become

$$\begin{aligned} p_1^- &= -\rho\, p_1^+ + (1+\rho) p_2^+, \\ p_2^- &= (1-\rho) p_1^+ + \rho\, p_2^+. \end{aligned} \tag{2.16}$$

These equations describe the Kelly-Lochbaum junction. A scattering diagram is depicted in Fig. 2.4.

This junction has been extensively used in what are sometimes termed "multitube lossless models" of the vocal tract [37]. These are basically articulatory models where the vocal tract shape is approximated as a series of concatenated cylindrical sections. Pressure wave propagation in each section is then described using digital waveguides, and interconnections are treated as Kelly-Lochbaum junctions. Remarkably, the same junction can be used to describe not only acoustic, but also mechanical structures. As an example, consider two strings with different densities, connected at one point: this can be thought of as a series junction, since the physical constraints impose that velocity (i.e., "current") has to be the same on the left and right sides, and the sum of forces (i.e., "voltages") from the two sides must be zero. Analogously to the above analysis, a series Kelly-Lochbaum junction can be derived in this case.

Terminations of a waveguide model are an interesting particular case of junctions. Consider an ideal cylindrical bore, closed at one end: this boundary condition corresponds

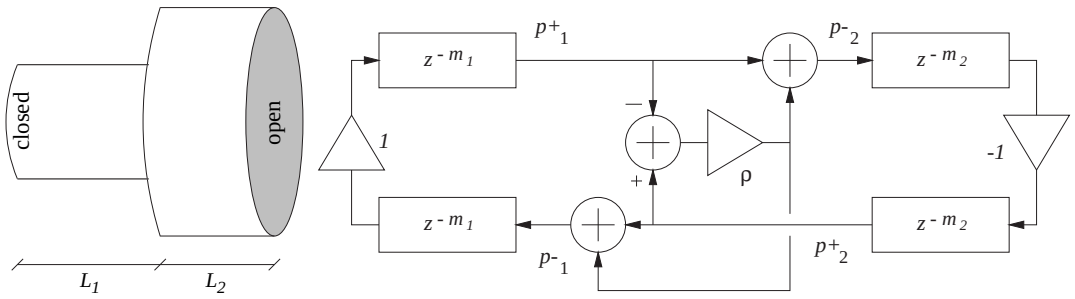

Figure 2.5: *Example of an acoustic model implemented with waveguide sections an junctions.*

[49] to an infinite impedance $Z_2 = \infty$ (i.e., $S_2 = 0$), and thus to a reflection coefficient $\rho = -1$. In other words, complete reflection occurs and the relation $p_1^-(0,t) = p_1^+(0,t)$ holds. Similarly, an ideally open end can be seen to correspond to $Z_2 = 0$ (i.e., $S_2 = \infty$), and thus to $\rho = 1$: this is a second case where complete reflection occurs, namely the relation $p_1^-(0,t) = -p_1^+(0,t)$ holds. Analogous considerations hold for string terminations. In non-ideal cases the strings can be not perfectly fixed at the termination, and tension modulations can be introduced [138].

Figure 2.5 shows an example where different junctions have been used and combined into a waveguide model. Note that in this example the scattering junction between the two cylindrical sections is not in the Kelly-Lochbaum form; instead, a *one-multiply scattering junction* is used, which allows more efficient implementation of Eqs. (2.16). Open- and closed-tube terminations are modeled according to the above remarks.

The result expressed in Eq. (2.16) can be easily extended to higher dimensions. Consider parallel junction of N acoustical bores. In this case a *scattering matrix* can be found, and Eq. (2.16) is generalized to

$$\boldsymbol{p}^- = \boldsymbol{A} \cdot \boldsymbol{p}^+, \tag{2.17}$$

where $\boldsymbol{p}^\pm$ are n-dimensional vectors whose elements are the incoming and outgoing pressure waves in the n bores. The physical constraints expressed in Eq. (2.14) are also generalized in an obvious way, and calculations analogous to those outlined for the Kelly-Lochbaum junction lead to the result

$$\boldsymbol{A} = \begin{bmatrix} \frac{2\Gamma_1}{\Gamma_J} - 1, & \frac{2\Gamma_2}{\Gamma_J}, & \cdots & \frac{2\Gamma_N}{\Gamma_J} \\ \frac{2\Gamma_1}{\Gamma_J}, & \frac{2\Gamma_2}{\Gamma_J} - 1, & \cdots & \frac{2\Gamma_N}{\Gamma_J} \\ \vdots & & \ddots & \vdots \\ \frac{2\Gamma_1}{\Gamma_J}, & \frac{2\Gamma_2}{\Gamma_J}, & \cdots & \frac{2\Gamma_N}{\Gamma_J} - 1 \end{bmatrix}, \qquad \text{where} \quad \Gamma_J = \sum_{l=1}^{N} \Gamma_l. \tag{2.18}$$

Note that when $N = 2$ Eq. (2.17) reduces to the Kelly-Lochbaum equations.

A final remark is concerned with junctions of conical elements: this is not a straightforward generalization of the cylindrical case. Indeed, the derivation of Kelly-Lochbaum

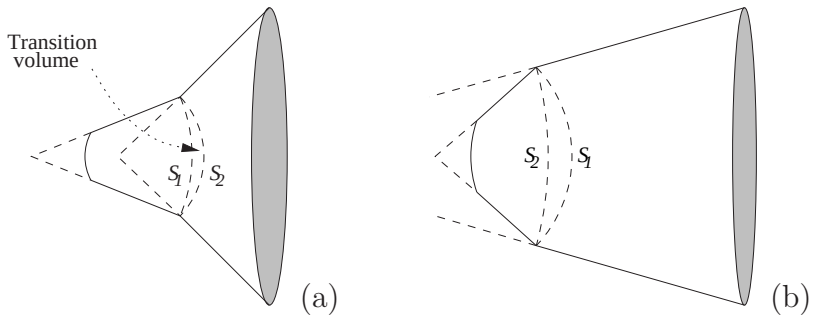

Figure 2.6: *Boundary regions for (a) non-convex and (b) convex conical junctions.*

equations is based on the implicit assumption of plane wave propagation. This assumption permits imposition of the constraints (2.14) on a flat scattering boundary, which is a wavefront for both p_1 and p_2. But wavefronts in conical sections are spherical and this circumstance makes it impossible to define a unique surface on which boundary conditions can be applied: Fig. 2.6(a) shows that there is a region between the two spherical wavefronts which is within neither conical segment. This ambiguity in the definition of the scattering boundary is usually overcome by assuming that the transition volume is small and thus pressure is constant inside the volume [3, 4, 87, 88]. A second and more serious problem arises when looking for a digital junction analogous to Eqs. (2.16). The reflection coefficient ρ is generalized to a first order IIR filter $R(z)$ (see Välimäki and Karjalainen [140]). However, this filter turns out to be unstable (non-causal growing exponential) in the case of the convex configuration depicted in Fig. 2.6(b). Berners [20] shows that this is ultimately due to the presence of a trapped mode, such that the Fourier traveling components do not constitute a complete basis set in the region surrounding the discontinuity.

2.2 Lumped models

Lumped models are well suited for describing systems whose spatial dimensions are small compared to acoustic wavelengths. A single reed in a wind instrument is a typical example: it is small enough to assume that pressure along the internal surface of the reed is constant; the frequency of the first oscillation mode of the reed is well above any playable frequency in a wind instrument; oscillations occur mainly in the vertical direction, and as a first approximation the system can be described with a single degree of freedom, i.e. the reed tip vertical displacement. Based on these considerations, many authors approximate the reed behavior with that of a lumped second order mechanical oscillator (Sec. 3.1 discusses this model in more detail).

Section 2.2.1 introduces analogies between mechanical, acoustic and electric lumped systems; the role of lumped non-linearities is discussed in Sec. 2.2.2.

2.2.1 Analogies

In a large class of systems it is possible to construct pairs of variables characterized by the property that their product has the dimensions of power [Kg m^2/s^3]. In electrical systems such a pair of variables is given by (v, i), voltage and current. Integro-differential relations can be found that relate these two variables, in particular three elementary relations define the fundamental quantities resistance R, inductance L and capacitance C. In the Laplace domain, the integro-differential equations are turned into simple algebraic relations:

$$V(s) = R \cdot I(s), \qquad V(s) = sL \cdot I(s), \qquad V(s) = \frac{1}{sC} I(s). \tag{2.19}$$

These are particular examples of a more general relation in linear electric circuits:

$$V(s) = Z(s)I(s), \tag{2.20}$$

where the quantity $Z(s)$ is called *impedance* of the circuit and is defined as the ratio between the Laplace transforms of voltage and current intensity. The inverse of $Z(s)$ is called *admittance*, and it is usually denoted as $\Gamma(s) = Z(s)^{-1}$.

An analogous pair of variables are found in mechanical systems: force f [Kg m/s^2] and velocity v [m/s] satisfy the same condition of voltage and current, i.e. their product is a power. Therefore, f and v are taken as mechanical Kirchhoff variables. Again, the ratio of these two variables in the Laplace domain is defined as (mechanical) *impedance*, and its inverse is the (mechanical) admittance. Using three notable relations between f and v, it is possible to introduce mechanical equivalents of resistance, capacitance and inductance. The simplest relation is direct proportionality: $f(t) = rv(t)$. This is used to define ideal linear viscous forces, and comparison with the first of Eqs. (2.19) permits r to be regarded as a mechanical resistance. Newton's second law of classical dynamics provides a second relation: the inertial mass m of a non-relativistic body is defined as the ratio between the total force acting on it and its acceleration, i.e. $f(t) = ma(t) = m\dot{v}(t)$. In the Laplace domain this is turned into $F(s) = msV(s)$, and from comparison with the second equation in (2.19) m is seen to be equivalent to an inductance. Finally, Hooke's law provide an analogy to electrical capacitance: in an ideal linear spring the elastic force is proportional to the elongation of the spring: $f(t) = kx(t) = k\int_0^t v(\tau)d\tau$. Again, in the Laplace domain this is turned into $F(s) = k/s\ V(s)$, and comparison with the third of Eqs. (2.19) shows that the stiffness constant k of the spring corresponds to the reciprocal of a capacitance. Summarizing, the analogies between mechanical and electrical elements are as follows:

$$\begin{array}{ccc} F(s) = r \cdot V(s), & F(s) = \dfrac{k}{s} V(s), & F(s) = ms \cdot V(s), \\ \Downarrow & \Downarrow & \Downarrow \\ r \sim R, & \dfrac{1}{k} \sim C, & m \sim L. \end{array} \tag{2.21}$$

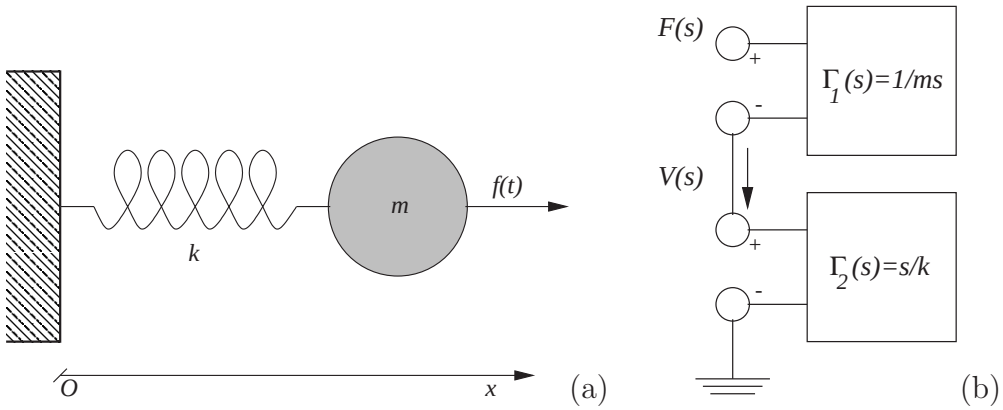

Figure 2.7: *A mass pulled by a linear spring; (a) mechanical system and (b) combination of one-ports in series.*

Figure 2.7(a) shows the simplest example of a "series" junction between these mechanical elements: a mass attached to an ideal linear spring and driven by an external force. The system's dynamics are described by the equation

$$m\ddot{x}(t) = -kx(t) + f(t), \qquad \Rightarrow \qquad F(s) = \left(ms + \frac{k}{s}\right) V(s). \tag{2.22}$$

The second equation (2.22) shows that the aggregate impedance $Z(s)$ of the system is the sum of the two elementary impedances $Z_1(s) = ms$ and $Z_2(s) = k/s$.

The above discussion is the starting point for developing one-port network theory for mechanical systems. The one port is defined as a black-box with a single pair of input/output terminals, as in Fig. 2.7(b). A force is applied at the terminals, analogously to an electrical potential, and velocity "flows" as electrical current. Instantaneous power and energy can be defined and used to characterize *passive* and *lossless* one-ports. Connections through ports can be made using Kirchhoff's Laws, so that series and parallel junctions are defined analogously to circuit theory. In circuit theory terminology, the two one-ports in Fig. 2.7(b) share a common velocity, thus they are connected in series. Bilbao [22] discusses these issues in detail.

It is worth mentioning that in certain situations, acoustical systems can be described in terms of lumped elements only. In particular, when the dimensions of an acoustical element are much less than the sound wavelength, then the acoustical pressure, p can be assumed constant In this case, the acoustic behavior of the element is, at least at low frequencies, very simple. Note that the acoustic pressure p [Kg/ms^2] and the volume velocity u [m^3/s] are such that their product is a power, and can therefore be used as a pair of Kirchhoff variables.

Resistive phenomena are observed during the passage of acoustic airflow through a small opening. In this case the flow behavior is dominated by viscous and thermal losses and it is reasonably assumed to be in phase with the acoustic pressure, therefore

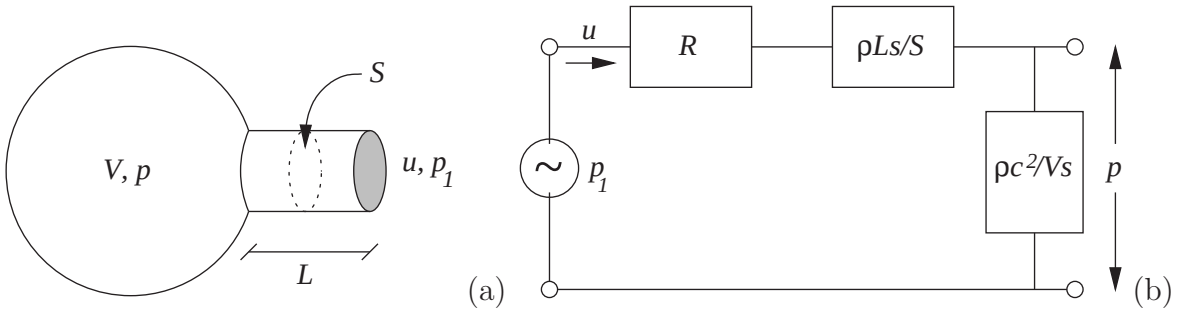

Figure 2.8: *A Helmholtz resonator driven by an external acoustic wave; (a) acoustic system and (b) circuit representation.*

the relation $p(t) = Ru(t)$ holds at the opening where the constant R is termed *fluid-dynamic resistance*. Fluid-dynamic inductance is defined in a short, open tube having cross-sectional area S and length L. The air mass inside the bore is then $m = \rho_{air}SL$ (ρ_{air} being the air density). Suppose that an acoustic pressure $p(t)$ is applied to one end of the tube; then the enclosed air behaves like a lumped mass driven by the force Sp, and Newton's law implies

$$Sp(t) = \rho_{air}SL \cdot \dot{v}(t), \qquad \Leftrightarrow \qquad P(s) = \frac{\rho_{air}L}{S} \cdot sU(s),$$

where the relation $u(t) = Sv(t)$ has been used, and $v(t)$ indicates particle velocity. Finally, capacitance is associated with air volumes. Consider the volume $V(t)$ of air inside a cavity; the contraction $dV(t)$ caused by an acoustic pressure $p(t)$ is such that $-\rho_{air}c^2 \cdot dV/V = p$, where $\rho_{air}c^2$ is the bulk modulus of air at atmospheric pressure. As a consequence, a new air volume $-dV$ can enter the cavity. By definition, this equals the integral of $u(t)$ over time, therefore

$$-dV(t) = \int_0^t u(t')dt' = \frac{V}{\rho_{air}c^2}p(t), \qquad \Leftrightarrow \qquad P(s) = \frac{\rho_{air}c^2}{Vs}U(s).$$

Comparison of this relation with the last of Eqs. (2.19) is then straightforward: it is immediately seen that the quantity $\rho_{air}c^2/Vs$ is the acoustical equivalent of a capacitive impedance.

Analogously to the mechanical case, simple acoustic systems can be described as combinations of these elementary impedances. Consider a Helmholtz resonator driven by an external sound wave, as in Fig. 2.8(a). Both the inductive impedance associated with the tube and the resistance associated with the opening impede the same flow u, and are therefore in series. This flow u enters the cavity, so that the capacitance associated with the volume is in series with the other two. The resulting acoustic circuit is depicted in Fig. 2.8(b).

Table 2.1 summarizes the main analogies between electrical, mechanical and acoustical systems, as discussed throughout this section.

Electrical		Mechanical		Acoustical	
Current i [A]		Velocity v [m/s]		Flow u [m^3/s]	
Voltage v [V]		Force f [N]		Pressure p [Pa]	
(Resistance) R	$\left[\frac{\mathrm{Kg}\cdot\mathrm{m}^2}{\mathrm{s}}\right]$	(Damping) r	$\left[\frac{\mathrm{Kg}}{\mathrm{s}}\right]$	(Opening) R	$\left[\frac{\mathrm{Kg}}{\mathrm{m}^4\cdot\mathrm{s}}\right]$
(Capacitance) $\frac{1}{sC}$		(Spring) $\frac{k}{s}$		(Cavity) $\frac{\rho_{air}c^2}{Vs}$	
(Inductance) $\frac{s}{L}$		(Mass) $m \cdot s$		(Bore) $\frac{\rho_{air}Ls}{S}$	

Table 2.1: *Summary of analogies in electrical, mechanical and acoustical systems.*

2.2.2 Non-linearities

As mentioned in chapter 2, musical oscillators are typically non-linear: non-linearities must be present for a system to reach stable self-sustained oscillations, as in the case of persistently excited instruments (e.g., winds and strings). Non-linear elements may also be present in other systems in order to account for accurate modeling of interaction mechanisms. As an example, collisions between lumped masses are often described through a non-linear contact force.

The previous section has outlined the formal analogies between linear mechanical and electrical systems. It is possible to extend the analogy to the non-linear case. Consider the well known *Chua-Felderhoff* electrical circuit: this is a RLC circuit, made of a series connection of a resistor R, an inductor L and a capacitor C. The elements R and L are constant, while this is not the case for C. More precisely, the characteristic of the capacitance is a function of the voltage v, so that the system is described as follows:

$$\begin{aligned} v(q) &= \frac{1}{2v_0C_0}\left(q^2 + q\sqrt{q^2 + 4C_0^2v_0^2}\right), \quad \Leftrightarrow \quad C(v) = \frac{C_0}{\sqrt{1+\dfrac{v}{v_0}}}, \\ v(q) &+ R\dot{q}(t) + L\ddot{q}(t) = v_e(t), \qquad (v > v_0). \end{aligned} \tag{2.23}$$

The variable $q(t)$ stands for the charge on the capacitor, and $v_e(t)$ is an applied voltage. It is easily verified that $C(v) \sim C_0$ when $v \to 0$, i.e. the system is a linear RLC circuit in the limit of small oscillations. However, for larger voltage v this approximation does not hold, and $C(v)$, $q(v)$ behave as depicted in Fig. 2.9(a) and (b), respectively. Note that there is no easy way to translate the non-linear relation (2.23) into the Laplace domain, because the definition of impedance given in Sec. 2.2.1 assumes linearity of the circuit elements. The Chua-Felderhoff circuit has been extensively studied and is one of the classical systems used for exemplifying transition to chaotic behavior: when the peak

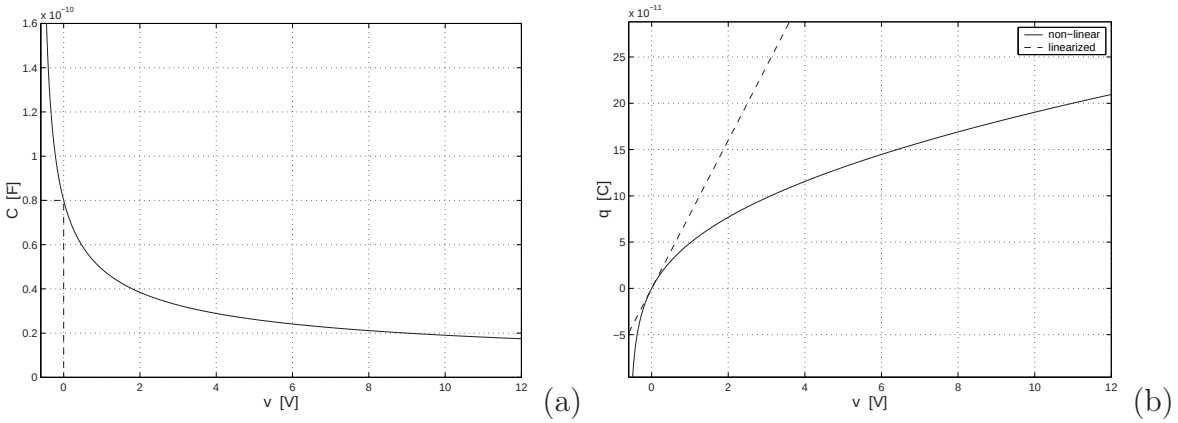

Figure 2.9: *Non-linear behavior of (a) capacitance $C(v)$ and (b) charge $q(v)$ in the Chua-Felderhoff circuit.*

of the voltage generator is increased, the behavior of the charge $q(t)$ on the capacitor undergoes successive bifurcations.

The Chua-Felderhoff circuit finds some analogous counterparts in mechanical and acoustic systems. As an example, a non-linear lumped model for a single reed is proposed in chapter 3, where the lumped parameters depend on the reed tip displacement:

$$m(y_L)\ddot{y}_L(t) + r(y_L)\dot{y}_L(t) + k(y_L)y_L(t) = S_d\Delta p(t), \tag{2.24}$$

where y_L is the reed tip displacement, Δp is the pressure drop along the slit and $m(y_L)$, $r(y_L)$, $k(y_L)$ are the reed mass, damping and stiffness, respectively (see Sec. 3.4 for details). In the limit of small oscillations, system (2.24) reduces to a lumped model with constant parameters. Self-sustained oscillations are reached due to the non-linear interaction between the displacement y_L and the pressure drop Δp (see Sec. 3.1). At larger amplitudes however, the lumped parameters are no longer constant and additional non-linearities come at play. The situation is analogous to that described for the Chua-Felderhoff circuit, where the non-constant capacitance $C(v)$ starts to affect the system behavior at large amplitudes.

A second mechanical example of non-linear elements is provided by an idealized contact model. In this model the contact restoring force is a non-linear elastic force of the form

$$f(x(t)) = \begin{cases} kx(t)^\alpha, & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{2.25}$$

where x is the penetration and k is an elastic constant. This model has been used for describing the compression characteristics of a piano hammer felt [66, 10]. The exponent α depends on the local geometry around the contact surface and can typically take values between 1.5 and 3.5. If the hammer is regarded as a lumped mass m and linear dissipation

r is taken into account, then the complete model is described by the equation of motion

$$m\ddot{x}(t) + r\dot{x}(t) + kx(t)^{\alpha} = f_e(t). \tag{2.26}$$

This is formally identical to Eq. (2.23): the non-linear hammer is a series connection of a mechanical resistance r and inductance m with a non-linear capacitance. One obvious structural difference with the Chua-Felderhoff circuit is given by the different shape of the non-linearities. A second, functional difference is that the external force in this case is due to the interaction with the piano string, while f_e in Eq. (2.23) is provided by an external generator. Chapter 5 discusses an interesting generalization of this contact model, and its main properties.

2.3 Stability, accuracy, computability

The lumped models described in Sec. 2.2 are developed in the continuous-time domain and are described through sets of ODEs. In order to be implemented as numerical algorithms for sound synthesis, the differential equations have to be discretized in an efficient and effective manner. In most cases, a trade-off has to be found between accuracy of the discretization technique and efficiency of the resulting algorithms.

From a historical perspective, the issue of discrete-time simulation of ordinary differential equations has been addressed by researchers both in digital signal processing (DSP) and in numerical analysis, and research in these two fields have to a large extent progressed with little exchange of information. Very roughly, it can be said that DSP techniques concentrate mainly on linear differential equations (a.k.a. filters), while methods developed in numerical analysis provide a unified discussion of linear and non-linear systems.

This section discusses the main aspects related to discretization of lumped models. In Sec. 2.3.1, numerical techniques and their main properties are reviewed. Attention is mainly focused on digital signal processing, rather than numerical analysis techniques. Section 2.3.2 addresses non-computability problems in non-linear numerical systems, and reviews a general method for dealing with these problems.

2.3.1 Numerical methods

When dealing with linear systems, such as the lumped elements of Sec. 2.2.1, the most elementary numerical technique is sampling. Given the admittance $\Gamma(s)$ of a linear system (corresponding to defining the input as the driving force and the output as the resulting velocity), its inverse Laplace transform $\gamma(t)$ is the continuous-time impulse response. The linear system can thus be digitized by defining the discrete response as $\gamma_d(n) := T_s\gamma(nT_s)$, i.e. by sampling $\gamma(t)$. This technique is widely used in the context of digital filter design [96, 92], and it is usually termed the *Impulse Invariant Method*. One quality of the method is that stability is guaranteed at any sampling rate. Indeed, if p_c is a pole of the continuous-time response, the corresponding pole of the discrete-time response is given

by $p_d = e^{p_c T_s}$. This implies that if $\mathrm{Re}(p_c) < 0$, then $|p_d| < 1$, i.e. the discrete-time pole lies inside the unit circle. On the other hand, a drawback of the method is *aliasing*. It is known that the discrete-time response Γ_d is obtained as a periodization of the continuous one:

$$\Gamma_d(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} \Gamma\left(\frac{j\omega}{T_s} + j\frac{2k\pi}{T_s}\right). \tag{2.27}$$

As a consequence, any Γ whose bandwidth is wider than $F_s/2$ introduces spurious components in Γ_d. Since no physical system is truly bandlimited, aliasing will always occur unless one can guarantee that the system is driven with bandlimited signals.

An approach alternative to sampling amounts to replacing time derivatives with finite differences, thus turning the differential equations directly into difference equations. In the Laplace domain, the derivation operator is turned to a multiplication by s. Similarly, in the Z domain the unit delay is turned into a multiplication by z^{-1}. Therefore, approximating derivatives with finite differences corresponds in the frequency domain to finding appropriate *s-to-z mappings*. Let $s = g(z)$ be such a mapping, then the discrete-time response is found as $\Gamma_d(z) = \Gamma\left(g(z)\right)$.

The simplest possible mapping is obtained by replacing the derivative with an incremental ratio. Let $x(t)$ be a smooth function of time, then

$$\begin{aligned} &\frac{d}{dt}x(t_n) := \lim_{h\to 0_+} \frac{x(t_n) - x(t_n - h)}{h} \approx \frac{x(t_n) - x(t_{n-1})}{T_s} := \delta_t x(n), \\ &\Rightarrow \qquad s \approx \frac{1 - z^{-1}}{T_s} := g_1(z). \end{aligned} \tag{2.28}$$

where $t_n = nT_s$. The mapping $g_1(z)$ is known in numerical analysis as the *backward Euler method* [80]. The adjective "backward" is used because the first derivative of x at time n is estimated through the values of x at time n and $n-1$. Note that the method is *implicit*, since it turns a generic first-order differential equation $\dot{x}(t) = f(x(t), t)$ into a difference equation of the form $x(n) = f_d(x(n), x(n-1), n)$, where $x(n)$ depends implicitly on itself. Higher-order derivatives can be estimated through iterate application of Eq. (2.28). The second derivative is found as

$$\frac{d^2}{dt^2}x(t_n) \approx \frac{1}{T_s}\left[\delta_t x(n) - \delta_t x(n-1)\right] = \frac{x(t_n) - 2x(t_{n-1}) + x(t_{n-2})}{T_s^2}. \tag{2.29}$$

However, a centered estimate is more often used:

$$\frac{d^2}{dt^2}x(t_n) \approx \frac{x(t_{n+1}) - 2x(t_n) + x(t_{n-1})}{T_s^2}. \tag{2.30}$$

A second, widely used s-to-z mapping is provided by the *bilinear transform*. Like the backward Euler method, this can be seen as a finite approximation of the time derivative:

$$\begin{aligned} &\frac{x(t_n) - x(t_{n-1})}{T_s} \approx \frac{\dot{x}(t_n) + \dot{x}(t_{n-1})}{2}, \\ &\Rightarrow \qquad s \approx \frac{2}{T_s}\frac{1 - z^{-1}}{1 + z^{-1}} := g_2(z). \end{aligned} \tag{2.31}$$

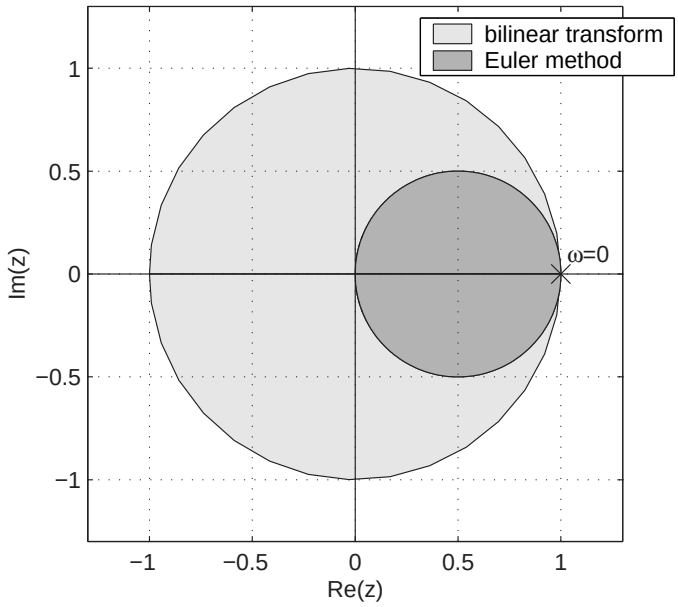

Figure 2.10: *Images of the vertical axis $s = j\omega$ (solid lines) and of the left-half s-plane (gray regions) using the backward Euler method g_1 and the bilinear transform g_2.*

The mapping $g_2(z)$ is known in numerical analysis as the one-step *Adams-Moulton method* [80]. Like the backward Euler method, it is implicit. By comparing the first estimate in Eq. (2.31) with the first in Eq. (2.28), it is intuitively seen that the bilinear transform provides a more accurate approximation than the Euler method. A rigorous analysis (e.g., Lambert [80]) would show that the order of accuracy of the bilinear transform is two, while that of the backward Euler method is one.

Another way of comparing the two techniques consists in studying how the frequency axis $s = j\omega$ and the left-half plane $\mathrm{Im}(s) < 0$ are mapped by g_1 and g_2 into the discrete domain. This gives information on the stability and accuracy properties of g_1 and g_2. Figure 2.10 illustrates this study. Two remarks must be made. First, both the methods define one-to-one mappings from $s = j\omega$, onto the two circles plotted in Fig. 2.10 (solid lines): thus, no frequency aliasing is introduced. Second, both the methods are stable, since the left-half s-plane is mapped inside the unit circle by both g_1 and g_2.

However, it is also seen both the mappings introduce frequency warping, i.e. the frequency axis is distorted. For the bilinear transform g_2, it can be seen that the $s = j\omega$ axis is mapped exactly onto the unit circle $z = e^{j\omega_d}$. Therefore, a direct mapping can be defined between the continuous frequencies ω and the discrete frequencies ω_d (see Mitra [92]): $\omega_d = 2 \cdot \arctan(\omega)$. At low frequencies, ω_d increases almost linearly with ω, while higher frequencies are progressively compressed (warped) and the Nyquist frequency $\omega_{Nyq} = \pi F_s$ is mapped to the point $z = -1$. Warping is the main drawback of the bilinear transform.

The Euler method maps the $s = j\omega$ axis onto the circle of radius 1/2 centered at

$z = 1/2$. Therefore no direct mapping is found from ω to ω_d. The function g_1 can be said to "doubly" warp the frequency axis: there is a progressive warping in the direction of increasing frequency (similarly to the bilinear transform), but there is also warping normal to the frequency axis. As far as stability is concerned, Fig. 2.10 shows that the poles of the discrete-time system obtained with g_1 are more "squeezed" inside the unit circle than the ones obtained with g_2. Furthermore, it can happen that continuous-time poles with positive real-part are turned by g_1 into discrete-time poles with modulus less than unity: in other words g_1 can turn unstable continuous systems into stable discrete systems. This *numerical damping* is a second major drawback of the Euler method. An example of such a damping property of the Euler method is provided in Sec. 3.2 (see in particular Figs. 3.4 and 3.5).

The bilinear transform finds application in Wave Digital Filters (*WDF*) theory [45]. These structures are the digital equivalent of the lumped circuits described in Sec. 2.2. Wave digital filters are constructed in two steps. The first step amounts to converting the continuous-time lumped circuits in wave variables. In this context, the definition of wave variables is identical to that used for waveguides models (see Eq. (2.8) in Sec. 2.1), namely:

$$f^+ = \frac{f + Z_0 v}{2}, \qquad f^- = \frac{f - Z_0 v}{2}, \tag{2.32}$$

where the mechanical Kirchhoff variables force f and velocity v have been used for clarity. The only and fundamental difference with Eq. (2.8) is that in this context the impedance Z_0 is a reference impedance that can be given any value and has no direct physical interpretation. Moreover, the variables $f^\pm$ themselves do not have a clear physical interpretation because in a lumped model they cannot be easily interpreted as traveling waves. Therefore in this context the Eqs. (2.32) have to be regarded as a mere change of coordinates.

Consider one of the elementary lumped elements analyzed in Sec. 2.2 and its associated impedance $Z(s)$. Then the new continuous-time variables $f^\pm$ are related to each other through a *reflectance* $R(s)$:

$$F(s) = Z(s)V(s), \quad \Rightarrow \quad F^-(s) = R(s)F^+(s), \quad \text{with} \quad R(s) := \frac{Z(s) - Z_0}{Z(s) + Z_0}. \tag{2.33}$$

The second step in WDF design is the discretization of $R(s)$. The equivalent wave digital filter $R_d(z)$ is then obtained using the bilinear transform: $R_d(z) = R(g_2(z))$. Note that since the reference impedance Z_0 can be given any value, this provides an additional degree of freedom for the design of R_d. In particular, Z_0 can be chosen such that R_d has no delay-free paths from input to output. This is an essential requirement for guaranteeing computability when connecting more than one element. A simple example will help clarify this concept: consider a mass mass m and its associated impedance $Z(s) = ms$, as found in Sec. 2.2. Then, from Eq. (2.33) the corresponding reflectance is $R(s) = (ms - Z_0)/(ms + Z_0)$. Choosing $Z_0 = m$ leads to the interesting result

$$R(s) = \frac{s-1}{s+1}, \quad \overset{s=g_2(z)}{\Rightarrow} \quad R_d(z) = -z^{-1}, \tag{2.34}$$

so no delay-free path is present in the wave digital filter R_d.

Wave digital filters can be generalized to incorporate non-linearities. This is a major issue, since non-linearities are predominant in musical acoustics and a self-oscillating system is necessarily described by some underlying non-linear phenomena. In particular, the treatment of non-linear resistors in a wave digital structure is well established, where the term "non-linear resistor" stands for a non-linear element which is described through an algebraic relation between the two Kirchhoff variables. Neither the Chua-Felderhoff circuit (2.23) nor the piano hammer (2.25) contain non-linear resistors, while the quasi-static reed model described in Sec. 3.1.2 introduces a non-linear resistor. Sarti and De Poli [112] have proposed an interesting generalization of the non-linear wave theory, that allows treatment of a more general class of *algebraic* non-linearities, i.e. non linear elements described through algebraic equations involving two Kirchhoff variables together with their time derivatives.

As a conclusion to this section, it should be remarked that the spectrum of available numerical techniques is much broader than those usually adopted in digital signal processing, and reviewed so far. As an example, higher-order Adams-Moulton methods are available from numerical analysis [80], and they can be easily interpreted as high-order s-to-z mappings (see Schneider *et al.* [118] for such an interpretation): high-order mappings are used in Sec. 3.2 for discretizing the single reed equations. Furthermore, so called *predictor-corrector* schemes [80] have been studied, that use an explicit numerical method (the predictor) in combination with an implicit one (the corrector): these schemes can in principle provide extremely efficient solutions to the computability problems found in implicit methods. As an example, high-order predictor-corrector schemes based on Obreckhoff methods are described by Lapidus and Seifeld [82]. Galkowski [56] has considered the use of high-order Obreckhoff methods from the viewpoint of systems theory. The possible applications of such methods for the design of physically-based numerical algorithms merit further investigation.

2.3.2 The K method

The discussion of wave digital filters in the last section has addressed the problem of non-computable loops in that particular context: wave variables rather than Kirchhoff variables are used to describe the components of the equivalent circuit, every component is treated as a scattering element with a reference impedance, and different components are connected by "adapting" Z_0. Wave methods can be said to be *local*, since non-computable paths are avoided by adapting the reference impedances of each element.

However, more severe computability problems can arise when simulating dynamic exciters, since the linear equations used to describe the system dynamics are tightly coupled with some non-linear map. These problems are addressed in this section, and a general method for solving non-computable paths (originally proposed by Borin *et al.* [24]) is reviewed. This method is termed the *K method*, since Kirchhoff variables are used. It operates on non-linearities using a global rather than a local approach. In particular, while the literature of wave digital filters deals only with non-linear bipoles

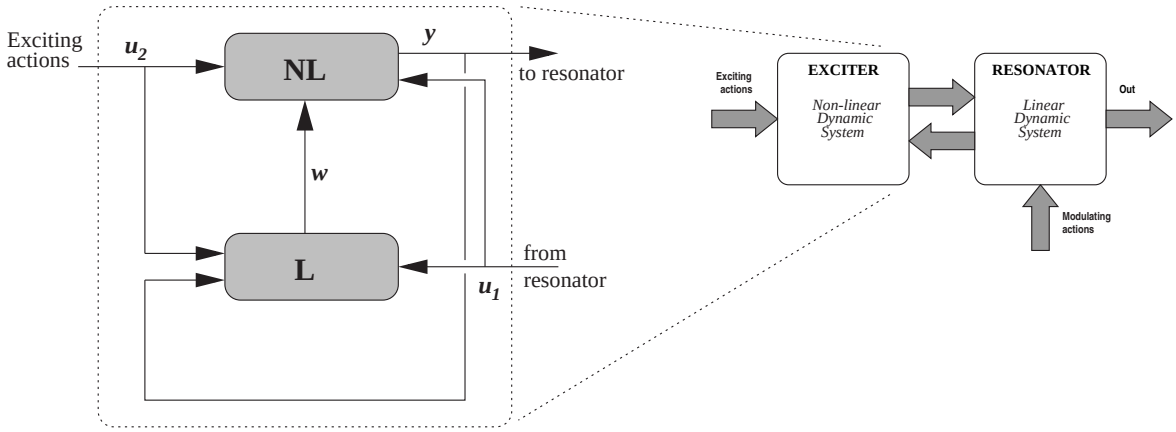

Figure 2.11: *Typical structure of a non-linear exciter.*

with or without memory, the K method can deal with non-linear functions of any set of Kirchhoff variables found in a circuit.

Figure 2.11 depicts a typical structure [24] which is found when "zooming" inside the excitation block of a musical instrument (cfr. Fig. 1.1). The elements denoted by **L** and **NL** represent a linear and a non-linear block, respectively. More precisely, **L** contains a linear filter that accounts for the system dynamics in terms of lumped elements, while **NL** contains a non-linear multiple-input, multiple-output (*MIMO*) map $\boldsymbol{f}(\cdot)$. Both of these blocks take exciting actions $\boldsymbol{u}_2$ and the resonator variables $\boldsymbol{u}_1$ as inputs, and they are connected to each other in a feedback loop. This representation does not seem to be restrictive for acoustic models. Indeed, each of the following chapters provide an example of excitation block that can be schematized as in Fig. 2.11: the dynamic single reed discussed in Sec. 3.1, the physical glottal models of Sec. 4.1, and the non-linear hammer presented in Sec. 5.1.

Without any loss in generality, it is assumed in the following that the non-linear map $\boldsymbol{f}$ depends on a linear combination $\boldsymbol{x}$ of its inputs $(\boldsymbol{w}, \boldsymbol{u}_1, \boldsymbol{u}_2)$. Thus, the continuous-time system of Fig. 2.11 is described through the equations

$$\begin{cases} \dot{\boldsymbol{w}}(t) = \boldsymbol{A}\boldsymbol{w}(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{C}\boldsymbol{y}(t), \\ \boldsymbol{x}(t) = \boldsymbol{D}\boldsymbol{w}(t) + \boldsymbol{E}\boldsymbol{u}(t) + \boldsymbol{F}\boldsymbol{y}(t), \\ \boldsymbol{y}(t) = \boldsymbol{f}\left(\boldsymbol{x}(t)\right), \end{cases} \tag{2.35}$$

where the vector $\boldsymbol{u} = [\boldsymbol{u}_1, \boldsymbol{u}_2]^T$ collects all the external inputs to the exciter. Suppose that system (2.35) is discretized using either the Euler method or the bilinear transform (see Sec. 2.3.1 for a discussion of these methods). Then it is easily verified that the

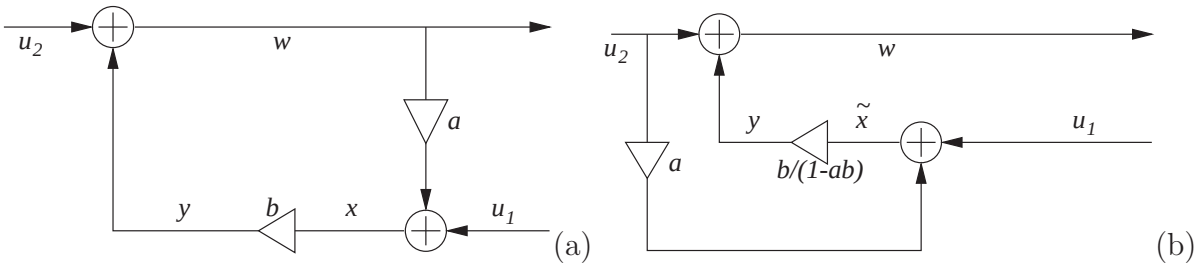

Figure 2.12: *A linear system; (a) delay-free path (figure based on Mitra [92]), and (b) an equivalent realization with no delay-free paths.*

discrete-time system takes the form

$$\begin{cases} \boldsymbol{w}(n) = \tilde{\boldsymbol{w}}(n) + \tilde{\boldsymbol{C}}\boldsymbol{y}(n), \\ \boldsymbol{x}(n) = \tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\boldsymbol{y}(n), \\ \boldsymbol{y}(n) = \boldsymbol{f}\left(\boldsymbol{x}(n)\right) = \boldsymbol{f}\left(\tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\boldsymbol{y}(n)\right), \end{cases} \tag{2.36}$$

where the vectors $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{x}}$ are computable vectors, i.e. they are linear combinations of $\boldsymbol{u}$ and past values of $\boldsymbol{w}$ and $\boldsymbol{y}$.

From Eqs. (2.36) it is clearly seen that there is a delay-free path connecting $\boldsymbol{y}$ to $\boldsymbol{x}$, and that the matrix $\boldsymbol{K}$ "weights" this path. Note that explicit expressions for the vectors $\tilde{\boldsymbol{w}}$, $\tilde{\boldsymbol{x}}$ and the matrices $\tilde{\boldsymbol{C}}$, $\boldsymbol{K}$ depend on whether the Euler method or the bilinear transform is used, but the overall system structure (2.36) is the same in both cases. Note also that the same structure is obtained when using other numerical methods: Sec. 3.2 shows that this occurs when discretizing a single reed model with Weighted Sample methods or higher-order Adams-Moulton methods.

A trivial example can help clarifying the computability problem in Eq. (2.36). Consider the system depicted in Fig. 2.12(a) (this is based on an example by Mitra [92]: see Sec. 6.1.3 and Fig. 6.5 in his book). Using the same notation as in Eq. (2.36) it is easily verified that the discrete-time system can be written as

$$\begin{cases} w(n) = \tilde{w}(n) + y(n), & \text{with} \quad \tilde{w} = u_2, \\ x(n) = \tilde{x}(n) + ay(n), & \text{with} \quad \tilde{x} = u_1 + au_2, \\ y(n) = f(x(n)) = bx(n), & \Rightarrow \quad y(n) = b[u_1(n) + au_2(n) + ay(n)]. \end{cases} \tag{2.37}$$

Note that in this example the function f is a linear map, i.e. a multiplication by b. It is clearly seen that a delay-free loop connects y to x, in particular the last of Eqs. (2.37) shows that y depends implicitly on itself.

However, the solution in this case is straightforward: the last of Eqs. (2.37) can be immediately inverted, yielding

$$y(n) = f(x(n)), \quad \longmapsto \quad y(n) = h(\tilde{x}(n)) = \frac{b}{1-ab}[u_1(n) + au_2(n)]. \tag{2.38}$$

The new map h relates y to the computable vector $\tilde{x}$. Therefore, an equivalent realization of the system is obtained as shown in Fig. 2.12(b). The key point in this example is that f is actually linear in this case, and this circumstance allows explicit inversion of the last of Eqs. (2.37).

However, it is obvious that such an analytical inversion is not always possible in the general non-linear case $\boldsymbol{y} = \boldsymbol{f}(\tilde{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{y})$. Many authors (e.g., [142, 62]) choose a rudimentary solution to restore computability in the numerical system: this amounts to inserting a fictitious delay element in the feedback loop, or in other words to assume that $\boldsymbol{y}(n) \approx \boldsymbol{y}(n-1)$. While this "trick" can be acceptable at significantly high sampling rates, it is known that at audio sampling rates ($F_s = 22.05, 44.1$ [kHz]) the insertion of such a delay element can deteriorate the accuracy and stability properties of the numerical system. Borin *et al.* [24] have shown that the insertion of a delay introduces instability in a numerical model of hammer-string interaction. Anderson and Spong [7] provide an analysis of instability due to the insertion of fictitious delays.

The K method provides a general technique that permits solution of delay-free paths in the general non-linear case, and to turn the implicit dependence $\boldsymbol{y} = \boldsymbol{f}(\tilde{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{y})$ into a new explicit dependence $\boldsymbol{y} = \boldsymbol{h}(\tilde{\boldsymbol{x}})$. This is achieved using the *implicit mapping theorem*. Define the function $\boldsymbol{g}$ as

$$\boldsymbol{g}(\tilde{\boldsymbol{x}}, \boldsymbol{y}) = \boldsymbol{f}(\tilde{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{y}) - \boldsymbol{y}, \tag{2.39}$$

and assume that there is a point $(\tilde{\boldsymbol{x}}_0, \boldsymbol{y}_0)$ such that $\boldsymbol{g}(\tilde{\boldsymbol{x}}_0, \boldsymbol{y}_0) = 0$. Moreover, assume that the following condition holds

$$\det[\boldsymbol{J}_{\boldsymbol{y}}(\boldsymbol{g})(\tilde{\boldsymbol{x}}_0, \boldsymbol{y}_0)] = \det\left[\frac{g_i}{y_j}(\tilde{\boldsymbol{x}}_0, \boldsymbol{y}_0)\right]_{i,j} \neq 0, \tag{2.40}$$

where the notation $\boldsymbol{J}_{\boldsymbol{y}}(\cdot)$ is used for denoting the Jacobian matrix with respect to the $\boldsymbol{y}$ variables. From the definition of $\boldsymbol{g}$ (2.39) it is seen that $\boldsymbol{J}_{\boldsymbol{y}}(\boldsymbol{g}) = \boldsymbol{J}_{\boldsymbol{x}}(\boldsymbol{f})\boldsymbol{K} - \boldsymbol{I}$. Therefore, the condition (2.40) implies that the matrix $[\boldsymbol{J}_{\boldsymbol{x}}(\boldsymbol{f})\boldsymbol{K} - \boldsymbol{I}]$ must be non-singular at the point $(\tilde{\boldsymbol{x}}_0, \boldsymbol{y}_0)$. If these conditions are fulfilled, then the implicit mapping theorem states that a function $\boldsymbol{h}(\tilde{\boldsymbol{x}})$ exists locally (i.e. for points $\tilde{\boldsymbol{x}}$ in a neighborhood of $\tilde{\boldsymbol{x}}_0$), with the properties

$$\boldsymbol{h}(\tilde{\boldsymbol{x}}_0) = \boldsymbol{y}_0 \quad \text{and} \quad \boldsymbol{g}(\tilde{\boldsymbol{x}}, \boldsymbol{h}(\tilde{\boldsymbol{x}})) = 0, \tag{2.41}$$

which is exactly the generalization of the mapping h that was found in Eq. (2.38). If the above conditions are fulfilled globally rather than in a neighborhood of $(\tilde{\boldsymbol{x}}_0, \boldsymbol{y}_0)$, then $\boldsymbol{h}$ is defined globally. At this point, the original discrete-time system (2.36) can be turned

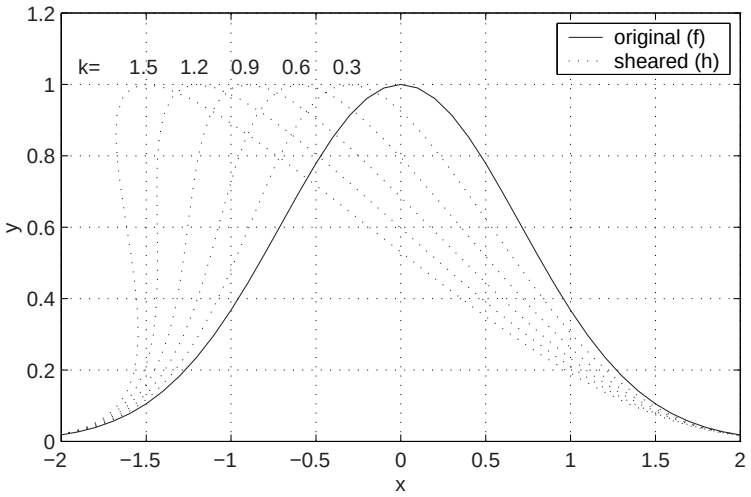

Figure 2.13: *Shear transformation of* $f(x) = e^{-x^2}$ *for various* k *values.*

into the new system

$$\begin{cases} \tilde{\boldsymbol{w}}(n) = \boldsymbol{L}_{\boldsymbol{w}}[u(n), u(n-1), w(n-1), y(n-1)], \\ \tilde{\boldsymbol{x}}(n) = \boldsymbol{L}_{\boldsymbol{x}}[u(n), u(n-1), w(n-1), y(n-1)], \\ \boldsymbol{y}(n) = \boldsymbol{h}\left(\tilde{\boldsymbol{x}}(n)\right), \end{cases} \tag{2.42}$$

where $\boldsymbol{L}_{\boldsymbol{w},\boldsymbol{x}}$ are two linear systems. Note that no delay-free paths are present in the equations.

A few geometrical considerations can help understanding the shape of the new function $\boldsymbol{h}$. Consider the coordinate transformation

$$\begin{bmatrix} \tilde{\boldsymbol{x}} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{K} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} = \mathcal{S} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}. \tag{2.43}$$

This defines a *shear* that leaves the $\boldsymbol{y}$ axes unchanged and distorts the $\boldsymbol{x}$ axis into the $\tilde{\boldsymbol{x}}$ axis. The plot of the function $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ "lives" in the $(\boldsymbol{x}, \boldsymbol{y})$ space. Then the plot of $\boldsymbol{y} = \boldsymbol{h}(\tilde{\boldsymbol{x}})$ is obtained by applying the coordinate transformation (2.43), and is therefore a sheared version of the former.

As an example, consider the following case where $f : \mathbb{R} \to \mathbb{R}$ is a Scalar Input-Scalar Output (*SISO*) function:

$$\begin{cases} x(n) = \tilde{x}(n) + ky(n), \\ y(n) = f\left(x(n)\right) = e^{-[x(n)^2]}. \end{cases} \tag{2.44}$$

The use of Eq. (2.40) results in the condition $f'(x) \neq 1/k$, which has a straightforward geometrical interpretation. The shear transformation $\mathcal{S}$ defined in Eq. (2.43) is such that the vector $[x, y]^T = [k, 1]^T$ (i.e. a point with tangent $1/k$) is transformed into the vector $[\tilde{x}, y]^T = [0, 1]^T$ (i.e. a point with vertical tangent). This explains why the derivative of f cannot equal $1/k$.

Figure 2.13 shows the original function $f(x)$, together with the sheared one $h(\tilde{x})$, for various k values. It can be seen that the horizontal coordinate is distorted when applying the shearing transformation. Moreover, note that for $k = 1.5$ the new function $h(\tilde{x})$ cannot be defined globally, because the condition $f'(x) \neq 1/k$ is not fulfilled globally in this case.

As a conclusion to this section, some remarks are provided about the implementation of the K method. Two main choices are available. The first one amounts to computing the new function $\boldsymbol{h}(\tilde{\boldsymbol{x}})$ in the range of interest, and then storing it in a look-up table. The main advantage of this approach is that $\boldsymbol{h}$ can be computed once forever off-line, thus the implementation of system (2.42) requires only linear operations and one look-up at each time step. However, there are at least two cases where this strategy exhibits drawbacks;

1. when $\boldsymbol{f}$ has high dimensions (for instance it is defined from $\mathbb{R}^3$ to $\mathbb{R}^2$), a table look-up also requires linear interpolation between neighboring points of the table and the efficiency might be deteriorated;

2. control parameters can affect the value of the matrix $\boldsymbol{K}$: therefore when such parameters are not constant over time (e.g. in a clarinet model with non-constant embouchure, see chapter 3), then $\boldsymbol{h}(\tilde{\boldsymbol{x}})$ needs to be computed on-line at the control rate.

In these cases, a different implementation strategy can be more efficient. This amounts to computing $\boldsymbol{h}(\tilde{\boldsymbol{x}})$ iteratively at the sampling rate, in two steps:

1. the linear part of system (2.42) is computed, and the current value of $\tilde{\boldsymbol{x}}$ is found;

2. the current value of $\boldsymbol{y}$ is found from Eq. (2.39) by imposing $\boldsymbol{g}(\tilde{\boldsymbol{x}}, \boldsymbol{y}) = 0$. The Newton-Raphson method (see e.g. Lambert [80]) can be used to iteratively solve this implicit non-linear equation.

Using the past value $\boldsymbol{y}(n-1)$ as the starting point in the Newton-Raphson iterations usually provides a fast convergence of the algorithm to the new value. This iterative procedure is used in the following chapters when implementing the K method.

Summary

This chapter has presented all the modeling approaches, the computational structures, and the numerical methods that are used in the rest of the thesis.

Section 2.1 has reviewed the principles of waveguide modeling of one-dimensional systems. First, it has been shown that wave propagation in cylindrical and conical sections

of acoustic bores can be in a first approximation assumed to be one-dimensional. Then the main components of 1-D waveguide structures have been introduced: sections, junctions, terminations, additional filtering elements for the modeling of dissipation and dispersion and for the fine tuning of the structures.

In Sec. 2.2, the lumped modeling approach has been introduced, stressing the analogies between electrical, mechanical and acoustical systems. In particular, series and parallel connections between lumped elements have been defined and analyzed. The role of non-linear elements in mechanical and acoustical circuits has been discussed.

Finally, Sec. 2.3 has addressed problems related to the discretization of continuous-time lumped systems. Various numerical techniques have been reviewed, focusing on their properties of stability, accuracy and efficiency. Then, the computability of the numerical structures has been discussed. It has been shown that delay-free paths are typically present in the structures, and that direct computation of these paths is not possible when non-linear elements are included in the system. A general method for solving this problem has been reviewed.

Chapter 3

Single reed models

Numerical simulations have been used for a long time by musical acousticians, for investigating experimentally the functioning of single reed wind instruments. More recently, the results from these studies have been exploited to develop numerical algorithms for sound synthesis. A widely accepted modeling approach divides the instrument into two main functional blocks, following the general scheme outlined in Sec. 1.2 (see in particular Fig. 1.1). The *resonator* is represented by the acoustic bore, and the *exciter* is the instrument reed. This chapter focuses on the excitation mechanism.

It is well understood that the single reed behaves as a pressure controlled, inward-striking valve that tends to close when an excess pressure p_m is provided by the player, and to open when the pressure p inside the mouthpiece exceeds p_m. Non-linear behaviors of the system are due to at least two different phenomena. First, the coupling between the reed displacement and the flow through the reed slit is non-linear. In a lumped modeling approach, this coupling is often treated using non-linear empirical equations derived from the Bernoulli law. Second, the interaction between the reed and the mouthpiece is also non-linear: the reed *bends* against the curved lay of the mouthpiece, and ultimately *beats* when complete closure occurs. Including in a lumped model this second non-linear mechanism is not trivial. In the literature, only beating is usually modeled, by imposing a maximum allowed displacement to the reed tip, while the effects of reed bending are neglected.

Section 3.1 reviews the lumped approach for single reed modeling. Section 3.2 addresses the problems related to discretization of a lumped model, and proposes an efficient and accurate technique for discretizing the continuous-time system. Section 3.3 discusses the problem of reed-lay interaction, and provides original results based on a distributed model of the reed-mouthpiece-lip system. These results are exploited in Sec. 3.4 to develop an improved lumped model, where the effects of reed-lay interaction are included in the lumped description. Table 3.1 summarizes the main variables and parameters used throughout the chapter.

This chapter is partially based on [8, 9, 13, 15, 146, 147].

quantity	symbol	unit
Reed length	L	[m]
Horizontal position	$x \in [0, L]$	[m]
Lay profile	$y_{lay}(x)$	[m]
Reed displacement	$y(x,t)$	[m]
Reed tip displ.	$y_L(t)$	[m]
Tip rest position	y_0	[m]
Max. tip displacement	$y_m = y_{lay}(L)$	[m]
Reed tip opening	$h(t) = y_m - y_L(t)$	[m]
Mouth pressure	p_m	[Pa]
Mouthpiece pressure	$p(t)$	[Pa]
Pressure drop	$\Delta p(t) = p_m - p(t)$	[Pa]
Mouthpiece flow	$u(t)$	[m^3/s]
Flow through the slit	u_f	[m^3/s]
Bore cross section	S	[m^2]
Bore wave impedance	$Z_0 = \rho_{air} c/S$	[Kg/m^4s]
Bore length	L_{bore}	[m]
Bell cutoff freq.	ω_{bell}	[rad/s]
Press. wave from the bore	$p^-(t)$	[Pa]
Press. wave to the bore	$p^+(t)$	[Pa]

Table 3.1: *Symbols used throughout the chapter.*

3.1 Lumped modeling

Section 3.1.1 presents a widely used lumped model. The reed is treated using three lumped mechanical elements. These are the reed stiffness, the reed mass, and a damping element that provides an approximate description of air friction, internal losses and lip damping. Section 3.1.2 derives a simplified version of this model, where only the reed stiffness is used (i.e. it is assumed that the reed moves in phase with the pressure signal inside the mouthpiece). This latter model is often referred to as the *quasi-static* approximation.

3.1.1 Reed-lip-mouthpiece system

The reed dimensions are small with respect to typical wavelengths in the resonator, thus pressure can be thought of as constant along the reed internal surface; under normal playing conditions, the first mode of the reed-mouthpiece-lip system is well above the main frequency component of the pressure signal that drives it; oscillations occur mainly in the vertical direction, and as a first approximation a single degree of freedom (i.e. the reed tip vertical displacement y_L) can be assumed. Figure 3.1 depicts such a schematized representation.

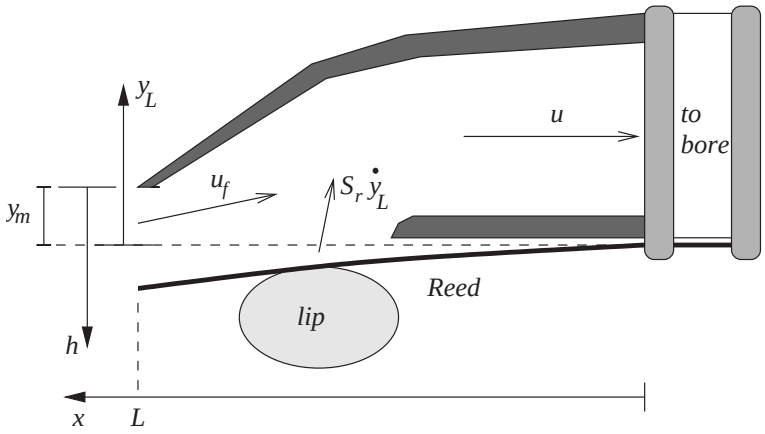

Figure 3.1: *Schematized representation of the reed-mouthpiece-lip system.*

Based on these considerations, many authors [16, 156, 154, 120] have approximated the reed as a lumped second-order mechanical oscillator, driven by the pressure drop Δp between mouth and mouthpiece:

$$m\ddot{y}_L(t) + r\dot{y}_L(t) + k[y_L(t) - y_0] = S_d \Delta p(t), \tag{3.1}$$

where m, r, k are the reed mass, damping, and spring constant, respectively. The parameter S_d is an effective driving surface on which the pressure Δp acts. In the Laplace domain, Eq. (3.1) can be rewritten as

$$Y_L(s) - y_0 = H_r(s)\Delta P(s), \quad \text{with} \quad H_r(s) = \frac{1}{\mu}\frac{1}{s^2 + gs + \omega_0^2}. \tag{3.2}$$

Therefore, H_r is the transfer function between Δp and the reed relative displacement. The parameter $\mu = m/S_d$ is the effective mass/area ratio, $g = r/m$ is the damping coefficient and $\omega_0 = \sqrt{k/m}$ is the resonance of the oscillator.

The phenomenon of reed beating (i.e. complete closure of the reed) is usually incorporated in the lumped model in a non-physical way, by imposing a "stop" when the reed tip reaches its maximum allowed displacement y_m. Equation (3.1) is thus turned into

$$\begin{cases} m\ddot{y}_L(t) + r\dot{y}_L(t) + k(y_L(t) - y_0) = S_d \Delta p(t), & \text{for} \quad y_L < y_m, \\ y_L(t) = y_m \quad \text{and} \quad \dot{y}_L(t) = 0, & \text{for} \quad y_L \geq y_m. \end{cases} \tag{3.3}$$

The value y_m is the maximum height of the lay and represents the maximum allowed displacement of the reed tip. Note that the system (3.3) models reed beating but does not take into account the effects due to reed bending: the reed behaves as a linear-oscillator with constant parameters in the entire allowed range of y_L. Note also that the dependence of the reed-mouthpiece system on the lip embouchure is not taken into account in this lumped representation.

Once the mechanical part has been modeled, the relation between the reed opening and the airflow through the slit u_f has to be found. As a first approximation, the pressure drop Δp can be assumed to obey the equation

$$\Delta p(t) = f(u_f(t), h(t)) = A^{-\alpha}\mathrm{sgn}[u_f(t)]\frac{|u_f(t)|^\alpha}{h(t)^2}. \tag{3.4}$$

A more accurate description should take into account the inertia of the air in the reed channel by adding a term of the form $M_{slit}du/dt$ to right-hand side of Eq. (3.4). However, many authors (see e.g. Nederveen [95], or Thompson [135]) have assumed the inertia to be small, and this additional term is usually neglected.

The exponent α in Eq. (3.4) merits some discussion. If the value $\alpha = 2$ is used, then Eq. (3.4) is the Bernoulli equation, that holds for the flow behavior of an ideal fluid in the static regime. It is assumed that the equation remains approximately valid also in the dynamic regime. Through experiments on real instruments, Backus [16] found empirically a non-linear equation of the form (3.4) However, Backus' experiments provided a value for α different from the theoretical one, namely $\alpha = 3/2$. Backus explained this discrepancy with the particular wedge-shape of the slit. Recent research has questioned the validity of Backus empirical equation. Hirschberg *et al.* [70] developed a flow model that uses the standard Bernoulli equation (with $\alpha = 2$), and experiments by Gilbert [63] did not confirm Backus' findings.

Equations (3.3) and (3.4) relate quantities at the reed slit. A third equation relates the flow u_f at the slit to the total flow u inside the instrument:

$$u(t) = u_f(t) + u_r(t), \quad \text{with} \quad u_r(t) = S_r\dot{y}_L(t). \tag{3.5}$$

This equation states that the total flow inside the instrument is affected by an additional component $u_r(t)$, induced by the reed motion and proportional to the reed tip velocity. The quantity S_r is the effective flow surface of the reed. As pointed out by Gazengel [61], S_r is not necessarily equal to S_d.

Equations (3.3), (3.4) and (3.5) provide a complete lumped description of a clarinet single reed. Figure 3.2 visualizes the model. Worman [156] was the first to formalize a complete non-linear theory using such a lumped description. It has been shown that this description is able to account for many distinctive features of the system, such as the existence of a threshold blowing pressure and the occurrence of self-sustained oscillations, as well as mode locking in a slightly inharmonic acoustical bore [46]. Wilson and Beavers [154] developed a theoretical analysis on operating modes of a clarinet by coupling a lumped reed model with a constant-area resonator, and showed qualitative agreement between theory and experiment. Schumacher [120] coupled an analogous model to a realistic clarinet bore, described by its reflection function, and was able to compute transient and steady-state oscillations of the mouthpiece pressure via time-domain simulations.

3.1.2 The quasi-static approximation

Suppose that a constant pressure drop Δp is applied and that the consequent reed tip displacement y_L is measured. Many authors (e.g. [130, 135]) introduce a new quantity,

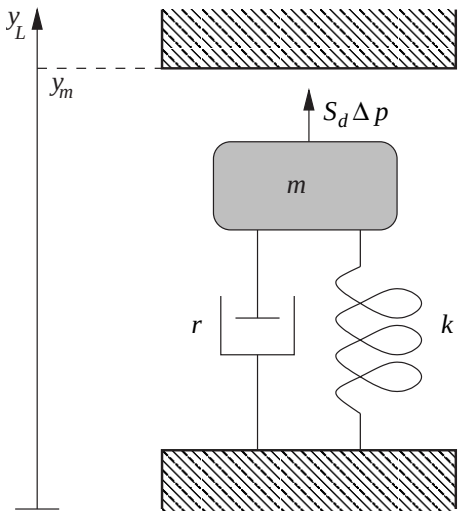

Figure 3.2: *Lumped model of a single reed.*

called reed *stiffness per unit area* K_a, to indicate the ratio between Δp and the relative displacement $(y_L - y_0)$:

$$\Delta p = K_a(y_L - y_0). \tag{3.6}$$

The inverse of K_a is referred to as reed *compliance* by Nederveen [95].

Equation (3.6) is a static relation. However, it can be assumed to be approximately valid when the signal pressure Δp is varying "slowly", precisely when Δp is a sufficiently smooth (bandlimited) signal and the reed mechanical resonance ω_0 is high compared to the fundamental frequency of Δp. This assumption seems reasonable in a clarinet or in other single reed instruments. In fact, in a single reed instrument the fundamental regime of oscillation is governed by the first resonance frequency of the pipe, and typical values for ω_0 are well above this pipe resonance.

Using Eq. (3.6), the reed opening h is computed as

$$h(t) = y_m - y_0 - \frac{\Delta p(t)}{K_a} = h_0 - \frac{\Delta p(t)}{K_a},$$

where $h_0 = y_m - y_0$ is the rest opening of the reed tip. From this equation, the condition for reed closure $h = 0$ turns out to be $\Delta p = h_0 K_a$. Using these relations, Eq. (3.4) can be turned into a new non-linear equation that relates the flow u_f through the slit and the pressure drop Δp. Straightforward calculations yield

$$u_f(t) = \begin{cases} A \cdot \mathrm{sgn}[\Delta p(t)] \cdot |\Delta p|^{1/\alpha} \left(h_0 - \dfrac{\Delta p(t)}{K_a}\right)^{2/\alpha} |\Delta p|^{1/\alpha}, & \text{for} \quad \Delta p < h_0 K_a, \\ 0, & \text{for} \quad \Delta p \geq h_0 K_a. \end{cases} \tag{3.7}$$

This relation is sometimes referred to as *quasi-static approximation* (see e.g. [69]), since it is derived by assuming that the static relation (3.6) also holds when the driving

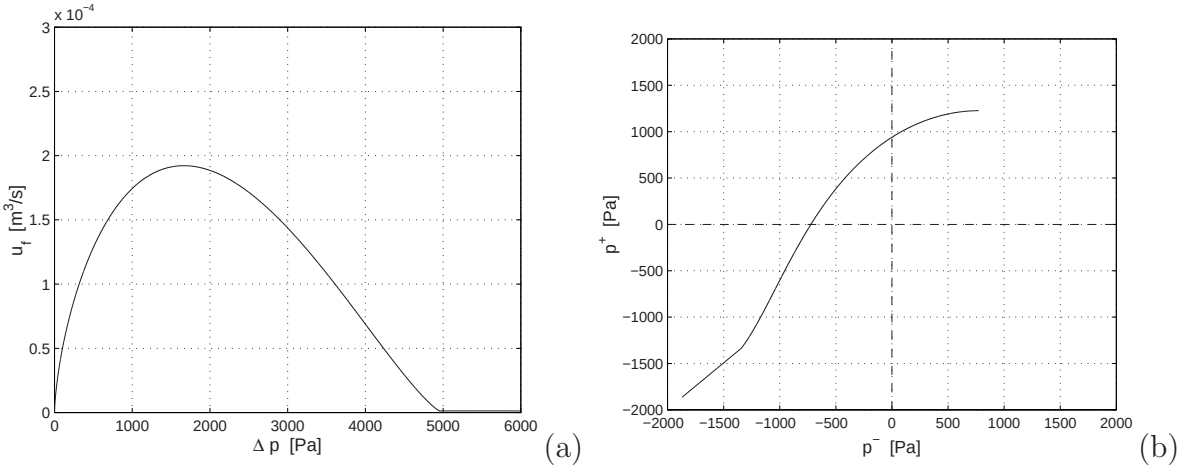

Figure 3.3: *Quasi-static approximation of a single reed; (a) u_f versus Δp and (b) rotated mapping $p^+ = R_{nl}(p^-)$.*

pressure Δp is a slowly varying signal. Figure 3.3(a) shows the plot of the function in Eq. (3.7).

Using the analogies described in Sec. 2.2, it can be said that the quasi-static single reed defined here is a non-linear resistor, i.e. a non-linear element which is described through an algebraic relation between the two Kirchhoff variables u_f and Δp. If wave variables $p^{\pm}$ are introduced, the non-linearity is turned in a new one. Namely, p^+ depends on p^- through a non-linear reflection function R_{nl}, i.e. $p^+ = R_{nl}(p^-)$. This is depicted in Fig. 3.3(b).

Suppose now that the quasi-static reed is coupled to an ideal bore of length L_{bore}, open at the end opposite to the reed. Then, the pressure wave p^- traveling from the bore to the reed is related to p^+ through the relation $p^-(t) = -p^+(t - T_0/2)$, where $T_0 = 4L_{bore}/c$ is the fundamental period of the bore. The negative sign comes from the ideal reflection at the open end (see Sec. 2.1.3 and Fig. 2.5). Then, the system is described by the equations

$$\begin{cases} p^-(t) = -p^+(t - T_0/2), \\ p^+(t) = R_{nl}(p^-(t)). \end{cases} \tag{3.8}$$

Therefore successive values of $p^{\pm}$ (equivalently, successive values of $u_f, \Delta p$) can be found by iterating the non-linear map R_{nl} (see e.g. Mc Intyre *et al.* [90] and Maganza *et al.* [85]). For low values of the mouth pressure p_m, the map has a stable fixed point, i.e. no oscillations occur. When p_m becomes larger than a certian threshold pressure p_t, the stable fixed point bifurcates in a stable orbit. It can be shown that the threshold pressure has the value $p_t = h_0 K_a/3$. For higher values of p_m, Maganza *et al.* [85] have proved that the system undergoes a Feigenbaum-type route to chaos.

In conclusion, the quasi-static model is able, despite its simplicity, to capture the basic non-linear mechanisms of self-sustained oscillations in a single reed instrument. Due to

its compactness and low number of parameters, this model has been also used for sound synthesis purposes. As an example, Smith [127]) has developed a clarinet model where the resonator is treated using waveguide structures. The waveguide termination at the reed end is described through a so-called *signal dependent reflection coefficient*, derived from the non-linear function R_{nl}. Smith shows that an extremely efficient numerical implementation can be obtained by precomputing the non-linearity: the computational costs of reed simulation reduces to only two subtractions, one multiply and one table look-up per sample.

3.2 An efficient and accurate numerical scheme

This section addresses the issue of discretization of the lumped model reviewed in Sec. 3.1.1. Two main problems have to be overcome. First, the coupling in the equations typically generates a delay-free loop in the computation; due to the presence of a non-linear equation, solving this loop is not trivial. Second, a numerical technique has to be found that preserves with reasonable accuracy the main properties of the physical system. It is shown in the following that the delay-free loop in the computational scheme can be solved using the *K method* (already described in Sec. 2.3.2). Then, four different techniques are used for for discretizing the mechanical differential equations. The so obtained "digital reeds" are coupled to a waveguide model of a cylindrical bore, and the resulting models are compared both in the frequency and the time domains.

These issues are often neglected in the literature of digital simulation of single reed instruments: the differential equations are usually discretized with simple numerical methods such as the Euler method or the Impulse Invariant method, and delay-free paths in the computation are dealt with by introducing fictitious delay elements (see e.g. Gazengel *et al.* [62]). One reason for this lack of interest is that the choice of the discretization method is usually considered to be non-critical at high sampling rates. However, as shown in the following, different methods produce numerical models whose behavior differ noticeably, even at a sampling rate of several tens of kHz (e.g., 44.1 [kHz]).

3.2.1 Numerical methods

The lumped reed model described in Sec. 3.1 can be restated in vector formulation as

$$\begin{cases} \dot{\boldsymbol{w}}(t) &= \boldsymbol{A}\boldsymbol{w}(t) + \boldsymbol{B}\boldsymbol{u}(t) + \boldsymbol{C}\Delta p(t), \\ \boldsymbol{x}(t) &= \boldsymbol{D}\boldsymbol{w}(t) + \boldsymbol{E}\boldsymbol{u}(t) + \boldsymbol{F}\Delta p(t), \\ \Delta p(t) &= f\left(\boldsymbol{x}(t)\right), \end{cases} \tag{3.9}$$

where the variables are given by

$$\boldsymbol{w} = \begin{bmatrix} h \\ \dot{h} \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} h_0 \\ p_m \\ p^- \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} u_f \\ h \end{bmatrix},$$

and where the matrices are

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -g \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ \omega_0^2 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 0 \\ -1/\mu \end{bmatrix},$$

$$\boldsymbol{D} = \begin{bmatrix} 0 & -S_r \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{E} = \begin{bmatrix} 0 & 1/Z_0 & -2/Z_0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} -1/Z_0 \\ 0 \end{bmatrix}.$$

The beating condition in Eq. (3.3) is rewritten as

$$\boldsymbol{w} = \boldsymbol{0}, \qquad \text{for} \quad h \leq 0.$$

When the first equation in system (3.9) is discretized, the structure of the resulting difference equation is found to be

$$\boldsymbol{w}(n) = \tilde{\boldsymbol{w}}(n) + \bar{\boldsymbol{C}}\Delta p(n). \tag{3.10}$$

Here the vector $\tilde{\boldsymbol{w}}(n)$ is a linear combination of all the terms that are computable at time n (namely $\boldsymbol{u}(n)$ and past values of $\boldsymbol{w}$, $\boldsymbol{u}$ and Δp) while the vector $\bar{\boldsymbol{C}}$ weights the dependence of $\boldsymbol{w}$ on $\Delta p(n)$. Explicit expressions for both $\tilde{\boldsymbol{w}}(n)$ and $\bar{\boldsymbol{C}}$ depend on the chosen method. The remaining equations in system (3.9) can thus be written as

$$\begin{cases} \boldsymbol{x}(n) &= \tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\Delta p(n), \\ \Delta p(n) &= f\left(\tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\Delta p(n)\right), \end{cases} \tag{3.11}$$

where $\boldsymbol{K} = (\boldsymbol{D}\bar{\boldsymbol{C}} + \boldsymbol{F})$ weights the delay-free loop connecting Δp to $\boldsymbol{x}$, while the vector $\tilde{\boldsymbol{x}}(n) = \boldsymbol{D}\tilde{\boldsymbol{w}}(n) + \boldsymbol{E}\boldsymbol{u}(n)$ has no instantaneous dependence on $\Delta p(n)$ and is therefore computable at each step. It is easily seen that this numerical system resembles the general structure described in Sec. 2.3.2 and depicted in Fig. 2.11. Therefore, the K method can be used in order to compute the delay-free path between $\boldsymbol{x}(n)$ and $\Delta p(n)$.

$$\Delta p(n) = f\left(\tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\Delta p(n)\right) \quad \overset{\text{K method}}{\longmapsto} \quad \Delta p(n) = h\left(\tilde{\boldsymbol{x}}(n)\right).$$

At each time step the vector $\tilde{\boldsymbol{x}}(n)$ is computed first, then $\Delta p(n)$ is obtained through the new non-linear relation h.

The K method provides a robust and general means to compute the difference equations (3.11) accurately. Given such a method, different discretization techniques for system (3.9) can be compared. In the literature of discrete-time simulations of single reed instruments, the Euler method and the Impulse Invariant method are the most commonly used numerical techniques. Few authors have tried to use different methods. Gazengel *et al.* [62] discuss the use of a fourth order Runge-Kutta solver. This method, although accurate for numerical simulations at high sampling rates, turns out to be unstable at low F_s. Moreover, it has high computational costs, since four evaluations of the non-linear function $f(\boldsymbol{x}(n))$ are needed at each time step. Van Walstijn [145, 147] uses a hybrid backward-centered scheme in which the derivatives $\dot{y}_L$ and $\ddot{y}_L$ in Eq. (3.52) are

approximated using the mixed backward-centered scheme described in Sec. 2.3.1. More precisely, the first derivative $\dot{y}_L$ is treated using the backward Euler approximation of Eq. (2.28), while $\ddot{y}_L$ is discretized with the centered scheme of Eq. (2.30). One advantage of this approach is that the vectors $\bar{\boldsymbol{C}}$ and $\boldsymbol{K}$ in Eqs. (3.10) and (3.11) are both zero, therefore no delay-free paths are created in the discrete-time equations. However, at each time step n the Newton-Raphson method is used for computing iteratively the flow $u(n)$, and nine iterations are typically required.

In this section the following numerical methods are used:

- *1- and 2-step Adams-Moulton Methods* (AM1,2 from now on). These are linear multistep methods, whose stability and accuracy properties are known from the numerical analysis literature [80]. As discussed in Sec. 2.3.1, the AM1 method is the bilinear transform.

- *1- and 2-step Weighted Sample Methods* (WS1,2 from now on). These have been recently introduced by Wan and Schneider [151]. They are designed for a generic linear system and are based on a polynomial interpolation of the system input.

The use of higher-order methods is not discussed here, for two main reasons: increasing the order causes *(1)* the stability properties to deteriorate, and *(2)* the computational costs to increase.

As Schneider *et al.* [118] have pointed out, AM methods can be easily seen as s-to-z mappings in the complex plane:

$$\begin{aligned} &\text{(AM1)} \quad s = 2F_s \frac{1 - z^{-1}}{1 + z^{-1}}, \\ &\text{(AM2)} \quad s = 12F_s \frac{1 - z^{-1}}{5 + 8z^{-1} - z^{-2}}. \end{aligned} \tag{3.12}$$

The discrete system is obtained from the Laplace-transformed continuous system (3.9) by substituting each occurrence of s with the corresponding mappings (3.12). Therefore, the first equation in system (3.9) is turned by the AM1 and the AM2 methods into a second-order and a fourth-order difference equation, respectively. Equivalently, the filter H_r in Eq. (3.2) is turned by the two methods into a second-order and a fourth-order digital IIR filter, respectively.

Wan and Schneider [151] have shown that the k-step WS method turn the first equation in system (3.9) into the difference equation

$$\boldsymbol{w}(n) = \boldsymbol{\Phi}(T_s)\boldsymbol{w}(n-1) + \boldsymbol{\mathcal{W}}_{\boldsymbol{u}}\boldsymbol{S}_{\boldsymbol{u}}{}^{(k)} \begin{bmatrix} \boldsymbol{u}(n) \\ \vdots \\ \boldsymbol{u}(n-k) \end{bmatrix} + \boldsymbol{\mathcal{W}}_{\Delta p}\boldsymbol{S}_{\Delta p}^{(k)} \begin{bmatrix} \Delta p(n) \\ \vdots \\ \Delta p(n-k) \end{bmatrix}, \tag{3.13}$$

where $k = 1, 2$. Therefore, the first equation in system (3.9) is turned by the WS1 and the WS2 methods into a second-order and a third-order difference equation, respectively.

Equivalently, the filter H_r in Eq. (3.2) is turned by the two methods into a second-order and a third-order digital IIR filter, respectively.

Details about the computation of the matrices can be found in [151, 8], here suffice it to mention that $\boldsymbol{\Phi}(t) = \exp(\boldsymbol{A}t)$ and that computation of $\boldsymbol{\mathcal{W}}\boldsymbol{u}_{,\Delta p}$ involves calculation of the $k+1$ integrals $\int_0^T \boldsymbol{\Phi}(T_s - t) \cdot t^l dt$ $(l = 1 \ldots k+1)$. Therefore, computing the coefficients of system (3.13) requires computation of transcendental functions. If system (3.9) is time-invariant, then computation of matrices $\boldsymbol{\Phi}(T_s)$ and $\boldsymbol{\mathcal{W}}$ can be performed off-line, while these matrices need to be updated at control rate when time-varying control parameters are used. In this latter case the WS methods have higher computational costs than the AM methods, and this is a potential drawback for real-time applications. However, for the low order methods ($k = 1, 2$) used here, only a small number of coefficients needs to be updated. Moreover, Wan and Schneider [151] show that the computational costs can be lowered using *ad hoc* techniques (e.g. the columns of $\boldsymbol{\mathcal{W}}$ can be computed iteratively).

It has already been shown in Sec. 2.3.1 that the bilinear transform AM1 preserves stability at any sampling rate. The AM2 method has worse stability properties, since its region of absolute stability is a finite subset of the left-half s-plane. This means that stability is preserved only at high sampling rates, so that the poles of the continuous system lie inside the region of absolute stability (see for instance [80].

Wan and Schneider have proven that, if p_c is a pole of the of the continuous-time equation, then the corresponding discrete-time pole p_d obtained with the k-step WS method is given by $p_d = e^{p_c T_s}$. Therefore, stability is preserved for any choice of T_s. Note that the same relation between discrete and continuous poles was found in Sec. 2.3.1 for the Impulse Invariant method. Indeed, it can be verified that the WS method with $k = 0$ is completely equivalent to the Impulse Invariant method.

Concerning accuracy, it is a general result that the k-step AM method has order $k+1$. This means that the method provides a global truncation error in time which has order T_s^{k+1}. For WS methods, Wan and Schneider have given experimental results showing that a k-step method has order $k + 1$, the same as the corresponding AM method.

3.2.2 Frequency-domain analysis

The four numerical methods described in the last section are used to discretize the single reed system (3.9). In this way, four different digital reeds are obtained. This section compares these digital reeds through analysis n the frequency domain. In this way, it is studied how the most significant physical parameters are mapped into the digital domain by each numerical method.

Consider the transfer function $H_r(s)$ in Eq. (3.2) and the corresponding frequency response $H_r(j\omega)$. A first comparison between the numerical methods amounts to analyze how they preserve this frequency response in the discrete-time domain. The study is performed for various sampling rates, and typical audio rates $F_s = 22.05$ [kHz] and $F_s = 44.1$ [kHz] are taken as reference values. Following ideas developed by Gazengel *et al.* [62], three physically meaningful parameters of the system are analyzed: the resonance frequency ω_0, the low-frequency magnitude response $|H_r(0)| = 1/\mu\omega_0^2$, and the damping

coefficient g.

Typical values for ω_0 lie in the high frequency region, and this parameter is therefore considered to be non-critical in helping self-sustained oscillations. Indeed, self-sustained oscillations occur even when there is no resonance at all, as in the quasi-static approximation (see Sec. 3.1.2). However, as Thompson [135] has pointed out, the reed resonance has a role in adjusting pitch, loudness and tone color, and in helping transitions to high regimes of oscillation, such as the clarion register or the reed regime ("squeaks"). The low-frequency reed response $|H_r(0)|$ is also an important parameter, since the fundamental frequency of the oscillation always lies in this region. Note that it can also be written as $|H_r(0)| = S_d/k$. Concerning the damping coefficient, the relation $g = \omega_0/q_0$ holds for a the oscillator (3.1), where $q_0 = \omega_0/(\omega_1 - \omega_2)$ is the quality factor and $\omega_{1,2}$ are the 3 dB cut-off frequencies. Therefore $g = \omega_1 - \omega_2$ for the continuous oscillator $H_r(j\omega)$.

When using AM and WS methods, $H_r(s)$ is turned into a digital filter which is not a necessarily a second-order oscillator. Therefore, the coefficients ω_0, μ, g cannot be deduced from the discrete transfer functions. Instead they are extrapolated from the shape of the frequency responses. In particular, following Gazengel *et al.* [62] the digital damping coefficient g_d is defined here as $g_d = \omega_{d1} - \omega_{d2}$, where $\omega_{d1,d2}$ are the 3 dB cut-off frequencies for the discrete frequency response.

Using AM methods, the digital transfer functions $H_{AM1}(z)$ and $H_{AM2}(z)$ are obtained by substitution of the corresponding s-to-z mapping (3.12) in $H_r(s)$. The corresponding frequency responses are given by evaluation at $z = \exp(j\omega_d T_s)$. Figure 3.4 shows the discrete responses H_{AM1} and H_{AM2} in the cases $F_s = 22.05$ [kHz] and $F_s = 44.1$ [kHz], for comparison with the original one. Responses obtained with the Euler method are also plotted as a term of comparison.

The Euler method is easily seen to provide poor accuracy even at $F_s = 44.1$ [kHz]. In particular, a noticeable numerical dissipation is introduced, so that the resonance is strongly attenuated. Results for the AM methods are in good agreement with theoretical predictions. Both the magnitude and the phase responses of H_{AM1} exhibit frequency warping (see the discussion in Sec. 2.3.1). At $F_s = 22.05$ [kHz], the resonance ω_0 has shifted from 23250 [rad/s] to 21300 [rad/s] (i.e. from 3700 [Hz] to 3390 [Hz]). The AM2 method provides different results: there is no significant warping, but the magnitude of the resonance is amplified. The amplification is small at $F_s = 44.1$ [kHz], but becomes unacceptable at $F_s = 22.05$ [kHz] (the peak magnitude is $4.7 \cdot 10^{-5}$ [m/Pa]). This phenomenon is a direct consequence of the stability properties of the method: it can be seen that, with the values used here, the method becomes unstable at $F_s \simeq 19$ [kHz]. This explains the strong amplification and the phase distortion exhibited by H_{AM2} at $F_s = 22.05$ [kHz].

Both methods preserve the low-frequency response. More specifically, the equalities $H_{AM1}(0) = H_{AM2}(0) = H_r(0)$ hold. Finally, qualitative analysis shows that both the 1- and 2-step AM methods lead to a digital damping $g_d = \omega_{d1} - \omega_{d2}$ which is smaller than the physical one, and decreases with decreasing F_s. For H_{AM1} this is a consequence of the frequency warping effect, which causes the resonance bandwidth to reduce and the quality factor to increase consequently. For H_{AM2} this is due to the resonance amplification rather

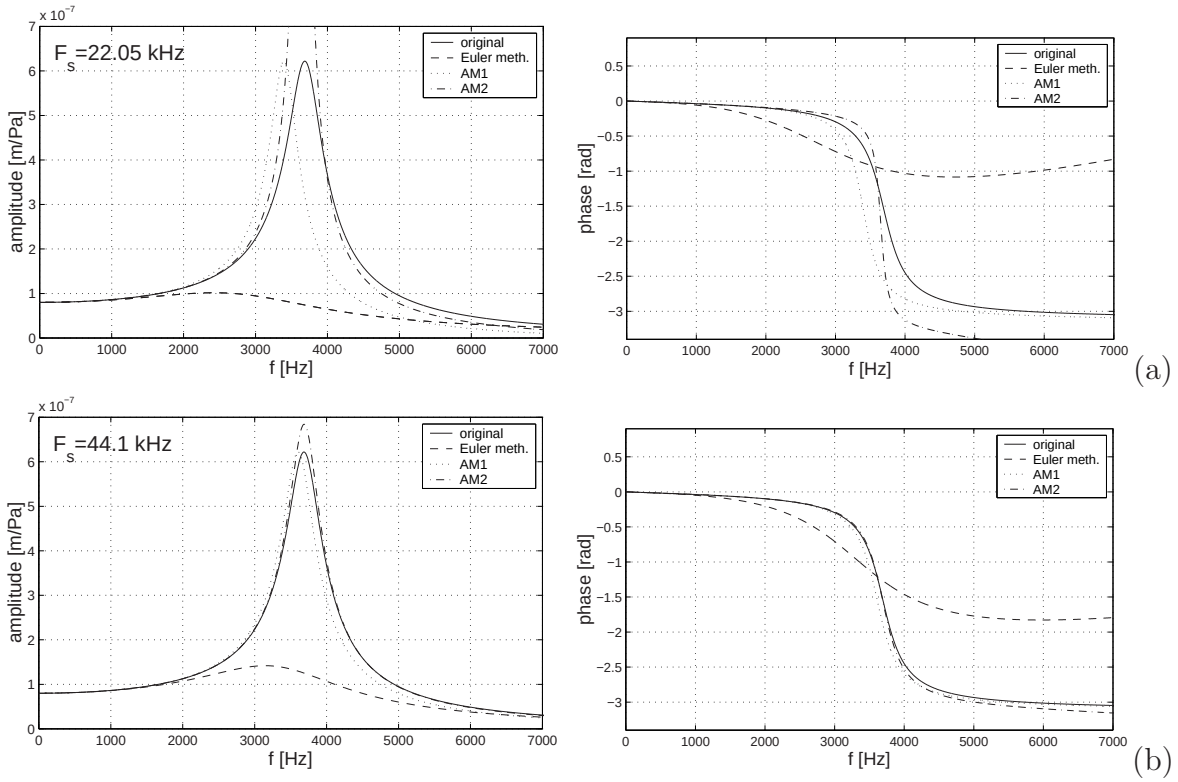

Figure 3.4: *Digital responses H_{AM1} and H_{AM2} in the cases (a) $F_s = 22.05$ [kHz] and (b) $F_s = 44.1$ [kHz]. Physical parameters are $\omega_0 = 2\pi \cdot 3700$ [rad/s], $g = 3000$ [rad/s], $\mu = 0.0231$ [Kg/m^2].*

than to warping effects.

Analogous analysis is performed for the WS methods. These methods do not define s-to-z mappings, therefore the transfer functions $H_{WS1}(z)$ and $H_{WS2}(z)$ are not obtained by substitution. Instead, they are computed directly from the general equation (3.13). Results are summarized in Fig. 3.5. Again, the responses obtained with the Euler method are plotted as a term of comparison.

The discrete responses H_{WS1} and H_{WS2} show excellent agreement with H_r, even at low sampling rates. Both methods preserve the resonance ω_0 without introducing warping. The low-frequency response is preserved as well. Numerical dissipation is introduced, which is more noticeable for the 1-step method. This effect can be noticed by observing that the amplitude responses $|H_{WS1}|$ and $|H_{WS2}|$ lie below the $|H_r|$. Due to this dissipation the digital damping coefficient g_d is larger than the physical one and increases with decreasing F_s, for both $H_{WS1}(z)$ and $H_{WS2}(z)$. The phase responses are well preserved by both methods. Summarizing, the frequency analysis developed in this section shows that the WS methods better approximate the reed frequency response than AM methods. It would appear that the WS methods are preferable. However, this conjecture is not confirmed from the time-domain analysis developed in the next section.

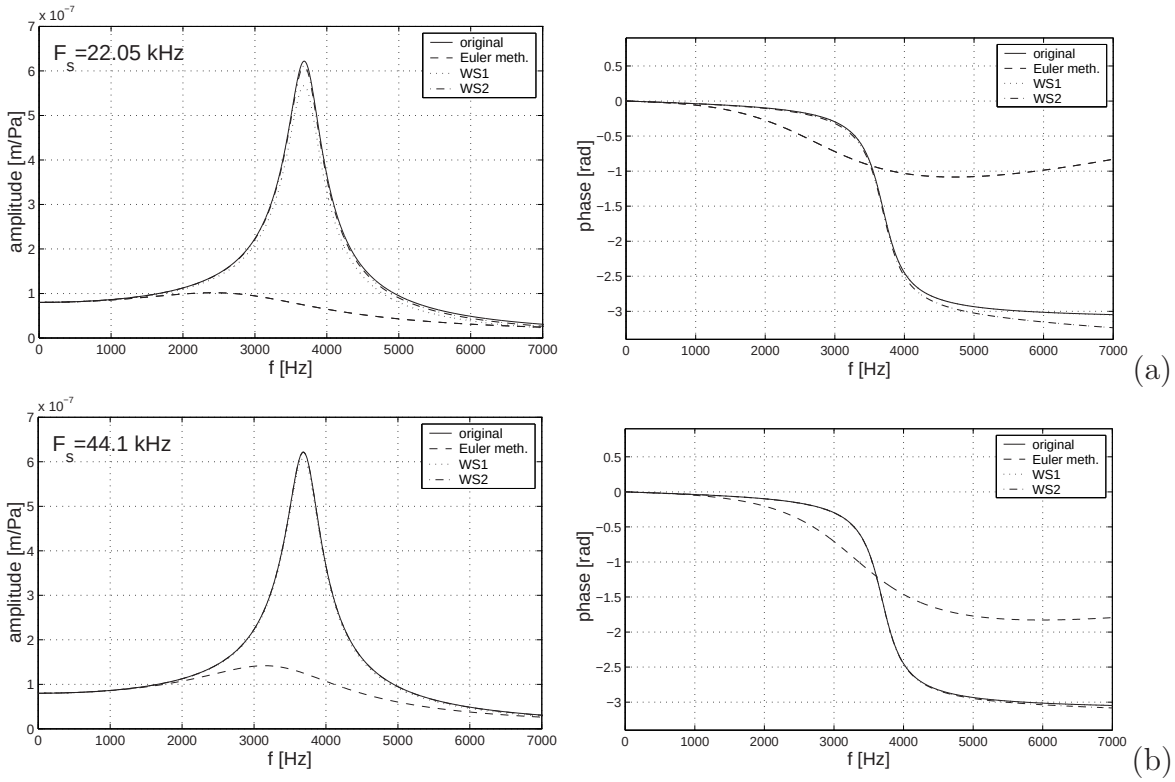

Figure 3.5: *Digital responses H_{WS1} and H_{WS2} in the cases (a) $F_s = 22.05$ [kHz] and (b) $F_s = 44.1$ [kHz]. Physical parameters are $\omega_0 = 2\pi 3700$ [rad/s], $g = 3000$ [rad/s], $\mu = 0.0231$ [Kg/m^2].*

3.2.3 Time-domain analysis

In order to further investigate the properties of the numerical methods, time-domain simulations are studied in which each of the four digital reed is connected to the same resonator. This is a basic waveguide model of the clarinet bore (see chapter 2 for details on one-dimensional waveguide structures). Propagation in the bore is simulated with two delay lines; the length m_{bore} of each line is chosen in such a way that $m_{bore}cT_s = L_{bore}$. The bell is modeled by inserting a low-pass filter R_d in the waveguide loop, and this filter is designed with standard techniques (see e.g. Mitra [92]): first the analog filter is designed using a fourth order Butterworth realization. Then a digital equivalent $R_d(z)$ is obtained with the bilinear transform:

$$R_d(z) = R_c\left(2F_s\frac{1-z^{-1}}{1+z^{-1}}\right).$$

Summarizing, the digital bore model takes the incoming pressure wave p^+ from the exciter, and returns to it an outgoing pressure wave p^- given by

$$P^-(z) = -R_d(z)\ z^{-2m_L}\ P^+(z). \tag{3.14}$$

	p_t [Pa]				
F_s [kHz]	quasi-static	$AM1$	$WS1$	$AM2$	$WS2$
20	1664	1816	1761	...	3346
25	1664	1808	1774	...	2842
30	1664	1807	1784	...	2554
35	1664	1807	1790	...	2365
40	1664	1807	1795	...	2233
45	1664	1804	1796	...	2136
50	1664	1804	1797	3781	2063
55	1664	1805	1798	3516	2008
60	1664	1805	1799	3278	1960
65	1664	1806	1799	3148	1932
70	1664	1803	1800	3026	1906
75	1664	1804	1800	2908	1881
80	1664	1804	1801	2841	1865
85	1664	1805	1801	2887	1848
90	1664	1803	1802	2737	1832
95	1664	1803	1802	2692	1816
100	1664	1803	1802	2643	1802

Table 3.2: *Measured threshold pressures from time-domain simulations. Values are rounded to the nearest integer*

A first study on time-domain simulations analyzes the threshold pressure p_t, i.e. the value for mouth pressure above which stable oscillations take place. A rough estimate for the threshold pressure, $p_t \simeq h_0 K_a/3$, has been given in Sec. 3.1.2 using the quasi-static approximation. Assuming that $K_a \approx 1/|H_r(0)| = k/S_d$, and using the values of Figs. 3.4 and 3.5, the quasi-static estimate is found to be $p_t \simeq 1664$ [Pa]. However, as observed by Keefe [77], this value underestimates the true p_t.

The quasi-static estimate is compared with experimental results from the simulations, in two steps. First, a "dynamic estimate" is found by running simulations at very high sampling rates (up to 500 [kHz]). For such sampling rates, all the systems are found to have the same threshold pressure, $p_t = 1802$ [Pa]. This can be therefore assumed to be the "true" value. Then simulations are run at lower sampling rates: the measured p_t are given in Table 3.2, from which a few remarks can be made.

- For all the digital reeds, p_t converges to the dynamic estimate 1802 [Pa] as the sampling rate is increased (the convergence of AM2 is not evident from Table 3.2, since it occurs at $F_s > 200$ [kHz]).

- The p_t estimates obtained from both the 1-step methods exhibit robustness with respect to the sampling rate. At $F_s = 30$ [kHz] the percentage deviation of p_t from

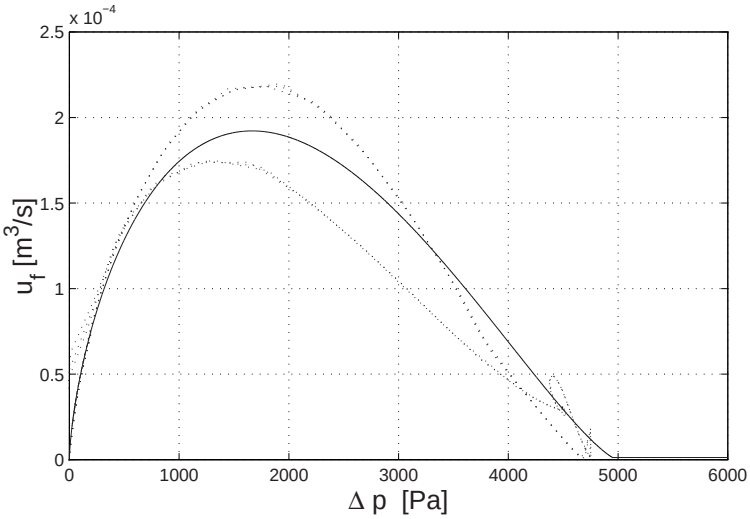

Figure 3.6: *Phase diagrams for u_f vs. Δp. Solid line: quasi-static model approximation (cfr. Eq. (3.7)). Dotted line: WS1 digital reed with $F_s = 44.1$ [kHz] and $p_m = 2265$ [Pa].*

the "true" value is less than 1% for both AM1 and WS1.

- The 2-step methods are less robust: even at high sampling rates, the threshold pressures for the corresponding systems are far from the "true" value. In particular, simulations with AM2 hardly reach steady state oscillations for $F_s < 35$ [kHz]. For this reason the AM2 column in Table 3.2 gives only the results for $F_s > 50$ [kHz].

As an example, Fig. 3.5(b) shows that at $F_s = 44.1$ [kHz] the frequency response H_{WS2} is indistinguishable from the original one. However, p_t is still noticeably higher than 1802 [Pa]. A similar remark holds for AM2 with $F_s = 100$ [kHz]. Therefore, time-domain simulations with the 2-step methods exhibit poor accuracy even when the reed response is approximated accurately. These results show that the study of the discrete response does not give sufficient information on reed behavior when the coupling with the non-linearity $f(u_f, h)$ and with the bore is considered. Due to the non-linearity, the whole system exhibits sensitive dependence to small deviations in the frequency response. The results from p_t analysis, together with the frequency-domain analysis of the previous section, show that the 1-step Weighted Sample method provides the most accurate reed simulations among the considered techniques.

A second study on time-domain simulations amounts to comparing the digital reeds with the quasi-static approximation described in Sec. 3.1.2. An interesting comparison is obtained by plotting the u_f vs. Δp phase diagrams for the steady state signals. An example of such phase diagrams is shown in Fig. 3.6, obtained by driving the WS1 digital reed with a mouth pressure $p_m = 2265$ [Pa]. This value is the limit value for non-beating conditions, while beating occurs for greater p_m.

The digital reed exhibits an interesting behavior: u_f and Δp move along a hysteretic

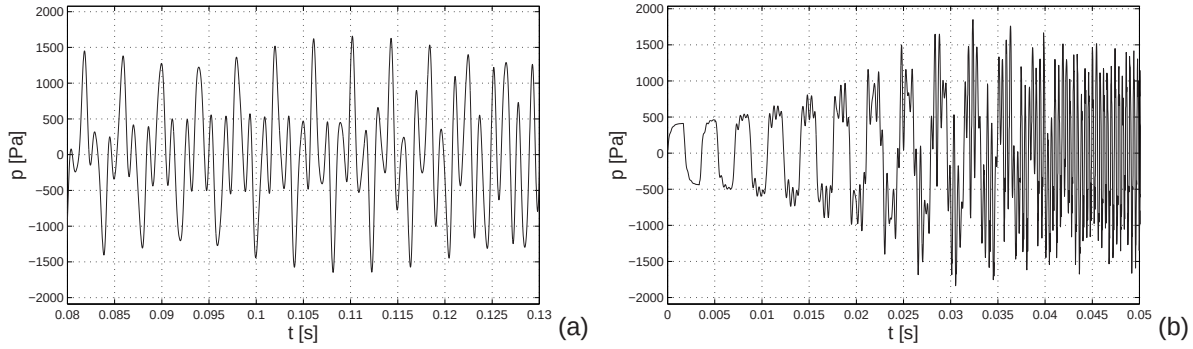

Figure 3.7: *Transitions to high regimes of oscillation (WS1 method, $F_s = 44.1$ [kHz]); (a) clarion register ($\omega_0 = 2\pi \cdot 2020$ [rad/s], $g = 1400$ [rad/s], $p_m = 1800$ [Pa]); (b) reed regime ($\omega_0 = 2\pi \cdot 3150$ [rad/s], $g = 300$ [rad/s], $p_m = 1900$ [Pa]).*

path. This phenomenon comes from the presence of memory in the equations. When the reed dynamics is taken into account, Eq. (3.7) does not hold, and h and u_f depend on Δp together with its derivatives. In other words, the attractor for the digital reed is not a curve in the plane, but instead a closed orbit in a higher dimensional phase space.

Another significant difference between the quasi-static and the dynamic model is concerned with transitions to high regimes of oscillation. As Thompson [135] and Wilson and Beavers [154] have pointed out, both ω_0 and g play a role in helping transition to the second register (clarion register). Experiments with artificial lips on real clarinets have proved that the clarion register can be produced without opening the register hole, if the reed resonance matches a low harmonic of the playing frequency and the damping is small enough. Moreover, an extremely low damping causes the reed regime ("squeaks") to be produced, i.e. the oscillation is governed by the reed resonance. From a musical standpoint, squeaks are often explained as a consequence of insufficient breathing, while the fundamental register comes in as the mouth pressure is increased. All these effects are seen to be well reproduced by numerical simulations with the WS1 digital reed, while on the contrary the quasi-static approximation does not allow control on such effects. Figure 3.7(a) shows an example of transition to the clarion register. This example has been obtained by matching ω_0 to the seventh harmonic of the playing frequency and by lowering g down to 1400 [rad/s]. Figure 3.7(b) shows a transition to the reed regime. This is achieved by giving g a value as low as 300 [rad/s]. Squeaks are more easily obtained in simulations by driving the reed with low blowing pressures.

3.3 Finite-difference modeling

The lumped model described in the previous section is appealing for sound synthesis purposes. The synthesis algorithm is compact enough to be implemented at low computational costs, and the conceptual simplicity of the physical description in terms of lumped

masses, springs and dampers, allows in principle intuitive control. However, a major limitation of the lumped model is that its validity can only be assumed for oscillations at small amplitudes. At larger amplitudes the reed *bends* against the mouthpiece lay, and the assumption of a single constant degree of freedom no longer holds. Furthermore, the phenomenon of reed beating (i.e. complete closure of the reed) is usually incorporated in the lumped model in a non-physical way, as seen in Sec. 3.1.1. Also, the dependence of the reed-mouthpiece system on the lip embouchure is not taken into account in the lumped representation.

A more realistic representation of the reed is that of a bar with non-uniform cross-sectional area, clamped to the mouthpiece at one end. Additional constraints on the reed motion are provided by the mouthpiece profile and by the interaction with the lip. Stewart and Strong [130] developed a numerical model based on such a distributed description, which incorporated automatically bending and beating phenomena. Later, Sommerfeldt and Strong [128] used the same model for studying a player-clarinet system where the player's air column was also included in the simulations. The distributed modeling approach was also adopted by Gilbert [63] and Gazengel [61], who developed analytical studies of the reed bending in the case of simplified geometries (wedge shaped reed, circular lay profile) and proposed a method for designing a lumped model approximation with non-constant parameters.

Distributed modeling of the reed-mouthpiece-lip system is also the topic of this section. The modeling principles presented here are similar to those adopted in [130, 128], but contain several improvements and refinements. The fourth-order partial differential equation (*PDE*) describing the reed is adapted in a such way that internal losses and damping due to the surrounding air, as well as the physical interaction with the mouthpiece lay and the player's lips, are modeled more accurately. Both geometrical and mechanical parameters are obtained via precise measurements on real clarinets. The equations are discretized using an implicit finite-difference scheme that guarantees unconditional stability and minimum frequency warping in the digital domain. This numerical technique was already used by Chaigne and Doutaut [28] for the modeling of idiophones. No attempt is made here to model either the air flow in the reed channel or the acoustical resonator. Instead attention is exclusively focused on the mechanical response of the system. The proposed distributed model is used in Sec. 3.4 as a numerical experimental set-up for the determination of the parameter functions of a lumped model approximation.

3.3.1 A distributed model

The reed is modeled here as a bar with length L, uniform width w and non-uniform thickness $b(x)$ (see Fig. 3.8). It is clamped to the mouthpiece at one end ($x = 0$) and free at the other one ($x = L$). Only flexural waves in the vertical direction are considered, and contribution of other waves are neglected. If the material is homogeneous and isotropic, then its density ρ and Young's modulus Y are constant, and the vertical displacement

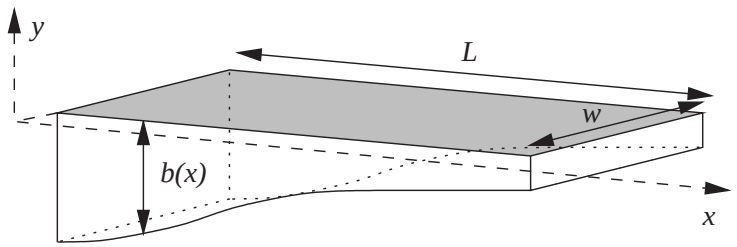

Figure 3.8: *Idealized geometry of a clarinet single reed.*

distribution $y(x,t)$ is governed by the equation [28]

$$F(x,t) = \frac{\partial^2}{\partial x^2}\left[Y\, I(x)\left(1+\eta\,\frac{\partial}{\partial t}\right)\frac{\partial^2 y}{\partial x^2}(x,t)\right] + \rho S(x)\left[\frac{\partial^2 y}{\partial t^2}(x,t) + \gamma_B \frac{\partial y}{\partial t}(x,t)\right], \quad (3.15)$$

where $F(x,t)$ is a driving force per unit length. The terms $S(x)$ and $I(x)$ are the cross sectional area and the moment of inertia about the longitudinal axis, respectively. If a rectangular cross-section is assumed, then $S(x) = wb(x)$ and $I(x) = wb(x)^3/12$ (see [49]). Note that Eq. (3.15) holds only if $S(x)$ is a slowly varying function [111]. The coefficient η represents the magnitude of the internal viscoelastic losses, and γ_B accounts for damping of the surrounding fluid. Chaigne and Doutaut [28] have shown that these two coefficients yield a satisfactory representation of bar losses. In particular, they found that the frequency dependent damping factor $\alpha(\omega)$ of the bar is fairly well fitted by the empirical law

$$\alpha(\omega) = \frac{1}{t_d(\omega)} = a_0 + a_2\omega^2, \quad (3.16)$$

where t_d is the decay time. It is shown in [28] that the $a_{0,2}$ coefficients are related to η, γ_B through the formulas

$$\eta = 2a_2 \quad \text{and} \quad \gamma_B = 2a_0. \quad (3.17)$$

Finally, clamped-free boundary conditions lead to the constraints (see [94])

$$\begin{aligned} y(0,t) &= \frac{\partial y}{\partial x}(0,t) = 0, \\ \frac{\partial^2 y}{\partial x^2}(L,t) &= \frac{\partial^3 y}{\partial x^3}(L,t) = 0. \end{aligned} \quad (3.18)$$

Besides being clamped at the mouthpiece, the reed is also sustained by the player's lip. Given lip coordinates (x_{lip}, y_{lip}) as in Fig. 3.9, it is assumed here that the reed and the lip are permanently in contact over a fixed segment $X_{lip} = (x_{lip} - L_{lip}, x_{lip} + L_{lip})$, and that there is an elastic lip force per unit length $F_{lip}(x,t)$ between them that is proportional to the lip compression:

$$F_{lip}(x,t) = \begin{cases} -K_{lip}\Delta y_{lip}(x,t), & x \in X_{lip}, \\ 0, & \text{otherwise}, \end{cases} \quad (3.19)$$

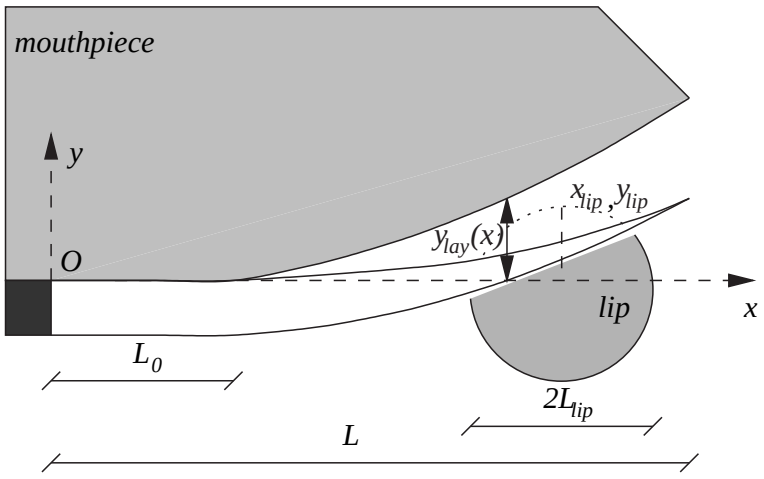

Figure 3.9: *Geometrical parameters for the mouthpiece lay and the lip.*

where $\Delta y_{lip}(x,t) = y_{lip} - y(x,t) + b(x)$ indicates lip compression and K_{lip} is the lip stiffness per unit length. Equation (3.19) provides the simplest possible restoring force, which is intended as a first-order approximation in the case of using an artificial lip (as in the measurements of Sec. 3.3.3). Note that no lip restoring forces are modeled in [130, 128]. As a second effect, the lip provides additional damping to the reed. Following [128], this is modeled by redefining the damping coefficient γ_B as position dependent:

$$\gamma_B(x) = \begin{cases} \gamma_{air} + \gamma_{lip}, & x \in X_{lip}, \\ \gamma_{air}, & \text{otherwise.} \end{cases} \tag{3.20}$$

Reed-lay interaction is slightly more complicated, since the portion of reed in contact with the mouthpiece varies over time. The contact force per unit length between reed and lay is defined as $F_{lay}(x,t) = F_{el}(x,t) + F_{dis}(x,t)$, where F_{el} and F_{dis} are an elastic and a dissipative component, respectively. The lay profile $y_{lay}(x)$ is given as in Fig. 3.9. Suppose that a section dx of the reed around x is hitting the lay, i.e. $y(x,t) = y_{lay}(x)$. Since collision is nearly inelastic [130], F_{dis} is chosen such that it nullifies the velocity of the section during the duration Δt of the collision. If Δt is short, the following relation is obtained from the law of impulsion:

$$dm(x)\ \dot{y}(x,t) = [F_{dis}(x,t)dx]\,\Delta t, \tag{3.21}$$

where $dm(x) = \rho b(x) w\ dx$ is the mass of the reed section. Therefore F_{dis} is non-zero only during the time Δt in which collision occurs. Equation (3.21) is exploited in the following section for deriving an expression for F_{dis} in the discrete space-time domain.

The elastic term F_{el} is nonzero only if the reed section is still in contact with the lay after the collision. Let $\Delta y_{lay}(x,t) = y(x,t) - y_{lay}(x)$ denote the reed compression. Then the set $X_{lay}(t) = \{x \in [0, L] \ :\ \Delta y_{lay}(x,t) > 0\}$ denotes the points where the reed and the lay are in contact. In this case, a linear elastic restoring force is applied, such that

F_{el} takes the form

$$F_{el}(x,t) = \begin{cases} -K_{lay}\Delta y_{lay}(x,t), & x \in X_{lay}(t), \\ 0, & \text{otherwise.} \end{cases} \tag{3.22}$$

Again, note that contact forces with the lay are not modeled in [130, 128]. In these papers, the interaction is simply treated by imposing a stop to the sections that hit the lay: the section velocity is set to zero and the section displacement is kept constant and equal to $y_{lay}(x)$ as long as the section tends to move into the mouthpiece. However, it is easily seen that this approach introduces artificial accelerations in the system. Imposing a stop on a reed section means adding forces also on the neighboring sections, and since these sections are not necessarily in contact with the mouthpiece, their accelerations are in many cases artificially affected. The explicit modeling of contact forces proposed here avoids such artificial accelerations and results in a more realistic interaction.

In conclusion, the total force F in Eq. (3.15) acting on the reed is given by

$$F(x,t) = F_{lip}(x,t) + F_{lay}(x,t) + F_{ext}(x,t), \tag{3.23}$$

where the term F_{ext} stands for an external driving force per unit length. The distributed model (3.15) is now completely defined in the continuous-time domain.

3.3.2 Numerical formulation

This section addresses the problem of discretization in time and space of the distributed system (3.15). Finite-difference methods (*FDM*) approximate the spatial domain with a grid of N sections (equivalently, $N+1$ points), and a corresponding spatial sampling step $X_s = L/N$. In the following the notation y_i^n is used to denote the value of the reed displacement y at the point $x_i = iX_s$ (for $i = 0 \dots N$) and at time $t_n = nT_s$ (with $n \geq 0$); the vector notation $\boldsymbol{y}^n = [y_i^n]_i$ is also used.

The selection of the two parameters X_s, T_s is usually imposed by stability and frequency warping criteria. *Explicit* FDM can be used, where discretization of the derivatives leads to an explicit dependence of y_i^{n+1} on known quantities. It is a general result that if the PDE does not model any damping, explicit schemes lead to a Von Neumann stability condition of the form $T_s/X_s^2 < c$, where c is a constant [81]. This condition takes on a slightly different form when dissipative elements are included.

Chaigne and Doutaut have shown that the stability condition for the numerical formulation of Eq. (3.15) with explicit methods can be written $N < N_{max}(F_s)$ i.e. the number of grid-points has an upper bound determined by the sampling rate. Moreover, it was shown in [28] that for a non-uniform bar modeled with equation (3.15) the function $N_{max}(F_s)$ is roughly proportional to $\sqrt{F_s}$ for low F_s, and tends to an asymptotic limit as F_s increases. Therefore, extremely high sampling rates should be used in order to achieve an acceptable spatial resolution, and consequently a reasonable estimation of the first natural frequencies of the bar.

For this reason a class of *implicit* methods, sometimes referred to as θ-schemes, is used. For clarity, equation (3.15) is rewritten as

$$\begin{aligned} F(x,t) = \; & \rho S(x)\left[\frac{\partial^2 y}{\partial t^2}(x,t) + \gamma_B \frac{\partial y}{\partial t}(x,t)\right] + Y\frac{\partial^2}{\partial x^2}\left[I(x)\frac{\partial^2 y}{\partial x^2}(x,t)\right] + \\ & + \eta Y \frac{\partial^2}{\partial x^2}\left[I(x)\frac{\partial^3 y}{\partial x^2 \partial t}(x,t)\right], \end{aligned} \tag{3.24}$$

and the discretization of each of these derivative terms is addressed separately in the following. Time derivatives in equation (3.24) are approximated using the centered difference scheme

$$\frac{\partial y}{\partial t}(x_i,t_n) \approx \frac{y_i^{n+1} - y_i^{n-1}}{2T_s}, \tag{3.25}$$

$$\frac{\partial^2 y}{\partial t^2}(x_i,t_n) \approx \frac{y_i^{n+1} - 2y_i^n + y_i^{n-1}}{T_s^2}, \tag{3.26}$$

The discrete operator δ_x^2, that approximates the second-order spatial derivatives, is again defined with a centered difference scheme. The second derivative of a generic function $g(x)$ is given by

$$\frac{\partial^2 g}{\partial x^2} \approx \delta_x^2 g, \quad \text{where} \quad (\delta_x^2 g)_i^n = \frac{g_i^{n+1} - 2g_i^n + g_i^{n-1}}{X_s^2}. \tag{3.27}$$

With this notation, the second-order derivative $\partial^2 y/\partial x^2$ is approximated as $\delta_x^2 y$ and the third-order derivative $\partial^3 y/\partial x^2 \partial t$ is approximated as

$$\frac{\partial^3 y}{\partial x^2 \partial t}(x_i,t_n) \approx \frac{(\delta_x^2 y)_i^{n+1} - (\delta_x^2 y)_i^{n-1}}{2T_s}. \tag{3.28}$$

The θ-scheme approximates the fourth-order spatial derivative in equation (3.24) using the following three-level time average:

$$\frac{\partial^2}{\partial x^2}\left[I\,\frac{\partial^2 y}{\partial x^2}\right](x_i,t_n) \approx (1-2\theta)\delta_x^2[I(\delta_x^2 y)]_i^n + \theta\left[\delta_x^2[I(\delta_x^2 y)]_i^{n+1} + \delta_x^2[I(\delta_x^2 y)]_i^{n-1}\right], \tag{3.29}$$

where $\delta_x^2[I(\delta_x^2 y)]_i^n$ can be written explicitly as

$$\begin{aligned} \delta_x^2[I(\delta_x^2 y)]_i^n = \; & \frac{I_{i+1}y_{i+2}^n - 2(I_{i+1}+I_i)y_{i+1}^n}{X_s^4} + \frac{(I_{i+1}+4I_i+I_{i-1})y_i^n}{X_s^4} + \\ & + \frac{-2(I_i+I_{i-1})y_{i-1}^n + I_{i-1}y_{i-2}^n}{X_s^4}. \end{aligned} \tag{3.30}$$

The last derivative in equation (3.24) is again obtained using the θ-scheme. Substituting $\partial^3 y/\partial x^2 \partial t$ with its approximation (3.28) and applying the θ-scheme (3.29) yields

$$\frac{\partial^2}{\partial x^2}\left[I\,\frac{\partial^3 y}{\partial x^2 \partial t}\right](x_i,t_n) \approx \theta\; f_s\left[\delta_x^2[I(\delta_x^2 y)]_i^{n+1} - \delta_x^2[I(\delta_x^2 y)]_i^{n-1}\right]. \tag{3.31}$$

Finally, extra grid-points have to be added at both ends, in order to express the boundary conditions (3.18) in discrete form. With the discretization scheme adopted here, the discrete boundary conditions are found to be

$$\begin{aligned} &(x=0) \qquad y_0^n = 0, \quad y_{-1}^n = y_1^n, \\ &(x=L) \qquad y_{N+1}^n = 2y_N^n - y_{N-1}^n, \quad y_{N+2}^n = y_N^n - 4y_{N-1}^n + y_{N-2}^n. \end{aligned} \tag{3.32}$$

Chaigne and Doutaut [28] exemplify the stability properties of this numerical scheme by applying it to the ideal bar equation $\partial^2 y/\partial t^2 = -a^2 \partial^4 y/\partial x^4$. Their analytical study shows that the resulting difference equations are unconditionally stable (i.e. they are stable for any combination of X_s and F_s) for any value $\theta \geq 1/4$. Moreover, minimum frequency warping is achieved with $\theta = 1/4$. Note that Stewart and Strong [130] and Sommerfeldt and Strong [128] use the θ-method with $\theta = 1/2$, although they do not mention it explicitly.

For clarity, let Δ_i^n denote the approximation of the spatial derivative term provided by Eqs. (3.29) and (3.31):

$$\frac{\partial^2}{\partial x^2}\left[I\left(1+\eta\,\frac{\partial}{\partial t}\right)\frac{\partial^2 y}{\partial x^2}\right](x_i,t_n) \approx \Delta_i^n. \tag{3.33}$$

Then the difference equation is seen to take the form

$$(1+\gamma)\,y_i^{n+1} - 2y_i^n + (1-\gamma)\,y_i^{n-1} = \frac{T_s^2}{\rho S_i}\left[F_i^n - Y\Delta_i^n\right], \tag{3.34}$$

where $\gamma = \frac{1}{2}\gamma_B T_s$. Note that this equation is implicit, since the term Δ_i^n depends on y_i^{n+1}. In matrix form, the numerical system is:

$$\boldsymbol{M}_1\boldsymbol{y}^{n+1} = \boldsymbol{M}_0\boldsymbol{y}^n + \boldsymbol{M}_{-1}\boldsymbol{y}^{n-1} + \boldsymbol{M}_F\boldsymbol{F}^n. \tag{3.35}$$

The coefficients of the matrices $\boldsymbol{M}_1$, $\boldsymbol{M}_0$, $\boldsymbol{M}_{-1}$ and $\boldsymbol{M}_F$ are found from the calculations outlined above. Thus, computation of $\boldsymbol{y}^{n+1}$ requires inversion of the matrix $\boldsymbol{M}_1$, and the final numerical sistem has the form:

$$\boldsymbol{y}^{n+1} = \boldsymbol{A}_0\boldsymbol{y}^n + \boldsymbol{A}_{-1}\boldsymbol{y}^{n-1} + \boldsymbol{A}_F\boldsymbol{F}^n, \tag{3.36}$$

where $\boldsymbol{A}_{[0,-1,F]} = \boldsymbol{M}_1^{-1}\boldsymbol{M}_{[0,-1,F]}$.

The elastic lip force $(\boldsymbol{F}_{lip})^n$ and the elastic component of the lay forces $(\boldsymbol{F}_{el})^n$ are easily computed directly from the continuous-time equations (3.19) and (3.22), respectively. Computation of the dissipative force $(\boldsymbol{F}_{dis})^n$ is less straightforward. As stated previously in this section, $(F_{dis})_i^n$ is non-zero only when a collision occurs, therefore one first needs to find the sections i that are colliding with the lay at each time step n. To this end the next value $\hat{\boldsymbol{y}}^{n+1}$ for the displacement distribution is predicted from the numerical system (3.36) using a force vector $(\hat{\boldsymbol{F}})^n$ that contains only lip and external forces. This way, it

is possible to predict which sections i are going to collide with the mouthpiece lay using the condition

$$C_i^n \quad : \quad \{\hat{y}_i^{n+1} > (y_{lay})_i \quad \text{and} \quad y_i^n < (y_{lay})_i\}. \tag{3.37}$$

Moreover, those sections that are colliding with the lay have an estimated velocity

$$\hat{v}_i^n = \frac{\hat{y}_i^{n+1} - y_i^n}{T_s}. \tag{3.38}$$

From these equations and from equation (3.21) an expression for $\boldsymbol{F}_{dis}$ can be derived in discrete time:

$$(F_{dis})_i^n = \begin{cases} \dfrac{m_i}{T_s}\hat{v}_i^n, & \text{if} \quad C_i^n \text{ holds,} \\ 0, & \text{otherwise,} \end{cases} \tag{3.39}$$

where $m_i = \rho b_i w$ is the mass per unit length of the i^{th} section. Once $(\boldsymbol{F}_{dis})^n$ has been estimated, the next displacement $\boldsymbol{y}^{n+1}$ is recomputed from equation (3.35) using the force vector $\boldsymbol{F}^n$ that includes lay forces.

Summarizing, at each time step n the new displacement $\boldsymbol{y}^{n+1}$ is computed as described in the following pseudocode lines:

```
for     n ≥ 1
        Compute     F̂^n = F^n_lip + F^n_ext
        Estimate    ŷ^{n+1} = A_0 y^n + A_{-1} y^{n-1} + A_F F̂^n
        Evaluate    condition C^n_i and velocities v̂^n_i
        Compute     F^n_lay and F^n = F̂^n + F^n_lay
        Compute     y^{n+1} = A_0 y^n + A_{-1} y^{n-1} + A_F F^n
end
```

3.3.3 Results from simulations

In order to obtain meaningful results from the numerical simulation of the reed-mouthpiece-lip system, it is essential to establish accurate estimates of the model parameters. Not all the parameters are easily found via direct measurements. Therefore some of them are estimated by adjusting their values while running numerical simulations, until the behavior is similar to that observed in real systems. All the parameter values listed in Table 3.3 are base on measurements carried out by Maarten van Walstijn, at the University of Edinburgh. Details about the measurements can be found in [145, 15].

Numerical simulations analyzed in the following use a Bundy mouthpiece. For this mouthpiece, the profile $y_{lay}(x)$ has been measured using a traveling microscope, and the data-points are fitted with a fourth-order polynomial [145]. The curved profile is found as

$$y_{lay}(x) = 1.6181(x - L_0)^2 + 1.8604(x - L_0)^3 + 5.5077 \cdot 10^2(x - L_0)^4 \quad [m], \tag{3.40}$$

Reed - (RICO plasticover, hardn. 2)	
Length (free part when clamped)	$L = 34 \cdot 10^{-3}$ [m]
Width	$w = 13 \cdot 10^{-3}$ [m]
Density	$\rho = 500$ [Kg/m^3]
Young's Modulus	$Y = 5.6 \cdot 10^9$ [N/m^2]
Viscoelastic constant	$\eta = 6.0 \cdot 10^{-7}$ [s]
Fluid damping coefficient	$\gamma_{air} = 100$ [s^{-1}]
Mouthpiece (Bundy)	
Lay length (total)	$L = 34 \cdot 10^{-3}$ [m]
Lay length (flat part)	$L_0 = 9 \cdot 10^{-3}$ [m]
Contact stiffness per unit length	$K_{lay} = 10^8$ [N/m^2]
Artificial lip (water-filled balloon)	
Horizontal position	$x_{lip} = 26 \cdot 10^{-3}$ [m]
Vertical position	$y_{lip} = 3.85 \cdot 10^{-3}$ [m]
Length of the contact segment	$2L_{lip} = 9 \cdot 10^{-3}$ [m]
Lip stiffness per unit length	$K_{lip} = 6.5 \cdot 10^4$ [N/m^2]
Additional lip damping	$\gamma_{lip} = 16000$ [s^{-1}]

Table 3.3: *Parameters of the distributed model used in the simulations.*

for $x > L_0$.

A plastic-coated reed is used in the following instead of a cane reed, because its mechanical properties remain approximately constant independently of humidity. In the approximation given in Fig. 3.8, the reed is assumed to have a rectangular cross-section. This does not fully correspond to reality, since in real reeds the thickness b is not constant over the width w. Therefore both the minimum and the maximum reed thickness are measured at position x, then an "effective" thickness is computed under the hypothesis that the curvature over w has a circular shape. Measurements have been taken for 18 points along the reed length, and the data-points are fitted again with a fourth-order polynomial [145]. The thickness profile for the plastic reed is

$$b(x) = 2.2633 \cdot 10^{-3} - 4.9483 \cdot 10^{-2} x - 4.444 x^2 + 2.0126 \cdot 10^2 x^3 - 2.4385 \cdot 10^3 x^4 \quad \text{[m]}. \tag{3.41}$$

Impulse responses of the numerical model are first analyzed by running the simulations without lip and lay interactions. Figure 3.10 shows the reed responses with $\theta = 1/4, 1/2$, when using $F_s = 200$ [kHz] and $N = 200$. Estimation of the resonances through FFT analysis shows that frequency warping starts to be significant above the first three resonances. In particular, $\theta = 1/4$ and $\theta = 1/2$ provide the same estimate $f_0 = 1405$ [Hz] for the first resonance, while the discrepancy in the estimation of the second and third resonances is less than 1%. Nonetheless, these experimental results confirm that $\theta = 1/4$

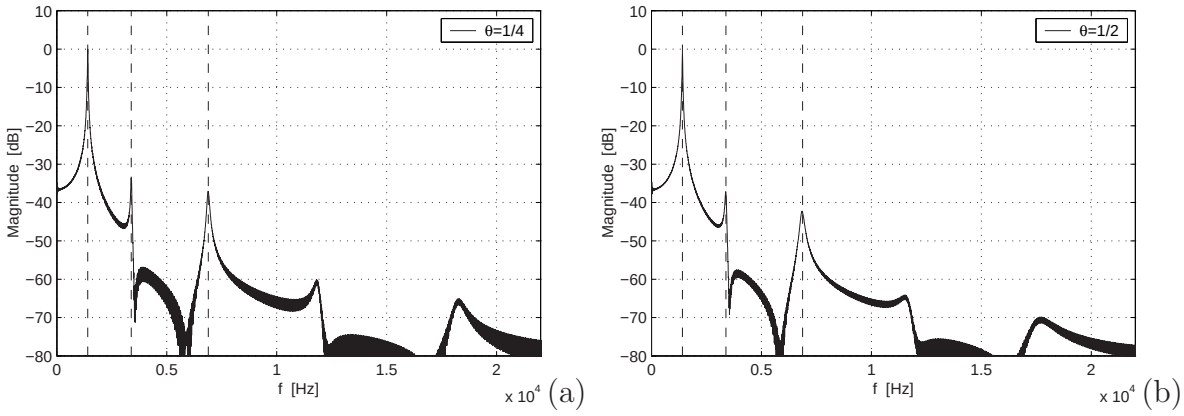

Figure 3.10: *Comparison of the reed responses for $\theta = 1/4$ and $\theta = 1/2$ (obtained from simulations with $F_s = 200$ [kHz] and $N = 200$.*

minimizes the frequency warping effects, therefore this value is used in the rest of this section.

In order to choose suitable values for F_s and N a commonly used tool is analysis of convergence, i.e. analysis of resonance estimation in the numerical system with respect to F_s and N. First, simulations are run with various F_s and N, without lip and lay interactions. The first resonance f_0 of the system is then obtained through FFT analysis. Higher reed resonances are not taken into account since they are strongly damped and their estimation does not affect the system behavior significantly.

The results are summarized in Table 3.4: it can be noticed that the f_0 estimated values always converge with increasing N. Moreover, the columns with $F_s = 200$ [kHz] and $F_s = 400$ [kHz] provide the same estimated f_0 values for every N. Also, the estimate with $F_s = 200$ [kHz] and $N = 200$ is very close to that found with $F_s = 200$ [kHz] and $N = 400$, the discrepancy being 0.2%. These results show that the values $F_s = 200$ [kHz] and $N = 200$ provide sufficiently accurate simulations. Therefore, these values are used in the rest of this section.

Using the numerical parameters θ, F_s, N as determined above, simulations of the complete system are run both in a *dynamic* and in a *quasi-static* manner. Dynamic simulations are obtained by driving the system with a sinusoidal force per unit length $(F_{ext})^n$ given by:

$$(F_{dyn})^n = a_{min} + \left(\frac{a_{max} - a_{min}}{2}\right) \cdot \left[1 + \sin\left(\omega_{dyn}\ nT_s + \phi_0\right)\right], \tag{3.42}$$

where ϕ_0 is such that $F_{dyn} = 0$ for $n = 0$. Figure 3.11 shows the driving signal together with the measured tip displacement $y(L, t)$. In this example the minimum and maximum force values a_{min} and a_{max} have been chosen in such a way that reed beating can be observed. These plots are qualitatively in agreement with findings by Idogawa *et al.* [72] on a real clarinet artificially blown, but provide a more detailed description of the closing phase. It is clearly seen that, due to interaction with the lay, the reed tip can

		F_s [kHz]			
N	X_s [mm]	50	100	200	400
20	1.7	1361	1363	1364	1364
50	0.68	1387	1389	1390	1390
100	0.34	1397	1399	1401	1401
150	0.227	1400	1403	1404	1404
200	0.17	1402	1405	1405	1405
250	0.136	1403	1405	1407	1407
300	0.113	1403	1406	1407	1407
350	0.097	1404	1407	1407	1407
400	0.085	1405	1407	1408	1408

Table 3.4: *Analysis of convergence for f_0, with varying F_s and N. Frequency resolution in the FFT is* 1.5 *[Hz], the f_0 values are rounded to the nearest integer.*

not exceed the value $y_m = 1.3 \cdot 10^{-3}$ [m]. However, the tip is not stopped suddenly but rather gradually, and a zoom-in of the plot in Fig. 3.11(b) in fact reveals that complete closure is not obtained. This suggests that the reed *stiffness per unit area* K_a as defined in Eq. (3.6) is increasing as y_L approaches its maximum value. Another phenomenon which is evident from the same plot is that y_L exhibits small additional oscillations when approaching the maximum.

A quantitative description of the dependence of K_a on the tip displacement can be obtained through numerical simulations. However, dynamic simulations as defined above are not suitable for studying such dependence, since the reed velocity can in principle induce hysteresis effects on the force-tip displacement characteristics. Therefore, *quasi-static* simulations are defined in the following way:
(1) the driving force is synthesized by taking sampled values of a sinusoidal signal and applying them in successive time steps of length n_0. If the integer part of n/n_0 is denoted as $\lfloor n/n_0 \rfloor$, then F_{ext} is defined in the quasi-static case as:

$$(F_{qs})^n = a_{min} + \left(\frac{a_{max} - a_{min}}{2}\right) \cdot \left[1 + \sin\left(\omega_{qs} \left\lfloor \frac{n}{n_0} \right\rfloor T_s + \phi_0\right)\right]. \qquad (3.43)$$

Therefore $(F_{qs})^n$ takes constant values in each time step of length n_0, and follows a sinusoidal envelope.
(2) The length n_0 is chosen such that reed oscillations decay completely during each step. When oscillations have decayed, the "quasi-static" reed displacement $\boldsymbol{y}^n$ distribution is measured.

Figure 3.12(a) shows the force-tip displacement characteristics as found from quasi-static simulations. As expected, the elastic behavior of the system is approximately linear for small relative displacements, and becomes increasingly non-linear as the reed

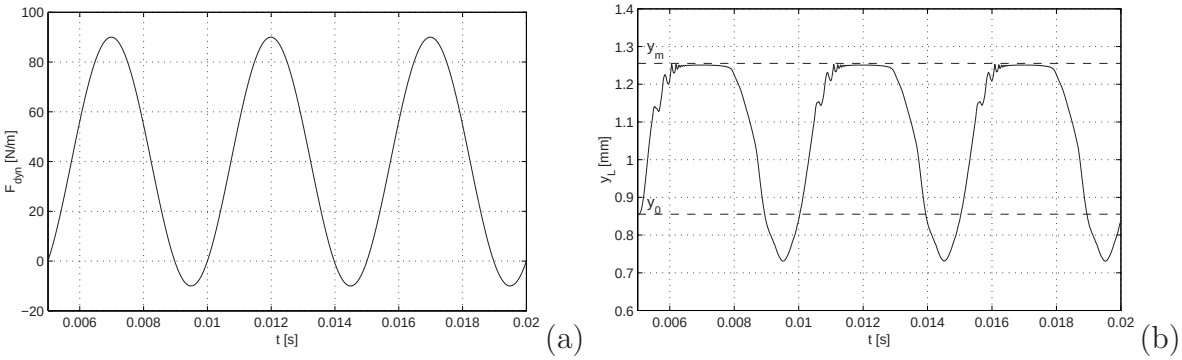

Figure 3.11: *Dynamic simulations; (a) external driving force and (b) reed tip displacement.*

moves toward closure. Figure 3.12(b) shows the behavior of K_a, computed as the ratio between Δp and $y_L - y_0$. Note that K_a is almost constant around y_0 and starts to increase significantly around the value $y_L \sim 1.08 \cdot 10^{-3}$ [m]. The additional oscillations observed in Fig. 3.11(b) occur approximately in the same range. This findings suggest that the mechanical properties of the system are changing dramatically in that range.

In order to further investigate these changes another numerical experiment is carried out, in which the *separation point* (defined as the point of contact between lay and reed which is closest to the tip) is measured as a function of the reed tip displacement. The separation point is denoted by x_{sep} in the following.

The dependence $x_{sep}(y_L)$ describes the way the reed bends against the lay, and in particular defines the part of the reed that can move freely during oscillation. Therefore it can be expected to be strongly correlated to the mechanical properties of the reed-mouthpiece system. The separation point is first studied using quasi-static simulations with driving forces F_{qs}. The measured values are plotted in Fig. 3.13(a), which shows an unexpected result: the reed does not bend against the lay in a smooth manner, instead x_{sep} undergoes a sudden jump around the value $y_L \sim 1.08 \cdot 10^{-3}$ [m]. This behavior corresponds to a reed section closer to the tip getting in contact with the lay before the "previous" segment has fully curled up onto the lay: a qualitative description of this behavior is given in Fig. 3.13(b).

Secondly, the separation point is studied using dynamic simulations. Simulations are run using driving forces F_{dyn} as in equation (3.42) with various driving frequencies ω_{dyn} ranging from $2\pi 200$ [rad/s] up to $2\pi 1500$ [rad/s]. The gray dots in Fig. 3.13(a) show results obtained with $\omega_{dyn} = 2\pi 200$ [rad/s]: slight hysteretic effects can be noticed, due to the fact that the opening and closing paths do not coincide exactly in dynamic simulations. However, the deviation from the static curve is generally small. Similar results are found for greater driving frequencies.

The separation point discontinuity explains the additional oscillations in the tip displacement, observed when running dynamical simulations (see Fig. 3.11(b)). When the discontinuity occurs, a section of the reed hits the lay as depicted in Fig. 3.13(b), and

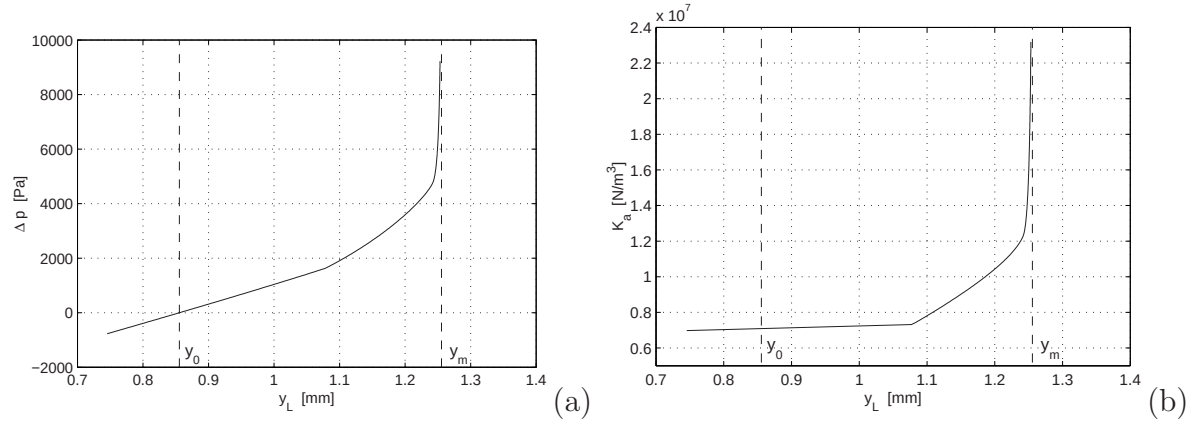

Figure 3.12: *Quasi-static simulations: (a) pressure-displacement characteristics of the system and (b) stiffness per unit area K_a.*

the collision causes the free part to oscillate. Furthermore, the frequency of these oscillations is significantly higher than the first resonance f_0 of the reed. Qualitatively, this is explained by observing that at discontinuity the length of the reed segment which is moving freely is significantly smaller than the original length L.

As a concluding remark, the analysis on $x_{sep}(y_L)$ shows that quasi-static simulations provide an accurate description of the reed motion, and can be used for approximating dynamic simulations.

3.4 A lumped non-linear model

The distributed model developed in the last section shows that the bending of the reed against the mouthpiece affects dramatically the mechanical properties of the system. In particular, the stiffness per unit area K_a increases as the reed tip moves toward closure. This non-linear behavior has important consequences on the sound and the playing characteristics of the instrument.

It is evident that such non-linear bending effects are not accounted by the lumped model described in Sec. 3.1. On the other hand, the distributed approach of Sec. 3.3 has extremely high computational loads, and is not suitable for sound synthesis purposes. An alternative approach, that allows a much more efficient implementation, amounts to describe the reed as one-mass lumped oscillator, whose mechanical parameters are not constant. As an example, the mass m of the oscillator in Eq. (3.1) can be thought as decreasing as the reed moves toward closure, because the part of the reed which is free to move becomes smaller as it bends against the lay.

In order to develop a realistic model, a quantitative approach for the determination of the non-constant lumped parameters has to be found. This section proposes one solution to this task [144].

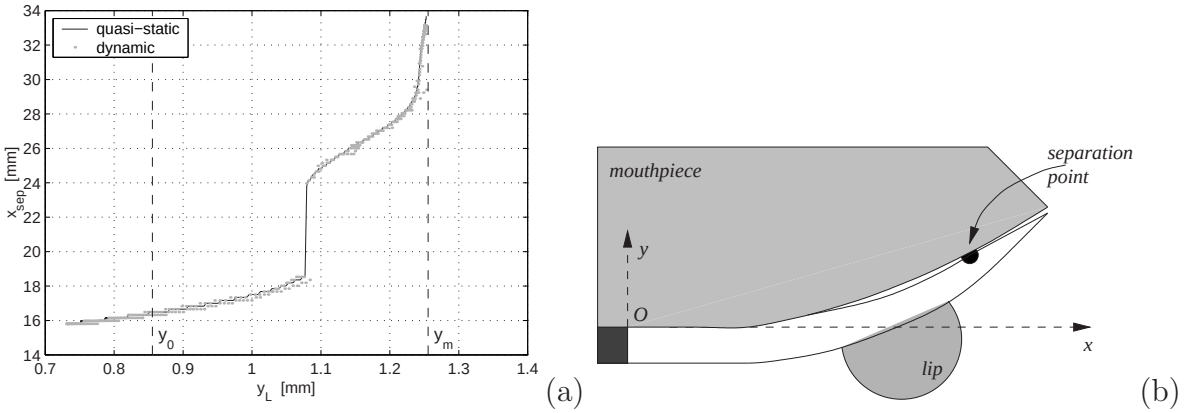

Figure 3.13: *Separation point versus tip displacement; (a) quasi-static simulations (black solid line) and dynamic simulations with $f_d = 200$ [Hz] (gray dots); (b) non-smooth reed curling.*

3.4.1 Parameter determination

In the remaining of this section the reed is considered to be described by the following equation:

$$m(y_L) \cdot \ddot{y}_L(t) + r(y_L) \cdot \dot{y}_L(t) + k(y_L) \cdot [y_L(t) - y_0] = S_d(y_L) \cdot \Delta p(t), \tag{3.44}$$

where the lumped parameters m, r, k are functions of the reed tip displacement. In other words, the reed is described as a *non-linear lumped oscillator*. Moreover, it is assumed that the effective flow surface S_r, defined in Eq. (3.5) is not constant either. Thus, the total flow u entering the instrument is given by

$$u(t) = u_f(t) + u_r(t), \quad \text{with} \quad u_r(t) = S_r(y_L) \cdot \dot{y}_L(t). \tag{3.45}$$

In order for Eqs. (3.44) and (3.45) to provide realistic description of the reed functioning, the variation of m, r, k, S_d, S_r with respect to y_L has to be established. The first attempt to use non-constant lumped parameters in single reed modeling appears to have been made by Adrien *et al.* [2]. However, the authors did not provide any theoretical or experimental method for the determination of the parameters. Gilbert [63] and Gazengel [61] developed an analytical method which is based on the assumption that at each time instant the lumped approximation (3.44) must have the same kinetic and potential energies of a distributed reed model. This approach is directly related to the so called *Rayleigh energy method*, often used for determining the eigenfrequencies of a beam under flexure.

The basic idea in the method is the following: if the potential and kinetic energies of the distributed model are known, then the lumped parameter functions can be established by imposing that the energies of the lumped model equal those of the distributed model at each time instant. Unfortunately, analytical results with this method can only be obtained

when the distributed model is given idealized geometries. As an example, Gazengel [61] was able to determine the lumped parameter using a distributed model in which a wedge shaped reed bends against a mouthpiece lay with circular curvature.

As shown in Sec. 3.3, real geometries of reed-mouthpiece systems are much more complicated than those used in [61], and the precise shape of the reed and the mouthpiece has a strong influence on the playing characteristics of the instrument. Therefore, an experimental approach based on numerical simulations of a distributed model may be preferable to a purely analytical analysis [144]. A similar experimental approach was adopted by De Vries *et al.* [36]: the authors determined the parameters of a lumped model of the vocal folds using numerical experiments with an accurate finite-element model of the glottis. In this section, the distributed model of Sec. 3.3 is used as a numerical experimental setup for deriving the lumped parameters in Eq. (3.44).

Consider the distributed reed model of Eq. (3.15). Given a displacement $y(x,t)$, the potential energy $E^*_{p,reed}$ and the kinetic energy $E^*_{k,reed}$ of the reed are found as (see [94])

$$\begin{aligned} E^*_{p,reed}(t) &= \frac{1}{2}\int_0^L YI(x)\left(\frac{\partial^2 y}{\partial x^2}(x,t)\right)^2 dx, \\ E^*_{k,reed}(t) &= E_k(t) = \frac{1}{2}\int_0^L \rho S(x)\left(\frac{\partial y}{\partial t}(x,t)\right)^2 dx. \end{aligned} \tag{3.46}$$

Note that $E^*_{k,reed}$ represents the total kinetic energy E^*_k of the reed-mouthpiece-lip system, since the reed is the only part which is actually moving. On the other hand, $E^*_{p,reed}$ does not represent the potential energy of the whole system, since the contributions $E^*_{p,lay}$ due to the mouthpiece and the contribution $E^*_{p,lip}$ due to the lip have to be taken into account. Using the same notation adopted in Eqs. (3.19) and (3.22), these contributions can be written as

$$E^*_{p,lay}(t) = \frac{K_{lay}}{2}\int_{X_{lay}} \Delta y_{lay}(x,t)^2 dx, \qquad E^*_{p,lip}(t) = \frac{K_{lip}}{2}\int_{X_{lip}} \Delta y_{lip}(x,t)^2 dx. \tag{3.47}$$

The total potential energy E^*_p of the distributed system is the sum of the three components: $E^*_p = E^*_{p,reed} + E^*_{p,lay} + E^*_{p,lip} - E^*_{p,0}$ where the constant $E^*_{p,0}$ is defined such that $E_p = 0$ when the system is at equilibrium.

Turning to the lumped model of Eq. (3.44), its potential energy E_p and kinetic energy E_k are defined as

$$E_p(t) = \frac{1}{2}k y_L(t)^2, \qquad E_k(t) = \frac{1}{2}m\dot{y}_L(t)^2. \tag{3.48}$$

Suppose now that E^*_p and E^*_k are measured from numerical simulations on the distributed model. The lumped approximation must have the same energies during motion, i.e. $E_p = E^*_p$ and $E_k = E^*_k$, and consequently the following equalities are derived:

$$k = \frac{2E^*_p}{y_L^2}, \qquad m = \frac{2E^*_k}{\dot{y}_L^2}. \tag{3.49}$$

As for the effective driving surface S_d, this can be determined by observing that in the static case ($\ddot{y}_L = \dot{y}_L = 0$) Eq. (3.44) reduces to the equality

$$k(y_L)[y_L(t) - y_0] = S_d \Delta p(t).$$

Recalling the static relation $\Delta p = K_a(y_L - y_0)$ (where K_a is the stiffness per unit area defined in Sec. 3.1.2), it follows that the effective driving surface S_d is given by

$$S_d = \frac{k}{K_a}. \tag{3.50}$$

Since k can be derived from Eq. (3.49) and K_a has been derived directly from numerical simulations in the last section, Eq. (3.50) can be used to compute S_d from known quantities. The flow effective surface S_r can be found as follows: the flow u_r^* induced by the reed motion is given in the distributed model as

$$u_r^*(t) = w \int_0^L \frac{\partial y}{\partial x}(x,t)dx.$$

Therefore S_r is found by imposing that $u_r = u_r^*$, and is therefore given by

$$S_r = \frac{u_r^*}{\dot{y}_L}. \tag{3.51}$$

Equations (3.49), (3.50) and (3.51) are used in the following section for determining the parameter functions. No method is proposed here for estimating the damping $r(y_L)$. A possible strategy may be found by exploiting the following remark: the dissipation $-r(y_L)\dot{y}_L$ and the driving force $S_d\Delta p$ are non-conservative forces acting on the lumped mass. Therefore the work of this dissipative terms must equal the variation of the total energy $E = E_p + E_k$. In formulas:

$$E(t + \Delta t) - E(t) = \int_{y_L(t)}^{y_L(t+\Delta t)} [S_d(y_L)\Delta p(t) - r(y_L)\dot{y}_L(t)]dy_L,$$

where $r(y_L)$ is the only unknown term. This equation can in principle be integrated numerically and used in combination with numerical experiments on the distributed system for determining $r(y_L)$. However, this approach is not investigated in the following. Instead the damping is arbitrarily assumed to be constant, and the exact determination of the function $r(y_L)$ is left for future work.

3.4.2 Properties of the non-linear oscillator

As already mentioned, the Rayleigh method relies on the assumption that the distributed model and the lumped approximation have the same energies at each time instant. However, it has been shown at the end of Sec. 3.3.3 that the distributed reed can be reasonably assumed to move in a "unique" way, i.e. hysteretic effects in reed motion are negligible (see in particular Fig. 3.13). This remark justifies the assumption made in

Sec. 3.4.1 that a one-to-one mapping relates the lumped parameter functions to the tip displacement. In other words, the term *time instant* can be substituted with the term *reed displacement*.

Given the above remark, quasi-static simulations are the most natural choice for measuring the energies of the distributed model. The advantage with respect to dynamic simulations is that more "clean" measures are obtained, since hysteretic effects are removed. Quasi-static simulations are run with the step-wise driving force F_{qs} given in Eq. (3.43), with $n_0 = 200$ and $\omega_{qs} = 2\pi 200$ [rad/s]. The force F_{qs} is given a total number $M = 1000$ of time steps. For each time step $m = 1 \dots M$ of the driving force, the following quantities are measured:

- the values of the reed displacement $\boldsymbol{y}^m$;
- the spatially discretized values of the "reed bend", $(\partial^2 \boldsymbol{y}/\partial x^2)^m$. These are needed for computing the reed potential energy in Eq. (3.46);
- the values of the lip compression $(\Delta y_{lip})^m$. These are needed for computing the lip potential energy in Eq. (3.47);
- the values of the lay compression $(\Delta y_{lay})^m$. These are needed for computing the lay potential energy in Eq. (3.47).

One last quantity, namely $(\partial \boldsymbol{y}/\partial t)^m$, is needed for computing the kinetic energy of the reed $E^*_{k,reed}$ and the reed-induced flow u^*_r. This is computed from the reed displacements as $(\partial \boldsymbol{y}/\partial t)^m = (\boldsymbol{y}^{m+1} - \boldsymbol{y}^{m-1})/2T_s$.

These measured quantities are used for computing the energies E^*_p, E^*_k, and the flow u^*_r. In order to compute the lumped parameter functions, one additional difficulty has to be overcome: in certain cases, the computations in Eqs. (3.49), (3.50) and (3.51) result in divisions where both the numerator and the denominator are very close to zero, which leads to numerical problems. Both the potential energy E^*_p and the relative displacement $(y_L - y_0)$ tend to zero when $y_L \to y_0$. Therefore, in the vicinity of y_0 the stiffness $k(y_L)$ is computed with a method that cross-fades between the values obtained from the ratio $E^*_p/(y_L - y_0)^2$ and the slope of a linear fit to the data $[(y_L - y_0)^2, E^*_p]$. Numerical problems are also found for m, S_d, S_r, in the vicinity of y_m, i.e. for very small reed openings: in this region the reed tip velocity tends to zero, and no direct computation of these parameter functions can be obtained. Instead, m, S_d, S_r are all assumed to go to zero when $y_L = y_m$. This is a plausible assumption, since no part of the reed is free to move for $y_L = y_m$. Figure 3.14 shows the resulting parameter functions.

Once the lumped parameters of the non-linear oscillator (3.44) have been found, this system has to be discretized. However, the discretization scheme proposed in Sec. 3.2 is not directly applicable in this context. The reason is that the K method in Sec. 3.2 is based on the assumption that the mechanical system has a linear part, while this is not the case for the system (3.44). The use of the K method for the accurate numerical simulation of the non-linear lumped oscillator is therefore an issue that needs further investigations.

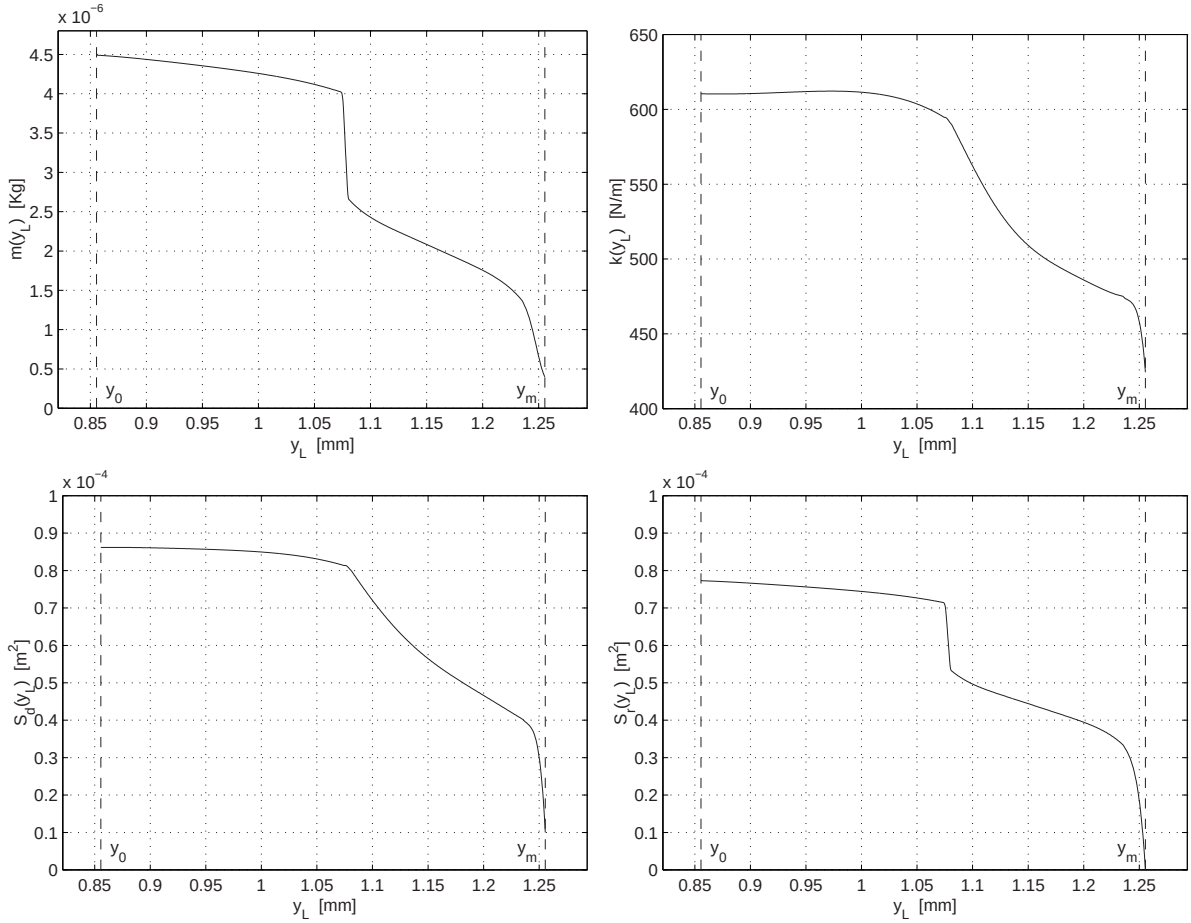

Figure 3.14: *Parameter functions of the lumped non-linear oscillator.*

An alternative discretization scheme, which does not make use of the K method, was proposed by van Walstijn [145, 147]. In his implementation, the non-linear oscillator is restated in the reduced form

$$\ddot{y}_L(t) + g(y_L) \cdot \dot{y}_L(t) + \omega_0(y_L)^2 \cdot [y_L(t) - y_0] = \frac{1}{\mu(y_L)} \cdot \Delta p(t), \tag{3.52}$$

where $g(y_L) = r(y_L)/m(y_L)$, $\omega_0(y_L) = \sqrt{k(y_L)/m(y_L)}$ and $\mu(y_L) = m(y_L)/S_d(y_L)$. The functions $\omega_0(y_L)$ and $\mu(y_L)$ are derived directly from the lumped parameters, and are stored as look-up tables (see Fig. 3.15). As already remarked, the function $g(y_L)$ is assumed to be constant ($g = 3000$ [rad/s]). For stability reasons, in the numerical implementation $g(y_L)$ is constant up to a certain displacement, and is set to increase according to a $\cosh(y_L)$ function for large displacements.

The derivatives $\dot{y}_L$ and $\ddot{y}_L$ in Eq. (3.52) are approximated using the mixed backward-centered scheme described in Sec. 2.3.1. As already remarked in Sec. 3.2.1, the resulting difference equation is explicit, i.e. the current value $y(n)$ depends only on past values

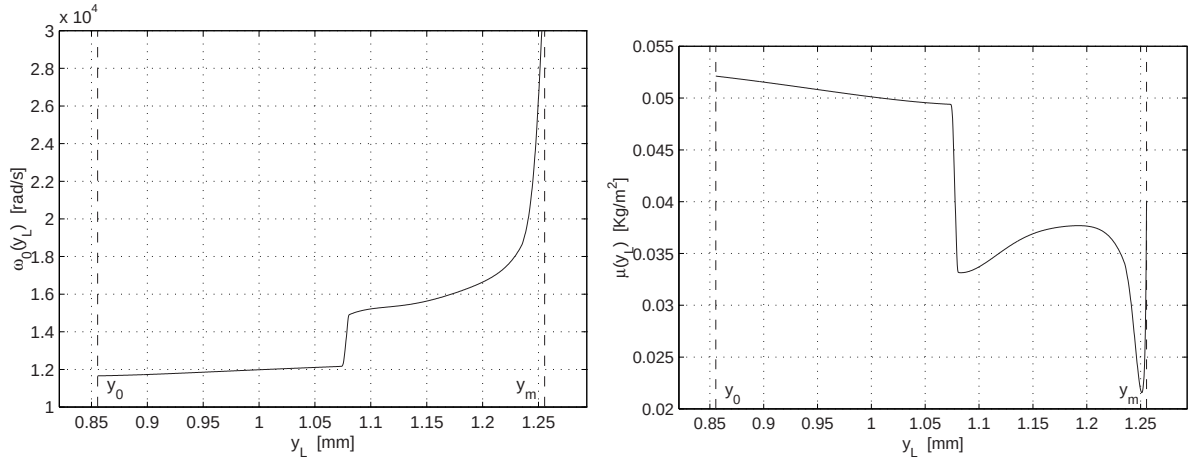

Figure 3.15: *Dependence of ω_0 and μ on the tip displacement.*

of the driving pressure Δp. Therefore, no delay-free paths are created in the difference equations when the non-linear reed is connected to a resonator. However, at each time step n nine iterations are typically required for computing iteratively the flow $u(n)$ with the Newton-Raphson method.

Summary

After reviewing the literature of lumped modeling for single reed systems, a simulation scheme has been proposed in Sec. 3.2. Through the analysis of the discrete frequency responses, it has been studied how theoretical properties of the numerical methods, such as stability, accuracy and frequency warping, affect the reed behavior in the digital domain. It has been shown that 1-step methods can approximate the system with good accuracy while keeping the computational costs low. Time-domain simulations have shown that the analysis of the reed frequency responses do not provide enough information on the properties of the whole system due to non-linear coupling in the equations. The discrete-time models have significantly different behaviors even when the frequency responses of the digital reeds are almost indistinguishable, and moreover the 2-step methods introduce artifacts in the digital domain. These two results show that the 1-step Weighted Sample method can be used as an efficient tool for sound synthesis purposes, in combination with the K method. Moreover, low sampling rates can be used without introducing instability or serious artifacts. This is important in applications such as Structured Audio coding (see Sec. 1.5), where instrument models are encoded and transmitted without precise knowledge of the computational power of the decoder. Sound examples obtained from numerical simulations show that the reed physical parameters allow effective and realistic control over the digital instrument.

Section 3.3 has proposed a distributed numerical model of the reed-lay-lip system.

It has been shown that the reed-lay, reed-lip interactions can be treated in a physically consistent manner by introducing appropriate conditional contact forces. The finite-difference scheme presented in Sec. 3.3.2 yields stable numerical simulations of the fourth-order reed equation and is suitable for the accurate simulation of the continuous system. The accuracy of the numerical system has been quantitatively assessed through convergence analysis. The effects of reed curling have been investigated using numerical simulations: it has been observed that the reed does not bend in a smooth way, and that the separation point undergoes a discontinuity during reed bending.

In Sec. 3.4, the distributed model has been employed as a numerical experimental set-up. In particular, it has been used to develop a non-linear lumped approximation in which the lumped parameters vary with reed tip displacement. The resulting lumped model is able to account for the effects of reed bending against the lay and of the player's lip. Due to its using a simple structure, it can be efficiently implemented in the discrete-time domain, with only a slight increase in the computational costs with respect to the model of Sec. 3.2.

Chapter 4

Source models for articulatory speech synthesis

Early research in analysis, synthesis and coding of voice has traditionally focused on the vocal tract filter, paying less attention to the source signal. Especially in the last decade, however, more emphasis has been given to the characteristics of the glottal source waveform: the development of a good model for the glottal excitation has been recognized to be a key feature for obtaining high quality speech synthesis, and for characterizing *voice quality* (e.g. modal voice, vocal fry, breathy voice [6, 32]).

Parametric models fit the glottal signal with piecewise analytical functions, and typically use a small number of parameters. As an example, the Liljencrants-Fant model [44] characterizes one cycle of the flow derivative using as few as four parameters. The Liljencrats-Fant model has been successfully used for fitting flow derivatives computed by inverse filtering real utterances [32, 101, 132].

Physical models describe the glottal system in terms of physiological quantities. The Ishizaka-Flanagan (*IF*) model [73] is a known example of lumped model of the vocal folds. Physical models capture the basic non-linear mechanisms that initiate self-sustained oscillations in the glottal system, and can simulate features (e.g. interaction with the vocal tract) that are not taken into account by parametric models. However they typically involve a large number of control parameters. As a consequence, the models are not easily controlled, since the model parameters do not map in an intuitive way into perceptual dimensions. Moreover, physical models are less suitable than parametric models for identification purposes.

Section 4.1 reviews the IF model in detail, and discusses the main qualities and drawbacks of lumped models of the glottis. Sections 4.2 and 4.3 describe two different approaches for simplifying the IF model. Both the proposed models rely on a common assumption, i.e. the vocal fold is treated as a single mass. However, the two models differ in the treatment of the non-linear block that accounts for the interaction between vocal folds and glottal airflow.

The *physically-informed* model presented in Sec. 4.2 describes the non-linear block

This chapter is partially based on [11, 41, 42].

using a black-box approach: no physical information is used to describe the interaction, instead it is modeled as a regressor-based mapping that relates the fold displacement to the glottal flow. This approach permits to use the model for identification purposes, i.e. the model parameters can be estimated in order to fit a given *target* glottal waveform. At the same time, the physical structure of the model can be exploited for controlling its behavior during resynthesis.

The *one-delayed-mass* model presented in Sec. 4.3 describes the non-linear block using a white-box (physical) approach. The pressure distribution along the glottis is modeled starting from the IF equations. The equations are then simplified by exploiting additional constraints in the vocal fold motion. As a result, the effects due to phase differences between the upper and lower margins of the folds are incorporated in the non-linear equations. Numerical simulations show the behavior of the one-delayed-mass model resembles that of the IF model, while using about half of the parameters and half of the parameters.

Table 3.1 summarizes the main variables and parameters that are used throughout the chapter.

4.1 Glottal models

Section 4.1.1 describes the Ishizaka-Flanagan model, where each vocal fold is treated using two lumped masses connected by springs and damping elements. Section 4.1.2 discusses the properties of the IF model and the use of lumped models for the synthesis of glottal signals.

4.1.1 The Ishizaka-Flanagan model

The human vocal folds consist of two opposing ligaments, that form a constriction at the beginning of the trachea. The orifice formed by the two folds is called the glottis. In the production of voiced sounds, the fundamental frequency and the quality of the sounds is to a large extent controlled by the quasi-periodic oscillations of the vocal folds. When sufficient subglottal pressure p_s is provided by the lungs, the folds start to vibrate and act as a pressure controlled valve, that tends to open for positive glottal pressures and to close for negative pressures (according to the notation introduced by Fletcher [48], this behavior corresponds to that of a $(+,+)$ valve). In normal phonation the folds typically collide with each other, and therefore the glottis is completely closed for a certain part of the oscillation. As a consequence, the quasi-periodic waveform describing the glottal airflow is made of a series of positive "puffs" produced when the glottis is open, and is zero when the glottis is closed.

The Ishizaka-Flanagan (IF) model describes each vocal fold using a two-mass approximation. Moreover, it assumes that the folds are bilaterally symmetric, so that only one of the needs to be modeled. As a consequence, the whole model is constructed using two masses, as depicted in Fig. 4.1. The masses are permitted only lateral motion (as in the coordinates x_1 and x_2 given in Fig. 4.1(b)). Along this lateral direction, the masses are

	quantity	symbol	unit
	Air shear viscosity	$\nu = 1.85 \cdot 10^{-5}$	[N· s/m^2]
	Vocal tract input area	$S = 5 \cdot 10^{-4}$	[m^2]
	Press. at tract entrance	$p(t)$	[Pa]
	Press. waves	$p^{\pm}(t)$	[Pa]
	Glottal flow	$u(t)$	[m^3/s]
	Vocal fold length	l_g	[m]
IF model			
	Vocal fold masses	$m_k, \quad k = 1, 2$	[Kg]
	Stiffness on m_k	$k_k(x_k)$	[N/m]
	Viscous resist. of m_k	$r_k = 0.1\sqrt{m_k k_k}$	[N· s/m]
	Thickness of m_k	d_k	[m]
	Displacement of m_k	$x_k(t)$	[m]
	Rest position of m_k	x_{0k}	[m]
	Glottal area under m_k	$A_k(t) = 2x_k(t)l_g$	[m^2]
	Area under m_k at rest	$A_{0k} = 2x_{0k}l_g$	[m^2]
1-mass models			
	Vocal fold mass	m	[Kg]
	Fold stiffness	k	[N/m]
	Fold viscous resist.	$r = 0.1\sqrt{m\ k}$	[N· s/m]
	Fold thickness	d	[m]
	Fold displacement	$x(t)$	[m]
	Fold rest position	x_0	[m]
	Glottal area	$A(t) = 2x(t)l_g$	[m^2]
	Glottal area at rest	$A_0 = 2x_0 l_g$	[m^2]

Table 4.1: *Symbols used throughout the chapter.*

assumed to behave as simple second-order mechanical oscillators, i.e. they are subject to elastic and dissipative forces. For the accurate simulation of the elastic properties of the fold, the springs are non-linear and $k_1(x_1)$ and $k_2(x_2)$ are modeled as quadratic functions of the corresponding displacements. In addition, the two masses are coupled through a third spring k_{12}.

Collisions between the folds are modeled by adding to the equations an additional restoring contact force, which is represented by an equivalent non-linear spring. In other words, when one of the masses m_k collides (i.e., when the condition $x_k < 0$ holds), its stiffness $k_k(x_k)$ increases. Summarizing, the equations for the mechanical system are

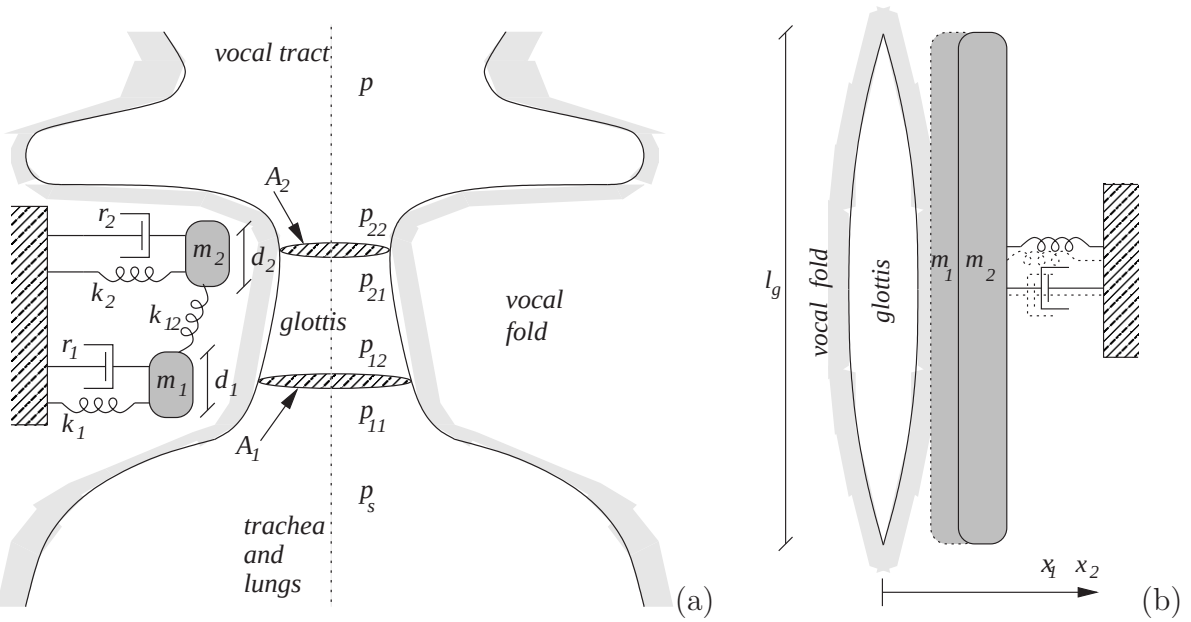

Figure 4.1: *Schematic representation of the Ishizaka-Flanagan model: (a) coronal view and (b) superior view.*

given by

$$\begin{cases} m_1\ddot{x}_1(t) + r_1\dot{x}_1(t) + k_1(x_1)[x_1(t) - x_{01}] + k_{12}[x_1(t) - x_2(t)] = l_g d_1 p_{m1}(t), \\ m_2\ddot{x}_2(t) + r_2\dot{x}_2(t) + k_2(x_2)[x_2(t) - x_{02}] - k_{12}[x_1(t) - x_2(t)] = l_g d_2 p_{m2}(t), \end{cases} \tag{4.1}$$

where $l_g d_1$ and $l_g d_2$ are the driving surfaces of the masses, on which the pressures p_{m1} and p_{m_2} act. The authors claim that using two masses can account for most of the relevant glottal detail, including the modeling of phase differences in the motion of the lower and upper edges of the folds.

The interaction of the mechanical model with the glottal pressure distribution is derived under the assumption of quasi-steady glottal flow. Ishizaka and Flanagan approximate the pressure distribution inside the glottis as successive discrete steps p_{ij} at each end j of each mass i (see Fig. 4.1(a)).

The first pressure drop $p_s - p_{11}$ is derived from the Bernoulli law for an ideal fluid in the static regime. According to the authors, the *vena contracta* produced by the abrupt contraction in the cross-sectional area makes the glottal area A_1 under the mass m_1 appears smaller than it is. As a consequence, the drop is greater than in the ideal case. Along the masses, the pressure drops $p_{11} - p_{12}$ and $p_{21} - p_{22}$ are governed by viscous losses and are proportional to the air shear viscosity ν. At the junction between the areas A_1 and A_2, the change in pressure $p_{12} - p_{21}$ equals the change in kinetic energy per unit volume of the fluid. Finally, the abrupt expansion at the upper edge of the

glottis causes the pressure to recover toward the atmospheric value. An estimate for the pressure recovery is found by Ishizaka and Flanagan by imposing continuity of the flow. Summarizing, the pressure distribution along the glottis is given by the following equations:

$$\begin{cases} p_s - p_{11}(t) = 0.69\rho_{air}\dfrac{u(t)^2}{A_1(t)^2}, \\ p_{11}(t) - p_{12}(t) = 12\nu d_1\dfrac{l_g^2 u(t)}{A_1(t)^3}, \\ p_{12}(t) - p_{21}(t) = \dfrac{1}{2}\rho_{air}u(t)^2\left(\dfrac{1}{A_2(t)^2} - \dfrac{1}{A_1(t)^2}\right), \\ p_{21}(t) - p_{22}(t) = 12\nu d_2\dfrac{l_g^2 u(t)}{A_2(t)^3}, \\ p_{22}(t) - p(t) = \dfrac{1}{2}\rho_{air}\dfrac{u(t)^2}{A_2(t)^2}\left[2\dfrac{A_2(t)}{S}\left(1 - \dfrac{A_2(t)}{S}\right)\right]. \end{cases} \tag{4.2}$$

The authors also discuss the inclusion of air inertance in the equations, when time-varying conditions are considered.

Given the pressure distribution of Eq. (4.2), the driving pressures p_{m1} and p_{m2} acting on the two masses have to be derived. In the IF model, these are simply assumed to be the mean pressures along each mass:

$$\begin{aligned} p_{m1}(t) = \frac{1}{2}[p_{11}(t) + p_{12}(t)] = \;& p_s - 0.69\rho_{air}\frac{u(t)^2}{A_1(t)^2} - 6\nu d_1\frac{l_g^2 u(t)}{A_1(t)^3}, \\ p_{m2}(t) = \frac{1}{2}[p_{21}(t) + p_{22}(t)] = \;& p(t) + \frac{1}{2}\rho_{air}\frac{u(t)^2}{A_2(t)^2}\left[2\frac{A_2(t)}{S}\left(1 - \frac{A_2(t)}{S}\right)\right] + \\ & + 6\nu d_2\frac{l_g^2 u(t)}{A_2(t)^3}. \end{aligned} \tag{4.3}$$

In conclusion, the IF model is completely described by Eqs. (4.1) and (4.3). Note that it is structurally very similar to the single reed model described in Sec. 3.1. The two main differences between the systems are that *(1)* the reed model uses only one mass while the IF model uses two, and *(2)* the single reed closes for positive pressures while the glottis opens.

4.1.2 Properties of lumped glottal models

The model described in the last section was implemented numerically by Ishizaka and Flanagan using the backward Euler method, and coupled to a vocal tract model. From the numerical simulations (run with sampling rates ranging from 10 to 30 [kHz]) the authors studied the dependence of the model behavior on the physical parameters and observed that realistic glottal signals can be obtained. The model can take into account subtle features that are not reproduced by a parametric model. In particular, acoustic

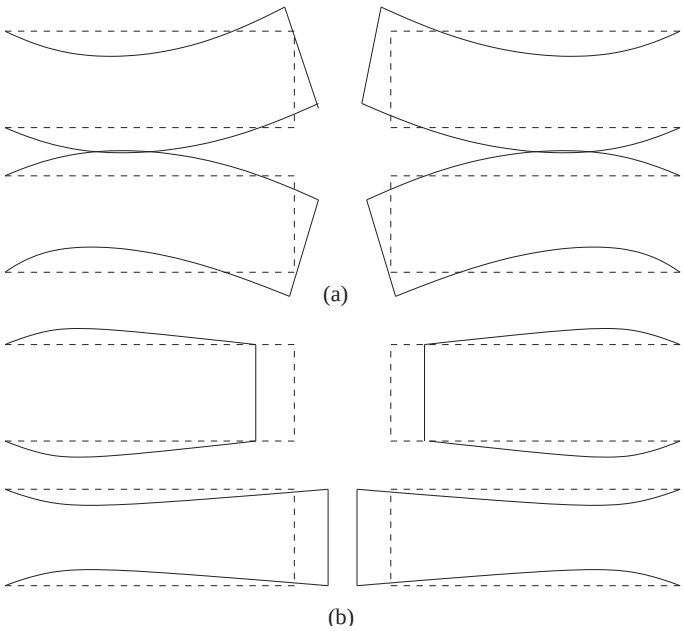

Figure 4.2: *First two excited modes in a distributed model of the vocal folds.*

interaction with the vocal tract is considered, thus allowing to develop a full articulatory model [129, 140]. This interaction gives rise to several "natural" effects: occurrence of oscillatory ripples on the glottal flow waveform, as well as a slight influence of the load characteristics on the pitch and the *open quotient* (i.e., the ratio between the open and closed phases in the glottal waveform).

Many refinements have been proposed to the IF model, in which the vocal folds are treated using a larger number of masses, or the description of the airflow though the glottis is modified. An example is the three-mass model by Story and Titze [131], where one large mass is used for describing the internal parts of the fold (the body), while two smaller masses account for the motion of the external part (the cover). On the other hand, simpler one-mass models are used by many authors in articulatory speech synthesizers (see e.g. [91]), or even in modeling non-human phonatory systems [47, 75]. One advantage of one-mass models lies in their reduced computational loads and better controllability. However, these models are not able to account for phase differences in the vocal fold motion.

Berry and Titze [21] studied the glottal system using a different modeling approach. In their study the vocal folds are modeled as a distributed, three-dimensional system, and is discretized using Finite Element methods. Using this numerical model the authors studied the normal modes of oscillation of the system. Assuming that the fold tissue is nearly incompressible, the first two excited modes in the model were found to be the ones depicted in Fig. 4.2.

Interestingly, the two masses in the IF model also have two eigenmodes which are conceptually equivalent to those of the distributed model. The IF mode where the two

masses move π-out of phase corresponds approximately to the one in Fig. 4.2(a), while the mode with the two masses in phase corresponds to that of Fig. 4.2(b). Berry and Titze suggest that the success of the IF model in describing the glottal behavior might be attributed to its ability to capture these two eigenmodes, and therefore facilitate self-oscillation.

A serious objection to all the models mentioned above has been raised by Villain *et al.* [150]. These authors remark that elementary mechanical constraints on the physiological problem are completely neglected in these models. As an example, in both the lumped and the distributed approach it is assumed that the elastic structure is fixed to a rigid wall, which is clearly a very crude approximation since in reality a significant radiation of surface waves from the throat can be noticed when voiced sounds are produced. The effect of this radiation may be significant in terms of energy loss in the system. Villain *et al.* used an experimental setup where the folds are modeled by thin latex tubes filled with water. Two mechanical boundary conditions were considered: one where the volume of the "water" fold is kept at a constant value and a second one where the internal water pressure remains constant independently on the blowing pressure. The experiments showed that the behavior of the latex valve is strongly affected by the mechanical constraints.

Another objection that can be raised to lumped models has to do with the glottal closure. As an example, in the IF model the glottal areas A_1 and A_2 are assumed to be rectangular. As a consequence, closure of the glottis occurs in an abrupt manner and the flow signals obtained from the model exhibit a sharp corner at the beginning of the closed phase. Equivalently, the airflow derivative exhibits a narrow negative peak at closure. This phenomenon affects the spectral tilt of the glottal source, introducing additional energy at high frequencies. In natural flow signals, a smoother glottal closure is usually observed. For example, stroboscopic measurements often show zipper-like movements of the glottal area during the closing phase. The IF model clearly does not take into account these phenomena.

Finally, when used in speech synthesis applications the IF model suffers from an over-parametrization: as many as 19 parameters have to be estimated in order to account for non-linear corrections in the elastic forces, for collisions between the two folds, and other features. This results in problems in tuning the parameters. Proposed refinements to IF (such as the three-mass model by Story and Titze [131]) involve an even larger number of parameters and are hardly controllable and more computationally expensive.

4.2 Non-linear block: identification

The model presented in this section relies on a hybrid approach. Its structure is similar to that of IF, but not all of the functional blocks are modeled following through a physical description.

The vocal fold is treated as a lumped mass subject to an elastic restoring force and to dissipation. The mass is driven by the pressure p_g at glottis. Therefore, in this

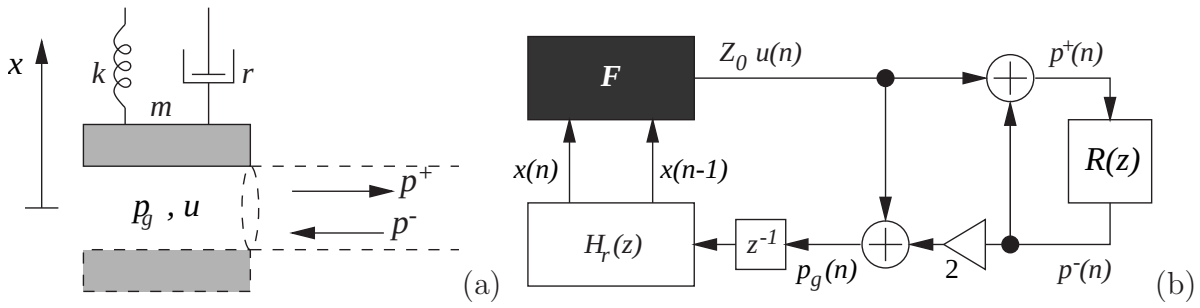

Figure 4.3: *One-mass vocal fold; (a) schematic representation and (b) block diagram of the model.* $Z_0 = \rho_{air} c/S$ *is the wave impedance of the vocal tract.* H_r, F, R *are the vocal fold oscillator, the non-linear black box, and the vocal tract reflectance, respectively. Other symbols are listed in Table 4.1.*

approximation the displacement x of the mass is governed by the second-order linear oscillator equation:

$$m\ddot{x}(t) + r\dot{x}(t) + k(x(t) - x_0) = l_g d\ p_g(t), \tag{4.4}$$

where $l_g d$ is the driving surface on which the pressure p_g acts.

A non-linear block accounts for interaction with glottal pressure. Unlike IF, however, no physical information is retained in the non-linear block. This is treated as a *black box* and described by a regressor-based mapping. As such, the model can be said to be *physically-informed* rather than really physical. This permits to exploit advantages of both the parametric and the physical approach. Namely, given a glottal flow signal the weights for the regressors can be estimated using non-linear identification techniques, in order to fit the waveform.

4.2.1 The identification procedure

In the following, the model is described in the discrete-time domain. A schematic representation and a block diagram are depicted in Fig. 4.3. Analogously to the IF model, it is assumed that the valve is perfectly symmetrical, so that only one fold needs to be modeled.

In the discrete-time domain, each vocal fold is described during the open and closed phases as a linear second-order oscillator, whose transfer function is denoted by H_r and given by

$$H_r(z) = \frac{\beta_0}{(1 + \alpha_1 z^{-1} + \alpha_2 z^{-2})}. \tag{4.5}$$

The filter $H_r(z)$ is the digital equivalent of the continuous-time system (4.4). It is completely defined by its gain factor β_0, its center frequency ω_0, and its quality factor q_0. Given the parameters (ω_0, q_0), the poles $p_d = re^{\pm j\phi}$ of the filter H_r are defined in the

following way [92]:

$$r = 1 - \frac{\omega_0}{2q_0}, \qquad \cos(\phi) = 2r\frac{\cos(\omega_0)}{1+r^2}.$$

The coefficients a_2, a_3 in the denominator of H_r are then found as $a_2 = -2r\cos(\phi)$ and $a_3 = r^2$.

The fold displacement x is related to the glottal pressure p_g through the equation $X(z) = H_r(z)P_g(z)$. Collisions between the two folds are treated using an approach analogous to that adopted in the IF model: during collisions, an elastic restoring contact force is added to the mass. Equivalently, during collisions the resonance ω_0 of H_r is increased.

The glottal flow is assumed to depend non-linearly on the glottal area, or equivalently on the fold displacement. A shown in Sec. 4.1, physical models such as IF describe this dependence analytically using very crude simplifications (e.g. quasi-steadiness of the flow). A different viewpoint is taken here: no attempt is made to model the glottal flow behavior in a physical way, and instead the dependence of the glottal flow on the displacement is modeled using a black-box approach. In this approach, a non-linear regressor-based mapping F relates flow to fold displacement. The inputs to the non-linear block are $x(n)$ and $x(n-1)$:

$$Z_0 u(n) = \begin{cases} F(n) = \sum_{i=0}^{M} w_i \psi_i(n), & \text{if} \quad x(n) > 0, \\ 0, & \text{if} \quad x(n) \leq 0, \end{cases} \tag{4.6}$$

where $Z_0 = \rho_{air} c/S$ is the wave impedance of the vocal tract, while the regressors are denoted as $\psi_i(n) = \psi_i\left(x(n), x(n-1)\right)$.

Several strategies are possible for the choice of the regressor set $\{\psi_i\}$. Local models, such as gaussian functions or any other radial basis function, are often used. Radial Basis Function Networks *(RBFN)* [31] are often used in the field of time series analysis and modeling. Alternatively, a polynomial expansion of the input can be used. This choice leads to a class of so-called NARMAX models [30], known in the fields of system identification and control. Finally, the regressors can be derived on the basis of physical considerations.

In the following, a third order polynomial expansion in $x(n), x(n-1)$ is used. One reason for using $x(n-1)$ is that the vocal fold velocity $\dot{x}$ is assumed to contribute to the total flow. This assumption is made in analogy to the single reed model described in chapter 3 (see in particular Eq. (3.5)). Taking into account $\dot{x}$ corresponds, in the discrete-time domain, to taking into account at least one past value of x. This qualitative discussion justifies the inclusion of the term $x(n-1)$ as an input variable to the regressor set $\{\psi\}_i$.

Interaction with the vocal tract is drastically simplified by neglecting pressure recovery at the vocal tract entrance. In other words, it is assumed that the pressure p at the vocal tract entrance equals the glottal pressure p_g. Again, the vocal tract model can only be said to be physically-informed, and not physical, since the assumption $p_g = p$ has

no counterpart in the physical reality. Under this assumption, the glottal pressure p_g is related to the glottal flow through the input impedance $Z_{in}(z)$ of the vocal tract. Equivalently, wave variables $p^{\pm} = (p_g \pm Z_0 u)/2$ can be used: then $p^{\pm}$ are related to each other through the vocal tract reflectance $R(z)$:

$$p_g(n) = z_{in}(n) * u(n), \quad \Leftrightarrow \quad p^-(n) = r(n) * p^+(n). \tag{4.7}$$

A final remark about Fig. 4.3 concerns the insertion of a fictitious delay element z^{-1}: this is needed in order to compute the delay-free feedback path in the numerical model. The K method has been used elsewhere in this thesis for dealing in a more accurate way with such computability problems. However, in this case the insertion of z^{-1} does not deteriorate or anyhow affect the properties of the numerical system. The reasons for this are explained at the end of this section.

The following problem is now addressed: given a *target* glottal flow waveform $Z_0\bar{u}$ (be it a synthetic signal or an inverse-filtered one), the physically-informed model has to be identified so that the output $Z_0 u$ from the non-linear block fits the target as closely as possible. Many works are available in the literature [32, 101, 132] where this problem has been studied using parametric models of the glottal source (e.g. the Liljencrants-Fant model). In the case under consideration here, a physical model of the source has to be identified rather than a parametric model of the signal, which makes the problem not trivial.

The identification procedure described in the following holds for both synthetic and real flow waveforms. In both cases a single period is chosen and a periodicized signal is constructed and used as target. The example utterance analyzed in this section is a sustained /a/ vowel produced in normal phonation by a male speaker. The inverse-filtered flow signals has been computed using an automatic method developed by Alku and described in [5]. This method estimates the glottal flow directly from the acoustic speech pressure waveform using a two-stage structure, where LP analysis is used as a computational tool.

The system is identified in three main steps.

Step 1 From the target $Z_0\bar{u}$, the corresponding glottal pressure signal $\bar{p}_g$ is computed using Eq. (4.7). In order to do that, the reflectance $R(z)$ of the vocal tract has to be known. In the following R is arbitrarily chosen to be that of a uniform vocal tract, and is implemented using waveguide modeling. Therefore, $R(z) = z^{-2m_L} R_{\text{load}}(z)$, where m_L defines the length (in samples) of the tract and R_{load} accounts for the reflection at mouth and has a low-pass characteristics. Note that the above defined $R(z)$ is formally identical to the one used for the idealized clarinet bore in Sec. 3.2 (see in particular Eq. (3.14)).

Step 2 The linear block $H_r(z)$ is driven using the synthesized target pressure $\bar{p}_g$, and the output $\bar{x}$ is computed. At this stage, resonances for H_r in the open phase and the closed phase are chosen interactively, in such a way that the open and closed phase for the target fold displacement $\bar{x}$ match those of the target flow $\bar{u}$ (in other

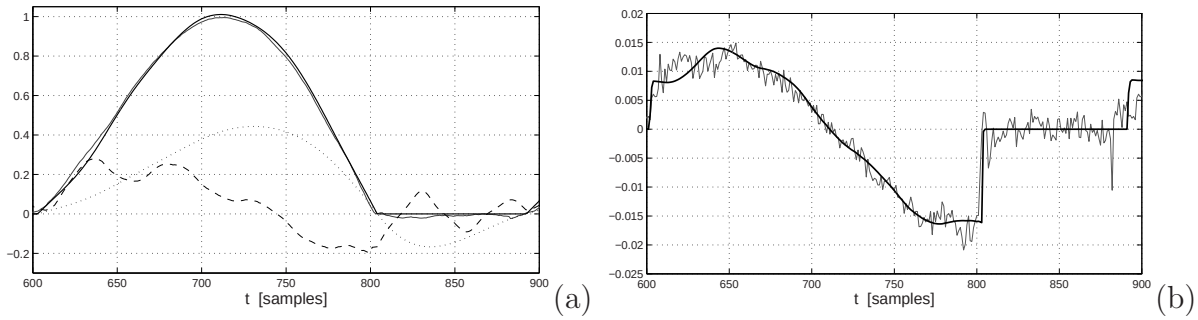

Figure 4.4: *The identification procedure; (a) target $Z_0\bar{u}$ (solid gray, computed from real speech by inverse filtering), synthesized pressure $\bar{p}_g$ (dashed) after Step 1, output $\bar{x}$ from linear block (dotted) after Step 2, output $Z_0 u$ from non-linear block (solid black), after Step 3; (b) derivative of the target flow $d\bar{u}/dt$ (solid gray) and derivative of output from the non-linear block du/dt (solid black).*

words, $\bar{x} = 0 \Leftrightarrow \bar{u} = 0$, see the example in Fig. 4.4). The value $q_0 = 10$ is chosen for the quality factor. This value is deduced from the parameters used by Ishizaka and Flanagan in [73].

Step 3 A complete input-output description of the non-linear block is available at this point, where the inputs are $\bar{x}(n), \bar{x}(n-1)$ and the output is the target $\bar{u}(n)$. Therefore, during the open phase the weights $\boldsymbol{w} = [w_0 \dots w_M]$ for the regressor set $\{\psi\}_i$ can be identified using the non-linear identification technique described in the following.

The assumption made in *Step 1* is questionable, since fro real utterances the vocal tract cross-section is obviously not uniform. However, in the physically-informed approach adopted here a pressure signal that provides *plausible* rather than realistic excitation to the vocal fold is needed. Figure 4.4(a) summarizes the three steps of the identification procedure. Figure 4.4(b) shows the time-derivatives of the target flow $\bar{u}$ and the identified flow u.

The identification of the weights $\boldsymbol{w}$ in *Step 3* is carried out as follows. A training window with starting time l and length N is chosen. Inside the window, two training sets $\boldsymbol{T}_u$ and $\boldsymbol{T}_\psi$ are defined as

$$\begin{aligned} \boldsymbol{T}_u &= [\bar{u}(l+1), \bar{u}(l+2), \dots, \bar{u}(l+N)]\,, \\ \boldsymbol{T}_\psi &= \begin{bmatrix} \psi_0(l+1) & \cdots & \psi_0(l+N) \\ \vdots & \ddots & \vdots \\ \psi_M(l+1) & \cdots & \psi_M(l+N) \end{bmatrix}. \end{aligned} \tag{4.8}$$

Using these definitions and Eq. (4.6), it is seen that the weights $\boldsymbol{w}$ must solve the system

$\boldsymbol{w} \cdot \boldsymbol{T}_\psi = \boldsymbol{T}_u$. The least-squares solution of such a system is known [83] to be

$$\boldsymbol{w} = \boldsymbol{T}_u \cdot \boldsymbol{T}_\psi^+, \tag{4.9}$$

where the symbol + has the meaning of *pseudo-inversion*.

It is now clear that the insertion of z^{-1} does not deteriorate the accuracy of the model: given the structure in Fig. 4.3 and a flow signal, the identification of the non-linear block automatically takes into account the z^{-1} element.

4.2.2 Results and applications

The identification procedure described in the last section has been tested using both real and synthetic signals. The performance is generally good and comparable to that obtained using signal-based models [32, 101, 132]. In particular, from the example in Fig. 4.4(b) one can see that in the closing phase both the width and the amplitude of the negative pulse in the flow derivative are well approximated. This portion is the most important to be fitted accurately, since it defines most of the spectral (and perceptual) features of the glottal source signal.

Moreover, large bandwidths and the consequent noise in the flow derivative waveform are seen not to affect the identification procedure. For the signal used in Fig. 4.4, the bandwidth is 11.025 [kHz], and considerable noise can be noticed in the flow derivative signal. However, the proposed identification procedure uses the glottal flow –rather than its derivative– as the target signal, while signal-based models (such as Liljencrants-Fant) typically try to fit the derivative of the glottal flow. It has been verified that large bandwidths can deteriorate the performance of identification techniques based on the Liljencrants-Fant model [101].

However, problems are encountered when strong ripples (due to interaction with tract formants) appear on the opening phase of the target signal: these can deteriorate the accuracy of identification drastically. This problem can be solved by focusing the accuracy of the identification on the portions of the signal (such as the closing phase) that are known to be the most important, and to loosen the accuracy requirements for those features (such as the ripples in the opening phase) that are known to be perceptually less relevant. One possible strategy for achieving this goal is to pre-process the target signal by applying time-warping, such that the relevant portions are magnified while the rest of the signal is time-compressed. However, at present no attempt has been made to implement this strategy.

The target signal to be identified is obtained through periodicization of one cycle of the flow waveform. Therefore one open problem is concerned with identification of non-periodic signals. A straightforward strategy is to identify the system parameters once for each period of analyzed data (as already done by Childers and Ahn [32] with the Liljencrants-Fant model). Alternatively, a single set $\boldsymbol{w}$ of weights may be used while adjusting only the filter parameters ω_0 and q_0 in order to account for changes in pitch and amplitude. However, one further problem is that the identification procedure is far from being automatic: while the weights $\{w_i\}$ are estimated automatically from Eq. (4.9),

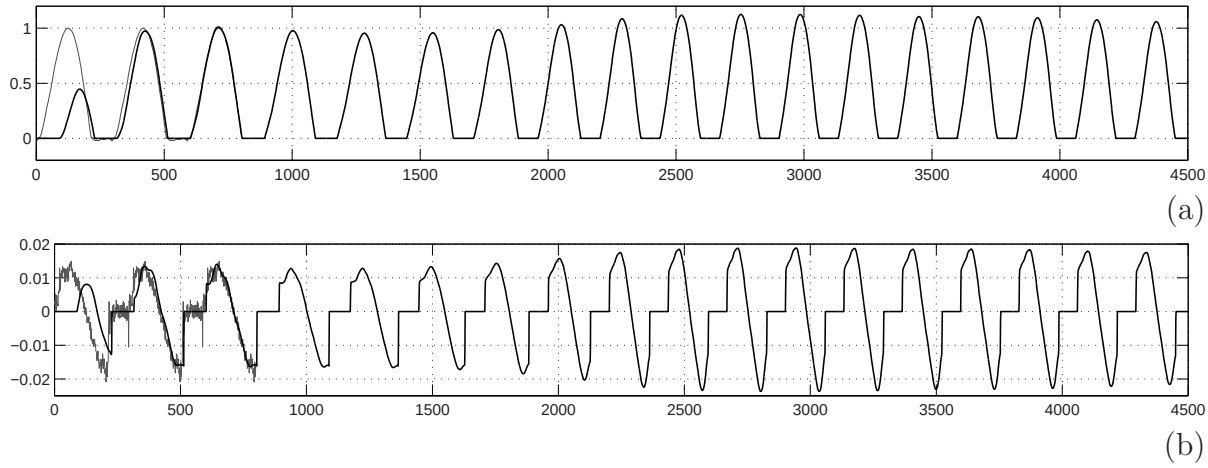

Figure 4.5: *Resynthesis and pitch shift; (a) target flow $\bar{u}$ (solid gray) and synthesized flow u (solid black); (b) target flow derivative $d\bar{u}/dt$ (solid gray) and synthesized flow derivative du/dt (solid black).*

the linear filter H_r is adjusted interactively. A technique for automatic estimation of the filter parameters has to be found in order to extend the identification procedure to the non-periodic case. The remaining of this section addresses the use of the proposed model for speech synthesis purposes and for voice quality assessment.

The physically-informed model can be used after identification for resynthesizing the target glottal flow signals. When studying the system behavior in autonomous evolution, it is seen that it reaches steady state self-sustained oscillations after a short transient. The steady state waveform coincides with the one obtained from identification. If the values of some of the model parameters are then adjusted, the system exhibits robustness to such changes. Figure 4.5 shows an example where ω_0 is increased after the system has reached the steady state: as a consequence the pitch of the signal increases correspondingly, while the flow shape is preserved. An increase in amplitude during pitch transition can also be noticed from Fig. 4.5. After the transition, the maximum amplitude turns back to its original value.

This example shows that the physically-informed parameters of the model can be used to perform transformations on the resynthesized signal (as an example, pitch shift or vibrato effects are obtained by adjusting ω_0). A major limitation is that the identification procedure is not able to guarantee *a priori* that the identified system is stable during resynthesis. Different approaches can be used in order to guarantee stability, such as the harmonic balance technique [64] or the imposition of additional constraints on the gradient ∇F [40].

A second potential application of the model is in the analysis and assessment of voice quality, including the detection and classification of voice pathologies. Voice quality assessment is traditionally based on subjective perceptual rating, which is still in current studies considered the only reasonable way to classify certain types of voice disorders.

The objective assessment of voice however, is still an open problem that calls for reliable analysis tools and algorithms. Many researchers [33, 60] use a set of parameters derived from analysis of the acoustic signal (i.e., the radiated pressure): examples of such parameters are time-domain measures such as *jitter* (pitch variation in successive oscillation periods) and *shimmer* (magnitude variation in successive oscillation periods), or frequency-domain parameters such as the spectral slope and the spectral flatness of the inverse-filtered signal. The main advantage of these techniques lies in their non-invasive nature.

However, all of the above mentioned parameters depend exclusively on the signal features, and no assumptions are made on the physiology of the source. As a consequence, signal-based analysis retains a mixed information in which the contributions of the vocal tract and the glottis are undistinguishable. Using a physically-informed model permits to localize the observation at the non-linear excitation mechanism, limiting the influence of other elements. The values of the identified weights $\boldsymbol{w}$ provide information on the glottal flow waveform, while analysis of their variations in time can be used to study the stability of the waveform. Early results on voice quality assessment using the proposed physically-informed model have been presented in [42].

4.3 Non-linear block: modified interaction

In Sec. 4.2, the IF model has been simplified by treating the vocal fold as a single mass, and by describing the interaction with the glottal flow using a black-box approach. The model presented in this section follows a different strategy. The vocal fold is still treated a single mass, but the interaction with the airflow is modeled through a physical description. The effect of the second mass of the IF model is taken into account by introducing a delay t_0 in the mass position. The glottal airflow is assumed to depend on this "delayed mass", and the non-linear aerodynamics equations are consequently modified.

Results from the simulations presented in Sec. 4.3.2 show that the model behaves qualitatively as IF, while using only one degree of freedom (one mass) instead of two. As a consequence, less than half of the IF parameters are needed. Among them, the delay t_0 gives control on the airflow skewness, which is known to be a perceptually relevant feature [32]. Having a small set of meaningful control parameters, the proposed physical model can be "competitive" with parametric ones, such as Liljencrants-Fant.

4.3.1 A one-delayed-mass-model

As shown in Sec. 4.1.1, Ishizaka and Flanagan describe the pressure drops p_{ij} along the vocal folds according to Eq. (4.2). Therefore the positions x_1 and x_2 of both masses are needed in order to compute the pressure drops and the resulting airflow. The "one-delayed-mass model" presented here avoids the use of a second mass by exploiting additional information on the system.

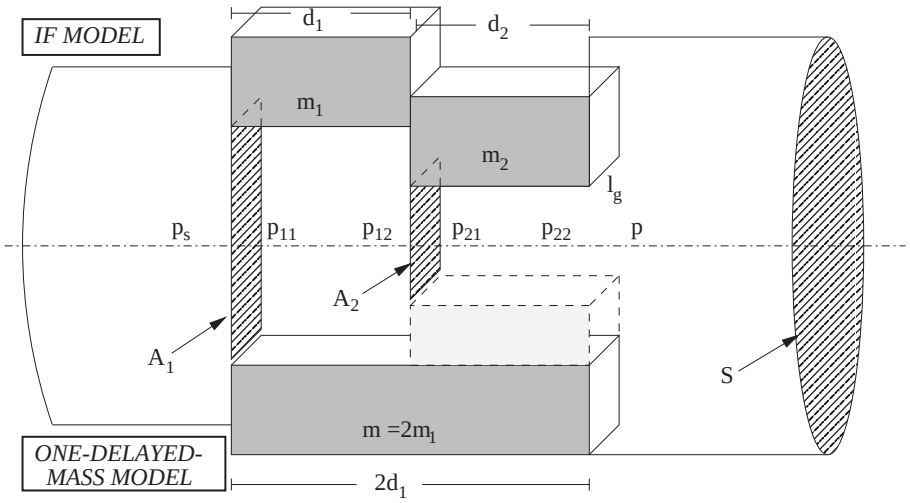

Figure 4.6: *Scheme of the one-delayed-mass model (lower half) as opposed to the IF model (upper half).*

- The IF model has two eigenmodes: the one with two masses in phase and the one with two masses π-out of phase. As already remarked in Sec. 4.1.2, these modes correspond roughly to the first two excited modes observed by Berry and Titze [21] using a distributed model of the vocal folds (see Fig. 4.3). Berry and Titze found that the two eigenfrequencies are very closely spaced. As a consequence, $1:1$ mode locking occurs during self-oscillation.

- In a recent paper, de Vries *et al.* [36] used a similar distributed model for estimating "correct" values for the IF parameters. Such values are found by requiring the behavior of the IF model to resemble as closely as possible that of the distributed model. Their results differ significantly with the values originally stated by Ishizaka and Flanagan. In particular the parameter values for the two masses are found in [36] to be much more symmetrical: the ratio between m_1 and m_2 is close to one (while it is close to five in the IF parameters), and the same holds for the spring constants, damping factors and geometrical parameters (d_1 and d_2 in Fig. 4.1 are found to be the same).

Using this additional information, the IF model can be consistently simplified using the following assumptions.

a1. The masses $m_{1,2}$ are taken to be equal, together with their thickness $d_{1,2}$ and their spring constants and damping factors.

a2. The two masses are taken to move with constant phase difference, because of mode locking; this means that the area $A_2(t)$ under the second mass follows the first on $A_1(t)$ with a constant phase difference.

The assumption *a2* can be restated as:

$$A_2(t) = A_1(t - t_0), \tag{4.10}$$

where t_0 represents the delay (in seconds) between the motion of the upper and lower edges of the fold. As a consequence, the pressure distribution p_{ij} in Eqs. (4.2) can be written as

$$\begin{cases} p_s - p_{11}(t) = 0.69\rho_{air}\dfrac{u(t)^2}{A(t)^2}, \\ p_{11}(t) - p_{12}(t) = 12\nu d_1\dfrac{l_g^2 u(t)}{A(t)^3}, \\ p_{12}(t) - p_{21}(t) = \dfrac{1}{2}\rho_{air}u(t)^2\left(\dfrac{1}{A(t-t_0)^2} - \dfrac{1}{A(t)^2}\right), \\ p_{21}(t) - p_{22}(t) = 12\nu d_2\dfrac{l_g^2 u(t)}{A(t-t_0)^3}, \\ p_{22}(t) - p(t) = \dfrac{1}{2}\rho_{air}\dfrac{u(t)^2}{A(t-t_0)^2}\left[2\dfrac{A(t-t_0)}{S}\left(1 - \dfrac{A(t-t_0)}{S}\right)\right], \end{cases} \tag{4.11}$$

where $A(t) \equiv A_1(t)$. In Eqs. (4.11), the pressure drops are non-linear functions of the area A and the same area delayed by t_0. This suggests that only one degree of freedom is needed in the model.

Therefore, in the following the fold is described using a single mass and is treated as a linear second-order oscillator. Similarly to IF, the driving pressure acting on the fold is chosen to be the mean pressure p_m at the glottis: $p_m = 1/4\sum_{i,j=1}^{2} p_{ij}$.

$$m\ddot{x}(t) + r\dot{x}(t) + k(x(t) - x_0) = l_g d p_m(t), \tag{4.12}$$

where $l_g d$ is the driving surface on which the pressure drop acts. From the assumption *a1*, the mass m is given by $m = 2m_1$. Explicit expressions for the driving pressure p_m and the pressure at vocal tract entrance p are derived from Eq. (4.11), and depend only on the variables $(A(t), A(t-t_0), u(t))$:

$$\begin{cases} p_m(t) = p_m(A(t), A(t-t_0), u(t)), \\ p(t) = p(A(t), A(t-t_0), u(t)). \end{cases} \tag{4.13}$$

In this way, the effects due to phase differences between the upper and lower margins of the folds are incorporated in the non-linear equations (4.13), and are controlled by the delay t_0. One last equation relates the glottal flow to the pressure p at the vocal tract entrance:

$$u(t) = z_{in}(t) * p(t), \tag{4.14}$$

where the load impedance z_{in} is the input impedance of the vocal tract.

Equations (4.12),(4.13),(4.14) describe the one-delayed-mass model. From Eq. (4.12), it is seen to be a one-mass model, but the dependence on the delayed area $A_1(t-t_0)$ in Eq. (4.13) results in a modified non-linear block. A graphic representation of the model, as opposed to IF, is depicted in Fig. 4.6. As shown in the following section, by introducing the parameter t_0 it is possible to preserve the main features of a two-mass model while using a single degree of freedom.

4.3.2 Numerical simulations

The linear differential Eq. (4.12) is discretized using the bilinear transform, i.e. each occurrence of s in the Laplace-transformed Eq. (4.12) is substituted with the mapping (see Sec. 2.3.1)

$$s = 2F_s \frac{1-z^{-1}}{1+z^{-1}}. \tag{4.15}$$

It is easily seen that the resulting numerical system resembles the general structure described in Sec. 2.3.2 and depicted in Fig. 2.11. Therefore, a delay-free path is generated between the linear and the non-linear blocks of the system:

$$\begin{cases} \boldsymbol{w}(n) &= \tilde{\boldsymbol{w}}(n) + \bar{\boldsymbol{C}}\boldsymbol{p}(n), \\ \boldsymbol{x}(n) &= \tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\boldsymbol{p}(n), \\ \boldsymbol{p}(n) &= \boldsymbol{f}\left(\tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\boldsymbol{p}(n)\right), \end{cases} \tag{4.16}$$

where the variables are given by

$$\boldsymbol{w}(n) = \begin{bmatrix} x(n) \\ \dot{x}(n) \end{bmatrix}, \quad \boldsymbol{u}(n) = \begin{bmatrix} x_0 \\ p_s \\ p^-(n) \end{bmatrix}, \quad \boldsymbol{x}(n) = \begin{bmatrix} u(n) \\ x(n) \\ x(n-n_0) \end{bmatrix}, \quad \boldsymbol{p}(n) = \begin{bmatrix} p_m(n) \\ p(n) \end{bmatrix},$$

and where the numerical delay n_0 is defined as $n_0 = t_0 F_s$. The vectors $\tilde{\boldsymbol{w}}(n)$ and $\tilde{\boldsymbol{x}}(n)$ have no instantaneous dependence on the pressures $\boldsymbol{p}(n)$, and are therefore computable at each step. The non-linear mapping $\boldsymbol{f} : \mathbb{R}^2 \to \mathbb{R}^2$ is simply Eq. (4.13) restated in vector notation.

The K method is used in order to compute the delay-free path. The mapping $\boldsymbol{f}$ is sheared and the non-linear equation in system (4.16) is turned into a new-one:

$$\boldsymbol{p}(n) = \boldsymbol{f}\left(\tilde{\boldsymbol{x}}(n) + \boldsymbol{K}\boldsymbol{p}(n)\right) \quad \overset{\text{K method}}{\longmapsto} \quad \boldsymbol{p}(n) = \boldsymbol{h}\left(\tilde{\boldsymbol{x}}(n)\right). \tag{4.17}$$

The new mapping $\boldsymbol{h}(\tilde{\boldsymbol{x}})$ is computed iteratively at sample rate, according to the discussion at the end of Sec. 2.3.2. The numerical implementation described above is used to study the properties of the one-delayed-mass model.

The dependence on $A_1(t-t_0)$ in Eq. (4.13) results in an additional delay loop in the system. This is a potential source of instability [7]. Due to the non-linear nature of

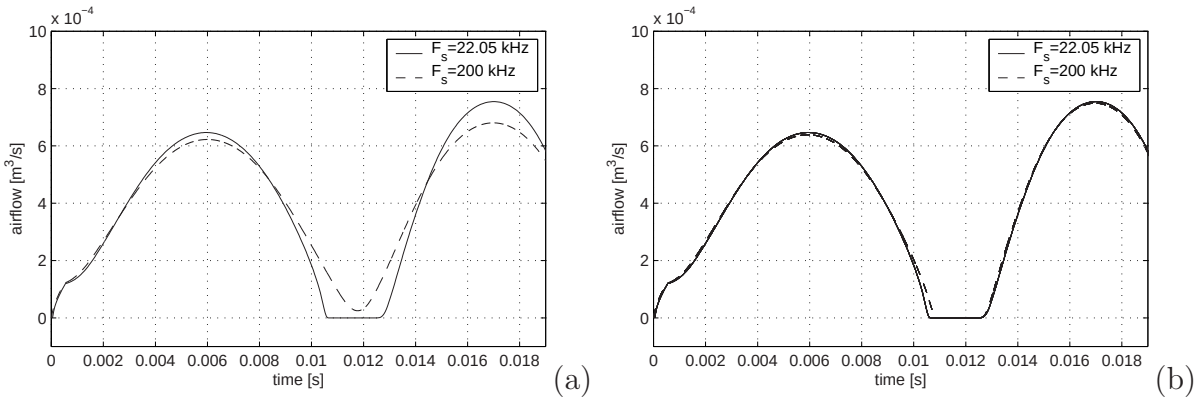

Figure 4.7: *Attack transients in the one-delayed-mass model; (a)* $t_0 = 1 \cdot 10^{-4}$ *[s], the numerical system at* $F_s = 22.05$ *[kHz] is unstable; (b)* $t_0 = 2 \cdot 10^{-4}$ *[s], the numerical system at* $F_s = 22.05$ *[kHz] is stable.*

the system, analytical conditions for stability are not easily found. Therefore, stability properties of the system have to be investigated experimentally. This can be done in two steps: first, simulations are run at a very high sampling rates ($F_s = 200$ [kHz]), and these simulations are taken as a reference for the behavior of the continuous-time system. In a second step, simulations are run at standard sampling rates ($F_s = 11.025, 22.05$ [kHz]) and compared with the reference.

The results for the case $F_s = 22.05$ [kHz] are plotted in Fig. 4.7: these show that the numerical delay $n_0 = t_0 F_s$ affects the system's stability. With very small delays ($t_0 < 2 \cdot 10^{-4}$ [s], i.e. $n_0 < 4$ at $F_s = 22.05$ [kHz]) the system is unstable, as seen from Fig. 4.7(a): the first few cycles in the oscillation show the increasing error with respect to the reference. In the following cycles, this trend continues until a steady state far from the reference is reached. In general, results from the simulations showed that for values $n_0 > 4$ the system appears to be stable. Figure 4.7(b) shows that with a delay $t_0 = 2 \cdot 10^{-4}$ [s] (i.e. $n_0 = 4$) the system at $F_s = 22.05$ [kHz] is stable. Analogous results are found for $F_s = 11.025$ [kHz].

A second experimental study with numerical simulations examines the effect of the delay t_0 in shaping the glottal waveform. Figure 4.8 shows the areas $A_1(t), A_1(t - t_0)$ and the airflow $u(t)$ for two different values of t_0 (for clarity, the signals are normalized in the figure). From these plots, it can be clearly seen that the skewness of the airflow is controlled by the delay t_0. A quantitative measure of the flow skewness is given by the *speed quotient*. This parameter is defined as the ratio between the opening phase ($\dot{u}(t) > 0$) and the closing phase ($\dot{u}(t) < 0$). The speed quotient is known to have perceptual relevance in characterizing different voice qualities: for instance, analysis on real signals by Childers and Ahn [32] show that the speed quotient ranges from about 1.6 to 3 when the voice quality changes from breathy voice to vocal fry and finally to modal voice.

In order to investigate quantitatively the influence of the delay t_0 on the signal pa-

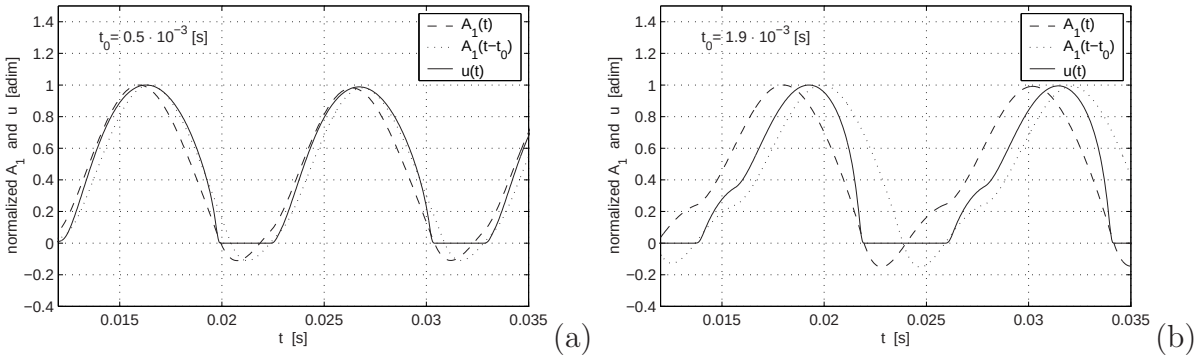

Figure 4.8: *Dependence of the flow skewness on the time delay t_0. Simulations are run at $F_s = 22.05$ [kHz].*

rameters (such as pitch, open quotient, speed quotient, maximum amplitude), automatic analysis of numerical simulations has to be developed. The following results are obtained by analyzing 0.3 [s] long flow signals, where the values of t_0 range from 0.1 to 1.9 [ms]. Figure 4.9(a) shows the dependence of the speed quotient on t_0: it is seen that, in the range under consideration, the speed quotient is approximately a linear function of t_0. By appropriately choosing t_0, one can range from very low up to extremely high values of the speed quotient.

Figure 4.9(b) shows another interesting feature of the system: the maximum amplitude for u exhibits a peak around $t_0 = 8 \cdot 10^{-4}$ [s]. This suggests the existence of an optimum delay t_0 that maximizes the aerodynamic input power (defined as mean subglottal pressure times mean glottal flow). The aerodynamic input power is in turn related to the glottal efficiency, usually defined as the ratio of radiated acoustic power to aerodynamic power (i.e., the power delivered to the vocal system by the lungs). Further analysis is needed in order to assess the precise influence of t_0 on the glottal efficiency.

The main advantages of the proposed model are its simple structure and its low number of control parameters. On the one hand, only one degree of freedom is needed, instead of two [73] or more [131] usually assumed in higher-dimensional lumped models of the vocal folds. On the other hand, the dependence on t_0 in Eq. (4.13) results in realistic glottal flow waveforms, that are not obtained with usual one-mass models [91]. In particular, the results given in this section show that t_0 provides control on the airflow skewness. The model is therefore a reasonable trade-off between accuracy of the description and simplicity of the structure.

Interaction of the model with vocal tract loads has not yet been investigated in detail. Preliminary results have been obtained by coupling the one-delayed-mass model with a uniform tube model, implemented as a digital waveguide line. The waveguide implementation is structurally identical to that used in chapter 2 for an ideal clarinet bore (see Eq. (3.14)). The simulations show the occurrence of ripples in the airflow signal, mainly due to interaction with the first resonance of the tract. Moreover, automatic analysis reveals a slight dependence of pitch on the vocal tract characteristics (length of the tube,

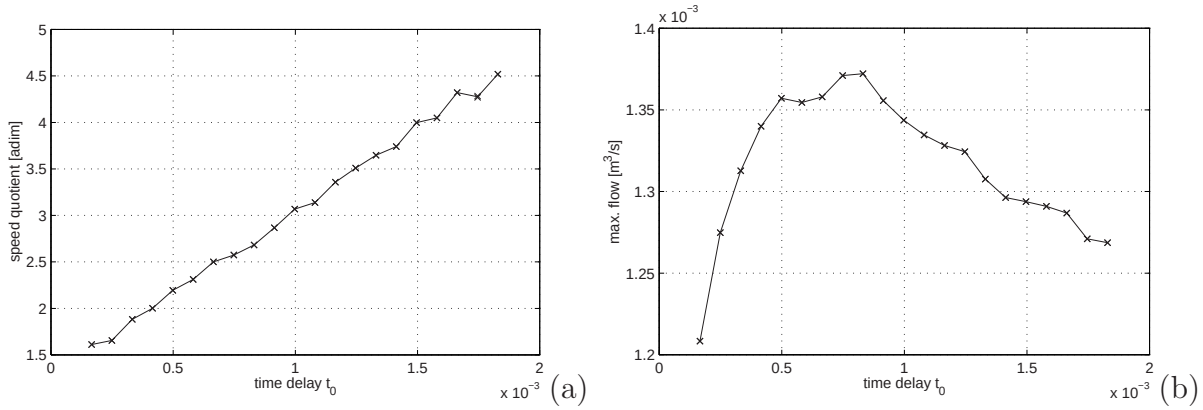

Figure 4.9: *Dependence of (a) the speed quotient SQ and (b) the maximum amplitude on t_0 for the airflow ($F_s = 22.05$ [kHz], $p_s = 1000$ [Pa]).*

low-pass characteristics of the reflection filter). Further efforts have to be devoted to this issue, in order to discuss applications of the proposed glottal model in articulatory speech synthesis.

One drawback of the model concerns closure: as the glottal area is assumed to be rectangular, closure of the glottis occurs abruptly, while in natural flow signals the closure is usually smoother due to, for example, zipper-like movements of the glottal area (see the discussion in Sec 4.1.2. Further studies must therefore concentrate on how to integrate such features into the model.

Summary

The use of lumped physical models for the synthesis of glottal flow signals has been discussed. Section 4.1 has reviewed the existing literature, focusing in particular on the Ishizaka-Flanagan model. The remaining of the chapter has presented two different attempts to simplify the IF model while preserving the physical description of the synthesis algorithms.

The physically-informed model of Sec. 4.2 can be used for identification, starting from real signals. The performance of the model has been shown to be comparable to that obtained from identification schemes that use the Liljencrants-Fant model. The proposed identification scheme is robust with respect to large bandwidths and noise in the target waveforms. After identification, the model can be used to resynthesize the target signals and to modify the voice quality by adjusting the physical parameters. A second application that has been discussed concerns the use of the model for voice quality assessment. Further improvements to the model must point towards a more rigorous study of the stability properties of the identified system, as well as complete automatization of the identification procedure.

The one-delayed-mass model proposed in Sec. 4.3 resembles closely the IF model,

but uses only one degree of freedom instead of two, and about half of the variables and the parameters. Analysis of the numerical simulations has shown that the model is a reasonable trade-off between accuracy of the description and simplicity of the structure. Due to its low computational costs it can be suitable for real-time applications. Further studies must investigate nature of the interaction between the glottal model and vocal tract loads. Another interesting direction for future research is concerned with the modeling of zipper-like movements of the vocal folds, and in general more accurate modeling of the closing phase.

Chapter 5

Contact models in multimodal environments

An important finding from the studies in ecological psychology [100, 152] is that listening subjects often tend to describe sounds in terms of causing events; Gaver [59] refers to this attitude as "everyday listening". According to these studies, the physical properties involved in sound generation can be grouped into two broad categories: *structural invariants* specify individual properties of objects such as size, shape, material. *Transformational invariants* characterize interactions between objects (e.g. collisions, frictions, and so on).

These experimental results can provide significant help in the development of sound models. The problem of sound-source determination is a fundamental question for the sonification of Virtual Environments and for the design of auditory icons [58]. An example application is that of simulated contact with objects in a Virtual Environment. This topic has already been discussed in Sec. 1.4, here it is enough to recall that previous research (see for e.g. [78]) have used signal models based on additive synthesis techniques. Physical models can be advantageous in that they can be easily synchronized with analogous graphic models, thus providing a high degree of coherence and consistency for the perception of objects and events [134]. Moreover, a description of synthesis algorithms in terms of physical generating phenomena can in principle help in conveying structural and transformational invariants in the synthesized sounds. Recent works [104, 54, 105, 107] have shown that oversimplified physical models are able to convey information on structural invariants related to geometrical properties of the resonators (such as shape and size) and to synthesize "cartoon" sounding objects where these invariants can be controlled. This chapter presents results about a third structural invariant, i.e. *material*, and about a structural invariant, namely *collision*.

A physical model is developed in Sec. 5.1, where a non-linear contact force excites a lumped resonator, and issues related to numerical implementation are discussed. The basic properties of the model are then investigated both analytically and experimentally in Sec. 5.2. Due to the simple structure of the system, it is possible to study the influence

This chapter is partially based on [12, 14, 108].

quantity	symbol	unit
No. of oscillators	N	
Oscill. positions	$x_{ol} \quad (l = 1 \dots N)$	[m]
Oscill. velocities	$\dot{x}_{ol} \quad (l = 1 \dots N)$	[m/s]
Hammer position	x_h	[m]
Hammer velocity	$\dot{x}_h$	[m/s]
Penetration	$x = x_h - \sum_{l=1}^{N} x_{ol}$	[m]
Penetration velocity	$\dot{x} = \dot{x}_h - \sum_{l=1}^{N} \dot{x}_{ol}$	[m/s]
Oscill. masses	m_o	[Kg]
Oscill. center freqs.	ω_{ol}	[rad]
Oscill. quality factor	q_o	
Oscill. damping coeffs.	$g_{ol} = \omega_{ol}/q_o$	[rad]
Hammer mass	m_h	[Kg]
Non-linear exponent	$\alpha = 2.8$	
Elastic constant	k	[N/m$^\alpha$]
Damping weight	λ	[Ns/m$^{\alpha+1}$]
Viscoel. charact.	$\mu = \lambda/k$	[s/m]

Table 5.1: *Symbols used throughout the chapter.*

of physical parameters (hammer and resonator masses, elasticity and damping coefficients of the non-linear contact force) on the system behavior. Contact time, in particular, can be an important cue for the perception of collision. Finally, Sec. 5.3 discusses the ability of the model to convey perception of material to the listener.

Although performed using elementary resonator models, this investigation can also be helpful for improving existing contact models in more complex systems: one example is hammer-string interaction in piano models, where contact time is a key feature for sound quality. The Stulov model [133] for piano hammer felts provides a realistic description of hysteretic contact forces, and is successful in fitting real data. However, recent research by Giordano and Mills [65] has questioned to some extent its general validity, suggesting the need for further investigations on alternative piano hammer models.

5.1 A hammer-resonator model

This section introduces a contact model made of two main blocks, namely a lumped hammer and a lumped resonator. The two blocks interact through a non-linear contact force. The contact force and its main properties are reviewed in Sec. 5.1.1. The interaction with a basic resonator model is discussed in Sec. 5.1.2, and details of the numerical implementation are discussed. Table 5.1 summarizes the main variables and parameters used throughout the chapter.

5.1.1 Non-linear contact force

Contact force models for collisions have been widely studied in musical acoustics, mainly in relation with hammer models in the piano. Assuming that the contact geometry is small (ideally, a point), an idealized model [66] states a polynomial dependence of the force f on the penetration x:

$$f(x(t)) = \begin{cases} -k[x(t)]^{\alpha}, & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{5.1}$$

where the exponent α depends on the local geometry around the contact surface. As an example, in an ideal impact between two spherical object α takes the value 1.5. Typical values in a piano hammer felt range from 1.5 to 3.5, with no definite trend from bass to treble. This simple model has been already briefly discussed in Sec. 2.2.2 as an example of non-linear lumped element.

More realistic models have to take into account hysteretic properties of the hammer material. In the case of a piano hammer, it is known that the force-compression characteristic exhibits a hysteretic behavior, such that loading and unloading of the hammer felt are not alike In particular, the dynamic force-compression characteristics is strongly dependence on the hammer normal velocity before collision. In order to account for these phenomena, Stulov [133] proposed an improved model where the contact force possesses history-dependent properties. The idea, which is taken from the general theory of mechanics of solids, is that the spring stiffness k in Eq. (5.1) has to be replaced by a time-dependent operator. Thus the contact force takes the form

$$f(x(t), t) = \begin{cases} -k[1 - h_r(t)] * [x(t)^{\alpha}], & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{5.2}$$

where $h_r(t) = \frac{\epsilon}{\tau} e^{-t/\tau}$ is a *relaxation function* that controls the "memory" of the material. By rewriting the convolution explicitly, the Stulov force is seen to be:

$$f(x(t), t) = -kx(t)^{\alpha} + \frac{\epsilon}{\tau} e^{-t/\tau} \int_0^t e^{\xi/\tau} x(\xi)^{\alpha} \, d\xi, \quad \text{for} \quad x > 0. \tag{5.3}$$

The Stulov model has proven to be successful in fitting experimental data where a hammer strikes a massive surface, and force, acceleration, displacement signal are recorded. Borin and De Poli [23] showed that it can be implemented numerically without significant losses in accuracy, stability and efficiency with respect to the simpler model (5.1). However, a recent study by Giordano and Mills [65] has investigated different experimental settings, where a hammer hits a vibrating string, and has showed that the Stulov model is not able to fit the data collected from such an experiment.

Useful results on contact models are also found from studies in robotics. Physical modeling of contact events is indeed an important aspect of dynamic simulations of robotic systems, when physical contact with the environment is required in order for

the system to execute its assigned task (for example, handling of parts by an industrial manipulator during assembly tasks, or manipulator collisions with unknown objects when operating in an unstructured environment). Marhefka and Orin [86] provide a detailed discussion of a collision model that was originally proposed by Hunt and Crossley [71]. Under the hypotesis that the contact surface is small, Hunt and Crossley proposed the following form for the contact force f:

$$f(x(t), \dot{x}(t)) = \begin{cases} -kx(t)^{\alpha} - \lambda x(t)^{\alpha} \cdot \dot{x}(t), & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{5.4}$$

where variables and parameters are listed in Table 5.1. Similarly to Eqs. (5.1) and (5.2), the value of the exponent α depends only on the local geometry around the contact surface. In the following the value $\alpha = 2.8$ is chosen, which is close to values found in piano hammer felts. Note that the force model (5.4) includes both an elastic component kx^{α} and a dissipative term $\lambda x^{\alpha}\dot{x}$. Moreover, the dissipative term depends on both x and $\dot{x}$, and is zero for zero penetration.

Marhefka and Orin have studied the collision of a hammer onto a massive surface, which is assumed not to move during collision. When the two objects collide, the hammer initial conditions are given by $x_h = 0$ and $\dot{x}_h = v_{in}$ (the latter quantity meaning normal velocity before collision). Since the surface is assumed not to move, the hammer position x_h coincides with the penetration x, and the hammer trajectory is therefore described by the differential equation $m_h \ddot{x}_h = f(x_h, \dot{x}_h)$. Define the hammer velocity as $v = \dot{x}_h$, then it is shown in [86] that

$$\frac{dv}{dx_h} = \frac{\dot{v}}{\dot{x}_h} = \frac{(\Lambda v + K)x_h^{\alpha}}{v}, \quad \Rightarrow \quad \int \frac{v\, dv}{(\Lambda v + K)} = \int x_h^{\alpha} dx, \tag{5.5}$$

where two auxiliary parameters $\Lambda = -\lambda/m_h$ and $K = -k/m_h$ have been introduced for clarity. The integral in Eq. (5.5) can be computed explicitly and x_h can be written as a function of v:

$$x_h(v) = \left[\left(\frac{\alpha + 1}{\Lambda^2} \right) \left(\Lambda(v - v_{in}) - K \log \left| \frac{K + \Lambda v}{K + \Lambda v_{in}} \right| \right) \right]^{\frac{1}{\alpha+1}}. \tag{5.6}$$

Another remark by Marhefka and Orin is concerned with "stickiness" properties of the contact force f. From Eq. (5.4), it can be seen that f becomes inward (or sticky) if $v < v_{lim} := -1/\mu$. However, this limit velocity is never exceeded on a trajectory with initial conditions $x_h = 0$, $v = v_{in}$, as shown in the phase portrait of Fig. 5.1(a). The upper half of the plot depicts the trajectories of a hammer striking the surface with various positive normal velocities (trajectories are traveled in clockwise direction). Note that the negative velocities after collision v_{out} are always smaller in magnitude than the corresponding v_{in}. Moreover, for increasing v_{in} the resulting v_{out} converges to the limit value v_{lim}. The horizontal line $v = v_{lim}$ corresponds to the trajectory where the elastic and dissipative terms cancel, and therefore the hammer travels from right to left with

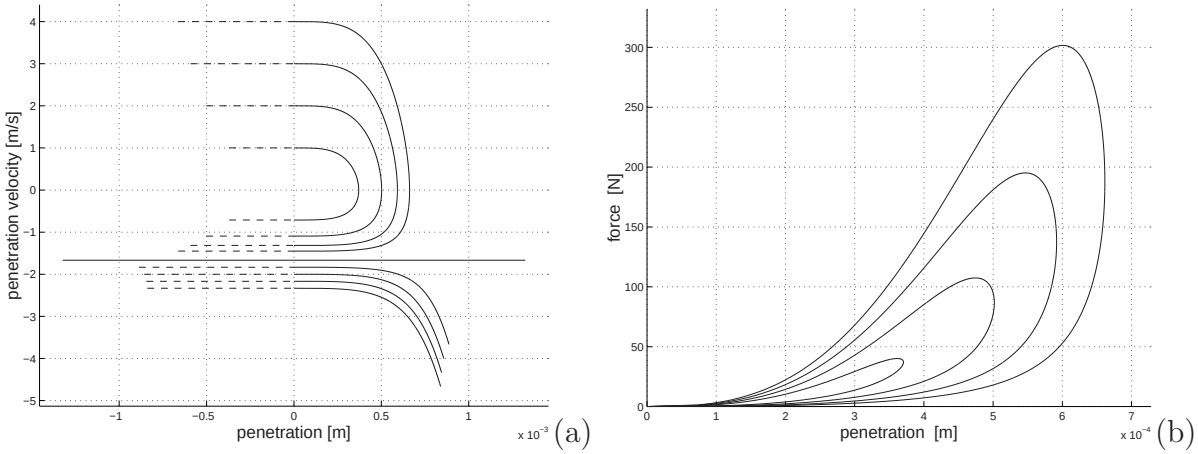

Figure 5.1: *Collision of a hammer with a massive surface for various v_{in}'s; (a) phase portrait, (b) penetration-force characteristics. Values for the hammer parameters are $m_h = 10^{-2}$ [Kg], $k = 1.5 \cdot 10^{11}$ [N/m$^\alpha$], $\mu = 0.6$ [s/m], $\alpha = 2.8$, $v_{in} = 1 \ldots 4$ [m/s].*

constant velocity. This horizontal line separates two regions of the phase space, and the lower region is never entered by the upper paths. The lower trajectories are entered when the hammer is given an initial displacement $x_h < 0$ and initial *negative* velocity $v_{in} < v_{lim}$. If such conditions are imposed, then one of the lower trajectories is traveled from left to right: the hammer bounces back from the surface, while its velocity decreases in magnitude, due to the sticky force f.

Figure 5.1(b) shows the penetration-force characteristics during collision. Note that the dissipative term $\lambda x^\alpha \dot{x}$ introduces hysteresis. It can also be noticed that f never becomes zero for positive penetrations. This is a significant advantage with respect to the Stulov model [133].

5.1.2 Interaction with resonators

The contact force discussed in the last section can be used to excite a resonator model made of lumped mechanical elements such as masses springs and dampers.

The simplest possible oscillating system made of these three lumped elements is a second order oscillator of the form

$$\ddot{x}_o(t) + g_o \dot{x}_o(t) + \frac{k_o}{m_o} x_o(t) = -\frac{1}{m_o} f_{ext}(t), \tag{5.7}$$

where variables and parameters for the oscillator are given in Table 5.1, while f_{ext} is any external driving force. Such a one-dimensional model provides a basic description of the resonator in terms of its pitch $\omega_o = \sqrt{k_o/m_o}$ and quality factor $q_o = \omega_o/g_o$. However, in many cases this is a too poor description, since it cannot provide rich spectral contents. A slightly more sophisticated model is obtained by connecting N of the oscillators given

in Eq. (5.7) in parallel (i.e., the oscillators are driven by the same force; see Sec. 2.2 for a discussion of parallel and series connections). By choosing a different stiffness k_{ol} ($l = 1 \dots N$) for each oscillator, it is possible to account for a set $\{\omega_{ol}\}_{l=1}^{N}$ of partials of the resonator spectrum. The resulting structure is similar to that of Eq. (5.7). In matrix form, it can be written as

$$\begin{bmatrix} \ddot{x}_{o1}(t) \\ \vdots \\ \ddot{x}_{oN}(t) \end{bmatrix} + \boldsymbol{G_o} \begin{bmatrix} \dot{x}_{o1}(t) \\ \vdots \\ \dot{x}_{oN}(t) \end{bmatrix} + \boldsymbol{\Omega_o}^2 \begin{bmatrix} x_{o1}(t) \\ \vdots \\ x_{oN}(t) \end{bmatrix} = \boldsymbol{M_o} f_{ext}(t), \tag{5.8}$$

where the vector the matrices are given by

$$\boldsymbol{\Omega_o} = \begin{bmatrix} \omega_{o1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \omega_{oN} \end{bmatrix}, \quad \boldsymbol{G_o} = \frac{1}{q_o}\boldsymbol{\Omega_o}, \quad \boldsymbol{M_o} = \begin{bmatrix} 1/m_o \\ \vdots \\ 1/m_o \end{bmatrix}.$$

The structure given in Eq. (5.8) provides a high degree of controllability. The displacement and the velocity of the resonator are obtained as $x_o = \sum_{l=1}^{N} x_{ol}$ and $\dot{x}_o = \sum_{l=1}^{N} \dot{x}_{ol}$, respectively. The quality factor q_o controls the decay time of the resonator response. The frequencies $\{\omega_{ol}\}_{l=1}^{N}$ can be chosen to reproduce spectra corresponding to various geometries of one-, two- and three-dimensional resonators. In particular, it was shown by Fontana and Rocchesso [54, 105] that spherical, cubic and intermediate-shaped 3-D cavities can be modeled in an efficient and effective way. The first N resonances of a cavity can be mapped into the spring constants k_{ol} of the N oscillators, and morphing between different shapes can be obtained by designing appropriate trajectories for each of these resonances.

When this resonator is coupled with the contact model of Sec. 5.1.1, the external force f_{ext} in Eq. (5.8) is substituted by the contact force $f(x, \dot{x})$ given in Eq. (5.4). In this case the hammer penetration x is given by $x = x_h - \sum_{l=1}^{N} x_{ol}$. Note that by allowing the mass m_o to vary for each oscillator, the matrix $\boldsymbol{M_o}$ can in principle be generalized to give control on the amounts of energy provided to each oscillator. This permits simulation of position-dependent interaction, in that different collision points excite the resonator modes in different ways. Early results about this generalized interaction were provided in [51].

The system is discretized using the bilinear transformation:

$$s = 2F_s \frac{1 - z^{-1}}{1 + z^{-1}}. \tag{5.9}$$

After this transformation, the resonator appears as a parallel filter bank of second-order bandpass filters, each one accounting for one specific mode of the resonator. Any bandpass filter can be accessed to its parameters of center-frequency and selectivity through the

physical quantities described above. The complete numerical system takes the form

$$\begin{cases} \boldsymbol{x_{ol}}(n) = \bar{\boldsymbol{A}}_{\boldsymbol{ol}}\boldsymbol{x_{ol}}(n-1) + \bar{\boldsymbol{b}}_{\boldsymbol{ol}}[y(n) + y(n-1)], \quad (\text{for} \quad l = 1 \dots N), \\ \boldsymbol{x_h}(n) = \bar{\boldsymbol{A}}_{\boldsymbol{h}}\boldsymbol{x_h}(n-1) + \bar{\boldsymbol{b}}_{\boldsymbol{h}}[y(n) + y(n-1)], \\ \boldsymbol{x}(n) = \boldsymbol{x_h}(n) - \sum_{l=1}^{N} \boldsymbol{x_{ol}}(n), \\ y(n) = f(\boldsymbol{x}(n)), \end{cases} \tag{5.10}$$

where

$$\boldsymbol{x_{ol}} = \begin{bmatrix} x_{ol} \\ \dot{x}_{ol} \end{bmatrix}, \quad \boldsymbol{x_h} = \begin{bmatrix} x_h \\ \dot{x}_h \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix},$$

$$\bar{\boldsymbol{A}}_{\boldsymbol{ol}} = \frac{1}{\Delta_{ol}} \begin{bmatrix} \Delta_{ol} - 2k_{ol} & 4F_s m_o \\ -4k_{ol}F_s & 8F_s^2 m_o - \Delta_{ol} \end{bmatrix}, \quad \bar{\boldsymbol{A}}_{\boldsymbol{h}} = \frac{1}{\Delta_h} \begin{bmatrix} \Delta_h & 4F_s m_h \\ 0 & \Delta_h \end{bmatrix}, \tag{5.11}$$

$$\bar{\boldsymbol{b}}_{\boldsymbol{ol}} = \frac{1}{\Delta_o} \begin{bmatrix} 1 \\ 2F_s \end{bmatrix}, \qquad \bar{\boldsymbol{b}}_{\boldsymbol{h}} = -\frac{1}{\Delta_h} \begin{bmatrix} 1 \\ 2F_s \end{bmatrix},$$

and where the quantities Δ_{ol} and Δ_h are given by $\Delta_{ol} = 4F_s^2 m_o + 2F_s m_o g_o + k_{ol}$ and $\Delta_h = 4F_s^2 m_h$, respectively. From Eqs. (5.10) it can be seen that the overall structure of the system resembles the block scheme depicted in Fig. 2.11. In particular, at each time step n the variables $\boldsymbol{x}(n)$ and $y(n)$ have instantaneous mutual dependence. This non-computable loop is solved with the K method as described in Sec. 2.3.2: the vector $\boldsymbol{x}(n)$ can be written as

$$\boldsymbol{x}(n) = \tilde{\boldsymbol{x}}(n) + \boldsymbol{K} f(n), \tag{5.12}$$

where

$$\boldsymbol{K} = -\left(\frac{1}{\Delta_h} + \frac{1}{\sum_{l=1}^{N} \Delta_{ol}} \right) \begin{bmatrix} 1 \\ 2F_s \end{bmatrix},$$

and $\tilde{\boldsymbol{x}}(n)$ is a computable vector (i.e. it is a linear combination of past values of the vectors $\boldsymbol{x_{ol}}, \boldsymbol{x_h}$ and y). Substituting the expression (5.12) in the non-linear contact force equation (5.4), and applying the implicit function theorem, y can be found as a function of $\tilde{\boldsymbol{x}}$:

$$y = f\left(\tilde{\boldsymbol{x}} + \boldsymbol{K} y\right) \quad \overset{\text{K method}}{\longmapsto} \quad y = h(\tilde{\boldsymbol{x}}). \tag{5.13}$$

As described in Sec. 2.3.2, two choices are available for efficient numerical implementation of the K method. The first choice amounts to pre-computing the new non-linear function h off-line and storing it in a look-up table. Alternatively, h can be found iteratively at each time step from the vector $\tilde{\boldsymbol{x}}$, using the Newton-Raphson method. This latter approach is adopted here. The same approach was already taken in chapter 3 and

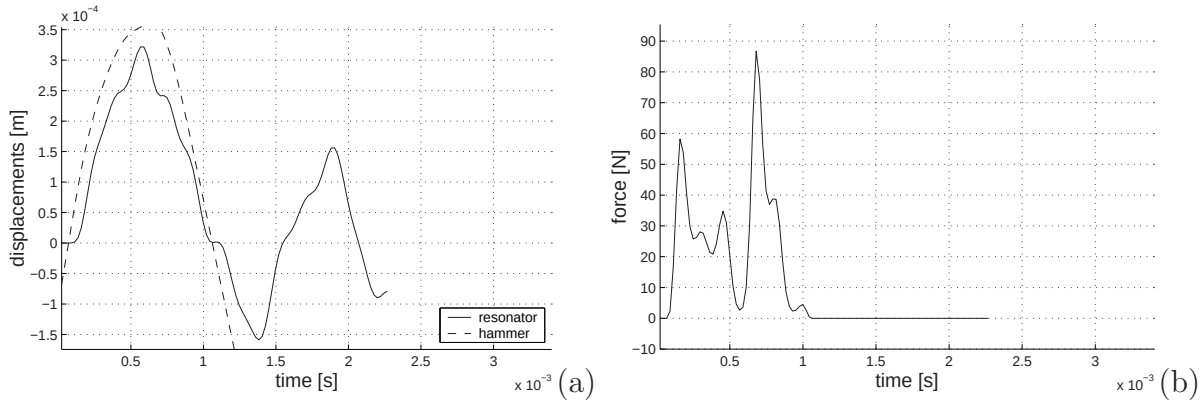

Figure 5.2: *A transient attack from the model: (a) hammer and resonator displacements* x_h *and* $\sum_{l=1}^{N} x_{ol}$*; (b) contact force* f *during interaction. The resonator is given here* $N = 3$ *partials.*

chapter 4 for dealing with non-linearities in a single reed model and in a model of the glottis, respectively.

Since most of the computational load in the numerical system comes from the non-linear function evaluation, the speed of convergence (i.e. the number of iterations) of the Newton-Raphson algorithm has a major role in determining efficiency of the simulations. In order to develop a real-time model, it is essential that the number of iterations remains small in a large region of the parameter space. To this end, analysis on the simulations has to be performed, where both the hammer and the resonator parameters are varied over a large range. Such analysis shows that in every conditions the algorithm exhibits a high speed of convergence. More precisely, the number of iterations is observed to be never higher than four, even when the Newton-Raphson algorithm is given extremely low tolerance errors ($\sim 10^{-13}$). Figure 5.2 displays an example of an attack transient, as obtained from the numerical model. The spikes in the force signal are due to the oscillations of the resonator partials during the collision.

5.2 Contact time: theory and simulations

The contact time t_0 (i.e. the time after which the hammer separates from the struck object) has a major role in defining the spectral characteristics of the initial transient. Qualitatively, a short t_0 corresponds to an impulse-like transient with a rich spectrum, and thus provides a bright attack. Similarly, a long t_0 corresponds to a smoother transient with little energy in the high frequency region. Therefore t_0 influences the spectral centroid of the attack transient.

It is known that the spectral content of the attack transient determines to a large extent the perceived quality of the impact. In a study on perceived mallet hardness, Freed [55], selected four acoustical parameters and studied their perceptual relevance through

listening tests. In his work, the acoustic stimuli were obtained from real recorded sounds of collision events, with mallets with varying hardness hitting various resonators. For each sound the subjects had to judge the perceived mallet hardness. It was found that the perception of hardness is strongly correlated to the spectral centroid of the attack transient. As already remarked, this latter parameter is in turn related to the contact time.

This section studies t_0 both analytically and experimentally (i.e. using numerical simulations). In Sec. 5.2.1, an equation is derived which relates t_0 to the physical parameters of the model, while Sec. 5.2.2 compares such analytical result to those obtained from simulations.

5.2.1 Analytical results

Using the contact model (5.4), Hunt and Crossley [71] studied the collision of a hammer with a rigid surface. They found an expression for the normal velocity after collision v_{out} in the limit of small μ (an analogous discussion is also reported in [86]). In this limit, v_{out} has a simple dependence on the normal velocity before collision, through a *coefficient of restitution* which in turns depends only on the hammer viscoelastic characteristics μ. When turning to the general case (i.e., when the parameter μ is allowed to take non-small values), studying the behavior of v_{out} is less trivial, and Hunt and Crossley do not address this case.

First of all, v_{in} and v_{out} correspond to the points where $x_h = 0$, i.e. to the roots of the right-hand side in Eq. (5.6). Therefore, from Eq. (5.6) v_{out} is found as

$$\begin{aligned} & x_h(v_{out}) = \Lambda(v_{out} - v_{in}) - K \log \left| \frac{K + \Lambda v_{out}}{K + \Lambda v_{in}} \right| = 0, \\ & \Rightarrow \frac{e^{\mu v_{out}}}{1 + \mu v_{out}} = \frac{e^{\mu v_{in}}}{1 + \mu v_{in}}. \end{aligned} \tag{5.14}$$

A first result is already evident from this equation: v_{out} depends only on the viscoelastic characteristics μ, and the input velocity v_{in}. There is no dependence on the spring stiffness k, the hammer mass m_h, the non-linear exponent α. A graphic study of the dependence of v_{out} on v_{in} and μ can be performed by plotting the function $\exp(\mu v)/(1 + \mu v)$, as in Fig. 5.3(a). It is seen that $v_{out} \to v_{lim}$ when v_{in} takes large values. This is consistent with the phase portrait in Fig. 5.1.

The second Eq. (5.14) can be rewritten as

$$e^{\mu v_{out}} = a\,(1 + \mu v_{out})\,, \quad \text{where} \quad a = \frac{e^{\mu v_{in}}}{1 + \mu v_{in}}. \tag{5.15}$$

Therefore v_{out} is the intersection of the exponential on the left-hand side and the linear function on the right-hand side, as shown in Fig. 5.3(b). The velocity v_{out} can be found numerically as the root of Eq. (5.15).

Having v_{out}, the contact time t_0 can now be computed. If collision occurs at $t = 0$, then the contact time is by definition given by $t_0 = \int_0^{t_0} dt$. Moreover, since $dt = dx_h/v$,

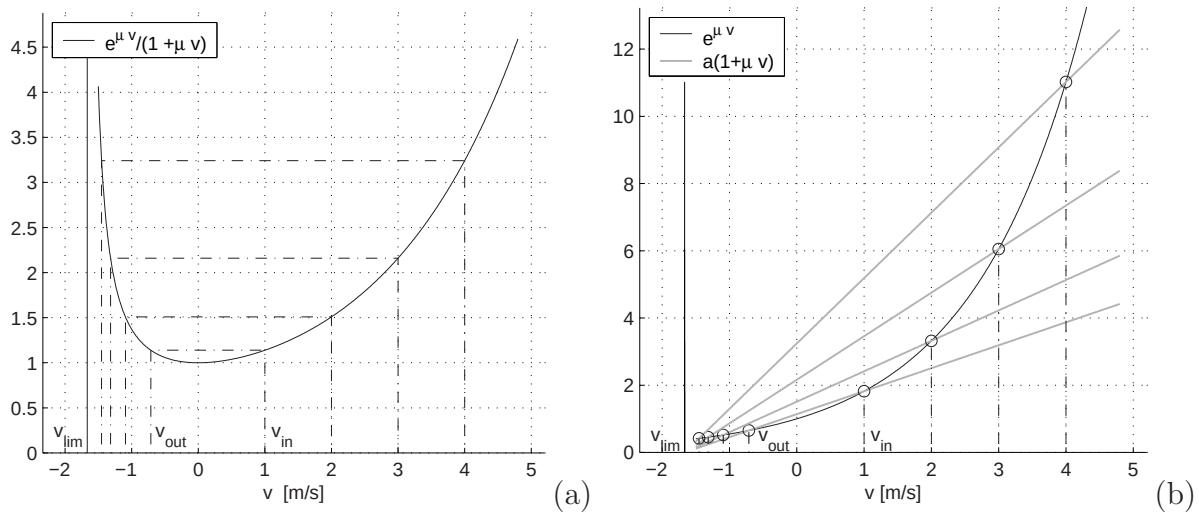

Figure 5.3: *Graphic study of v_{out} for various v_{in}'s. Values for the parameters are the same used in Fig. 5.1.*

also by definition, it is easily seen from Eq. (5.5) that

$$dt = \frac{dx_h}{v} = \frac{dv}{(\Lambda v + K)x_h^{\alpha}},$$

from which

$$t_0 = \int_0^{t_0} dt = \int_{v_{in}}^{v_{out}} \frac{dv}{(\Lambda v + K)x_h^{\alpha}}. \tag{5.16}$$

Recalling Eq. (5.6), x^{α} can be rewritten in the integral as a function of the velocity v. Thus, the integrand function depends only on v. Then straightforward calculations lead to the expression

$$t_0 = \left(\frac{\alpha+1}{\Lambda^2}\right)^{-\frac{\alpha}{\alpha+1}} \cdot \int_{v_{in}}^{v_{out}} \frac{dv}{(\Lambda v + K)\left[\Lambda(v - v_{in}) - K\log\left|\frac{K+\Lambda v}{K+\Lambda v_{in}}\right|\right]^{\frac{\alpha}{\alpha+1}}}.$$

A more useful expression is obtained by substituting the parameter $\mu = \Lambda/K$ in this equation. In this way, t_0 can be computed as a function of the parameter set (m, k, μ, α), together with the normal velocities before/after collision, (v_{in}, v_{out}). Again, few calculation steps yield

$$t_0 = \left(\frac{m_h}{k}\right)^{\frac{1}{\alpha+1}} \cdot \left(\frac{\mu^2}{\alpha+1}\right)^{\frac{\alpha}{\alpha+1}} \cdot \int_{v_{out}}^{v_{in}} \frac{dv}{(1+\mu v)\left[-\mu(v - v_{in}) + \log\left|\frac{1+\mu v}{1+\mu v_{in}}\right|\right]^{\frac{\alpha}{\alpha+1}}}. \tag{5.17}$$

It can be checked that the constant outside the integral has dimension [s^2/m], while the integral itself is a velocity [m/s]. Therefore, the whole expression on the right-hand side

has dimension [s]. Equation (5.17) states an important result: the contact time t_0 depends only on v_{in} and two parameters, i.e. the viscoelastic characteristic μ and the ratio m_h/k. A few remarks about Eq. (5.17):

- the integral has two singularities at the boundaries v_{out} and v_{in}. However, it can be checked that at these boundaries the integrand function converges asymptotically to $1/(v - v_{out})^{\alpha/(\alpha+1)}$ and $1/(v - v_{in})^{\alpha/(\alpha+1)}$, respectively. Therefore the integral always takes finite values;
- the integral depends only on v_{in} and μ. This is a consequence of Eq. (5.15), which states that v_{out} depends only on μ and v_{in};
- the constant outside the integral depends only on μ and the ratio m_h/k. Since neither m_h nor k affect the value of the integral, it follows that the power-law dependence $t_0(m_h/k) \sim (m_h/k)^{1/(\alpha+1)}$ holds;
- the dependence $t_0(\mu)$ is less easily established analytically; however, numerical integration of Eq. (5.17) can be used in order to study such dependence. Note that the singularities at v_{out}, v_{in} require additional care while integrating near the boundaries.

The results presented in this section emphasize a second advantage in using the contact model (5.4) instead of the Stulov model [133]: the explicit dependence of the force f on the system state $(x, \dot{x})$, as stated in Eq. (5.4), allows the analytical study resulting in Eq. (5.17). A similar analysis is not possible with the Stulov model. In this case, the only results about the contact time characteristics are obtained experimentally, from the numerical simulations.

5.2.2 Comparisons with simulations

This section analyzes experimentally the influence of the model parameters on the contact time. Following the work and the terminology by Giordano and Mills [65], two types of numerical experiments are analyzed in the rest of the section. In a first setup the hammer strikes a rigid surface and rebounds from it: note that this is the same setting used in Sec. 5.2.1 for deriving Eq. (5.17). In the following this is termed a "Type I" experiment. A second experimental setup involves collision between the non-linear hammer and the resonator described in Sec. 5.1.2; in the following, this is referred to as a "Type II" experiment.

The strategy adopted here amounts to run several simulations in which the parameters m_h/k and μ are varied over a large range. Automatic analysis on these simulations computes the contact time t_0 and compares the obtained values with those resulting from numerical integration of Eq. (5.17). The analyzed simulations are run at a sampling rate $F_s = 44.1$ [kHz], and each simulation is $5 \cdot 10^{-2}$ [s] long. For Type II simulations, the resonator is given $N = 3$ partials.

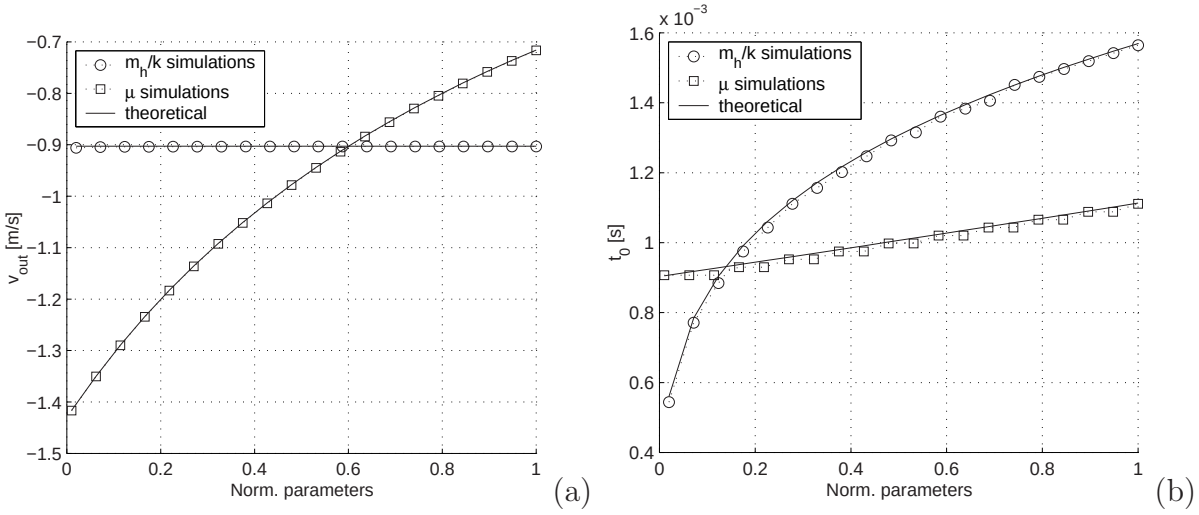

Figure 5.4: *Dependence of (a) v_{out} and (b) t_0 on parameters m_h/k and μ for Type I simulations (solid lines computed from Eq. (5.17), discrete points obtained from simulations). The horizontal axes are in normalized coordinates, ranges for the two parameters are $m_h/k \in [6, 300] \cdot 10^{-12}$ [Kg m^{α}/N], $\mu \in [0.01, 1]$ [s/m]. Other parameters are as in Fig. 5.1.*

Results for Type I experiments are summarized in Fig. 5.4, where both the theoretical behavior predicted by Eq. (5.17) and extracted data from numerical simulations are plotted. It can be seen that there is excellent accordance between theory and experiment. In particular, Fig. 5.4(a) shows that the values for v_{out} computed from the simulations do not depend on the ratio m_h/k, as predicted by Eq. (5.14). Moreover, from Fig. 5.4(b) it can be seen that the dependence $t_0(\mu)$ is almost linear in the observed range, and that the slope of this curve is small when compared to that of $t_0(m_h/k)$. In other words, the contact time t_0 is primarily affected by the value of m_h/k, and varies more slowly with respect to μ.

On the one hand, the accordance between the theoretical and the experimental t_0 values confirms the validity of the analytical study presented in Sec. 5.2.1. On the other hand, it assesses quantitatively the accuracy of the numerical implementation described in Sec. 5.1.2.

When analyzing Type II simulations, somewhat different results are found. This is due to the fact that in this case also the struck object is moving, therefore the contact time depends on both the hammer and the resonator parameters. In particular it is observed from Type II simulations that t_0 is always longer than in the Type I case, for any parameter setting.

Figure 5.5 plots the functions $t_0(\mu)$ and $t_0(m_h/k)$, for various resonator masses m_o. It can be noticed that the general dependence is always similar to that observed in Type I simulations. Moreover, the contact time is longer for light resonators and tends to the theoretical curve of Eq. (5.17) as m_o increases. This is an expected result, since Type I

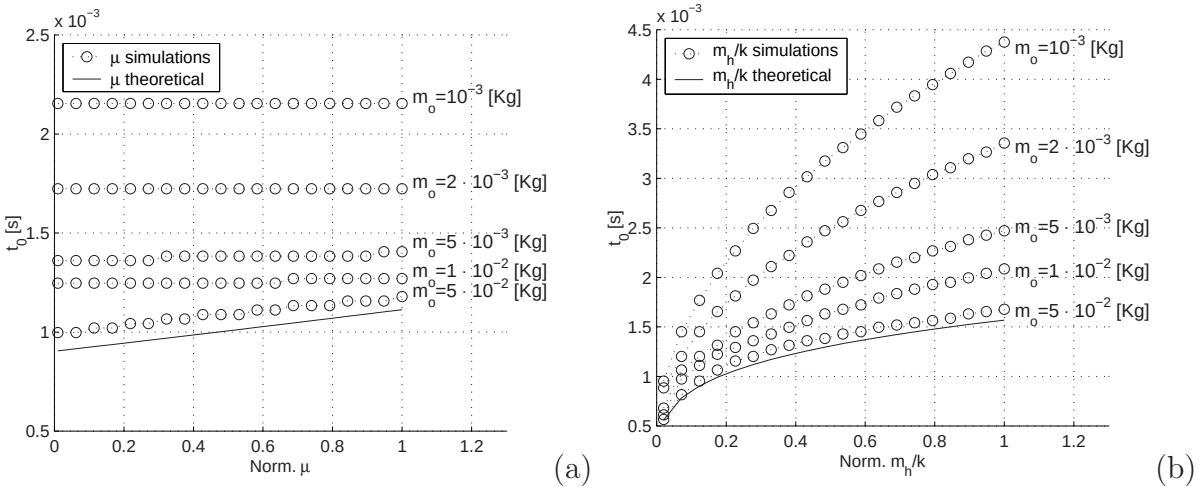

Figure 5.5: *Dependence of t_0 on the parameters (a) μ and (b) m_h/k, for Type II experiments (solid lines computed from Eq. (5.17), discrete points obtained from numerical simulations). The horizontal axes are in normalized coordinates, with $\mu \in [0.01, 1]$ [s/m] and $m_h/k \in [6, 300] \cdot 10^{-12}$ [Kg m^{α}/N]. Other parameters are as in Fig. 5.1.*

simulations are equivalent to Type II simulations where the resonator mass m_o is given an infinite value.

The results outlined in this section, and in particular Eq. (5.17), can be regarded as complementary to the findings by Freed [55]. In his work, Freed studied experimentally the correlation between signal parameters (such as spectral centroid of the attack transient) and perceived hammer hardness, without any assumption on the physical description. Conversely, here it has been established a connection between signal parameters (namely t_0, which is strongly correlated with spectral centroid) and an underlying physical model. Further study is needed in order to find a direct correlation between physical parameters and perceived hammer hardness.

5.3 Material perception

Rendering an impression of object material is not always cost effective –or even possible– in graphic rendering. On the other hand, physically-based sound models can give the possibility to embed material properties with almost no computational overhead. Section 5.3.1 discusses the acoustic cues exploited by the auditory system for the perception of material, and examines how these cues are controlled in the hammer-resonator model. Section 5.3.2 presents the results obtained from listening tests with synthetic stimuli.

5.3.1 Acoustic cues

At present, few studies have investigated what acoustic cues (if any) are exploited by the auditory system in order to recognize materials of sound sources. Based on theoretical considerations, Wildes and Richards [153] suggested the overall decay time as a significant cue, since it is a direct measure of internal friction in a given material. However, this is only true when a standard inelastic linear solid model is assumed.

Two recent studies with listening subjects provided some experimental basis to this conjecture, but results were not in accordance. Lutfi and Oh [84] found that changes in the decay time are not easily perceived by listeners, while changes in the fundamental frequency seem to be a more salient cue. On the other hand, Klatzky *et al.* [78] showed that decay plays a much larger role than pitch in affecting judgment. Both these studies made use of synthetic stimuli obtained using additive synthesis algorithms. Therefore, no realistic attack transients could be obtained.

Even less clear is how to embed material properties in physically-based sound synthesis algorithms, and how to control them by accessing the physical parameters of the sound models. Djoharian [38] showed that finite difference models of resonators can be covered by a "viscoelastic dress" to fit a given frequency-damping characteristic, which is taken as the sound signature of the material. This approach relies on a low-level physical description and, as a result, very accurate yet computationally expensive algorithms are obtained. The sound examples provided by Djoharian convinced many researchers of the importance and the effectiveness of materials in sound communication. No impact model was used in this study. Therefore the sounds are impulse responses with no physical attack transients.

This section investigates the use of the contact model presented in Sec. 5.1 for conveying material perception to the listener. Using such an accurate yet efficient physical model for subjective tests is advantageous over using damped sinusoids or other signal-based sound models, in that realistic interactions can be reproduced. As a result, complex and realistic attack transients can be kept in the stimuli, thus eliminating possible biases due to oversimplified test sounds.

The results presented in the next section are based on a listening test in which experimental subjects are asked to listen to acoustic stimuli synthesized using the hammer-resonator model. When designing the acoustic stimuli for the test, it has to be considered that the contact sound produced by hammer-resonator interaction can give information on both the hammer and the resonator properties. This perceptual effect is known as *phenomenical scission* in experimental psychology. As an example, Freed [55] showed that not only properties of the resonator but also the hammer hardness (see Sec. 5.2) can be perceived from percussive sounds.

Since the perception of material relates to the properties of the resonator, the acoustic stimuli for the test have to be synthesized using the same hammer, i.e. the same set of coefficients m_h, k, μ, α. The impact velocity v_{in} of the hammer has to be fixed as well. In this way, a constant excitation is provided to the resonator. The resonator is constructed from Eq. (5.8). For the actual stimuli, the value $N = 1$ is chosen.

The choice of using only one partial is justified by the fact that previous studies [153, 84, 78] have concentrated exclusively on two acoustic parameters of the stimuli, namely pitch and decay time. Therefore, a single second-order oscillator is enough for controlling these two parameters: pitch is defined using the center frequency $\omega_o = \sqrt{k_o/m_o}$ of the oscillator, while decay characteristics are controlled through the quality factor q_o. More precisely, it is known that for a second-order oscillator q_o relates to decay via the equation $q_o = \omega_o t_e/2$, where t_e is defined as the time for the sound to decay by a proportion $1/e$.

The stimuli used in the test described in the next section are a set of 100 sounds. These are obtained using five equally log-spaced pitches from 1000 to 2000 [Hz], together with 20 equally log-spaced quality factors from 5 to 5000. These extremal q_o values correspond to typical values found in rubber and aluminium, respectively. In a recent study on plucked string sounds, Tolonen and Järveläinen [137] found that relatively large deviations (between -25% and $+40\%$) in the decay time are not perceived by listeners. With the values that have been chosen here, the relative lower/upper spacing between q_o values is $-31\%/+44\%$.

The mapping from the two acoustic parameters and the physical parameters of the resonator is chosen as follows:

- the stiffness k_o is given a fixed value for all the stimuli, so that for each ω_o the corresponding mass is computed as $m_o = k_o/\omega_o^2$.

- For each quality factor q_o, the corresponding damping coefficient is computed as $g_o = \omega_o/q_o$.

5.3.2 Experimental results

The results presented in this section are based on a listening test with 22 experimental subjects. The subjects had to listen to the 100 sounds described in the previous section, and to indicate what material each sound was coming from, choosing from a set of four material classes: rubber, wood, glass and steel (this approach is analogous to that adopted in [78]). Each sound was played once and followed by a pause in which subjects had to choose the corresponding material class. All of the 22 listeners were volunteers, both expert and non-expert listeners, all reported normal hearing.

Figure 5.6 summarizes results of the experiment: it shows the proportion of subjects who assigned each sound to a given material category, as a function of the two acoustic cues (pitch and quality factor). The intersubject agreements (proximity of the response proportions to 0 or 1) are qualitatively consistent with indications given by Wildes and Richards [153], namely

- The responses tend to cluster in horizontal stripes. This result suggests that the quality factor q_o is the most significant cue, while the listeners' choices are less affected by the pitch ω_o.

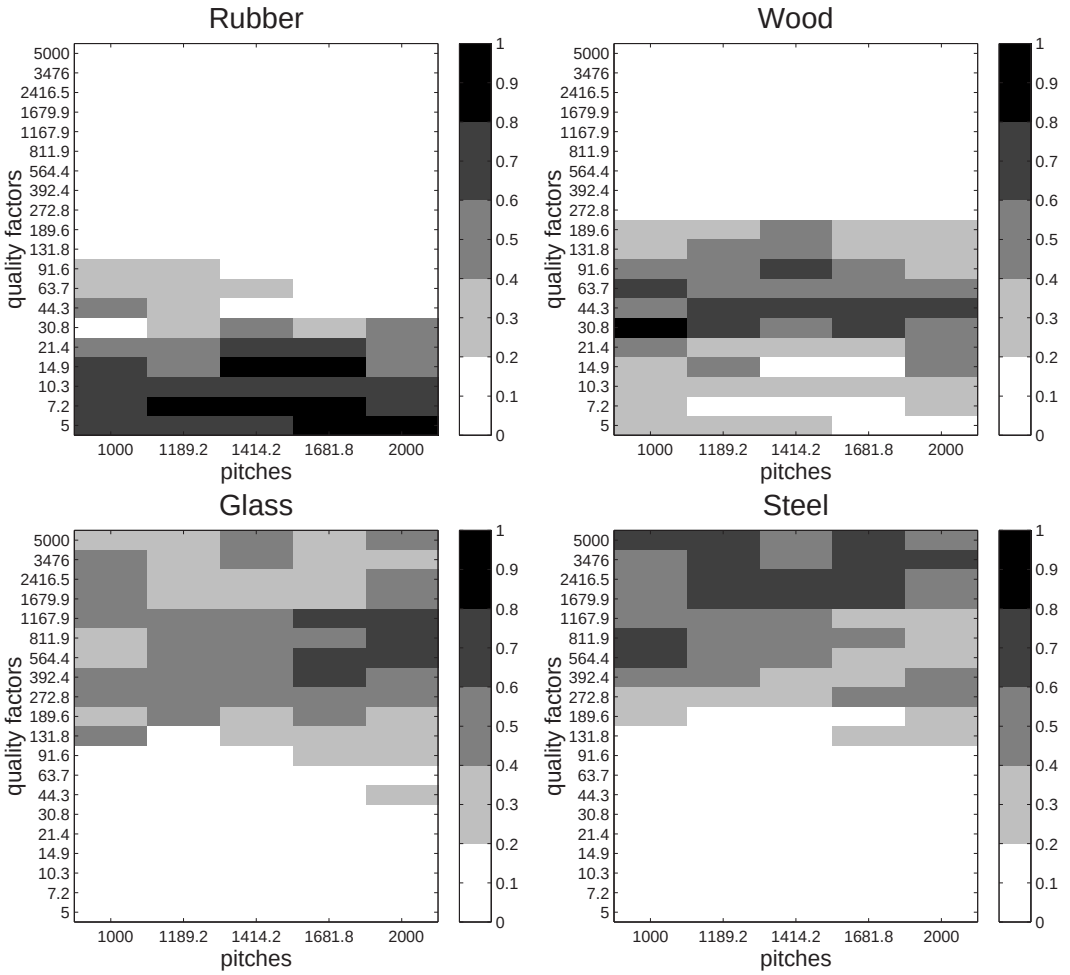

Figure 5.6: *Proportion of subjects who recognized a certain material for each sound example. Pitches (in [Hz]) and quality factors are on the horizontal and vertical axes, respectively.*

- The quality factor q_o (and therefore the decay time t_e) is in increasing order for rubber, wood, glass and steel.

A slight dependence on pitch can be noticed: rubber and glass tend to be preferred at high pitches, while wood and steel are more often chosen at low pitches. Klatzky *et al.* [78] found a similar trend for glass and steel, while they observed the opposite for rubber and wood (i.e. these two materials were preferably chosen at low and high pitches, respectively). However, the findings in [78] are not easily comparable to those illustrated in Fig. 5.6, since a different pitch range was chosen in [78] (namely, five equally log-spaced frequencies from 100 to 1000 [Hz]).

Table 5.2 collects the q_o ranges for each material. The boundaries of these ranges are found by computing the minimum and maximum values where more than 50% of the

Material	q_o	t_e [s]
Rubber	$[5, 44.3]$	$[8 \cdot 10^{-4}, 1.41 \cdot 10^{-2}]$
Wood	$[14.9, 131.8]$	$[2.3 \cdot 10^{-3}, 3.53 \cdot 10^{-2}]$
Glass	$[189.6, 5000]$	$[4.34 \cdot 10^{-2}, 1.1254]$
Steel	$[272.8, 5000]$	$[4.34 \cdot 10^{-2}, 1.5915]$

Table 5.2: *Minimum and maximum values of the quality factor and the decay time for which more than* 50% *of the experimental subjects have chosen a given material.*

subjects chose one of the four materials. The corresponding ranges for the decay time t_e are also given.

From both Fig. 5.6 and Table 5.2, it appears that the upper and lower halves of the q_o range are well separated, while materials within each of these two regions are less easily discriminated. In particular, it is evident from the figure that the regions corresponding to glass and steel are largely overlapping, while ranges for rubber and wood are better delimited. After the test, many subjects reported that the indication "glass" was not not immediately clear to them, since they could not guess what sound is produced by a bar made of glass.

Another possible explanation for the worse performance in the high q_o-range has to do with the synthesis model: for long decay times (such as those of glass and steel) an single partial with exponential decay envelope is probably a too poor approximation of reality, and a more accurate description is needed. Figure 5.7 plots the same data as in Table 5.2 on the $q_o/\omega_o, t_e$ plane, thus allowing direct comparison with the qualitative plot reported by Wildes and Richards [153]. Again, the separation between the low and high q_o-ranges appears clearly. Rubber and wood are better discriminated, while the ranges for glass and steel are largely overlapping.

These findings show that decay (or quality factor q_o) plays a much larger role than pitch ω_o in material perception. Moreover, material classification by subjects is qualitatively in accordance with reported measures of internal friction coefficients for these material classes. This indicates that even the extremely simple resonator model (5.8) with $N = 1$ can elicit perception of material, provided that it allows for control over the salient acoustical cues. When using a higher number N of partials, more realistic and less "cartoonized" sounds are obtained while control over material perception is maintained through the parameter q_o.

However, measures on the intersubject agreement show that the classification is inaccurate for high quality factors (glass and steel), thus suggesting that the overall decay time does not fully account for material properties and that control on the decay shape of

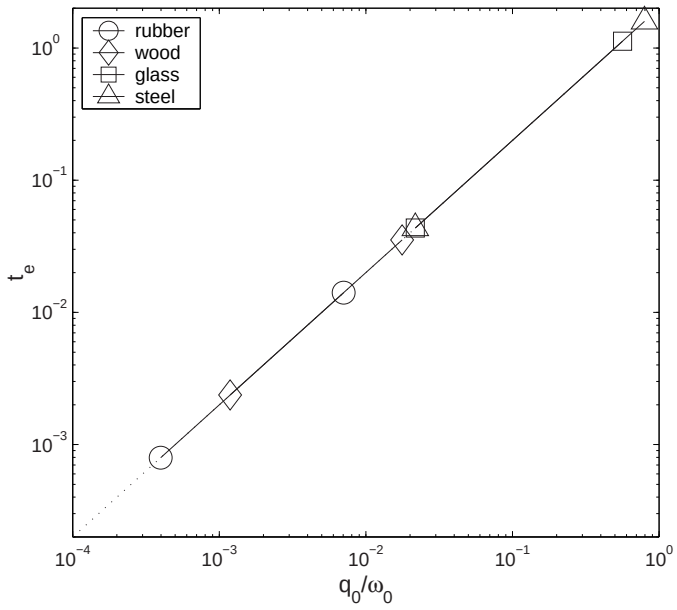

Figure 5.7: *Distribution of materials on the $q_o/\omega_o, t_e$ plane.*

each partial would be needed in order to allow for a more accurate description of reality. Analysis and subjective experiments with real sounds have to be performed, in order to understand how the classification is improved and to investigate whether decay shape can play a role in helping material perception. This information can then be exploited to develop an improved resonator model. Preliminary observations [149] suggest that experiments with real sounds yield perceptual results that are qualitatively similar to those summarized in Fig. 5.6.

Summary

A hammer-resonator model has been developed, and compared to existing contact models in the literature. In particular, it has been shown that the proposed contact model has common features with the piano hammer felt model proposed by Stulov. However, further research is needed to compare the two, and to discuss the possible use of the model in physically-based synthesis of the piano.

In Sec. 5.1 it has been shown that an accurate and efficient numerical implementation of the model can be obtained by using the bilinear transform in combination with the K method. The resulting numerical algorithm can be easily implemented in real-time on general-purpose hardware.

Section 5.2 has discussed the properties of the contact force, focusing on the influence of physical parameters in perceptual features of the interaction. Contact time has been investigated in detail, since this parameter affects the spectral centroid of the transient

attack and influences the perceived quality of the collision. Further research and psychophysical experiments are needed in order to validate the ability of the model to convey perception of impact hardness.

Section 5.3 has focused on the resonator properties, and has explored the ability of the model to elicit perception of the resonator material to the listener. First, it has been shown that the physical parameters give control over the salient acoustic cues. Second, listening tests with synthetic stimuli obtained from the model have provided perceptual validation of the model. Results from the tests have been compared to the existing literature.

Bibliography

[1] S. Adachi and M. Sato. Trumpet Sound Simulation Using a Two-dimensional Lip Vibration Model. *J. Acoust. Soc. Am.*, 99(2):1200–1209, Feb. 1996.

[2] J. M. Adrien, R. Caussé, and E. Ducasse. Dynamic Modeling of Stringed and Wind Instruments, Sound Synthesis by Physical Models. In *Proc. Int. Computer Music Conf. (ICMC'88)*, pages 265–276, 1988.

[3] J. Agulló, A. Barjau, and J. Martinez. Alternatives to the Impulse Response *h(t)* to Describe the Acoustical Behavior of Conical Ducts. *J. Acoust. Soc. Am.*, 84(5):1606–1612, Nov. 1988.

[4] J. Agulló, S. Cardona, and D. H. Keefe. Time-domain Deconvolution to Measure Reflection Functions for Discontinuities in Waveguides. *J. Acoust. Soc. Am.*, 97(3):1950–1957, March 1995.

[5] P. Alku, H. Tiitinen, and R. Näätänen. A Method for Generating Natural Sounding Speech Stimuli for Cognitive Brain Research. *Clinical Neurophysiology*, 110:1329–1333, 1999.

[6] P. Alku and E. Vilkman. A Comparison of Glottal Voice Quantification Parameters in Breathy, Normal and Pressed Phonation of Female and Male Speakers. *Folia Phoniatrica et Logopaedica*, 48:240–254, 1996.

[7] R. J. Anderson and M. W. Spong. Bilateral Control of Teleoperators with Time Delay. *IEEE Trans. Automat. Contr.*, 34(5):494–501, May 1989.

[8] F. Avanzini. Stable and Accurate Numerical Methods in Physical Models of Musical Instruments. A Study on the Clarinet. In *Proc. Diderot Forum on Mathematics and Music*, pages 11–19, Vienna, Dec. 1999.

[9] F. Avanzini. On the Use of Weighted Sample Methods in Digitizing the Clarinet Equations. In *Proc. Int. Computer Music Conf. (ICMC'00)*, pages 46–49, Berlin, Aug. 2000.

[10] F. Avanzini, B. Bank, G. Borin, G. De Poli, F. Fontana, and D. Rocchesso. Musical Instrument Modeling: the Case of the Piano. In *Proc. MOSART Int. Workshop on Current Research Directions in Comp. Mus.*, Barcelona, Nov. 2001. Available at `http://www.iua.upf.es/mtg/mosart/`.

[11] F. Avanzini, C. Drioli, and P. Alku. Synthesis of the Voice Source Using a Physically Informed Model of the Glottis. In *Proc. Int. Symp. Mus. Acoust. (ISMA'01)*, pages 31–34, Perugia, Sept. 2001.

[12] F. Avanzini and D. Rocchesso. Controlling Material Properties in Physical Models of Sounding Objects. In *Proc. Int. Computer Music Conf. (ICMC'01)*, La Habana, Sept. 2001. Available at `http://www.soundobject.org`.

[13] F. Avanzini and D. Rocchesso. Efficiency, Accuracy, and Stability Issues in Discrete Time Simulations of Single Reed Wind Instruments. 2001. Submitted for publication.

[14] F. Avanzini and D. Rocchesso. Modeling Collision Sounds: Non-linear Contact Force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFX-01)*, pages 61–66, Limerick, Dec. 2001.

[15] F. Avanzini and M. van Walstijn. Modelling the Mechanical Response of the Reed-mouthpiece-lip System of a Clarinet. Part I. The Finite-difference Approach. 2001. Submitted for publication.

[16] J. Backus. Small Vibration Theory of the Clarinet. *J. Acoust. Soc. Am.*, 35:305–313, 1963.

[17] D. R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, 1994.

[18] A. H. Benade. *Fundamentals of Musical Acoustics*. Oxford University Press, New York, 1976.

[19] A. H. Benade. Equivalent Circuits for Conical Waveguides. *J. Acoust. Soc. Am.*, 83(5):1764–1769, May 1988.

[20] D. P. Berners. *Acoustics and Signal Processing Techniques for Physical Modeling of Brass Instruments*. PhD thesis, CCRMA, Stanford University (CA), 1999. Available at `http://www-ccrma.stanford.edu/~dpberner/`.

[21] D. A. Berry and I. R. Titze. Normal Modes in a Continuum Model of Vocal Fold Tissues. *J. Acoust. Soc. Am.*, 100(5):3345–3354, Nov. 1996.

[22] S. Bilbao. *Wave and Scattering Methods for the Numerical Integration of Partial Differential Equations*. PhD thesis, CCRMA, Stanford University (CA), 2000. Available at `http://www-ccrma.stanford.edu/~bilbao`.

[23] G. Borin and G. De Poli. A Hysteretic Hammer-String Interaction Model for Physical Model Synthesis. In *Proc. Nordic Acoustical Meeting*, pages 399–406, Helsinki, June 1996.

[24] G. Borin, G. De Poli, and D. Rocchesso. Elimination of Delay-free Loops in Discrete-Time Models of Nonlinear Acoustic Systems. *IEEE Trans. Speech Audio Process.*, 8(5):597–606, Sep. 2000.

[25] G. Borin, G. De Poli, and A. Sarti. Algorithms and Structures for Synthesis Using Physical Models. *Computer Music J.*, 16(4):30–42, 1992.

[26] K. Brandenburg. Perceptual Coding of High Quality Digital Audio. In M. Kahrs and K. Brandenburg, editors, *Applications of DSP to Audio and Acoustics*, pages 39–83. Kluwer Academic Publishers, 1998.

[27] M. A. Casey. Understanding Musical Sound with Forward Models and Physical Models. *Connection Sci.*, 6:355–371, 1994.

[28] A. Chaigne and V. Doutaut. Numerical Simulations of Xylophones. I. Time-domain Modeling of the Vibrating Bar. *J. Acoust. Soc. Am.*, 101(1):539–557, Jan. 1997.

[29] A. Chaigne and V. Doutaut. Numerical Simulations of Xylophones. II. Time-domain Modeling of the Resonator and of the Radiated Sound Pressure. *J. Acoust. Soc. Am.*, 104(3):1633–1647, Sep. 1998.

[30] S. Chen and S. A. Billings. Representation of Non-linear Systems: Narmax Model. *Int. J. of Control*, 49(3):1013–1032, 1989.

[31] S. Chen, C. F. Cowan, and P. M. Grant. Orthogonal Least Squares Learning Algorithm for Radial Basis Functions Networks. *IEEE Trans. Neural Networks*, 2(2):302–309, March 1991.

[32] D. G. Childers and C. Ahn. Modeling the Glottal Volume-Velocity Waveform for Three Voice Types. *J. Acoust. Soc. Am.*, 97(1):505–519, Jan. 1995.

[33] M. de Oliveira Rosa, J. C. Perreira, and M. Grellet. Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis. *IEEE Trans. Biomed. Engineering*, 47(1):96–104, Jan. 2000.

[34] G. De Poli. A Tutorial on Digital Sound Synthesis Techniques. In C. Roads, editor, *The Music Machine*, pages 429–447. MIT Press, 1991.

[35] G. De Poli and D. Rocchesso. Physically Based Sound Modelling. *Organized Sound*, 3(1):61–76, 1998.

[36] M. P. de Vries, H. K. Schutte, and G. J. Verkerke. Determination of Parameters for Lumped Parameter Model of the Vocal Fold Using a Finite-Element Method Approach. *J. Acoust. Soc. Am.*, 106(6):3620–3628, Dec. 1999.

[37] J. R. Deller, J. G. Proakis, and J. H. L. Hansen. *Discrete-Time Processing of Speech Signals.* Macmillan, New York, 1993.

[38] P. Djoharian. Shape and Material Design in Physical Modeling Sound Synthesis. In *Proc. Int. Computer Music Conf. (ICMC'00)*, pages 38–45, Berlin, Sept. 2000.

[39] P. Doenges, F. Lavagetto, J. Ostermann, I.S. Pandzic, and E. Petajan. Mpeg-4: Audio/Video and Synthetic Graphics/Audio in Mixed Media. *Image Communication*, 9(4):433–463, May 1997.

[40] P. G. Drazin. *Nonlinear Systems*. Cambridge Univ. Press, 1992.

[41] C. Drioli and F. Avanzini. Model-Based Synthesis and Transformation of Voiced Sounds. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-00)*, pages 45–49, Verona, Dec. 2000.

[42] C. Drioli and F. Avanzini. A Physically Informed Model of the Glottis with Application to Voice Quality Assessment. In *Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA'01)*, Firenze, Sept. 2001.

[43] G. Fant. *Acoustic Theory of Speech Production*. Mouton, Paris, 2nd edition, 1970.

[44] G. Fant, J. Liljencrants, and Q. Lin. A Four-Parameter Model of Glottal Flow. In *Speech Transmiss. Lab. Q. Prog. Stat. Rep.*, pages 1–13, 1985.

[45] A. Fettweis. Wave Digital Filters: Theory and Practice. *Proc. IEEE*, 74(2):270–327, Feb. 1986.

[46] N. H. Fletcher. Mode Locking in Nonlinearly Excited Inharmonic Musical Oscillators. *J. Acoust. Soc. Am.*, 64(6):1566–1569, Dec. 1978.

[47] N. H. Fletcher. Bird Song–A Quantitative Acoustic Model. *J. Theor. Biol*, 135:455–481, 1988.

[48] N. H. Fletcher. Autonomous Vibration of Simple Pressure-Controlled Valves in Gas Flows. *J. Acoust. Soc. Am.*, 93(4):2172–2180, Apr. 1993.

[49] N. H. Fletcher and T. D. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, New York, 1991.

[50] J.L. Florens and C. Cadoz. The Physical Model: Modeling and Simulating the Instrumental Universe. In G. De Poli, A. Piccialli, and C. Roads, editors, *Representations of Musical Signals*, pages 227–268. MIT Press, 1991.

[51] F. Fontana, L. Ottaviani, and D. Rocchesso. Recognition of Ellipsoids from Acoustic Cues. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 160–164, Limerick, Dec. 2001.

[52] F. Fontana and D. Rocchesso. Physical Modeling of Membranes for Percussion Instruments. *Acustica – acta acustica*, 84:529–542, May 1998.

[53] F. Fontana and D. Rocchesso. Signal-Theoretic Characterization of Waveguide Mesh Geometries for Models of Two-Dimensional Wave Propagation in Elastic Media. *IEEE Trans. Speech Audio Process.*, 9(2):152–161, Feb 2001.

[54] F. Fontana, D. Rocchesso, and E. Apollonio. Acoustic Cues from Shapes between Spheres and Cubes. In *Proc. Int. Computer Music Conf. (ICMC'01)*, La Habana, Sept. 2001. Available at `http://www.soundobject.org`.

[55] D. J. Freed. Auditory Correlates of Perceived Mallet Hardness for a Set of Recorded Percussive Events. *J. Acoust. Soc. Am.*, 87(1):311–322, Jan. 1990.

[56] K. Galkowski. Higher Order Discretization of 2-D Systems. *IEEE Trans. Circuits and Systems–I: Fund. Theory and Applications*, 47(5):713–722, May 2000.

[57] W. G. Gardner. Reverberation Algorithms. In M. Kahrs and K. Brandenburg, editors, *Applications of DSP to Audio and Acoustics*, pages 85–131. Kluwer Academic Publishers, 1998.

[58] W. Gaver. Using and Creating Auditory Icons. In G. Kremer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 417–446. Addison-Wesley, 1994.

[59] W. W. Gaver. How Do We Hear in the World? Explorations in Ecological Acoustics. *Ecological Psychology*, 5(4):285–313, Apr. 1993.

[60] L. Gavidia-Ceballos and J. H. L. Hansen. Direct Speech Feature Estimation Using an Iterative EM Algorithm for Vocal Fold Pathology Detection. *IEEE Trans. Biomed. Engineering*, 43(4):373–383, Apr. 1996.

[61] B. Gazengel. *Caractérisation Objective de la Qualité de Justesse, de Timbre et d'Émission des Instruments à Vent, à Anche, Simple.* PhD thesis, Laboratoire d'Acoustique de l'Université du Maine, Le Mans, France, 1994.

[62] B. Gazengel, J. Gilbert, and N. Amir. Time Domain Simulation of Single Reed Wind Instruments. From the Measured Impedance to the Synthesis Signal. Where are the Traps? *Acta Acustica*, 3:445–472, Oct. 1995.

[63] J. Gilbert. *Etude des Instruments de Musique à Anche, Simple: Extension de la Méthode d'Équilibrage Harmonique, Rôle de l'Inharmonicité des Résonances, Mesure des Grandeurs d'Entrée.* PhD thesis, Laboratoire d'Acoustique de l'Université du Maine, Le Mans, France, 1991.

[64] J. Gilbert and J. Kergomard. Calculation of the Steady-State Oscillations of a Clarinet Using the Harmonic Balance Technique. *J. Acoust. Soc. Am.*, 86(1):35–41, July 1989.

[65] N. Giordano and J. P. Mills. Hysteretical Behavior of Piano Hammers. In *Proc. Int. Symp. Mus. Acoust. (ISMA'01)*, pages 237–241, Perugia, Sept. 2001.

[66] D. E. Hall. Piano String Excitation VI: Nonlinear Modeling. *J. Acoust. Soc. Am.*, 92:95–105, July 1992.

[67] L. Hiller and P. Ruiz. Synthesizing Musical Sounds by Solving the Wave Equation for Vibrating Objects: Part I. *J. Audio Engin. Soc.*, 19(6):462–470, June 1971.

[68] L. Hiller and P. Ruiz. Synthesizing Musical Sounds by Solving the Wave Equation for Vibrating Objects: Part II. *J. Audio Engin. Soc.*, 19(7):542–551, July 1971.

[69] A. Hirschberg, X. Pelorson, and J. Gilbert. Aeroacoustics of Musical Instruments. *Meccanica*, 31:131–141, 1996.

[70] A. Hirschberg, R. W. A. van de Laar, J. P. Marrou-Mauriéres, A. P. J. Wijnands, H. J. Dane, S. G. Kruijswijk, and A. J. M. Houtsma. A Quasi-stationary Model of Air Flow in the Reed Channel of Single-reed Woodwind Instruments. *Acustica*, 70:146–154, 1990.

[71] K. H. Hunt and F. R. E. Crossley. Coefficient of Restitution Interpreted as Damping in Vibroimpact. *ASME J. Applied Mech.*, pages 440–445, June 1975.

[72] T. Idogawa, T. Kobata, K. Komuro, and M. Iwaki. Nonlinear Vibrations in the Air Column of a Clarinet Artificially Blown. *J. Acoust. Soc. Am.*, 93(1):540–551, Jan. 1993.

[73] K. Ishizaka and J. L. Flanagan. Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords. *Bell Syst. Tech. J.*, 51:1233–1268, 1972.

[74] M. Kahrs. Digital Audio System Architecture. In M. Kahrs and K. Brandenburg, editors, *Applications of DSP to Audio and Acoustics*, pages 195–234. Kluwer Academic Publishers, 1998.

[75] M. Kahrs and F. Avanzini. Computer Synthesis of Bird Songs and Calls. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 23–27, Limerick, Dec. 2001.

[76] K. Karplus and A. Strong. Digital Synthesis of Plucked String and Drum Timbres. *Computer Music J.*, 7:43–55, 1983.

[77] D. H. Keefe. On Sound Production in Reed-Driven Wind Instruments. Technical report, (No. 9003), Systematic Musicology Program. School of Music, Univ. of Washington, Seattle, 1990.

[78] R. L Klatzky, D. K. Pai, and E. P. Krotov. Perception of Material from Contact Sounds. *Presence*, 9(4):399–410, Aug. 2000.

[79] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine. Splitting the Unit Delay Tools for Fractional Delay Filter Design. *IEEE Sig. Process. Magazine*, 13(1):30–60, Jan. 1996.

[80] J. D. Lambert. *Numerical Methods for Ordinary Differential Systems.* John Wiley & Sons, 1993.

[81] L. Lapidus and G. Pinder. *Numerical solution of partial differential equations in science and engineering.* John Wiley & Sons, New York, 1982.

[82] L. Lapidus and J. H. Seinfeld. *Numerical Solution of Ordinary Differential Equations.* Academic Press, New York, 1971.

[83] L. Ljung. *System Identification. Theory for the User.* Prentice-Hall, Englewood Cliffs, NJ, 1999.

[84] R. A. Lutfi and E. L. Oh. Auditory Discrimination of Material Changes in a Struck-Clamped Bar. *J. Acoust. Soc. Am.*, 102(6):3647–3656, Dec. 1997.

[85] C. Maganza, R. Caussé, and L. Laloë. Bifurcations, Period Doublings and Chaos in Clarinetlike Systems. *Europhys. Lett.*, 1(6):295–302, March 1986.

[86] D. W. Marhefka and D. E. Orin. A Compliant Contact Model with Nonlinear Damping for Simulation of Robotic Systems. *IEEE Trans. Systems, Man and Cybernetics-Part A*, 29(6):566–572, Nov. 1999.

[87] J. Martinez and J. Agulló. Conical Bores. Part I: Reflection Functions Associated with Discontinuities. *J. Acoust. Soc. Am.*, 84(5):1613–1619, Nov. 1988.

[88] J. Martinez, J. Agulló, and S. Cardona. Conical Bores. Part II: Multiconvolution. *J. Acoust. Soc. Am.*, 84(5):1620–1627, Nov 1988.

[89] D. W. Massaro. Auditory Visual Speech Processing. In *Proc. Eurospeech Conf.*, pages 1153–1156, Aalborg, Sept. 2001.

[90] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse. On the Oscillations of Musical Instruments. *J. Acoust. Soc. Am.*, 74(5):1325–1345, Nov. 1983.

[91] P. Meyer, R. Wilhelms, and H. W. Strube. A Quasiarticulatory Speech Synthesizer for German Language Running in Real Time. *J. Acoust. Soc. Am.*, 86(2):523–539, Aug. 1989.

[92] S. K. Mitra. *Digital Signal Processing. A Computer Based Approach.* Mc Graw Hill, 1998.

[93] J. A. Moorer. About This Reverberation Businness. *Computer Music J.*, 2:13–28, 1979.

[94] P. M. Morse and K. U. Ingard. *Theoretical Acoustics.* Princeton University Press, Princeton, New Jersey, 1968.

[95] C. J. Nederveen. *Acoustical Aspects of Woodwind Instruments*. F. Knuf, Amsterdam, 1969.

[96] A. V. Oppenheim and R. W. Schafer. *Discrete-time Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[97] T. Painter and A. Spanias. Perceptual Coding of Digital Audio. *Proc. IEEE*, 88(4):451–513, Apr. 2000.

[98] F. Pedersini, A. Sarti, and S. Tubaro. Object-Based Sound Synthesis for Virtual Environments. *IEEE Sig. Process. Magazine*, 17(6):37–51, Nov 2000.

[99] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals*. Prentice-Hall, Englewood Cliffs, NJ, 1978.

[100] W. A. Richards. Sound Interpretation. In W. A. Richards, editor, *Natural Computation*, pages 301–308. MIT Press, 1988.

[101] E. L. Riegelsberger and A. K. Krishnamurthy. Glottal Source Estimation: Methods of Applying the LF-Model to Inverse Filtering. In *Proc. IEEE Int. Conf. Acoust. Speech. and Signal Process. (ICASSP'93)*, volume II, pages 542–545, Minneapolis, 1993.

[102] D. Rocchesso. The Ball within the Box: a Sound-Processing Metaphor. *Computer Music J.*, 19(4):47–57, 1995.

[103] D. Rocchesso. Fractionally-addressed Delay Lines. *IEEE Trans. Speech Audio Process.*, 8(6):717–727, Nov. 2000.

[104] D. Rocchesso. Acoustic Cues for 3-D Shape Information. In *Proc. Int. Conf. Auditory Display (ICAD'01)*, Espoo, Finland, July 2001.

[105] D. Rocchesso. Simple Resonators with Shape Control. In *Proc. Int. Computer Music Conf. (ICMC'01)*, La Habana, Sept. 2001.

[106] D. Rocchesso and P. Dutilleux. Generalization of a 3-D Resonator Model for the Simulation of Spherical Enclosures. *Applied Sig. Process.*, 2001(1):15–26, 2001.

[107] D. Rocchesso and L. Ottaviani. Can One Hear the Volume of a Shape? In *Proc. IEEE Workshop on Applications of Sig. Process. to Audio and Acoustics (WASPAA'01)*, New Paltz (NY), Oct. 2001.

[108] D. Rocchesso, L. Ottaviani, F. Avanzini, F. Fontana, and M. Rath. Sonic Rendering of Object Features. Submitted for publication.

[109] D. Rocchesso and F. Scalcon. Accurate Dispersion Simulation for Piano Strings. In *Proc. Nordic Acoustical Meeting*, pages 407–414, Helsinki, June 1996.

[110] D. Rocchesso and J. O. Smith III. Circulant and Elliptic Feedback Delay Networks for Artificial Reverberation. *IEEE Trans. Speech Audio Process.*, 5(1):51–60, Jan. 1997. Available at `http://www-ccrma.stanford.edu/~jos/cfdn/`.

[111] M. Roseau. *Vibrations in Mechanical Systems: Analytical Methods and Applications*. Springer-Verlag, Berlin, 1987.

[112] A. Sarti and G. De Poli. Toward Nonlinear Wave Digital Filters. *IEEE Trans. Sig. Process.*, 47(6):1654–1668, June 1999.

[113] L. Savioja and V. Välimäki. Reducing the Dispersion Error in the Digital Waveguide Mesh Using Interpolation and Frequency-Warping Techniques. *IEEE Trans. Speech Audio Process.*, 8(2):184–194, Mar. 2000.

[114] G. P. Scavone. *An Acoustic Analysis of Single-Reed Woodwind Instruments with an Emphasis on Design and Performance Issues and Digital Waveguide Modeling Techniques.* PhD thesis, CCRMA, Stanford University (CA), 1997. Available at `http://www-ccrma.stanford.edu/~gary`.

[115] E. D. Scheirer. The MPEG-4 Structured Audio Standard. In *Proc. IEEE Int. Conf. Acoust. Speech. and Signal Process. (ICASSP'98)*, pages 3801–3804, Seattle, 1998.

[116] E. D. Scheirer. Structured Audio and Effects Processing in the MPEG-4 Multimedia Standard. *Multimedia Systems*, 7:11–22, 1999.

[117] E. D. Scheirer, R. Väänänen, and J. Huopaniemi. Audiobifs: Describing Audio Scenes with the MPEG-4 Multimedia Standard. *IEEE Trans. Multimedia*, 1(3):237–250, Sept. 1999.

[118] A. M. Schneider, J. T. Kaneshige, and F. D. Groutage. Higher Order s-to-z Mapping Functions and their Application in Digitizing Continuous-Time Filters. *Proc. IEEE*, 79(11):1661–1674, Nov. 1991.

[119] J. Schroeter and M. M. Sondhi. Techniques for Estimating Vocal-Tract Shapes from the Speech Signal. *IEEE Trans. Speech Audio Process.*, 2(1):133–150, Jan. 1994.

[120] R. T. Schumacher. *Ab Initio* Calculations of the Oscillations of a Clarinet. *Acustica*, 48(2):71–85, 1981.

[121] S. Serafin, J. O. Smith III, and H. Thornburg. A Pattern Recognition Approach to Invert a Bowed String Physical Model. In *Proc. Int. Symp. Mus. Acoust. (ISMA'01)*, pages 241–244, Perugia, 2001.

[122] X. Serra. Musical Sound Modeling With Sinusoids Plus Noise. In C. Roads, S. Pope, A. Picialli, and G. De Poli, editors, *Musical Signal Processing*. Swets & Zeitlinger.

[123] X. Serra, J. Bonada, P. Herrera, and R. Louriero. Integrating Complementary Spectral Models in the Design of a Musical Synthesizer. In *Proc. Int. Computer Music Conf. (ICMC'97)*, pages 152–159, Thessaloniki, 1997.

[124] J. O. Smith III. Viewpoints on the History of Digital Synthesis. In *Proc. Int. Computer Music Conf. (ICMC'91)*, pages 1–10, Montreal, Oct. 1991.

[125] J. O. Smith III. Physical Modeling Using Digital Waveguides. *Computer Music J.*, 16(4):74–87, 1992.

[126] J. O. Smith III. Physical Modeling Synthesis Update. *Computer Music J.*, 20(2):44–56, 1996.

[127] J. O. Smith III. Principles of Digital Waveguide Models of Musical Instruments. In M. Kahrs and K. Brandenburg, editors, *Applications of DSP to Audio and Acoustics*, pages 417–466. Kluwer Academic Publishers, 1998.

[128] S. D. Sommerfeldt and W. J. Strong. Simulation of a Player-Clarinet System. *J. Acoust. Soc. Am.*, 83(5):1908–1918, May 1988.

[129] M. M. Sondhi and J. Schroeter. A Hybrid Time-Frequency Domain Articulatory Speech Synthesizer. *IEEE Trans. Acoust., Speech, and Sig. Process.*, 35(7):955–967, July 1987.

[130] S. E. Stewart and W. J Strong. Functional Model of a Simplified Clarinet. *J. Acoust. Soc. Am.*, 68(1):109–120, July 1980.

[131] B. H. Story and I. R. Titze. Voice Simulation with a Body-Cover Model of the Vocal Folds. *J. Acoust. Soc. Am.*, 97(2):1249–1260, Feb. 1995.

[132] H. Strik. Automatic Parametrization of Differentiated Glottal Flow: Comparing Methods by Means of Synthetic Flow Pulses. *J. Acoust. Soc. Am.*, 103(5):2659–2669, May 1998.

[133] A. Stulov. Hysteretic Model of the Grand Piano Hammer Felt. *J. Acoust. Soc. Am.*, 97(4):2577–2585, Apr. 1995.

[134] K. Tadamura and E. Nakamae. Synchronizing Computer Graphics Animation and Audio. *IEEE Multimedia*, 5(4):63–73, Oct. 1998.

[135] S. C. Thompson. The Effect of the Reed Resonance on Woodwind Tone Production. *J. Acoust. Soc. Am.*, 66(5):1299–1307, Nov. 1979.

[136] I. R. Titze, B. H. Story, and E. Hunter. Rules for Controlling Low-Dimensional Vocal Fold Models with Muscle Activities. NCVS Status and Progress Report, 12:43-53, 1998.

[137] T. Tolonen and H. Järveläinen. Perceptual Study of Decay Parameters in Plucked String Synthesis. Presented at the AES 109th Int. Conv., preprint no. 5205, 2000.

[138] T. Tolonen, V. Välimäki, and M. Karjalainen. Modeling of Tension Modulation Nonlinearity in Plucked Strings. *IEEE Trans. Speech Audio Process.*, 8(3):300–310, May 2000.

[139] V. Välimäki. *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters.* PhD thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, 1995. Available at `http://www.acoustics.hut.fi/publications/theses.html`.

[140] V. Välimäki and M. Karjalainen. Improving the Kelly-Lochbaum Vocal Tract Model Using Conical Tube Sections and Fractional Delay Filtering Techniques. In *Proc. Int. Conf. Spoken Language Process.*, pages 615–618, Sept. 1994.

[141] K. van del Doel and D. K. Pai. The Sounds of Physical Shapes. *Presence*, 7(4):382–395, Aug. 1998.

[142] S. A. van Duyne, J. R. Pierce, and J. O. Smith III. Traveling Wave Implementation of a Lossless Mode-Coupling Filter and the Wave Digital Hammer. In *Proc. Int. Computer Music Conf. (ICMC'94)*, pages 411–418, Århus, Denmark, Sept. 1994.

[143] S. A. van Duyne and J. O. Smith III. The 2-D Digital Waveguide Mesh. In *Proc. IEEE Workshop on Applications of Sig. Process. to Audio and Acoustics (WASPAA'93)*, pages 177–180, New Paltz (NY), Oct. 1993.

[144] M. van Walstijn. Private communication, 2000.

[145] M. van Walstijn. *Discrete-Time Modelling of Brass and Reed Woodwind Instruments with Applications to Musical Sound Synthesis.* PhD thesis, Dept. of Physics and Astronomy, University of Edinburgh, 2001. Available at `http://www.music.ed.ac.uk/student/pages/pg/maarten/`.

[146] M. van Walstijn and F. Avanzini. Determination of the Lumped Parameters of a Non-linear Clarinet Reed Oscillator Using a Finite-Difference Approach. In *Proc. Int. Symp. Mus. Acoust. (ISMA'01)*, pages 221–224, Perugia, Sept. 2001.

[147] M. van Walstijn and F. Avanzini. Modelling the Mechanical Response of the Reed-mouthpiece-lip System of a Clarinet. Part II. A Lumped Model Approximation. 2001. In preparation.

[148] B. L. Vercoe, W. G. Gardner, and E. D. Scheirer. Structured Audio: Creation, Transmission and Rendering of Parametric Sound Representations. *Proc. IEEE*, 86(5):922–940, May 1998.

[149] G. B. Vicario. Phenomenology of Sounding Objects. In D. Rocchesso, editor, *Report of the Sounding Object Project Consortium*. 2001. Available at `http://www.soundobject.org`.

[150] C. Villain, L. Le Marrec, W. Op't Root, J. Willems, X. Pelorson, and A. Hirschberg. Towards a New Brass Player's Lip Model. In *Proc. Int. Symp. Mus. Acoust. (ISMA'01)*, pages 107–110, Perugia, Sept. 2001.

[151] C. Wan and A. M. Schneider. Further Improvements in Digitizing Continuous-Time Filters. *IEEE Trans. Signal Process.*, 45(3):533–542, March 1997.

[152] W. H. Warren and R. R. Verbrugge. Auditory Perception of Breaking and Bouncing Events: Psychophysics. In W. A. Richards, editor, *Natural Computation*, pages 364–375. MIT Press, 1988.

[153] R. P. Wildes and W. A. Richards. Recovering Material Properties from Sound. In W. A. Richards, editor, *Natural Computation*, pages 357–363. MIT Press, 1988.

[154] T. A. Wilson and G. S. Beavers. Operating Modes of the Clarinet. *J. Acoust. Soc. Am.*, 56(2):653–658, Aug. 1974.

[155] J. Wolfe. Seminar on wind instrument acoustics and live performance with a "flarinet". Scuola di S. Giorgio Maggiore, Fond. Cini, Venice, June 2000.

[156] W. E. Worman. *Self-sustained Nonlinear Oscillations of Medium Amplitude in Clarinet-like Systems*. PhD thesis, Case Western Reserve University, Cleveland, OH, 1971.

[157] U. Zölzer. *Digital Audio Signal Processing*. John Wiley & Sons, Chichester, 1997.