

Article

Not peer-reviewed version

Design of a Computational Procedure for Optimizing the Acoustic Outputs of the Human Vocal Tract

[Jaroslav Štorkán](#)^{*} and Tomáš Vampola

Posted Date: 22 September 2023

doi: 10.20944/preprints202309.1579.v1

Keywords: vocal tract; singer formant; finite element method; isoparametric element



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Design of a Computational Procedure for Optimizing the Acoustic Outputs of the Human Vocal Tract

Jaroslav Štorkán ^{1,*} and Tomáš Vampola ²

¹ Department of Mechanics, Biomechanics and Mechatronics, Faculty of Mechanical Engineering, Czech Technical University in Prague, Technická 4, 160 00 Prague 6, Czech Republic; jaroslav.storkan@fs.cvut.cz

² Department of Mechanics, Biomechanics and Mechatronics, Faculty of Mechanical Engineering, Czech Technical University in Prague, Technická 4, 160 00 Prague 6, Czech Republic; tomas.vampola@fs.cvut.cz

* Correspondence: jaroslav.storkan@fs.cvut.cz; Tel.: +420 721825942

Abstract: In order to examine the impact of geometric configuration of the human vocal tract on the redistribution of acoustic energy in the frequency spectrum, it is necessary to assemble computation models, which facilitate simple modification of the model's geometric characteristics. This article describes how the human vocal tract model was constructed using so-called super-elements using the improved reduced system method. The mentioned procedure allows using such a model for demanding optimization calculations, for which the model is modified on the super-element level. Furthermore, the article presents an alternative approach, in which an isoparametric element of a higher degree polynomial was derived. Both approaches were compared and used to find the geometric configuration of the human vocal tract amplifying acoustic energy values on the frequency spectrum to around 3kHz.

Keywords: vocal tract; singer formant; finite element method; isoparametric element

1. Introduction

The subject of interest of this work is explaining the impact of geometric configuration of the human vocal tract (hereinafter referred to as the VT) on the distribution of acoustic energy on the frequency spectrum of vocal speech. For voice teachers, it is necessary to analyse the procedures facilitating the impact of geometric configuration of the VT to amplify acoustic energy values in the required frequency range [1], [2]. In order to shed light on such modifications of human VT on the characteristics of vocal speech, it is necessary to assemble mathematical calculation models enabling effective examination of sensitivity on individual geometric changes in the VT [3], [4]. Assembling a fully parametric model at the level of describing boundary surfaces defining the examined acoustic system with regard to the complexity of the VT is a very demanding problem [5], [6], [7]. Currently, the FEM is almost exclusively used for analysing the acoustic characteristics of VT, the computation complexity of which increases with the number of finite elements used, which discredits the examined volume of the acoustic system. Direct geometric reconfiguration at the level of the finite element mesh represents a very complex problem, which requires a detailed analyses of the derived model [8]. However, due to the computation demands, such models are oftentimes not suitable for subsequent optimization tasks, whose aim it is to finetune the acoustic characteristics of phonation to the required parameters.

One of the possible approaches for assembling effective computation models is to use the super-elements method [9], [10], which allows for modifying the analysed acoustic system only at the point of the geometric change. To speed up the calculation, the mesh of individual elements is reduced at the level of mass, damping and stiffness matrices to a number of degrees of freedom (hereinafter referred to only as "DoF") ensuring compliance of the frequency characteristics of a full and reduced model in the frequency range, which is important for the comprehension of voice speech.

The article also presents an alternative, yet very effective approach to assembling an isoparametric element of a higher polynomial, using typical VT geometric configuration. Such an element even allows parallel acoustic cavities to be connected. Formulating the calculation model for

[a:] was subsequently used to examine the permissible geometric modifications of VT, aimed at connecting the 3rd and 4th formant and shifting it into the 2.5kHz – 3.5kHz frequency range. Bringing the formants closer together makes it possible to significantly increase the amount of acoustic energy in the field of the so-called singing formant.

2. Finite Element Model of the Vocal Tract

The VT geometry is unique for every individual. A human's ability to change the size and shape of the VT cross-section in different places, allows us to modulate the basic signal generated by the vocal cords and create different sounds. For the purpose of VT modelling, VT geometry [11] and [12] were measured. The CT LightSpeed VCT GE-64 was used for imaging. The measurement resulted in 181 sections, 0.625mm thick. The image resolution is 512x512 pixels. It was measured in the supine position, while sounding out [a:]. The geometric model was created by analysing sagittal CT cross-sections. The surface of the geometric model is used to create the FEM of the VT model (Figure 1). The geometric model contains the parallel cavities of the piriform sinuses and the epiglottic vallecula.

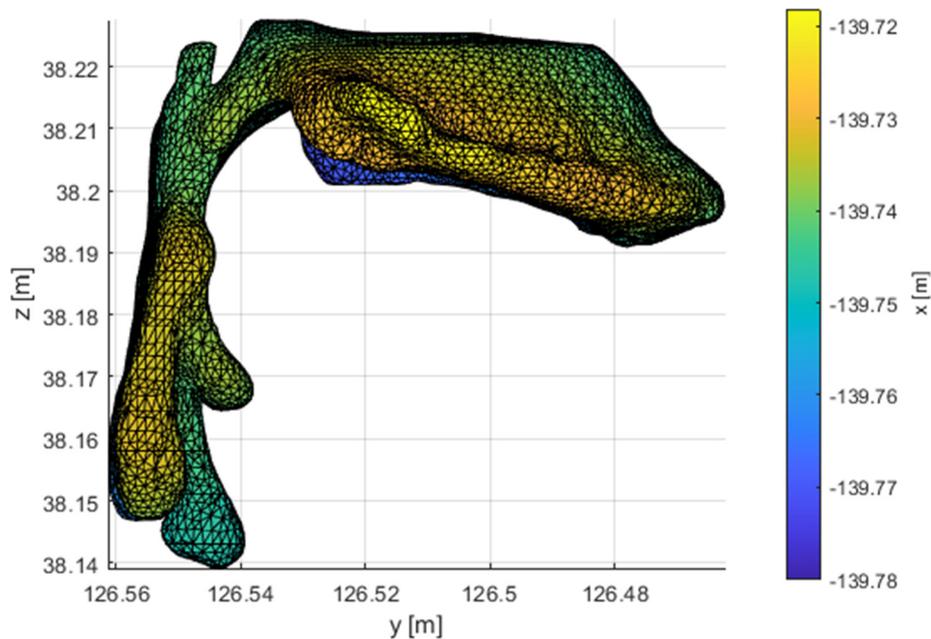


Figure 1. FEM mesh model for vowel [a:].

The formulated calculation model for [a:] was then used to examine the permissible geometric modifications of the VT, with the aim of connecting the 3rd and 4th formants and shifting them into the 2.5kHz – 3.5kHz frequency range. Bringing these formants closer together makes it possible to significantly increase the amount of acoustic energy in the field of the so-called singing formant.

The foundation of FEM is the discretization of the solved domain with the finite number of elements, which disjointly fill the entire solved volume of VT. The mathematical description of the elements is provided by the $N_e [-]$, shape functions, which distribute the nodal values of the unknown quantities into the element's entire volume. Multiplying by the column vector of the nodal values of the sound pressure of element $P_e [Pa]$ results in the following acoustic pressure distribution function in the volume of the element

$$p'(t, x, y, z) = N_e(x, y, z)P_e(t). \quad (1)$$

The finite element method uses a so-called weak formulation and the Galerkin Method [13]. The differential equation of the strong formulation is a scalarly multiplied by testing $\delta p'$ function and the resulting expression is integrated via the solved volume

$$\int_{V_e} \left(\delta p' \frac{1}{c_0^2} \frac{\partial^2 p'}{\partial t^2} + \delta p' \frac{r}{c_0^2 \rho_0} \frac{\partial p'}{\partial t} \right) dV = \int_{V_e} (\delta p' \Delta p') dV, \quad (2)$$

where $\rho_0 [kgm^{-3}]$ is the density of the environment, $c_0 [ms^{-1}]$ is the velocity of sound, $r [kgm^{-3}s^{-1}]$ is the volume damping coefficient, $V_e [m^3]$ is the element volume and $\Delta [m^{-2}]$ is the Laplace operator. The acoustic pressures and their variations expressed using nodal values and their values are put into the equation. The equation is converted into the frequency domain and Green's first identity is applied. The pressure gradient in the resulting surface integral can be expressed using acoustic velocity

$$\delta \mathbf{P}_e^T \int_{V_e} \left(-\omega^2 \frac{1}{c_0^2} \mathbf{N}_e^T \mathbf{N}_e + i\omega \mathbf{N}_e^T \frac{r}{c_0^2 \rho_0} \mathbf{N}_e + (\nabla \mathbf{N}_e)^T \nabla \mathbf{N}_e \right) dV \mathbf{P}_e = -\delta \mathbf{P}_e^T \int_{\partial V_e} (\mathbf{N}_e^T \mathbf{n}^T \rho_0 i\omega \mathbf{v}') dS, \quad (3)$$

where $\omega [s^{-1}]$ is the angular frequency, $\nabla [m^{-1}]$ is the nabla operator, $\mathbf{n} [-]$ is the unit normal to the boundary $\partial V_e [m^2]$ and $\mathbf{v}' [ms^{-1}]$ is the acoustic velocity. The expressions can be simplified as follows

$$\delta \mathbf{P}_e^T \left((-\omega^2 \mathbf{M}_e + i\omega \mathbf{B}_e + \mathbf{K}_e) \mathbf{P}_e - \mathbf{V}_e \right) = 0, \quad (4)$$

where the $\mathbf{M}_e [s^2m]$ is the element's mass matrix

$$\mathbf{M}_e = \frac{1}{c_0^2} \int_{V_e} (\mathbf{N}_e^T \mathbf{N}_e) dV, \quad (5)$$

the $\mathbf{B}_e [sm]$ is the element's damping matrix (the relationship between the specific acoustic impedance Z_s and acoustic velocity \mathbf{v}' was applied here)

$$\mathbf{B}_e = \frac{1}{c_0^2 \rho_0} \int_{V_e} (r \mathbf{N}_e^T \mathbf{N}_e) dV + \rho_0 \int_{\partial V_e} \left(\mathbf{N}_e^T \frac{1}{Z_s} \mathbf{N}_e \right) dS, \quad (6)$$

the $\mathbf{K}_e [m]$ is the element's stiffness matrix

$$\mathbf{K}_e = \int_{V_e} ((\nabla \mathbf{N}_e)^T (\nabla \mathbf{N}_e)) dV, \quad (7)$$

and $\mathbf{V}_e [kgs^2]$ is the vector of the right sides

$$\mathbf{V}_e = \int_{\partial V_e} (\mathbf{N}_e^T \mathbf{n}^T \rho_0 i\omega \mathbf{v}') dS. \quad (8)$$

Relationship (4) mathematically defines the acoustic properties of one element. After discretization of the space of finite elements, global matrices describing the whole system solved are then assembled.

Figure 1 illustrates the FEM mesh model of the VT for vowel [a:]. This model has a total of 36553 nodes. The mesh model is not structured and the types of elements used are quadratic tetrahedrons with ten nodes. There are 186570 elements in the model. The model is created using commercial software Ansys.

Air flow is not stationary due to vocal cord oscillations. The air flow volume is periodic and can be approximated using an analytical model. The Liljencrants-Fant model (L-F model) [14] is one of the basic models of air pulses. This model has four parameters. The first derivative of this pulse is in Figure 2. A velocity-type boundary condition is modelled at the level of the vocal cords.

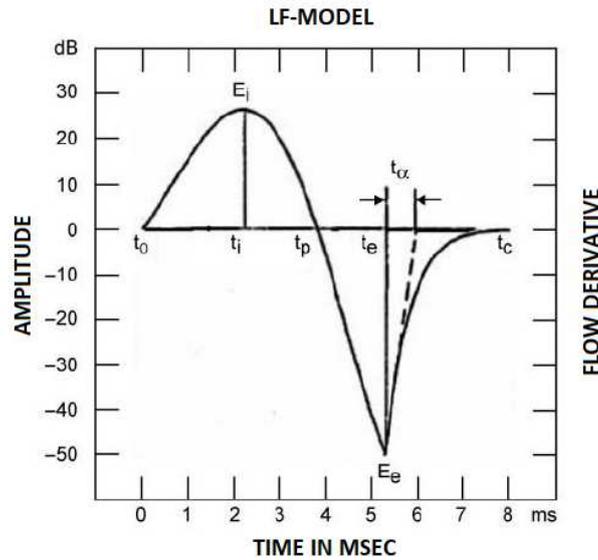


Figure 2. Significance of control parameters of the L-F model [14].

Acoustic energy is emitted by the mouth into the surrounding space. The external environment model is replaced by the emitted acoustic impedance. This is derived under the assumption of a circular oral opening, constant acoustic velocity throughout the oral opening and an infinite external environment [15]. In such a case, the specific acoustic impedance at the output of the VT can be defined by the following relationship.

$$Z_s(\omega) = c_0 \rho_0 \left(1 - \frac{J_1\left(\frac{2\omega R}{c_0}\right)}{\frac{\omega R}{c_0}} + i \frac{H_1\left(\frac{2\omega R}{c_0}\right)}{1 + \frac{\omega R}{c_0}} \right), \quad (9)$$

where J_1 [-] and H_1 [-] are the first order Bessel and Struve functions and R [m] is the radius of the considered oral opening. The disadvantage of this procedure is the non-linear frequency dependance of the acoustic impedance.

A disadvantage of the described FEM model is the large number of DoF. The dimensions of the model's mass, damping and stiffness matrices equal to the number of the DoF model. Although the matrices are symmetrical and sparse, their dimensions are generally so large (in this case: 36 553) that the model can be described as very computationally demanding. The computation demands depend on the type of analysis, the computer's performance and the performance optimization of the solver.

3. Aim of This Work

Calculating the sensitivity of acoustic properties on the geometric modifications of the VT or calculating the optimal geometry VT for achieving the required properties are important tasks for clinical practice. Such tasks are generally more demanding and are oftentimes solved using optimization approaches, which require ample calculations given various modification models. The number of analyses is then considerable, requiring more time for calculations.

The aim of this work is to optimize the model's computation demands, allowing complex optimization calculations to be done. The first procedure uses concept of super-elements obtained by reducing an extensive FEM model divided into smaller subsystems. The second method uses a specially designed element for FEM. This element is designed in such a way so that it enables the discretization of the VT geometry. The models assembled using this element have a very low number of DoF. They only contain the necessary nodes on the surface of the model for defining the geometry and nodes on the curve passing through the centre of the VT cross-section. Both procedures generate models with a similar number of DoF and node spacing.

4. IRS Reduction

The commonly used method for reducing the computing demands of large models are reductions at the level of the M and K matrices. The principle consists in decreasing the number of DoF. The resulting reduced model can be understood as one super-element. The basic reduction method is the Guyan Reduction Method [16], which is generalized by the IRS (improved reduced system) reduction.

IRS reduction is based on a model containing stiffness and mass matrices. The model's DoF is divided into those that are preserved, P_m and those that are reduced, P_s [17]

$$P = \begin{bmatrix} P_m \\ P_s \end{bmatrix}. \quad (10)$$

The mathematical model transforms into the following shape

$$\left(-\omega^2 \begin{bmatrix} M_{m,m} & M_{m,s} \\ M_{s,m} & M_{s,s} \end{bmatrix} + \begin{bmatrix} K_{m,m} & K_{m,s} \\ K_{s,m} & K_{s,s} \end{bmatrix} \right) \begin{bmatrix} P_m \\ P_s \end{bmatrix} = \begin{bmatrix} \hat{V}_m \\ \mathbf{0} \end{bmatrix}. \quad (11)$$

For the reduction, it is necessary to find map P in the P_m set. This is ensured by transformation matrix T_{IRS} by the following relationship

$$P = \begin{bmatrix} P_m \\ P_s \end{bmatrix} = T_{IRS} P_m. \quad (12)$$

Transformation matrix T_{IRS} is used to transform the whole mathematical description into the reduced DoF

$$K_{red} = T_{IRS}^T K T_{IRS}, \quad (13)$$

$$M_{red} = T_{IRS}^T M T_{IRS} \quad (14)$$

is defined by the following relationship [18]

$$T_{IRS} = T_s + S M T_s M_{red}^{-1} K_{red}, \quad (15)$$

where T_s is the transformation matrix of the static Guyan reduction

$$T_s = \begin{bmatrix} E \\ -K_{s,s}^{-1} K_{s,m} \end{bmatrix}, \quad (16)$$

and matrix S is defined by the following relationship

$$S = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & K_{s,s}^{-1} \end{bmatrix}. \quad (17)$$

The formula for T_{IRS} is implicit because it depends on the reduced system. Therefore, it has to be calculated iteratively. The first iteration step is solved using the Guyan reduction $T_{IRS,1} = T_s$. After obtaining the first reduced mass and stiffness matrices, iteration can continue using the IRS method

$$T_{IRS,i+1} = T_s + S M T_s M_{red,i}^{-1} K_{red,i}. \quad (18)$$

The convergence of this iteration scheme is accelerated if T_s is replaced by $T_{IRS,i}$ in the second part. The model's final reduction matrix is determined after the convergence. The shape of the model is then

$$(-\omega^2 M_{red} + K_{red}) P_m = V_m. \quad (19)$$

5. Properties of IRS Reduction

The derivation of IRS reduction [18] utilizes simplified assumptions, which cause errors in the calculated model. This error depends on the amount and location of the reduced and preserved nodes and on the number of iterations of the IRS method. Therefore, the IRS reduction effect test is conducted on the model's properties. The motivation to use reduction methods to modify models was to reduce computation demands, which allow optimal calculations to be carried out in

reasonable times. The expression for the IRS method transformation matrix is implicit and is solved iteratively, thus making the calculation more complex. Therefore, the convergence test of the IRS iteration scheme was conducted. This test is important for optimally selecting the number of iterations, which guarantee compromise between computation demands and the generated error.

The dominant frequency formants are used to study the importance of VT for forming vowels. From the acoustics viewpoint, the formant is a local frequency extreme in the voice spectrum, which is created by amplifying the acoustic pressure by transfer function of the VT between the vocal cords and the mouth. These local extremes are always in the area around eigenfrequency of the acoustic oscillations of the solved geometry. Therefore, the task then turns into finding eigenfrequencies.

For the purposes of this work, the geometric model of the measured VT had to be divided into subsequent volumes. The cross-sections are always planar and the resulting geometry is illustrated in Figure 3. The main VT cavity is made up of 16 volumes. Furthermore, the model contains another two parts forming the piriform sinuses and 4 parts that close the model. The vallecula are one part of the main cavity. All parts of the divided VT were reduced by the IRS method. The aim of the reduction was to only preserve the nodes at the intersection of the dividing planes with the VT surface. This reduced the total number of DoF from 36553 to 1154. Almost 97% of the model size is reduced. Figure 4 illustrates the size of the relative error of eigenfrequency models in relation to the number of reduced DoF. These graphs are used to evaluate the smallest eigenfrequencies for every VT part and every analysed degree of reduction. To reduce the amount of data in the image, the eigenfrequency with the worst result is always displayed. It can be seen that there is practically no error up to a reduction of 70%-80% DoF. In the case of only preserving the DoF on the perimeter of the dividing surface, the errors are in tenths of percent. Different degrees of reduction were achieved at different parts of the VT. This is given by the fact that the ratio between the number of preserved DoF and the total number of DoF is different. In some parts, only 86% of DoF were reduced and maximum reduction reached 97%.

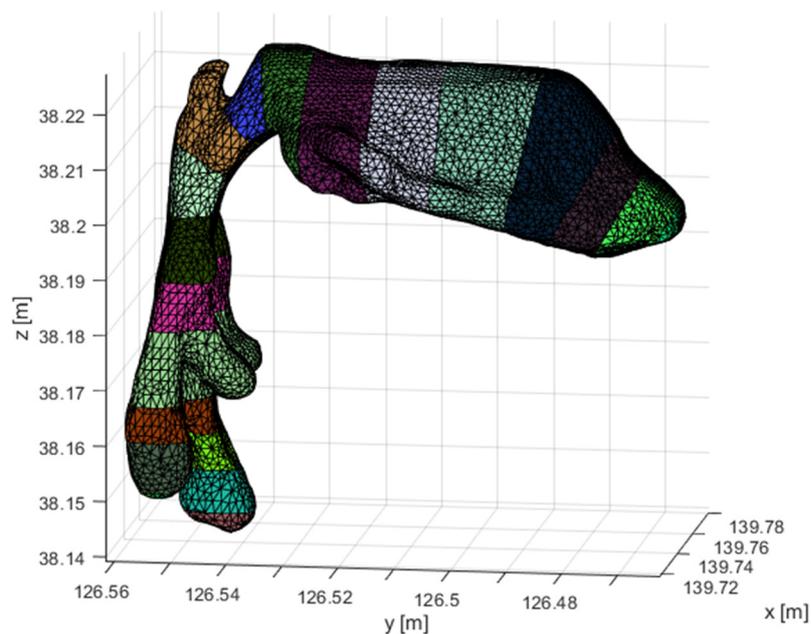


Figure 3. Vocal tract for vowel [a:] divided into subsystems.

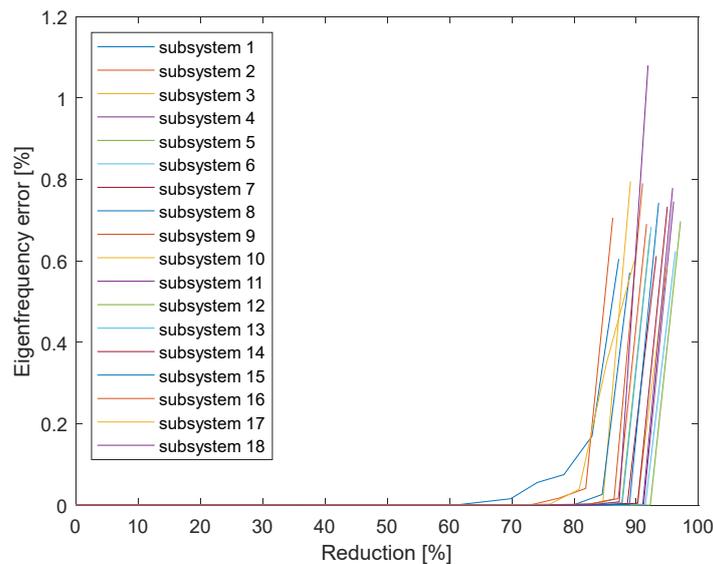


Figure 4. The impact of reduction on the eigenfrequencies in general geometry.

The computation times increase with the number of iterations. The optimal number of iterations was determined by a numerical experiment. The relative error of eigenfrequency was monitored depending on the number of iterations (Figure 5). The eigenfrequencies were re-evaluated and the worst result was always used for illustration purposes. The error is up to 68% after the first iteration and the average error is approximately 20%. The error decreases steeply with the number of iterations. The decrease decelerates after the fifth iteration. The error does not exceed 1.7% after the 10th iteration and the maximum error after the 20th iteration is 1.08%.

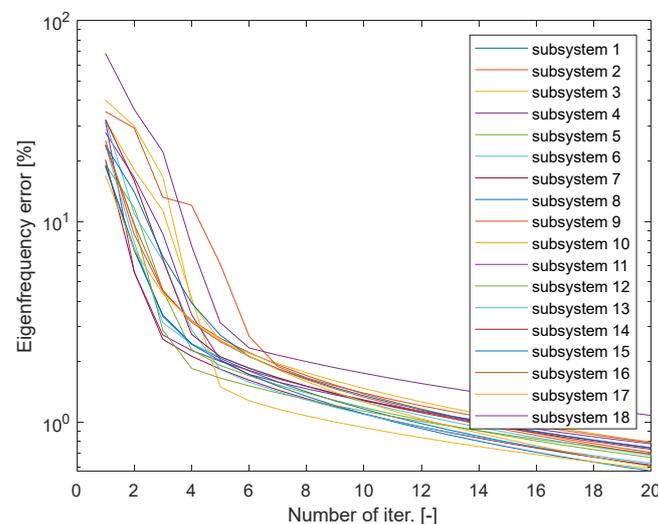


Figure 5. The impact of the number of IRS iterations on the eigenfrequencies.

The test conducted using the IRS method only evaluated eigenfrequency, showing a low distortion of results. After the IRS reduction, the model has a lower number of DoF and thereby even a lower number of eigenfrequencies and oscillation shapes. The IRS preserves the lowest frequency. The model preserves the lowest eigenfrequency even when it does not contain the nodes necessary to describe the respective oscillation shape.

Super-elements for all parts of the VT were created. According to Figure 3, this represented a total of 22 super-elements, based on which the entire VT calculation model was assembled.

Such a created model has 1154 nodes (Figure 6), which represents about 3% of the original model. The model's computation demands increase with its size faster than linearly. The processing time savings are high. In this case, no emitted acoustic impedance was used at the level of the mouth. This boundary condition was modelled as the null sound pressure. This can be interpreted as emission without any losses. The quality of the results is evaluated by comparing the values of eigenfrequencies. The frequencies of the reduced models are as follows

$$f = [600, 1721, 1904, 2704, 3071, 3193, 3689, 4257] \text{ Hz.} \quad (20)$$

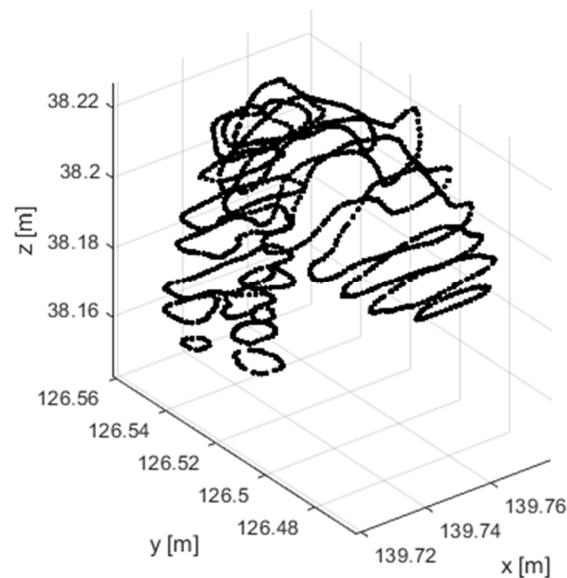


Figure 6. Nodes of a complete model assembled from superelements.

Using super-elements and reduction methods confirmed the applicability for decreasing computation demands. Comparing the size of the original model with the reduced model is not sufficient for evaluating the effectiveness of these procedures. It is also necessary to evaluate the computation demands of the reduction itself and the number of runs of the created model. If the number of runs of the reduced calculation model is high, it can be assumed that this procedure is worthwhile. However, it is necessary to add up the reduction demands for a single run and any subsequent calculations. Before starting the iteration, it is necessary to calculate $\mathbf{K}_{s,s}^{-1}$ and it is necessary to calculate the reduced mass and stiffness matrices (simply by multiplying the matrices) and invert the reduced mass matrix for every iteration. Most of the computing time during the IRS is spent on these inversions. The $\mathbf{K}_{s,s}$ matrix is square-shaped with a dimension corresponding to the number of reduced DoF. On the contrary, the dimension of the reduced mass matrix is the same as the number of preserved DoF. If the number of reduced DoF is relatively low, calculating $\mathbf{K}_{s,s}^{-1}$ is not demanding but repeatedly calculating \mathbf{M}_{red}^{-1} is very demanding. On the contrary, if the DoF is primarily reduced, then calculating $\mathbf{K}_{s,s}^{-1}$ is demanding and calculating \mathbf{M}_{red}^{-1} is not demanding. Generally speaking, a reduction in which less DoF are preserved is even more effective from the viewpoint of the demands of the reduction itself. In this case, approximately 97% of the DoF are reduced, which means that the dimensions of the $\mathbf{K}_{s,s}$ matrix is almost the entire stiffness matrix. If the majority of the DoF are reduced, then the demands correspond to the number of inversions of the stiffness matrix. This calculation is comparable to the demanding calculations of general eigenvalues or eigenfrequencies. The IRS method of creating super-elements is only practical when the resulting model is to be used repeatedly. In addition, implementing IRS reduction makes most modifications of the model impossible.

6. The Designed Isoparametric Element

The human vocal cord has a characteristic geometry, which can be described as a curved canal, with a dominant longitudinal dimension and less dominant transverse dimensions. The cross-section is different from the circular one. The requirements for a new FEM element are that it is suitable for VT discretization given the model's low number of DoF. With regard to the defined requirements, the designed element has a triangular base in the reference state, which extends to the third dimension (Figure 7). The triangular base is based on the shape functions of a higher degree polynomial but most of the nodes are analytically eliminated. The base only has nodes at the vertices and in one wall.

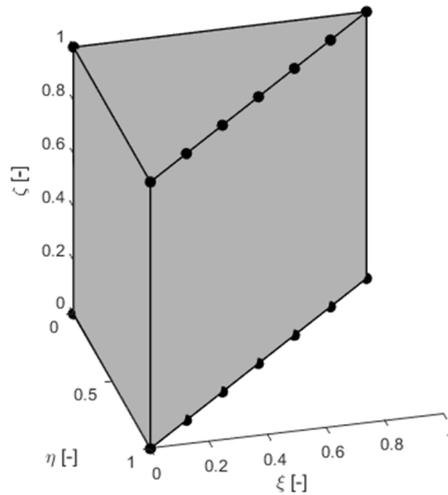


Figure 7. Derived isoparametric element.

The element's topology is chosen in such a way, so that the side, where more nodes are located, forms the surface of the VT and the nodes that are outside of it lie in the centre of the cross-section. The VT is longitudinally divided, with two cross-sections along every part, which is discretized by several elements. Since the outer sides of the surface contain many nodes, it is possible to take into consideration the complex shape of the VT's cross-section. The base functions of the 6th degree and 6 elements were used in every cross-section. Overall, every cross-section is described using 36 nodes on the surface and one central node. The elements in the longitudinal direction are linear. Therefore, it is necessary to cut a sufficient number of cross-sections in the VT geometry.

The derivation of the base functions is based on a 2D triangular 6th degree polynomial element (Figure 8 on the left). The derivation of such a 2D element is described in [19]. Area coordinates are used here (Figure 8 on the right), which are defined as a proportion of the defined area to the area of the element [20]. The first step in the algorithm is assigning the L_1, L_2, L_3 coordinates to all nodes. Parameters α, β, γ for each node are calculated from the node coordinates

$$\alpha = nL_1, \quad \beta = nL_2, \quad \gamma = nL_3, \quad (21)$$

where n is the degree of the element. It is now possible to express an explicit relationship for the node's shape function

$$N_i(L_1, L_2, L_3) = \eta_{\alpha_i}(L_1)\eta_{\beta_i}(L_2)\eta_{\gamma_i}(L_3), \quad (22)$$

where

$$\eta_{\alpha_i}(L_1) = \prod_{i=1}^{\alpha_i} \left(\frac{nL_1 - i + 1}{i} \right) \text{ pro } \alpha_i \geq 1, \quad (23)$$

$$\eta_{\alpha_i}(L_1) = 1 \text{ pro } \alpha_i = 0.$$

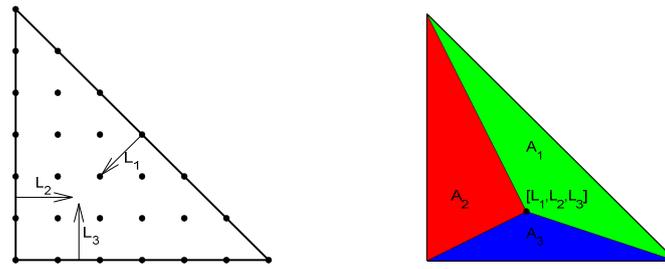


Figure 8. Left: Full planar 6th order element. Right: Definition of surface coordinates.

Functions η_{β_i} and η_{γ_i} are defined analogically to function η_{α_i} [19]. Parameters α, β, γ determine for which node the given shape function is valid. This procedure creates a full-fledged element. The desired element does not contain any nodes. Therefore, the redundant nodes are eliminated. Redundant nodes are eliminated by dividing its shape function into proportions, which are added, in the correct ratio, to the preserved nodes.

Once the redundant nodes are eliminated, the redundant surface coordinates are converted into independent coordinates η, ξ [20]

$$L_1 = 1 - \eta - \xi, \quad L_2 = \eta, \quad L_3 = \xi. \quad (24)$$

The resulting system of the triangular base shape functions of the derived element is as follows

$$N^T(\eta, \xi) = \begin{bmatrix} 1 - \xi - \eta \\ -\frac{\eta}{20}(648\eta^4\xi - 972\eta^3\xi + 558\eta^2\xi - 117\eta\xi + 648\xi^5 - 1620\xi^4 + 1530\xi^3 - 675\xi^2 + 157\xi - 20) \\ \frac{9}{5}\eta\xi(108\eta^4 - 180\eta^3 + 105\eta^2 - 25\eta + 108\xi^4 - 252\xi^3 + 213\xi^2 - 77\xi + 12) \\ -\frac{9}{4}\eta\xi(216\eta^4 - 396\eta^3 + 246\eta^2 - 61\eta + 216\xi^4 - 468\xi^3 + 354\xi^2 - 107\xi + 15) \\ 2\eta\xi(324\eta^4 - 648\eta^3 + 441\eta^2 - 117\eta + 324\xi^4 - 648\xi^3 + 441\xi^2 - 117\xi + 20) \\ -\frac{9}{4}\eta\xi(216\eta^4 - 468\eta^3 + 354\eta^2 - 107\eta + 216\xi^4 - 396\xi^3 + 246\xi^2 - 61\xi + 15) \\ \frac{9}{5}\eta\xi(108\eta^4 - 252\eta^3 + 213\eta^2 - 77\eta + 108\xi^4 - 180\xi^3 + 105\xi^2 - 25\xi + 12) \\ -\frac{\xi}{20}(648\eta^5 - 1620\eta^4 + 1530\eta^3 - 675\eta^2 + 648\eta\xi^4 - 972\eta\xi^3 + 558\eta\xi^2 - 117\eta\xi + 157\eta - 20) \end{bmatrix}. \quad (25)$$

The 2D triangular element is extracted into the spatial dimension by a linear shape function. The planar element exists in the η, ξ plane and its extraction is on the $\zeta \in \langle 0, 1 \rangle$ axis. Two new shape functions are created from the shape function of every node. This results in a final system of 16 shape functions of the element from Figure 7

$$N^T(\eta, \xi, \zeta) = \begin{bmatrix} (1 - \zeta)N^T(\eta, \xi) \\ \zeta N^T(\eta, \xi) \end{bmatrix}. \quad (26)$$

The derived shape functions are valid in the element's volume. The element's volume is defined by the set, V in the space of local coordinates as

$$V = \left\{ \begin{array}{l} \eta \in \mathbb{R} | 0 \leq \eta \leq 1 \\ \xi \in \mathbb{R} | 0 \leq \xi \leq 1 - \eta \\ \zeta \in \mathbb{R} | 0 \leq \zeta \leq 1 \end{array} \right\}. \quad (27)$$

The resulting element is a polynomial of the 6th degree in a planar base and the 1st degree in the third dimension. This allows complex shapes to be modelled by only using 6 elements in every cross-section. In order for the derived element to be correct for FEM, its conditionality has to be checked in all geometric configurations, in which it is used. This was verified by numeric tests, which verified its conditionality on the tested cross-sectional shapes. Conditionality was only tested in 2D. Figure 9

illustrates the conditionality of the elements with a circular cross-section shape. The conditionality ranges at intervals of approximately (1.7, 2.6). Ideally, the conditionality is 1 and infinite conditionality means singularity. When using numbers with a floating decimal point and 8 bytes per number, the accuracy is approximately 15 decimal places. The conditionality is proportional to the increase in the range of the matrix values during inversion and thus decreasing calculation accuracy. The conditionality values in tens are not an obstacle for the calculation.

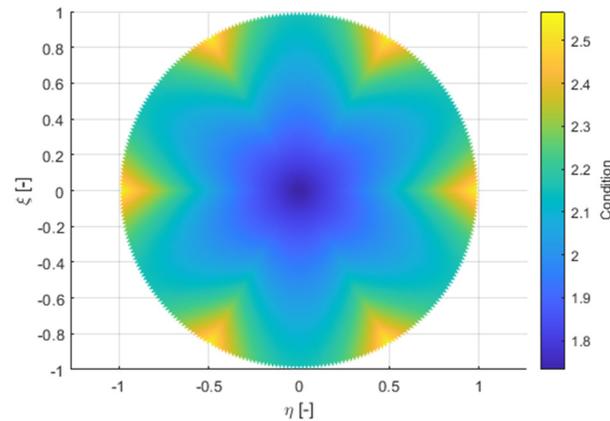


Figure 9. Conditionality of elements in circular geometry.

A human's vocal tract is geometrically more complex than a rotating cylinder (circular cross-section). Therefore, a second conditionality test was conducted. Figure 10 illustrates the test for significantly non-convex geometry. Here, it depends on how the elements are distributed in the geometry. Therefore, 2 variants are shown. The worst conditionality found is illustrated in 10.3.

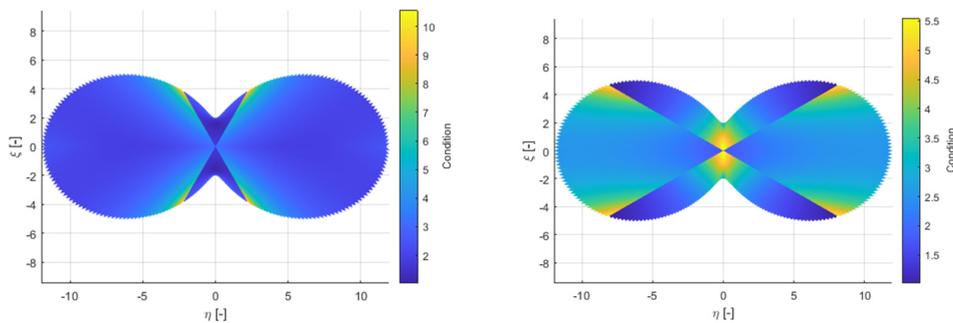


Figure 10. Conditionality of elements in non-convex geometry.

The element mass matrices and stiffness matrices for $c_0 = 1 \text{ ms}^{-1}$ are full. The structure of the mass and stiffness matrices are illustrated in Figure 11. This shows that the larger values of the stiffness matrix are located in the node positions, which lies in the element's wall, creating the VT shell. On the contrary, larger mass matrix values are found in the node positions that do not lie in this wall.

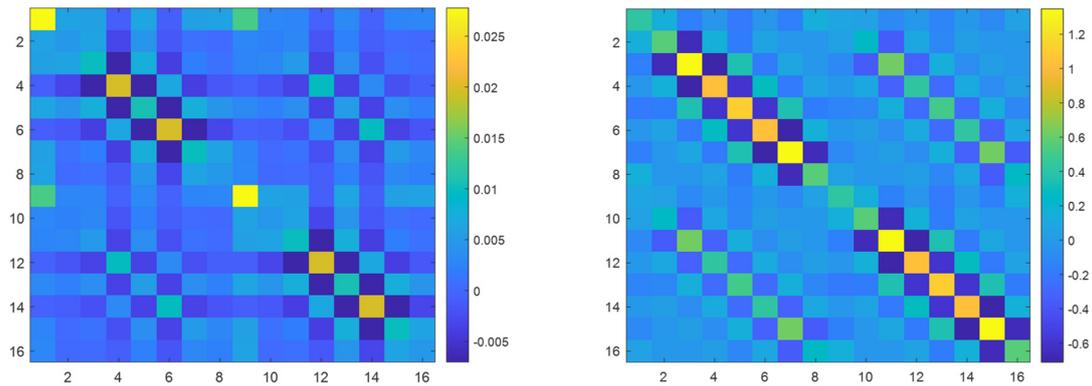


Figure 11. Visualization of the derived element mass (left) and stiffness (right) matrices values.

7. Properties of the Derived Isoparametric Element

Combining six derived elements results in a prismatic volume, with a node in the centre of the bases and 36 nodes along the perimeter of the bases, so 74 nodes in total. This is similar to the super-element amount created by the IRS method. Each of these groups of six elements can fill the VT volume surrounded by two cross-sections.

There is a significant difference between both types of elements. The IRS approach explicitly does not contain internal nodes, but their effect is included in the resulting super-element, while the derived isoparametric element does not contain these nodes. This results in it being able to only affect such oscillation shapes, which can be described using the existing nodes. This means that when only one group of six elements is used, it is not possible to find a higher longitudinal oscillation shape. The same problem arises with the transverse oscillation shapes. When using a sufficient number of layers assembled of a group of six derived elements, the qualitative results are the same as for the full-scale FEM models reduced by the IRS method. An advantage of the super-elements from the IRS method is the ability to find eigenfrequency of the oscillation shape, which cannot be described by the nodes used. In contrast, the model from the derived elements is incomparably numerically more effective. Assembling one group of six derived elements is direct. The shape functions are derived in general and transforming them into the necessary geometry is very easy. Therefore, creating their stiffness and mass matrices is also easy. This is done only 6x for one layer, while creating a super-element represents assembling a full FEM model from ordinary elements contains thousands of DoF. When reducing nodes, it is necessary to invert the matrix by a dimension equal to the number of reduced DoF and then invert the resulting mass matrix in every iteration. If tens of nodes are to exist after the reduction, it is then necessary to invert the matrix by thousands of rows. This is computationally demanding. Overall, assembling a model directly from derived elements is 250x more effective than simply reducing a full model, where the time it takes to assemble a full model is not taken into consideration. Moreover, if the time to assemble a new model were taken into consideration, this ratio would be even greater. This result was calculated from real computing times on a regular computer. Since it is a ratio, the same ratio can be expected on different hardware and on differently optimal codes.

The VT model (Figure 12) for vowel [a:], divided into parts according to Figure 3, can be discretized by the derived isoparametric elements, using sixteen layers of elements. The VT is modelled to include parallel cavities. The piriform sinuses and the vallecula are modelled by six additional elements that are attached to the main cavity from the side. Overall, the model contains 949 nodes. There are 17 planes with nodes in the main VT cavity, which is sufficient for modelling lower longitudinal oscillation shapes, which are the most important for voice creation. The velocity of sound is chosen at 350m/s, the VT walls are assumed to be stiff. The VT is excited by speed pulses of the L-F model. It is not necessary to know the shape of these pulses to find eigenfrequencies because the boundary condition used will be of homogeneous velocity. The boundary condition at the VT output in the mouth is defined as the emitted acoustic impedance.

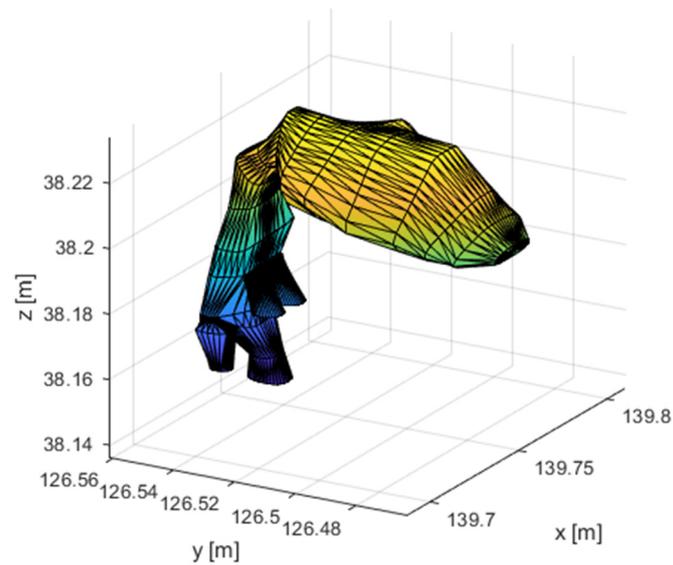
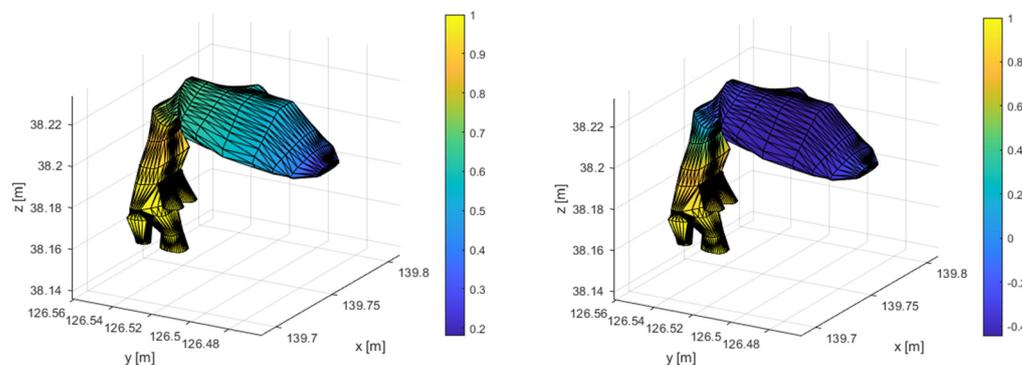


Figure 12. The finite element mesh of the vocal tract for vowel [a:] assembled from derived isoparametric elements.

The model's eigenfrequencies are illustrated in Table 1, along with a description of the respective oscillation shape. The first 4 oscillation shapes are longitudinal. The first transverse oscillation shape is in the widest part of the VT with a frequency of 4261.9Hz. For the human voice, the most significant frequency ranges between 300Hz to 5kHz. There are 5 longitudinal oscillations and one transverse oscillation in this area. The first 4 longitudinal shapes are illustrated in Figure 13 and the first transverse shape is illustrated in Figure 14.

Table 1. Eigenfrequency and Description of the Shape of the Model's Oscillations.

Number of Shape	Frequency [Hz]	Description
1	530.4	1. Longitudinal shape
2	1146.6	2. Longitudinal shape
3	3061.2	3. Longitudinal shape
4	3876.1	4. Longitudinal shape
5	4261.9	1. Transverse shape in the mouth
6	5006.3	5. Longitudinal shape



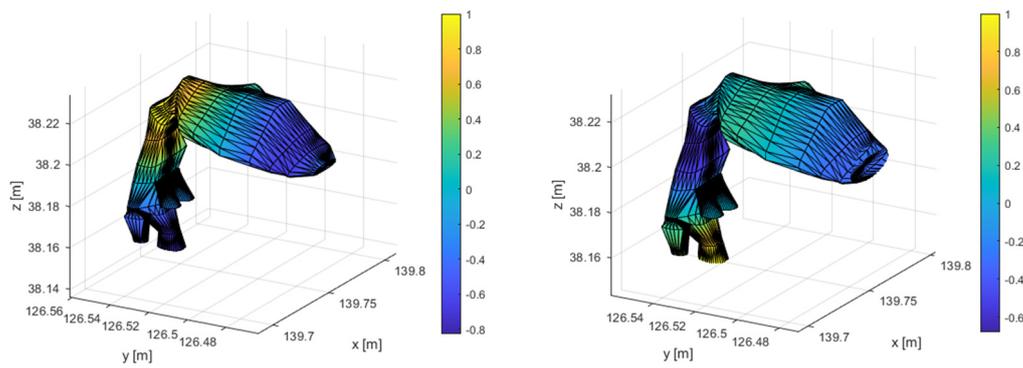


Figure 13. First 4 longitudinal oscillation shapes.

$$F_1 = 530.4 \text{ Hz}, F_2 = 1146.6 \text{ Hz}, F_3 = 3061.2 \text{ Hz}, F_4 = 3876.1 \text{ Hz}$$

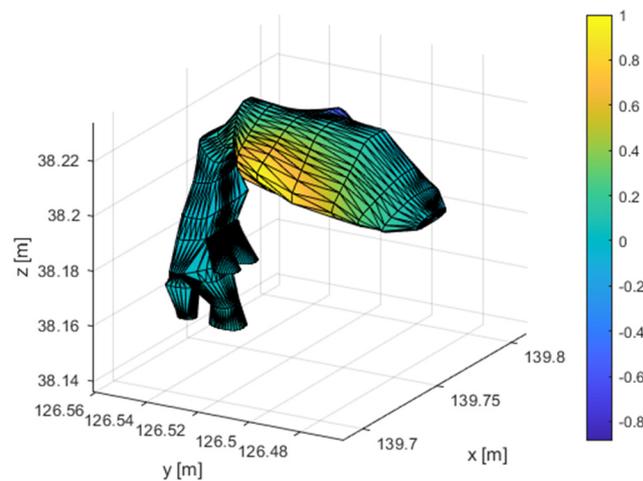


Figure 14. Transverse oscillation shape $F = 4261.9 \text{ Hz}$.

The character of the vowel is given by the formants. These are the most represented frequencies in the voice spectrum. When the VT is excited with speed pulses, the spectrum of the pulses is modulated by the VT transmission characteristic. The frequency peaks represent formants if the excitation is broadband. The transmission characteristic is in Figure 15, which shows the significance of the character of the oscillation shape. The first 6 resonance peaks correspond to the longitudinal oscillation frequencies. There are also 2 transverse oscillations among these two shapes, which are not visible on the transmission characteristics. This is given by the position of their nodes at the excitation and emission points. Therefore, the longitudinal oscillation shapes are important for voice creation.

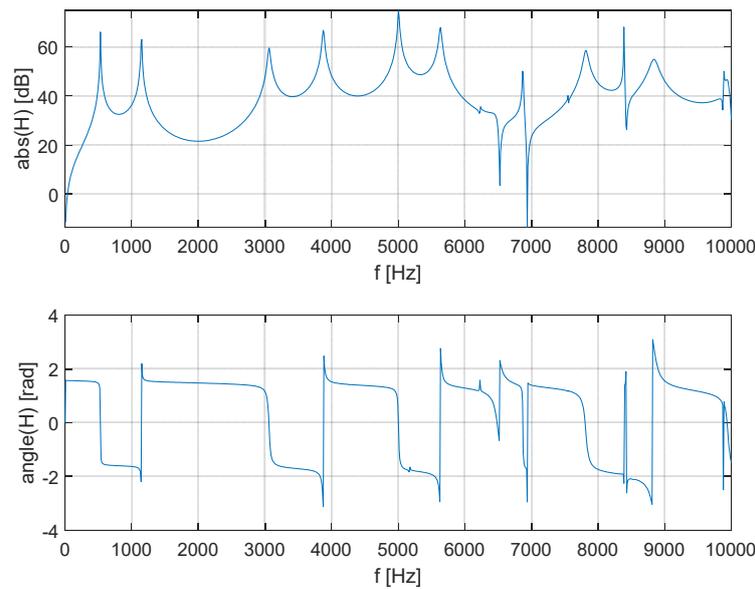


Figure 15. Transmission characteristics between vocal cord velocity and emitted acoustic pressure.

8. Energy Maximalization in the Frequency Band

From the viewpoint of the vocal abilities of professional singers, it's advantageous to be able to amplify certain frequencies of the spectrum when singing. The human ear is most sensitive around 4kHz. If the singing dominantly represents frequencies around 3kHz, then this singing is much clearer and when sung with an orchestra, the orchestra does not drown out the singing. This is called a singing formant and it is primarily important for opera singers.

The singing formant is the subject of research [21], [22]. It was found that it can be achieved by an appropriate ratio of the VT cross-section in the epiglottis and the piriform sinuses cross-section [23]. Therefore, the assembled model was used to verify the hypothesis. The FEM model was parametrized to allow for changes in the cross-section of the main VT cavity in all cross-sections and the piriform sinuses cross-section. The ratio of these cross-sections was changed and the effect on the frequency of the formants was monitored (Figure 16).

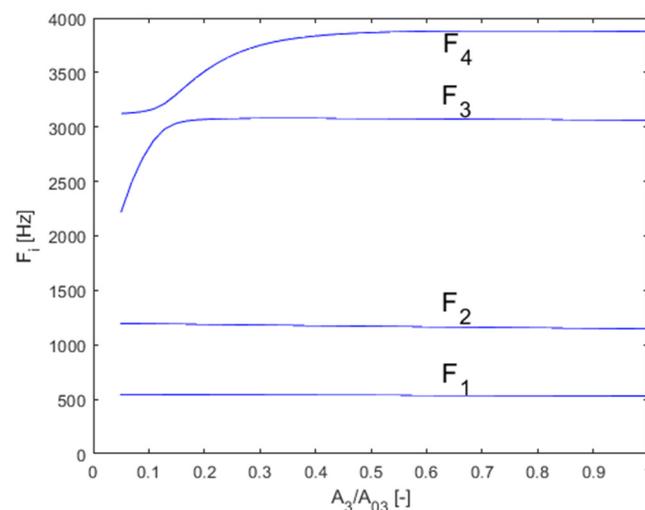


Figure 16. Position of formants depending on the vocal tract cross-section and the piriform sinuses ratio.

The analysis showed that given a suitable cross-section ratio, the third and fourth formants are brought closer together, to a distance of 224Hz from the original 815Hz. Bringing the two resonance peaks closer together results in a higher amplification in the transmission characteristics in Figure 17. This can be interpreted as the singing formant. This double peak is located in the area around 3kHz, which is consistent with observations done on opera singers.

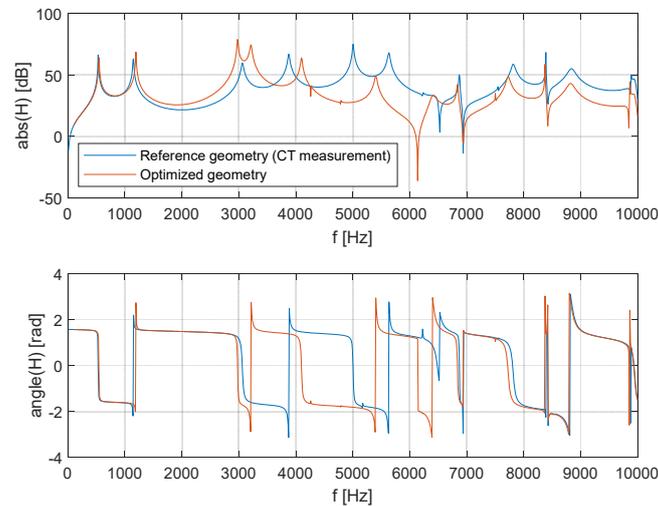


Figure 17. Compare transmission characteristic reference model with optimized model.

Figure 17 compares the frequency characteristics of the model modified for creating the singing formant with the original geometry model. Here, it can be seen that the geometric modification neither affected the area below 2 kHz nor the area above 7 kHz. Since it did not affect the frequency of the first 2 formants, the phonated sound did not change. The maximum amplification in the area of the singing formant (Figure 18) increased from the original 60 dB to 79 dB, which corresponds to a ten-fold amplification. The average amplification in the 2750-3250 Hz band increased from 43 dB by 17 dB to 60 dB, which also corresponds to about a ten-fold increase in acoustic pressure.

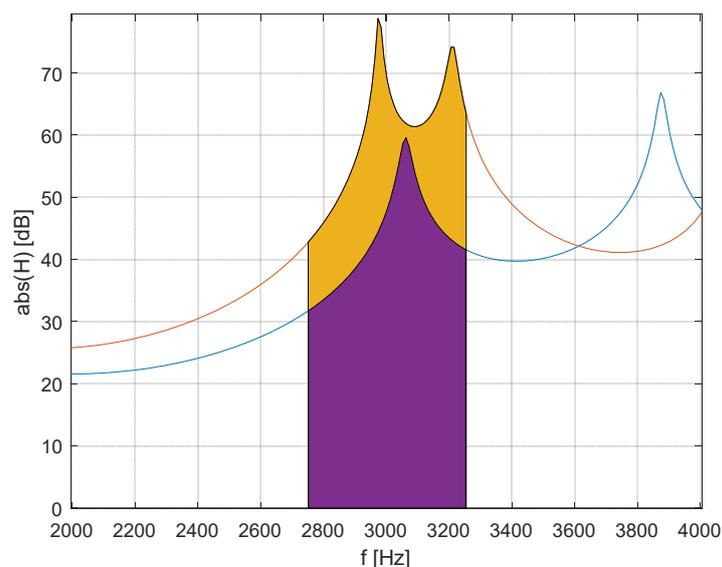


Figure 18. Area of the singing formant in the transmission characteristic.

To evaluate the amount of acoustic energy in the monitored frequency band, it is necessary to excite the transmission function. Setting up the excitation of the L-F model was selected by basic

frequency pulses of 100Hz, $t_e = 0.006s$, $t_a = 0.00023s$, $t_p = 0.0045s$ and $E_e = 0.4 m^3s^{-2}$. Acoustic energy is defined by the following relationship [15]

$$w = \frac{1}{2} \rho_0 |\mathbf{v}'(t)|^2 + \frac{1}{2} \frac{p'(t)^2}{\rho_0 c_0^2}. \quad (28)$$

Acoustic energy is made up of a potential (pressure) and a kinetic part, which changes over time. Only the pressure part of the energy is used to evaluate acoustic energy. After converting it to the frequency domain, acoustic energy is defined as follows

$$w(\omega) = \frac{1}{2} \frac{p'(\omega) \tilde{p}'(\omega)}{\rho_0 c_0^2}. \quad (29)$$

Exciting VT pulses generated by the vocal cords are periodic, they contain a fundamental frequency and higher harmonic components. Their spectrum is not continuous. Acoustic energy is then simply a sum of the frequency components. A comparison on the singing formant shows a 111x higher amount of energy in the monitored frequency band.

9. Conclusion

The new isoparametric element was derived for the direct assembly of FEM VT models. In parallel, a spatial FEM VT model was assembled from ordinary elements. Super-elements were created from it using the IRS method. The created VT models were compared. The aim of these procedures is to reduce the computation demands of the model. The VT geometry for vowel [a:] was obtained from CT measurements and it is very detailed. All models even include piriform sinuses and epiglottic vallecula. All of the models used achieve satisfactory compliance in terms of the results of the formant position of [a:]. However, the calculation demands vary significantly.

The FEM model created from ordinary elements has 36553 DoF and 186570 number of elements. Assembling the model requires calculating the stiffness and mass matrices for each element. This is done by numerically integrating the shape functions over a selected number of integration points. The solution then lies in analysing the resulting matrices.

The analyses of the resulting matrices can be more effective by reducing the matrices using reduction methods. In such a case, it is necessary to first assemble a full FEM model and reduce it. If the degree of reduction is high, then the largest number of demands used on the inversion of the stiffness sub-matrix takes place during the IRS procedure. In terms of size, it is close to a full matrix. Even though the IRS method reduces the DoF from 36553 to 1154, thereby decreasing the calculation demands of the formants to negligible, this saving is negatively compensated by calculating the inversion stiffness sub-matrix. As a result, this procedure is not suitable for a one-time analysis because the savings correspond to the demands for the reduction itself.

The last option used is applying a new element for FEM. The obvious advantage of this procedure is that it does not require assembling the model from ordinary elements and a calculation-efficient model is created directly, not by simplifying an extensive model. This model contains 132 elements and a 7th accuracy order integration scheme is used (4 layers with 13 integration points in a triangular plane). The total number of integration points is 6864, which is more than 100x less compared to the assembled ordinary model. The total number of DoF is 949, which represents significant savings even in assembling the global matrix since there are lower demands on the computer's memory. When comparing the demands at the computation time level, the model assembled from the new isoparametric elements is approximately 250x faster than the simple IRS reduction. If the time needed to assemble the full FEM model were added, the ratio would be even greater.

This work compares two approaches for creating effective FEM VT models. It turns out that the generally known IRS method is less beneficial than using the new element described here. There is a considerable difference in efficiency. The model with the new element is an order of magnitude faster. One disadvantage of the new element is its design for the VT geometry. In the acoustic model, which does not geometrically resemble the VT, this element does not offer any advantages. In contrast the

IRS method is applicable everywhere. The designed isoparametric element was tested by practical calculation of the singing formant. The calculation confirmed generally valid assumptions for its creation.

The model using the derived isoparametric element was used for a sensitivity study aimed at finding the optimal geometric configuration for creating a singing formant. It has been confirmed that a suitable ratio between piriform sinuses cross-section and the laryngeal part of the VT will bring the 3rd and 4th formant closer together in the area of 3kHz. This singing formant provides a 17dB higher amplification than the original model with the geometry obtained from CT measurement. The amount of acoustic energy is 111x greater than the original model when the LF model is excited by the model.

Author Contributions: Conceptualization, T.V.; methodology, J.Š.; software, J.Š.; validation, J.Š., and T.V.; formal analysis, J.Š.; investigation, T.V.; resources, J.Š.; data curation, J.Š.; writing—original draft preparation, J.Š.; writing—review and editing, T.V.; visualization, J.Š.; supervision, T.V.; project administration, T.V.; funding acquisition, T.V. All authors have read and agreed to the published version of the manuscript.

Funding: The research has been supported by the grant CTU in Prague No.930-122-5201X000 Optimizing the acoustic output with minimal possible load on the vocal cords due to changes in the human vocal tract using computer physical modeling.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to reason non-existent infrastructure for their distribution.

Acknowledgments: The collective of authors would like to thank their colleagues for a good working environment and cooperation in the translation of this article into English.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Johnson-Read, L.; Chmiel, A.; Schubert, E.; Wolfe, J. Performing Lieder: Expert Perspectives and Comparison of Vibrato and Singer's Formant With Opera Singers, *Journal of Voice*, Volume 29, Issue 5, 2015, Pages 645.e15-645.e32, ISSN 0892-1997, <https://doi.org/10.1016/j.jvoice.2014.10.020>
2. Sundberg, J. Level and Center Frequency of the Singer's Formant, *Journal of Voice*, Volume 15, Issue 2, 2001, Pages 176-186, ISSN 0892-1997, [https://doi.org/10.1016/S0892-1997\(01\)00019-4](https://doi.org/10.1016/S0892-1997(01)00019-4)
3. Arnela, M.; Dabbaghchian, S.; Blandin, R.; Guasch, O.; Engwall, O., et al., Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds, *The Journal of the Acoustical Society of America*, 2016, 140. pp. 1707-1718. 10.1121/1.4962488
4. Motoki, K. Three-dimensional acoustic field in vocal-tract, *Acoustical Science and Technology*, 2002, 20. 10.1250/ast.23.207
5. Ribeiro, V.; Isaieva, K.; Leclere, J.; Vuissoz, P.; Laprie, Y. Automatic generation of the complete vocal tract shape from the sequence of phonemes to be articulated, *Speech Communication*, Vol. 141, 2022, pp. 1-13, ISSN 0167-6393, <https://doi.org/10.1016/j.specom.2022.04.004>
6. Dabbaghchian, S.; Arnela, M.; Engwall, O.; Guasch, O. Reconstruction of Vocal Tract Geometries from Biomechanical Simulations, *International Journal for Numerical Methods in Biomedical Engineering*, 2018, 10.1002/cnm.3159
7. Kagawa, Y.; Ohtani, Y.; Shimoyama, R. Vocal tract shape identification from formant frequency spectra—a simulation using three-dimensional boundary element models, *Journal of Sound and Vibration*, Vol. 203, Issue 4, 1997, pp. 581-596, ISSN 0022-460X
8. Arnela, M.; Ureña, D. Tuned two-dimensional vocal tracts with piriform fossae for the finite element simulation of vowels, *Journal of Sound and Vibration*, Vol. 537, 2022, ISSN 0022-460X, doi: 10.1016/117168
9. Vampola, T.; Horáček, J.; Švec, J. G. FE Modeling of Human Vocal Tract Acoustics. Part I: Production of Czech Vowels, *Acta Acustica united with Acustica* Vol. 94, 2008, pp. 433-447
10. Vampola, T.; Horáček, J.; Švec, J. G. Modeling the Influence of Piriform Sinuses and Valleculae on the Vocal Tract Resonances and Antiresonances, *Acta Acustica united with Acustica* Vol. 101, No. 3, 2015, pp. 594-602

11. Vampola, T.; Laukkanen, A. M.; Horáček, J.; Švec, J. G. Vocal tract changes caused by phonation into a tube: a case study using computer tomography and finite-element modeling, *The Journal of the Acoustical Society of America*, 2011, Vol. 129 1, pp. 310-5
12. Vampola, T.; Laukkanen, A. M.; Horáček, J.; Švec, J. G. Finite element modelling of vocal tract changes after voice therapy, *Applied and Computational Mechanics* 5, 2011, pp. 77–88
13. Johnson, C. *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Cambridge University Press, Cambridge, 1987
14. Fant, G.; Liljencrants, J.; Lin, Q. A four-parameter model of glottal flow, *STL-QPSR*, vol. 26, issue 4, 1985, pp. 1-13
15. Pierce, S. *Acoustic An introduction to Its Physical Principles and Applications*, Springer, 1981, ISBN 978-3-030-11214-1
16. Guyan, R. J. Reduction of stiffness and mass matrices, *American Institute of Aeronautics and Astronautics Journal* 3(2), 380, 1965
17. Blair, M. A.; Camino, T. S.; Dickens, J. M. An iterative approach to a reduced mass matrix, *Proceedings of the 8th International Modal Analysis Conference*, Florence Italy, April 1991, 621-626
18. Friswell, M. I.; Garvey, S. D.; Penny, J. E. T. Model reduction using dynamics and iterated IRS techniques, *Journal of Sound and Vibration*, 1995, 186(2), pp. 311-323
19. Silvester, P. High-order polynomial triangular finite elements for potential problems, *International Journal of Engineering Science*, Volume 7, Issue 8, 1969, Pages 849-861, ISSN 0020-7225
20. Bathe, K. J. *Finite Elements Procedures*, Prentice Hall Inc., Upper Saddle River, New Jersey, 1996, ISBN 0-13-301458-4
21. Ikävalko, T.; Laukkanen, A. M.; Mcallister, A., et al., Three Professional Singers' Vocal Tract Dimensions in Operatic Singing, Kulning, and Edge—A Multiple Case Study Examining Loud Singing, *Journal of Voice*, 2022, 10.1016/j.jvoice.2022.01.024
22. Story, B. The Vocal Tract in Singing, *The Oxford Handbook of Singing*, Oxford Library of Psychology, 2016, <https://doi.org/10.1093/oxfordhb/9780199660773.013.012>
23. Yanagisawa, E.; Estill, J.; Kmucha, S.; Leder, S. The contribution of aryepiglottic constriction to “ringing” voice quality—A videolaryngoscopic study with acoustic analysis, *Journal of Voice*, Volume 3, Issue 4, 1989, P. 342-350, ISSN 0892-1997

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.