# FINITE ELEMENTS IN VOCAL TRACT ACOUSTICS: GENERATION OF VOWELS, DIPHTHONGS AND SIBILANTS

Oriol Guasch[1], Marc Arnela[1], Arnau Pont[2], Joan Baiges[2] and Ramon Codina[2]

[1]GTM - Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull
C/ Quatre Camins 2, 08022 Barcelona, Catalonia
Email: oguasch@salleurl.edu


[2] Universitat Politècnica de Catalunya,
C/ Jordi Girona 1-3, Edifici C1, 08034, Barcelona, Catalonia

## Abstract

Most physics related to voice production takes place in our larynx and in our vocal tract. In this work we will focus on the latter and show its role in the generation of vowels, diphthongs and sibilants. A review will be made of the involved partial differential equations and the finite element methods (FEM) used to solve them. These equations may range from the irreducible wave equation in the case of vowels, to its mixed formulation in an Arbitrary Eulerian-Lagrangian (ALE) framework in the case of diphthongs, or to the incompressible Navier-Stokes equations, which are solved to obtain the acoustic source terms of acoustic analogies in the numerical generation of sibilants. Yet, it is well-known that for mixed problems in general, the standard Galerkin FEM suffers from oscillations which make necessary to resort to some kind of numerical stabilization. The variational multiscale methods (VMM), also often referred to as subgrid scale stabilization (SGS) methods, offer a nice way out to this situation by splitting the problem unknowns into large scales, resolvable by the computational mesh, and small scales whose effects onto the large ones have to be modeled. The additional terms in the variational equations arising from the modeled subscales not only account for stabilization but also offer many other advantages that will be outlined in the present work. As regards the numerical examples, three-dimensional simulations of vowels and diphthongs will be presented, as well as a simulation on sound generated by flow past a sharp edge at the exit of a rectangular duct, which is important for understanding some basic features of sibilant production.

## 1.    Introduction

The production of voice involves many different physical phenomena that can be described by means of partial differential equations (PDEs). For instance, vowels can be generated by solving the linear wave equation for the acoustic pressure inside the vocal tract (VT), once prescribed, at its entrance, a train of glottal pulses generated by the vocal folds. The shape of the VT changes for each vowel pronunciation and its resonances (known as formants in the voice and speech scientific communities) allow one to distinguish one vowel from another. Many numerical simulations (most using the finite element method, FEM) have been performed to date so as to analyze several aspects related to vowel production. These simulations are motivated either for better understanding the underlying physics (e.g., determine the role played by the VT lateral cavities like the piriform sinuses and valleculae, the effects of radiation and VT wall losses, etc., see [1, 2, 3, 4, 5] among others), or for medical applications as well [6, 7].
    As we move to diphthongs it becomes necessary to deal with time depending VTs that evolve from

an initial vowel to a final one. The wave equation in irreducible form used for vowels becomes no longer useful and one has to work with the wave equation in mixed form instead. The reason for that is the possibility to express the mixed wave equation for the acoustic pressure and acoustic particle velocity in an arbitrary Eulerian-Lagrangian (ALE) frame of reference that moves with the VT, see [8]. To date, little work has been done as regards diphthong sound generation in dimensions higher than one (see e.g., [8, 9, 10] for some exceptions) and numerical efforts are currently being placed on using realistic MRI geometries, as well as on attempting an articulatory driving of the VT motion.

When it comes to sibilant production (e.g., /s/) further PDEs are to be solved. The basic generation mechanism is well understood and attributed to the diffraction by the teeth of the aerodynamic sound generated by the turbulent glottal airflow passing the gap between incisors [11]. Low Mach number computational aeroacoustics (CAA) strategies can be applied to simulate that process and shed light into the relative influence of the upper/lower incisors and lips in the generated sound [12, 13, 14]. The CAA strategy usually involves first solving the incompressible Navier-Stokes equations to extract the acoustic source term for an acoustic analogy, being Lighthill's the standard one [15].

In this paper an outline will be made of the various numerical difficulties encountered when solving vowels, diphthongs and sibilants. In particular, it will be exposed why the standard Galerkin FEM fails when applied to mixed problems, like the mixed wave equation or the Navier-Stokes equations. The spurious oscillations and/or blown up of the Galerkin FEM solutions can however be fixed by resorting to numerical stabilization strategies like the variational multiscale method (VMM) (see [16, 17]). Some examples of numerical simulations will be presented that show the good performance of such methods.

## 2.    Vowels: the wave equation in irreducible and mixed form

### 2.1    The irreducible wave equation for the acoustic pressure

A vowel can be generated to a good degree of accuracy by directly solving the wave equation in a computational domain $\Omega$, which comprises the vocal tract and an outer region where waves exiting the mouth can propagate. Alternatively, one could work in the frequency domain and solve the Helmholtz equation instead. It is quite customary then to emulate radiation losses at the mouth by means of impedance load models corresponding to radiation of a piston set in an infinite plane, or in a sphere. Yet, we will be concerned throughout this work with numerical simulations in the time domain to naturally account for moving domains (e.g., diphthongs) as well as to directly convert the computed acoustic pressure at a given point into an audio file, and listen to the generated sounds.

As for vowel production, the wave equation in $\Omega$ has to be supplemented with appropriate initial and boundary conditions on $\partial\Omega$. Let us split the latter into four non-intersecting regions (see Fig. 1). $\Gamma_\mathrm{G}$ in the figure stands for the cross-section where the vocal folds are located. Besides, $\Gamma_\mathrm{W}$ denotes the vocal tract walls characterized by their impedance, $\Gamma_\mathrm{H}$ stands for the head contour, which can be considered rigid to a good extent, and $\Gamma_\infty$ is a fictitious non-reflecting boundary where the computational domain ends.

The acoustic pressure $p(\boldsymbol{x}, t)$ corresponding to vowel generation can then be obtained from the solution of

$$\left(\partial_{tt}^2 - c_0^2 \nabla^2\right) p = 0 \qquad\qquad \text{in } \Omega,\ t > 0, \tag{1a}$$

with boundary and initial conditions

$$\nabla p \cdot \boldsymbol{n} = -\rho_0 \partial_t u_g \qquad\qquad \text{on } \Gamma_\mathrm{G},\ t > 0, \tag{1b}$$
$$\nabla p \cdot \boldsymbol{n} = -\mu/c_0 \partial_t p \qquad\qquad \text{on } \Gamma_\mathrm{W},\ t > 0, \tag{1c}$$
$$\nabla p \cdot \boldsymbol{n} = 0 \qquad\qquad \text{on } \Gamma_\mathrm{H},\ t > 0, \tag{1d}$$
$$\nabla p \cdot \boldsymbol{n} = 1/c_0 \partial_t p \qquad\qquad \text{on } \Gamma_\infty,\ t > 0, \tag{1e}$$
$$p = 0,\ \partial_t p = 0 \qquad\qquad \text{in } \Omega,\ t = 0. \tag{1f}$$

In (1), $c_0$ designates the speed of sound, $\rho_0$ the air density, $u_g(t)$ the train of glottal pulses and $\mu$ is a constant admittance for the losses at the vocal tract walls. $\boldsymbol{n}$ stands for the normal vector pointing outwards $\partial\Omega$. The explicit dependence of variables on space $\boldsymbol{x}$ and time $t$ will be omitted throughout the paper to alleviate notation.

The method of lines to solve equation (1) consists in using a finite element strategy to discretize the weak form of (1) in space, and then using a finite difference scheme for the time discretization of the resulting algebraic matrix system. The weak form is found as usual: we first multiply (1) by a test function $q$, we integrate the resulting expression in $\Omega$ and then, integrating by parts we transfer a nabla operator from
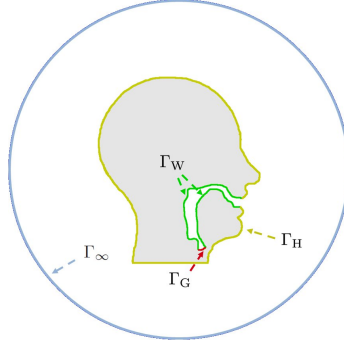
Figure 1 – A sketch of the computational domain $\Omega$ of Eq. (1) in text. $\Gamma_G$ represents the glottal cross-sectional area, $\Gamma_W$ the vocal tract walls, $\Gamma_H$ the human head and $\Gamma_\infty$ a non-reflecting boundary.

the problem unknown to the test function. If we consider, for the ease of exposition, homogeneous Dirichlet conditions on $\partial\Omega$, the Laplacian term in (1) simply results in a term $a(q,p) \equiv \int_\Omega \nabla q \cdot \nabla p d\Omega$ in the problem weak form. It turns out that $a(q,p)$ is a coercive bilinear form (i.e., $\exists \alpha > 0, \ \forall p \in V, \ a(p,p) \geq \alpha \|p\|_V^2$, with $V \equiv H_0^1(\Omega)$ in the present context). For a stationary problem, the Lax-Milgram lemma then guarantees that the problem is well-posed in the Hadamard sense (unique solution bounded by the problem data, see e.g., [18]). Coercivity has the nice property that becomes transferred to the standard Galerkin FEM discretization of $a(q,p)$. That is to say, if the continuous bilinear form is coercive, then the stiffness matrix of the resulting FEM algebraic matrix system is automatically positive definite and a unique bounded discrete solution can be found for it. Therefore, as regards the numerical solution of the wave equation (1), most problems will come from its time discretization rather than from the spatial one.

For vowel production, however, the acoustic pressure is usually required at a location quite close to the mouth exit. This means that waves generated at the glottis have not to travel many wavelengths from where they originate to the point of interest, even for high frequencies (a vocal tract has a typical length of 17 cm). Therefore, time discretization is not that critical and a simple second order finite difference central scheme to approximate the second order time derivative in (1) provides fairly good results. Problems may arise though, when using very detailed MRI vocal tract geometries. If very fine meshes are needed for them, the CFL stability condition will impose very small time steps resulting in lengthy computations. Besides, needless to say that a lumped approximation for the mass matrix is employed, as usual, to avoid full matrix inversion at each time step of the simulation.

Rather than numerical instabilities related to time and space discretizations, probably most difficulties regarding vowel generation stem from the need to resort to Perfectly Matched Layers (PMLs) to avoid spurious reflections at the boundary of the computational domain $\partial\Omega$. This is so because, numerically, the Sommerfeld condition (1e) only performs well for waves impinging in the direction normal to the boundary. Implementing a PML for the irreducible wave equation is rather intricate and involves the introduction of several auxiliary variables to the problem [4]. Another non-straightforward topic for vowel production is that of implementing frequency dependent wall impedances in time domain simulations instead of the constant value used in (1c).

## 2.2 The mixed wave equation for the acoustic pressure and acoustic particle velocity

Instead of solving the wave equation (1) to generate a vowel, one could have also directly resort to the linearized continuity and momentum equations that describe sound propagation in a quiescent fluid, namely

$$\frac{1}{\rho_0 c_0^2}\partial_t p + \nabla \cdot \boldsymbol{u} = 0, \tag{2a}$$

$$\rho_0 \partial_t \boldsymbol{u} + \nabla p = \boldsymbol{0}, \tag{2b}$$

where $p(\boldsymbol{x},t)$ stands for the acoustic pressure, $\boldsymbol{u}(\boldsymbol{x},t)$ for the acoustic particle velocity and $\rho_0$ for the air mean density. Equation (2b) is to be supplemented with appropriate initial and boundary conditions not detailed here due to space restrictions. Note that (1) is obtained by subtracting the divergence of (2b) from $\rho_0$ times the time derivative of (2a). Dealing with (2) may be advantageous for the computation of impedances and/or of the acoustic intensity inside the vocal tract, given that the acoustic particle velocity becomes a problem unknown, rather than a quantity to be derived from the computed FEM

approximation to the acoustic pressure, as in (1). Yet, the plenty justification for (2) manifests when simulating diphthongs and equation (2) needs to be expressed in a frame of reference that follows the mesh movement when transitioning from one vowel to another (see next subsection). It is however interesting to have a glance at the numerical difficulties encountered when trying to solve (2). Those will also manifest in the PDEs of forthcoming sections.

Equation (2) is a particular example of the so called mixed problems. The well-posedness of the weak form of stationary mixed problems relies on the fulfillment of an appropriate inf-sup (or LBB) compatibility condition. However, fulfillment of this condition does not actually translate from the continuum to the discrete level, which causes typical spurious oscillations when solving mixed problems with the standard Galerkin FEM method. In other words, even if the continuum problem is well posed, the stiffness matrix of its corresponding Galerkin FEM matrix system does not necessarily fulfill the conditions (null kernel and full range) that guarantee existence of a unique solution. As we shall see, this will require some type of strategy to avoid instabilities in the numerical solution of mixed formulations.

In the case of time-depending problems as (2), well-posedness is stated by means of the Hille-Yosida theorem (see e.g., [18]). Let us have a closer inspection at how it works for the present case, in order to better understand the ultimate reason for the failure of the Galerkin FEM approach. We will closely follow [19] for that purpose.

Equation (2) can be written in matrix form as

$$\boldsymbol{\mu}\partial_t \boldsymbol{U} + \mathcal{A}\boldsymbol{U} = \boldsymbol{0}, \tag{3}$$

where we have made the following identifications

$$\boldsymbol{U} \equiv \begin{bmatrix} p \\ \boldsymbol{u} \end{bmatrix}, \quad \boldsymbol{\mu} \equiv \begin{bmatrix} (\rho_0 c_0^2)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \rho_0 \boldsymbol{I} \end{bmatrix}, \quad \mathcal{A} \equiv \begin{bmatrix} 0 & \mathrm{div} \\ \nabla & \boldsymbol{0} \end{bmatrix}. \tag{4}$$

Note that $\mathcal{A}$ can also be written as $\mathcal{A} \equiv \boldsymbol{A}_i \partial_i$ where the elements $A_i^{mn}$ of matrices $\boldsymbol{A}_i$ have values $A_i^{i\,i+1} = A_i^{i+1\,i} = 1$ and zero otherwise. Next, assume homogeneous Dirichlet boundary conditions ($p = 0$ on $\Gamma_p$, $\boldsymbol{n} \cdot \boldsymbol{u} = 0$ on $\Gamma_{\boldsymbol{u}}$, $\partial\Omega = \Gamma_p \cup \Gamma_{\boldsymbol{u}}$, $\Gamma_p \cap \Gamma_{\boldsymbol{u}} = \emptyset$) again for the ease of exposition, as well as initial boundary conditions $\boldsymbol{U}(\boldsymbol{x}, 0) = \boldsymbol{U}_0 = [p_0, \boldsymbol{u}_0]^\top$ in $\Omega$. Let us introduce the functional spaces $V_p = \{q \in H^1(\Omega) | q = 0 \text{ on } \Gamma_p\}$, $\boldsymbol{V_u} = \{\boldsymbol{v} \in H(\mathrm{div}, \Omega) | \boldsymbol{n} \cdot \boldsymbol{v} = 0 \text{ on } \Gamma_{\boldsymbol{u}}\}$ as well as the product spaces $V = V_p \times \boldsymbol{V_u}$ and $L = L^2(\Omega) \times \boldsymbol{L}^2(\Omega)$. The weak formulation of problem (2) in a time interval $[0, T]$ can then be posed as that of finding $\boldsymbol{U} \in C^0([0, T]; V) \cap C^1([0, T]; L)$ such that

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}, \boldsymbol{V})_L + (\mathcal{A}\boldsymbol{U}, \boldsymbol{V})_L = 0 \quad \forall \boldsymbol{V} = [q, \boldsymbol{v}]^\top \in L, \tag{5a}$$

$$(\boldsymbol{U}(0), \boldsymbol{V})_L = (\boldsymbol{U}_0, \boldsymbol{V})_L = 0 \quad \forall \boldsymbol{V} \in L, \tag{5b}$$

where the inner product in $L$ is defined as

$$(\boldsymbol{U}, \boldsymbol{V})_L := \int_\Omega \boldsymbol{U}^\top \boldsymbol{V} \, d\Omega = \int_\Omega qp \, d\Omega + \int_\Omega \boldsymbol{u} \cdot \boldsymbol{v} \, d\Omega. \tag{6}$$

The Hille-Yosida theorem guarantees that if $\mathcal{A}$ is monotone and maximal (which can be proved for the present case) then (5) is well posed and has a unique solution bounded by the initial conditions (we are not considering external forces in this example). The following bounds follow

$$\sup_{t\in[0,T]} \left\{ \frac{1}{\rho_0 c_0} \|p\|^2 + \rho_0 c_0 \|\boldsymbol{u}\|^2 \right\}^{1/2} \leq C \left\{ \frac{1}{\rho_0 c_0} \|p_0\|^2 + \rho_0 c_0 \|\boldsymbol{u}_0\|^2 \right\}^{1/2} \tag{7a}$$

$$\sup_{t\in[0,T]} \left\{ \frac{1}{\rho_0 c_0} \left[ \|p\|^2 + \|\nabla p\|^2 \right] + \rho_0 c_0 \left[ \|\boldsymbol{u}\|^2 + \|\nabla \cdot \boldsymbol{u}\|^2 \right] \right\}^{1/2}$$

$$\leq C \left\{ \frac{1}{\rho_0 c_0} \left[ \|p_0\|^2 + \|\nabla p_0\|^2 \right] + \rho_0 c_0 \left[ \|\boldsymbol{u}_0\|^2 + \|\nabla \cdot \boldsymbol{u}_0\|^2 \right] \right\}^{1/2}, \tag{7b}$$

for a positive real constant $C$ and with $\|\boldsymbol{f}(\boldsymbol{x}, t)\|^2 \equiv \int_\Omega \boldsymbol{f}^2(\boldsymbol{x}, t) \, d\Omega$. Note that (7a) provides control on the pressure and the velocity and prevents that these variables could attempt an infinite value in $\Omega$ at a
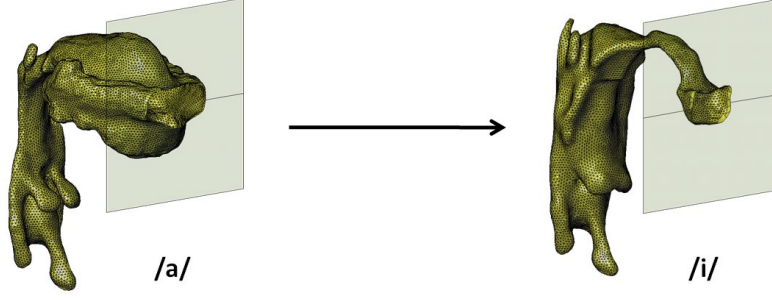
Figure 2 – The generation of diphthong /ai/ involves solving the wave equation in mixed form in a vocal tract that transitions from /a/ to /i/ (geometries adapted from [20]).

given time instant. The same will occur for the time derivatives of these quantities because an analogous bound for $\partial_t \boldsymbol{U}$ to that for $\boldsymbol{U}$ in (7a) can be derived. The bound (7a) can be proved by testing the weak form (5) with $\boldsymbol{V} = \boldsymbol{U}$, using the monotonicity of $\mathcal{A}$ and integrating in time from $t = 0$ to an arbitrary value $t'$. Further, the second bound (7b) is crucial because, in addition to (7a), it provides control on the gradient of the acoustic pressure and on the divergence of the acoustic particle velocity. It can be proved by testing the weak form (5) with $\boldsymbol{V} = \text{diag}[\rho_0 c_0, (\rho_0 c_0)^{-1}\boldsymbol{I}]\mathcal{A}\boldsymbol{U}$, using the maximality of $\mathcal{A}$ and the bound (7a) for $\boldsymbol{U}$, together with its analogous for $\partial_t \boldsymbol{U}$. As we shall next see, the bound (7b) is what precisely fails when trying to find a numerical solution to (5) by means of the standard Galerkin FEM.

The conformal Galerkin FEM approach to solve (5) consists in finding a discrete finite dimensional space $V_h \subset V$ where to interpolate the acoustic pressure and velocity fields. The problem reads: find $\boldsymbol{U}_h = [p_h, \boldsymbol{u}_h]^\top \in C^1([0, t]; V_h)$ such that

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}_h)_L + (\mathcal{A}\boldsymbol{U}_h, \boldsymbol{V}_h)_L = 0 \quad \forall \, \boldsymbol{V}_h = [q_h, \boldsymbol{v}_h]^\top \in V_h, \tag{8a}$$

$$(\boldsymbol{U}_h(0), \boldsymbol{V}_h)_L = (\boldsymbol{U}_{0h}, \boldsymbol{V})_L = 0 \quad \forall \, \boldsymbol{V}_h \in V_h. \tag{8b}$$

The reason why the Galerkin FEM cannot provide an accurate solution to the problem can now be made apparent. It is possible to obtain an analogous bound to (7a) for the discrete Galerkin solution $\boldsymbol{U}_h$ by testing (8a) with $\boldsymbol{V}_h = \boldsymbol{U}_h$ and therefore control the $L_2$-norm of the approximated acoustic pressure and acoustic velocity. However, we have no means to prove a bound analogous to (7b) for the pressure gradient and the velocity divergence, given that we cannot test (8a) with $\boldsymbol{V}_h = \mathcal{A}\boldsymbol{U}_h$. This is so because $\boldsymbol{U}_h \in V_h$ but $\mathcal{A}\boldsymbol{U}_h \notin V_h$. There is thus no guarantee that the term $\int_\Omega \boldsymbol{U}_h^\top \mathcal{A}\boldsymbol{U}_h$ will be bounded by the problem initial conditions. Note that just bounding the $L_2$-norm of the interpolated solution does not prevent this solution from oscillations. Actually this is what one gets in practice: spurious oscillations appear in the numerical acoustic pressure and acoustic particle velocity which make them totally unrealistic.

At this point two possibilities emerge. Either one designs a specific finite element with tailored interpolation spaces for the pressure and the velocity such that $\int_\Omega \boldsymbol{U}_h^\top \mathcal{A}\boldsymbol{U}_h$ can be bounded, or, alternatively, one resorts to strategies which enlarge the Galerkin discrete weak form with some additional stabilization terms. These terms are to be such that an analogous bound to (7b) can be proved for the new stabilized weak form. The Variational MultiScale (VMS) finite element method to be outlined in Section 5 proposes a general framework to allow one deriving these stabilization terms for any partial differential equation.

## 3. Diphthongs: the mixed wave equation in an ALE frame of reference

In the preceding section it has been shown that vowels can be produced either by resorting to the irreducible wave equation for the acoustic pressure, or by means of the wave equation in mixed form for the acoustic pressure and particle velocity. In view of the numerical difficulties associated with the later, one may prefer to deal with the former. However, as soon as we get to diphthong generation one has to deal with the mixed wave equation. This is so because diphthongs involve solving the wave equation in a moving domain that changes from the vocal tract geometry of say, an /a/, to that of an /i/, for the particular case of diphthong /ai/ (see Fig. 2).

If the vocal tract is moving/distorting at a velocity $\boldsymbol{u}_d(\boldsymbol{x}, t)$, it becomes convenient to express the mixed wave equation (2) in a framework moving with the computational domain. A quasi-Eulerian ALE formulation (see e.g., [21] and references therein) can be used for that purpose which accounts for

replacing the time derivative of any fluid property, say $f$, by

$$\partial_t f \leftarrow \partial_t f - \boldsymbol{u}_d \cdot \nabla f. \tag{9}$$

Given that when the equations become discretized $\boldsymbol{u}_d(\boldsymbol{x}, t)$ corresponds to the interpolation from the node mesh velocities, $\boldsymbol{u}_d(\boldsymbol{x}, t)$ can directly be referred to as the mesh velocity. Inserting (9) into (2) results in

$$\frac{1}{\rho_0 c_0^2} \partial_t p - \frac{1}{\rho_0 c_0^2} \boldsymbol{u}_d \cdot \nabla p + \nabla \cdot \boldsymbol{u} = 0, \tag{10a}$$

$$\rho_0 \partial_t \boldsymbol{u} - \rho_0 \boldsymbol{u}_d \cdot \nabla \boldsymbol{u} + \nabla p = \boldsymbol{0}, \tag{10b}$$

which is a convective version of (2) driven by the mesh velocity. There is no straightforward way to combine (10a) and (10b) to obtain an irreducible equation for the acoustic pressure. Therefore, one is forced to solve the mixed problem (10) to produce diphthong sounds.

To the best of the authors knowledge and quite amazingly, there is no existent proof for the well-posedness of the weak formulation of problem (10). Equation (10) closely resembles that of wave propagation in shallow waters and to date only very partial results have been proved for it. In particular it has been shown that the continuous linearized two-dimensional shallow water equations are well posed only for some specific boundary conditions in a very simple rectangular geometry [22]. Despite of this remarkable fact, it is quite reasonable to expect that (10) will present very similar numerical problems to those of the mixed wave equation (2), further reinforced by the presence of the mesh velocity convective terms. Resorting to stabilization strategies will then become a must.

On the other hand, the generation of diphthongs also involves further numerical difficulties due to the computational mesh movement. If the geometries of the vocal tract are not very complex (e.g., simplified vocal tracts of circular cross section) the mesh movement can be accounted for by solving a Laplacian equation for the node displacements. However, if one deals with vocal tracts from MRI data remeshing strategies become necessary. This is also the case if one wishes to solve outward wave propagation due to the mouth exit distortion. In that regard, it should be mentioned that implementing a PML to avoid reflections at the boundary of the computational domain becomes much easier for the wave equation in mixed form than for its irreducible counterpart.

## 4. Sibilants: the Navier-Stokes equations and the acoustic analogies

To generate a sibilant sound like /s/, the anterior portion of the tongue closest to the lips approaches the hard palate creating a small constriction which drives a jet airflow through it. This turbulent jet impinges on the upper incisors and it is deflected downwards to pass through the gap between the tips of the upper and lower incisors. According to [11], it is precisely the diffraction of the turbulent jet pressure aerodynamic sound by the sharp edges of the upper and lower incisors which generate most of the fricative sound. Yet, many research is still needed to determine the relative values of such incisor contributions, as well as the influence of the lips in the production of sibilants. All in all it turns that the basic mechanism of sibilant production relies on aeroacoustics. Given that the flow coming from the lungs, inducing the vocal folds self-oscillations, and finally emanating from the mouth does so at a very low speed, low Mach number computational aeroacoustics (CAA) strategies can be applied to reproduce the phenomenon.

The most common approach for low Mach number CAA is a hybrid procedure which consists in first solving the incompressible Navier-Stokes equations

$$\partial_t \boldsymbol{u}^0 - \nu \Delta \boldsymbol{u}^0 + \boldsymbol{u}^0 \cdot \nabla \boldsymbol{u}^0 + \nabla p^0 = \boldsymbol{f}, \tag{11a}$$

$$\nabla \cdot \boldsymbol{u}^0 = 0, \tag{11b}$$

with $\boldsymbol{u}^0(\boldsymbol{x}, t)$ and $p^0(\boldsymbol{x}, t)$ respectively standing for the incompressible flow velocity and pressure, $\nu$ for the kinematic flow viscosity and $\boldsymbol{f}$ for an external force. In the second step of the approach a wave operator is solved using a source term built from the outputs $\boldsymbol{u}^0$ and/or $p^0$ of the first step. Typically, the celebrated Lighthill acoustic analogy is applied and one obtains the acoustic pressure from the irreducible wave equation (1), with $\rho_0 \partial_i \partial_j (u_i^0 u_j^0)$ as the source term. Alternatively, one could also attempt a mixed formulation, namely,

$$\frac{1}{\rho_0 c_0^2} \frac{\partial p}{\partial t} + \nabla \cdot \boldsymbol{u} = 0, \tag{12a}$$

$$\rho_0 \frac{\partial \boldsymbol{u}}{\partial t} + \nabla p = -\rho_0 \frac{\partial (u_i^0 u_j^0)}{\partial x_j}. \tag{12b}$$

It should be remarked that Lighthill's analogy in irreducible or mixed form (12) is valid as long as one is interested in the sibilant sound recorded at a position well-located outside the mouth, in a quiescent position not influenced by the expelled flow. If one was interested in computing the acoustic field inside the vocal tract, or at a point close to the mouth exit, then one should better solve instead the acoustic perturbation equations.

In Section 2.2 we have already commented on the numerical difficulties associated to the Galerkin FEM for (12), so let us next briefly focus on the Navier-Stokes equations (11). The well-posedness of (11), or of its weak formulation counterpart, remains today an open problem the importance of which has been reflected by its inclusion as one of the Millennium Prize Problems stated by the Clay Mathematics Institute. To date, it has been only proved that a weak solution to the equations exist for finite time intervals, and that a classical solution for the differential problem also exists, yet for a time interval which is inversely proportional to the flow Reynolds number (this makes the result meaningless for most flow situations of practical interest).

In his classical proof for the existence of weak solutions Leray was able to prove the following energy bound (now known as Leray's inequality),

$$\frac{1}{2}\left\|\boldsymbol{u}^0\right\|^2 + \nu \int_0^t \left\|\nabla \boldsymbol{u}^0\right\|^2 ds \leq \frac{1}{2}\left\|\boldsymbol{u}_0^0\right\|^2 + \int_0^t \int_\Omega \boldsymbol{f} \cdot \boldsymbol{u}^0 d\Omega ds. \tag{13}$$

An analogous bound to (13) in terms of the initial problem conditions and external forces can be derived for the Galerkin FEM approach to the problem. However, one can devise why such a bound will be unsatisfactory for the numerical solution. First, the bound provides no control on the incompressible pressure. Second, the control on the velocity gradient becomes irrelevant for small viscosity values (which would correspond to high Reynolds numbers had been the equations adimensionalized). As commented in Section 2.2, only bounding the $L_2$-norm of the velocity (incompressible in this case) does not prevent spurious oscillations to appear. Therefore, it is also clear that some type of stabilization strategy becomes imperative for the numerical solution of the Navier-Stokes equations.

## 5. Subgrid scale stabilized finite elements

The numerical difficulties reported for the wave equation in mixed form (12), or for the Navier-Stokes equations (11), can be addressed in two different ways. As mentioned at the end of section 2, one option is that of designing specific finite elements that fulfill the particular discrete LBB-like conditions of the original problem. This usually requires using different interpolation spaces for the variables at hand (e.g., acoustic/incompressible pressure and velocity), which can make the code implementation costly. Another option is that of adding some stabilization terms to the Galerkin FEM variational form. The additional terms should make possible to find appropriate bounds for the stabilized weak form, in terms of the problem initial data and external forces.

A general framework where to set the stabilization problem is the *variational multiscale method* (VMM) which was originally proposed in [16, 17] (also known as the *subgrid scale* method, or, for some particular choices of the subgrid scales, as the *residual-based stabilized* method). The key idea of the approach is simple. The spurious oscillations encountered in the numerical solution of an equation must come from those small scales (*subscales* or *subgrid scales*) that we disregard in the discretization process. Therefore, the effects of the subscales onto the computed scales should be taken into account to improve the solution.

Let us see how the method proceeds for the weak formulation of the mixed wave equation (12). Again, we will closely follow [19]. We start by decomposing the exact solution of (12), $\boldsymbol{U}$, into a finite element component $\boldsymbol{U}_h$ that can be resolved by the computational mesh, plus a subscale $\boldsymbol{U}'$, which will have to be modelled. Substitution of $\boldsymbol{U} = \boldsymbol{U}_h + \boldsymbol{U}'$ into (12) yields two equations. The first one governs the dynamics of the large scales. Considering that the subscales vanish at the interelemental boundaries, that they do not change with time (quasi-static assumption) and integrating by parts to transfer the spatial derivatives to the test function, we get

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}_h) + (\boldsymbol{A}_i \partial_i \boldsymbol{U}_h, \boldsymbol{V}_h) - (\boldsymbol{U}', \boldsymbol{A}_i^\top \partial_i \boldsymbol{V}_h) = (\boldsymbol{F}, \boldsymbol{V}_h). \tag{14}$$

Besides, the second equation, governing the behavior of the static subscales, is given by

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}') + (\boldsymbol{A}_i \partial_i \boldsymbol{U}_h, \boldsymbol{V}') + (\boldsymbol{A}_i \partial_i \boldsymbol{U}', \boldsymbol{V}') = (\boldsymbol{F}, \boldsymbol{V}'), \tag{15}$$
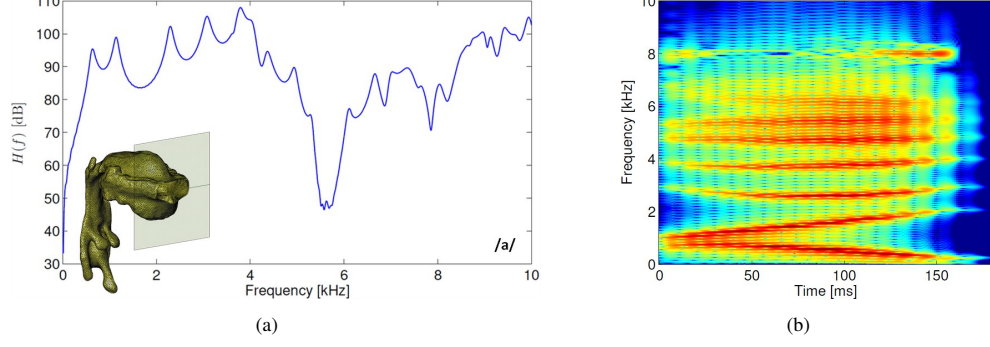
Figure 3 – (a) Vocal tract transfer function for vowel /a/. (b) Spectrogram for diphthong /ai/.

and corresponds to the $L^2$-projection of $\boldsymbol{\mu}\partial_t \boldsymbol{U}_h + \boldsymbol{A}_i\partial_i \boldsymbol{U}_h + \boldsymbol{A}_i\partial_i \boldsymbol{U}' = \boldsymbol{F}$ onto the subscale space. Let $\mathcal{P}'$ denote this projection and rewrite (15) as

$$\mathcal{P}'(\boldsymbol{A}_i\partial_i \boldsymbol{U}') = \mathcal{P}'[\boldsymbol{F} - (\partial_t \boldsymbol{U}_h + \boldsymbol{A}_i\partial_i \boldsymbol{U}_h)] =: \boldsymbol{R}_h, \tag{16}$$

where we have introduced the residual $\boldsymbol{R}_h$ in the last equality. $\mathcal{P}'$ is taken as the identity in the algebraic subgridscale method ASGS, or as $\mathcal{P}' = \boldsymbol{I} - \Pi_h$ in the orthogonal subgrid scale method OSS [23] ($\Pi_h$ stands for the projection onto the finite element space). The solution to (16) is unknown and has to be somehow modelled. A reasonable option consists in taking $\mathcal{P}'(\boldsymbol{A}_i\partial_i \boldsymbol{U}') \approx \boldsymbol{\tau}^{-1}\boldsymbol{U}'$ so that $\boldsymbol{U}' = \boldsymbol{\tau}\boldsymbol{R}_h$ (see e.g., [23, 24, 19]). $\boldsymbol{\tau} \equiv \mathrm{diag}(\tau_p, \tau_{\boldsymbol{u}})$ stands for a matrix of stabilization parameters to be determined, which has been chosen diagonal for simplicity. The values for $\tau_p$ and $\tau_{\boldsymbol{u}}$ can be found from an analysis of the subscale equation in the wavenumber domain, which ensures that despite of not having an exact solution for $\boldsymbol{U}'$, the amount of kinetic energy $\boldsymbol{U}'$ transfers to the large scales is correct. For the present problem it turns that $\tau_p = C\rho_0 c_0 h$ and $\tau_{\boldsymbol{u}} = C(\rho_0 c_0)^{-1}h$, $C$ being a constant to be determined from numerical experiments.

Substituting $\boldsymbol{U}' = \boldsymbol{\tau}\boldsymbol{R}_h$ into the large scale equation (14) results in

$$(\boldsymbol{\mu}\partial_t \boldsymbol{U}_h, \boldsymbol{V}_h) + (\boldsymbol{A}_i\partial_i \boldsymbol{U}_h, \boldsymbol{V}_h) + \sum_{e=1}^{n_{el}}(\boldsymbol{A}_i^\top\partial_i \boldsymbol{V}_h, \boldsymbol{\tau}\mathcal{P}[(\partial_t \boldsymbol{U}_h + \boldsymbol{A}_i\partial_i \boldsymbol{U}_h) - \boldsymbol{F}])_{\Omega_e} = (\boldsymbol{F}, \boldsymbol{V}_h). \tag{17}$$

with $n_{el}$ standing for the number of elements in the computational mesh and $\Omega_e$ for the $e$-th element domain. Equation (17) is the stabilized weak form we were looking for. Note that the first two terms in (17) correspond to the standard Galerkin FEM approach in (8a), whereas the third one is the additional stabilization term that facilitates using equal interpolations for the acoustic pressure and acoustic particle velocity. As discussed at the end of Section 2.2, the inability of the Galerkin FEM to solve the wave equation in mixed form was attributed to the fact that it was not possible to prove any bound to control the pressure gradient and the velocity divergence. As opposed, the following bound can be proved for the stabilized weak form (17),

$$\tau_{\boldsymbol{u}} \int_0^t \|\nabla p_h\|^2 \, ds + \tau_{\boldsymbol{p}} \int_0^t \|\nabla \cdot \boldsymbol{u}_h\|^2 \, ds$$
$$\leq C\big(\frac{1}{\rho_0 c_0^2}\|p_{h0}\|^2 + \rho_0\|\boldsymbol{u}_{h0}\|^2 + \tau_{\boldsymbol{u}}\|\nabla p_{h0}\|^2\, t + \tau_{\boldsymbol{p}}\|\nabla \cdot \boldsymbol{u}_{h0}\|^2\, t\big). \tag{18}$$

Therefore, the stabilized formulation can get rid off the spurious oscillations which make the Galerkin FEM useless.

Following the same subgrid scale strategy for the Navier-Stoles equations, one can also prevent the numerical instabilities in that equation (see e.g.,[24]). Actually, not only instabilities can be avoided but it can also be shown that the subgrid scale stabilization terms act, in fact, as an implicit large eddy simulation (LES) model satisfying the appropriate statistics of turbulent flows [25].

## 6.    Numerical Results

Let us next briefly show some of the results that can be obtained by solving the equations presented in the previous sections. In Fig. 3a we plot a vocal tract transfer function (quotient between the acoustic pressure
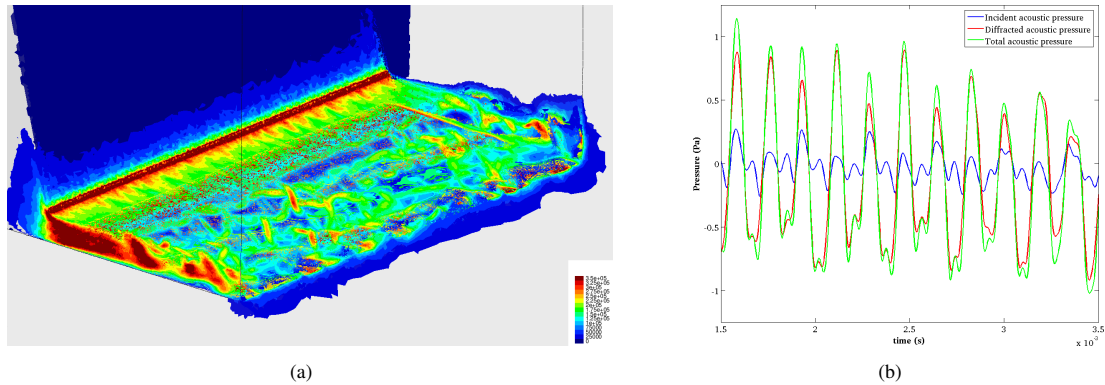
<div style="text-align:center">(a)                      (b)</div>

Figure 4 – (a) Vorticity isosurfaces shedding from the tooth-shaped obstacle. (b) Direct (blue), diffracted (red) and total (green) sound presure contributions at a far field point.

at a point close to the mouth exit and the glottal pulse volume velocity at the glottis) postprocessed from the solution of the irreducible wave equation (1). The VT geometry has been obtained from MRI data [20]. The formants due to plane wave propagation can be appreciated for low frequencies, as well as the large antiresonance due to the lateral cavities between $5 - 6$ kHz, and the high-order modes above that range. In Fig. 3b a spectrogram showing the evolution of formants from /a/ to /i/ for diphthong /ai/ is presented, which has been obtained from the solution of the mixed equation (2).

In Fig. 4 we present some aerodynamic and aeroacoustics results according to the methodology described in Section 4, for a configuration frequently used to analyze the generation of sibilants. That consists of a rectangular duct with an input flow at the entrance and a tooth-shaped obstacle placed close to the exit. Turbulent vortex shedding occurs past the obstacle resulting in the generation of aerodynamic sound which is diffracted by the tooth. A snapshot of the isovorticty surfaces past the obstacle is shown in Fig. 4a, while the direct, diffracted and total acoustic pressure contributions at a far field point are presented in Fig. 4b. These have been computed following the approach recently presented in [14]. As expected, the contribution from the diffracted component is the dominant one.

## 7. Conclusions

In this work we have made a survey of some of the equations involved in vocal tract acoustics for vowel, diphthong and sibilant generation. The numerical difficulties encountered when approximating them by means of the finite element method have been highlighted, and special emphasis has been put on mixed formulations found when producing diphthongs and sibilants. The reasons for the failure of the standard Galerkin FEM have been indicated as well as the possibility to circumvent them by resorting to subgrid scale stabilized finite element methods.

## 8. Acknowledgements

## References

[1] Motoki, K., "Three-dimensional acoustic field in vocal-tract", *Acoust. Sci. & Tech.* 23(4) (2002), 207–212.

[2] Vampola, T., Horáček, J. and Švec, J. G., "FE Modeling of Human Vocal Tract Acoustics. Part I: Production of Czech vowels", *Acta Acust.* 94(5) (2008), 433–447.

[3] Takemoto, H., Mokhtari, P. and Kitamura, T., "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method", *J. Acoust. Soc. Am.* 128(6) (2010), 3724–3738.

[4] Arnela, M. and Guasch, O., "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method", *J. Acoust. Soc. Am.* 133(6) (2013), 4197–4209.

[5]  Arnela, M., Guasch, O. and Alías, F., "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations", *J. Acoust. Soc. Am.* 134(4) (2013), 2946–2954.

[6]  Švancara, P. and Horáček, J., "Numerical modelling of effect of tonsillectomy on production of czech vowels", *Acta Acust.* 92 (2006), 681–688.

[7]  Vampola, T., Horáček, J., Voǩrál, J. and Černý, L., "FE Modeling of Human Vocal Tract Acoustics. Part II: Influence of velopharyngeal insufficiency on phonation of vowels", *Acta Acust.* 94(5) (2008), 448–460.

[8]  Guasch, O., Arnela, M., Codina, R. and Espinoza, H., "A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production", *Submitted* (2015).

[9]  Mullen, J., Howard, D. M. and Murphy, D. T., "Real-Time Dynamic Articulations in the 2-D Waveguide Mesh Vocal Tract Model", *IEEE Trans. Audio Speech Lang. Process.* 15(2) (2007), 577–585.

[10]  Arnela, M., Guasch, O., Codina, R. and Espinoza, H., "Finite element computation of diphthong sounds using tuned two-dimensional vocal tracts", *Proc. of 7th Forum Acousticum,* Kraków, Poland, 2014.

[11]  Howe, M.S. and McGowan, R.S., "Aeroacoustics of [s]", *Proc. R. Soc. A* 461 (2005), 1005–1028.

[12]  Cisonni, J., Nozaki, K., Hirtum, A. Van, Grandchamp, X. and Wada, S., "Numerical simulation of the influence of the orifice aperture on the flow around a teeth-shaped obstacle", *Fluid Dyn. Res* 45 (2) (2013).

[13]  Yoshinaga, T., Koike, N., Nozaki, K. and Wada, S., "Study on production mechanisms of sibilant /s/ using simplified vocal tract model", *Proc. of Inter-noise'15,* San Francisco, California (USA), 2015.

[14]  Guasch, O., Pont, A., Baiges, J. and Codina, R., "Concurrent finite element simulation of incident and diffracted flow noise in computational aeroacoustics", *Proc. of Inter-noise'15,* San Francisco, California (USA), 2015.

[15]  Lighthill, M. J., "On sound generated aerodynamically. I. General theory", 211(1107) (1952), 564–587.

[16]  Hughes, T.J.R., "Multiscale phenomena: Green's function, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized formulations", *Comput. Methods Appl. Mech. Engrg.* 127 (1995), 387–401.

[17]  Hughes, T.J.R., Feijóo, G., Mazzei, L. and Quincy, J., "The variational multiscale method, a paradigm for computational mechanics", *Comput. Methods Appl. Mech. Engrg.* 166 (1998), 3–24.

[18]  Alexandre, E. and Guermond, J.L., *Theory and practice of finite elements,* vol. 159, Springer Science & Business Media, 2013.

[19]  Codina, R., "Finite element approximation of the hyperbolic wave equation in mixed form", *Comput. Methods Appl. Mech. Engrg.* 197(13–16) (2008), 1305–1322.

[20]  Aalto, D. et al., "Large scale data acquisition of simultaneous {MRI} and speech", *Appl. Acoust.* 83(0) (2014), 64–75.

[21]  Donea, J., Huerta, A., Ponthot, J.-Ph. and Rodríguez-Ferran, A., *Arbitrary Lagrangian Eulerian Methods, in: Encyclopedia of Computational Mechanics,* John Wiley & Sons, Ltd, 2004.

[22]  Huang, A. and Temam, R., "The linearized 2D inviscid shallow water equations in a rectangle: boundary conditions and well-posedness", *Arch. Rational Mech. Anal.* 211 (2014), 1027–1063.

[23]  Codina, R., "Stabilized finite element approximation of transient incompressible flows using orthogonal subscales", *Comput. Methods Appl. Mech. Engrg.* 191 (2002), 4295–4321.

[24]  Codina, R., Principe, J., Guasch, O. and Badia, S., "Time dependent subscsales in the stabilized finite element approximation of incompressible flow problems", *Comput. Methods Appl. Mech. Engrg.* 196(21–24) (2007), 2413–2430.

[25]  Guasch, O. and Codina, R., "Statistical behavior of the orthogonal subgrid scale stabilization terms in the finite element large eddy simulation of turbulent flows", *Comput. Methods Appl. Mech. Engrg.* 261 (2013), 154–166.