METHODS

Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

RÉMI BLANDIN[1], **MARC ARNELA**[2], **SIMON FÉLIX**[3], **JEAN-BAPTISTE DOC**[4], **AND PETER BIRKHOLZ**[1], (Member, IEEE)

[1]Institute of Acoustics and Speech Communication, Technische Universität Dresden (TU Dresden), 01062 Dresden, Germany
[2]Grup de recerca en Tecnologies Mèdia (GTM), La Salle - Universitat Ramon Llull, 08022 Barcelona, Catalonia, Spain
[3]Laboratoire d'Acoustique de l'Université du Mans (LAUM), UMR 6613, CNRS, Institut d'Acoustique—Graduate School (IA-GS), Le Mans Université, 72085 Le Mans, France
[4]Laboratoire de Mécanique des Structures et des Systèmes Couplés, Conservatoire National des Arts et Métiers, 75003 Paris, France

Corresponding author: Rémi Blandin (remi.blandin@tu-dresden.de)

ABSTRACT Acoustic simulation of sound propagation inside the vocal tract is a key element of speech research, especially for articulatory synthesis, which allows one to relate the physics of speech production to other fields of speech science, such as speech perception. Usual methods, such as the transmission line method, have a very low computational cost and perform relatively good up to 4-5 kHz, but are not satisfying above. Fully numerical 3D methods such as finite elements achieve the best accuracy, but have a very high computational cost. Better performances are achieved with the state of the art semi-analytical methods, but they cannot describe the vocal tract geometry as accurately as fully numerical methods (e.g. no possibility to take into account the curvature). This work proposes a new semi-analytical method that achieves a better description of the three-dimensional vocal-tract geometry while keeping the computational cost substantially lower than the fully numerical methods. It is a multimodal method which relies on two-dimensional finite elements to compute transverse modes and takes into account the curvature and the variations of cross-sectional area. The comparison with finite element simulations shows that the same degree of accuracy (about 1% of difference in the resonance frequencies) is achieved with a computational cost about 10 times lower.

INDEX TERMS Acoustics, acoustic waves, simulations, waveguide, human voice, speech synthesis, vocal tract.

I. INTRODUCTION

A. GENERAL CONTEXT

Speech sounds are produced by sound sources (vocal fold oscillations and/or turbulent flow) placed inside a waveguide, the vocal tract, defined as the air volume between the vocal folds and the lips. The motion of the articulators, such as the tongue or the jaw, modify the vocal tract shape and the speech sound produced. Resonances are generated inside the vocal tract by the wave propagating forward and backward from

the glottis to the mouth and the mouth to the glottis. They enhance some part of the speech spectrum, usually referred to as formants, which constitute acoustic cues allowing us to differentiate phonemes.

Acoustic simulation of the vocal tract is useful to establish a relationship between the vocal tract shape and the acoustic properties of speech. Thus, it is possible to synthesise speech directly from the vocal tract shape using articulatory synthesis [8], [23], [36], [48], [49], [51].

The most popular method to simulate vocal tract acoustics is the transmission line model (TLM) [49]. It takes advantage of the fact that acoustic waves are guided along the vocal tract,

The associate editor coordinating the review of this manuscript and approving it for publication was Guido Lombardi .

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE *Access*

and that under a low frequency assumption (valid up to about 5 kHz), only plane waves propagate. This means that the local cross-sectional shape has little impact on the propagation and that one needs to consider the cross-sectional area only. Thus, under these hypotheses, the curvature of the vocal tract can be ignored and the description of its shape can be limited to the variation of its cross-sectional area. The TLM can be implemented in both frequency and time domain. It requires very few computational resources and can run in real time.

B. LIMITS OF CURRENT SIMULATION METHODS

A first limitation of the TLM is the impossibility to take into account accurately the three-dimensional (3D) shape of the vocal tract. A second one is that it can intrinsically not describe a non uniform transverse acoustic field. The consequence of this is particularly visible at high frequencies, above about 4-5 kHz, when other modes than plane waves (higher order modes) can propagate [11], [40], [41]. This induces additional resonances and antiresonances in the vocal tract that cannot be predicted by TLM. However, it has also consequences at lower frequencies in the vicinity of disconti- nuities where the transverse acoustic field can be slightly non- uniform. This induces inaccuracies in the prediction of the resonance frequencies. Note that it is theoretically possible to reduce these inaccuracies by applying length corrections to the TLM segments [49].

These limitations motivated the use of 3D simulation methods to study vocal tract acoustics. For this purpose, finite elements (FEM) [2], [24], [55], boundary elements (BEM) [34], finite differences (FDM) [53], waveguide meshes [20], [31] and the multimodal method [11], [42], [56] have been used.

The consequences of each of the two limitations of the TLM have been explored. On one hand, the effect of geo- metrical details such as the piriform fossae,[1] the vallecula,[2] the inter-dental space, the nasal cavity or more generally the precise 3D vocal-tract shape have been observed to affect resonance frequencies and induce additional resonances and anti-resonances [53], [54]. On the other hand, the higher-order modes have been shown to induce addi- tional resonances and anti-resonances at frequencies above 4-5 kHz [11] and affect speech radiation patterns [12], [16]. Note that side branches can be modelled with TLM, however, to our knowledge, no correction has been proposed to take into account transverse modes.

So far, most of the works using 3D acoustics simulations have been limited to small numbers of static geometries. Some of the most recent works are focused on the generation of articulated sounds such as specific diphthongs [3], [5], [18], [31] or consonant-vowel sequences [4], [32].

The main reason for this limitation is that the 3D numerical methods are strongly limited by their high computational cost, which prevent to use them as a substitute for the TLM.

[1]Two small sides cavities located at the base of the vocal tract.
[2]A groove between the base of the tongue and the epiglottis.

To illustrate that, using the FEM in the frequency domain, about 10 hours are required to simulate 1000 frequencies using a standard single processor (without meshing the outer space though) [10]. The BEM may have a lower compu- tational cost because it does not require to mesh the outer space and involves a smaller number of degree of freedom compared to FEM and FDM [35]. However, its application to vocal tract acoustics simulation is very anecdotal [34].

One possibility to reduce this cost is to use analytical or semi-analytical methods which have a lower computational cost, but are limited in term of geometrical accuracy. As for the TLM, it is possible to take advantage of the elongated shape of the vocal tract to simulate the propagation of higher- order modes, in addition to the plane waves simulated with the TLM. This type of method, which simulates several transverse modes instead of just the plane mode is called multimodal method. For example Motoki [42], [43] have taken advantage of the lower computational cost of this type of method to run numerous simulations with small perturba- tions of the geometry. However, the use of rectangular cross- sections, which allowed a fully analytical solution with a very small computational cost, implied a strong geometrical approximation of the vocal tract shape. Blandin *et al.* [11] improved this method by introducing a numerical compu- tation of the transverse modes that substantially increased the realism of the cross-sectional shapes. However, the cur- vature was not taken into account, the cross-sectional area was piecewise constant and no wall losses were simulated. The multimodal method can also help to understand the physical phenomena involved in speech production better than fully numerical methods, as they provide a theoretical background (higher-order modes propagation) that can be related to experimental observations. For example, it was useful to Yoshinaga *et al.* [56] for understanding better the mechanisms of sound generation of the fricative /s/. However, in this case as well, the use of rectangular cross-sections made the geometry used less realistic than what could be achieved with fully numerical methods.
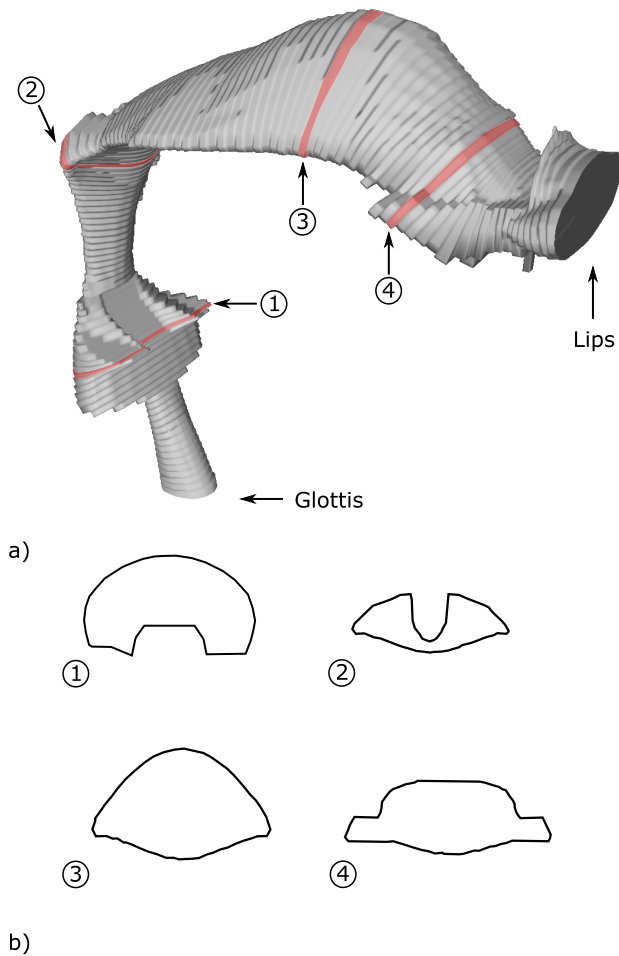
C. PROPOSED APPROACH

The objective of this work is to achieve a better trade-off between geometrical accuracy and computational cost by combining the advantages of numerical methods and the state-of-the-art multimodal method.

For this purpose the vocal tract geometry is sliced in segments in which the cross-sectional shape is considered constant, as illustrated in Fig. 1. The local transverse modes are computed with two-dimensional (2D) FEM, and the prop- agation along the waveguide axis is solved using a multi- modal formulation. The wave problem is solved for a static geometry. Similar semi-analytical finite-elements (SAFE) methods have already been applied to the simulation of street acoustics [47] or vibrations in rails [6], [26], [33], [38]. The multimodal formulation of the propagation along the wavegu- ide axis allows us to take into account the effects of the curvature and cross-section area variations, as detailed below

IEEE Access·

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

FIGURE 1. a) Vocal tract geometry corresponding to the vowel /a/ generated by the articulatory model implemented in the speech synthesizer Vocaltractlab 2.3 [8] sliced into 129 segments, b) examples of cross-sectional shapes of different segments at various locations in the vocal tract geometry. (Figure adapted from [14]).

FIGURE 2. Constant cross-sectional shape waveguide segment represented a) in Cartesian coordinates (X, Y, Z) and b) in the transformed coordinates (x, y, z) removing curvature and cross-sectional area variations (Figure adapted from [14]).
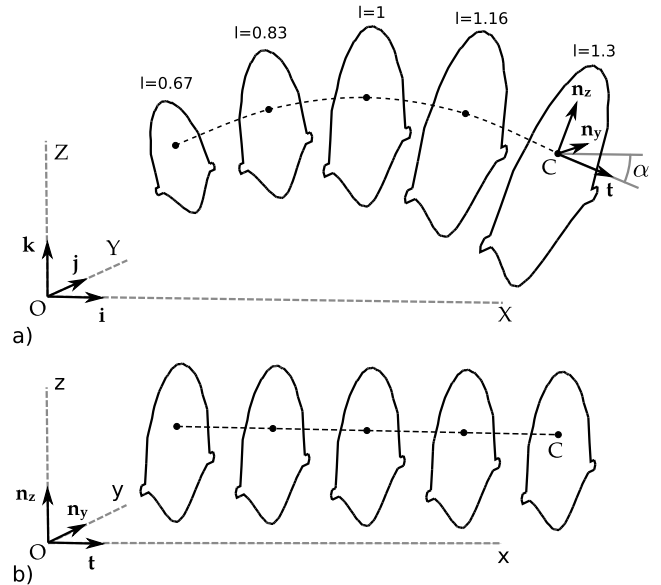
in Sec. II. An implementation of this method is compared with a full FEM solution of the wave equation in a single segment and in three vocal tract geometries corresponding to the vowels /a/, /i/ and /u/. The geometries, the parameters used for the multimodal simulations and the method used for the full FEM simulations are presented in Section III. The outcomes of both methods are compared in Section IV.

II. MULTIMODAL METHOD
Throughout the paper, small bold letters denote vectors, capital bold letters denote matrices and ∂_x is understood as $\frac{\partial}{\partial x}$.

A. SEGMENTATION OF THE 3D GEOMETRY
The 3D vocal tract geometry can be provided as a 3D surface mesh (e.g. from medical image) or by a set of functions defining 3D surfaces (e.g. from an articulatory model). In order to apply the proposed method, they are first segmented in portions with constant cross-sectional shape, as illustrated in Fig. 1. For this purpose, a curve, referred to as the centerline,

passing through the vicinity of the centers of the transverse cross-sections and contained in the sagittal plane is defined [9]. The 3D geometry is cut in several planes perpendicular to the centerline. A segment is defined between two consecutive cuts. It is characterized by a contour extracted from one of the cuts, the local curvature of the centerline and a scaling factor describing the area variation along the centerline (see Fig. 2a). Keeping the contour shape constant along the segment allows one to use the same transverse modes along the segment. Several methods have been proposed to implement this segmentation process [7], [9], [28], [36]. For classical TLM simulation, the number of segments generated for vocal tract geometries ranges between 40 and 80 [50], [52]. For 3D simulations, the comparison of FEM simulations of geometries without segmentation and segmented with 40, 60 and 80 segments showed that the segmentation process induces an upward shift of the resonance frequencies (lower than 5% below 5 kHz for 80 segments), the global shape of the transfer functions being preserved [2].

B. EXPRESSION OF THE WAVE EQUATION INSIDE THE SEGMENTS
1) GEOMETRICAL TRANSFORMATION
Following earlier works on the multimodal formulation in complex-shaped waveguides (2D [39] or 3D with circular cross-section [30]), we define a geometrical transformation from the Cartesian coordinates (X, Y, Z) in which the segments are curved with a varying cross-sectional area (see Fig. 2a), to the coordinates (x, y, z) in which they are straight with a constant cross-sectional area, as illustrated in Fig. 2b. The position m of a point can be described by the coordinates (x, y, z) in a Frenet-Serret frame (t, n_y, n_z), whose origin is a

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE *Access*

point C of the centerline, to which is associated the position vector s. The vector t is tangent to the centerline, and the vectors n_y and n_z are orthogonal to each other and contained in the local transverse plane (y, z). x is the curvilinear abscissa along the centerline curve. A scaling function $l(x)$ is defined that describes the variations of the cross-section dimensions along the centerline.

The position vector m of a point in both coordinate systems can be expressed as

$$m = Xi + Yj + Zk = s(x) + yl(x)n_y + zl(x)n_z. \quad (1)$$

An angle $\alpha(x)$ is defined between t and i (see Fig. 2a). The differentials of the vectors s, n_y and n_z are

$$\frac{ds}{dx} = t, \quad \frac{dn_y}{dx} = 0 \text{ and } \frac{dn_z}{dx} = -\kappa(x)t, \quad (2)$$

where $\kappa(x)$ is the local curvature of the centerline.

In order to apply the proposed transformation to the wave equation, it is necessary to compute its Jacobian matrix J. For this purpose, using the derivatives given in Eq. (2), Eq. (1) is differentiated

$$\begin{aligned} dm &= dXi + dYj + dZk \\ &= \left(ft + yl'n_y + zl'n_z\right) dx + ln_y dy + ln_z dz, \quad (3) \end{aligned}$$

where $f(x) = 1 - z\kappa l$ and $l' = dl/dx$. The vectors t, n_y and n_z are expressed as a function of the vectors i, j, k and the angle $\alpha(x)$, and Eq. (3) is expressed in a matrix format

$$\begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix} = J^{-1} \begin{pmatrix} \partial_X \\ \partial_Y \\ \partial_Z \end{pmatrix}, \quad (4)$$

with the Jacobian matrix

$$J = \frac{\partial(x, y, z)}{\partial(X, Y, Z)} = R_\alpha^{-1} \begin{pmatrix} \dfrac{1}{f} & -\dfrac{yl'}{fl} & -\dfrac{zl'}{fl} \\ 0 & \dfrac{1}{l} & 0 \\ 0 & 0 & \dfrac{1}{l} \end{pmatrix}, \quad (5)$$

where R_α is a rotation matrix of angle α. The determinant of this Jacobian is $\det J = \frac{1}{fl^2}$.

2) WAVE EQUATION
The wave equation in Cartesian coordinates (X, Y, Z) is

$$(\Delta_X + k^2)p = 0, \quad (6)$$

where $\Delta_X p = \partial_X^2 p + \partial_Y^2 p + \partial_Z^2 p$ is the Laplacian operator, p the acoustic pressure, $k = \omega/c$ the wavenumber, ω the angular frequency, c the sound speed and a time factor $e^{j\omega t}$ is understood. The wall boundary condition is expressed in Cartesian coordinates as

$$\nabla_X p \cdot n_X + jk\zeta p = 0, \quad (7)$$

where $\nabla_X p = \left(\partial_X p \ \partial_Y p \ \partial_Z p\right)^t$ is the gradient operator, n_X the outward pointing normal to the boundary and $\zeta = Z_0/Z_w$ is a boundary admittance coefficient related to the air

impedance $Z_0 = \rho c$, with ρ the volumic mass and c the sound speed. The wall impedance Z_w is defined as

$$u \cdot n_X = \frac{p}{Z_w}, \quad (8)$$

where u is the particle velocity. According to Eq. (2.1) in [44], the Laplacian Δ_X can be expressed as a function of the divergence div and the gradient ∇ expressed in the transformed coordinates as

$$\Delta_X p = \det(J)\text{div}(H\nabla p) \quad (9)$$

where

$$H = \frac{J^t J}{\det J} = \frac{1}{f} \begin{pmatrix} l^2 & -yl'l & -zl'l \\ -yl'l & f^2 + (yl')^2 & yzl'^2 \\ -zl'l & yzl'^2 & f^2 + (zl')^2 \end{pmatrix}. \quad (10)$$

Substituting directly this expression in the wave equation Eq. (6) gives the wave equation in transformed coordinates

$$\text{div}\,(H\nabla p) + \frac{k^2}{\det J}p = 0. \quad (11)$$

According to Eq. (2.10) in [44], the outward pointing normal n_X in Cartesian coordinates can be expressed as a function of the outward pointing normal n in the transformed coordinates as

$$n_X = \frac{J}{\det J}n. \quad (12)$$

Knowing that from Eq. (4) the gradient in Cartesian coordinates is related to the gradient in the transformed coordinates as $\nabla_X = J\nabla$, the boundary condition equation Eq. (7) can be written as

$$(J\nabla p) \cdot \left(\frac{J}{\det J}n\right) + jk\zeta p = 0. \quad (13)$$

After rearrangement, it can be rewritten as

$$H\nabla p \cdot n + jk\zeta p = 0, \text{ for } y, z \in \Gamma_W, \quad (14)$$

where Γ_W is the surface constituting the wall boundary of the waveguide. In order to transform the wave equation Eq. (11) in a first order evolution equation along x, a secondary field q is introduced

$$q \equiv (H\nabla p) \cdot t = \frac{l}{f}\left(l\partial_x p - yl'\partial_y p - zl'\partial_z p\right), \quad (15)$$

Identifying its expression in the wave equation, it can be rearranged as

$$\partial_x \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \dfrac{l'}{l}v \cdot \nabla_\perp & \dfrac{f}{l^2} \\ -\text{div}_\perp\left(f\nabla_\perp\right) - f(kl)^2 & \dfrac{l'}{l}\text{div}_\perp(v.) \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}, \quad (16)$$

where the local frame transverse coordinate vector is $v = (y\ z)^t$, the transverse gradient is $\nabla_\perp = \left(\partial_y\ \partial_z\right)^t$ and

the transverse divergence is $\text{div}_\perp \boldsymbol{u} = \partial_y u_y + \partial_z u_z$. Likewise, an expression is obtained for the boundary condition

$$n_x q + f\left(\boldsymbol{\nabla}_\perp p\right) \cdot \boldsymbol{n}_\perp - \frac{l'}{l}\boldsymbol{v} \cdot \boldsymbol{n}_\perp q + jk\zeta p = 0, \quad (17)$$

where n_x is the x component of the outward pointing normal $\boldsymbol{n}$ and $\boldsymbol{n}_\perp$ is the projection of the normal $\boldsymbol{n}$ in the local frame transverse plane (y, z).

3) MODAL FORMULATION
The fields p and q are decomposed into the transverse modes $\varphi_n(y, z)$

$$\begin{cases} p(x, y, z) = \sum\limits_{n=0}^{\infty} p_n(x)\varphi_n(y, z) = \boldsymbol{p} \cdot \boldsymbol{\varphi}\,, \\ q(x, y, z) = \sum\limits_{n=0}^{\infty} q_n(x)\varphi_n(y, z) = \boldsymbol{q} \cdot \boldsymbol{\varphi}\,, \end{cases} \quad (18)$$

where $\boldsymbol{p} \equiv \{p_n(x)\}$, $\boldsymbol{q} \equiv \{q_n(x)\}$ and $\boldsymbol{\varphi} \equiv \{\varphi_n(y, z)\}$. The transverse propagation modes φ_n fulfill the orthogonality relationships

$$\begin{cases} \langle \varphi_m, \varphi_n \rangle = \delta_{mn}\,, \\ \langle \boldsymbol{\nabla}_\perp \varphi_m, \boldsymbol{\nabla}_\perp \varphi_n \rangle = \gamma_m^2 \delta_{mn}\,, \end{cases} \quad (19)$$

where $\langle a, b \rangle \equiv \int_S \bar{a}b\, dS$ and $\langle \boldsymbol{a}, \boldsymbol{b} \rangle \equiv \int_S \bar{\boldsymbol{a}} \cdot \boldsymbol{b}\, dS$ is the scalar product, S is the cross-sectional surface, $\bar{a}$ is the complex conjugate of a and γ_m^2 is the eigenvalue associated to the transverse mode φ_m.

Combining Eqs. (16) and (17) and projecting them onto the transverse propagation modes φ_n yields a set of coupled equations describing the evolution of the vectors $\boldsymbol{p}$ and $\boldsymbol{q}$

$$\partial_x \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix} = \mathbf{M}(x) \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix} = \begin{pmatrix} \mathbf{M_1} & \mathbf{M_2} \\ \mathbf{M_3} & \mathbf{M_4} \end{pmatrix} \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix}, \quad (20)$$

where

$$\mathbf{M_1} = \frac{l'}{l}\boldsymbol{E}, \quad (21)$$

$$\mathbf{M_2} = \frac{1}{l^2}\left(\boldsymbol{I} - \kappa l \boldsymbol{C}\right), \quad (22)$$

$$\mathbf{M_3} = \boldsymbol{K}^2 + \kappa l\left(\boldsymbol{C}\,(kl)^2 - \boldsymbol{D}\right), \quad (23)$$

$$\mathbf{M_4} = -\frac{l'}{l}\boldsymbol{E}^t, \quad (24)$$

where

$$E_{mn} = \langle \varphi_m \boldsymbol{v}, \boldsymbol{\nabla}_\perp \varphi_n \rangle, \quad (25)$$

$$C_{mn} = \langle \varphi_m z, \varphi_n \rangle, \quad (26)$$

$$K_{mn}^2 = \left(\gamma_m^2 - (kl)^2\right)\delta_{mn} + jkl\zeta_n K_{mn}^{R2}, \quad (27)$$

$$K_{mn}^{R2} = \int_\Gamma \bar{\varphi}_m \varphi_n d\Gamma, \quad (28)$$

$$D_{mn} = \langle \boldsymbol{\nabla}_\perp \varphi_m, z \boldsymbol{\nabla}_\perp \varphi_n \rangle, \quad (29)$$

where Γ is the contour of the segment and the component ζ_n corresponds to the boundary impedance specific to the

mode φ_n. This allows one to account for the tangential velocity specific to each mode in the computation of viscous losses as proposed in [17]. For the full development of the modal projection, see Appendix V-B.

C. COMPUTATION OF THE ACOUSTIC FIELD
1) COMPUTATION OF THE TRANSVERSE MODES AND THE PROPAGATION MATRICES
Since the vocal tract cross-sectional shapes are quite different from simple shapes such as ellipses or rectangles, the transverse eigenvalues problem giving the transverse modes φ_n and the associated eigenvalues γ_n^2 is solved using 2D finite elements. A Neumann boundary condition is considered as it is closer to the properties of the vocal tract walls and thus, ensure a faster convergence of the method. Note that an improved formulation using a so called supplementary mode would further improve the convergence [21]. The solution obtained is used to compute the propagation matrices $\boldsymbol{E}$, $\boldsymbol{C}$, $\boldsymbol{K}^2$ and $\boldsymbol{D}$, defined in Eqs. (25), (26), (27) and (29).

The transverse functions φ_n can be expressed as the summation of shape functions $e_i(y, z)$ corresponding to the finite elements

$$\varphi_n = \sum_{i=1}^{N} \xi_{ni} e_i = \boldsymbol{\xi_n} \cdot \boldsymbol{e}, \quad (30)$$

where ξ_{ni} are the amplitudes of the shape functions. The transverse modes φ_n and their associated eigenvalues γ_n^2 are obtained solving the eigenvalue problem

$$\boldsymbol{\xi_n}^t \boldsymbol{A} = k_n^2 \boldsymbol{\xi_n}^t \boldsymbol{M}, \quad (31)$$

where $\boldsymbol{A}$ is the stiffness matrix whose terms are

$$A_{ij} = \langle \boldsymbol{\nabla}_\perp e_i, \boldsymbol{\nabla}_\perp e_j \rangle, \quad (32)$$

and $\boldsymbol{M}$ is the mass matrix whose terms are

$$M_{ij} = \langle e_i, e_j \rangle. \quad (33)$$

The propagation matrices $\boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}$ and $\boldsymbol{K}^{R2}$ can be obtained from the eigenvectors $\boldsymbol{\xi_n}$ solutions of Eq. (31) as

$$C_{mn} = \boldsymbol{\xi_m}^t \boldsymbol{M_Z} \boldsymbol{\xi_n}, \quad (34)$$

$$D_{mn} = \boldsymbol{\xi_m}^t \boldsymbol{A_Z} \boldsymbol{\xi_n}, \quad (35)$$

$$E_{mn} = \boldsymbol{\xi_m}^t \boldsymbol{B} \boldsymbol{\xi_n}, \quad (36)$$

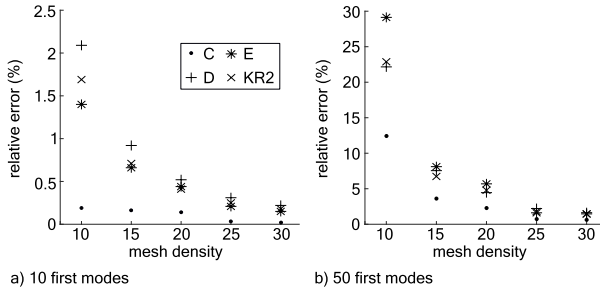$$K_{mn}^{R2} = \boldsymbol{\xi_m}^t \boldsymbol{R} \boldsymbol{\xi_n}, \quad (37)$$

where

$$M_{Zij} = \langle z e_i, e_j \rangle, \quad (38)$$

$$A_{Zij} = \langle z \boldsymbol{\nabla}_\perp e_i, \boldsymbol{\nabla}_\perp e_j \rangle, \quad (39)$$

$$B_{ij} = \langle \boldsymbol{v} e_i, \boldsymbol{\nabla}_\perp e_j \rangle, \quad (40)$$

$$R_{ij} = \int_\Gamma e_i e_j d\Gamma. \quad (41)$$

For the details of the implementation of the 2D finite element method the reader is referred to classical textbooks such as [37].

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE*Access*

a) 10 first modes b) 50 first modes

FIGURE 3. Relative error between the matrices C, D, E and K^{R2} computed numerically with Eqs. (34), (35), (36) and (37) and analytically with the expressions provided in Appendix V-A for the first (a) 10 and (b) 50 transverse modes.

FIGURE 4. Junction between two waveguide segments with different cross-sectional contours and area varying along the propagation axis.

The computation of the transverse modes and the propagation matrices has been validated comparing it with analytical solutions obtained for a rectangular shape of dimensions 5.5 cm × 3.2 cm. The expression of the transverse modes and the propagation matrices for a rectangular shape are provided in Appendix V-A.

The relative error of the matrices C, D, E and K^{R2} has been computed as the ratio of the norm of the difference between the numerical and the analytical solutions over the norm of the analytical solution. For example, for the matrix C the relative error is $||C_A - C||/||C_A||$, with C_A the analytical solution. It is presented for various densities in Figs. 3a and 3b for the 10 and 50 first modes respectively. The mesh density is defined as the ratio of the square root of the cross-sectional surface over the average side length of the triangular elements. It can be seen as the number of elements per characteristic length.

As expected, the relative error decreases when the mesh density is increased. For the 10 first modes a density of 15 is sufficient to obtain an error lower than 1%. But for the 50 first modes, one need a density of 30 to achieve this. The error depends on the type of matrix: the matrix C has the lowest errors, and the matrix D has the highest. This may be due to the approximation of the gradient of the modes $\nabla_\perp \varphi$ for the matrices D and E which could potentially generate more errors, and of the integration on the contour for the matrix K^{R2} which involves less elements than surface integration. The impact of the mesh density on the computation of transfer functions is evaluated in Section IV-A.

2) CONNECTION OF THE SEGMENTS
As shown in Fig. 1, the vocal tract geometry is described by a succession of segments which need to be connected to each other. This is done by assuming the continuity of the acoustic field at the junctions, as proposed by Pagneux *et al.* [45].

Fig. 4 illustrates a junction between two waveguide segments with different cross-sectional contours and area varying along the propagation axis. The cross-section of the segment a is contained in the one of the segment b, and, hence, the surface S_a (in grey in Fig. 4) is smaller than surface S_b. The variations of cross-sectional area inside the
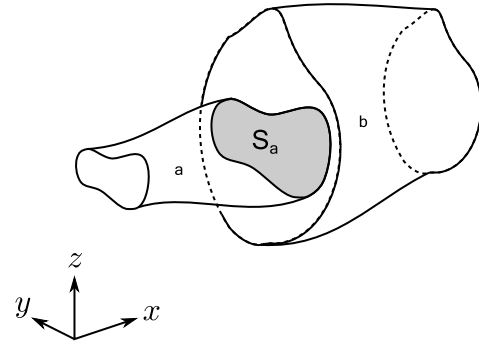
segments a and b are described by the scaling functions $l_a(x_a)$ and $l_b(x_b)$ respectively.

Relationships between the modal amplitudes at both sides of the interface between the segments a and b can be derived from the condition of continuity of the acoustic field on the common surface S_a and the boundary condition $q_b = 0$ on the surface $S_b - S_a$ of the wall of the largest segment

$$p_a = \frac{l_b}{l_a}Fp_b, \tag{42}$$

$$q_b = \frac{l_b}{l_a}F^tq_a, \tag{43}$$

where

$$F = \int_{S_{aX}} \varphi_a\varphi_b{}^t \frac{dYdZ}{l_al_b}, \tag{44}$$

where S_{aX} is the surface S_a expressed in Cartesian coordinates (X, Y, Z). The scaling factors l_a and l_b must be introduced as the integration is done in Cartesian coordinates (X, Y, Z) (see Appendix V-C).

Combining Eqs. (42) and (43) and introducing the impedance and admittance matrices $\mathcal{Z}$ and $\mathcal{Y}$ defined as $p = \mathcal{Z}q$ and $q = \mathcal{Y}p$ allows one to derive relationships between the impedance and admittance matrices on side a and b

$$\mathcal{Z}_a = \left(\frac{l_b}{l_a}\right)^2 F\mathcal{Z}_bF^t, \tag{45}$$

$$\mathcal{Y}_b = \left(\frac{l_b}{l_a}\right)^2 F^t\mathcal{Y}_aF. \tag{46}$$

However, the segment interfaces obtained by slicing vocal tract geometries do not always correspond to this case: the cross-section of side a may not be fully contained in the cross-section of side b. According to Ginsberg [27], this case cannot be treated directly using an orthogonality based method as used here. To overcome this limitation, zero length segments whose cross-section is the intersection between the cross-section a and b are introduced in this case. This allows one to use the expression derived above. Note that, according to Ginsberg [27], a collocation based method would not have this limitation, and could be used as an alternative.
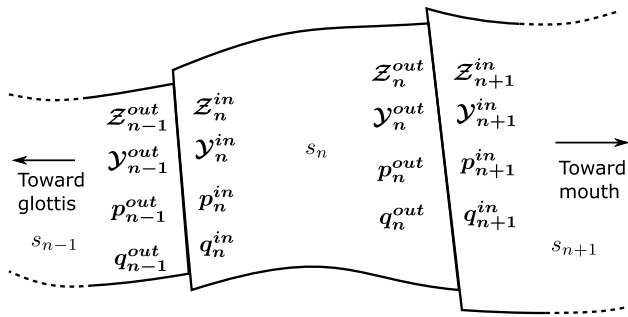
FIGURE 5. Disposition of the physical quantities solution of the wave problem: the impedance $\mathcal{Z}$ and admittance $\mathcal{Y}$ matrices, the modal amplitude of the acoustic pressure p and the related quantity q, at the input and output of the segment s_n.

3) PROPAGATION

The vectors p and q and the impedance and admittance matrices $\mathcal{Z}$ and $\mathcal{Y}$ can be computed solving numerically the Ricatti equation with a Magnus-Möbius scheme as proposed by Pagneux [46]. The vectors $p(x_{n+1})$ and $q(x_{n+1})$ can be obtained at a position x_{n+1} from their value at the position x_n through the relationship

$$\begin{pmatrix} p(x_{n+1}) \\ q(x_{n+1}) \end{pmatrix} = \Lambda \begin{pmatrix} p(x_n) \\ q(x_n) \end{pmatrix} = \begin{pmatrix} \Lambda_1 & \Lambda_2 \\ \Lambda_3 & \Lambda_4 \end{pmatrix} \begin{pmatrix} p(x_n) \\ q(x_n) \end{pmatrix} \quad (47)$$

where Λ is, for the fourth order of the Magnus scheme

$$\Lambda = \exp\left(\frac{x_{n+1} - x_n}{2} (\mathbf{M}(x_a) + \mathbf{M}(x_b)) \right.$$
$$\left. + \frac{\sqrt{3}(x_{n+1} - x_n)^2}{12} [\mathbf{M}(x_b), \mathbf{M}(x_a)] \right), \quad (48)$$

where $[\mathbf{M}(x_b), \mathbf{M}(x_a)] = \mathbf{M}(x_b)\mathbf{M}(x_a) - \mathbf{M}(x_a)\mathbf{M}(x_b)$ is a commutator, $\mathbf{M}(x)$ is defined in Eq. (20) and

$$x_a = x_n + \left(\frac{1}{2} - \frac{\sqrt{3}}{6} \right)(x_{n+1} - x_n), \quad (49)$$

$$x_b = x_n + \left(\frac{1}{2} + \frac{\sqrt{3}}{6} \right)(x_{n+1} - x_n). \quad (50)$$

As the boundary condition at the mouth is a radiation impedance matrix, it is necessary to compute the impedance matrix $\mathcal{Z}_{n+1}$ at a position x_{n+1} from the impedance $\mathcal{Z}_n$ at a position x_n. This is obtained through the relationship

$$\mathcal{Z}_{n+1} = (\Lambda_1 \mathcal{Z}_n + \Lambda_2)(\Lambda_3 \mathcal{Z}_n + \Lambda_4)^{-1}. \quad (51)$$

Likewise, the admittance $\mathcal{Y}_{n+1}$ can be related to $\mathcal{Y}_n$ by

$$\mathcal{Y}_{n+1} = (\Lambda_3 + \Lambda_4 \mathcal{Y}_n)(\Lambda_1 + \Lambda_2 \mathcal{Y}_n)^{-1}. \quad (52)$$

D. GENERAL WORKFLOW

The whole process of solving the wave problem is described in the pseudo-code algorithm 1. The different physical quantities implied in solving the wave problem are shown in Fig. 5.

The equation (20) is solved by setting a radiation impedance boundary condition at the mouth end and a particle velocity field at the glottis end. The radiation impedance condition is described by a radiation impedance matrix which can be numerically integrated following the method described in [13]. It is necessary to compute this radiation impedance matrix in the Cartesian coordinates system. Thus, the distances must be multiplied by the factor l at the end of the last of the N_s segments. In an infinite baffle, the radiation impedance matrix $\mathcal{Z}_{N_s}^{out}$ is given by

$$\mathcal{Z}_{N_s}^{out} = \frac{-1}{2\pi} \int_S \int_{S_0} \varphi(y, z)\varphi^t(y_0, z_0) \frac{e^{-jklh}}{lh} dS_0 \, dS, \quad (53)$$

where $h = \sqrt{(y - y_0)^2 + (z - z_0)^2}$. Alternatively, the radiation impedance matrix can be computed using the method proposed by Felix and Doc [22].

The impedance and admittance matrices are computed from the mouth to the glottis, and the acoustic field is computed from the glottis input particle velocity to the mouth using the impedance and admittance matrices computed previously.

The acoustic pressure radiated to an external point (x_r, y_r, z_r) is computed with the Rayleigh-Sommerfeld integral in the same scaled space

$$p(x_r, y_r, z_r) = \frac{-1}{2\pi} \int_S q_{N_s}^{out} \cdot \varphi(y, z) \frac{e^{jklr}}{lr} dS, \quad (54)$$

where $r = \sqrt{x_r^2 + (y - y_r)^2 + (z - z_r)^2}$ and $(y, z) \in S$.

This simulation method was implemented as a new functionality of the articulatory synthesizer VocalTractLab [8], which will be made available to the public soon on the website www.vocaltractlab.de.

III. SIMULATIONS

The method presented in section II has been applied to test geometries and compared to FEM simulations. This section presents the test geometries used, the parameters for the multimodal simulations and the FEM used. The results of these simulations are compared and discussed in Section IV.

A. GEOMETRIES
1) SINGLE SEGMENT

The first test geometry is a single horn shaped segment with a circular cross-section. Its centerline is defined as a circle arc of radius 7.5 cm and angle 130° yielding a total length $L = 16.95$ cm (see Fig. 8). These dimensions are chosen to be of the same order as vocal tract dimensions. The scaling factor $l(x)$ along the centerline is

$$l(x) = \frac{1}{4} + \frac{9}{4}\left(\frac{x}{L}\right)^2 - \frac{3}{2}\left(\frac{x}{L}\right)^3. \quad (55)$$

2) VOCAL TRACT GEOMETRIES

After testing the method on one segment, it was tested on more realistic vocal tract geometries constituted of multiple segments. They were generated by the articulatory model

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE *Access*

Algorithm 1 Solve Wave Problem in a Segmented Geometry

1: Slice the vocal tract geometry in N_s segments
2: **for** Each segment **do**
3:     Mesh the contour
4:     Compute the matrices A, M, M_z, A_z, B and R integrating Eqs. (32), (33), (38), (39), (40) and (41)
5:     Compute the transverse modes φ_m and their eigenvalues γ_m^2 solving Eq. (31)
6:     Compute the matrices C, D, E and K^{R2} using Eqs. (34), (35), (36) and (37)
7: **end for**
8: **for** Each segment **do**
9:     **if** the segment is connected to another one **then**
10:         Compute the mode matching matrix F integrating Eq. (44)
11:     **end if**
12: **end for**
13: **for** each frequency **do**
14:     Compute the radiation impedance matrix $\mathcal{Z}_{N_s}^{out}$ following [22] or [13]
15:     $n = N_s$
16:     **while** $n > 0$ **do**
17:         Propagate either $\mathcal{Z}_n$ or $\mathcal{Y}_n$ depending on the input using Eqs. (51) or (52) inside the segment s_n
18:         **if** exit area of s_{n-1} > entrance area of s_n **then**
19:             If necessary, compute $\mathcal{Z}_n^{in} = \mathcal{Y}_n^{in-1}$
20:             Compute the impedance matrix $\mathcal{Z}_{n-1}^{out}$ at the exit of the next segment using Eq. (45)
21:         **else**
22:             If necessary compute $\mathcal{Y}_n^{in} = \mathcal{Z}_n^{in-1}$
23:             Compute the admittance matrix $\mathcal{Y}_{n-1}^{out}$ at the exit of the next segment using Eq. (46)
24:         **end if**
25:         $n = n - 1$
26:     **end while**
27:     Compute the input modal amplitude $q_0^{in} = \int_{S_0} \varphi_0 q_0^{in} dS_0$
28:     Compute $p_0^{in} = \mathcal{Z}_0^{in} q_0^{in}$ (if necessary compute $\mathcal{Z}_0^{in} = \mathcal{Y}_0^{in-1}$)
29:     $n = 0$
30:     **while** $n < N_s$ **do**
31:         Propagate p_n and q_n using Eq. (47)
32:         **if** entrance area of s_{n+1} > exit area of s_n **then**
33:             Compute q_{n+1}^{in} at the entrance of the next section using Eq. (43)
34:             Compute $p_{n+1}^{in} = \mathcal{Z}_{n+1}^{in} q_{n+1}^{in}$
35:         **else**
36:             Compute p_{n+1}^{in} at the entrance of the next section using Eq. (42)
37:             Compute $q_{n+1}^{in} = \mathcal{Y}_{n+1}^{in} p_{n+1}^{in}$
38:         **end if**
39:         $n = n + 1$
40:     **end while**
41:     If necessary, compute $q_{N_s}^{out} = \mathcal{Z}_{N_s^{out}}^{-1} p_{N_s}^{out}$
42:     Compute the radiated acoustic pressure using the Rayleigh-Sommerfeld integral Eq. (54)
43: **end for**

---

implemented in VocalTractLab 2.3 (www.vocaltractlab.de). The geometries provided by this software have been fitted on MRI scans of a male subject [8]. The vowels /a/, /i/ and /u/ were selected as they represent well the variety of vocal tract shapes for vowels. To ensure the best geometrical accuracy possible, the maximal number of segments provided by the segmentation algorithm implemented in VocalTractLab 2.3 was used. This results in 129 segments whose length ranges between 0.55 mm and 1.6 mm. The corresponding contours were exported in text files so that they could be used with Matlab and GID[3] to generate finite element meshes. During this process, the consecutive contours were linked by straight segments. This introduces slight differences with the geometry actually simulated with the multimodal method. In fact, since the proposed method considers the contour shape as constant within each segment, small discontinuities are introduced, making the geometry simulated rough compared to the one simulated with FEM. The potential consequences of such discontinuities are discussed in the section IV-B.

B. PARAMETERS FOR THE MULTIMODAL SIMULATIONS

For both the single segment and the vowel geometries, the multimodal simulations were performed using the same value of wall boundary admittance coefficient $\zeta = 0.005$ as in [2]. A uniform input particle velocity $G(f) = q^{in} \cdot \varphi$, with f the frequency, was imposed on the glottal cross-sectional surface Γ_G. This was done setting the amplitude of the plane mode to $q_0^{in} = -j2\pi f \rho$ and the other ones to zero. A vocal tract transfer function was computed as

$$H(f) = \frac{P_o(f)}{A_g G(f)} [\text{Pa/m}^3 \text{s}^{-1}], \qquad (56)$$

where $P_o(f)$ is the acoustic pressure tracked inside the geometries at 3 mm from the center of the exit surface, $G(f) = -j2\pi f \rho \varphi_0$ the uniform input velocity (φ_0 being the plane mode) and A_g the glottal cross-sectional area.

1) SINGLE SEGMENT

A zero pressure boundary condition was simulated at the open end of the single segment. It was approximated by setting a uniform impedance of 10^{-16} kg.m^{-2}.s^{-1} on the exit surface. This was done by defining $\mathcal{Z}_{N_s}^{out}$ as a diagonal matrix whose diagonal terms are all equal to this value. The impact of different numbers of modes, mesh densities and number of integration points were evaluated. Simulations were performed using the 6, 17, 32 and 53 first modes (ordered by increasing cutoff frequency). The choice of these numbers is done to include the second, third, fourth and fifth axi-symmetric modes which have significant contribution to the acoustic field in geometries with circular cross-section. The mesh density was varied from 10 to 30 and the number of integration points from 25 to 200 (about 1.5 to 12 points per centimeter).

---

[3]https://www.gidhome.com/

IEEE *Access*

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

2) VOWELS

The boundary condition at the open end was simulated using a radiation impedance matrix numerically integrated using the method proposed in [13]. The number of modes in each segment was determined by a maximal cutoff frequency criterion: all the transverse modes having a cutoff frequency lower than a chosen value were included. This results in a reduced number of modes for the smaller segments, which allows one to minimise the number of modes to be simulated over the entire geometry for a given accuracy. For a maximal cutoff frequency of 40 kHz the number of transverse modes ranges between 2 and 42. Different values of this criterion ranging from 20 kHz to 60 kHz were used to evaluate the influence of this parameter (see section IV-B).

Different geometrical approximations were explored:

- straight segments with constant cross-sectional area ($\kappa = 0$, $l(x) = 1$), corresponding to the state of the art multimodal simulations of vocal tract geometries [11],
- segments with curvature ($\kappa = 1/R_c$, $l(x) = 1$), where R_c is the curvature radius determined by the angle between the normals of the cutting planes at both ends of the segment,
- segments with curvature ($\kappa = 1/R_c$) and area variations defined as a linear interpolation between the area at the entrance and the end of the segments (case presented in Fig. 2).

C. FINITE ELEMENT SIMULATIONS

FEM simulation were performed to compare to the multimodal simulations. It was used to numerically solve the wave equation in the time domain for the acoustic pressure $p(\boldsymbol{m}, t)$ and particle velocity $\boldsymbol{u}(\boldsymbol{m}, t)$

$$\frac{1}{\rho c^2} \partial_t p + \nabla \cdot \boldsymbol{u} = 0, \tag{57a}$$

$$\rho \partial_t \boldsymbol{u} + \nabla p = 0, \tag{57b}$$

in a computational domain Ω. Equation (57) is supplemented with the following set of boundary and initial conditions (see Fig. 6)

$$\boldsymbol{u} \cdot \boldsymbol{n}_X = g(t) \text{ on } \Gamma_G, \ t > 0, \tag{58a}$$

$$\boldsymbol{u} \cdot \boldsymbol{n}_X = p/Z_w \text{ on } \Gamma_W, \ t > 0, \tag{58b}$$

$$\boldsymbol{u} \cdot \boldsymbol{n}_X = 0 \text{ on } \Gamma_H, \ t > 0, \tag{58c}$$

$$\boldsymbol{u} \cdot \boldsymbol{n}_X = p/Z_0 \text{ on } \Gamma_\infty, \ t > 0, \tag{58d}$$

$$p = 0, \quad \boldsymbol{u} = 0 \text{ in } \Omega, \ t = 0. \tag{58e}$$

$g(t)$ is used to introduce a particle acoustic velocity at the glottal cross-sectional area Γ_G, Z_w is the wall impedance imposed at the vocal tract walls Γ_W to introduce losses, and $Z_0 = \rho c$ is the air impedance used in (58d) to implement a Sommerfeld boundary condition so as to absorb sound waves reaching the outer boundary Γ_∞. Z_w is computed from the boundary admittance coefficient $\zeta = Z_0/Z_w$, which was set to the same value as for the multimodal simulations: $\zeta = 0.005$. A rigid wall boundary condition was enforced on the baffle surface Γ_H of the external domain.
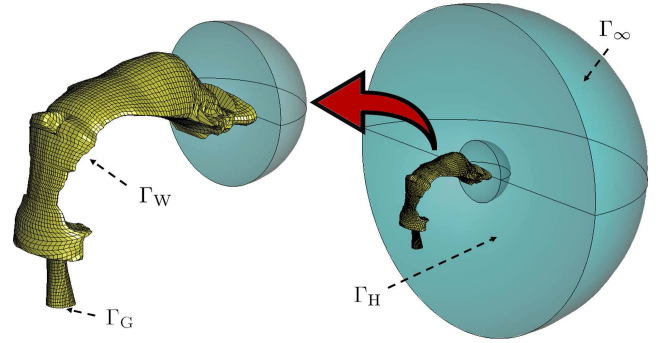

FIGURE 6. Computational domain Ω used for the FEM simulation of vowel /a/, with boundaries Γ_G for the glottis, Γ_W for the vocal tract walls, Γ_H for the circular baffle representing the head, and Γ_∞ for the outer boundary absorbing all incoming sound waves as if they were propagating towards infinity.
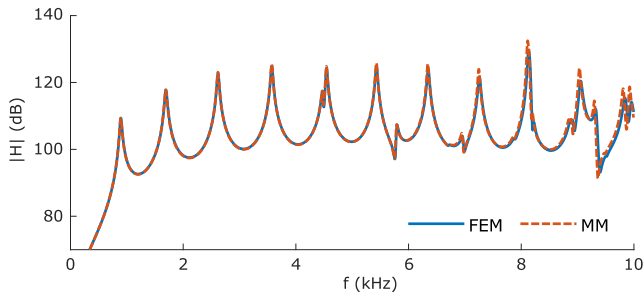
For the vocal tract geometries, the computational domain Ω consists of a vocal tract set in a hemisphere of radius 0.16 m to allow sound waves radiate outwards from the mouth (see Figure 6). In the single segment geometry, however, the radiation domain is removed and Ω terminates at the horn exit. A zero pressure release condition ($p = 0$) is imposed on it to emulate an open-end condition. A volume mesh was next generated discretizing Ω in a set of tetrahedra of average size 0.003 m in the horn or vocal tract, and {0.004, 0.0065} m in the inner and outer radiation domains of the vocal tract geometries.

The wave equation (57) with boundary and initial conditions (58) was next solved in Ω using a custom FEM code based on the stabilization strategy defined in [29]. The stabilization parameter was reduced from 0.05 to 0.01 compared to [29], which allowed us to use larger time steps while preserving a low numerical dissipation. The following Gaussian pulse was used as input for the vocal tract,

$$g(t) = e^{-[(t-T_{gp})/0.29\,T_{gp}]^2} \text{[m/s]}, \tag{59}$$

with $T_{gp} = 0.646/f_c$ and $f_c = 10$ kHz. This pulse was low-pass filtered at 10 kHz to avoid numerical errors above the maximum frequency of analysis (10 kHz). A numerical simulation of a time event lasting 0.05 s was then performed using a time step of 2e-6 s. The acoustic pressure $p_o(t)$ was tracked close to the vocal tract exit, inside of it, at 3 mm from the mouth aperture center. This helped us to minimize possible spurious reflections coming from Γ_∞, compared to capturing the acoustic pressure in the radiation domain. A vocal tract transfer function was finally computed in the same way as for the multimodal simulations defined in Eq. (56). $P_o(f)$ and $G(f)$ were computed as the Fourier transform of $p_o(t)$ and $g(t)$ respectively.

This simulation method showed in previous studies a good agreement with experimental data [2], [11]. More generally, the FEM can be trusted to describe accurately vocal tract acoustics, as illustrated by the good agreement obtained by Fleischer *et al.* [25] between a different implementation of FEM and experimental data.

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE *Access*

FIGURE 7. Amplitude of the transfer function between the input volume velocity and the acoustic pressure at a point located inside 3 mm away from the center of the exit of a single segment computed with finite element and the proposed multimodal method.
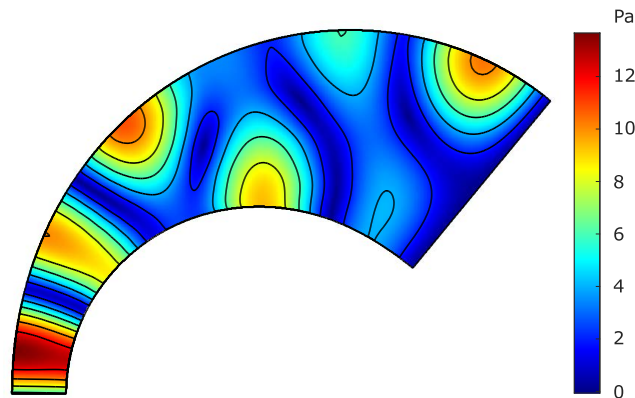

FIGURE 8. Acoustic pressure amplitude (Pa) computed for an input volume velocity of $1\,\mathrm{cm}^3$/s in the plane (X, Z) with the proposed method inside a single segment at the frequency 5755 Hz which corresponds to a trough in the transfer function presented in Fig. 7. The iso-amplitude lines are higlighted in black.

IV. COMPARISON WITH FINITE ELEMENT SIMULATIONS

A. SINGLE SEGMENT

The figure 7 shows the transfer function of the single segment computed with FEM and the multimodal method with 53 modes, a mesh density of 30 and a 200 integration points (about 12 points per centimeter). It shows the typical resonances produced by the plane mode, but also exhibits troughs and additional resonances related to transverse resonances induced by the curvature and higher order modes propagation. See as an example the trough located at 5755 Hz and the corresponding acoustic field, computed with the multimodal method, presented in Fig. 8, in which these effects are clearly observable. The wall boundary condition described in Eq. (7) implies that in the vicinity of the walls the normal gradient of the acoustic pressure is very small. This implies that the iso-amplitude lines are almost perpendicular to the walls. This can be observed in Fig. 8, thus, confirming that this boundary condition is properly simulated.

The influence of the number of modes, the mesh density and the number of integration points on the simulations was examined by computing the transfer function with different values of these parameters. The resonance frequencies, -3 dB bandwidths and amplitude of the resonances of the obtained
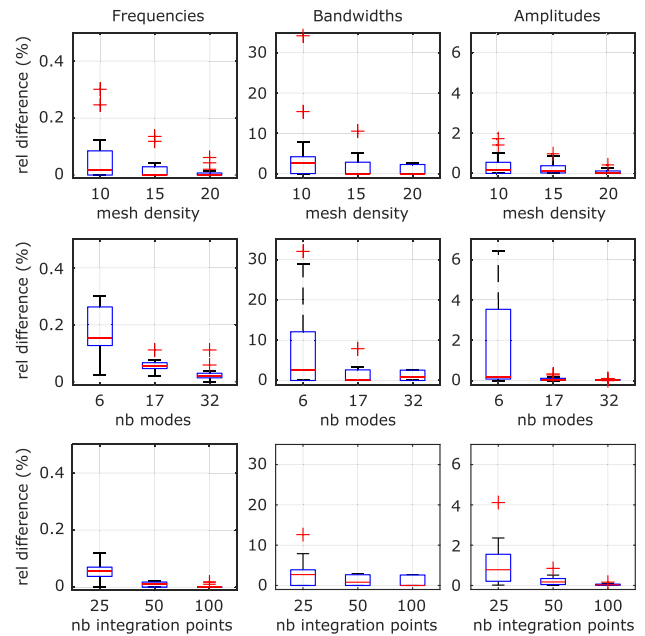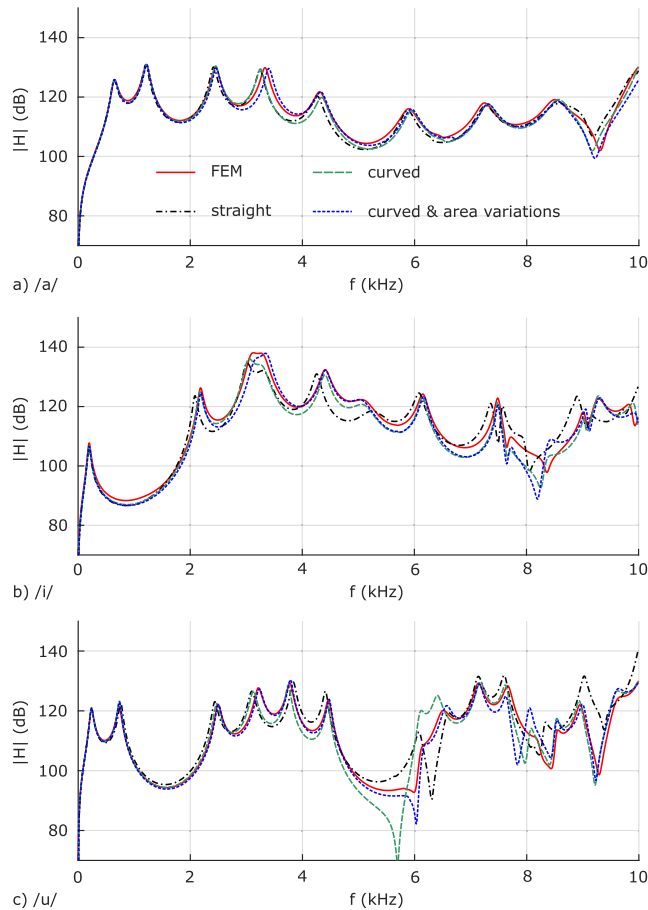

FIGURE 9. Relative difference (percentage) between the resonance frequencies, bandwidth and amplitude of the transfer functions of a segment with curvature and varying cross-sectional area computed with different number of modes, mesh density and number of integration points. The transfer function corresponding to the reference case (53 modes, a mesh density of 30 and 200 integration points) is presented in Fig. 7. The outliers are plotted individually as crosses.

transfer functions were compared with a reference case of 53 modes, a mesh density of 30 and 200 integration points (corresponding to about 1.5 to 12 points per centimeter).

This reference case was first compared with the FEM simulation with which a very good agreement was observed. As can be seen in Fig. 7, the amplitude of both transfer functions is almost perfectly superimposed, with some slight differences increasing toward the high frequencies which consists mainly in stronger losses and slightly higher resonance frequencies for FEM. The latter can be attributed to the stiffening effect of FEM [19]. More precisely, the average differences are 0.16%, 15.6% and 1.2% for the resonance frequencies, bandwidth and amplitudes respectively. The maximal differences are 0.33%, 51.3% and 3.7% for the frequencies, bandwidth and amplitudes respectively. The significant differences in bandwidth can be attributed to the time and spatial discretization of FEM, which tends to artificially increase the losses with frequency. Indeed, note that below 8 kHz the FEM and MM curves match to a large extend, reducing the average differences to 0.1%, 6.8% and 0.45% for the resonance frequencies, bandwidths, and amplitudes, respectively. It is to be mentioned that FEM results could be improved using a finer mesh and smaller time step, but at the price of increasing the computational cost.

The relative difference between different values of parameters and the reference case (mesh density of 30, 53 modes and 200 integration points) are presented as box plots in Fig. 9. When a parameter is changed, it is implicitly understood that

IEEE *Access*

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements


FIGURE 10. Transfer functions between the input volume velocity at the glottis and the acoustic pressure at a point located inside 3 mm away from the center of the exit surface for vocal tract geometries corresponding to the vowels /a/, /i/ and /u/ of a male speaker (129 segments). The transfer functions have been computed with FEM and with the proposed multimodal method using different geometrical approximations for the segments, a mesh density of 15 for the computation of the transverse modes, 3 integration points per segment and a maximal cutoff frequency of 60 kHz.

the other parameters are set to their maximal values (e.g. the mesh density and the number of integration points are set to 30 and 200 when the number of modes is varied). The differences reduce as the parameters get closer to the reference case. The parameter which induces the greatest impact is the number of modes. Large differences are observed for the bandwidth which is equally influenced by the three tested parameters. 32 modes are sufficient to achieve an average difference lower than 0.1% for the frequencies and amplitudes, but still more than 1% for the bandwidths.

B. VOCAL TRACT GEOMETRIES
1) TRANSFER FUNCTIONS
The transfer functions of the vowel geometries obtained with different geometrical approximations and with FEM are presented in Fig. 10. There is an overall good agreement for the 3-4 first resonances between the different geometrical approximations: the difference in resonance frequencies with
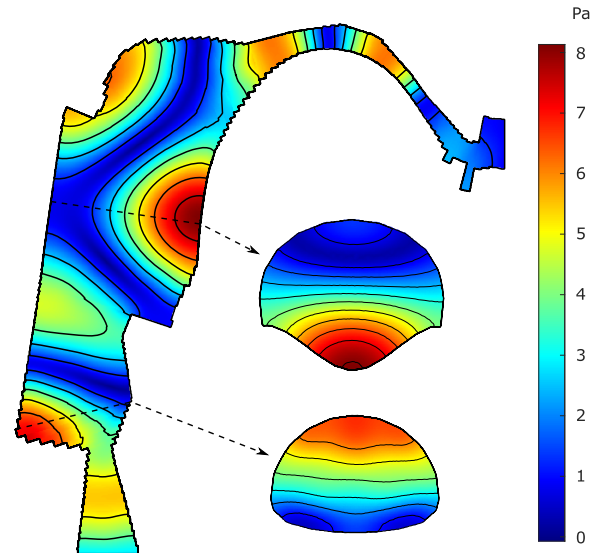
FIGURE 11. Acoustic pressure amplitude computed with the proposed method in the sagittal plane and in two transverse cutting planes indicated as dashed lines in a vocal tract geometry corresponding to the vowel /i/ at 7500 Hz for an input volume velocity of $1 cm^3/s$. This corresponds to the 6th resonance of the transfer function obtained with curved segments in Fig. 10. The iso-amplitude lines are highlighted in black.

FEM remains lower than 5% in all cases. This is in line with the fact that the plane wave assumption is a good approximation up to 4-5 kHz [15].

The agreement with the FEM solution is better when the curvature is considered, and further improved when the variations of cross-sectional area are taken into account. The resonance frequencies are generally lower with the straight segment approximation. In particular for the second resonance of the /i/ (see Fig. 10b). This is in line with the findings of Arnela *et al.* [2]. Furthermore, taking into account the curvature just by changing a parameter of the model confirms that this effect is induced by the curvature alone. As a matter of fact, in the study of Arnela *et al.* [2] this could have also been induced by the method used to straighten the vocal tract geometries.

There is overall more difference between the geometrical approximations above 4-5 kHz, in particular for /i/ and /u/. Taking into account the variation of cross-sectional area substantially improves the agreement with FEM in this frequency range for /u/ (see Fig. 10c). Thus, the high frequency acoustic properties of the vocal tract appear to be sensitive to small geometrical details. This is in line with the findings of Motoki [43] who showed that small geometrical perturbations of simplified vocal tract geometries induced more important variations of the transfer function above 4 kHz. Despite the roughness introduced by the contour change, there is a relatively good agreement with FEM simulation performed with a smoother geometry.

Substantial differences remain between FEM and the multimodal method for the case accounting for the curvature and the cross-sectional area variations (about 1% of difference in

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE *Access*

the resonance frequencies), see also around 8 kHz for /i/ and /u/ (see Figs. 10b and 10c). This can be attributed to remaining geometrical differences due to the discontinuities introduced by the contour changes, and to intrinsic differences between both methods. The small geometrical differences could be slightly reduced, for example by using a better interpolation of the cross-sectional area, or by refining the segmentation of parts having significant changes of cross-sectional shape. However, the gain in accuracy may be of the order of the uncertainties in the extraction of the vocal tract geometries from medical images. It would also be possible to optimise the segmentation by using longer segments in the parts having little variations of cross-sectional shape, and thus further reduce the computational cost. The algorithm used to generate the centerline curve can have an impact as well, as it was observed by Arnela *et al.* [5]. However, it is expected to be smaller for 3D simulations than for TLM simulations.

The losses are slightly lower with the multimodal method: the resonances have smaller -3 dB bandwidths (of the order of 10%). This can be attributed to the perfectly reflective surfaces which are considered at the discontinuities between the segments (see Section II-C2). On the other hand, the numerical scheme used for the FEM could tend to slightly artificially increase the losses as the frequency increases due to the adopted stabilization strategy and the time and space discretization (see Section III-C).

2) ACOUSTIC FIELD
The Fig. 11 illustrates the acoustic field computation inside a segmented geometry with the field corresponding to the 6th resonance of the vowel /i/ (7500 Hz). As for the single segment, the effect of the transverse modes is clearly visible as transverse variations both in the sagittal and transverse planes. The iso-amplitude lines tend to be perpendicular to the walls of the geometry, confirming that the boundary condition is properly simulated. When looking carefully, the effect of the segmentation is visible as small discontinuities in the iso-amplitude lines where there is a strong discontinuity at the interface between two segments (see as an example at the expansion at the base of the geometry). This may be improved by using more transverse modes to describe the strong discontinuities.

3) INFLUENCE OF THE NUMBER OF MODES
Multimodal simulations were run with values of the maximal cutoff frequency ranging from 20 kHz to 60 kHz using the best geometrical approximation (taking into account the curvature and the variations of cross-sectional area). The resonance frequencies of the transfer functions were compared with the solution obtained with 60 kHz and the FEM solution. The distribution of the relative error is presented as box plots in Fig. 12. For the 3 vowels tested the simulations converge toward the 60 kHz solution, with less than 1% of difference from 40 kHz. It does not strictly converge toward the FEM solution, as from 40 kHz onward the difference with FEM remains stable or slightly increases, but remains smaller
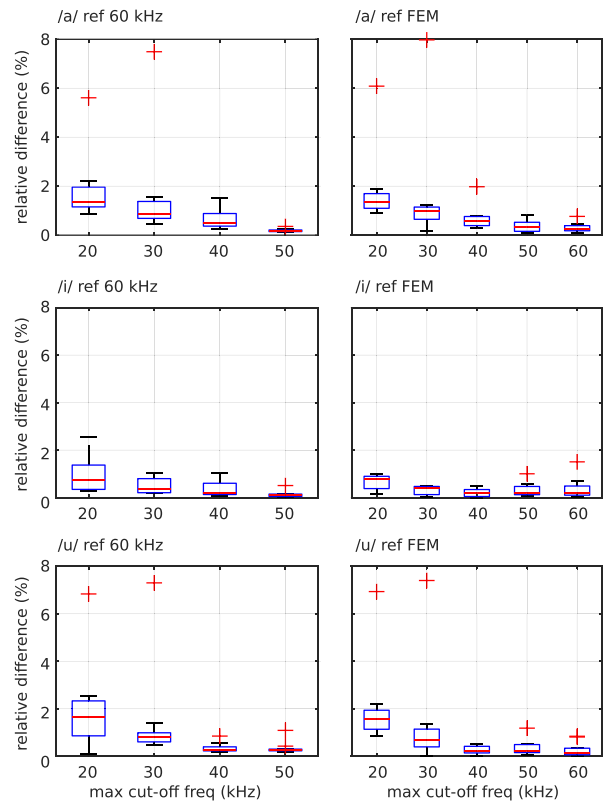
FIGURE 12. Relative error of the resonance frequencies for different values of the maximal cutoff frequencies for segmented vocal tract geometries corresponding to the vowels /a/, /i/ and /u/. The outliers are plotted individually as crosses.
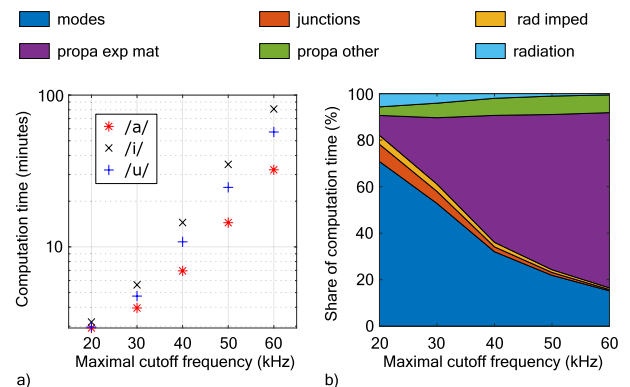
FIGURE 13. a) Computation times for the transfer function of the vowels /a/, /i/ and /u/ for different values of the maximal cutoff frequency, a mesh density of 15, about 20 integration point per centimeter and 1000 frequencies. b) Share of the computation time taken by the different tasks for the vowel /i/: transverse modes, junction matrices, radiation impedance matrix, computation of the matrix exponential for the propagation, other tasks for the propagation and radiated field.

than 1%. Note also that for some resonances the difference does not decrease rigorously monotonously: it can increase for some specific values of the maximal cutoff frequency (see as an example the outlier at 30 kHz for /a/).

4) COMPUTATION TIME
Fig. 13a presents the computation times for the transfer functions of the vowels /a/, /i/ and /u/ for different values of the maximal cutoff frequency, a mesh density of 15, 3 integration

IEEE *Access*

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

points per segment (about 20 points per centimeter) and 1000 frequencies. The computations were run in a single thread on a processor Intel® Xeon® W-2145 3.7 Ghz. The computation time varies with the vowel geometries, and from 30 kHz onward, it increases exponentially.

Fig. 13b shows the distribution of the computation time over the different tasks required to solve the wave problem for the vowel /i/ (vowels /a/ and /u/ show similar trends). They consist in the computation of the transverse modes, the junction matrices (Eq. (44)), the radiation impedance matrix (Eq. (53)), the propagation of the physical quantities ($\mathcal{Z}$, $\mathcal{Y}$, p and q) and the computation of the radiated field (Eq. (54)). The computation time for the propagation is split into the computation of the matrix exponential in Eq. (48) and the other parts of the task since this part of the process requires a lot of computation time. Most of the computation time is dedicated to the computation of the transverse modes and the matrix exponential. When few modes are used, the computation time is mainly spent in the computation of the modes. As the number of modes used is increased, the time spent in the computation of the matrix exponential increases and becomes the task occupying most of the computation time. Note that the computation of the transverse modes, the junction matrices and the radiation impedance are done only at the beginning of the whole process. Thus, the time required for these tasks is not affected by the number of frequencies at which the transfer function is computed. A maximal cutoff frequency of 40 kHz appears to be a good compromise between the accuracy and the computation time.

For frequency domain FEM, 10 hours were necessary to compute 1000 frequencies with a single processor [10]. However, in this example a simplifying assumption was made for the radiation and the external space was not meshed, which reduces the computation time in comparison to a more accurate simulation. Thus, it can be concluded that the proposed method provides simulations with the same degree of accuracy as FEM, but with a computation time reduced by about a factor 10. The implementation used can be optimised to achieve even shorter computation times. One can also mention that it requires far less exchange memory, and it is thus easier to run on a standard laptop without special equipment for simulations (e.g. additional memory).

V. CONCLUSION

The proposed method can simulate curved vocal tract geometries with realistic cross-sectional shapes and cross-sectional area variations within the segments. The same degree of accuracy as FEM is achieved for the computation of transfer functions of a single segment (about 0.2% of differences in the resonance frequencies). Accounting for the curvature and the cross-sectional area variation substantially improves the agreement with FEM, particularly above 4-5 kHz. The segmentation process of realistic 3D vocal-tract geometries introduces small differences, but the agreement with FEM simulations remains very good (less than 1% of difference in the resonance frequencies). A good trade-off between

accuracy and computation time is achieved using a mesh density of 15, about 10 integration points per centimeter and a maximal cutoff frequency of 40 kHz. With these parameters, the proposed method achieves computation times about 10 times shorter than FEM while preserving the same degree of accuracy. The proposed method can be used to synthesize articulated sounds by simulating a succession of static geometries corresponding to various instants of the articulation motion and following the approach proposed by Sondhi and Schroeter [49].

APPENDIX

A. ANALYTICAL EXPRESSIONS OF THE TRANSVERSE MODES AND PROPAGATION MATRICES FOR A RECTANGULAR SHAPE

1) TRANSVERSES MODES

The transverse modes corresponding to a rectangular shape of side lengths a and b along y and z respectively, can be expressed as

$$
\varphi_{mn}(y, z) = \sqrt{\frac{2 - \delta_{0m}}{a}} \cos\left(\frac{m\pi y}{a}\right)
$$
$$
\times \sqrt{\frac{2 - \delta_{0n}}{b}} \cos\left(\frac{n\pi z}{b}\right), \qquad (60)
$$

and the corresponding eigenvalues are

$$
\gamma_{mn}^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2. \qquad (61)
$$

2) MATRIX C

The terms of the matrix C are

$$
C_{ij,mn} = \int_S z\varphi_{ij}\varphi_{mn}dS = C_y C_z, \qquad (62)
$$

where C_y and C_z are the integrals over y and z respectively.

$$
C_y = \delta_{im} \qquad (63)
$$

$$
C_z = \begin{cases} \dfrac{b}{2} & \text{if } j = n, \\[2mm] \dfrac{b}{\pi^2}\left(\dfrac{\cos((j+n)\pi) - 1}{(j+n)^2} + \dfrac{\cos((j-n)\pi) - 1}{(j-n)^2}\right) \\ & \text{if } j \neq n \neq 0, \\[2mm] \dfrac{\sqrt{2}b}{(n\pi)^2}(\cos(n\pi) - 1) & \text{if } j = 0 \text{ and } n \neq 0, \\[2mm] \dfrac{\sqrt{2}b}{(j\pi)^2}(\cos(j\pi) - 1) & \text{if } j \neq 0 \text{ and } n = 0. \end{cases}
$$
$$
\qquad (64)
$$

3) MATRIX D

The terms of the matrix D are

$$
D_{ij,mn} = \int_S z(\partial_y\varphi_{ij}\partial_y\varphi_{mn} + \partial_z\varphi_{ij}\partial_z\varphi_{mn})dS
$$
$$
= D_{1y}D_{1z} + D_{2y}D_{2z}, \qquad (65)
$$

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE *Access*

where

$$D_{1y} = \begin{cases} \left(\dfrac{i\pi}{a}\right)^2 & \text{if } i = m \neq 0, \\ 0 & \text{if } i \neq m \text{ or } i = m = 0, \end{cases} \quad (66)$$

D_{1z} has the same expression as C_z of the matrix $\boldsymbol{C}$ (Eq. (64)),
D_{2y} has the same expression as C_y of the matrix $\boldsymbol{C}$ (Eq. (63)),

$$D_{2z} = \begin{cases} 0 & \text{if } j = 0 \text{ and } n = 0, \\ \dfrac{jn}{b}\left(\dfrac{\cos((j-n)\pi) - 1}{(j-n)^2} - \dfrac{\cos((j+n)\pi) - 1}{(j+n)^2}\right) \\ \qquad \text{if } j \neq n \neq 0, \\ \dfrac{(j\pi)^2}{2b} & \text{if } j = n \neq 0. \end{cases} \quad (67)$$

4) MATRIX E

The terms of the matrix $\boldsymbol{E}$ are

$$E_{ij,mn} = \int_S \left(y\varphi_{ij}\partial_y\varphi_{mn} + z\varphi_{ij}\partial_z\varphi_{mn}\right) dS$$
$$= E_{1y}E_{1z} + E_{2y}E_{2z}, \quad (68)$$

where

$$E_{1y} = \begin{cases} 0 & \text{if } m = 0, \\ \sqrt{2}\cos(m\pi) & \text{if } i = 0 \text{ and } m \neq 0, \\ \dfrac{1}{2} & \text{if } i = m \neq 0, \\ m\left(\dfrac{\cos((i+m)\pi)}{i+m} - \dfrac{\cos((i-m)\pi)}{i-m}\right) \\ \qquad \text{if } i \neq m \neq 0 \end{cases} \quad (69)$$

$$E_{1z} = \begin{cases} 1 & \text{if } j = n, \\ 0 & \text{if } j \neq n, \end{cases} \quad (70)$$

$$E_{2y} = \begin{cases} 1 & \text{if } i = m, \\ 0 & \text{if } i \neq m, \end{cases} \quad (71)$$

$$E_{2z} = \begin{cases} 0 & \text{if } n = 0, \\ \sqrt{2}\cos(n\pi) & \text{if } j = 0 \text{ and } n \neq 0, \\ \dfrac{1}{2} & \text{if } j = n \neq 0, \\ n\left(\dfrac{\cos((j+n)\pi)}{j+n} - \dfrac{\cos((j-n)\pi)}{j-n}\right) \\ \qquad \text{if } j \neq n \neq 0. \end{cases} \quad (72)$$

5) MATRIX K^{R2}

The terms of the matrix $\boldsymbol{K}^{R2}$ are

$$K_{ij,mn}^{R2} = (1 + (-1)^i(-1)^m)\frac{\sqrt{1 - \delta_{0i}}\sqrt{1 - \delta_{0m}}}{b}\delta_{jn}$$
$$+ (1 + (-1)^j(-1)^n)\frac{\sqrt{1 - \delta_{0j}}\sqrt{1 - \delta_{0n}}}{b}\delta_{im} \quad (73)$$

B. MODAL PROJECTION

Both sides of Eq. (16) are multiplied by $\bar{\varphi}_m$ and p and q are decomposed using the expression in Eq. (18). This yields for the first line of Eq. (16)

$$\bar{\varphi}_m\partial_x\varphi_n p_n = \bar{\varphi}_m\frac{l'}{l}\boldsymbol{v}\cdot\boldsymbol{\nabla}_\perp\varphi_n p_n$$
$$+ \frac{1}{l^2}(\bar{\varphi}_m\varphi_n - \bar{\varphi}_m\kappa lz\varphi_n)q_n. \quad (74)$$

Integrating over S and identifying Eq. (19) yields

$$\delta_{mn}\partial_x p_n = \frac{l'}{l}\langle\varphi_m\boldsymbol{v}, \boldsymbol{\nabla}_\perp\varphi_n\rangle p_n$$
$$+ \frac{1}{l^2}(\boldsymbol{I} - \kappa l\langle\varphi_m z, \varphi_n\rangle), \quad (75)$$

where the expressions Eqs. (25) and (26) of the matrices $\boldsymbol{E}$ and $\boldsymbol{C}$ can be identified

$$\partial_x\boldsymbol{p} = \frac{l'}{l}\boldsymbol{E}\boldsymbol{p} + \frac{1}{l^2}(\boldsymbol{I} - \kappa l\boldsymbol{C})\boldsymbol{q}. \quad (76)$$

The second line of Eq. (16) is

$$\partial_x q = -\text{div}_\perp(f\boldsymbol{\nabla}_\perp p) - f(kl)^2 p + \frac{l'}{l}\text{div}_\perp(\boldsymbol{v}q). \quad (77)$$

Multiplying by $\bar{\varphi}_m$ yields

$$\bar{\varphi}_m\partial_x q = -\bar{\varphi}_m\text{div}_\perp(f\boldsymbol{\nabla}_\perp p)$$
$$- \bar{\varphi}_m f(kl)^2 p + \bar{\varphi}_m\frac{l'}{l}\text{div}_\perp(\boldsymbol{v}q). \quad (78)$$

Using the properties of the divergence operator, it can be expressed as

$$\bar{\varphi}_m\partial_x q = -\text{div}_\perp(\bar{\varphi}_m f\boldsymbol{\nabla}_\perp p)$$
$$+ (\boldsymbol{\nabla}_\perp\bar{\varphi}_m)\cdot(f\boldsymbol{\nabla}_\perp p) - \bar{\varphi}_m f(kl)^2 p$$
$$+ \frac{l'}{l}(\text{div}_\perp(\bar{\varphi}_m\boldsymbol{v}q) - (\boldsymbol{\nabla}_\perp\bar{\varphi}_m)\cdot(\boldsymbol{v}q)). \quad (79)$$

Integrating over S and applying the divergence theorem yields

$$\delta_{mn}q_n = \langle\boldsymbol{\nabla}_\perp\varphi_m, \boldsymbol{\nabla}_\perp\varphi_n\rangle p_n - \kappa l\langle\boldsymbol{\nabla}_\perp\varphi_m z, \boldsymbol{\nabla}_\perp\varphi_n\rangle p_n$$
$$- (kl)^2\delta_{mn}p_n + \kappa l(kl)^2\langle\varphi_m z, \varphi_n\rangle p_n$$
$$- \frac{l'}{l}\langle\boldsymbol{\nabla}_\perp\varphi_m, \boldsymbol{v}\varphi_n\rangle q_n$$
$$+ \int_\Gamma\left(\frac{l'}{l}\bar{\varphi}_m\boldsymbol{v}q - \bar{\varphi}_m f\boldsymbol{\nabla}_\perp p\right)\cdot l\boldsymbol{n}_\perp d\Gamma, \quad (80)$$

where Γ is the contour. The normal must be written as $l\boldsymbol{n}_\perp$ in the contour integral because the other vectors are expressed as a function of the unit vectors $\boldsymbol{n}_y$ and $\boldsymbol{n}_z$ which are normalized in the Cartesian space (X, Y, Z) whereas the normal $\boldsymbol{n}$ is normalized in the transformed space (x, y, z), and $||\boldsymbol{n}_\perp||_X = l||\boldsymbol{n}_\perp||$. Considering that the contour is the same all over the waveguide, the component n_x of the normal $\boldsymbol{n}$ to the boundary is 0 and $\boldsymbol{n} = \boldsymbol{n}_\perp$. The expression of the boundary condition Eq. (17) can be identified in the last integral, and the last integral term of Eq. (80) becomes

$$jkl\zeta_n K_{mn}^{R2} = jkl\zeta_n\int_\Gamma\bar{\varphi}_m\varphi_n d\Gamma. \quad (81)$$

IEEE Access

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

Identifying the expressions of the matrices $\boldsymbol{K}^2$, $\boldsymbol{C}$, $\boldsymbol{D}$ and $\boldsymbol{E}$ defined in Eqs. (27), (26), (29) and (25) yields

$$\partial_x \boldsymbol{q} = \left(\boldsymbol{K}^2 + \kappa l \left((kl)^2 \boldsymbol{C} - \boldsymbol{D} \right) \right) \boldsymbol{p} - \frac{l'}{l} \boldsymbol{E}^t \boldsymbol{q}. \qquad (82)$$

C. CONNECTION OF THE SEGMENTS

The condition of continuity of the acoustic pressure is

$$p_a = p_b \qquad (83)$$

on the common surface S_a. Multiplying both sides of this equation by $\boldsymbol{\varphi_a}$ and integrating over S_a yield

$$\int_{S_a} \boldsymbol{\varphi_a} p_a dS = \int_{S_a} \boldsymbol{\varphi_a} p_b dS, \qquad (84)$$

in which the expression of the modal amplitude $\boldsymbol{p_a} = \int_{S_a} \boldsymbol{\varphi_a} p_a dS$ can be identified. In order to integrate p_b one must operate a change of variable from the transformed coordinate (x_a, y_a, z_a) to the Cartesian coordinates (X, Y, Z). On the other hand, $p_b = \boldsymbol{\varphi_b}^t \boldsymbol{p_b}$, which yield

$$\boldsymbol{p_a} = \int_{S_{aX}} \boldsymbol{\varphi_a} \left(\frac{Y}{l_a}, \frac{Z}{l_a} \right) \boldsymbol{\varphi_b}^t \left(\frac{Y}{l_b}, \frac{Z}{l_b} \right) \frac{dYdZ}{l_a^2} \boldsymbol{p_b}, \qquad (85)$$

where the expression of the matrix $\boldsymbol{F}$ (Eq. (44)) can be identified and Eq. (42) is obtained.

Multiplying q_b by $\boldsymbol{\varphi_b}$ and integrating over S_{bX} in the Cartesian coordinate system (X, Y, Z) yield

$$\int_{S_{bX}} \boldsymbol{\varphi_b} q_b \frac{dYdZ}{l_b^2}$$
$$= \int_{S_{bX} - S_{aX}} \boldsymbol{\varphi_b} q_b \frac{dYdZ}{l_b^2} + \int_{S_{aX}} \boldsymbol{\varphi_b} q_b \frac{dYdZ}{l_b^2}. \qquad (86)$$

Identifying the expression of the modal amplitude vector $\boldsymbol{p_b} = \int_{S_b} \boldsymbol{\varphi_b} p_b dS$, applying the boundary condition on the surface $S_{bX} - S_{aX}$, $q_b = 0$, applying the condition of continuity of q on the surface S_{aX}, $q_b = q_a$ and expressing p_a as $p_a = \boldsymbol{\varphi_a}^t \boldsymbol{p_a}$ yield

$$\boldsymbol{q_b} = \int_{S_{aX}} \boldsymbol{\varphi_b} \left(\frac{Y}{l_b}, \frac{Z}{l_b} \right) \boldsymbol{\varphi_a}^t \left(\frac{Y}{l_a}, \frac{Z}{l_a} \right) \frac{dYdZ}{l_b^2} \boldsymbol{q_a}, \qquad (87)$$

where the expression of the matrix $\boldsymbol{F}$ (Eq. (44)) can be identified and Eq. (43) is obtained. The Eqs. (45) and (46) are obtained by combining Eqs. (42) and (43).

REFERENCES

[1] M. Arnela, R. Blandin, S. Dabbaghchian, O. Guasch, F. Alías, X. Pelorson, A. Van Hirtum, and O. Engwall, "Influence of lips on the production of vowels based on finite element simulations and experiments," *J. Acoust. Soc. Amer.*, vol. 139, no. 5, pp. 2852–2859, May 2016, doi: 10.1121/1.4950698.

[2] M. Arnela, S. Dabbaghchian, R. Blandin, O. Guasch, O. Engwall, A. Van Hirtum, and X. Pelorson, "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds," *J. Acoust. Soc. Amer.*, vol. 140, no. 3, pp. 1707–1718, Sep. 2016, doi: 10.1121/1.4962488.

[3] M. Arnela and O. Guasch, "Finite element synthesis of diphthongs using tuned two-dimensional vocal tracts," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 10, pp. 2013–2023, Oct. 2017, doi: 10.1109/TASLP.2017.2735179.

[4] M. Arnela and O. Guasch, "Finite element simulation of (ASA) in a three-dimensional vocal tract using a simplified aeroacoustic source model," in *Proc. 23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019, pp. 1–8.

[5] M. Arnela, S. Dabbaghchian, O. Guasch, and O. Engwall, "MRI-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 12, pp. 2173–2182, Dec. 2019, doi: 10.1109/TASLP.2019.2942439.

[6] I. Bartoli, A. Marzani, F. L. D. Scalea, and E. Viola, "Modeling wave propagation in damped waveguides of arbitrary cross-section," *J. Sound Vibrat.*, vol. 295, no. 3, pp. 685–707, Aug. 2006, doi: 10.1016/j.jsv.2006.01.021.

[7] D. Beautemps, P. Badin, and G. Bailly, "Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling," *J. Acoust. Soc. Amer.*, vol. 109, no. 5, pp. 2165–2180, Jun. 2001, doi: 10.1121/1.1361090.

[8] P. Birkholz, "Modeling consonant-vowel coarticulation for articulatory speech synthesis," *PLoS ONE*, vol. 8, no. 4, Apr. 2013, Art. no. e60603, doi: 10.1371/journal.pone.0060603.

[9] P. Birkholz, "Enhanced area functions for noise source modeling in the vocal tract," in *Proc. 10th Int. Seminar Speech Prod.*, Cologne, Germany, May 2014, pp. 1–4.

[10] P. Birkholz, S. Kürbis, S. Stone, P. Häsner, R. Blandin, and M. Fleischer, "Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties," *Sci. Data*, vol. 7, no. 1, pp. 1–16, Aug. 2020, doi: 10.6084/m9.figshare.12591554.

[11] R. Blandin, M. Arnela, R. Laboissière, X. Pelorson, O. Guasch, A. V. Hirtum, and X. Laval, "Effects of higher order propagation modes in vocal tract like geometries," *J. Acoust. Soc. Amer.*, vol. 137, no. 2, pp. 832–843, Feb. 2015, doi: 10.1121/1.4906166.

[12] R. Blandin, A. Van Hirtum, X. Pelorson, and R. Laboissière, "The effect on vowel directivity patterns of higher order propagation modes," *J. Sound Vibr.*, vol. 432, pp. 621–632, Oct. 2018, doi: 10.1016/j.jsv.2018.06.053.

[13] R. Blandin, A. Van Hirtum, X. Pelorson, and R. Laboissière, "Multimodal radiation impedance of a waveguide with arbitrary cross-sectional shape terminated in an infinite baffle," *J. Acoust. Soc. Amer.*, vol. 145, no. 4, pp. 2561–2564, Apr. 2019, doi: 10.1121/1.5099262.

[14] R. Blandin, S. Félix, and J. B. Doc, and P. Birkholz, "Combining multimodal and 2D finite elements for the efficient simulation of vocal tract acoustics," in *Proc. 27th Int. Congr. Sound Vib.*, Jul. 2021, pp. 1–8.

[15] R. Blandin, M. Arnela, S. Félix, and J. B. Doc andP Birkholz, "Comparison of the finite element method, the multimodal method and the transmission-line model for the computation of vocal tract transfer functions," in *Proc. 22nd Conf. Int. Speech Commun. Assoc.*, Brno, Czechia, Aug. 2021, pp. 1–5.

[16] M. Brandner, R. Blandin, M. Frank, and A. Sontacchi, "A pilot study on the influence of mouth configuration and torso on singing voice directivity," *J. Acoust. Soc. Amer.*, vol. 148, no. 3, pp. 1169–1180, Sep. 2020, doi: 10.1121/10.0001736.

[17] A. M. Bruneau, M. Bruneau, P. Herzog, and J. Kergomard, "Boundary layer attenuation of higher order modes in waveguides," *J. Sound Vib.*, vol. 119, no. 1, pp. 15–27, Nov. 1987, doi: 10.1016/0022-460X(87)90186-6.

[18] S. Dabbaghchian, M. Arnela, O. Engwall, and O. Guasch, "Simulation of vowel-vowel utterances using a 3D biomechanical-acoustic model," *Int. J. Numer. Methods Biomed. Eng.*, vol. 37, no. 1, pp. 1–20, Oct. 2020, doi: 10.1002/cnm.3407.

[19] J. O. Dow and D. E. Byrd, "The identification and elimination of artificial stiffening errors in finite elements," *Int. J. Numer. Methods Eng.*, vol. 26, no. 3, pp. 743–762, Mar. 1988, doi: 10.1002/nme.1620260316.

[20] S. El-Masri, X. Pelorson, P. Saguet, and P. Badin, "Development of the transmission line matrix method in acoustics applications to higher modes in the vocal tract and other complex ducts," *Int. J. Numer. Model., Electron. Netw. Devices Fields*, vol. 11, no. 3, pp. 133–151, Dec. 1998, doi: 10.1002/(SICI)1099-1204(199805/06)11:3<133::AID-JNM298>3.0.CO;2-H.

[21] S. Félix, A. Maurel, and J.-F. Mercier, "Improved multimodal methods for the acoustic propagation in waveguides with finite wall impedance," *Wave Motion*, vol. 54, pp. 1–10, Apr. 2015, doi: 10.1016/j.wavemoti.2014.11.007.

[22] S. Félix, J.-B. Doc, and M. A. Boucher, "Modeling of the multimodal radiation from an open-ended waveguide," *J. Acoust. Soc. Amer.*, vol. 143, no. 6, pp. 3520–3528, Jun. 2018, doi: 10.1121/1.5041268.

[23] S. Fels, F. Vogt, K. Van Den Doel, J. Lloyd, and I. Stavness, and E. Vatikiotis-Bateson, "Artisynth: A biomechanical simulation platform for the vocal tract and upper airway," in *Proc. 7th Int. Seminar Speech Prod.*, Ubatuba, Brazi, Dec. 2006, pp. 1–7.

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

IEEE *Access*

[24] M. Fleischer, S. Pinkert, W. Mattheus, A. Mainka, and D. Mürbe, "Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall," *Biomech. Model. Mechanobiol.*, vol. 14, no. 4, pp. 719–733, Aug. 2015, doi: 10.1007/s10237-014-0632-2.

[25] M. Fleischer, A. Mainka, S. Kürbis, and P. Birkholz, "How to precisely measure the volume velocity transfer function of physical vocal tract models by external excitation," *PLoS ONE*, vol. 13, no. 3, Mar. 2018, Art. no. e0193708, doi: 10.1371/journal.pone.0193708.

[26] L. Gavric, "Computation of propagative waves in free rail using a finite element technique," *J. Sound Vib.*, vol. 185, no. 3, pp. 531–543, Aug. 1995, doi: 10.1006/jsvi.1995.0398.

[27] J. H. Ginsberg, "Application of collocation and orthogonality-based techniques to satisfy junction conditions in an arbitrary segmented waveguide," *J. Acoust. Soc. Amer.*, vol. 145, no. 4, pp. 2022–2031, Apr. 2019, doi: 10.1121/1.5094777.

[28] U. G. Goldstein, "An articulatory model for vocal tracts growing children," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 1980.

[29] O. Guasch, M. Arnela, R. Codina, and H. Espinoza, "A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production," *Acta Acustica United Acustica*, vol. 102, no. 1, pp. 94–106, Jan. 2016, doi: 10.3813/AAA.918927.

[30] T. Guennoc, J.-B. Doc, and S. Félix, "Improved multimodal formulation of the wave propagation in a 3D waveguide with varying cross-section and curvature," *J. Acoust. Soc. Amer.*, vol. 149, no. 1, pp. 476–486, Jan. 2021, doi: 10.1121/10.0003336.

[31] A. J. Gully, H. Daffern, and D. T. Murphy, "Diphthong synthesis using the dynamic 3D digital waveguide mesh," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 2, pp. 243–255, Feb. 2018, doi: 10.1109/TASLP.2017.2774921.

[32] A. J. Gully and B. Tucker, "Modeling voiced stop consonants using the 3D dynamic digital waveguide mesh vocal tract model," in *Proc. 19th Int. Congr. Phonetic Sci.*, Melbourne, VIC, Australia, Aug. 2019, pp. 1–6.

[33] T. Hayashi, W.-J. Song, and J. L. Rose, "Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example," *Ultrasonics*, vol. 41, pp. 175–183, May 2003, doi: 10.1016/S0041-624X(03)00097-0.

[34] Y. Kagawa, R. Shimoyama, T. Yamabuchi, and T. Murai, and K. Takarada, "Boundary element models of the vocal tract and radiation field and their response characteristics," *J. Sound Vib.*, vol. 157, no. 3, pp. 385–403, Sep. 1992, doi: 10.1016/0022-460X(92)90523-Z.

[35] S. Kirkup, "The boundary element method in acoustics: A survey," *Appl. Sci.*, vol. 9, no. 8, p. 1642, Apr. 2019, doi: 10.3390/app9081642.

[36] Y. Laprie, M. Loosvelt, S. Maeda, R. Sock, and F. Hirsch, "Articulatory copy synthesis from cine X-ray films," in *Proc. 14th Conf. Int. Speech Commun. Assoc.*, Lyon, France, Aug. 2013, pp. 1–6.

[37] M. G. Larson and F. Bengzon, *The Finite Element Method: Theory, Implementation, and Practice*. Berlin, Germany: Springer-Verlag, 2010.

[38] P. W. Loveday, "Semi-analytical finite element analysis of elastic waveguides subjected to axial loads," *Ultrasonics*, vol. 49, pp. 298–300, Mar. 2009, doi: 10.1016/j.ultras.2008.10.018.

[39] A. Maurel, J.-F. Mercier, and S. Félix, "Propagation in waveguides with varying cross section and curvature: A new light on the role of supplementary modes in multi-modal methods," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 470, no. 2166, Jun. 2014, Art. no. 20140008, doi: 10.1098/rspa.2014.0008.

[40] K. Motoki, N. Miki, and N. Nagai, "Measurement of sound-pressure distribution in replicas of the oral cavity," *J. Acoust. Soc. Amer.*, vol. 92, no. 5, pp. 2577–2585, Nov. 1992, doi: 10.1121/1.404430.

[41] K. Motoki, "Three-dimensional acoustic field in vocal-tract," *Acoust. Sci. Technol.*, vol. 23, no. 4, pp. 207–212, 2002, doi: 10.1250/ast.23.207.

[42] K. Motoki, "Three-dimensional rectangular vocal-tract model with asymmetric wall impedances," in *Proc. 14th Conf. Int. Speech Commun. Assoc.*, Lyon, France, Aug. 2013, pp. 138–142.

[43] K. Motoki, "Effects of wall impedance on transmission and attenuation of higher-order modes in vocal-tract model," in *Proc. 11th Conf. Int. Speech Commun. Assoc.*, Makuhari, Japan, Sep. 2010, pp. 1013–1016.

[44] A. N. Norris, "Acoustic cloaking theory," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 464, no. 2097, pp. 2411–2434, Apr. 2008, doi: 10.1098/rspa.2008.0076.

[45] V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation," *J. Acoust. Soc. Amer.*, vol. 100, no. 4, pp. 2034–2048, Oct. 1996, doi: 10.1121/1.417913.

[46] V. Pagneux, "Multimodal admittance method in waveguides and singularity behavior at high frequencies," *J. Comput. Appl. Math.*, vol. 234, no. 6, pp. 1834–1841, Jul. 2010, doi: 10.1016/j.cam.2009.08.034.

[47] A. Pelat, S. Felix, and V. Pagneux, "A coupled modal-finite element method for the wave propagation modeling in irregular open waveguides," *J. Acoust. Soc. Amer.*, vol. 129, no. 3, pp. 1240–1249, Mar. 2011, doi: 10.1121/1.3531928.

[48] C. H. Shadle and R. I. Damper, "Prospects for articulatory synthesis: A position paper," in *Proc. 4th ISCA Tutorial Res. Workshop (ITRW) Speech Synthesis*, 2001, pp. 121–126.

[49] M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. Audio Speech Language Process.*, vol. ASLP-35, no. 7, pp. 955–967, Jul. 1987, doi: 10.1109/TASSP.1987.1165240.

[50] B. H. Story, "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," *J. Acoust. Soc. Amer.*, vol. 123, no. 1, pp. 327–335, Jan. 2008, doi: 10.1121/1.2805683.

[51] B. H. Story, "Phrase-level speech simulation with an airway modulation model of speech production," *Comput. Speech Lang.*, vol. 27, no. 4, pp. 989–1010, Jun. 2013, doi: 10.1016/j.csl.2012.10.005.

[52] H. Takemoto, K. Honda, S. Masaki, and Y. I. S. Fujimoto, "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 1037–1049, Jan. 2006, doi: 10.1121/1.2151823.

[53] H. Takemoto, P. Mokhtari, and T. Kitamura, "Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method," *J. Acoust. Soc. Amer.*, vol. 128, no. 6, pp. 3724–3738, Dec. 2010, doi: 10.1121/1.3502470.

[54] T. Vampola, J. Horáček, and J. G. Švec, "Modeling the influence of piriform sinuses and valleculae on the vocal tract resonances and antiresonances," *Acta Acustica United Acustica*, vol. 101, no. 3, pp. 594–602, May 2015, doi: 10.3813/AAA.918855.

[55] T. Vampola, J. Horáček, V. Radolf, J. G. Švec, and A.-M. Laukkanen, "Influence of nasal cavities on voice quality: Computer simulations and experiments," *J. Acoust. Soc. Amer.*, vol. 148, no. 5, pp. 3218–3231, Nov. 2020, doi: 10.1121/10.0002487.

[56] T. Yoshinaga, A. Van Hirtum, and S. Wada, "Multimodal modeling and validation of simplified vocal tract acoustics for sibilant/s," *J. Sound Vibrat.*, vol. 411, pp. 247–259, Dec. 2017, doi: 10.1016/j.jsv.2017.09.004.

RÉMI BLANDIN was born in Pau, France, in 1989. He received the master's degree in acoustics from the University Pierre and Marie Curie, Paris, France, in 2013, and the Ph.D. degree in acoustics from the University of Grenoble, France, in 2016.

Since 2019, he has been working as a Postdoctoral Researcher with the Dresden University of Technology (TU Dresden). His research interests include waveguide acoustics, semi-analytical simulation methods, speech production, articulatory synthesis, and speech directivity.

MARC ARNELA received the degree in telecommunications engineering, with a specialization on image and sound, the M.Sc. degree in networks and telecommunications engineering, and the Ph.D. degree in acoustic engineering from La Salle-Universitat Ramon Llull (URL), Barcelona, Spain, in 2009, 2010, and 2015, respectively, and the professional degree in music education, with a specialization on piano, from the Conservatori Municipal de Música de Barcelona, in 2009. He is currently a Lecturer and a Researcher on acoustics with La Salle-URL, where he is also the Coordinator of the Acoustics Laboratory. His main research interests include the numerical production of voice using finite element methods, the design and experimental testing of new loudspeakers made of ultrasonic sensors, and the acoustics of ducts and mufflers with a special interest on the acoustic black hole effect.

IEEE *Access*

R. Blandin *et al.*: Efficient 3D Acoustic Simulation of the Vocal Tract by Combining the Multimodal Method and Finite Elements

SIMON FÉLIX received the M.S. degree in fundamental physics from Université Paris-Sud, France, and the Ph.D. degree in acoustics from Le Mans University. He is currently working as a Senior Researcher with the Laboratory of Acoustics, French National Centre for Scientific Research (CNRS), Institute of Acoustics—Graduate School, Le Mans, France. His research interests include wave propagation in inhomogeneous waveguides, or structured complex media whose heterogeneity, anisotropy, or regularity (symmetries, periodicity, and disorder) are used for wave control applications. The fields of application of this work cover various fields of wave physics and various problems, such as urban acoustics, musical acoustics, metamaterials, or wave control.

JEAN-BAPTISTE DOC received the Ph.D. degree in acoustics from the University of Le Mans, France, in 2012. He is currently an Associate Professor with the Laboratoire de Mécanique des Structures et des Systèmes Couplés, Conservatoire National des Arts et Métiers, Paris, France. His research interests include the modeling of complex shaped waveguide and the analysis of sound production mechanisms in wind instruments.

PETER BIRKHOLZ (Member, IEEE) was born in Rostock, Germany, in 1978. He received the Diploma degree in computer science and the Ph.D. degree (Hons.) in signal processing from the University of Rostock, Germany, in 2002 and 2005, respectively.

He worked as a Research Associate with the University of Rostock, from 2005 to 2009; and with the Department of Phoniatrics, Pedaudiology, and Communication Disorders, RWTH Aachen University, Germany, from 2009 to 2014. He became a Junior Professor for cognitive systems and a Full Professor for speech technology and cognitive systems with the Dresden University of Technology (TU Dresden), Germany, in 2014 and 2020, respectively. His research interests include speech production, articulatory speech synthesis, computational neuroscience, silent speech interfaces, and measurement techniques for speech research.

Dr. Birkholz was awarded the Joachim–Jungius Prize by the University of Rostock for his dissertation on articulatory speech synthesis, in 2006; and the Klaus-Tschira Award for achievements in public understanding of science, in 2006.

● ● ●