

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, 2002

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 gestural, 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

Prefazione	i
Ringraziamenti	v
Preface	vii
Acknowledgments	xi
1 Physically-based models	1
1.1 Introduction: sounds, sources	1
1.2 Musical instruments	3
1.3 Voice production and speech synthesis	7
1.4 Sound sources in virtual environments	10
1.5 Structured audio coding	13
Summary	16
2 Structural and computational aspects	19
2.1 Distributed models: the waveguide approach	19
2.1.1 One-dimensional wave propagation	19
2.1.2 Waveguide structures	22
2.1.3 Junctions and networks	26
2.2 Lumped models	29
2.2.1 Analogies	30
2.2.2 Non-linearities	33
2.3 Stability, accuracy, computability	35
2.3.1 Numerical methods	35
2.3.2 The K method	39
Summary	44

3	Single reed models	47
3.1	Lumped modeling	48
3.1.1	Reed-lip-mouthpiece system	48
3.1.2	The quasi-static approximation	50
3.2	An efficient and accurate numerical scheme	53
3.2.1	Numerical methods	53
3.2.2	Frequency-domain analysis	56
3.2.3	Time-domain analysis	59
3.3	Finite-difference modeling	62
3.3.1	A distributed model	63
3.3.2	Numerical formulation	66
3.3.3	Results from simulations	69
3.4	A lumped non-linear model	74
3.4.1	Parameter determination	75
3.4.2	Properties of the non-linear oscillator	77
	Summary	80
4	Source models for articulatory speech synthesis	83
4.1	Glottal models	84
4.1.1	The Ishizaka-Flanagan model	84
4.1.2	Properties of lumped glottal models	87
4.2	Non-linear block: identification	89
4.2.1	The identification procedure	90
4.2.2	Results and applications	94
4.3	Non-linear block: modified interaction	96
4.3.1	A one-delayed-mass-model	96
4.3.2	Numerical simulations	99
	Summary	102
5	Contact models in multimodal environments	105
5.1	A hammer-resonator model	106
5.1.1	Non-linear contact force	107
5.1.2	Interaction with resonators	109
5.2	Contact time: theory and simulations	112
5.2.1	Analytical results	113
5.2.2	Comparisons with simulations	115
5.3	Material perception	117
5.3.1	Acoustic cues	118
5.3.2	Experimental results	119
	Summary	122
	Bibliography	125

List of Figures

1.1	<i>Exciter-resonator interaction scheme for a musical instrument, to be compared to Fig. 1.3. Figure based on De Poli and Rocchesso [35].</i>	4
1.2	<i>Feedforward source-filter block scheme associated to LPC based speech synthesis.</i>	7
1.3	<i>General block scheme of an articulatory speech synthesizer, to be compared to Fig. 1.1. Figure based on Sondhi and Schroeter [129].</i>	8
1.4	<i>Schematic representation of the main components in a virtual environment or multimedia system.</i>	11
1.5	<i>A source-medium-receiver model, based on Begault [17]. Vertical numbered sections represent n_i processing stages associated to each block</i>	12
2.1	<i>Illustration of cylindrical and spherical coordinates.</i>	21
2.2	<i>Lossless waveguide sections with observation points at position $x = 0$ and $x = mX_s = L$; (a) cylindrical section; (b) conical section.</i>	23
2.3	<i>Waveguide simulation of non-ideal media; (a) frequency independent dissipation; (b) dispersion.</i>	25
2.4	<i>Kelly-Lochbaum junction for two cylindrical bores with different areas.</i>	27
2.5	<i>Example of an acoustic model implemented with waveguide sections and junctions.</i>	28
2.6	<i>Boundary regions for (a) non-convex and (b) convex conical junctions.</i>	29
2.7	<i>A mass pulled by a linear spring; (a) mechanical system and (b) combination of one-ports in series.</i>	31
2.8	<i>A Helmholtz resonator driven by an external acoustic wave; (a) acoustic system and (b) circuit representation.</i>	32
2.9	<i>Non-linear behavior of (a) capacitance $C(v)$ and (b) charge $q(v)$ in the Chua-Felderhoff circuit.</i>	34
2.10	<i>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.</i>	37
2.11	<i>Typical structure of a non-linear exciter.</i>	40
2.12	<i>A linear system; (a) delay-free path (figure based on Mitra [92]), and (b) an equivalent realization with no delay-free paths.</i>	41
2.13	<i>Shear transformation of $f(x) = e^{-x^2}$ for various k values.</i>	43

3.1	<i>Schematized representation of the reed-mouthpiece-lip system.</i>	49
3.2	<i>Lumped model of a single reed.</i>	51
3.3	<i>Quasi-static approximation of a single reed; (a) u_f versus Δp and (b) rotated mapping $p^+ = R_{nl}(p^-)$.</i>	52
3.4	<i>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/m²].</i>	58
3.5	<i>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 $\omega_0 = 2\pi \cdot 3700$ [rad/s], $g = 3000$ [rad/s], $\mu = 0.0231$ [Kg/m²].</i>	59
3.6	<i>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].</i>	61
3.7	<i>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]).</i>	62
3.8	<i>Idealized geometry of a clarinet single reed.</i>	64
3.9	<i>Geometrical parameters for the mouthpiece lay and the lip.</i>	65
3.10	<i>Comparison of the reed responses for $\theta = 1/4$ and $\theta = 1/2$ (obtained from simulations with $F_s = 200$ [kHz] and $N = 200$).</i>	71
3.11	<i>Dynamic simulations; (a) external driving force and (b) reed tip displacement.</i>	73
3.12	<i>Quasi-static simulations: (a) pressure-displacement characteristics of the system and (b) stiffness per unit area K_a.</i>	74
3.13	<i>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.</i>	75
3.14	<i>Parameter functions of the lumped non-linear oscillator.</i>	79
3.15	<i>Dependence of ω_0 and μ on the tip displacement.</i>	80
4.1	<i>Schematic representation of the Ishizaka-Flanagan model: (a) coronal view and (b) superior view.</i>	86
4.2	<i>First two excited modes in a distributed model of the vocal folds.</i>	88
4.3	<i>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.</i>	90
4.4	<i>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_0u 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).</i>	93

4.5	<i>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).</i>	95
4.6	<i>Scheme of the one-delayed-mass model (lower half) as opposed to the IF model (upper half).</i>	97
4.7	<i>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.</i>	100
4.8	<i>Dependence of the flow skewness on the time delay t_0. Simulations are run at $F_s = 22.05$ [kHz].</i>	101
4.9	<i>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]).</i>	102
5.1	<i>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], $\alpha = 2.8$, $v_{in} = 1 \dots 4$ [m/s].</i>	109
5.2	<i>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.</i>	112
5.3	<i>Graphic study of v_{out} for various v_{in}'s. Values for the parameters are the same used in Fig. 5.1.</i>	114
5.4	<i>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$^\alpha$/N], $\mu \in [0.01, 1]$ [s/m]. Other parameters are as in Fig. 5.1.</i>	116
5.5	<i>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$^\alpha$/N]. Other parameters are as in Fig. 5.1.</i>	117
5.6	<i>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.</i>	120
5.7	<i>Distribution of materials on the $q_0/\omega_0, t_e$ plane.</i>	122

List of Tables

1.1	<i>A tentative taxonomy of sound/speech synthesis techniques.</i>	2
1.2	<i>Properties of various structured representations of sound. Table based on Vercoe et al. [148].</i>	15
2.1	<i>Summary of analogies in electrical, mechanical and acoustical systems. . .</i>	33
3.1	<i>Symbols used throughout the chapter.</i>	48
3.2	<i>Measured threshold pressures from time-domain simulations. Values are rounded to the nearest integer</i>	60
3.3	<i>Parameters of the distributed model used in the simulations.</i>	70
3.4	<i>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.</i>	72
4.1	<i>Symbols used throughout the chapter.</i>	85
5.1	<i>Symbols used throughout the chapter.</i>	106
5.2	<i>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.</i>	121

List of symbols

<i>Symbol</i>	<i>Quantity</i>	<i>Unit</i>
$t \in \mathbb{R}$	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	$\text{Im}(s)$ [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]

