

ESCOLA UNIVERSITÀRIA D'ENGINYERIA
TÈCNICA DE TELECOMUNICACIÓ LA SALLE

TREBALL FINAL DE MÀSTER

MÀSTER EN ENGINYERIA DE XARXES I TELECOMUNICACIONS



ALUMNE

PROFESSOR PONENT

Marc Arnela Coll

Dr. Oriol Guasch Fortuny

ACTA DE L'EXAMEN DEL TREBALL FINAL DE MÀSTER

Reunit el Tribunal qualificador en el dia de la data, l'alumne

Marc Arnela Coll

va exposar el seu Treball Final de Màster, el qual va tractar sobre el tema següent:

Articulatory synthesis of vowels using finite element methods

Acabada l'exposició i contestades per part de l'alumne les objeccions formulades pels Srs. membres del tribunal, aquest valorà l'esmentat Treball amb la qualificació de

Barcelona,

VOCAL DEL TRIBUNAL

VOCAL DEL TRIBUNAL

PRESIDENT DEL TRIBUNAL

Als meus pares

Abstract

In this work we propose to develop a computational strategy based on the use of Finite Element Methods (FEM) for articulatory speech synthesis. In this context, we follow a bottom-up strategy to deal with the physics involved in speech generation. First, we solve the acoustic wave equation using FEM in space and finite differences in time. Next, boundary losses due to viscous friction and heat conduction are taken into account. Then, a radiation condition is introduced to simulate outward propagation waves to freespace. A Perfectly Matched Layer (PML) is used for this purpose. Finally we apply the finite element method to the vowel synthesis problem, taking as an example the case of vowel /e/. The quality of the results is analyzed by means of objective measurements.

Summary

This project is focused on human computer interaction (HCI), where speech plays a key role, both for general users and, in particular, for users with particular accessibility needs (people with sensory disabilities and the aged). In this context, the automatic generation of speech signals (i.e. speech synthesis) has made several significant steps so as to move from poorly intelligible to very natural systems. However, this process has almost ignored the seminal idea of modeling the human vocal tract and the articulation processes (articulatory techniques) to propose more practical techniques based on actual speech recordings, typically from a professional speaker. With the increasing capacity of computers, these recordings (speech corpora or databases) have become larger and larger, leaving the signal processing techniques practically aside i.e. following the so called "choose the best to modify the least". Nevertheless, this corpus based techniques must record a new database for each new voice they want to synthesize, which is a very expensive process, both in terms of time and economical cost.

In contrast, this project retakes the idea to generate speech from scratch by means of articulatory speech synthesis techniques in order to generate any kind of speech (gender, age, speaking style, etc.). Although the physics involved in speech generation is quite complex, the current capacity of computers, combined with recent advances on numerical mathematics (in particular the Finite Element Method (FEM)) makes possible to address it. Since this is a very challenging goal, we will start from the most simple speech sound: the synthesis vowels.

So, one of the techniques that deals with the above purpose is articulatory speech synthesis. In order to better understand it, we present a brief review. The articulatory synthesis concept is explained by means of a generic block diagram, whose main blocks (geometrical, glottal and vocal tract) are detailed.

This work focus on the vocal tract modelling. For this purpose, we use a computational model based on finite element methods in the time domain. Two dimensional geometries are considered. In order to deal with the complexities of the acoustic modelling, we have started by solving the standard acoustic wave equation using FEM in space and finite differences in time, deriving an explicit scheme. Next, we have increased the complexity of the problem taking into account boundary losses due to viscous friction and heat conduction. Finally, a non-reflection condition has been included, which allows to consider free space propagation of sound waves. A Perfectly Matched Layer (PML) is used for this

purpose. The above approaches are tested using benchmark problems such as the wave propagation in a tube and in a membrane.

Once solved the complexities of the acoustic modelling, we have applied the finite element method to the vowel problem. Given the /e/ vocal tract geometry and a glottal source, we synthesize as an example the vowel /e/. Then, its quality is analyzed by means of objective measurements.

Contents

1	An introduction to articulatory speech synthesis	1
1.1	Introduction	1
1.2	Vocal tract geometry	2
1.2.1	Articulatory data	2
1.2.2	Geometry models	9
1.3	Glottal models	11
1.3.1	Waveform and self-oscillating models	11
1.3.2	Computational models	12
1.4	Vocal tract acoustic models	13
1.4.1	Tube models	13
1.4.2	Computational models	15
2	Finite Element Method for acoustics	17
2.1	The acoustic wave equation	17
2.1.1	Strong form	17
2.1.2	Variational problem statement	18
2.1.3	Space and time discretization	20
2.1.4	Benchmark examples	22
2.2	The acoustic wave equation with boundary losses	27
2.2.1	Variational problem statement	27
2.2.2	Space and time discretization	28
2.2.3	Numerical example	29
2.3	The acoustic wave equation with a Perfectly Matched Layer (PML)	34
2.3.1	The Perfectly matched layer	34
2.3.2	Strong form	34
2.3.3	Variational problem statement	36
2.3.4	Space and time discretization	37
2.3.5	Numerical example	39
3	Applying FEM to the synthesis of vowels	43
3.1	Introduction	43
3.2	Some considerations on the acoustic modeling	44
3.2.1	Geometry	44

3.2.2	Glottal source	46
3.2.3	Losses	49
3.3	Computational model for the vocal tract	50
3.3.1	Model description	50
3.3.2	Numerical scheme	52
3.4	An example: synthesis of vowel /e/	53
3.4.1	Synthesis	53
3.4.2	Analysis of the results	57
3.4.3	Some remarks on vowel synthesis quality	59
4	Conclusions and future work	63
4.1	Conclusions	63
4.2	Future work	65
A	Numerical computation in 2D	69
A.1	Introduction	69
A.2	The element point of view	69
A.3	Numerical computation over a master element	70
A.3.1	Coordinate transformation and shape functions	70
A.3.2	Mapping the integrals to the reference domain	73
A.3.3	Numerical Integration	74
A.4	Calculation of the Stiffness matrix	75
A.4.1	Stiffness matrix	75
A.4.2	Memory efficiency	75
	Bibliography	77

Chapter 1

An introduction to articulatory speech synthesis

In this introductory chapter, a brief review of different issues and approaches related to articulatory speech synthesis will be presented. We will begin by explaining what is articulatory speech by means a generic block diagram. Then, in the following sections the main blocks of this diagram will be described (geometry, glottal and acoustic). First, a discussion of the most relevant techniques used to acquire articulatory data is given, as well as the main models used to construct the vocal tract geometry by means of these data. Second, some models used to emulate the behavior of the vocal cords are introduced. These models aim to provide the input airflow of the vocal tract. Finally, the different acoustic models for the vocal tract are presented and roughly compared. These models describe the acoustic behavior of the waves propagating within the vocal tract through which the synthesized speech can be obtained. This is the main goal of this work. Special attention will be paid to computational models, which constitute the core of the proposed model.

1.1 Introduction

Articulatory speech synthesis aims at producing the speech sounds by means of modelling the human vocal tract and the acoustic behavior of the sound waves travelling inside it. An articulatory speech synthesis system roughly comprise [34]: i) a module for the generation of vocal tract dynamics (control model), ii) a module for converting this dynamics information into a continuous succession of vocal tract geometries (vocal tract geometry model), and iii) a module for the generation of acoustic signals on the basis of this articulatory information (glottal and vocal tract acoustic models) (see Figure 1.1).

This project mainly focuses on this last step (acoustic modelling), and specifically on the use of finite element methods (FEM) to compute the acoustic pressure at the exit of the vocal tract (speech). However, before going into detail in the proposed model (see chapter 3), it is interesting to make a brief overview of the main elements of an articulatory speech synthesizer. In what follows, we will describe each block in Figure 1.1.

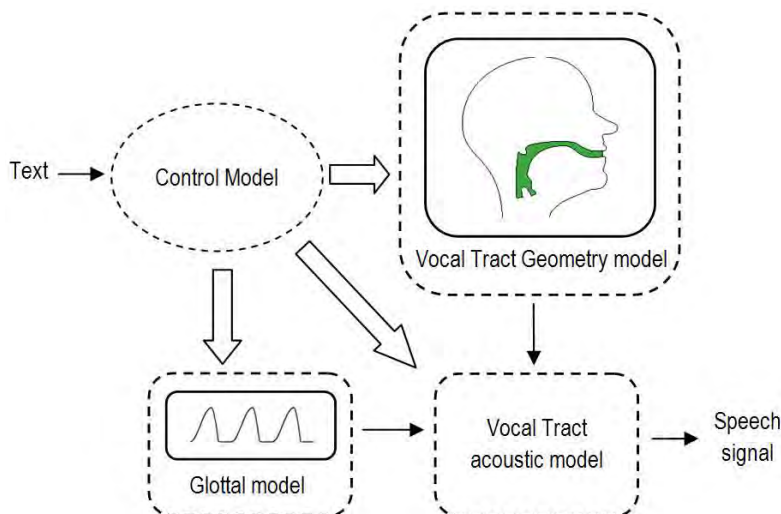


Figure 1.1: Generic block diagram of an articulatory speech synthesis system

1.2 Vocal tract geometry

If speech is to be generated from scratch, the first thing we need is a geometry that resembles the human vocal tract, in order to later compute its acoustic behavior. That is, we need to know the inner boundary surface of the vocal tract cavity when a given sound is produced. This implies knowing the shapes and positions of all vocal tract organs such as lips, jaw, tongue, palate, nasal cavity, etc. If the sound is not steady, then the geometry will evolve with time (e.g., when pronouncing a syllable) and this evolution will have to be modelled to.

To build the vocal tract geometry, two issues have to be taken into account: i) how to obtain the articulatory data and ii) how to reconstruct or approximate the vocal tract geometry from this articulatory information. In the following subsections, a brief description of the most relevant techniques is provided.

1.2.1 Articulatory data

For the construction of the geometry model, a database with articulatory information (e.g. position of all vocal tract organs) is necessary. In order to build it, there are several techniques to collect the articulatory information. The main features that should have these techniques are a good spatial and temporal resolution, and health safety. The spatial resolution is necessary for a clear distinction of all articulatory organs, and the temporal resolution for a good observation of the speech dynamics (healthy restrictions are obvious). The most relevant techniques are the X-ray, Magnetic Resonance Imaging (MRI), Computed Tomography (TC), ultrasounds and Electromagnetic Articulography

(EMA). Below there is a brief review of these tools applied to speech production (see [47]). Moreover, links to some existing databases are provided.

X-ray

X-ray was the first technique applied to articulatory speech analysis and was widely used in the earlier years of the XX century until the appearance of new tools such as MRI or EMA. This technique has a good temporal resolution, but a poor spatial resolution. Moreover, it can only acquire articulatory data in the midsagittal plane. However, X-ray can be used to acquire information of all speech organs.

Nowadays X-ray has fallen into disuse due to health safety reasons. However, some databases developed in the 60's and 70's were digitalized and are now available to the speech research community. One of them is *the X-ray film database for speech research*, also known as *ATR database*, developed by Queen's University and ATR Laboratories. This database offers 25 X-ray films in the midsagittal plane (totalling 55 minutes of footage) acquired at a rate of 50 images/s [26], with a DAT recording of the original audio tracks (see Figure 1.2a). The subjects are 14 native speakers of Canadian English or French, reading phonetically contrastive sentences. The X-ray film database is available to researchers at no cost, with a limitation of one disk per institution. Only a small fee have to be paid for the DAT recording. Some information about this database can be found in [43] and in its webpage (http://psyc.queensu.ca/~munhallk/05_database.htm). It is to be noted that this database has been under-exploited due to the tedious hand tracing needed for the analysis of such data. However, in recent articles (e.g. [26]) semi-automatic methods for extracting tract movements from X-ray films have been developed (see Figure 1.2b). This could awake more interest for ATR database given the difficulties of finding open access databases.

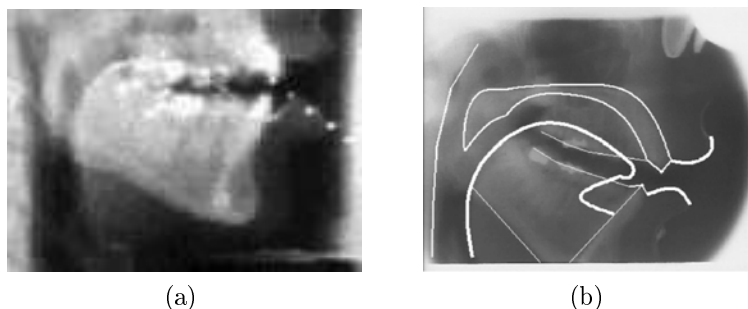


Figure 1.2: (a) An X-ray image of a 38 year old male native speaker of Canadian English producing the sentence “Why did Ken set the soggy net on top of his deck”, from the *ATR database*. (b) Complete vocal tract contour from the semi-automatic method described in [26].

Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is the most common technique given that it is the only technique that can provide three-dimensional (3D) data of the whole vocal tract, without involving any known radiation risk. Moreover, in contrast to X-ray, this technique provides a high signal-to-noise ratio and a high spatial resolution (e.g. 0.1 cm/pixel in [49]).

The MRI scan can be done in any plane: axial, coronal or sagittal. Then, using the set of images acquired in the different slices, a reconstruction in the other planes or even in any direction can be done. For example, in [49] a stack of 25 sagittal images with an inter-slice space of 0.4 cm were acquired (see Figure 1.3).

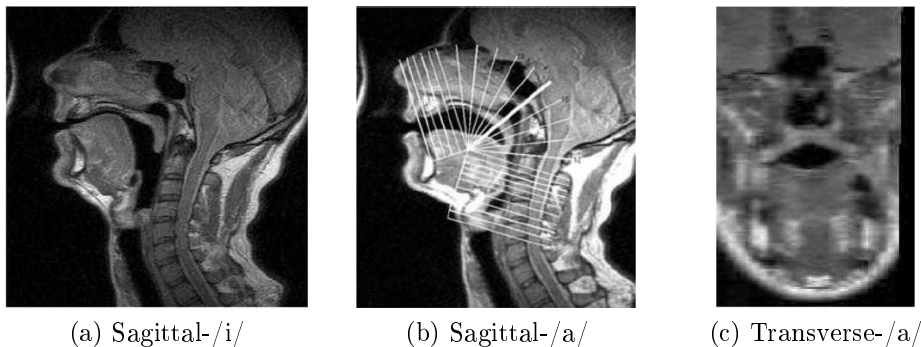


Figure 1.3: Examples of MRI images for /i/ and /a/ articulations (a) and (b) and (c) of a transverse image for /a/ reconstructed along the thick white line in (b) [49]

The main drawback of this technique, in contrast to X-ray, is the slow acquisition speed, which implies that the subject has to sustain an articulatory position artificially (e.g. 20 seconds in [53] for each vowel, 35 seconds in [49] for each articulator), restricting the use of MRI to static speech sounds (e.g. vowels, fricatives, etc.). However, this time resolution restriction has been overcome by some recent advances on MRI technology and signal processing (at the price of decreasing the signal-to-noise ratio) leading to the so-called *real-time MRI* or *cine-MRI* (e.g. in [44], 8-9 images/s are acquired and 20-24 images/s are reconstructed, overcoming the limit of 20 images/s necessary for the observation of the speech production dynamics [44]). However, these new techniques can only acquire data in the midsagittal plane, which only allows to construct two-dimensional models.

On the one hand, it is to be noted that MRI images can only clearly distinguish between soft tissues and air, but not bones. To compensate it, CT (Computed Tomography) scans are frequently used to identify the bony structures (e.g. the teeth). On the other hand, MRI also needs to be complemented with other measurements with higher temporal resolution such as EMA, ultrasounds and/or X-ray, to correctly replicate the articulatory movements.

Although there exist almost no public databases, many studies have been conducted with the aim of achieving MRI data. The most representative are:

- A database of MRI vowels called *MRI vowels image database* developed in 1998. Coronal and axial slices with a thickness of 3mm are provided for different vowels corresponding to five subjects (three males and two females). (<http://www.isle.illinois.edu/mri/>) (see Figure 1.4a).
- Some MRI images can be found in the *3D Vocal Tract project* developed in the *Center for Speech Technology (KTH)*. (<http://www.speech.kth.se/multimodal/vocaltract.html>) (see Figure 1.4b).
- Some cine-MRI examples (up to 25Hz) among others can be found in the *Vocal Tract Visualization Laboratory (University of Maryland, Baltimore)*. (<http://speech.umaryland.edu/>) (see Figure 1.4c).
- Some examples of cine-MRI with synchronized audio of sentences can be found in the *Phonetics Laboratory (Faculty of Linguistics, Philology and Phonetics, University of Oxford)*. (<http://www.phon.ox.ac.uk/mri>) (see Figure 1.4d).
- Several cine-MRI with synchronized audio of syllables can be found in the *Speech Production and Articulation kNoledge Group (SPAN) (University of Southern California)* (see Figure 1.4d). (<http://sail.usc.edu/span/video.php>) (see Figure 1.4e).
- Some MRI images corresponding to all catalan sounds (vocals and consonants) can be found in the *Laboratori de Fonètica (Universitat de Girona)*. (<http://web.udg.edu/labfon/imatge.htm>) (see Figure 1.4f).

CT (Computed Tomography)

Computed Tomography (CT) is another technique that can provide 3D data, but with an existing radiation risk. In contrast to MRI, this technique can discriminate bones and has a higher spatial resolution (e.g. 0.05 cm/pixel in CT in contrast to 0.01 cm/pixel in MRI [49]). Due to healthy restrictions, this technique can be only applied with the human vocal tract in a rest position. So, CT scans are used to obtain much detail and to distinguish the bony structures, in contrast to MRI scans that are used to acquire several articulatory positions. As for MRI, the temporal resolution is also small, but this is not a drawback given that this technique is not used for acquiring speech dynamics.

The CT scan also can be done in anyone of the three planes (as in MRI): axial, coronal and sagittal. Then, using the set of images acquired in the different slices, a reconstruction in the other planes can be done. For example, in [49] a stack of 149 axial CT images with an inter-slice space of 0.13 cm are acquired. Then the sagittal and coronal views are reconstructed. (see Figure 1.5).

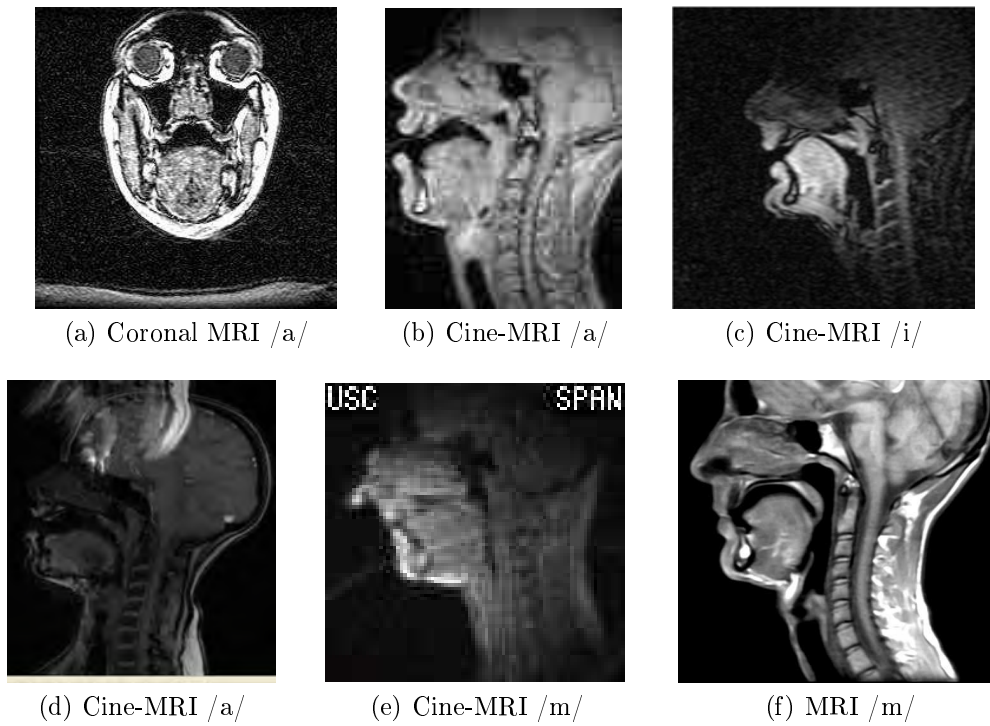


Figure 1.4: (a) Coronal MRI of /a/. (b) First frame of a Cine-MRI corresponding to the sentence “matt”. (c) First frame of a Cine-MRI (/i/) of the diphthong /ia/. (d) First frame (/a/) of the sentence “answer a door”. (e) First frame of the syllable “pai”. (f) Static MRI of /m/.

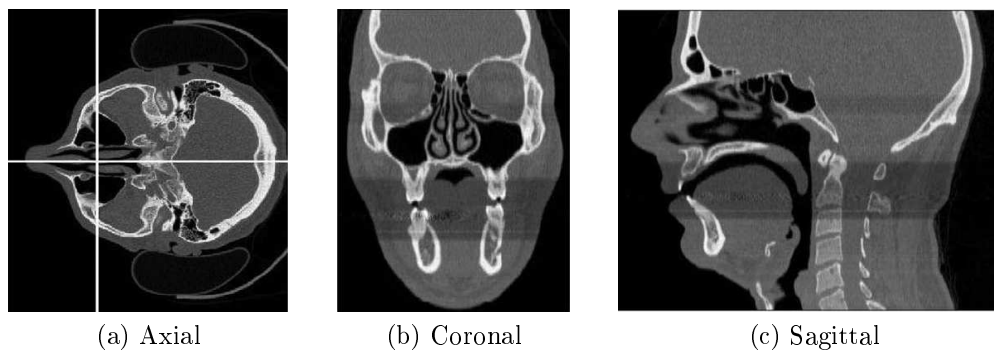


Figure 1.5: (a) An axial CT image of a rest vocal tract and the reconstructed images in the (b) coronal and (c) sagittal planes [49].

Ultrasounds

Ultrasound is a technique used to collect real-time data of the tongue surface. In the reconstructed image, the tongue surface is (more or less) visible as a white line on a black background (see Figure 1.6a). Then, a 3D reconstruction of the tongue surface can be done (see Figure 1.6b).

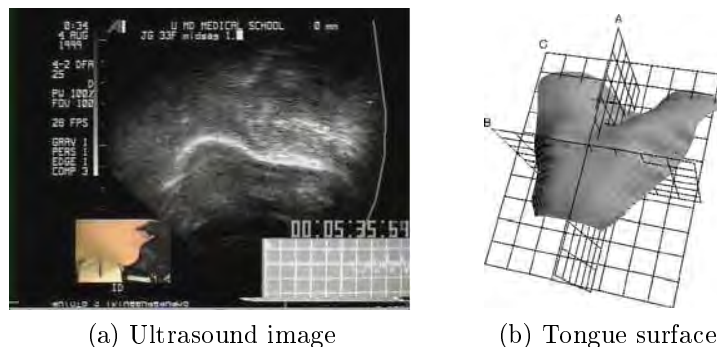


Figure 1.6: (a) /i/ midsagittal ultrasound image corresponding to the sentence “It ran a lot”, (b) a 3D reconstruction of the tongue surface during /i/, from the *Vocal Tract Visualization Laboratory*

Some videos can also be found in the *Vocal Tract Visualization Laboratory* (University of Maryland, Baltimore). (<http://speech.umaryland.edu/>).

EMA (ElectroMagnetic Articulography)

The ElectroMagnetic Articulography is a minimally invasive real-time technique for transducing the movements of specific points of the active speech organs, as the tongue, palate, lips, etc. A finite number of small coils are located inside the mouth and on some points of the face (see Figure 1.7). Then, using magnetic fields, the position of each coil can be deduced, obtaining a time evolution function of each measured point (see Figure 1.8a).

In the beginnings, this technique was only able to measure in the midsagittal plane (2D data), but nowadays, new EMA generation allows to collect real-time 3D data, like the Carstens AG500 (<http://www.articulograph.de>) (see Figure 1.7a). The main advantage of this technique is that is real-time, portable and cheap (in contrast to X-ray, MRI, CT, etc.). Although it cannot provide data for a full 3D reconstruction (it can only acquire a small finite number of points), this kind of data is useful, for example, when used for articulatory inversion (e.g. in [29,40]) or in hybrid parametric synthesis systems (e.g. in [52]).

There are many EMA database. One of them is the MultiChanel Articulatory (MOCHA) database (<http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html>), recorded at Queen Margaret University College in 1999. Its main purpose is create

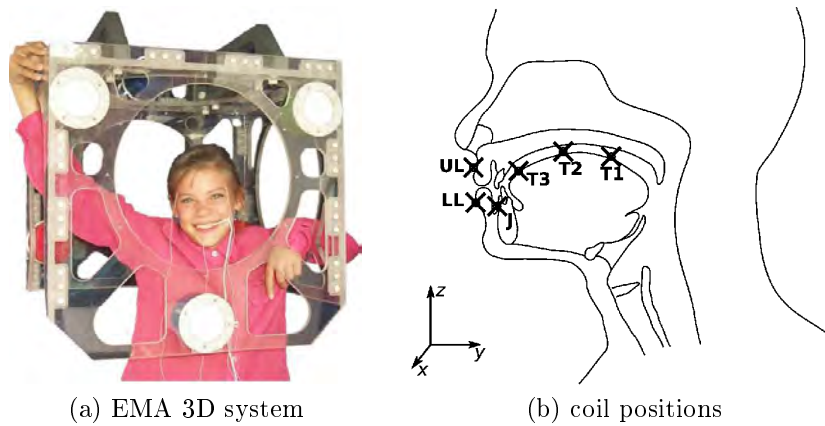


Figure 1.7: (a) EMA AG500 3D system, from Carstens Medizinelektronik GmbH (<http://www.articulograph.de>). (b) EMA coils configuration used in [52], where T1 is the tongue dorsum, T2 the tongue body, T3 the tongue tip, J the jaw, LL the Lowe lip and UL the upper lip.

a phonetically balanced dataset for training an automatic speech recognition system, but it can also be used for speech synthesis systems. This database provide 2D EMA data, acquired with the Carstens AG100 (midsagittal plane, see Figure 1.8b for coil positions) at a sample rate of 50Hz, acoustic speech waveform recorded a 16KHz, laryngograph waveform at 16KHz and Electropalatography (EPG). In Figure 1.8, the interface of the EMA database and the EMA coil configuration can be seen.

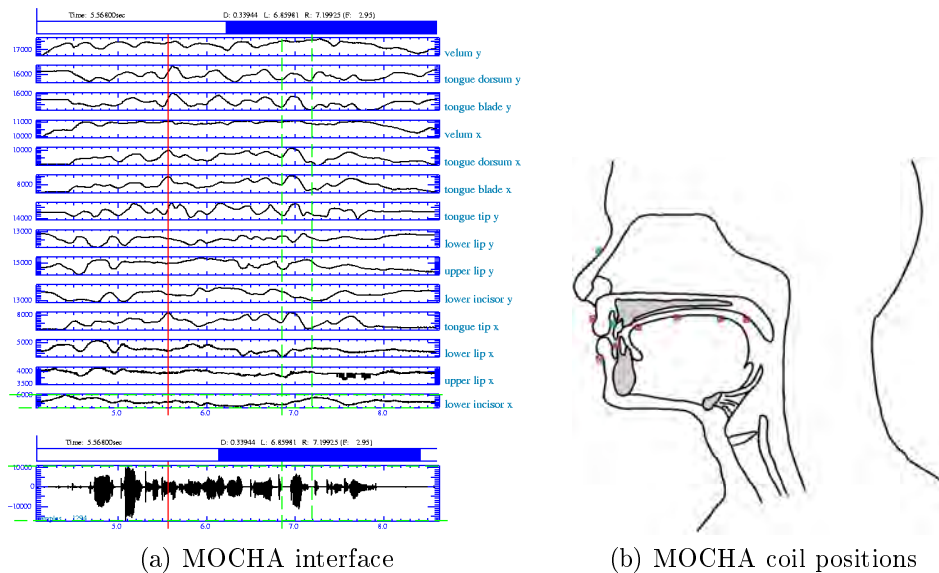


Figure 1.8: (a) MOCHA interface and (b) coil positions, where the magenta coils are the articulatory points and the cyan ones are used for head correction respect to the helmet in the EMA processing. From [46].

1.2.2 Geometry models

Depending on how the vocal tract geometry is approximated, the corresponding models can be classified as being geometrical, statistical or biomechanical [34]. Alternatively, some authors play emphasis on the distinction between 2-dimensional models that only take into account the midsagittal plane (e.g. [39, 51]), and 3-dimensional models (see e.g. [5, 11, 16, 49]), which consider the whole geometry.

Statistical models

Statistical models (see e.g. [49, 51]) obtain the vocal tract geometry from huge databases measured using different techniques such MRI (Magnetic Resonance Imaging), CT (Computed Tomography), X-ray or/and EMA (ElectroMagnetic Articulography) (see Figure 1.9). Although very precise and realistic, statistical models just reproduce the characteristics of a single speaker and become much less flexible than geometrical models. Their advantage is that they deal with a relatively small set of uncorrelated parameters.

Geometrical models

Geometrical models (see e.g. [5, 39]) rely on simulating the complex airflow in the vocal tract linking simple geometric elements (i.e., circumferences, arcs, squares...) (see Figure 1.10). The various parameters of these elements (i.e. radius, dimensions, etc.) can be modified to simulate the articulatory movements of the vocal tract. Geometric models are the most flexible ones and can be adapted to mimic any speaker's vocal tract (different age and sex) [7].

Biomechanical models

Biomechanical models (see e.g. [11, 16]) commonly make use of FEM to simulate the dynamics of the vocal tract. Therefore physiological knowledge on the relation between muscle activation and articulatory movements for speech synthesis is required. These models usually involve a large number of parameters and are difficult to control. An example of a biomechanical model developed by Artisynth staff [15] can be seen in Figure 1.11.

Additionally, some further approaches to obtain the vocal tract geometry have been recently devised such as the audiovisual-to-articulatory inversion process (e.g. [29]), which obtains articulatory data combining EMA measurements with face video acquisition and speech recording.

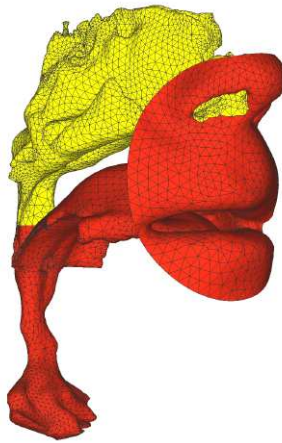


Figure 1.9: Statistical model developed by Serrurier [37]

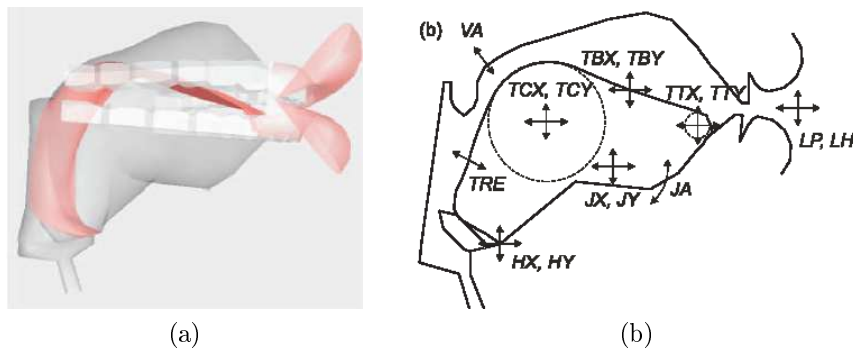


Figure 1.10: Geometrical model developed by Birkholz [7], where (a) is the 3D rendering of the geometrical model and (b) corresponds to the vocal tract parameters used by this model.

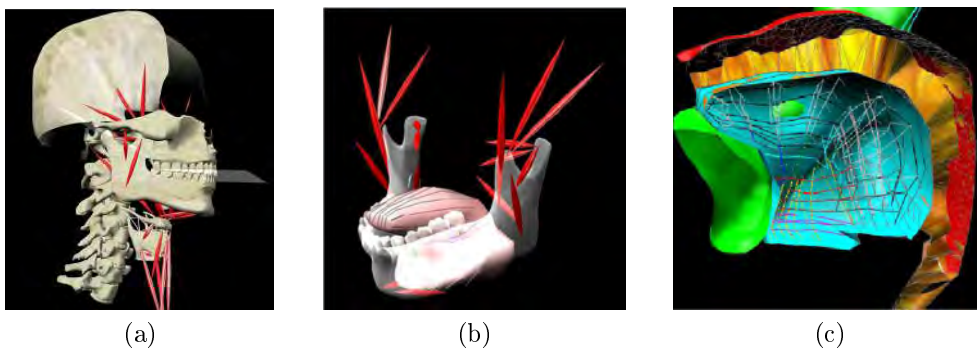


Figure 1.11: Biomechanical model developed by Artisynt staff, where (a) are the jaw and laryngeal models, (b) is the jaw model connected to the tongue model and (c) is the airway model coupled to the tongue, palate and jaw meshes [15].

1.3 Glottal models

Once defined the vocal tract geometry and prior to the simulation of its acoustics response, we need to know how sound is generated in it. This is the main goal of glottal models, which aim at simulating the behavior of the phonatory organs (vocal cords). The latter are responsible for the characteristics of the input airflow into the vocal tract. For voiced sounds (e.g. vowels), this air inflow corresponds to a train of pseudoperiodic pulses known as glottal pulses. Glottal pulses are generated by vocal cords, which act on the steady airflow coming from the lower respiratory tract (trachea, lungs, etc.). Basically, glottal models can be divided into waveform models, self-oscillating models and computational models.

1.3.1 Waveform and self-oscillating models

Waveform models

Waveform models approximate the velocity waveform of the airflow generated by the phonatory organs by means of trigonometric functions. The input parameters of these models coincide with some articulatory parameters (e.g. fundamental frequency or pitch, amplitude, etc.). The most celebrated waveform model is that of Rosenberg [48] (e.g. in [12]). Rosenberg's waveform shapes (glottal pulses) can be seen in Figure 1.12.

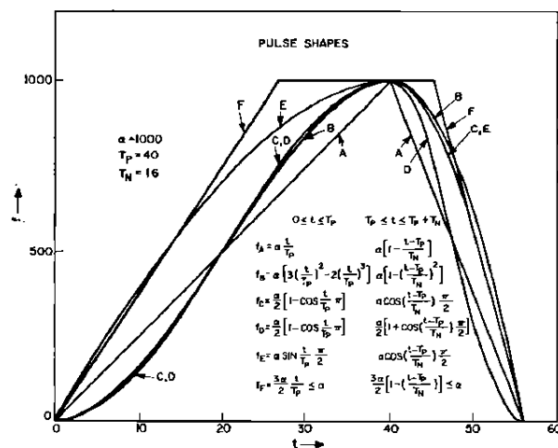


Figure 1.12: Waveform shapes from the Rosenberg model [48]

Self-oscillating models

In the case of self-oscillating models, the vocal cords behavior is modelled by means of a mechanical analogy (e.g., coupled mass-spring systems). The system solution gives place to self-sustained oscillations determining the glottal aperture (aperture between the vocal folds, aka glottis) and the glottal waveform (glottal pulses). Self-oscillating models are

controlled by biomechanical parameters like the air pressure provided by the lungs, the tension of the vocal cords, etc. The simplest of these models are the one-mass model [17] (see Figure 1.13) and the two-mass model [25] (e.g. in [50]).

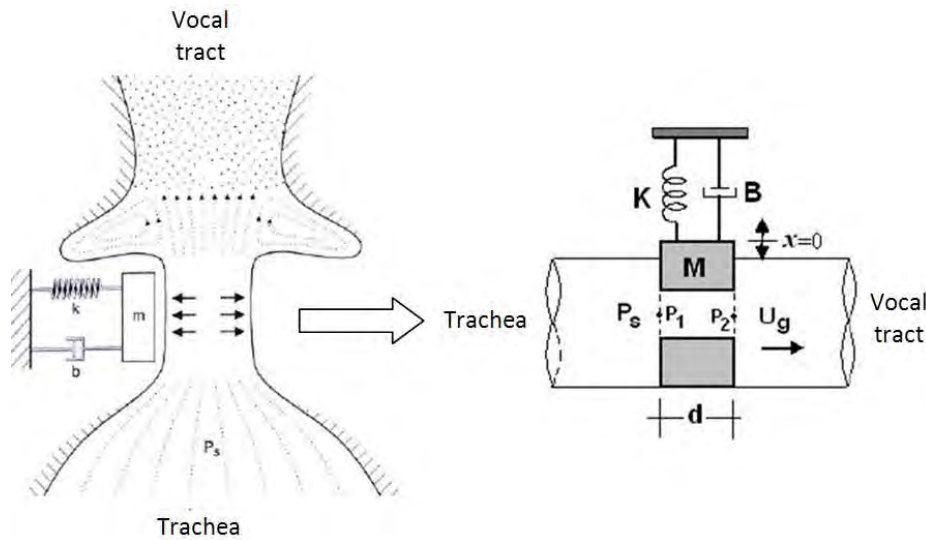


Figure 1.13: One-mass model scheme [8]

It is worthwhile noting that although Rosenberg and two-mass models were developed quite a long time ago, they are still very popular because they combine simplicity with very acceptable results (see e.g. [12]).

1.3.2 Computational models

Much research has also been placed to understand the fluid mechanics and aeroacoustics of the vocal tract airflow, both from a theoretical analysis of the involved physics [30, 32], and from the results of simple scale models as well [31]. Analytical approaches based on the Green's function solution of the corresponding partial differential equations (vorticity formulation of Lighthill's analogy approach) have been also attempted [21, 22, 38]. These approaches have the advantage of easy parameter space exploration, but are only suitable for very simplified geometries. Further remarkable insight has been gained by resorting to computational approaches to glottal models. For instance, in [56, 57], a direct numerical simulation of the compressible Navier-Stokes equations using a finite difference scheme was carried out, and its results compared to those of applying an acoustic analogy (the Ffwocs-Williams Hawkins analogy was used in this case). However, it has not been until very recently, that a finite element method to solve the coupled equations for the mechanics, fluid dynamics and the acoustics of a 2-dimensional glottal system has been presented [35]. The use of FEM to address the problem seems to be the most promising way to deal with all the complex physical phenomena involved in the generation of human voice.

1.4 Vocal tract acoustic models

Once having the vocal tract geometry and once implemented the glottal model to simulate the source of sound (vocal tract inflow), we can focus on simulating the acoustics of the vocal tract, which will finally yield the synthesized voice. Basically, two main types of acoustic models can be distinguished: tube models and computational models.

1.4.1 Tube models

For historical reasons and thanks to their simplicity, tube models have become widely known and used. Duct models approximate the vocal tract geometry as a finite set of concatenative tubes, each one having constant cross section [34] (see Figure 1.14).

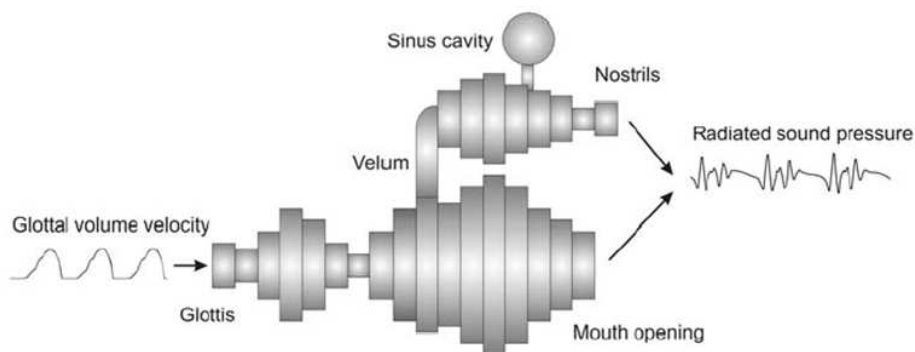


Figure 1.14: Schematics of a tube model for a whole vocal tract (nasal and vocal cavities are considered) [33]

Usually, the geometry block only provides an area function: a time-varying function that describes the constant cross section of each tube, which is computed as the area of each section perpendicular to the midline of the vocal tract airway (see e.g. [2]). Duct or tube models can be mainly subdivided into the ABCD matrix based models, the Digital Waveguides models (aka KL models) and the circuit analogy models. The former correspond to hybrid time-frequency domain models, the second to reflection type line models, and the latter mimic transmission line circuit models [34].

ABCD matrix based models

Tube models of the ABCD matrix type (e.g. [50]) compute the acoustic transfer function of the vocal tract as the products of individual transfer functions for each elemental tube. To obtain them, a matrix that links the tube's input with its output, known as the ABCD matrix, is used. This matrix is such that

$$\begin{pmatrix} P_{out} \\ U_{out} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} P_{in} \\ U_{in} \end{pmatrix} \equiv \mathbf{K} \begin{pmatrix} P_{in} \\ U_{in} \end{pmatrix}, \quad (1.1)$$

where P is the pressure, U is the volume velocity, the subscripts *in* and *out* refer to the tube's input and output, and \mathbf{K} is the ABCD matrix with components A , B , C and D , which are computed using classical acoustic duct theory. Then, using the ABCD matrix K , the duct transfer function is computed. Finally, the speech signal is calculated in the frequency domain using the vocal tract transfer function and the flow source provided by the glottal model. In contrast to time domain models, these models do not directly provide the acoustic pressure or the acoustic velocity within the vocal tract. Consequently, time-frequency transformations must be used to obtain their time variations.

Waveguide models

In waveguide models, the d'Alembertian solution (backward and forward signals) of the one-dimensional acoustic equation is used to emulate the behavior of the acoustic wave propagation within the vocal tract. Hence, each tube is approximated as a digital waveguide, made of bidirectional delay units to emulate the wave propagation in the time domain. Then, on the basis of scattering equations that reflect the impedance discontinuity at tube junctions, forward and backward travelling flow waves are computed in each tube (see Figure 1.15). The main problem of these models is that the geometry cannot be changed smoothly [12], which is essential for coarticulation processes such as the synthesis of diphthongs (e.g. /ei/) and syllables (e.g. /na/, /sa/).

The simplest model is the Kelly-Lochbaum model [28] (aka KL model), which can only use one-dimensional geometries. However, some complex models have been developed in order to achieve synthesised speech with more complex geometries (e.g. [42] for 2-dimensional geometries).

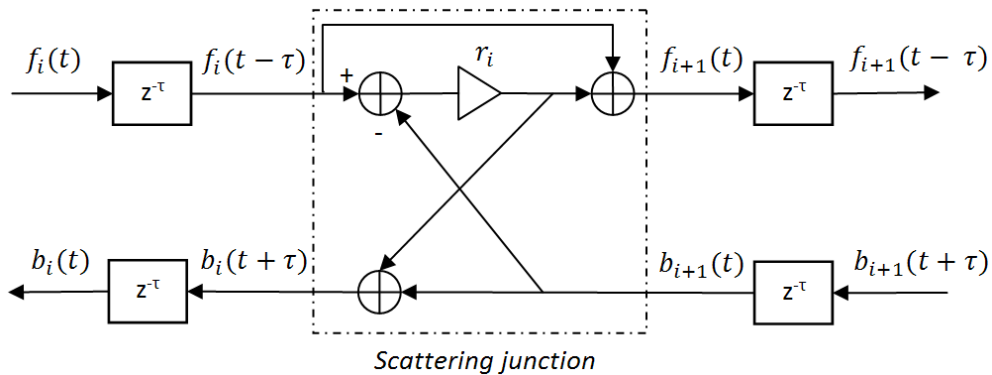


Figure 1.15: KL schematics of two concatenated tubes with different section, where r is the reflection coefficient of the scattering junction, τ is the delay introduced by the delay line, and f and b denote the forward and backward signals respectively.

Circuit analogy models

In tube models based on circuit analogies (see e.g. [6,49]), the volume velocity and pressure waves are respectively interpreted as intensity and voltage signals [6], and the acoustic properties of each tube are modelled by an electrical circuit analogy, obtaining a two-port network representation for each tube (see Figure 1.16a). Then, the whole vocal tract can be represented by a chain of these two-port networks (see Figure 1.16b).

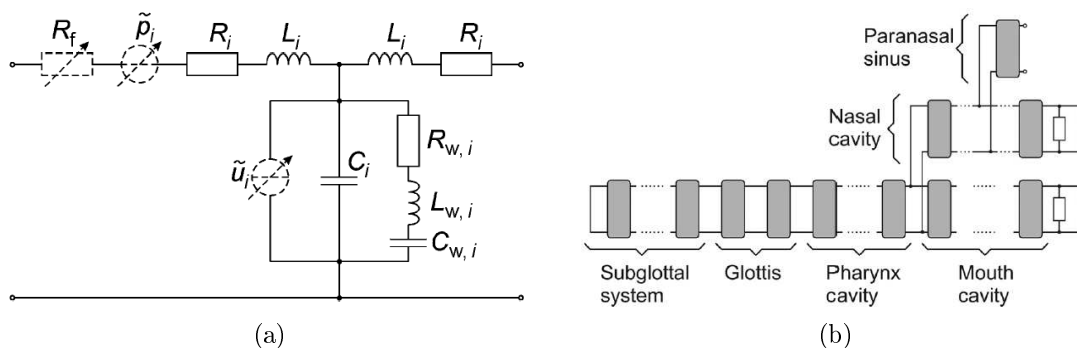


Figure 1.16: (a) Two-port circuit network equivalent to one tube section. L_i is the inertance of the mass of air in the tube section i , C_i represents its compressibility, and R_i accounts for energy lost to viscous friction at the tube walls. The $R_{w,i} - L_{w,i} - C_{w,i}$ circuit models the elasticity of the vocal tract walls. The optional elements \tilde{u}_i , \tilde{p}_i and R_f constitute a volume velocity source, a pressure source and a resistance for the kinetic pressure drop at the main constriction [6]. (b) Circuit model for the entire vocal tract system. Each gray box represents a two-port network [6].

These models compute the speech signal in the time domain, but they need to do a lot of approximations in the electric analogy process [34] (this is the case for instance, of the electrical analogue for the propagation of frequency-dependent waves into free space, which is necessary to account for the propagation of speech emanating from the mouth). In contrast to waveguide models, circuit models can assume time-varying geometry lengths [34].

1.4.2 Computational models

The use of computational models for the acoustics of the vocal tract offers again wider possibilities than the presented models. Complex geometries can be implemented in full detail, coarticulation can be included, and the aeroacoustics involved in the generation of many sounds can be taken into account.

Simple 1-dimensional models that require low computational cost have been already applied to the synthesis of diphthongs [12]. These were based on the solution of the flow momentum equations using a finite volume approach to perform a space-time discretization. However, the use of 1-dimensional models requires many artifacts to correctly reproduce the physics of speech synthesis. Although finite difference schemes

have been commonly used at an initial stage, the appropriate numerical method to address the whole complexity of the vocal tract acoustics clearly seems to be the finite element method (FEM), applied to 2 and 3-dimensional geometries. Given that the application of FEM to speech synthesis is rather new, there is still much work to be done. For the moment, efforts seem to have been focused on frequency domain methods that perform a modal analysis of the vocal tract using the Helmholtz equation (see Figure 1.17). This is used to compute the frequency response function of the latter, which can be combined with a glottal model to obtain a source-filter type model for speech synthesis [27]. Hence, it should be noted that frequency domain models can not directly synthesize speech.

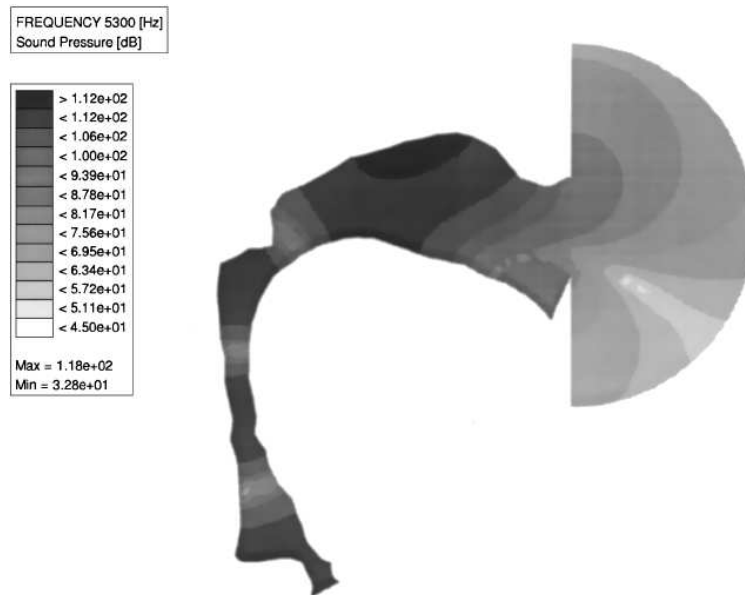


Figure 1.17: An example of the pressure distribution for vowel /a/ at 5300 Hz computed through the Helmholtz equation [41].

On the other hand, frequency domain models cannot deal with the most natural aspects of speech production such as coarticulation of sounds and time variations of the glottal inflow (pitch, intensity) [4]. Only steady states can be considered. Consequently, if the above aspects are to be considered, it will be necessary to work in the time domain. This is the aim of this project. Some steps in this direction has been recently done. In [53] 3-dimensional models using finite element methods have been developed in time domain and static vowels are synthesized solving the classical acoustic equation. However, as for as we know, no dynamic computational models have been developed.

It is the main goal of this project to use FEM to directly compute the time dependent acoustic pressure at the output of the vocal tract (synthesized speech).

Chapter 2

Finite Element Method for acoustics

In this chapter, we will describe some fundamentals on finite element methods (FEM) that will be later needed for the synthesis of vowels. First, we will show how to solve the acoustic wave equation using FEM. Second, we will deal with losses in the domain boundaries due to friction and heat conduction of the acoustic waves. Finally, a non reflection condition will be addressed, which will allow to consider propagation of sound waves towards infinity. A perfectly matched layer (PML) will be used for this purpose. For each involved equation the corresponding weak form will be solved. These will be discretized in space (FEM) and time (finite differences), resulting in an explicit scheme. Finally, these numerical schemes will be tested using benchmark problems such as wave propagation in a membrane or in a tube.

2.1 The acoustic wave equation

2.1.1 Strong form

As first step, we will need to compute the sound wave propagation in a given open or closed domain. We will consider sound waves being solution of the hyperbolic wave equation

$$(\partial_{tt}^2 - c_0^2 \nabla^2) p(x, t) = c_0^2 f(x, t), \quad \text{in } \Omega, \quad t > 0 \quad (2.1)$$

where c_0 stands for the sound speed, $p(x, t)$ is the sound pressure, $f(x, t)$ is the external force and $\partial_t = \partial/\partial t$. To solve (2.1) both initial and boundary conditions are required. Let $\partial\Omega$ denote the boundary of the domain Ω . The boundary $\partial\Omega$ can be split into two boundary regions Γ_D and Γ_N such that $\partial\Omega = \Gamma_D \cup \Gamma_N$ (see Figure 2.1), with Γ_D and Γ_N respectively standing for the Dirichlet and Neumann boundaries.

The Dirichlet and Neumann boundary conditions are defined as

$$p(x, t) = g_D(x, t) \quad \text{on } \Gamma_D, \quad t > 0, \quad (2.2)$$

$$\nabla p(x, t) \cdot n = g_N(x, t) \quad \text{on } \Gamma_N, \quad t > 0. \quad (2.3)$$

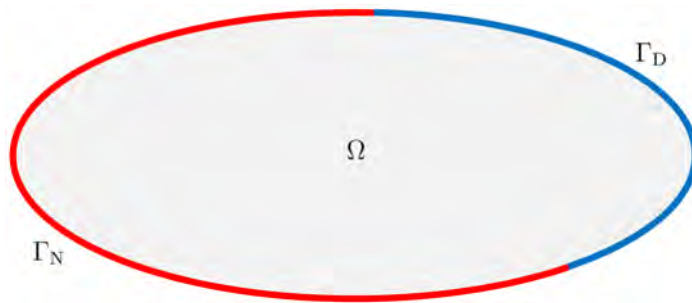


Figure 2.1: Boundaries Γ_D (blue) and Γ_N (red) of a domain Ω

The first is used to impose pressure conditions on the boundary and the later to impose the pressure gradient, which corresponds to velocity fluctuations. For simplicity, we consider that $g_D(x, t)$ and $g_N(x, t)$ are constant in their boundaries. So, the boundary conditions that we will consider are

$$p(x, t) = g_D(t) \quad \text{on } \Gamma_D, \quad t > 0, \quad (2.4)$$

$$\nabla p(x, t) \cdot n = g_N(t) \quad \text{on } \Gamma_N, \quad t > 0. \quad (2.5)$$

On the other hand, for simplicity we will consider the following initial conditions

$$p(x, 0) = 0, \quad \text{in } \Omega, \quad (2.6)$$

$$\partial_t p(x, 0) = 0, \quad \text{in } \Omega. \quad (2.7)$$

2.1.2 Variational problem statement

Functional framework

Prior to establish the weak or variational form of the problem, it is necessary to introduce the functional framework that will be used through this work. As usual $L^2(\Omega)$ will stand for square integrable functions,

$$L^2(\Omega) := \left\{ f : \Omega \rightarrow \mathfrak{R} \mid \int_{\Omega} |f|^2 d\Omega < \infty \right\} \quad (2.8)$$

$L^2(\Omega)$ is a Barach space with norm

$$\|\mathbf{u}\|_{L^2} := \left(\int_{\Omega} |\mathbf{u}(\mathbf{x})|^2 d\mathbf{x} \right)^{\frac{1}{2}} \quad (2.9)$$

and also a Hilbert space with the inner product of two functions $f, g \in L^2(\Omega)$ being given by

$$(f, g) := \int_{\Omega} fg d\Omega. \quad (2.10)$$

The L^2 norm can then be written as $\|\mathbf{u}\|_{L^2} = (\mathbf{u}, \mathbf{u})^{\frac{1}{2}}$. On the other hand, if f, g are functions such that the product fg is integrable, we will denote by $\langle \cdot, \cdot \rangle$ the integral

$$\langle f, g \rangle := \int_{\Omega} fg d\Omega. \quad (2.11)$$

In the particular case of $f, g \in L^2$, $\langle \cdot, \cdot \rangle$ becomes the inner product (\cdot, \cdot) . Next, let us introduce the Sobolev spaces H^1 and H_0^1

$$H^1(\Omega) := \{f \mid f \text{ and } \partial_i f \in L^2(\Omega)\}, \quad (2.12)$$

$$H_0^1(\Omega) := \{f \in H^1(\Omega) \mid f = 0 \text{ in } \Gamma_D\}, \quad (2.13)$$

i.e. H^1 is a space of functions whose elements and elements first derivatives are both square integrable, and $H_0^1 \subset H^1$ contains the functions in H^1 that also vanish on the boundary Γ_D . Its associated inner product is

$$(\mathbf{u}, \mathbf{v})_{H^1} = \frac{1}{L}(\mathbf{u}, \mathbf{v}) + (\nabla \mathbf{u}, \nabla \mathbf{v}), \quad (2.14)$$

where L is a characteristic length. The H^1 norm is given by $\|\mathbf{u}\|_{H^1} = (\mathbf{u}, \mathbf{u})_{H^1}^{\frac{1}{2}}$. Finally, to take into account the time evolution of the pressure, use will be made of the space

$$L^2(0, T; X(\Omega)) := \left\{ f : (0, T) \rightarrow X(\Omega) \left| \int_0^T \|f\|_X^2 dt < \infty \right. \right\}, \quad (2.15)$$

where $X(\Omega)$ can be any of the above spatial functional spaces.

Pressure and trial space of functions

Once defined the general functional framework, we can identify the space of functions \mathcal{P} for the pressure and the space \mathcal{Q} for the test function such that

$$\mathcal{P} = L^2(0, T; H^1(\Omega)), \quad (2.16)$$

$$\mathcal{Q} = H_0^1(\Omega). \quad (2.17)$$

These space of functions will be needed for the weak formulation of the acoustic wave equation.

Variational form

Taking into account the above functional framework, the variational or weak problem consists in finding $p \in \mathcal{P}$ such that

$$(q, \partial_{tt}^2 p) + c_0^2 (\nabla q, \nabla p) - c_0^2 \langle q, g_N \rangle_{\Gamma_N} = c_0^2 \langle q, f \rangle, \quad \forall q \in \mathcal{Q}. \quad (2.18)$$

(2.18) is found as follows. Multiply the wave equation (2.1) by the test function $q \in \mathcal{Q}$ and integrate it on Ω to get

$$\int_{\Omega} q \partial_{tt}^2 p \, d\Omega - c_0^2 \int_{\Omega} q \nabla^2 p \, d\Omega = c_0^2 \int_{\Omega} q f \, d\Omega. \quad (2.19)$$

Then, integrating (2.19) by parts and applying the divergence theorem it follows

$$\int_{\Omega} q \partial_{tt}^2 p \, d\Omega - c_0^2 \int_{\Gamma_D} q \nabla p \cdot \hat{n} \, d\Gamma_D - c_0^2 \int_{\Gamma_N} q \nabla p \cdot \hat{n} \, d\Gamma_N + \int_{\Omega} \nabla q \nabla p \, d\Omega = c_0^2 \int_{\Omega} q f \, d\Omega. \quad (2.20)$$

Given that $q \in H_0^1$, inserting boundary condition (2.5) and using definitions (2.10) and (2.11), yields

$$(q, \partial_{tt}^2 p) + c_0^2 (\nabla q, \nabla p) - c_0^2 \langle q, g_N \rangle_{\Gamma_N} = c_0^2 \langle q, f \rangle. \quad (2.21)$$

2.1.3 Space and time discretization

Spatial discretization: Galerkin approximation

Being $\mathcal{P}_h \subset \mathcal{P}$ and $\mathcal{Q}_h \subset \mathcal{Q}$ finite subspaces, the spatial discretized scheme consists in finding $p_h \in \mathcal{P}_h$ such that

$$(q_h, \partial_{tt}^2 p_h) + c_0^2 (\nabla q_h, \nabla p_h) - c_0^2 \langle q_h, g_{N,h} \rangle_{\Gamma_N} = c_0^2 \langle q_h, f_h \rangle, \quad \forall q_h \in \mathcal{Q}_h \quad (2.22)$$

where $g_{N,h}$ and f_h are discretizations of g_N and f respectively. If we expand $p_h \in \mathcal{P}_h$ and $q_h \in \mathcal{Q}_h$ as

$$p_h(\mathbf{x}, t) = \sum_b N^b(\mathbf{x}) P^b(\mathbf{x}, t), \quad (2.23)$$

$$q_h(\mathbf{x}) = \sum_a N^a(\mathbf{x}) Q^a(\mathbf{x}), \quad (2.24)$$

where $N(\boldsymbol{x})$ are the shape functions and P^b and Q^a are the nodal values, and introduce the above equations (2.23–2.24) into (2.22) we get

$$\begin{aligned} \sum_a \sum_b Q^a (N^a, N^b) \ddot{P}^b &= -c_0^2 \sum_a \sum_b Q^a (\nabla N^a, \nabla N^b) P^b \\ c_0^2 + \sum_a Q^a \langle N^a, g_{N,h} \rangle_{\Gamma_N} &+ c_0^2 \sum_a Q^a \langle N^a, f_h \rangle, \end{aligned} \quad (2.25)$$

(the double dot denotes second order time derivative). Expressing (2.25) in matrix form

$$\mathbf{Q}^\top \mathbf{M} \ddot{\mathbf{P}} = c_0^2 \mathbf{Q}^\top \mathbf{L} - c_0^2 \mathbf{Q}^\top \mathbf{K} \mathbf{P}, \quad (2.26)$$

where \top denotes the transpose of a vector and \mathbf{P} and \mathbf{Q} are vectors of nodal values

$$\mathbf{P} = [P^b] = (P^1 \ P^2 \ \dots \ P^N)^\top, \quad (2.27)$$

$$\mathbf{Q} = [Q^a] = (Q^1 \ Q^2 \ \dots \ Q^N)^\top, \quad (2.28)$$

being N the total number of nodes in the mesh. Cancelling \mathbf{Q}^\top in (2.25) yields the final algebraic system

$$\mathbf{M} \ddot{\mathbf{P}} = c_0^2 \mathbf{L} - c_0^2 \mathbf{K} \mathbf{P}, \quad (2.29)$$

where \mathbf{M} is the mass matrix, \mathbf{K} is the stiffness matrix and \mathbf{L} the load vector. The corresponding entries are (see appendix A for its numerical computation)

$$\mathbf{M} = [M^{ab}], \quad M^{ab} = (N^a, N^b), \quad (2.30)$$

$$\mathbf{K} = [K^{ab}], \quad K^{ab} = (\nabla N^a, \nabla N^b), \quad (2.31)$$

$$\mathbf{L} = [L^a], \quad L^a = \langle N^a, g_{N,h} \rangle_{\Gamma_N} + \langle N^a, f_h \rangle. \quad (2.32)$$

Time discretization: Finite differences

Next, a time discretization is carried out. We use a second order finite difference scheme to approximate the second time derivative

$$\ddot{\mathbf{P}} = \frac{\mathbf{P}^{n+1} - 2\mathbf{P}^n + \mathbf{P}^{n-1}}{\Delta t^2} + \Theta(\Delta t^2), \quad (2.33)$$

where the superindex n denotes the time step, and $\Theta(\Delta t^2)$ is the error introduced by the scheme, which is proportional to Δt^2 . Introducing the second order finite difference scheme (2.33) into the matrix Garlekin expression (2.29), we get the following explicit scheme for the evolution of the nodal acoustic pressure:

$$\mathbf{P}^{n+1} = c_0^2 \Delta t^2 \mathbf{M}^{-1} (\mathbf{L}^n - \mathbf{K} \mathbf{P}^n) + 2\mathbf{P}^n - \mathbf{P}^{n-1} \quad (2.34)$$

2.1.4 Benchmark examples

In this section we will apply the scheme 2.34 to solve some benchmark examples.

Square membrane

This example consists of a squared membrane of 1m^2 fixed by the edges with a inner circle where some oscillations are introduced. This system could represent, for example, a loudspeaker being a squared membrane.

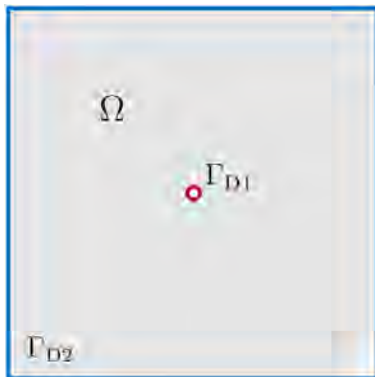


Figure 2.2: Domain Ω with inner (Γ_{D1} , colour red) and outer (Γ_{D2} , colour blue) boundaries.

So, we have a 1m^2 squared domain Ω with a circle of radius 1cm in its center. Boundaries Γ_{D1} and Γ_{D2} correspond the internal and external boundaries respectively (see Figure 2.2). Because the membrane has fixed edges, the displacement on Γ_{D2} will be zero. On the other hand, oscillations are introduced at the inner boundary, which correspond time dependent Dirichlet condition. Hence,

$$p(x, t) = g_D(t) \quad \text{on } \Gamma_{D1}, \quad t > 0, \quad (2.35)$$

$$p(x, t) = 0 \quad \text{on } \Gamma_{D2}, \quad t > 0. \quad (2.36)$$

In this example, the following smooth function [18] is used for g_D

$$g_D(t) = \frac{d}{dt} \left(e^{-\pi^2(F_0 t - 1)^2} \right), \quad (2.37)$$

where F_0 is the frequency of the pulse. In Figure 2.3 the above pulse with $F_0 = 1000\text{Hz}$ (the used frequency in this example) can be observed. This pulse has one positive peak followed by a negative one, and finally tends to zero. This type of pulses are often used to test numerical schemes (e.g. [18]), because they resemble a punctual force. However, the key point of this pulse is that it is very smooth (i.e. it contains no abrupt transitions). This ensures that no high frequency energy is introduced in the numerical scheme, which could cause numerical instabilities and numerical errors in case the mesh not being fine enough to capture this information.

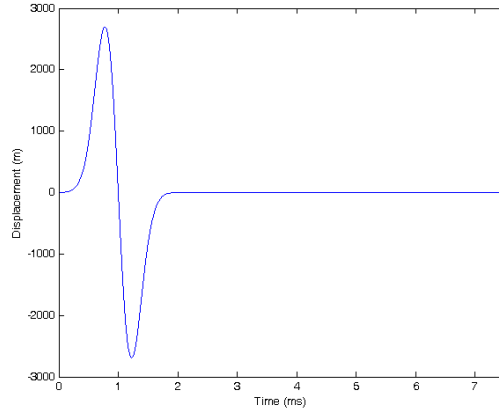


Figure 2.3: Representation of the pulse (2.37) with frequency $F_0 = 1000\text{Hz}$.

Once the problem statement has been defined, we will proceed to construct the computational domain Ω and to mesh it (see Figure 2.4). The geometry has been divided in different surfaces to ensure a good transitioning mesh. A non-uniform mesh of triangular elements have been built with element sizes $h=0.015\text{m}$ for surfaces and $h=0.003\text{m}$ for the inner boundary (over lines). Finally, we have got a mesh with 7310 nodes and 14332 elements.

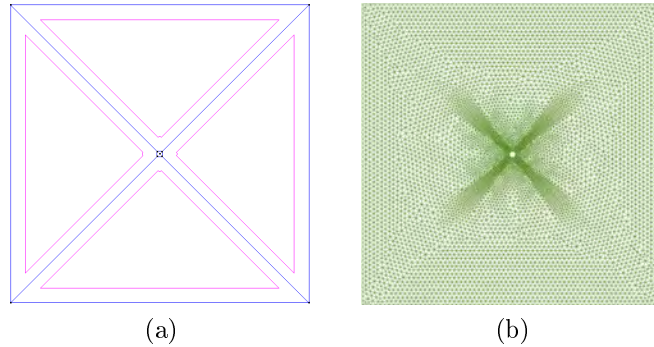


Figure 2.4: Geometry (a) and mesh (b) obtained for the membrane example.

Because our final goal is to be able to perform speech synthesis, we use as sound speed c the velocity of sound waves in the air at a temperature of 24°C ($c = 345\text{m/s}$). On the other hand, given that we use an explicit scheme, we have to use a sampling frequency f_s (the sampling frequency is the inverse of the time step Δt) such that the stability condition is satisfied. Stability is guaranteed wherever the link between Δt and the element size h given by the Courant-Friedrich-Levy (CFL) number [1]

$$\text{CFL} = c \frac{\Delta t}{h} \quad (2.38)$$

fulfills $CFL < 1$. If $CFL > 1$ the numerical scheme becomes unstable. However, it has to be noted that a $CFL < 1$ does not fully ensures stability. In this example we take $f_s = 200\text{KHz}$. Using $h = 0.003\text{m}$ and $c = 345\text{m/s}$, the CFL number will be $CFL = 0.575 < 1$. We have tested in a long simulation (1 second) that the numerical solution under the above description becomes stable in time. It has to be noted, that given that the used time step is so small ($\Delta t = 1/200\text{KHz}$), 1 second of simulation implies 200000 time steps, which means that a simulation of 1s is a long simulation.

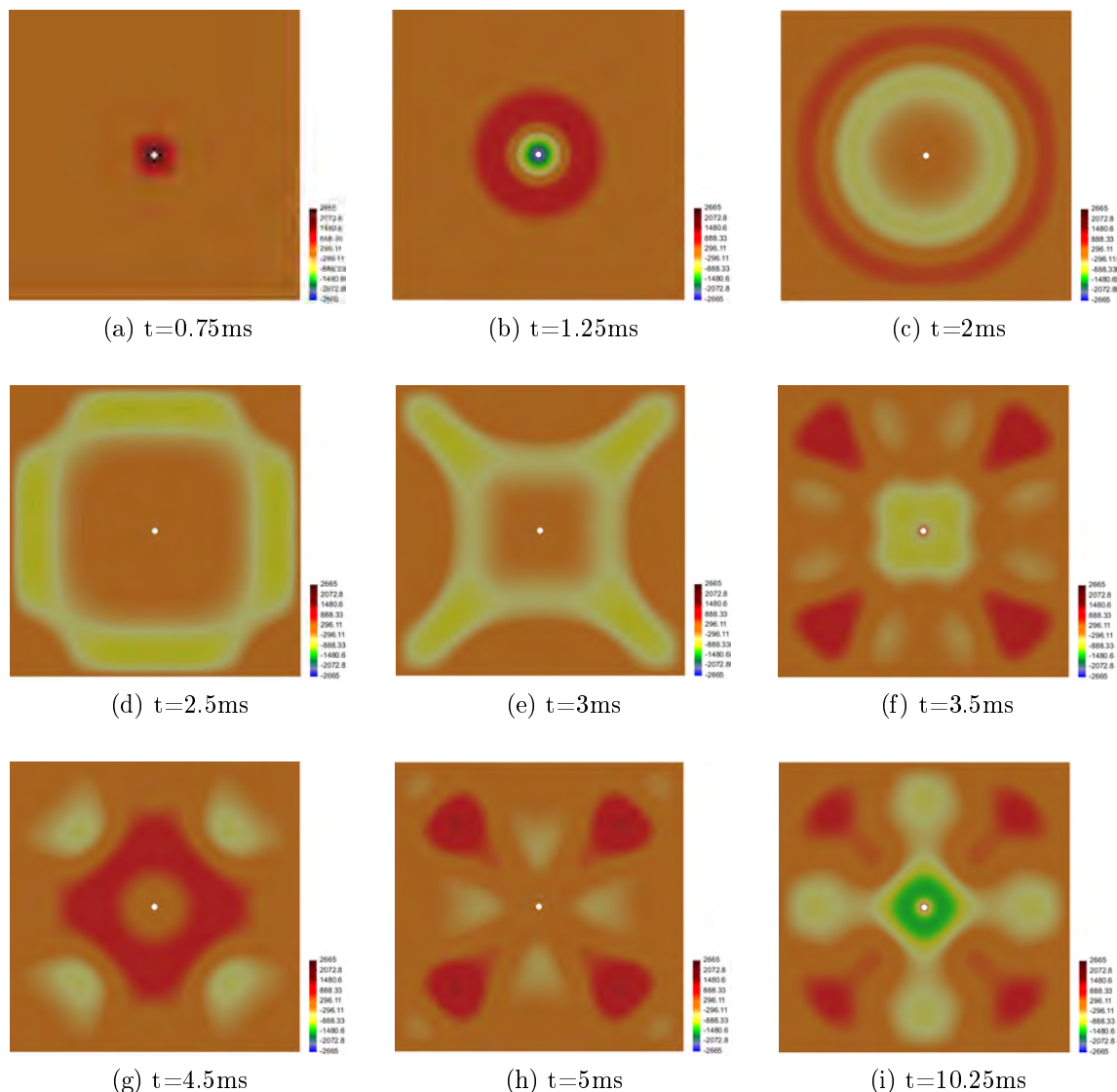


Figure 2.5: Snapshots of the numerical solution of the membrane at different time instants.

Once explained the framework and configuration of the example, in Figure 2.5 some results corresponding to the first milliseconds can be seen. In the first frame (a), we can observe the initial spherical wave front generated by the boundary Γ_{D2} . By this instant,

the positive peak of the smooth function (2.37) has been generated. In the next frame (b), the negative peak appears in the domain. Then, in (c) the whole pulse has been generated by Γ_{N2} and the wave front is near the boundary Γ_{D1} , where it will reflect. In the frames (c)-(h) we can see how the wave front reflects on the domain edges and interact with the other reflected waves, causing constructive and destructive interferences. Finally, given that the used formulation does include any dissipation mechanism, these interactions in (h) do not vanish with time.

Tube

As seen in chapter 1, the human vocal tract can be modeled as a finite number of concatenated tubes. So, we have considered that the acoustic behavior of a tube could be an interesting benchmark problem. In this example, we study the acoustic behavior of a tube closed by its extremes and excited in one end by means of a piston. The dimensions of this tube are 0.5m x 0.1m. We have considered a rectangular domain Ω with boundaries Γ_{N1} and Γ_{N2} , where Γ_{N2} corresponds to the rigid walls of the tube and Γ_{N1} to the piston wall (see Figure 2.6).

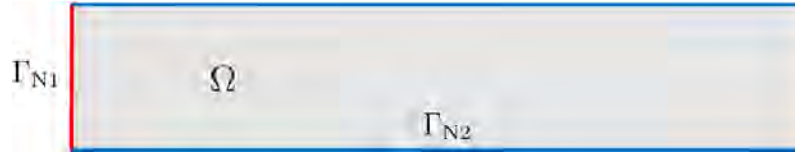
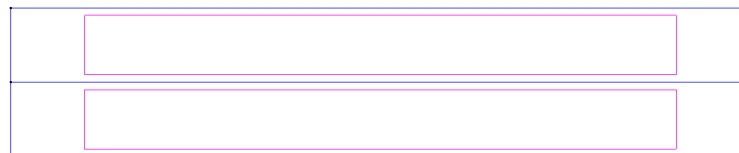
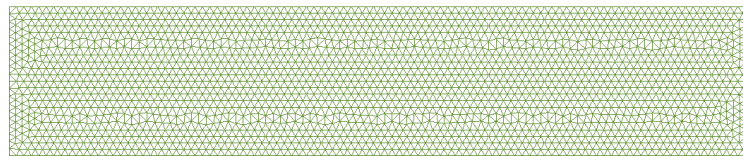


Figure 2.6: Domain Ω corresponding to a tube with rigid walls (Γ_{N2} , colour blue) and a piston in its beginning (Γ_{N1} , colour red).

Once again, like in the previous example, we have used the smooth pulse (2.37), but in this case with a frequency $F_0 = 2000\text{Hz}$ and inverted (i.e. multiplied by -1). Moreover we take $c = 345 \text{ m/s}$ and $f_s = 200\text{KHz}$.



(a)



(b)

Figure 2.7: Geometry (a) and mesh (b) obtained for the tube example.

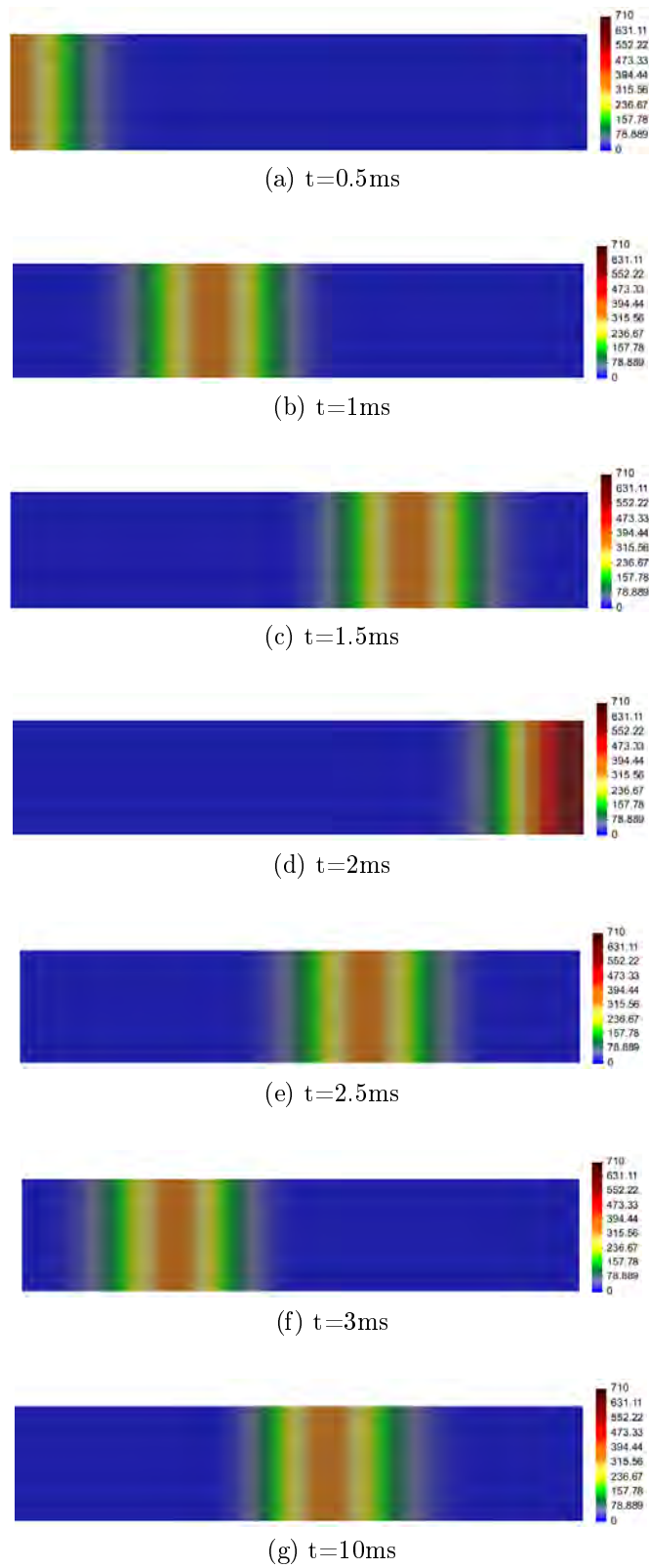


Figure 2.8: Snapshots of the acoustic pressure at the tube for different time instants.

We have also generated a non structured mesh of triangular elements with $h=0.005\text{m}$ in surfaces (no specification is done over lines). The final geometry and mesh can be seen in Figure 2.7.

Some numerical results for different time steps can be seen in Figure 2.8. In the first frame (a), a plane wave front is generated at boundary Γ_{N1} . This plane wave front is done due to Huygens phenomena. Then, this plane wave travels through the tube (b)-(c) and arrives to the end (d), where it reflects and return to the beginning (e)-(f). Once again, given that there are no losses, this wave front will go back and forward without deformations (g).

2.2 The acoustic wave equation with boundary losses

2.2.1 Variational problem statement

It will be interesting for our purposes to consider the acoustic wave equation with boundary losses due to viscous frictions and heat conduction. We will directly address them in the weak form of the problem. With respect to the weak formulation of the acoustic wave equation, we only have to add [53]

$$\mu c_0(q, \partial_t p)_{\Gamma_W} \quad (2.39)$$

to the left hand side of the acoustic wave equation in weak form (2.18), where

$$(q, \partial_t p)_{\Gamma_W} = \int_{\Gamma_W} q \partial_t p \, d\Gamma_W. \quad (2.40)$$

q is the test function, μ stands for the coefficient of the boundary admittance and Γ_W is the lossy boundary. The boundary coefficient μ can be computed as [53]

$$\mu = \frac{r}{\rho_0 c_0}, \quad (2.41)$$

where ρ_0 stands for air density and r corresponds to the real component of the boundary impedance (resistance term). The term (2.39) corresponds to the representation in the weak formulation of the boundary condition

$$\partial_t p = \mu c_0, \quad \text{on } \Gamma_W. \quad (2.42)$$

So, the variational or weak problem consists in finding $p \in \mathcal{P}$ such that

$$(q, \partial_{tt}^2 p) + c_0^2 (\nabla q, \nabla p) - c_0^2 \langle q, g_N \rangle_{\Gamma_N} + \mu c_0 (q, \partial_t p)_{\Gamma_W} = c_0^2 \langle q, f \rangle, \quad \forall q \in \mathcal{Q}, \quad (2.43)$$

where \mathcal{P} and \mathcal{Q} are the same space of functions that in the acoustic wave equation (see section 2.1).

2.2.2 Space and time discretization

Spatial discretization: Galerkin approximation

Being $\mathcal{P}_h \subset \mathcal{P}$ and $\mathcal{Q}_h \subset \mathcal{Q}$ finite subspaces, the spatial discretized scheme consists in finding $p_h \in \mathcal{P}_h$ such that

$$(q_h, \partial_{tt}^2 p_h) + c_0^2 (\nabla q_h, \nabla p_h) - c_0^2 \langle q_h, g_{N,h} \rangle_{\Gamma_N} + \mu c_0 (q_h, \partial_t p_h)_{\Gamma_W} = c_0^2 \langle q_h, f_h \rangle, \quad (2.44)$$

$\forall q_h \in \mathcal{Q}_h$, where $g_{N,h}$ and f_h are discretizations of g_N and f respectively. If we expand $p_h \in \mathcal{P}_h$ and $q_h \in \mathcal{Q}_h$ in terms of shape functions $N(x)$ and insert them into (2.44) we get

$$\begin{aligned} & \sum_a \sum_b Q^a (N^a, N^b) \ddot{P}^b + \mu c_0 (N^a, N^b)_{\Gamma_W} + c_0^2 \sum_a \sum_b Q^a (\nabla N^a, \nabla N^b) P^b = \\ & + \sum_a Q^a \langle N^a, g_{N,h} \rangle_{\Gamma_N} + c_0^2 \sum_a Q^a \langle N^a, f_h \rangle, \end{aligned} \quad (2.45)$$

where P^b and Q^a are the nodal values. Expressing (2.45) in matrix form and canceling the test function vector \mathbf{Q}^\top gives the final algebraic system

$$\mathbf{M}\ddot{\mathbf{P}} + \mu c_0 \mathbf{B}\dot{\mathbf{P}} + c_0^2 \mathbf{K}\mathbf{P} = c_0^2 \mathbf{L}, \quad (2.46)$$

where \mathbf{M} is the mass matrix, \mathbf{B} is the damping matrix, \mathbf{K} is the stiffness matrix and \mathbf{L} the load vector. The corresponding entries are

$$\mathbf{M} = [M^{ab}], \quad M^{ab} = (N^a, N^b), \quad (2.47)$$

$$\mathbf{B} = [B^{ab}], \quad B^{ab} = (N^a, N^b)_{\Gamma_W}, \quad (2.48)$$

$$\mathbf{K} = [K^{ab}], \quad K^{ab} = (\nabla N^a, \nabla N^b), \quad (2.49)$$

$$\mathbf{L} = [L^a], \quad L^a = \langle N^a, g_{N,h} \rangle_{\Gamma_N} + \langle N^a, f_h \rangle. \quad (2.50)$$

Time discretization: Finite differences

Using second order finite differences for the time discretization,

$$\mathbf{M} \frac{\mathbf{P}^{n+1} - 2\mathbf{P}^n + \mathbf{P}^{n-1}}{\Delta t^2} + \mu c_0 \mathbf{B} \frac{\mathbf{P}^{n+1} - \mathbf{P}^{n-1}}{2\Delta t} + c_0^2 \mathbf{K}\mathbf{P}^n = c_0^2 \mathbf{L}^n. \quad (2.51)$$

Grouping some terms we get

$$\left(\frac{\mathbf{M}}{\Delta t^2} + \frac{\mu c_0 \mathbf{B}}{2\Delta t} \right) \mathbf{P}^{n+1} = \left(\frac{2\mathbf{M}}{\Delta t^2} \right) \mathbf{P}^n - \left(\frac{\mathbf{M}}{\Delta t^2} - \frac{\mu c_0 \mathbf{B}}{2\Delta t} \right) \mathbf{P}^{n-1} + c_0^2 \mathbf{L}^n - c_0^2 \mathbf{K}\mathbf{P}^n. \quad (2.52)$$

Defining the matrices

$$\mathbf{C}_1 = \left(\frac{\mathbf{M}}{\Delta t^2} + \frac{\mu c_0 \mathbf{B}}{2\Delta t} \right), \quad (2.53)$$

$$\mathbf{C}_2 = \left(\frac{2\mathbf{M}}{\Delta t^2} \right), \quad (2.54)$$

$$\mathbf{C}_3 = \left(\frac{\mathbf{M}}{\Delta t^2} - \frac{\mu c_0 \mathbf{B}}{2\Delta t} \right), \quad (2.55)$$

we finally arrive at the following explicit scheme for the evolution of the nodal acoustic pressure:

$$\mathbf{P}^{n+1} = \mathbf{C}_1^{-1} (\mathbf{C}_2 \mathbf{P}^n - \mathbf{C}_3 \mathbf{P}^{n-1} + c_0^2 \mathbf{L}^n - c_0^2 \mathbf{K} \mathbf{P}^n). \quad (2.56)$$

2.2.3 Numerical example

Tube with wall losses

As a example we reconsider the same tube used in the second benchmark problem for the acoustic wave equation (section 2.1), but including losses in the walls. We also use $c = 345\text{m/s}$, $f_s = 200\text{KHz}$ and the same mesh. However, in this case we study its behavior using two different pulses: a transient pulse and a stationary signal.

Let us first consider the case of transient signals. To do so we use the smooth pulse (2.37) with $F_0 = 2000\text{Hz}$. We use it to test the model for different boundary admittance coefficients: $\mu = 0.005, 0.001, 0.0005$. Some time instants of the numerical solution for $\mu = 0.001$ can be seen in Figure 2.9. In contrast to the solution obtained for the lossless example (see Figure 2.8), it can be seen how the pulse is attenuated at each time step.

Point number	x (m)	y (m)
1	1.1	1.05
2	1.2	1.05
3	1.3	1.05
4	1.4	1.05

Table 2.1: Coordinates of four points in the tube (it has to be noted that the lower corner of the tube is at $(x, y) = (1, 1)$). These points are used to capture the time. In Figure 2.10 there is a representation of these points in the tube.

In order to better observe the attenuation effect, we have captured the time evolution of the solution at four points (see Figure 2.10). Their location is given in Table 2.1.

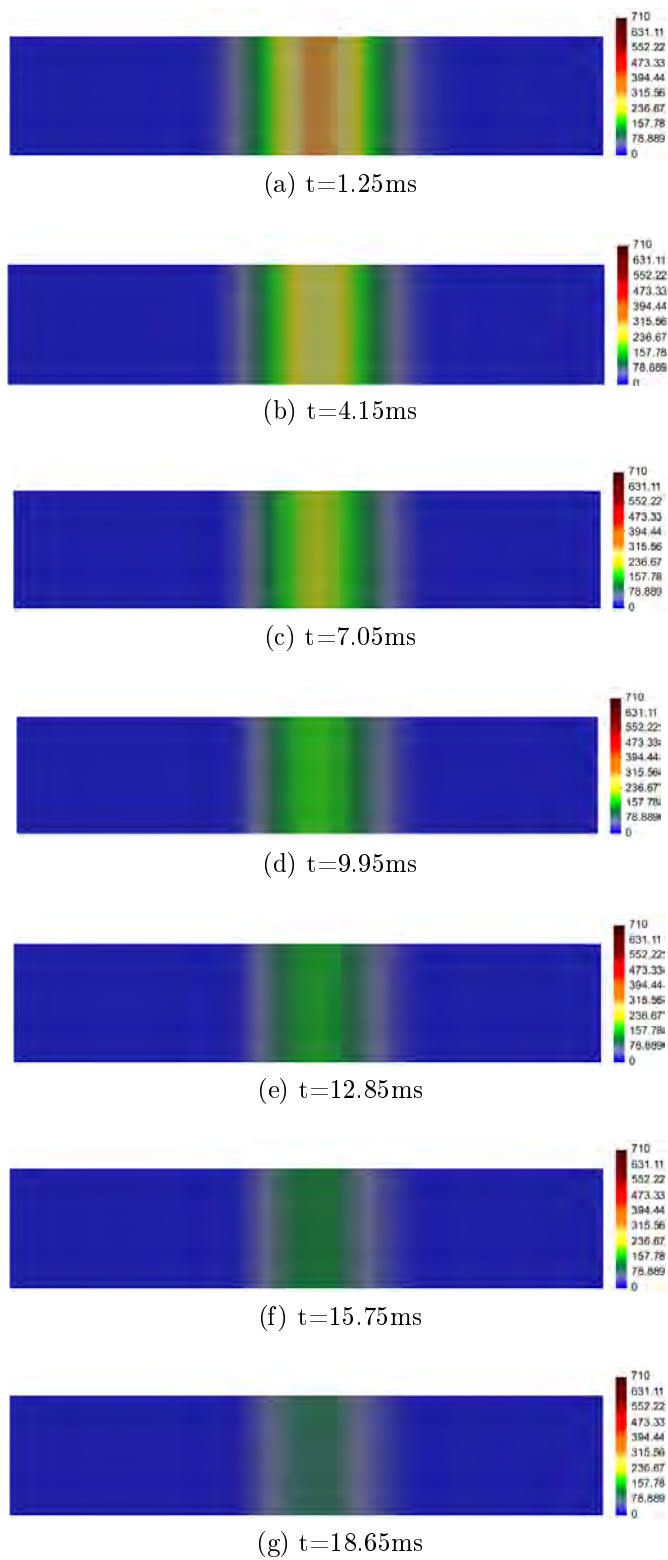


Figure 2.9: Snapshots of the tube acoustic pressure with wall losses and $\mu = 0.001$ for different time values.

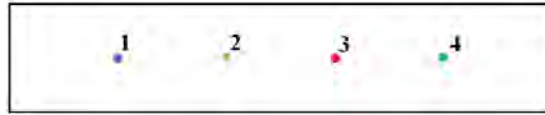


Figure 2.10: Location of the four points in the tube used to capture the time evolution of the different solutions. Their coordinates are provided in Table 2.1.

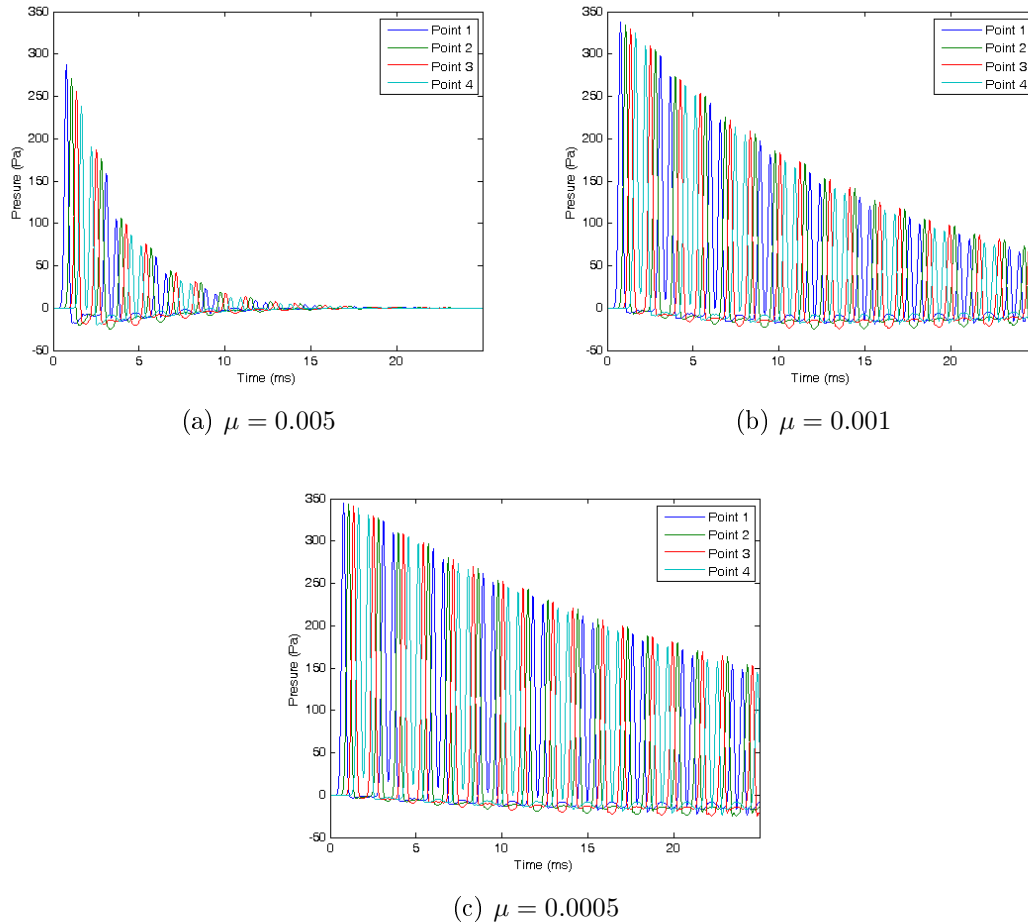


Figure 2.11: Time evolution of the solution of the tube in the transient case with $\mu = 0.005$ (a), 0.01 (b) and 0.005 (c) for 4 points (see Table 2.1 or Figure 2.10 for their location).

First of all, we will describe the wave propagation in the graphic of Figure 2.11. We can observe some wave packets formed by four pulses (one for each point). Each packet corresponds to the front wave passing through the points. The direction of this front wave can be seen according to when the peaks captured in each point appear. For example, when the front wave comes from the left side, we will see that the peak arrives first at point

1, second at point 2, then at point 3 and finally at point 4. However, when the wave front comes from the right side, we will observe the opposite effect ($4 \rightarrow 3 \rightarrow 2 \rightarrow 1$). Once analyzed the meaning of this graphic, it can be observed that the higher the boundary admittance coefficient μ the higher the speed at which the signal is attenuated (i.e. the higher the boundary admittance coefficient the higher the losses). For example, with $\mu = 0.005$, with only five reflections at the ends of the tube (six packets) there remains no signal in the tube, while for $\mu = 0.01$ and $\mu = 0.005$ the wave signal still has not lost the half of the initial amplitude.

Next we study the case of stationary signals generating a sinusoidal signal of fundamental frequency $F_0 = 2070\text{Hz}$, which corresponds to the 6th resonance mode of the tube without losses. The goal of this benchmark is to observe the tube behavior for stationary signals in a resonance mode. In a lossless model with constant energy input, the solution will grow to infinity because there is no energy dissipation. Furthermore, given that we are close to a resonance mode it will grow up faster. However, in a model with losses we could expect that the solution converges to a certain value. In Figure 2.12 the time evolution of point 1 (see Table 2.1 or Figure 2.10 for its location) is plotted for different μ . It can be observed how for all tested μ the different solutions converges to a value. On the other hand, for the tested cases, the higher the losses the lower the converge value. This is a normal behavior, because the higher the losses the lower the remaining signal, which is directly related to the observed “offset”. We refer to remaining signal as that signal that corresponds to a previous time step and still have an important amplitude. We have seen in Figure 2.11 that the higher the losses the higher the speed at which the signal is attenuated. If we introduce energy constant in time, the remaining signal will never be zero. This signal will be proportional to the speed at which the waves are attenuated, i.e. to the losses. So, the higher the losses the lower the convergence value.

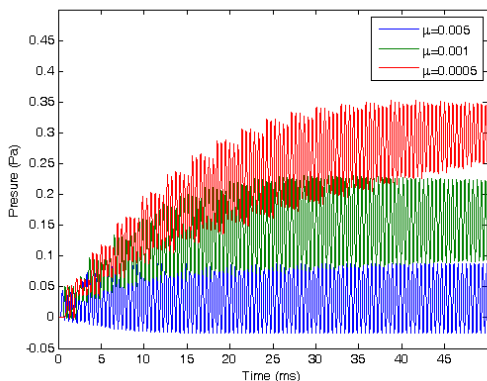


Figure 2.12: Time evolution of point 1 (see Table 2.1 or Figure 2.10 for its location) corresponding to the the solution of the tube with losses excited with a stationary signal and using different boundary coefficients (μ).

Another interesting effect can be seen in Figure 2.13. When attenuation is introduced in the walls, the pressure level close to the walls decays because viscous friction and heat conduction. This causes that plane wave front to curve. For example, with $\mu = 0.01$ (a) (we have added to see this phenomena easier), this effect is high, but as the losses decrease, this effect is reduced. In fact, the last observable case is $\mu = 0.005$ (b). With $\mu = 0.001$ (c) and $\mu = 0.0005$ (d) this phenomena can not be observed. On the other hand, it has to be noted some asymmetries on the solution of (a) and (b). This is done because we are using a non structured mesh that cause not uniform losses on the boundaries.

(a) $\mu = 0.01$ (b) $\mu = 0.005$ (c) $\mu = 0.001$ (d) $\mu = 0.0005$

Figure 2.13: Snapshots of the numerical solution of the tube with wall losses with different μ in the transient case at $t=32\text{ms}$.

2.3 The acoustic wave equation with a Perfectly Matched Layer (PML)

2.3.1 The Perfectly matched layer

Finally, we will also need to consider radiation of sound waves towards infinity. The main problem is that the computational domain must be finite, so it will be necessary to truncate it. To emulate outward waves to infinity, a radiation boundary condition has to be introduced to avoid reflections of the acoustic waves at the domain boundary. In the continuous, it takes the Sommerfeld boundary condition

$$\nabla p(\mathbf{x}, t) \cdot \mathbf{n} = 1/c_0 \partial_t p \quad \text{on } \Gamma_\infty \quad (2.57)$$

where Γ_∞ is the non reflection boundary condition. However, (2.57) is not optimal [10]. In this work it will be replaced by a Perfectly Matched Layer (PML). A PML is an artificial region where incident sound waves are absorbed. The key property is that incident waves coming from a non PML region do not reflect at the PML interface. Then, these waves are attenuated in the PML region (see Figure 2.14). However, in practice some residual signal remains. Therefore, an important aspect when designing a PML region is that it provides enough absorption for this residual to be negligible.

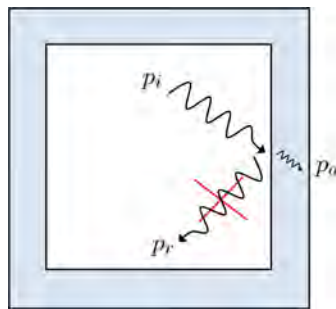


Figure 2.14: Truncation of a domain using a Perfectly Matched Layer (PML). The PML region corresponds to the shaded area. In this example, an incident wave p_i reaches the PML interface and becomes absorbed inside it p_a . No reflection p_r takes place at the PML interface.

2.3.2 Strong form

Let us introduce the formulation of the acoustic wave equation with a perfectly matched layer (PML) in the two-dimensional case [18]

$$\partial_{tt} p - c_0^2 \nabla^2 p = c_0^2 f + \nabla \cdot \boldsymbol{\phi} - (\xi_1 + \xi_2) \partial_t p - \xi_1 \xi_2 p, \quad (2.58)$$

$$\partial_t \boldsymbol{\phi} = \boldsymbol{\Gamma}_1 \boldsymbol{\phi} + c_0^2 \boldsymbol{\Gamma}_2 \nabla p, \quad (2.59)$$

where ξ_i are the damping profiles, $\boldsymbol{\phi}$ is an auxiliary vector of the form $\boldsymbol{\phi} = [\phi_x, \phi_y]^\top$, and $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are matrices of the form

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} -\xi_1 & 0 \\ 0 & -\xi_2 \end{pmatrix}, \quad \boldsymbol{\Gamma}_2 = \begin{pmatrix} \xi_2 - \xi_1 & 0 \\ 0 & \xi_1 - \xi_2 \end{pmatrix}. \quad (2.60)$$

Let us define the continuous functions $\alpha(\mathbf{x}) = \xi_1(x) + \xi_2(y)$, $\beta(\mathbf{x}) = \xi_1(x) \xi_2(y)$ and $\gamma(\mathbf{x}) = \xi_2(y) - \xi_1(x)$. In components, the above expressions (2.58) and (2.59) can be written as

$$\partial_{tt}p - c_0^2 \nabla^2 p = c_0^2 f + \partial_x \phi_x + \partial_y \phi_y - (\xi_1 + \xi_2) \partial_t p - \xi_1 \xi_2 p, \quad (2.61)$$

$$\partial_t \phi_x = -\xi_1 \phi_x + c_0^2 \gamma \partial_x p, \quad (2.62)$$

$$\partial_t \phi_y = -\xi_2 \phi_y - c_0^2 \gamma \partial_y p. \quad (2.63)$$

The damping profiles ξ_i can be constant, linear, or quadratic among others. Following Grote and Sim [18], we use

$$\xi_i(x_i) = \begin{cases} 0 & \text{for } |x_i| < a_i, \quad i = 1, 2, \\ \hat{\xi}_i \left(\frac{|x_i - a_i|}{L_i} - \frac{\sin\left(\frac{2\pi|x_i - a_i|}{L_i}\right)}{2\pi} \right) & \text{for } a_i \leq |x_i| \leq a_i + L_i, \quad i = 1, 2, \end{cases} \quad (2.64)$$

being $\hat{\xi}_i$ a constant to control the damping effect, a_i the i -th coordinate of the PML layer and L_i the thickness of the PML region in the i -th direction. Because $\xi_i(x)$ is twice continuously differentiable throughout the interface at $|x_i| = a_i$, no special transmission conditions are needed there [18].

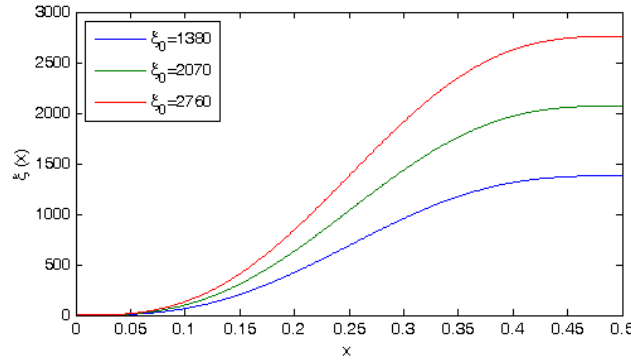


Figure 2.15: Damping profile $\xi(x)$ in the PML region for different values of $\hat{\xi}$ (in the graphic is ξ_0). The length of the PML region is $L = 0.5\text{m}$ and $c = 345\text{m/s}$.

The constant $\hat{\xi}_i$ depends on the discretization and thickness of the layer and can be computed as [18]

$$\hat{\xi}_i = \frac{c}{L_i} \log\left(\frac{1}{R}\right), \quad (2.65)$$

where R is the relative reflection of the boundary of the PML, which usually is truncated using a Dirichlet or Neumann condition. A good value for R could be $R = 10^{-4}$, which for a $c = 345\text{m/s}$ corresponds to $\hat{\xi} = 2760$ (see Figure 2.15).

On the other hand, we consider the following boundary and initial conditions

$$p(\mathbf{x}, t) = g_D(t) \quad \text{on } \Gamma_D, \quad t > 0, \quad (2.66)$$

$$\nabla p(\mathbf{x}, t) \cdot \mathbf{n} = g_N(t) \quad \text{on } \Gamma_N, \quad t > 0, \quad (2.67)$$

$$p = 0, \quad \text{in } \Omega, \quad t = 0, \quad (2.68)$$

$$\partial_t p = 0, \quad \text{in } \Omega, \quad t = 0. \quad (2.69)$$

2.3.3 Variational problem statement

Space of functions

The space of functions \mathcal{P} , \mathcal{Q} , \mathcal{U} and \mathcal{V} for the pressure, pressure test function, auxiliary function and auxiliary test function respectively are

$$\mathcal{P} = L^2(0, T; H^1(\Omega)), \quad (2.70)$$

$$\mathcal{Q} = H_0^1(\Omega), \quad (2.71)$$

$$\mathcal{U} = L^2(0, T; H^1(\Omega)), \quad (2.72)$$

$$\mathcal{V} = H_0^1(\Omega). \quad (2.73)$$

See section 2.1 for the definition of the above space of functions.

Variational form

Defining $\alpha(\mathbf{x}) = \xi_1(\mathbf{x}) + \xi_2(\mathbf{x})$, $\beta(\mathbf{x}) = \xi_1(\mathbf{x})\xi_2(\mathbf{x})$ and $\gamma(\mathbf{x}) = \xi_2(\mathbf{x}) - \xi_1(\mathbf{x})$, the variational form of the problem consist in find $p \in \mathcal{P}$ and $\phi \in \mathcal{U}$ such that

$$\begin{aligned} (q, \partial_{tt}^2 p) + c_0^2(\nabla q, \nabla p) &= c_0^2 \langle q, g_N \rangle_{\Gamma_N} + c_0^2 \langle q, f \rangle \\ &+ (q, \partial_x \phi_x) + (q, \partial_y \phi_y) - (q, \alpha \partial_t p) - (q, \beta p), \end{aligned} \quad (2.74)$$

$$(v_x, \partial_t \phi_x) = -(v_x, \xi_1 \phi_x) + c_0^2(v_x, \gamma \partial_x p), \quad (2.75)$$

$$(v_y, \partial_t \phi_y) = -(v_y, \xi_2 \phi_y) - c_0^2(v_y, \gamma \partial_y p), \quad (2.76)$$

$\forall q \in \mathcal{Q}$, $\forall v_x \in \mathcal{V}$ and $\forall v_y \in \mathcal{V}$.

2.3.4 Space and time discretization

Spatial discretization: Galerkin approximation

Being \mathcal{P}_h , \mathcal{Q}_h , \mathcal{U}_h and \mathcal{V}_h finite subspaces of \mathcal{P} , \mathcal{Q} , \mathcal{U} and \mathcal{V} respectively ($\mathcal{P}_h \subset \mathcal{P}$, $\mathcal{Q}_h \subset \mathcal{Q}$, $\mathcal{U}_h \subset \mathcal{U}$ and $\mathcal{V}_h \subset \mathcal{V}$), the spatial discretized scheme of the problem consist in finding $p_h \in \mathcal{P}_h$ and $\phi_h \in \mathcal{U}_h$ such that

$$(q_h, \partial_{tt}^2 p_h) + c_0^2 (\nabla q_h, \nabla p_h) = +c_0^2 \langle q_h, g_{N,h} \rangle_{\Gamma_N} + c_0^2 \langle q_h, f_h \rangle + (q_h, \partial_x \phi_{x,h}) + (q_h, \partial_y \phi_{y,h}) - (q_h, \alpha_h \partial_t p_h) - (q_h, \beta_h p_h), \quad (2.77)$$

$$(v_{x,h}, \partial_t \phi_{x,h}) = -\xi_{1,h}(v_{x,h}, \phi_{x,h}) + c_0^2 (v_{x,h}, \gamma_h \partial_x p_h), \quad (2.78)$$

$$(v_{y,h}, \partial_t \phi_{y,h}) = -\xi_{2,h}(v_{y,h}, \phi_{y,h}) - c_0^2 (v_{y,h}, \gamma_h \partial_y p_h), \quad (2.79)$$

$\forall q_h \in \mathcal{Q}_h$, $\forall v_{x,h} \in \mathcal{V}_h$ and $\forall v_{y,h} \in \mathcal{V}_h$. Using the shape functions $N(x)$; p_h , q_h , ϕ_h and \mathbf{v}_h can be written as

$$p_h(\mathbf{x}, t) = \sum_b N^b(\mathbf{x}) P^b(\mathbf{x}, t), \quad (2.80)$$

$$q_h(\mathbf{x}) = \sum_a N^a(\mathbf{x}) Q^a(\mathbf{x}), \quad (2.81)$$

$$\phi_h(\mathbf{x}, t) = \left[\sum_b N^b(\mathbf{x}) \Phi_x^b(x, t), \sum_b N^b(\mathbf{x}) \Phi_y^b(y, t) \right], \quad (2.82)$$

$$\mathbf{v}_h(\mathbf{x}) = \left[\sum_a N^a(\mathbf{x}) V_y^a(y), \sum_a N^a(\mathbf{x}) V_x^a(x) \right]. \quad (2.83)$$

Using the above expression and inserting into the scheme, the final algebraic system is

$$\mathbf{M} \ddot{\mathbf{P}} + c_0^2 \mathbf{K} \mathbf{P} = c_0^2 \mathbf{L} + \mathbf{B}_x \Phi_x + \mathbf{B}_y \Phi_y - \mathbf{M}_\alpha \dot{\mathbf{P}} - \mathbf{M}_\beta \mathbf{P}, \quad (2.84)$$

and the algebraic system for the auxiliary functions

$$\mathbf{M} \dot{\Phi}_x = -\mathbf{M}_{\xi_1} \Phi_x + c_0^2 \mathbf{B}_{x,\gamma} \mathbf{P}, \quad (2.85)$$

$$\mathbf{M} \dot{\Phi}_y = -\mathbf{M}_{\xi_2} \Phi_y - c_0^2 \mathbf{B}_{y,\gamma} \mathbf{P}, \quad (2.86)$$

where \mathbf{M} is the mass matrix, \mathbf{M}_α and \mathbf{M}_β are the mass matrices with the spacial functions $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$, \mathbf{K} is the stiffness matrix, \mathbf{L} the load vector, \mathbf{B}_x and \mathbf{B}_y are the damping matrices in the x and y directions, and $\mathbf{B}_{x,\gamma}$ and $\mathbf{B}_{y,\gamma}$ are the damping matrices with the spacial function $\alpha(\mathbf{x})$. The entries of these matrices and vectors are

$$\mathbf{M} = [M^{ab}], \quad M^{ab} = (N^a, N^b), \quad (2.87)$$

$$\mathbf{M}_\alpha = [M_\alpha^{ab}], \quad M_\alpha^{ab} = \langle N^a, \alpha_h N^b \rangle, \quad (2.88)$$

$$\mathbf{M}_\beta = [M_\beta^{ab}], \quad M_\beta^{ab} = \langle N^a, \beta_h N^b \rangle, \quad (2.89)$$

$$\mathbf{K} = [K^{ab}], \quad K^{ab} = (\nabla N^a, \nabla N^b), \quad (2.90)$$

$$\mathbf{L} = [L^a], \quad L^a = \langle N^a, g_{N,h} \rangle_{\Gamma_N} + c_0^2 \langle N^a, f_h \rangle, \quad (2.91)$$

$$\mathbf{B}_x = [B_x^{ab}], \quad B_x^{ab} = (N^a, \partial_x N^b), \quad (2.92)$$

$$\mathbf{B}_y = [B_y^{ab}], \quad B_y^{ab} = (N^a, \partial_y N^b), \quad (2.93)$$

$$\mathbf{B}_{x,\gamma} = [B_{x,\gamma}^{ab}], \quad B_{x,\gamma}^{ab} = (N^a, \gamma_h \partial_x N^b), \quad (2.94)$$

$$\mathbf{B}_{y,\gamma} = [B_{y,\gamma}^{ab}], \quad B_{y,\gamma}^{ab} = (N^a, \gamma_h \partial_y N^b). \quad (2.95)$$

Time discretization: Finite differences

Using second order finite difference schemes for the time discretization

$$\begin{aligned} & \mathbf{M} \frac{\mathbf{P}^{n+1} - 2\mathbf{P}^n + \mathbf{P}^{n-1}}{\Delta t^2} + c_0^2 \mathbf{K} \mathbf{P}^n = \\ & + c_0^2 \mathbf{L}^n + \mathbf{B}_x \Phi_x^n + \mathbf{B}_y \Phi_y^n - \mathbf{M}_\alpha \frac{\mathbf{P}^{n+1} - \mathbf{P}^{n-1}}{2\Delta t} - \mathbf{M}_\beta \mathbf{P}^n. \end{aligned} \quad (2.96)$$

Grouping some terms

$$\begin{aligned} \left(\frac{\mathbf{M}}{\Delta t^2} + \frac{\mathbf{M}_\alpha}{2\Delta t} \right) \mathbf{P}^{n+1} &= \left(\frac{2\mathbf{M}}{\Delta t^2} + \mathbf{M}_\beta \right) \mathbf{P}^n \\ &- \left(\frac{\mathbf{M}}{\Delta t^2} - \frac{\mathbf{M}_\alpha}{2\Delta t} \right) \mathbf{P}^{n-1} + c_0^2 \mathbf{L}^n - c_0^2 \mathbf{K} \mathbf{P}^n + \mathbf{B}_x \Phi_x^n + \mathbf{B}_y \Phi_y^n. \end{aligned} \quad (2.97)$$

We define the auxiliary matrix \mathbf{C}_1 , \mathbf{C}_2 and \mathbf{C}_3 such that

$$\mathbf{C}_1 = \left(\frac{\mathbf{M}}{\Delta t^2} + \frac{\mathbf{M}_\alpha}{2\Delta t} \right), \quad (2.98)$$

$$\mathbf{C}_2 = \left(\frac{2\mathbf{M}}{\Delta t^2} - \mathbf{M}_\beta \right), \quad (2.99)$$

$$\mathbf{C}_3 = \left(\frac{\mathbf{M}}{\Delta t^2} - \frac{\mathbf{M}_\alpha}{2\Delta t} \right). \quad (2.100)$$

With the above definitions, we get the following explicit scheme

$$\mathbf{P}^{n+1} = \mathbf{C}_1^{-1} (\mathbf{C}_2 \mathbf{P}^n - \mathbf{C}_3 \mathbf{P}^{n-1} + c_0^2 \mathbf{L}^n - c_0^2 \mathbf{K} \mathbf{P}^n + \mathbf{B}_x \Phi_x^n + \mathbf{B}_y \Phi_y^n). \quad (2.101)$$

Note that if we are not in the PML region, $\xi_{1,h}$, $\xi_{2,h}$, Φ_x and Φ_y are zero and the above scheme reduces to the explicit scheme for the acoustic wave equation. On the other hand,

the explicit schemes for the auxiliary functions Φ_x and Φ_y are

$$\Phi_x^{n+1} = \Phi_x^{n-1} - 2\Delta t M^{-1} (M_{\xi_1} \Phi_x^n - c_0^2 B_{x,\gamma} P^n) \quad (2.102)$$

$$\Phi_y^{n+1} = \Phi_y^{n-1} - 2\Delta t M^{-1} (M_{\xi_2} \Phi_y^n + c_0^2 B_{y,\gamma} P^n) \quad (2.103)$$

2.3.5 Numerical example

Infinite membrane

In this example we use the same squared membrane that in the example of section 2.1, but surrounded by a perfectly matched layer of thickness $L=0.5\text{m}$ (see Figure 2.16) to emulate free space propagation. So we get an infinite membrane.

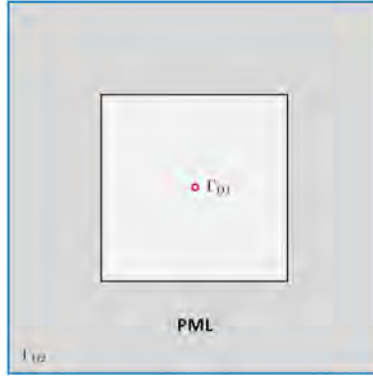


Figure 2.16: Membrane of the example of section 2.1 surrounded by a perfectly matched layer (PML) of thick $L=0.5$. The boundaries Γ_{D1} and Γ_{D2} are marked in red and blue respectively.

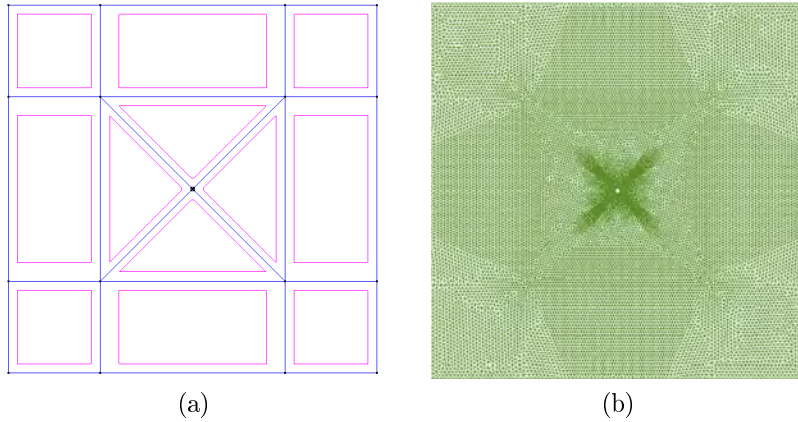


Figure 2.17: Obtained Geometry (a) and mesh (b) for the infinite membrane example.

The smooth pulse (2.37) with $F_0 = 1000\text{Hz}$ is also used as an excitation signal in the contour Γ_{D1} . We take the same wave propagation ($c = 345\text{m/s}$) and sampling frequency

($f_s = 200\text{KHz}$). The adopted meshing criteria is also $h = 0.015\text{m}$ over surfaces and $h = 0.003\text{m}$ over the inner boundary Γ_{D1} . We have obtained a mesh of 17560 nodes and 34700 elements (see Figure 2.17 for the obtained geometry and mesh). However, in this case the boundary Γ_{D2} corresponds to the outer boundary of the PML region, where we have imposed an homogeneous Dirichlet condition ($p = 0$). On the other hand, we take the damping factor $\hat{\xi}_i = 2760$ being the same scale factor for x and y . This factor implies a relative reflection coefficient R of 10^{-4} (see (2.65)).

The obtained numerical results can be seen in Figure 2.18. In (a) and (b) the wave front is generated by the contour Γ_{D1} and advances towards the PML interface. In (c) the wave front has just entered the PML region, where it will be absorbed. This effect can be seen in the frames (d)-(h), until it reaches (i), where there is no appreciable signal.

To be able to properly evaluate the quality of the perfectly matched layer, we have captured the time evolution for four membrane points. Their coordinate values and locations on the membrane are given in Table 2.2 and Figure 2.19a respectively. It has to be noted that the lower left corner of the PML is in $(x, y) = (0.5, 0.5)\text{m}$.

Point number	x (m)	y (m)
1	1.53122	1.49994
2	2	1.49254
3	2.25053	1.5
4	2.45862	1.5

Table 2.2: Coordinates of four analysis points in the infinite membrane. These points are used to capture the time evolution signal in some locations of interest: in the domain of interest, in the PML interface and in the PML region (see Figure 2.19a.)

It can be seen in Figure 2.19b that point 1 has the higher value given that it is the nearest point to the source. Just in the PML interface there is point 2, whose signal will be taken as a reference to evaluate the PML efficiency. It has to be noted that this signal is lower in amplitude with respect to point 1 because geometrical divergence effect. Time delay of signal 2 respect to signal 1 is because sound propagation. Once the wave front has gone inside the perfectly matched layer, it is captured by point 3, which is located in the middle of the PML. Finally, close to the end of the PML, there is point 4. It can be seen how the front wave arrives at point 4 with a small signal, close to zero. With the aim to evaluate the PML efficiency, we have computed the difference between the maximum value of signal 2 with respect the maximum value of signal 4. This difference is about 15dB, which in acoustics means that this signal can be omitted. Note that this front wave still has to reflect in the boundary Γ_{N2} and return to the domain of interest, in which path it will be even more attenuated. Finally, after the front wave arrives at point 4,

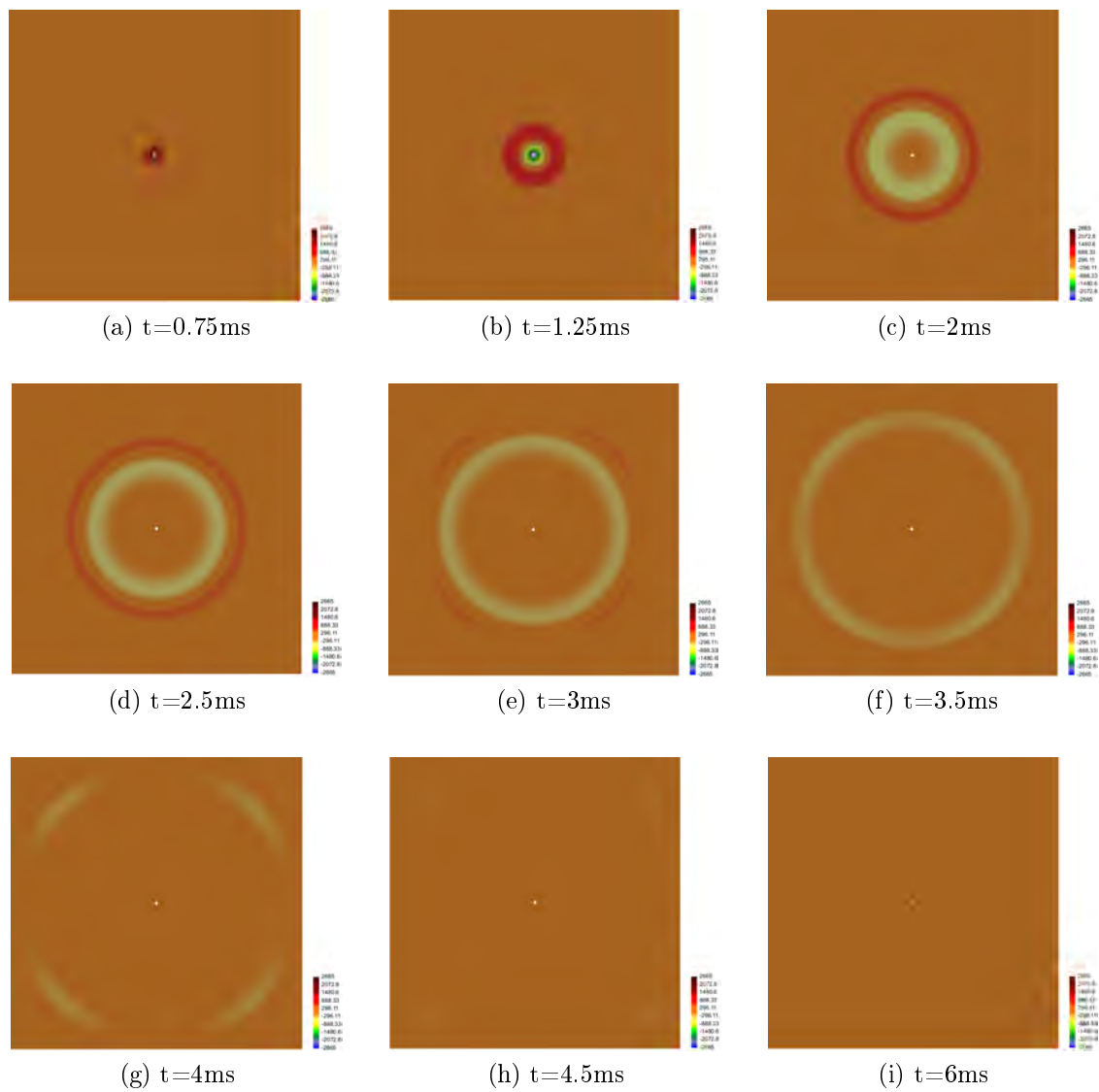


Figure 2.18: Snapshots of the numerical solution of the infinite membrane example for different time instants.

no more peaks can be observed. So, it can be said that no significant reflections appear in the measurement. On the other hand, this effect can also be observed in the residual

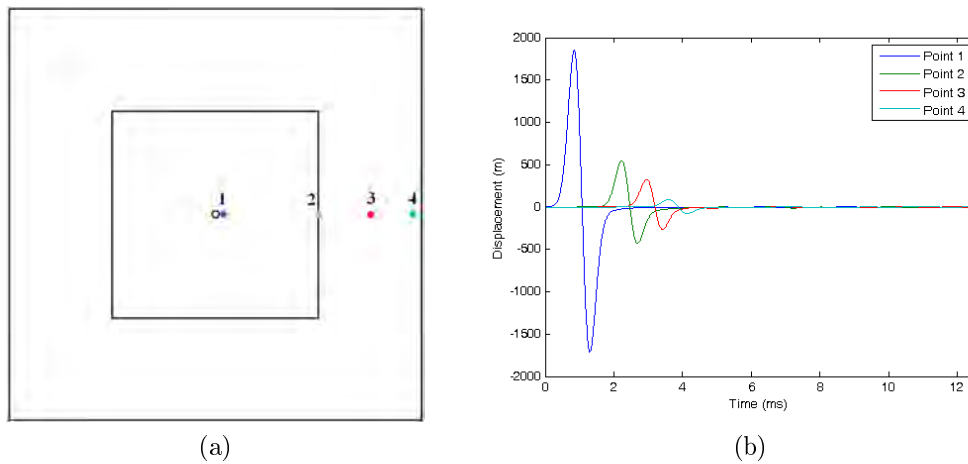


Figure 2.19: Time evolution of four points (b) located according to Table 2.2 for the infinite membrane example. The representation of the localization of these points in the domain can be seen in (a).

signal of the numerical simulation at time instants higher than 5ms (when the whole wave front has left the domain of interest). For example, in Figure 2.20, we have plotted the residual signal at time $t=6\text{ms}$. Its amplitude is about 3 order of magnitude less than the signal of interest. This also means that these reflections can be neglected.

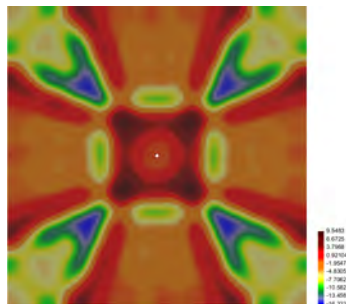


Figure 2.20: Residual signal for the infinite membrane example at time $t=6\text{ms}$.

Chapter 3

Applying FEM to the synthesis of vowels

In this chapter the finite element method will be applied to the synthesis of vowels problem. First, we will put in context the synthesis problem by means the particularization of the generic block diagram used in chapter 1 to introduce the articulatory speech synthesis. Second, some important remarks concerning the acoustic modelling will be provided (geometry, glottal source and losses). Next, the main features that should have the acoustic wave equation within the vocal tract will be discussed. Then, the obtained numerical scheme will be provided. Finally, as an example, we will synthesize the catalan vowel /e/. Its transfer function will be calculated, from which an objective study of the intelligibility will be done. Some remarks on vowel quality will also be done.

3.1 Introduction

In chapter 1 we have introduced the concept of articulatory speech synthesis by means of a generic block diagram (see Figure 1.1). In this chapter, we will use the same generic block diagram, particularizing it to the synthesis of vowels (see Figure 3.1).

Special attention has to be paid to the vocal tract acoustic block, which is at the core of this work. This block has become a computational model based on finite element methods (FEM). This block, using as inputs the airflow provided by the glottal model and the vocal tract geometry, will simulate the time evolving pressure distribution within the vocal tract by solving the underlying physics equations using FEM. It is subdivided into three blocks: the pre-processing block where the computational domain becomes meshed and the boundary conditions imposed (e.g., inflow from the glottal model, Sommerfield free space radiation at the outlet of the computational domain, rigid or elastic conditions for the vocal tract walls, etc.); then, in the second block the FEM code is used to solve the acoustic and flow dynamics equations using FEM. In the post-processing block, the time dependent acoustic pressure at the outlet of the lips is converted to an audio file.

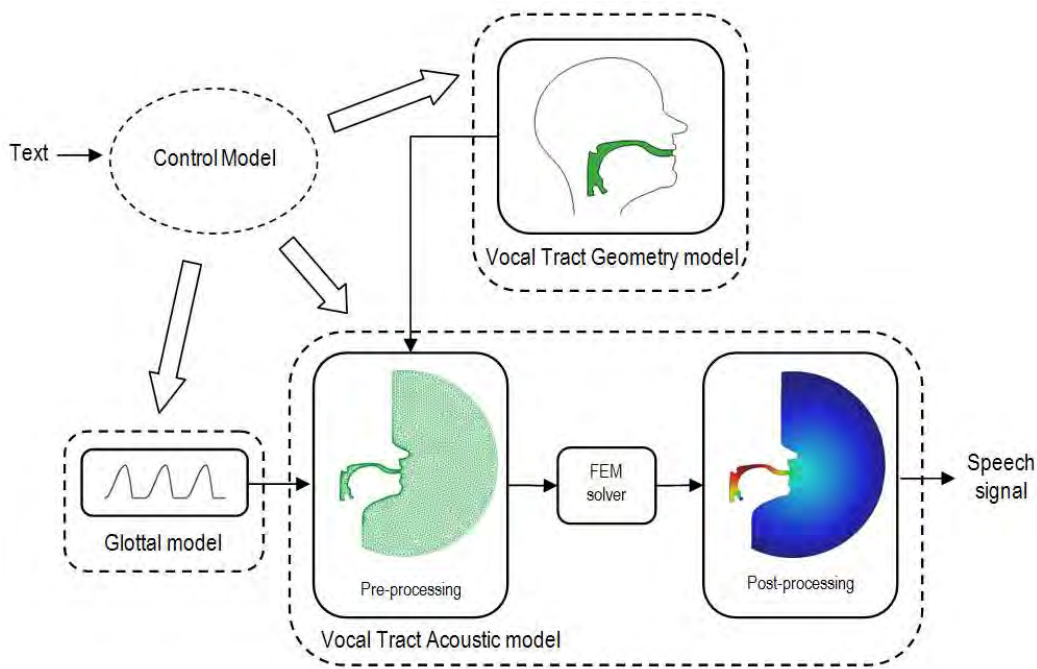


Figure 3.1: Block diagram for the proposed articulatory speech synthesis system

Prior to describing the core of the model, in the following section some important remarks on the acoustic modelling will be done.

3.2 Some considerations on the acoustic modeling

To construct the acoustic model, it is necessary to study all aspects involved in speech production, as well as the physical phenomena that occur. In this section some important considerations to be taken into account will be described as well as the assumptions we will make. The most relevant key points are the geometry, the glottal source and the losses, described below.

3.2.1 Geometry

One of the most important issues in the acoustic modeling process is the geometry, which is, in the current case, the human vocal tract. The geometry used can be one, two or three dimensional. Depending on the geometry accuracy, a higher speech quality will be obtained. In fact, the higher the dimension, the better the spectrum accuracy, which means better speech quality. However, in the case of synthesis of vowels, the geometry dimensionality hardly affects the low frequency range (below 3KHz) of the speech spectrum [53]. This assures a good intelligibility but not a good quality. So, it is

therefore desirable to use geometries of higher dimensions, i.e. two or three dimensions.

On the other hand, the numerical method to use depends, among other factors, on the complexity of the geometry. For example, in the one-dimensional case, a finite difference method is widely used (e.g. in [12], [1]). Moreover, in the two or three dimensional cases, due to the complexity of the geometry, the above method can not be used. In these cases, the most widely used numerical method seems to be the finite element method (e.g. [53]).

Another important aspect is the geometrical model to use (see chapter 1 for a review). Given that we aim at synthesize high quality vowels, we will require the geometry to be accurate and realistic as possible. The geometry model that satisfy these requirements is the statistical model, which needs a huge database of articulatory data. In our case, we will use a statistical model corresponding of a collection of static magnetic resonance images. These images correspond to the midsagittal plane of the vocal tract and they were captured from a male native catalan speaker producing five catalan vowels (/a/, /e/, /i/, /o/ and /u/). In Figure 3.2 there are some examples of the MRI collection. Then, the geometry has been obtained by hand tracing the inner boundary of the vocal tract.

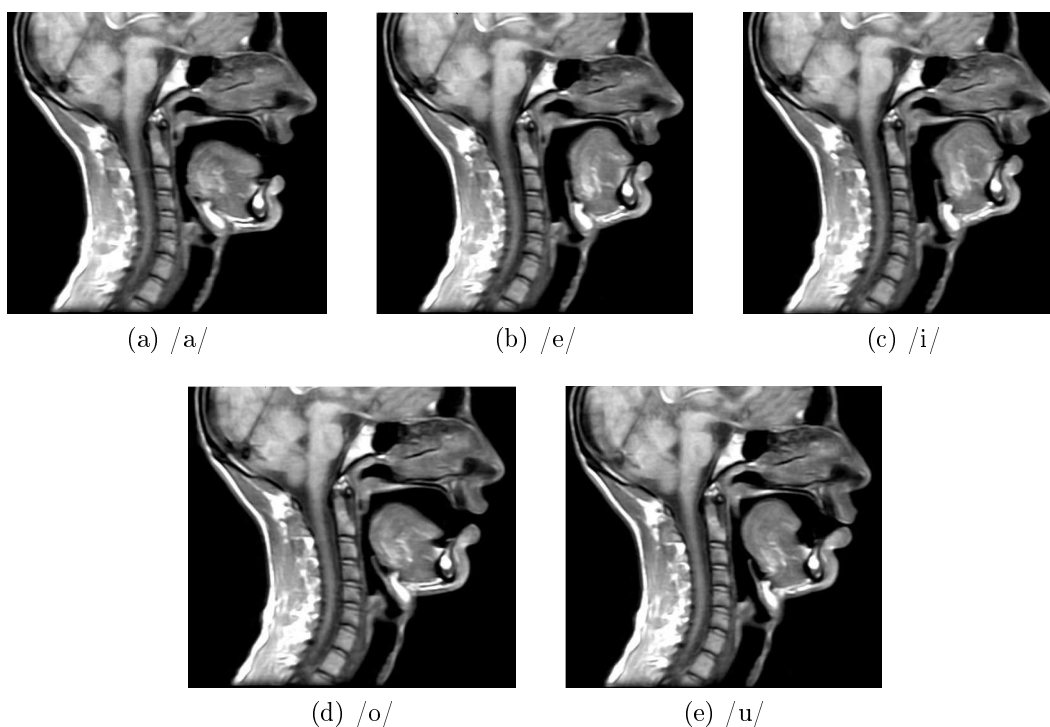


Figure 3.2: MRI images used to synthesize the vowels /a/, /e/, /i/, /o/ and /u/.

So, we will use a statistical model that will provide two-dimensional geometries to the vocal tract acoustic model. Given the complexity of these geometries, will use as numerical method the finite element method.

3.2.2 Glottal source

Although the glottal model is usually an independent block of an articulatory synthesizer, it needs to be coupled into the vocal tract acoustic model. Generally, glottal models are coupled to the vocal tract acoustic models by imposing the glottal output airflow (glottal pulses) as the input airflow of the vocal tract, e.g. the Rosenberg model [48]. However, some more complex models require some additional data in order to estimate the glottal pulses. For example, the two-mass models require the cross-sectional area of the input of the vocal tract [50].

In our case, we first considered using the C Rosenberg model [48] as a glottal model. This model stands out for its simplicity and its proper behavior for the synthesis of vowels. This model parametrizes the waveform of the volumetric glottal velocity $u_g(t)$ as

$$u_g(t) = \begin{cases} \frac{a}{2} \left[1 - \cos\left(\pi \frac{t}{T_p}\right) \right] & 0 \leq t \leq t_p \\ a \cos\left(\frac{\pi}{2} \frac{t-T_p}{T_n}\right) & t_p < t \leq t_p + t_n \\ 0 & t_p + t_n < t < T_0 \end{cases}, \quad (3.1)$$

where a is the amplitude, t_p is the positive slope, t_n the negative slope and T_0 the period of the glottal pulses (i.e. the inverse of the fundamental frequency F_0 or pitch) (see Figure 3.3). The times t_p and t_n are related to the pitch as

$$t_p = 40\% T_0, \quad (3.2)$$

$$t_n = 16\% T_0. \quad (3.3)$$

In figure 3.3 we show one pulse generated by the C Rosenberg model, using $a = 4 \cdot 10^{-4} \text{m}^3/\text{s}^2$ and a pitch of $F_0 = 100 \text{ Hz}$.

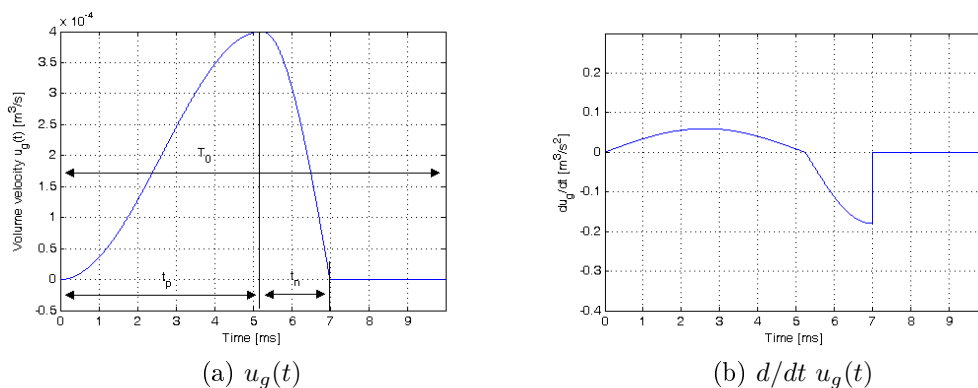


Figure 3.3: Rosenberg model of the C type, where u_g (a) is the volumetric velocity of the glottal pulse and $d/dt u_g$ its time derivative.

However, this pulse contains high frequency information because it has abrupt transitions (i.e. it is not smooth). In numerical schemes, high frequency information can pollute the

solution because they can not be captured by the mesh (the size of the elements can not be fine enough to capture it). This may result in errors and instabilities of the solution.

Given the above drawbacks, we thought about using the LF model [14]. This models also describe the flow variations of the glottal source, but it does so acting on its time derivative. This is an important feature because imposing the time derivative of the glottal flow is the most natural way to couple the glottal model with the vocal tract acoustic model (we will see it in (3.10)). However, the most important advantage is that this model describes a smooth pulse, closer to reality. So, no excessive high frequency are expected to be introduced into the numerical scheme.

In a LF model, the time derivative of the volumetric velocity $u_g(t)$ is

$$\frac{d}{dt}u_g(t) = \begin{cases} E_0 e^{\alpha t} \sin(w_g t) & 0 \leq t \leq t_e \\ -\frac{E_e}{\epsilon t_a} (e^{-\epsilon(t-t_e)} - e^{-\epsilon(T_0-t_e)}) & t_e < T_0 \end{cases}, \quad (3.4)$$

where T_0 is the inverse of the fundamental frequency or pitch F_0 . Prior to define the wave parameters E_0 , α , w_g and t_e , it is necessary to introduce some dimensionless quantities

$$R_a = \frac{t_a}{T_0 - t_e}, \quad R_k = \frac{t_e - t_p}{t_p}, \quad R_g = \frac{T_0}{2t_p}, \quad (3.5)$$

where the definitions for t_e , t_p , t_a and T_0 can be seen in Figure 3.4b. The angular frequency w_g is related to the fundamental frequency F_0 by $w_g = 2\pi F_0 R_g$. On the other hand, E_e is the minimum value of the time derivative of the glottal pulse (see Figure 3.4). For a male voice, $E_e = 0.4\text{m}^3\text{s}^{-2}$, $R_g = 1.12$ and $R_k = 0.34$ [54]. The time t_a , for small values of ϵ can be computed as

$$t_a = \frac{U_e}{E_e}, \quad (3.6)$$

where U_e is the mean flow volume rate in the glottis, which in a male is $U_e = 0.12\text{l/s}$ [54]. So, using (3.6) we can compute the time $t_a = 0.3\text{ms}$. On the other hand, imposing that the time derivative of the glottal pulse (3.4) at time t_p must be zero and imposing its continuity at time t_e , we get the following non linear equations

$$E_0 e^{\alpha t_p} \sin(w_g t_p) = 0, \quad (3.7)$$

$$E_0 e^{\alpha t_e} \sin(w_g t_e) = E_e, \quad (3.8)$$

$$-\frac{E_e}{\epsilon t_a} (1 - e^{-\epsilon(T_0-t_e)}) = E_e. \quad (3.9)$$

Solving the non-linear system of equations (3.7–3.9), the parameters E_e , α and ϵ can be determined.

We will use the LF model as the glottal model. However, it would have been interesting to use a mechanical model, because they work with physical parameters closer to the process

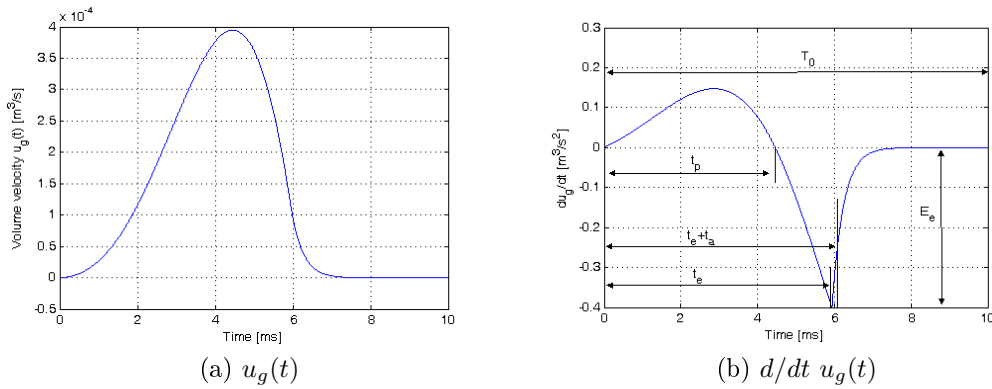


Figure 3.4: LF model, where (a) is the volume velocity of the glottal source and (b) is its time derivative. In (b) the four wave shape parameters t_p , t_e , t_a and E_e that uniquely determine the pulse can also be seen.

of phonation, such as the tension of the vocal cords, the air injected by the lungs, etc. With such parameters, it is easier to synthesize natural voice, because they are better defined than the shape of the glottal pulses.

On the other hand, given that the unknown of the equation is the acoustic pressure, it is necessary to express the velocity fluctuations of the glottal pulses as a pressure condition. This step can be done using the Newton equation

$$\nabla p = -\rho \frac{\partial u}{\partial t}, \quad (3.10)$$

where ρ is the air density. Thanks to the LF model, we do not need to discretize the temporal term of (3.10) because we directly know its time derivative. However, we have to keep in mind that the LF model provides a volumetric velocity and the Newton equation requires a punctual velocity. The conversion can be easily done if we know the area where the source have to be imposed.

Finally, this glottal source will be introduced into the acoustic model using the following Neumann boundary condition

$$\nabla p = g(t) = -\rho \frac{\partial u_g}{\partial t} \quad \text{on } \Gamma_G, \quad (3.11)$$

where Γ_G is the Neumann boundary. Expressing (3.11) in its time discrete form we obtain

$$\nabla p^n \cdot n = g^n = -\rho \frac{\partial u_g^n}{\partial t} \quad \text{on } \Gamma_G, \quad (3.12)$$

where the superscript n denotes the time step n . So, the glottal model will have to provide to the vocal tract acoustic model the glottal source g^n at each time step.

3.2.3 Losses

The modeling of losses is an important issue for achieving good quality in the synthesized speech. These losses mainly contribute to the bandwidth of the formants, specially for middle and high frequencies, and in less extent, to their location. The most significant are the radiation loss, the wall losses and the losses introduced by the vibration of the walls.

Radiation loss

The radiation loss refers to the energy dissipated when the sound field generated inside the vocal tract leaves the mouth, spreading towards free space. This is the most important loss effect in the speech production process [3]. Many artifacts have been devised to model this phenomena. The most commonly used is the approximation of the mouth as a circular piston in an infinite baffle, from which the value of a load impedance can be obtained ([25]). Then, the domain is truncated at the mouth, using the load impedance to emulate the acoustic radiation into free space. The main drawback is that this impedance is frequency dependent, which cause several difficulties for time domain methods. However, some approximations to the load impedance can be made (e.g. in [3]), but decreasing the quality of the radiation losses.

In our case, we will not use any approximation. We will calculate the acoustic field without truncating the domain at the mouth. That is it will consider a domain with the vocal tract and the free space. However, the domain must be finite, it will be necessary truncate the free space domain. To simulate this behavior we will use the perfectly matched layer (PML) approach, previously introduced in chapter 2.

Wall losses: viscous friction and heat conduction

The wall losses are losses due to the propagation of the acoustic waves close to the vocal tract walls. The most important are due to viscous friction and the heat conduction. Compared to other losses, heat conduction can generally be neglected [3]. Moreover, compared to radiation loss, viscous friction losses contribute the second to the bandwidth of formants. So, it is desirable to model this loss in the vocal tract. In general, when wall losses are considered, we refer to soft walls, whereas if they are not considered, we refer to hard walls [53]. In Figure 3.5 we show the vocal tract transfer function corresponding to a Japanese /a/. Hard and soft walls are considered, but not radiation losses. The nasal cavity is also included in the study. Vocal tract functions have been computed using 3D finite element methods in frequency domain [36]. It can be seen how sharp peaks vanish and formant bandwidth increase because of wall losses.

In one-dimensional cases, some modification to the conservation of momentum equation can be done to include viscous friction effects (e.g. [3,12]). However, for higher dimensions (2D or 3D), these losses are introduced as boundary conditions (e.g. [53]).

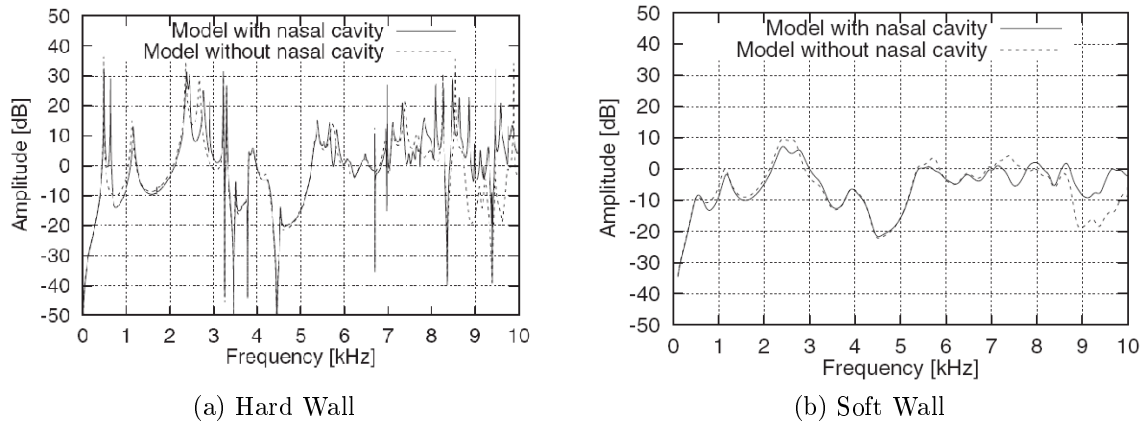


Figure 3.5: Vocal tract transfer functions of a male producing Japanese /a/, considering nasal coupling and hard (a) and soft (b) walls [36].

In our acoustic model we will take them into account, i.e. we consider soft walls. Because we are using two-dimensional geometries, we will consider the following Neumann condition

$$\frac{\partial p}{\partial t} = \mu c_0 \quad \text{on } \Gamma_W, \quad (3.13)$$

where Γ_W is the vocal tract boundary.

Wall vibration losses

A usual assumption in the acoustic modelling process is to consider that the vocal tract walls are rigid (e.g. in [58]). However, a non rigid model of the vocal tract (yielding walls) can also be used. This introduces some losses to the model. One technique is simulate this vibration by means of mechanical models (e.g. in [12]). This simulate the wall vibration using mechanical analogies, which depend on different physical properties such as the elasticity, the pressure within the vocal tract, the cross-sectional area, etc.

In our case, we will consider rigid walls for simplicity.

3.3 Computational model for the vocal tract

3.3.1 Model description

In the previous section, we have seen some of the most important considerations that have to be taken into account in the acoustic modelling process. In the computational model framework, we have finally decided the following aspects:

- a) Geometry model: A statistical geometry model providing 2D static geometries.
- b) Glottal source: a FL model generating glottal pulses.

- c) Radiation Loss: a Perfectly matched layer (PML) for taking into account free space propagation.
- d) Wall loss: wall losses due to viscous friction and heat conduction.
- e) Wall vibration losses: a rigid model of the vocal tract (so no wall vibration losses are introduced).

First, we have seen that given that we want to synthesize high quality vowels, we need a geometry with all possible details. So, the appropriate geometry model to use is of the statistical type. It has to provide 2D geometries to the vocal tract acoustic model. Thanks to their simplicity with complex geometries, we have decided to use the finite element method (FEM). Second, as a glottal source, we use a FL model. We have seen that this model directly provides the time derivative of the volume velocity of the glottal pulses, which simplifies the coupling between the glottal and vocal tract acoustic models. Moreover the generated pulse is smooth, which ensures that no exceed high frequency energy is introduced in the numerical scheme. Finally, as we make synthesis in the time domain (in time domain we can get better quality than in frequency domain [4]), we have to use the acoustic wave equation. We have addressed this issue in section 2.1. However, if we want to deal with losses, we will also have to consider the wall losses and radiation losses. The former have been addressed in section 2.2 and the latter in section 2.3 by means of a Perfectly Matched Layer (PML).

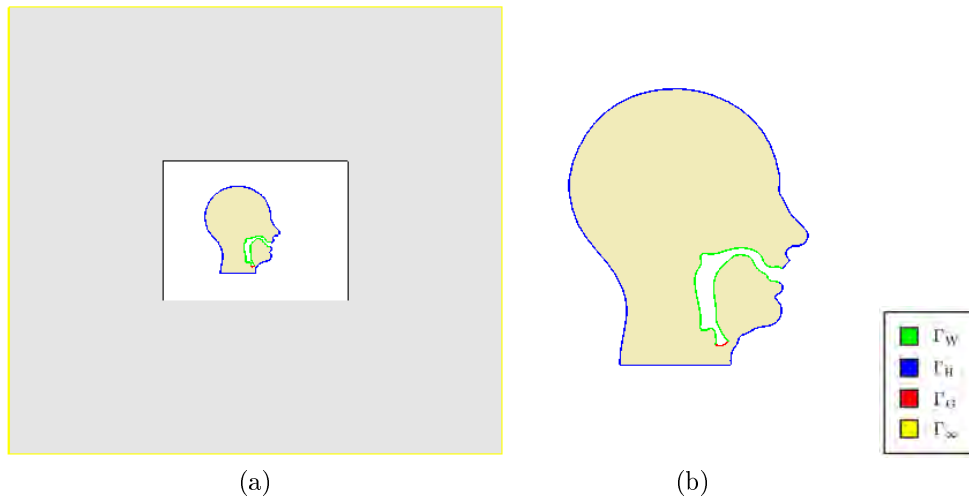


Figure 3.6: Boundary conditions (BC) for the acoustic wave equation within the vocal tract, where Γ_W , Γ_H , Γ_G and Γ_∞ are the Wall, Head, Glottal and Infinity boundaries respectively. The shaded area in (a) denotes the PML region. In (b) can be seen in detail the head region.

With regards to the boundary conditions, we suppose a soft vocal tract wall (Γ_W) and a hard head, i.e. the gradient of the pressure is zero on the boundary Γ_H . On the other hand, we introduce the variable airflow $g(t)$ produced by the glottal model by means of an

inhomogeneous Neumann condition (Γ_G). Finally, we truncate the PML domain by means of an homogeneous Neumann condition (see Figure 3.6). So, the boundary conditions are

$$\frac{\partial t}{\partial t} p(\mathbf{x}, t) \cdot \mathbf{n} = \mu c_0 \quad \text{on } \Gamma_W, t > 0, \quad (3.14)$$

$$\nabla p(\mathbf{x}, t) \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_H, t > 0, \quad (3.15)$$

$$\nabla p(\mathbf{x}, t) \cdot \mathbf{n} = g(t) \quad \text{on } \Gamma_G, t > 0, \quad (3.16)$$

$$\nabla p(\mathbf{x}, t) \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_\infty, t > 0, \quad (3.17)$$

where c_0 stands for the sound speed and μ stands for the coefficient of the boundary admittance (see section 2.2). On the other hand, for simplicity we take the following initial conditions

$$p(x, 0) = 0, \quad \text{in } \Omega, \quad (3.18)$$

$$\partial_t p(x, 0) = 0, \quad \text{in } \Omega. \quad (3.19)$$

3.3.2 Numerical scheme

Next, we will directly present the final explicit scheme for the wave equation within the vocal tract. This problem has been addressed in the different sections of chapter 2. The final explicit scheme for the vector of nodal pressures \mathbf{P} reduces to use the explicit scheme for the PML approach (see section 2.3)

$$\mathbf{P}^{n+1} = \mathbf{C}_1^{-1} (\mathbf{C}_2 \mathbf{P}^n - \mathbf{C}_3 \mathbf{P}^{n-1} + c_0^2 \mathbf{L}^n - c_0^2 \mathbf{K} \mathbf{P}^n + \mathbf{B}_x \Phi_x^n + \mathbf{B}_y \Phi_y^n), \quad (3.20)$$

but redefining the auxiliary matrices \mathbf{C}_1 , \mathbf{C}_2 and \mathbf{C}_3 as

$$\mathbf{C}_1 = \left(\frac{\mathbf{M}}{\Delta t^2} + \frac{\mathbf{M}_\alpha}{2\Delta t} + \frac{\mu c_0 \mathbf{B}}{2\Delta t} \right), \quad (3.21)$$

$$\mathbf{C}_2 = \left(\frac{2\mathbf{M}}{\Delta t^2} - \mathbf{M}_\beta \right), \quad (3.22)$$

$$\mathbf{C}_3 = \left(\frac{\mathbf{M}}{\Delta t^2} - \frac{\mathbf{M}_\alpha}{2\Delta t} - \frac{\mu c_0 \mathbf{B}}{2\Delta t} \right). \quad (3.23)$$

Note that the boundary losses terms (see section 2.2) have been added to the original definition of \mathbf{C}_1 (2.98), \mathbf{C}_2 (2.99) and \mathbf{C}_3 (2.100). On the other hand, no changes are introduced to the explicit schemes for the auxiliary functions Φ_x and Φ_y (see section 2.3), which yields

$$\Phi_x^{n+1} = \Phi_x^{n-1} - 2\Delta t \mathbf{M}^{-1} (\mathbf{M}_{\xi_1} \Phi_x^n - c_0^2 \mathbf{B}_{x,\gamma} \mathbf{P}^n) \quad (3.24)$$

$$\Phi_y^{n+1} = \Phi_y^{n-1} - 2\Delta t \mathbf{M}^{-1} (\mathbf{M}_{\xi_2} \Phi_y^n + c_0^2 \mathbf{B}_{y,\gamma} \mathbf{P}^n) \quad (3.25)$$

Considering that we have no external forces, the above matrices and vectors can be computed as

$$\mathbf{M} = [M^{ab}], \quad M^{ab} = (N^a, N^b), \quad (3.26)$$

$$\mathbf{M}_\alpha = [M_\alpha^{ab}], \quad M_\alpha^{ab} = \langle N^a, \alpha_h N^b \rangle, \quad (3.27)$$

$$\mathbf{M}_\beta = [M_\beta^{ab}], \quad M_\beta^{ab} = \langle N^a, \beta_h N^b \rangle, \quad (3.28)$$

$$\mathbf{K} = [K^{ab}], \quad K^{ab} = (\nabla N^a, \nabla N^b), \quad (3.29)$$

$$\mathbf{L}^n = [L^{n,a}], \quad L^{n,a} = \langle N^a, g^n \rangle_{\Gamma_G}, \quad (3.30)$$

$$\mathbf{B} = [B^{ab}], \quad B^{ab} = (N^a, N^b)_{\Gamma_W}, \quad (3.31)$$

$$\mathbf{B}_x = [B_x^{ab}], \quad B_x^{ab} = (N^a, \partial_x N^b), \quad (3.32)$$

$$\mathbf{B}_y = [B_y^{ab}], \quad B_y^{ab} = (N^a, \partial_y N^b), \quad (3.33)$$

$$\mathbf{B}_{x,\gamma} = [B_{x,\gamma}^{ab}], \quad B_{x,\gamma}^{ab} = (N^a, \gamma_h \partial_x N^b), \quad (3.34)$$

$$\mathbf{B}_{y,\gamma} = [B_{y,\gamma}^{ab}], \quad B_{y,\gamma}^{ab} = (N^a, \gamma_h \partial_y N^b), \quad (3.35)$$

where N stands for the shape function, g^n is the glottal source at time step n and the discrete functions α_h , β_h and γ_h are functions that control the behavior of the perfectly matched layer (see section 2.3).

3.4 An example: synthesis of vowel /e/

3.4.1 Synthesis

In this section, we will present an example of the running of the articulatory synthesizer. We will synthesize the Catalan vowel /e/. First, the geometry is constructed and then meshed (see Figure 3.7). The former has been done by hand tracing the inner boundaries of the vocal tract. Then, an artificial head has been included. Finally, we have surrounded the domain by a PML region of thick $L = 0.5\text{m}$. The computational mesh has been assigned a non structured mesh of triangle elements with size of $h=0.003\text{m}$ over the vocal tract surface and a $h=0.02\text{m}$ to the rest of surfaces. To get a smooth transition between the two surface zones, we have imposed a $h=0.003\text{m}$ over the head boundary. With the above specifications, we have obtained a mesh of 9997 nodes and 19276 elements.

We have used a wave speed $c = 345\text{m/s}$ and a sampling frequency $f_s = 200\text{KHz}$ (i.e the time step $\Delta t = 1/200\text{KHz}$). The PML has been configured to have a reflection coefficient $R = 10^{-4}$, which with the above c and f_s we have got a damping coefficient $\hat{\xi} = 2760$. On the other hand, we consider a wall loss coefficient $\mu = 0.0005$.

As a first step, we have tested the model behavior using a broadband frequency pulse [54]. This pulse has energy until 10KHz, which ensures a good coherence below this range (see Figure 3.8). This pulse has been introduced on the boundary Γ_G (glottal cords). Then, the transfer function between the broadband pulse and the acoustic pressure at the lips

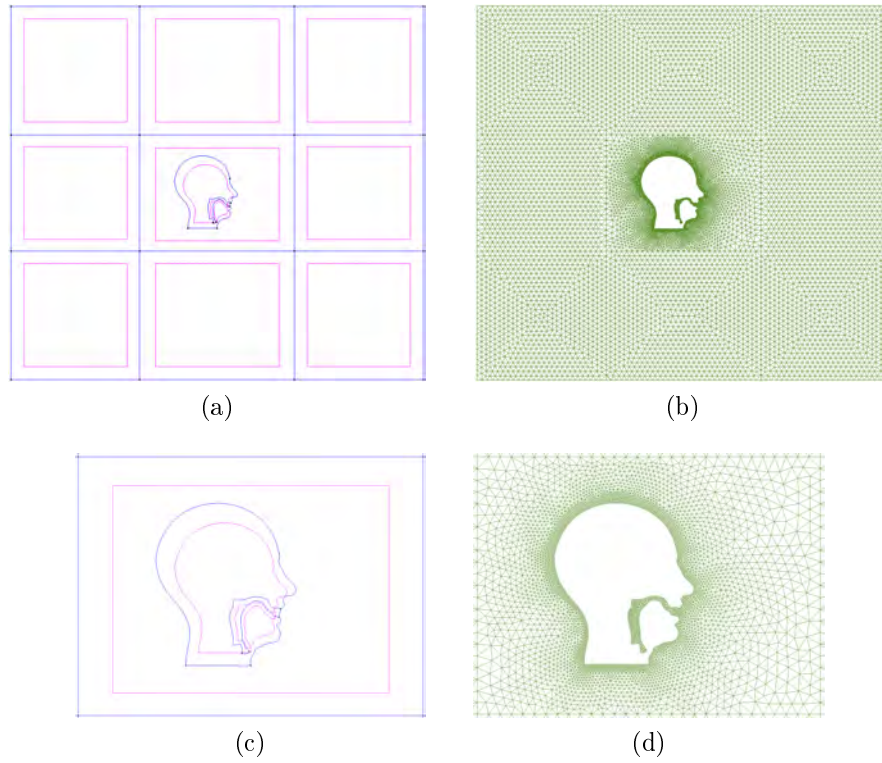


Figure 3.7: Geometry (a) and mesh (b) for the /e/ example. In (c) and (d) the head region is shown in detail.

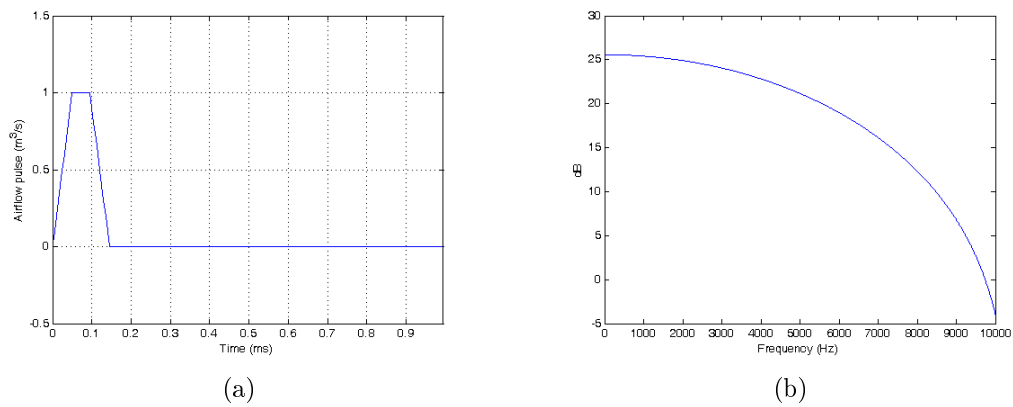


Figure 3.8: Broadband frequency pulse used to test the FE model. (a) is the time signal and (b) its spectrum.

has been computed (see Figure 3.9b). In order to obtain the acoustic pressure at the lips, we have chosen an arbitrary point close to the output of the vocal tract (see Figure 3.9a).

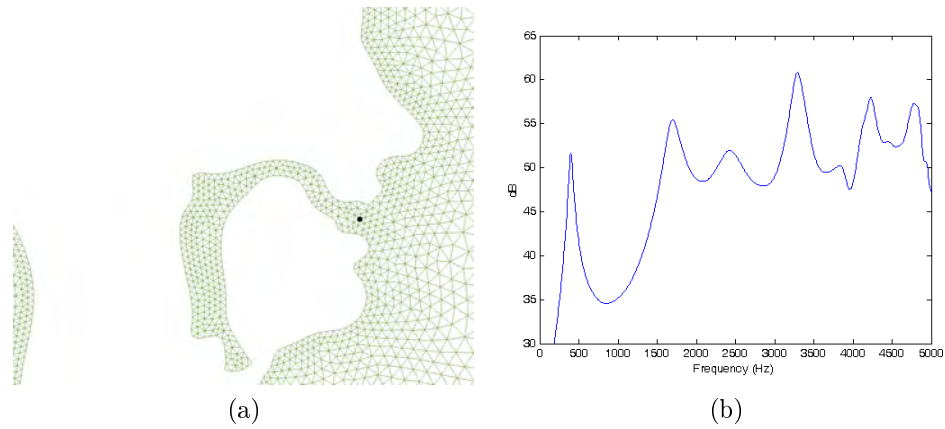


Figure 3.9: (a) Location of the point used to capture the speech signal. (b) Transfer function for vowel /e/.

Second, we have synthesized the vowel /e/ using the glottal pulse provided by the LF model (see subsection 3.2.2). In this example, we have configured this glottal pulse with a fundamental frequency or pitch $F_0 = 100\text{Hz}$ and a duration time $t_a = 1\text{ms}$. In figure 3.10 you can see the acoustic pressure obtained at the lips and its spectrum. Its envelope has also been calculated and overprinted using a LPC analysis. In order to do the LPC analysis, the time signal has been downsampled to $f_s = 40\text{KHz}$. Then, 50 coefficients has been used to estimate the signal.

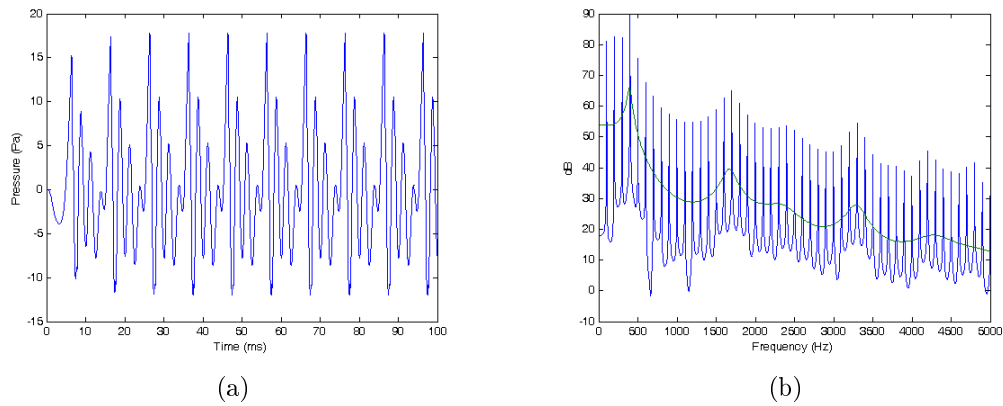


Figure 3.10: Acoustic pressure (a) of /e/ and its spectrum and envelope (b). The time signal corresponds to the first 100 milliseconds of vowel /e/. The spectrum has been computed in a stable region of the signal and then its envelope has been obtained using a LPC analysis.

Finally, some snapshots for the /e/ vowel acoustic pressure can be observed in Figure 3.11. Note that we have adjusted their amplitude for a clear visualization of the sound waves coming from the mouth. So, sometimes the acoustic pressure in the vocal tract can not be observed because its amplitude is too high.

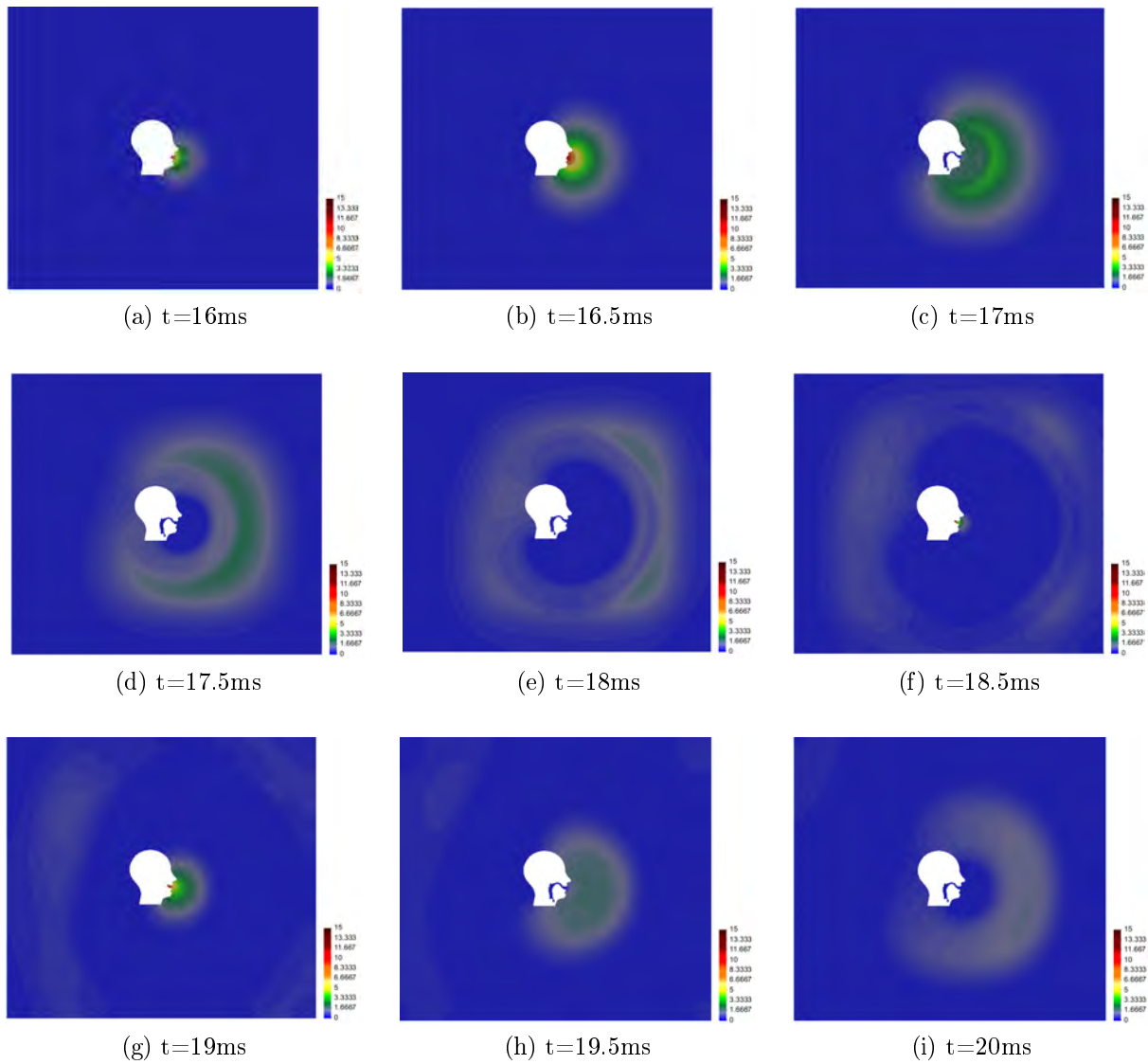


Figure 3.11: Snapshots of the /e/ acoustic pressure for different time instants.

3.4.2 Analysis of the results

In order to verify the behavior of the FE model, we have compared the obtained frequency formants with other formants of real voice. We have used some results of a frequency analysis study of Catalan vowels done by Recasens *et al.* [45]. Recasens *et al.* had studied the variations on vowel formants for four catalan dialects: majorcan, valencian, western catalan and eastern catalan. On the other hand, given that in this work we have focused on the FE modelling, we have only evaluated it measuring the location of the first two formants, which could tell us objectively which vowel has been generated.

We have computed the transfer function of the FE model for /e/ using different wall loss coefficients (μ) (see Figure 3.12). Then, the location of the first two formants has

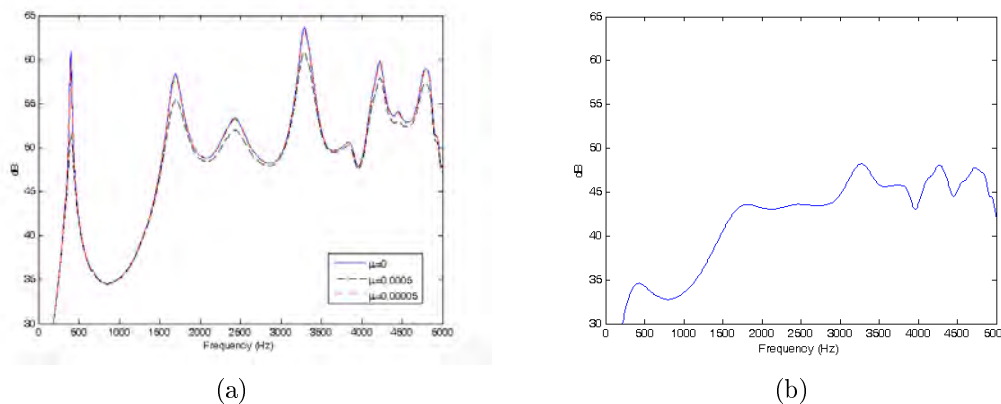


Figure 3.12: Transfer functions of vowel /e/ for different wall loss coefficients (μ). In (b) there is an example of over-damped transfer function.

been calculated (see Table 3.1). Their bandwidths have also been computed for a better understanding of the effects of wall losses. It can be seen how the bigger the losses the higher the formant bandwidth. No significant variations on the formant location can be observed, except when wall losses are too large, in which case the formants practically vanish (see Figure 3.12b). So, this seems that wall losses hardly affect to the location of first two formants. On the other hand, if we compare the obtained formant values with any of the real voice formants, we can see small variations (50Hz) in the location of the first formant and bigger variations (150Hz) for the second.

On the other hand, if we plot the formant location in a vowel chard (F1 vs F2), we can see that the obtained results are not so far from /e/ (see Figure 3.13). Moreover, doing an informal perceptual test, the vowel /e/ has been detected. The problem on the formant location could be caused by the use of a rough geometry, by some problems on the FE modelling, by problems on the source coupling, etc. These points will be studied in detail in future works.

	F1/BW1 (Hz)	F2/BW2 (Hz)	F3/BW3 (Hz)
$\mu=0$	397/21	1692/142	2426/380
$\mu=0.00005$	397/24	1692/148	2426/388
$\mu=0.0005$	397/58	1698/203	2427/486
Majorcan /e/	489/ -	1905/ -	2656/ -
Valencian /e/	460/ -	1837/ -	2575/ -
Western /e/	448/ -	1854/ -	2552/ -
Eastern /e/	450/ -	1839/ -	2571/ -

Table 3.1: Formant location (Fi) and bandwidth (BW_i) of the first three formants (i=1,2,3) corresponding to the 2D vocal tract when /e/ is synthesized. Different wall losses have been considered. The formant values of /e/ for four catalan dialects have been also provided [45]

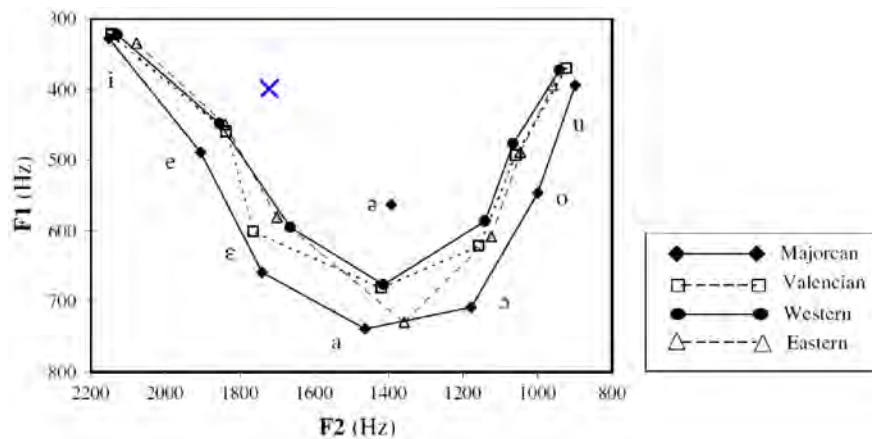


Figure 3.13: Vowel triangle for the four dialects presented in [45]. The synthesized /e/ is plotted as a blue cross.

3.4.3 Some remarks on vowel synthesis quality

In the above subsection we have studied the location of the first two formants. However, the quality of the synthesized speech not only depends on the formant location but also on their bandwidth and energy. Moreover the whole spectrum has to be analyzed.

We have also seen that the higher the wall losses the lower the formant bandwidth and energy. So, tuning the wall losses seems to be a necessary tool to adjust the formant bandwidth, and therefore to improve the quality of synthesized vowel.

On the other hand, the vowel quality not only depends on the acoustic vocal tract model but also on the glottal and geometry model. To illustrate it, we present some examples changing the geometry and the glottal source.

First, we want to observe spectrum changes due to geometry changes. To build the geometry, we have transformed the /e/ geometry of a 1D synthesizer into 2D. We have supposed a 2D tube model such that its radius is

$$r(x) = \sqrt{S(x)/\pi}, \quad (3.36)$$

where $S(x)$ is the cross sectional area of the 1D tube used in [1, 12]. The geometry and mesh used can be seen in Figure 3.14. We also use the same domain and parameter configuration than in the real vocal tract example (PML region, speed wave, sampling frequency, mesh criteria, etc.). Then, we have computed the transfer function of the tube example using the above methodology, but in this case, for simplicity, we have not considered wall losses. The obtained transfer function can be seen in Figure 3.15 overprinted with the transfer function of the 2D vocal tract (“Tube” vs “2D” in Figure 3.15). The formant values are given in Table 3.2. We can see how the location of the first two formants are the same in both examples. This seems to indicate that different geometries corresponding to the same vowel do not change the location of the first two formants. On the other hand, it seems that the 2D vocal better reproduce the higher frequency range (above 3KHz) because some formants not shown in tube becomes clearly visible. All these effects may be studied in detailed in future works.

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)
/e/ tube	429	1706	2635
/e/ 2D	397	1692	2426

Table 3.2: Formant location of the 2D vocal tract vs tube when /e/ is synthesized. No wall losses have been considered.

Second, we have computed the acoustic pressure at the lips for the 2D vocal tract using different glottal sources. We have used the same configuration described in the synthesis

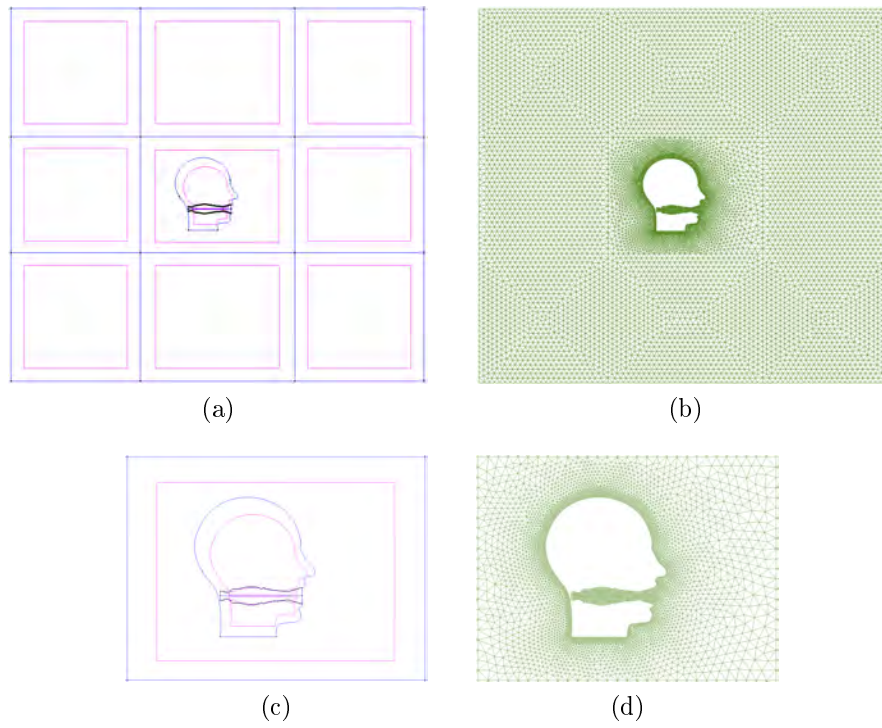


Figure 3.14: Obtained geometry (a) and mesh (b) for the /e/ tube example. The head region can be seen in detail in (c) and (d).

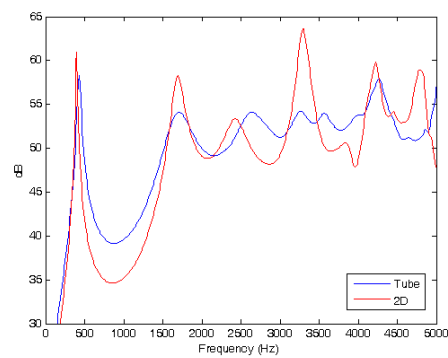


Figure 3.15: Transfer function of the tube and the 2D vocal tract.

example for vowel /e/ (see subsection 3.4.1). Then, we have calculated its spectrum and applied a LPC analysis with 50 coefficients to obtain its envelope. We have done it with the LF model with $t_a = 3\text{ms}$, $t_a = 1\text{ms}$ and the Rosenberg model (see subsection 3.2.2). The results are shown in Figure 3.16. We can see that for different glottal sources we get different spectra. The location of the first two formants do not change, but their bandwidth and energy do. Moreover, the spectrum above the second formant changes. So, different glottal sources will result in different vowel quality.

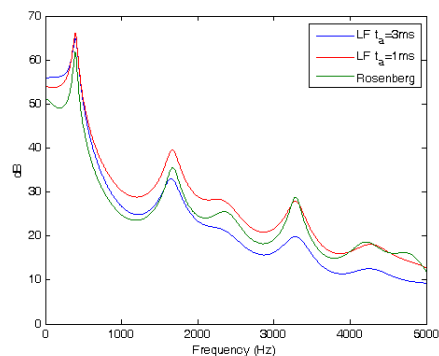


Figure 3.16: Envelope of vowel /e/ using different glottal sources: LF model with $t_a = 3\text{ms}$ and $t_a = 1\text{ms}$ and Rosenberg model (see subsection 3.2.2).

Chapter 4

Conclusions and future work

4.1 Conclusions

In this work, we have proposed to develop a computational strategy based on the use of finite element methods for articulatory speech synthesis. We have dealt with the problem of synthesis of vowels and as an example, we have synthesized the catalan vowel /e/.

First, a brief review related to articulatory speech synthesis has been provided. We have discussed about geometry, glottal and vocal tract acoustic models, being the latter at the core of this work. Basically, two main types of vocal tract models can be distinguished:

- Tube models

The duct or tube models can be subdivided into the ABCD matrix based models, the Digital Waveguides models and the circuit analogy models. These models approximate the vocal tract geometry as a finite set of concatenative tubes, each one having constant cross section, and need to do a lot of approximations in the acoustic modelling process.

- Computational models

We have computational models, which offer wider possibilities than the above duct models. Complex geometries can be implemented in full detail, coarticulation can be included, and the aeroacoustics involved in the generation of many sounds can be taken into account. We basically can distinguish between frequency and time domain models. The former perform a modal analysis through which a frequency response is obtained. Then, this is introduced in a source-filter type model for speech synthesis. In contrast to frequency domain models, time domain models compute directly the acoustic pressure at the output of the vocal tract. Moreover, we have also seen that frequency domain methods need to work in steady state regime. So, they can not deal with the most natural aspects of speech production such as pitch or amplitude variations of the glottal source, coarticulation of sounds, etc. The above aspects can only be considered in time domain methods.

Consequently we have chosen a computational model due to its versatility with complex geometries and its wide possibilities in the acoustic modelling. Moreover we have decided to work in the time domain because we aim at synthesizing natural speech. To address this problem we have chosen the finite element method.

On the one hand, among the existing geometrical models (statistical, geometrical or biomechanical), we have decided to use a statistical model because they are more precise and realistic than the others. On the other hand, as a glottal model we have used a parametric model because its simplicity and its proper behavior for vowel synthesis. Two models have been hardly compared: the C Rosenberg model and the LF model. The former is the simplest but introduces excessive high frequency content, which could cause numerical errors. However, the latter is more complex but better models the waveform of the glottal source, ensuring that no excess of high frequency is introduced.

Second, given the complexity of the speech problem, we have considered simpler approaches, following a bottom-up strategy. These problems have been solved using the finite element method for the spatial discretization and finite differences for the time discretization, obtaining an explicit scheme. Then, these schemes have been validated by means of benchmark problems. The considered approaches have been the following:

- The acoustic wave equation

As a first step, we have computed the sound wave propagation in a closed domain by solving the hyperbolic wave equation.

- The acoustic wave equation with boundary losses

Next, losses due to viscous friction and heat conduction of the acoustic waves at the boundaries have been introduced into the acoustic wave equation. We have used an approach that assumes a constant frequency absorption.

- The acoustic wave equation with a Perfectly Matched Layer (PML)

Finally, a non-reflection condition has been considered to simulate propagation of sound waves towards infinity. A Perfectly Matched Layer (PML) has been used for this purpose.

Third, once solved the numerical difficulties of finite element methods for acoustics, we have applied them to the problem of the synthesis of vowels. We have seen some important features that have to be considered in the acoustic modelling. We have implemented radiation losses (PML), wall losses (boundary losses) and a rigid model of the vocal tract.

As an example, we have synthesized the catalan vowel /e/. Then, we have evaluated its quality using objective measurements. We have computed its transfer function using a broadband pulse. These pulses are often used in transient analysis. Thanks to this kind of pulses, we have obtained an accurate transfer function for the vocal tract. Some important conclusions have been obtained:

- Location of the first two formants

We have measured the location of the first two formants for /e/ which have been compared with the formant location of different catalan dialects. The obtained results have not been so good as we initially expected. However, when plotted it into a vowel graph (F1 vs F2) the results were not far from a real /e/. We have also tested another geometry of an /e/ corresponding to a 1D tube synthesizer, transforming its geometry to two dimensions. We have seen than the first two formants have also coincided with the obtained results in the real vocal tract. These problems could be caused by a rough geometry, by some problems on the FE modelling, by problems on the source coupling, etc. These points will be solved in future works.

- Other aspects in vowel quality

We have given some key points to adjust the vowel spectrum and therefore to improve higher synthesised vowel quality. We have seen that the higher the losses at the boundaries the higher the bandwidth and the lower the energy formant. Moreover, geometries variations for a given vowel hardly affect the location of the first two formants. However, their energy and bandwidth change and the others formants also do. On the other hand, from the comparison of the tube example with the real vocal tract, it can be observed that the real vocal tract yields better results for the high frequency range (above 3KHz).

4.2 Future work

Let us next focus on possible research lines.

First, some improvements regarding the computational modelling of the vocal tract could be done. The most important are:

- Boundary losses

The boundary losses are a kind of boundary condition that is difficult to impose in time domain models because it has a frequency dependent behavior. In this work we have used the same value of boundary impedance for all frequencies. More complex approaches have already done do deal with this problem [13]. Studying these techniques could be interesting.

- Truncation of the PML

To reduce the computational cost, we could try to limit the extend of the computational domain. For example, it could be interesting to only consider a semi-infinite computational domain in the forward direction of the sound waves (i.e. backward waves are not computed because they are absorbed by the PML). With this approach the number of elements of the PML could be half reduced. However, some simulations have been done and numerical instabilities appear in the corners with truncated PML. This could be worth analysing.

Next, the quality of the synthesized speech could be evaluated objectively and subjectively:

- Objective measurements

Objective measurements generally need an acoustical reference, which is not always a simple task [1, 12]. Typical objective measurements are the formant location and their bandwidth and energy. The whole spectrum should have to be analyzed using the above measurements,

- Subjective measurements

Subjective evaluations could be done, such as MOS (Mean Opinion Score) or PSEQ (Perceptual Evaluation of Speech Quality) [9], are often used to evaluate the overall quality of the synthetic speech.

By means of the vowel quality analysis, the articulatory model could be tuned to achieve the highest possible quality. As a first step, given a glottal source and a geometry, the internal parameters of the vocal tract acoustic model would have to be adjusted. This strategy could be as follows [1]:

- Generate transfer functions and synthesized speech for several configurations of the vocal tract acoustic model

We have seen that the transfer function depends among other factors on the losses. So, we could calculate the transfer function for different values of the losses, obtaining a set of candidates to be the best configuration. The corresponding synthesized vowel would also be computed.

- Voice quality analysis using objective measurements

As a second stage, the above set of transfer functions would have to be analyzed using objective measurements such as location, bandwidth and energy of the formants. Then, these results would have to be compared with a reference model analyzed under the same conditions. At this stage, a first estimation of the proper parameters may be obtained. All results that are far from this estimation may be discarded.

- Voice quality analysis using subjective measurements

Third, with the preselected configurations in the above stage, subjective measurements would be carry out. Formal perceptual tests would be performed with the corresponding synthesized vowels, from which a set of candidates may be obtained.

- Choose the configuration that satisfies both measurements

Finally, the best configuration would be the intersection between the objective and subjective candidates.

Once the vocal tract model had been adjusted, we could additionally carry out a study of the vowel quality for different glottal sources and geometries. This process could be similar to the realized above for the internal parameters of the vocal tract model.

Finally, once the vowel problem had been solved, we could increase the complexity of the problem and proceed to the synthesis of other sounds. The proper steps could be

- Synthesis of diphthongs

First, we could deal with the synthesis of diphthongs (e.g. /ai/, /ei/, etc.). In contrast to static vowels, the coarticulation of vowels will require a time-varying geometry. This implies in turn the use of an ALE (Arbitrary Lagrangian-Eulerian) formulation (see e.g., [20], [24], and [55] for recent advances) for the involved equations, to be solved within the FEM framework. Moreover, the complexity of the geometry model will also increase, now being dynamic.

- Synthesis of syllables

Second, we would address the synthesis of syllables (e.g. /na/, /sa/, etc.). In this process nasal consonants would be generated (e.g. /n/), which requires a more complex geometry given that the nasal cavity has to be included. Syllables containing fricatives consonants (e.g. /sa/) would pose more challenging problems as we would need to consider aeroacoustic effects. An acoustic analogy approach (e.g., Lighthill's acoustic analogy, [19]) would be followed involving a first Computational Fluid Dynamic (CFD) simulation of the airflow in the vocal tract, from which an acoustic source term could be extracted. Then, the modified wave equation with the source term from the CFD acting as an inhomogeneous term could be solved, to obtain the acoustic pressure at the exit of the vocal tract. All involved equations would be solved again using FEM in an ALE formulation.

For the above processes, the speech quality of the synthesized speech should be also evaluated.

Appendix A

Numerical computation in 2D

In this appendix we show some basic concepts and tools that are used to compute the integrals that appear in the finite element formulation. We focus on the 2D problem and we use as example the computation of the stiffness matrix. Prior to describe it, we present the element point of view which allows us to perform integrals over an element and then over a master element, where standard numerical integration techniques can be applied.

A.1 Introduction

One of the most common matrices that appear in the 2D problem is the stiffness matrix \mathbf{K} with entries

$$K_{ab} = \int_{\Omega} \nabla N_a \cdot \nabla N_b \, dx \, dy, \quad (\text{A.1})$$

where N_a and N_b stand for the shape functions. In this appendix we describe some numerical approaches used to compute K_{ab} .

A.2 The element point of view

Up to this point, we have worked with the global point of view, where the properties of the finite element problem has been imposed. This point of view consider the whole domain. However, in order to compute the integrals that appear in the finite element formulation, another point of view that focuses on each element is necessary. This is the so called element or local point of view [23] (see Figure A.1).

By means of the element point of view, an integral over a domain Ω can be computed as the sum of the individual integrals over each element. So, we can use the decomposition

$$\int_{\Omega} f \, d\Omega = \sum_{e=1}^{N_e} \int_{\Omega_e} f \, d\Omega_e, \quad (\text{A.2})$$

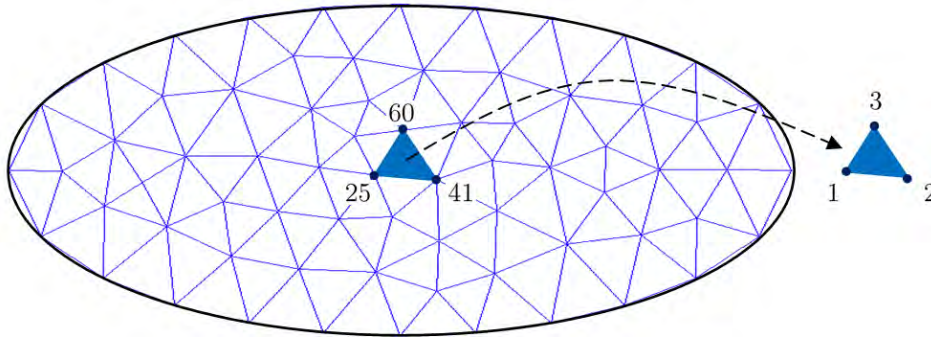


Figure A.1: A 2D triangle mesh: global and local point of view. The global point of view corresponds to the whole domain while the local point can be understood as one triangle. Note also the nodal numeration used in each case.

where N_e is the number of elements of the mesh and Ω_e is the element domain. In the 2D problem, the above expression is

$$\int_{\Omega} f d\Omega = \sum_{e=1}^{N_e} \int_{\Omega_e} f dx dy, \quad (\text{A.3})$$

where x and y are the local or element coordinates in the element domain Ω_e .

A.3 Numerical computation over a master element

In this section we describe how to perform the integration of the shape function and its first derivatives, i.e.

$$\int_{\Omega_e} N^e(x, y) dx dy, \quad (\text{A.4})$$

$$\int_{\Omega_e} \nabla N^e(x, y) dx dy, \quad (\text{A.5})$$

which appear in the construction of the stiffness and mass matrix.

A.3.1 Coordinate transformation and shape functions

In order to numerically compute integrals over an element domain Ω_e , it is necessary to make a coordinate transformation to a new domain where the integration rules can be applied. For example, a typical domain in one-dimension numerical integrals is $[-1, 1]$. This new domain is known as the master element domain, which will be referred in this work as $\hat{\Omega}_e$. So, we need a transformation function T^e that maps the reference coordinates $\hat{\mathbf{x}} = \{\hat{x} \hat{y}\}^T$ of the master element to the physical coordinates $\mathbf{x} = \{x y\}^T$ of an element

e (see Figure A.2).

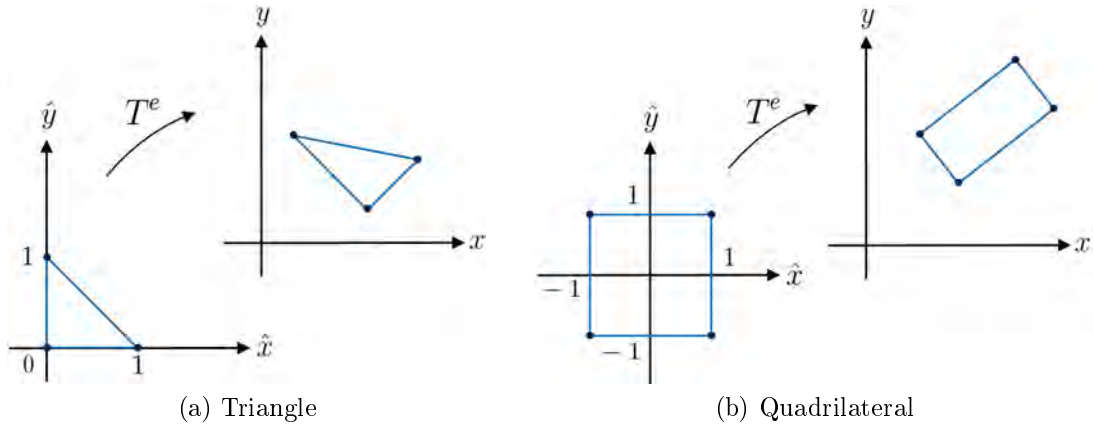


Figure A.2: Master elements and coordinate transformation T^e

The transformation function can be written as

$$\mathbf{x} = T^e(\hat{\mathbf{x}}) = \sum_{j=1}^m \hat{N}_j^e(\hat{\mathbf{x}}) \mathbf{x}_j^e, \quad (\text{A.6})$$

where \mathbf{x}_j^e are the physical coordinates of the local node j of the element e , and $\hat{N}_j^e(\hat{\mathbf{x}})$ is the finite element interpolation function corresponding to the node j of the master element $\hat{\Omega}_e$. The above expression is called the geometrical interpolation.

For example, for the following master elements

- Triangular elements with 3 nodes

The master element is such that $0 \leq \hat{\mathbf{x}} \leq 1$, and if this has 3 nodes ($m = 3$), the transformation function is

$$\mathbf{x} = T^e(\hat{\mathbf{x}}) = \sum_{j=1}^3 \hat{N}_j^e(\hat{\mathbf{x}}) \mathbf{x}_j^e, \quad (\text{A.7})$$

and the interpolation functions are [23]

$$\hat{N}_j^e(\hat{x}, \hat{y}) = \begin{cases} \hat{x} & j = 1 \\ \hat{y} & j = 2 \\ 1 - \hat{x} - \hat{y} & j = 3 \end{cases} \quad (\text{A.8})$$

- Quadrilateral elements with 4 nodes

In the case of a quadrilateral element $-1 \leq \hat{\mathbf{x}} \leq 1$, and if 4 nodes are defined ($m = 4$), the transformation function is

$$\mathbf{x} = T^e(\hat{\mathbf{x}}) = \sum_{j=1}^4 \hat{N}_j^e(\hat{\mathbf{x}}) \mathbf{x}_j^e, \quad (\text{A.9})$$

and the interpolation functions [23] are

$$\hat{N}_j^e(\hat{x}, \hat{y}) = \frac{1}{4} \begin{cases} (1 - \hat{x})(1 - \hat{y}) & j = 1 \\ (1 + \hat{x})(1 - \hat{y}) & j = 2 \\ (1 + \hat{x})(1 + \hat{y}) & j = 3 \\ (1 - \hat{x})(1 + \hat{y}) & j = 4 \end{cases} \quad (\text{A.10})$$

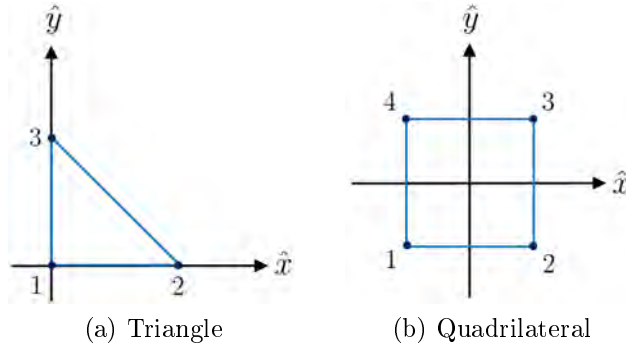


Figure A.3: Node numeration of the master elements.

There also exist triangular and quadrilateral elements with higher order of interpolations and corresponding master elements (see e.g. [23]), but the above are quite common and will serve our purposes. On the other hand, given that the polynomials that have been used for the geometrical interpolation are first order, and the used interpolation for the pressure is first order too, we can use the same shape functions in both, hence

$$\hat{N}^e(\hat{\mathbf{x}}) = N^e(\hat{\mathbf{x}}). \quad (\text{A.11})$$

This is called the isoparametric case, and it is the most used in finite element methods.

A.3.2 Mapping the integrals to the reference domain

Using the coordinate transformation T^e , we can compute the integral of the shape function $N^e(x, y) \in \Omega_e$ on the element domain $\hat{\Omega}_e$ as

$$\int_{\Omega_e} N^e(x, y) dx dy = \int_{\hat{\Omega}_e} N^e(\hat{x}, \hat{y}) |J| d\hat{x} d\hat{y}, \quad (\text{A.12})$$

where $|J|$ is the Jacobian determinant of the transformation T^e :

$$J = \frac{\partial x_i}{\partial \hat{x}_i} = \begin{pmatrix} \frac{\partial x}{\partial \hat{x}} & \frac{\partial y}{\partial \hat{x}} \\ \frac{\partial x}{\partial \hat{y}} & \frac{\partial y}{\partial \hat{y}} \end{pmatrix}. \quad (\text{A.13})$$

This can be computed by means of the coordinate transformation as

$$J = \frac{\partial x_i}{\partial \hat{x}_i} = \sum_{j=1}^m \left(\frac{\partial N(\hat{x}, \hat{y})}{\partial \hat{x}_i} \right) x_i. \quad (\text{A.14})$$

On the other hand, we can map the integral of the shape function derivative to the master domain as

$$\int_{\Omega_e} \partial_{x_i} N^e(x, y) dx dy = \int_{\hat{\Omega}_e} \partial_{x_i} N^e(\hat{x}, \hat{y}) |J| d\hat{x} d\hat{y}. \quad (\text{A.15})$$

Using the chain rule for partial differentiation, we have

$$\frac{\partial N^e(\hat{x}, \hat{y})}{\partial x_i} = \frac{\partial N^e(\hat{x}, \hat{y})}{\partial \hat{x}_i} \frac{\partial \hat{x}_i}{\partial x_i}. \quad (\text{A.16})$$

The first term $\partial_{\hat{x}_i} N^e(\hat{x}, \hat{y})$ can be simply computed deriving the shape function for the master element. Regarding to the second term, it can be computed as

$$\frac{\partial \hat{x}_i}{\partial x_i} = \left(\frac{\partial x_i}{\partial \hat{x}_i} \right)^{-1} = (J)^{-1}. \quad (\text{A.17})$$

Hence, the problem reduces to computing integrals over the master element. In the following subsection, the numerical method used to solve these integrals is described.

A.3.3 Numerical Integration

Between the most populars rules used for numerical integration (e.g. trapezoidal rule, Simpson's rule,...), finite element methods normally use the Gaussian Quadrature rule as to it is just accurate as the others but involve fewer integration points [23]. This issue is important in practice since the fewer the integration points the less the cost [23].

So, using a Gauss quadrature, an integral of a function $F(\hat{x}, \hat{y}) \in \hat{\Omega}_e$ can be computed as

$$\int_{\hat{\Omega}_e} F(\hat{x}, \hat{y}) d\hat{x} d\hat{y} \approx \sum_{k=1}^{n_{int}} F(\hat{\xi}_k, \hat{\eta}_k) w(k), \quad (\text{A.18})$$

where n_{int} is the number of integration points, $\hat{\xi}_k$ and $\hat{\eta}_k$ are the coordinates of the integration point, and $w(k)$ is the weight.

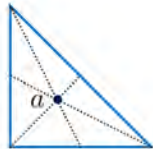
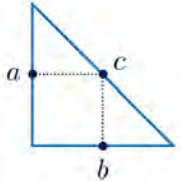
n_{int}	Degree	$\hat{\xi}$	$\hat{\eta}$	w	Point	Geometry
1	1	1/3	1/3	1	a	
3	2	0	0.5	1/3	a	
		0.5	0	1/3	b	
		0.5	0.5	1/3	c	

Table A.1: Gauss coordinates $(\hat{\xi}, \hat{\eta})$ and weights (w) for one and three integration points (n_{int}) over a triangle. "Degree" stands for the degree of the polynomial used in the geometrical interpolation.

A.4 Calculation of the Stiffness matrix

A.4.1 Stiffness matrix

The stiffness matrix $\mathbf{K} = [K_{ab}]$ is

$$K_{ab} = \int_{\Omega} \nabla N_a(x, y) \cdot \nabla N_b(x, y) \, dx \, dy. \quad (\text{A.19})$$

Using the element point of view, the above expression can be computed as the sum of the integrals in each element

$$K_{ab} = \sum_e \int_{\Omega_e} \nabla N_a^e(x, y) \cdot \nabla N_b^e(x, y) \, dx \, dy. \quad (\text{A.20})$$

In order to compute the integral in Ω_e , we have seen that it has to be computed over the master element domain $\hat{\Omega}_e$

$$K_{ab} = \sum_e \int_{\hat{\Omega}_e} \nabla N_a^e(\hat{x}, \hat{y}) \cdot \nabla N_b^e(\hat{x}, \hat{y}) |J| \, d\hat{x} \, d\hat{y}. \quad (\text{A.21})$$

Then, using a quadrature rule for the computation of the integral over the master element, the stiffness matrix can be computed as

$$K_{ab} \approx \sum_e \sum_k \nabla N_a^e(\hat{\xi}, \hat{\eta}) \cdot \nabla N_b^e(\hat{\xi}, \hat{\eta}) |J(\hat{\xi}, \hat{\eta})| w(k). \quad (\text{A.22})$$

A.4.2 Memory efficiency

On the other hand, for memory efficiency reasons, it could be interesting to directly compute

$$K_{ab} P_b \approx \sum_b \sum_e \sum_k \nabla N_a^e(\hat{\xi}, \hat{\eta}) \cdot \nabla N_b^e(\hat{\xi}, \hat{\eta}) |J(\hat{\xi}, \hat{\eta})| w(k) P_b. \quad (\text{A.23})$$

Finally, rearranging some terms we set

$$K_{ab} P_b \approx \sum_e \sum_k \sum_b \nabla N_a^e(\hat{\xi}, \hat{\eta}) \cdot \nabla N_b^e(\hat{\xi}, \hat{\eta}) P_b |J(\hat{\xi}, \hat{\eta})| w(k). \quad (\text{A.24})$$

Bibliography

- [1] M. Arnela. *Síntesi de vocals mitjançant diferències finites*. Master Thesis, La Salle - Universitat Ramon Llull, 2009.
- [2] P. Birkholz and D. Jackèl. A three-dimensional model of the vocal tract for speech synthesis. In *15th ICPhS*, pages 2597–2600, Barcelona, Spain, 2003.
- [3] P. Birkholz and D. Jackèl. Boundary-layer resistance in time-domain simulations of the vocal tract system. In *European Signal Processing Conference (EUSIPCO'04)*, pages 999–1002, Vienna, Austria, 2004.
- [4] P. Birkholz and D. Jackèl. Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system. In *Proc. of Interspeech*, pages 1125–1128, Jeju Island, Korea, 2004.
- [5] P. Birkholz, D. Jackèl, and K. Kroger. Construction and control of a three-dimensional vocal tract model. In *Proc. of ICASSP'06*, pages 873–876, Toulouse, France, 2006.
- [6] P. Birkholz, D. Jackèl, and K. Kroger. Simulation of losses due to turbulence in the time-varying vocal system. *IEEE Trans. Audio, Speech, Lang. Process.*, 15:1218–1225, 2007.
- [7] P. Birkholz and K. Kroger. Vocal tract model adaptation using magnetic resonance imaging. In *Proc. of ISSP'06*, pages 493–500, Ubatuba, Brazil, 2006.
- [8] E. Cataldo, R. Sampaio, and L. Nicolato. Uma discussão sobre modelos mecânicos de laringe para síntese de vogais. *Engevista*, 6(1):47–57, 2004.
- [9] M. Cernak and M. Rusko. An evaluation of synthetic speech using the pseq measure. In *Proc. of Forum Acusticum*, volume 4, pages 2725–2728, Budapest, Hungary, 2005.
- [10] G. C. Cohen. *Higher-Order Numerical Methods for Transient Wave Equations*. Scientific Computation, Springer, 2002.
- [11] J. Dang and K. Honda. Construction and control of a physiological articulatory model. *J. Acoust. Soc. Am.*, 115(2):853–870, 2004.

- [12] K. V. den Doel and M. Ascher. Real-time numerical solution of webster's equation on a nonuniform grid. *IEEE Trans. Speech Audio Process.*, 16(6):1163–1172, 2008.
- [13] B. V. den Nieuwenhof and J.-P. Coyette. Treatment of frequency-dependent admittance boundary conditions in transient acoustic finite/infinite-element models. *J. Acoust. Soc. Am.*, 110(4):1743–1751, 2001.
- [14] G. Fant, J. Liljencrants, and Q. Lin. A four-parameter model of glottal flow. *STL-QPSR, KTH, Stockholm, Sweden*, 77(4):1–13, 1985.
- [15] S. Fels, J. E. Lloyd, K. van den Doel, F. Vogt, I. Stavness, and E. Vatikiotis-Bateson. Developing physically-based, dynamic vocal tract models using artisynt. In *7th International Seminar on Speech Production*, volume 4, pages 419–426, Ubatuba, Brazil, 2006.
- [16] S. Fels, I. Stavness, A. Hannam, J. Lloyd, P. Anderson, C. Batty, H. Chen, C. Combe, T. Pang, T. Mandal, B. Teixeira, S. Green, R. Bridson, A. Lowe, F. Almeida, J. Fleetham, and R. Abugharbieh. Advanced tools for biomechanical modeling of the oral, pharyngeal, and laryngeal complex. In *International Symposium on Biomechanics, Healthcare and Information Science. Pages electronic proceedings*, pages 2725–2728, 2009.
- [17] J. Flanagan and L. Landgraf. Self-oscillating source for vocal-tract synthesizers. *IEEE Trans. Speech Audio Process.*, 16(1):57–64, 1968.
- [18] M. Grote and I. Sim. Efficient pml for the wave equation. *Global Science Preprint, arXiv:1001.0319v1 [math.NA]*, pages 1–15, 2010.
- [19] O. Guasch and R. Codina. Computational aeroacoustics of viscous low speed flows using subgrid scale finite element methods. *J. Acoust. Soc. Am.*, 17(3):309–330, 2009.
- [20] C. Hirt, A. Amsden, and J. Cook. An arbitrary lagrangian-eulerian computing method for all flow speeds. *J. Comput. Phys.*, 14:227–254, 1974.
- [21] M. Howe and R. McGowan. Aeroacoustics of [s]. *Proc. R. Soc. A*, 461:1005–1028, 2005.
- [22] M. Howe and R. McGowan. Sound generated by aerodynamic sources near a deformable body, with application to voiced speech. *J. Fluid Mech.*, 592:367–392, 2007.
- [23] T. Hughes. *The finite element method. Linear Static and Dynamic Finite Element Analysis*. Dover publications, INC. Mineola, New York, 2000.
- [24] T. Hughes, W. Liu, and T. Zimmermann. Lagrangian-eulerian finite-element formulation for compressible viscous flows. *Comput. Methods Appl. Mech. Engrg.*, 29:329–349, 1981.

- [25] K. Ishizaka and J. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst. Tech. Journal*, 51:1233–1268, 1972.
- [26] J. F. Jallona and F. Berthommier. A semi-automatic method for extracting vocal tract movements from x-ray films. *Speech Commun.*, 51(2):97–115, 1968.
- [27] T. Kako and K. Touda. Numerical method for voice generation problem. *J. Acoust. Soc. Am.*, 14(1):45–56, 2006.
- [28] J. Kelly and C. Lochbaum. Speech synthesis. In *Proc. Fourth ICA*, pages 1–4, Copenhagen, Denmark, 1962.
- [29] H. Kjellström and O. Engwall. Audiovisual-to-articulatory inversion. *Speech Commun.*, 51(3):195–209, 2009.
- [30] M. Krane. Aeroacoustic production of low-frequency unvoiced speech sounds. *J. Acoust. Soc. Am.*, 118(1):410–427, 2005.
- [31] M. Krane, M. Barry, and T. Wei. Unsteady behavior of flow in a scaled-up vocal folds model. *J. Acoust. Soc. Am.*, 122(6):3659–3670, 2007.
- [32] M. Krane and T. Wei. Theoretical assessment of unsteady aerodynamic effects in phonation. *J. Acoust. Soc. Am.*, 120(3):1578–1588, 2006.
- [33] B. Kröger and P. Birkholz. *A gesture-based concept for speech movement control in articulatory speech synthesis*, A. Esposito et al. (Eds.): *Verbal and Nonverbal Communication Behaviours*, pages 174–189. Springer, 2007.
- [34] B. Kröger and P. Birkholz. *Articulatory Synthesis of Speech and Singing: State of the Art and Suggestions for Future Research*, A. Esposito et al. (Eds.): *Multimodal Signals: Cognitive and Algorithmic Issues*, volume 5398/2009, pages 306–319. Springer, 2009.
- [35] G. Link, M. Kaltenbacher, M. Breuer, and M. Döllinger. A 2d finite-element scheme for fluid-solid-acoustic interactions and its application to human phonation. *Comput. Methods Appl. Mech. Engrg.*, 198(41–44):3321–3334, 2009.
- [36] H. Matsuzaki and K. Motoki. Study of acoustic characteristics of vocal tract with nasal cavity during phonation of japanese /a/. *Acoust. Sci. & Tech.*, 28(2):124–127, 2007.
- [37] H. Matsuzaki, A. Serrurier, P. Badin, and K. Motoki. Time-domain FEM simulation of japanese and french vowel /a/ with nasal coupling. In *2008 Spring Meeting of the Acoustical Society of Japan*, pages 331–332, Japan, March 2008.
- [38] R. McGowan and M. Howe. Compact green’s functions extend the acoustic theory of speech production. *J. Phonetics*, 35:259–270, 2007.

- [39] P. Mermelstein. Articulatory model for the study of speech production. *J. Acoust. Soc. Am.*, 53(4):1070–1082, 1973.
- [40] V. Mitra, Y. Özbek, H. Nam, X. Zhou, and C. Espy-Wilson. From acoustics to vocal tract time functions. In *Proc. of ICASSP'09*, pages 4497–4500, Taipei, Taiwan, 2009.
- [41] K. Motoki. Three-dimensional acoustic field in vocal-tract. *Acoust. Sci. & Tech.*, 23(4):207–212, 2002.
- [42] J. Mullen, D. Howard, and D. Murphy. Real-time dynamic articulations in the 2-d waveguide mesh vocal tract model. *IEEE Trans. Audio, Speech, Lang. Process.*, 15(2):577–585, 2007.
- [43] K. Munhall, E. Vatikiotis-Bateson, and Y. Tohkura. X-ray film database for speech research. *J. Acoust. Soc. Am.*, 98(2):1222–1224, 1995.
- [44] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd. An approach to real-time magnetic resonance imaging. *J. Acoust. Soc. Am.*, 115(4):1771–1776, 2004.
- [45] D. Recasens and A. Espinosa. Dispersion and variability of catalan vowels. *Speech Commun.*, 48:645–666, 2006.
- [46] K. Richmond. *Estimating Articulatory Parameters from the Acoustic Speech Signal*. PhD thesis, The Centre for Speech Technology Research, Edinburgh University, 2002.
- [47] R. Ridouane. Investigating speech production: A review of some techniques. *Laboratoire de Phonétique et Phonologie*, [Online Document] Available at http://lpp.univ-paris3.fr/equipe/rachid_ridouane/Ridouane_Investigating.pdf, 2006.
- [48] A. Rosenberg. Effect of glottal pulse shape on the quality of natural vowels. *J. Acoust. Soc. Am.*, 49(2):583–590, 1971.
- [49] A. Serrurier and B. Badin. A three-dimensional articulatory model of the velum and nasopharyngeal wall based on mri and ct data. *J. Acoust. Soc. Am.*, 123(4):2335–2355, 2008.
- [50] M. M. Sondhi and J. Schroeter. A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Trans. Audio, Speech, Lang. Process.*, 35(7):955–967, 1987.
- [51] J. Stark, C. Ericsson, P. Branderud, P. Branderurd, J. Sundberg, and J. L. H. Lundberg. The apex model as a tool in the specification of speakerspecific articulatory behavior. In *Proc. from the XIVth ICPHS*, pages 1–7, San Francisco, USA, 1999.
- [52] I. A. F. I. H.-B. P. S. Synthesis. Zhen-hua ling and k. richmond and j. yamagishi and ren-hua wang. *IEEE Trans. Audio, Speech, Lang. Process.*, 17(6):1171–1185, 2009.

- [53] T. Vampola, J. Horacek, and J. Svec. Fe modeling of human vocal tract acoustics. part i: Production of czech vowels. *Acta Acustica united with Acustica*, 94(5):433–447, 2008.
- [54] T. Vampola, J. Horacek, J. Vokral, and L. Cerny. Fe modeling of human vocal tract acoustics. part ii: Influence of velopharyngeal insufficiency on phonation of vowels. *Acta Acustica united with Acustica*, 94(5):448–460, 2008.
- [55] W. Wall, A. Gerstenberger, P. Gammnitzer, C. Förster, and E. Ramm. *Large deformation fluid-structure interaction - advances in ALE methods and new fixed grids approaches*, in: *H.J. Bungartz, M. Schäfer (Eds.), Fluid-Structure Interaction: Modeling, Simulation, Optimization, LNCSE*. Springer, 2006.
- [56] W. Zhao, C. Zhang, S. Frankel, and L. Mongeau. Computational aeroacoustics of phonation, part i: Computational methods and sound generation mechanisms. *J. Acoust. Soc. Am.*, 112(5):2134–2146, 2002.
- [57] W. Zhao, C. Zhang, S. Frankel, and L. Mongeau. Computational aeroacoustics of phonation, part ii: Effects of flow parameters and ventricular folds. *J. Acoust. Soc. Am.*, 112(5):2147–2154, 2002.
- [58] X. Zhou, C. Espy-Wilson, S. Boyce, M. Tiede, C. Holland, and A. Choe. A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” american english /r/. *J. Acoust. Soc. Am.*, 123(6):4466–4481, 2008.